Technische Universität München
Lehrstuhl für Theoretische Informationstechnik

# Multiple Access Channels with Cooperating Encoders

**Moritz Wiese**

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

**Doktors der Naturwissenschaften**

genehmigten Dissertation.

| | |
|---|---|
| Vorsitzender: | Univ.-Prof. Dr. sc. techn. Gerhard Kramer |
| Prüfer der Dissertation: | |
| | 1. Univ.-Prof. Dr.-Ing. Dr. rer. nat. Holger Boche |
| | 2. Univ.-Prof. Dr. rer. nat. Michael Marc Wolf |

Die Dissertation wurde am 17.01.2013 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 19.06.2013 angenommen.

# Multiple Access Channels with Cooperating Encoders

Moritz Wiese

# Acknowledgment

This PhD thesis was written during my time in Berlin at the Heinrich-Hertz-Lehrstuhl für Informationstheorie und theoretische Informationstechnik of TU Berlin and then, starting from October 2010, in Munich at the Lehrstuhl für Theoretische Informationstechnik of TU München. During the whole period, my supervisor was Holger Boche. I would like to thank him very much for giving me the opportunity to work with him and for all his advice and motivation. I also have to thank my colleagues, working and discussing with them was fun and a constant source of inspiration. Off work, they even taught me skiing, but I haven't yet advanced to the fun part of that.

I am also grateful to Prof. Michael Wolf, who has agreed to be the second referee of my thesis. Finally, I owe my gratitude to Prof. Gerhard Kramer for serving as the chairman of the dissertation committee.

# Contents

# 1. Introduction

In 1982/83, Willems [63, 64] introduced an extension of standard Multiple Access Channel (MAC) codes. The resulting channel is called the MAC with conferencing encoders. In this model, the encoders want to transmit one message each. However, instead of being completely ignorant of the other encoder's message as in standard MACs, each encoder has partial knowledge about the message the other encoder would like to transmit. This partial knowledge is generated by a "conference", i.e. a protocol to iteratively exchange information noiselessly subject to a rate constraint. Only the general structure of conferencing and the rate constraints are part of the model, the actual protocol is part of the code. One of the main questions is how the capacity regions obtained with different conferencing capacities compare.

In this work, we extend both the discrete memoryless MAC with common message and the discrete memoryless MAC with conferencing encoders to more general channels models: the compound MAC, the Arbitrarily Varying MAC (AV-MAC), and the wiretap MAC. We do not consider the AV-MAC with common message as this would lead us too far away from conferencing. The main reason for considering the various MACs with common message is that in many cases they are the basis for achievability results for the MAC with conferencing encoders. Intuitively, the information exchanged between the encoders during a conference form a common message, so the results known for the corresponding non-cooperative channel with common message can be applied to find the achievable rates of the conferencing setting. This transition has been used in [63, 64, 15, 62], and [41]. The AV-MAC with conferencing encoders can however be treated based on the compound MAC with conferencing encoders.

The practical relevance of MACs with conferencing encoders lies in the connection to base station cooperation in wireless networks. This has been considered for future network standards such as LTE-Advanced. The main goals of base station cooperation are interference mitigation, improving the spectral efficiency of mobile networks, enhancing the performance of cell-edge users, and resolving fairness issues more easily. In standardization oriented literature, the assumptions on models incorporating base station cooperation generally are very strict. The cooperation backbones, i.e. the wires linking the base stations, are assumed to have infinite capacity. Full Channel State Information (CSI) is assumed to be present at all cooperating base stations. Then, Multiple-Input-Multiple-Output (MIMO) optimization techniques can be used for designing the system [34]. However, while providing a useful theoretical benchmark, the results thus obtained are not accepted by the operators as reliably predicting the performance of actual networks.

There have been more realistic practical studies of base station cooperation. In [44], the cooperation of base stations in an uplink network is analyzed. A turbo-like decoding

scheme is proposed. Different degrees of cooperation and different cooperation topologies are compared in numerical simulations. In [33], the implementation of a real-time distributed cooperative system for the downlink of LTE-Advanced was presented. CSI at the transmitters was assumed imperfect, the limited-capacity glass fibers between the transmitting base stations were used to exchange CSI and data information. A feeder distributed the data among the transmitting base stations.

In order to obtain a more realistic theoretical assessment of the performance of cellular networks with base station cooperation, the optimistic assumptions of infinite cooperation capacity and perfect CSI need to be adapted to reality. First, it is well-known that one cannot really assume perfect CSI in mobile communication networks. Second, glass fibers or any medium used for the backbones never have infinite capacity. The assumption of finite cooperation capacity will also lead to a better understanding of the amount of cooperation necessary to achieve a certain performance. Vice versa, one would like to know which capacity can be achieved with the backhaul found in heterogeneous networks using microwave, optical fibers and other media. Such insights would get lost when assuming infinite cooperation capacity.

The question arises how much cooperation is needed in order to achieve the same performance as would be achievable with infinite cooperation capacity. For general interference networks with multiple receivers, the analysis is very difficult. Thus it is natural to start by taking a closer look at component networks which together form a complete interference network. Among these components are the subnetworks formed by a subset of the base stations and with only one receiving mobile. Then there is no more interference, so one can concentrate on finding out by how much the capacity increases by limited base station cooperation. This result can be seen as a first step towards a complete rigorous analysis of general interference networks.

A MAC with conferencing encoders models a very simple component network of a wireless network with cooperating base stations. The two senders of the MAC are interpreted as base stations, the conferencing capacities depend on the physical properties of the real cooperation backbone. The MAC's receiver is a mobile terminal. This reduction allows for precise results which can not yet be obtained when the whole network is considered.

Parts of this work have been published or are about to be published in [58, 55, 56, 57, 59, 60, 61]. There are some differences between these publications and the thesis. Mostly they just concern notation, but in the compound and arbitrarily varying cases (Chapters 3-5), the content has also changed partially. This is due to new insights gained during the reviewing process through reviewers' comments and afterwards.

## Notation

For sets $\{1, \ldots, M\}$, where $M$ is a positive integer, we use the combinatorial shorthand $[M]$. For a real number $x$ we define $[x]_+ := \max\{x, 0\}$, $\lfloor x \rfloor$ is the largest integer $m$ with $m \leq x$. For a subset $A$ of a topological space, we denote its topological closure by closure$(A)$. The convex hull of a set $A$ is denoted by conv$(A)$. The logarithm denoted

by log is to base 2. Analogously, by writing $\exp(x)$ we mean $2^x$. The natural logarithm to base $e$ is denoted by ln.

For any set $\mathscr{X}$ and subset $A \subset \mathscr{X}$ we write $A^c := \mathscr{X} \setminus A$. For elements of $\mathscr{X}^n$ we write $\mathbf{x}$ and implicitly understand that $\mathbf{x} = (x_1, \ldots, x_n)$. We let $1_A : \mathscr{X} \to \{0, 1\}$ be the indicator function of $A$ which takes on the value 1 at $x \in \mathscr{X}$ if and only if $x \in A$. For a set $A \subset \mathscr{X} \times \mathscr{Y}$, we write $A_{|y} \subset \mathscr{X}$ for the set defined by $A_{|y} := \{x \in \mathscr{X} : (x, y) \in A\}$. Given a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ we write $\mathbb{E}$ for the expectation corresponding to $\mathbb{P}$ and for $A \in \mathscr{A}$ and a real-valued random variable $X$ we write $\mathbb{E}[X; A] := E[X1_A]$.

The space of probability distributions on the finite set $\mathscr{X}$ is denoted by $\mathscr{P}(\mathscr{X})$. In particular, it contains for every $x \in \mathscr{X}$ the *Dirac measure* $\delta_x$ defined by $\delta_x(x) = 1$. The product of two probability distributions $P$ and $Q$ is denoted by $P \otimes Q$. The $n$-fold product of $P$ with itself is called $P^{\otimes n}$. A stochastic matrix with input alphabet $\mathscr{X}$ and output alphabet $\mathscr{T}$ is written as a mapping $W : \mathscr{X} \to \mathscr{P}(\mathscr{T})$. The $n$-fold memoryless extension of a channel $W : \mathscr{X} \to \mathscr{P}(\mathscr{T})$ is denoted by $W^{\otimes n}$, so that for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathscr{X}^n$ and $\mathbf{t} = (t_1, \ldots, t_n) \in \mathscr{T}^n$,

$$W^{\otimes n}(\mathbf{t}|\mathbf{x}) = \prod_{i=1}^{n} W(t_i|x_i).$$

We also define for $P \in \mathscr{P}(\mathscr{X})$ and $W : \mathscr{X} \to \mathscr{P}(\mathscr{T})$ the probability distribution $P \otimes W \in \mathscr{P}(\mathscr{X} \times \mathscr{T})$ by $(P \otimes W)(x, t) = P(x)W(t|x)$.

On the set of measures on $\mathscr{X}$, we define the *total variation distance* by

$$\|\vartheta_1 - \vartheta_2\| := \sum_{x \in \mathscr{X}} |\vartheta_1(x) - \vartheta_2(x)|.$$

The support $\operatorname{supp}(\vartheta)$ of a measure $\vartheta$ on $\mathscr{X}$ is defined as the set of those $x$ with $\vartheta(x) \neq 0$.

Given a random variable $X$ living on $\mathscr{X}$ and a $P \in \mathscr{P}(\mathscr{X})$, we mean by $X \sim P$ that $P$ is the distribution of $X$. Given a pair of random variables $(X, Y)$ taking values in the finite Cartesian product $\mathscr{X} \times \mathscr{Y}$, we write $P_X \in \mathscr{P}(\mathscr{X})$ for the distribution of $X$ and $P_{X|Y}$ for the conditional distribution of $X$ given $Y$. The support of $X$, denoted by $\operatorname{supp}(X)$, is the support of $P_X$.

For random variables $X, Y, Z$ we write $H(X)$ for the entropy of $X$, $H(X|Y)$ for the conditional entropy of $X$ given $Y$, $I(X \wedge Y)$ for the mutual information of $X$ and $Y$ and $I(X \wedge Y|Z)$ for the conditional mutual information of $X$ and $Y$ given $Z$. We write the mutual information between $X$ and $Y$ conditioned on the event that $Z = z$ as $I(X \wedge Y|z)$.

# 2. Preliminaries

In this chapter we recall some results from the literature and add a couple of remarks. The largest part of the chapter is concerned with the discrete memoryless Multiple-Access Channel (MAC) with common message and with conferencing encoders.

First we recall the coding theorem of the discrete memoryless MAC with common message. The capacity regions of the traditional classical discrete memoryless MACs, both where two senders have one message each and where they have an additional common message, have been characterized in [2, 38] and [50], respectively. There is a vast literature on generalizations in all kinds of directions. Apart from conferencing, which will be treated extensively in this work, Willems [63] also considered various feedback models for the discrete memoryless MAC as well as "cribbing" encoders. The capacity of the Gaussian MAC was found in [66].

Next we consider discrete MACs without a common message whose encoders may exchange some information in an iterative manner. This concept was introduced by Willems [63, 64], so we call it a *Willems conference*. Its relevance was not recognized until some years ago, so most of the literature is fairly recent. Gaussian MACs using Willems conferencing between the encoders were analyzed in [15] and [62]. As the traditional way of proving results for conferencing encoders is to reduce these to a situation with common message, the two aforementioned works also provide the corresponding results where the encoders do not cooperate, but have a common message. A variant of unidirectional cooperation was investigated in [48], where the three encoders of a Gaussian MAC can cooperate over a ring of unidirectional links. However, only lower and upper bounds which are not tight were found for the maximum achievable equal rate.

Further literature exists for Willems conferencing on the decoding side of a multi-user network. For degraded discrete memoryless broadcast channels, the capacity region was found in [22] if the receivers can exchange information about the received codewords in a single conference step. For general broadcast and multicast channels, achievable regions were determined. For the Gaussian relay channel, the dependence of the performance on the number of conferencing iterations between the receiver and the relay was investigated in [45]. For the Gaussian $Z$-interference channel, outer and inner bounds to the capacity region where the decoders can exchange information about the channel outputs are provided in [24]. Finally, for discrete and Gaussian memoryless interference channels with conferencing decoders, [47] determines achievable regions. Exact capacity regions are determined if the channel is physically degraded.

In the last part of the chapter, we collect some facts about types and typical sequences which we will use frequently. Most of them are well-known, properties that are used less are mentioned with proof. We also state two important lemmas of information theory: the fact that entropy is uniformly continuous with respect to total variation distance,

and Fano's lemma, the standard lemma applied in proofs of weak converses.

## 2.1. The Discrete Memoryless MAC with Common Message

**Definition 2.1.** Let $\mathscr{X}, \mathscr{Y}, \mathscr{T}$ be finite sets. A *Multiple-Access Channel (MAC)* with alphabets $\mathscr{X}, \mathscr{Y}$ and $\mathscr{T}$ is a set of stochastic matrices

$$W_s^n : \mathscr{X}^n \times \mathscr{Y}^n \to \mathscr{P}(\mathscr{T}^n), \qquad s \in \mathscr{S}_n,\ n = 1, 2, \ldots,$$

where $\mathscr{S}_n$ may be any set.

In order to enable the reliable transmission of message over MACs, one employs codes. Here we assume that there is one sender for each input alphabet $\mathscr{X}$ and $\mathscr{Y}$. Each of these senders has a private message and together they have a common message to send to a receiver with alphabet $\mathscr{T}$. A code$_{\mathrm{CM}}$ as defined in the following definition is independent of the family $\{W_s^n\}$, it only depends on the input- and output alphabets.

**Definition 2.2.** Let $n$ be a positive integer. A *deterministic $n$-code*$_{\mathrm{CM}}$ with alphabets $\mathscr{X}, \mathscr{Y}$ and $\mathscr{T}$ is a triple of mappings

$$f_1 : [K_0] \times [K_1] \to \mathscr{X}^n, \quad f_2 : [K_0] \times [K_2] \to \mathscr{Y}^n, \quad \varphi : \mathscr{T}^n \to [K_0] \times [K_1] \times [K_2],$$

where $K_0, K_1, K_2$ are arbitrary positive integers. $f_1, f_2$ are the *encoding functions* and $\varphi$ is the *decoding function*. The triple $(K_0, K_1, K_2)$ is called the *codelength triple* and $n$ is called the *blocklength* of $(f_1, f_2, \varphi)$.

We denote the set of deterministic $n$-codes$_{\mathrm{CM}}$ with codelength triple $(K_0, K_1, K_2)$ by $\Gamma_{\mathrm{CM}}(n, K_0, K_1, K_2)$.

The codes$_{\mathrm{CM}}$ will usually be applied in conjunction with a given channel, so the code alphabets will be clear and do not have to be mentioned. Every $(k_0, k_1, k_2) \in [K_0] \times [K_1] \times [K_2]$ is called a *message triple*. For an $n$-code$_{\mathrm{CM}}$ $\gamma = (f_1^\gamma, f_2^\gamma, \varphi^\gamma)$, we call the values $f_1^\gamma(k_0, k_1) =: \mathbf{x}_{k_0k_1}(\gamma)$ and $f_2^\gamma(k_0, k_2) =: \mathbf{y}_{k_0k_2}(\gamma)$ *codewords*. Every message triple $(k_0, k_1, k_2)$ gives rise to a *decoding set* $D_{k_0k_1k_2}(\gamma) := (\varphi^\gamma)^{-1}(k_0, k_1, k_2) \subset \mathscr{T}^n$.

The first class of MACs we consider are discrete memoryless channels. Every such channel is determined by a single stochastic matrix

$$W : \mathscr{X} \times \mathscr{Y} \to \mathscr{P}(\mathscr{T}), \tag{2.1}$$

where $\mathscr{X}, \mathscr{Y}, \mathscr{T}$ are finite alphabets.

**Definition 2.3.** Let $W$ be as in (2.1). The MAC

$$W^{\otimes n} : \mathscr{X}^n \times \mathscr{Y}^n \to \mathscr{P}(\mathscr{T}^n), \qquad n = 1, 2 \ldots$$

is called the *discrete memoryless MAC* DMAC($W$).

There are two standard ways of measuring the reliability of a deterministic code$_{\mathrm{CM}}$ when applied for the transmission over DMAC($W$).

**Definition 2.4.** Let $W : \mathscr{X} \times \mathscr{Y} \to \mathscr{P}(\mathscr{T})$ and let $\gamma$ be a deterministic $n$-code$_{\mathrm{CM}}$. Its *DM-average error* is defined as

$$\bar{e}^{\mathrm{DM}}(\gamma, W) := \frac{1}{K_0 K_1 K_2} \sum_{k_0,k_1,k_2} W^{\otimes n}\big(D_{k_0k_1k_2}(\gamma)^c | \mathbf{x}_{k_0k_1}(\gamma), \mathbf{y}_{k_0k_2}(\gamma)\big). \tag{2.2}$$

Its *DM-maximal error* is defined by

$$e^{\mathrm{DM}}(\gamma, W) := \max_{k_0,k_1,k_2} W^{\otimes n}\big(D_{k_0k_1k_2}(\gamma)^c | \mathbf{x}_{k_0k_1}(\gamma), \mathbf{y}_{k_0k_2}(\gamma)\big). \tag{2.3}$$

**Definition 2.5.** A triple $(R_0, R_1, R_2)$ of nonnegative real numbers is called a *deterministically CM-achievable rate triple for* DMAC$(W)$ *under the average (maximal) error criterion* if for every $\lambda \in (0,1)$ and $\varepsilon > 0$ and $n \geq n_0(\lambda, \varepsilon)$ there is a deterministic $n$-code$_{\mathrm{CM}}$ $\gamma$ with $\bar{e}^{\mathrm{DM}}(\gamma, W) \leq \lambda$ ($e^{\mathrm{DM}}(\gamma, W) \leq \lambda$) and

$$\frac{1}{n} \log K_\nu \geq R_\nu - \varepsilon \qquad (\nu = 0, 1, 2). \tag{2.4}$$

The set of deterministically CM-achievable rate triples is called the *deterministic CM-capacity region of* DMAC$(W)$ *under the average (maximal) error criterion* and is denoted by $\overline{\mathscr{C}}^{\mathrm{DM}}_{\mathrm{CM}}(W)$ ($\mathscr{C}^{\mathrm{DM}}_{\mathrm{CM}}(W)$).

The definition immediately implies the closedness of both capacity regions. There are different kinds of outer bounds on a given capacity region. The most prominent among these are the *weak* and the *strong converse*.

**Definition 2.6.** Let $\|\cdot\|$ be any norm on $\mathbb{R}^3$. Let $\mathscr{C} \in \{\overline{\mathscr{C}}^{\mathrm{DM}}_{\mathrm{CM}}(W), \mathscr{C}^{\mathrm{DM}}_{\mathrm{CM}}(W)\}$.

1) There exists a *weak converse* for $\mathscr{C}$ if for every $\varepsilon > 0$ there is a $\lambda(\varepsilon) > 0$ such that every $n$-code$_{\mathrm{CM}}$ $\gamma \in \Gamma_{\mathrm{CM}}(n, K_0, K_1, K_2)$ with

$$\left\| \frac{1}{n}(\log K_0, \log K_1, \log K_2) - \mathscr{C} \right\| > \varepsilon \tag{2.5}$$

and sufficiently large blocklength satisfies $\bar{e}^{\mathrm{DM}}(\gamma, W) \geq \lambda(\varepsilon)$ or $e^{\mathrm{DM}}(\gamma, W) \geq \lambda(\varepsilon)$ depending on whether $\mathscr{C} = \overline{\mathscr{C}}^{\mathrm{DM}}_{\mathrm{CM}}(W)$ or $\mathscr{C} = \mathscr{C}^{\mathrm{DM}}_{\mathrm{CM}}(W)$.

2) There exists a *strong converse* for $\mathscr{C}$ if for every $\lambda \in (0,1)$, every $n$-code$_{\mathrm{CM}}$ $\gamma$ satisfying (2.5) and with sufficiently large $n$ has $\bar{e}^{\mathrm{DM}}(\gamma, W) \geq \lambda$ or $e^{\mathrm{DM}}(\gamma, W) \geq \lambda$ depending on whether $\mathscr{C} = \overline{\mathscr{C}}^{\mathrm{DM}}_{\mathrm{CM}}(W)$ or $\mathscr{C} = \mathscr{C}^{\mathrm{DM}}_{\mathrm{CM}}(W)$.

The difference between these concepts is that the weak converse does not rule out the possibility that there are rate triples outside the capacity region that can be achieved with small, but not arbitrarily small error.

**Definition 2.7.** Let $W : \mathscr{X} \times \mathscr{Y} \to \mathscr{T}$. We set

$$\Pi(W) := \{p \in \mathscr{P}(\mathscr{U} \times \mathscr{X} \times \mathscr{Y} \times \mathscr{T}) : \mathscr{U} \text{ finite subset of the integers}, \\ p = P_U \otimes (P_{X|U} \otimes P_{Y|U}) \otimes W\}.$$

Let a random vector $(U, X, Y, T)$ take values in $\mathscr{U} \times \mathscr{X} \times \mathscr{Y} \times \mathscr{T}$ such that $P_{UXYT} \in \Pi(W)$. Define $\mathscr{R}_{\mathrm{CM}}(p)$ to be the set of those triples $(R_0, R_1, R_2)$ of nonnegative reals satisfying

$$R_1 \le I(T \wedge X|YU), \tag{2.6}$$
$$R_2 \le I(T \wedge Y|XU), \tag{2.7}$$
$$R_1 + R_2 \le I(T \wedge XY|U), \tag{2.8}$$
$$R_0 + R_1 + R_2 \le I(T \wedge XY). \tag{2.9}$$

**Theorem 2.8** (Slepian, Wolf, Willems)**.** *Let $W : \mathscr{X} \times \mathscr{Y} \to \mathscr{P}(\mathscr{T})$. For* DMAC$(W)$*, we have*

$$\overline{\mathscr{C}}^{\mathrm{DM}}_{\mathrm{CM}}(W) = \mathrm{closure}\Big(\bigcup_{p \in \Pi(W)} \mathscr{R}(p)\Big).$$

*The cardinality of $\mathscr{U}$ can be restricted to be at most* $\min\{|\mathscr{X}||\mathscr{Y}| + 2, |\mathscr{T}| + 3\}$*. There exists a weak converse for $\overline{\mathscr{C}}^{\mathrm{DM}}_{\mathrm{CM}}(W)$.*

This theorem is a special case of our Theorem 3.11. Its proof is essentially due to Slepian and Wolf [50], the bound on $|\mathscr{U}|$ is due to Willems [63]. Simpler versions of the proof can be found in, e.g., [63, 65], Wolfowitz also shows that without loss of generality $|\mathscr{U}| \le |\mathscr{X}||\mathscr{Y}| + 2$. The proofs apply the standard methods of random coding in the direct part and Fano's inequality in the converse. Dueck [26] and Ahlswede [6] have shown a strong converse for the discrete memoryless MAC without common message.

Not much is known for the maximal error criterion. Dueck [25] has shown that there is a discrete memoryless MAC without common message and without conferencing which contains in its capacity region for the average error (the case described above) a rate pair which is not achievable under the maximal error criterion, i.e. if the maximal error needs to be arbitrarily small. This extends immediately to the discrete memoryless MACs with common message.

## 2.2. The Discrete Memoryless Multiple Access Channel with Conferencing Encoders

We start with a description of Willems' concept of conferencing encoders. Let finite sets $\mathscr{K}_1, \mathscr{K}_2$ be given. These sets contain the encoders' messages and any further information whose exchange could be useful to enhance transmission over the MAC. Let $J_1$ and $J_2$ be positive integers which can be written as products

$$J_\nu = J_{\nu,1} \cdots J_{\nu,I} \qquad (\nu = 1, 2)$$

for some positive integer $I$ which without loss of generality does not depend on $\nu$. A pair of Willems conferencing functions $(c_1, c_2)$ completely describing such a conference is determined in an iterative manner via sequences of functions $c_{1,1}, \ldots, c_{1,I}$ and $c_{2,1}, \ldots, c_{2,I}$. That means that $c_{1,i}$ determines what the first encoder tells the second in the $i$-th conferencing iteration given the knowledge accumulated so far at encoder 1. Thus, using the notation

$$\bar{\nu} := \begin{cases} 1 & \text{if } \nu = 2, \\ 2 & \text{if } \nu = 1, \end{cases} \tag{2.10}$$

these functions need to satisfy for $\nu = 1, 2$ and $i = 2, \ldots, I$,

$$\begin{aligned} c_{\nu,1} &: \mathscr{K}_\nu \to [J_{\nu,1}], \\ c_{\nu,i} &: \mathscr{K}_\nu \times [J_{\bar{\nu},1}] \times \ldots \times [J_{\bar{\nu},i-1}] \to [J_{\nu,i}]. \end{aligned}$$

For $\nu = 1, 2$ and $i = 2, \ldots, I$, one then recursively defines functions

$$\begin{aligned} c^*_{\nu,1} &: \mathscr{K}_\nu \to [J_{\nu,1}], \\ c^*_{\nu,i} &: \mathscr{K}_1 \times \mathscr{K}_2 \to [J_{\nu,i}] \end{aligned}$$

by

$$\begin{aligned} c^*_{\nu,1}(\kappa_\nu) &= c_{\nu,1}(\kappa_\nu), \\ c^*_{\nu,i}(\kappa_1, \kappa_2) &= c_{\nu,i}\big(\kappa_\nu, c^*_{\bar{\nu},1}(\kappa_{\bar{\nu}}), \ldots, c^*_{\bar{\nu},i-1}(\kappa_1, \kappa_2)\big). \end{aligned}$$

Finally the functions $c_1, c_2$ are obtained by setting

$$c_\nu := (c^*_{\nu,1}, \ldots, c^*_{\nu,I}).$$

As both $c_1$ and $c_2$ may depend on both encoders' messages and additional information, the codewords determined by the encoders after the conference may also depend on both encoders' messages and additional information. Thus if conferencing were unrestricted, this would transform the MAC into a single-sender channel with input alphabet equal to the Cartesian product of the two input alphabets of the MAC. However, Willems introduces a rate restriction for the amount of information exchanged during the conference. For arbitrary fixed numbers $C_1, C_2 \geq 0$ called *conferencing capacities*, he requires that for a blocklength-$n$ code, only those conferencing protocols may be used that satisfy

$$\frac{1}{n} \log J_\nu \leq C_\nu \qquad (\nu = 1, 2). \tag{2.11}$$

**Definition 2.9.** Let $n$ be a positive integer and $C_1, C_2$ nonnegative real numbers. A pair of functions

$$(c_1, c_2) : \mathscr{K}_1 \times \mathscr{K}_2 \to [J_1] \times [J_2]$$

as described above which satisfies (2.11) is called an $(n, C_1, C_2)$-*Willems conference*. $C_1, C_2$ are called the *conferencing capacities*. If $I = 1$, we call $(c_1, c_2)$ a *one-shot Willems conference*.

**Definition 2.10.** Let $n$ be a positive integer and $C_1, C_2 \geq 0$. A *deterministic* $(n, C_1, C_2)$-*code*$_{\mathrm{CONF}}$ with alphabets $\mathscr{X}$, $\mathscr{Y}$ and $\mathscr{T}$ is a quintuple of mappings $(c_1, c_2, f_1, f_2, \varphi)$, where

$$(c_1, c_2) : [K_1] \times [K_2] \to [J_1] \times [J_2]$$

is an $(n, C_1, C_2)$-Willems conference and

$$f_1 : [K_1] \times [J_2] \to \mathscr{X}^n, \quad f_2 : [K_2] \times [J_1] \to \mathscr{Y}^n, \quad \varphi : \mathscr{T}^n \to [K_1] \times [K_2].$$

As for codes$_{\mathrm{CM}}$, we call $f_1, f_2$ the *encoding functions* and $\varphi$ the *decoding function*. The pair $K_1, K_2$ is called the *codelength pair* and $n$ the *blocklength* of $(c_1, c_2, f_1, f_2, \varphi)$.

We denote the set of deterministic $(n, C_1, C_2)$-codes$_{\mathrm{CONF}}$ with codelength pair $(K_1, K_2)$ by $\Gamma_{\mathrm{CONF}}(n, K_1, K_2, C_1, C_2)$.

As for codes$_{\mathrm{CM}}$, it will generally not be necessary to mention the code alphabets as they will be clear from the channel the code is applied to. Thus conferencing actually is part of the encoding procedure, the conferencing functions may be varied as long as they satisfy the conferencing capacity constraint (2.11). Clearly, the decoding function does not differ from that used in the previous section (except that it does not have to decode a common message), but encoding depends on the outcome of the conference. Note that an $(n, 0, 0)$-code$_{\mathrm{CONF}}$ is a traditional blocklength-$n$ MAC code without common message and conferencing.

For the transmission of the message pair $(k_1, k_2)$ using the $(n, C_1, C_2)$-code$_{\mathrm{CONF}}$ $\gamma = (c_1^\gamma, c_2^\gamma, f_1^\gamma, f_2^\gamma, \varphi^\gamma)$, the encoders first hold the conference determined by $(c_1^\gamma, c_2^\gamma)$. Then they form the codewords $\mathbf{x}_{k_1k_2}(\gamma) := f_1^\gamma(k_1, c_2(k_1, k_2))$ and $\mathbf{y}_{k_1k_2}(\gamma) := f_2^\gamma(k_2, c_1(k_1, k_2))$. Thus both codewords generally depend on both senders' messages. The decoding sets $D_{k_1k_2}(\gamma)$ are obtained analogous to the common message case by $D_{k_1k_2}(\gamma) = (\varphi^\gamma)^{-1}(k_1, k_2)$.

**Definition 2.11.** Let $W : \mathscr{X} \times \mathscr{Y} \to \mathscr{P}(\mathscr{T})$ and let $\gamma$ be an $(n, C_1, C_2)$-code$_{\mathrm{CONF}}$. Its *DM-average error* is defined as

$$\bar{e}^{\mathrm{DM}}(\gamma, W) := \frac{1}{K_1 K_2} \sum_{k_1, k_2} W^{\otimes n}\big(D_{k_1k_2}(\gamma)^c | \mathbf{x}_{k_1k_2}(\gamma), \mathbf{y}_{k_1k_2}(\gamma)\big). \tag{2.12}$$

Its *DM-maximal error* is defined by

$$e^{\mathrm{DM}}(\gamma, W) := \max_{k_1, k_2} W^{\otimes n}\big(D_{k_1k_2}(\gamma)^c | \mathbf{x}_{k_1k_2}(\gamma), \mathbf{y}_{k_1k_2}(\gamma)\big). \tag{2.13}$$

**Definition 2.12.** A pair $(R_1, R_2)$ of nonnegative real numbers is called a *deterministically CONF-achievable rate pair for* DMAC$(W)$ *with conferencing capacities* $C_1, C_2 \geq 0$ *under the average (maximal) error criterion* if for every $\lambda \in (0, 1)$ and $\varepsilon > 0$ and $n \geq n_0(\lambda, \varepsilon)$ there is a $(n, C_1, C_2)$-code$_{\mathrm{CONF}}$ $\gamma$ with $\bar{e}^{\mathrm{DM}}(\gamma, W) \leq \lambda$ ($e^{\mathrm{DM}}(\gamma, W) \leq \lambda$) and

$$\frac{1}{n} \log K_\nu \geq R_\nu - \varepsilon \qquad (\nu = 0, 1, 2). \tag{2.14}$$

The set of deterministically CONF-achievable rates is called the *deterministic CONF-capacity region of* DMAC($W$) *with conferencing capacities* $C_1, C_2$ *under the average (maximal) error criterion* and denoted by $\overline{\mathscr{C}}^{\mathrm{DM}}_{\mathrm{CONF}}(W, C_1, C_2)$ ($\mathscr{C}^{\mathrm{DM}}_{\mathrm{CONF}}(W, C_1, C_2)$).

The definition of weak and strong converse is analogous to that in the common message case.

Now let $p \in \Pi(W)$ (see Definition 2.7) and $(U, X, Y, T)$ the corresponding random vector. We define $\mathscr{R}_{\mathrm{CONF}}(p, C_1, C_2)$ to be the set of those rate pairs that satisfy

$$R_1 \le I(T \wedge X|YU) + C_1, \tag{2.15}$$

$$R_2 \le I(T \wedge Y|XU) + C_2, \tag{2.16}$$

$$R_1 + R_2 \le \min\{I(T \wedge XY|U) + C_1 + C_2, I(T \wedge XY)\}. \tag{2.17}$$

**Theorem 2.13** (Willems)**.** *Let* $W : \mathscr{X} \times \mathscr{Y} \to \mathscr{P}(\mathscr{T})$. *For* DMAC($W$) *with conferencing capacities* $C_1, C_2 \ge 0$, *we have*

$$\overline{\mathscr{C}}^{\mathrm{DM}}_{\mathrm{CONF}}(W, C_1, C_2) := \mathrm{closure}\Bigg( \bigcup_{p \in \Pi(W)} \mathscr{R}_{\mathrm{CONF}}(p, C_1, C_2) \Bigg).$$

*The cardinality of* $\mathscr{U}$ *can be restricted to be at most* $\min\{|\mathscr{X}||\mathscr{Y}| + 2, |\mathscr{T}| + 3\}$. *There exists a weak converse for* $\overline{\mathscr{C}}^{\mathrm{DM}}_{\mathrm{CONF}}(W, C_1, C_2)$.

This theorem was proved in [63, 64]. The direct part of its proof bases on Theorem 2.8. The idea is that the outcome of the conference can be treated as a common message. For the converse it is essential that the random variables $M_1$ and $M_2$ which are uniformly distributed on $[K_1]$ and $[K_2]$, respectively, are conditionally independent given the outcome of the conference. Theorem 2.13 is a special case of Theorem 4.6, but the latter will be proved directly using random coding. However, for the wiretap MAC, we will exploit the above connection between the common message and the conferencing problems.

One of the questions solved in Willems' coding theorem [63, 64] is how the capacity region of the discrete memoryless MAC with conferencing encoders scales with $C_1$ and $C_2$. In particular, it is interesting to ask whether the performance of DMAC($W$) with sufficiently large conferencing capacities comes close to the performance of $W$ when regarded as a single-sender channel with input alphabet $\mathscr{X} \times \mathscr{Y}$, i.e. of DMC($W$) (see Appendix A). There are two answers to this question. The notation for single-sender discrete memoryless channels is defined in Appendix A.

DMC($W$) has capacity $\mathscr{C}_{1\mathrm{S}}(W)$ (both under the average and the maximum error criterion). We define the set $\mathscr{M}$ as the set of those $p \in \Pi(W)$ that satisfy $I(T \wedge XY) = \mathscr{C}_{1\mathrm{S}}(W)$.

**Lemma 2.14.** $\mathscr{M}$ *is nonempty.*

*Proof.* Let an arbitrary pair of random variables $(X', Y')$ with values in $\mathscr{X} \times \mathscr{Y}$ be given. We define a triple of random variables $(U, X, Y)$ on $\mathscr{U} \times \mathscr{X} \times \mathscr{Y}$, $\mathscr{U}$ a finite set, satisfying

$$P_{UXY} = P_U \otimes (P_{X|U} \otimes P_{Y|U}) \quad \text{and} \quad P_{XY} = P_{X'Y'}. \tag{2.18}$$

Just set $P_U = P_{X'}$, $P_{X|U} = \delta_U$, and $P_{Y|U} = P_{Y'|X'}$. Then $(U, X, Y)$ satisfies (2.18).

Now assume that $I(T' \wedge X'Y') = \mathscr{C}_{1\mathrm{S}}(W)$, where $(X', Y')$ is a pair of random variables on $\mathscr{X} \times \mathscr{Y}$ and $P_{T'|X'Y'} = W$. Then construct $(U, X, Y)$ from $(X', Y')$ as above and define a random variable $T$ by $P_{T|UXY} = W$. We have $P_{UXYT} \in \Pi(W)$ and $I(T \wedge XY) = I(T' \wedge X'Y')$. □

**Lemma 2.15.** *$\mathscr{C}_{1\mathrm{S}}(W)$ is achieved by the maximal sum rate of* $\mathrm{DMAC}(W)$ *with conferencing encoders if and only if $C_1 + C_2 \geq \min_{p \in \mathscr{M}} I(T \wedge U)$.*

*Proof.* The maximal achievable sum rate for the discrete memoryless MAC with conferencing capacities $C_1, C_2$ equals

$$\max_{p \in \Pi(W)} \min\bigl\{I(T \wedge XY|U) + C_1 + C_2, I(T \wedge XY)\bigr\}.$$

This is at most $\mathscr{C}_{1\mathrm{S}}(W)$. If $C_1 + C_2 < \min_{p \in \mathscr{M}} I(T \wedge U)$, then we have for every $p \in \mathscr{M}$ that

$$\min\bigl\{I(T \wedge XY|U) + C_1 + C_2, I(T \wedge XY)\bigr\} < I(T \wedge XY),$$

and as $\mathscr{M}$ is closed, the maximal achievable sum rate with conferencing capacities $C_1 + C_2 < \min_{p \in \mathscr{M}} I(T \wedge U)$ cannot equal $\mathscr{C}_{1\mathrm{S}}(W)$.

On the other hand, assume that $C_1 + C_2 \geq \min_{p \in \mathscr{M}} I(T \wedge U)$. Assume that $p^* = P_{U^*X^*Y^*T^*} \in \mathscr{M}$ attains this minimum. Then

$$\min\bigl\{I(T^* \wedge X^*Y^*|U^*) + C_1 + C_2, I(T^* \wedge X^*Y^*)\bigr\} = I(T^* \wedge X^*Y^*) = \mathscr{C}_{1\mathrm{S}}(W).$$

This proves the lemma. □

The next question is how large $C_1, C_2$ have to be in order for the complete total cooperation region to be attained, i.e. in order for the set

$$\bigl\{(R_1, R_2) : 0 \leq R_1 + R_2 \leq \mathscr{C}_{1\mathrm{S}}(W)\bigr\}$$

to be contained in $\overline{\mathscr{C}}^{\mathrm{DM}}_{\mathrm{CONF}}(W, C_1, C_2)$.

**Lemma 2.16.** *The complete total cooperation region is attained by the discrete MAC with conferencing encoders if and only if*

$$C_1 \geq \mathscr{C}_{1\mathrm{S}}(W) - \max_{p \in \Pi} I(T \wedge X|Y) \quad \textit{and} \quad C_2 \geq \mathscr{C}_{1\mathrm{S}}(W) - \max_{p \in \Pi} I(T \wedge Y|X). \tag{2.19}$$

*Proof.* As the capacity region is convex, we only have to find the values $C_1, C_2$ where the single-rate bounds on $R_1$ and $R_2$ equal $\mathscr{C}_{1\mathrm{S}}(W)$. As the maxima of $I(T \wedge X|YU)$ and $I(T \wedge Y|XU)$ are attained for single-valued $U$, it is immediate that the conditions (2.19) are necessary and sufficient. □

The two above lemmas show that the performance of the discrete memoryless MAC with conferencing encoders can equal that of the single-sender Discrete Memoryless Channel (DMC) determined by $W$ already with finite $C_1, C_2$. Equality of "performance" here can mean either of the two criteria of the above lemmas. Note that Willems has a result similar to Lemma 2.16 but with possibly larger $C_1, C_2$.

The discrete memoryless MAC with conferencing encoders becomes a DMC if $C_1, C_2 \geq \mathscr{C}_{1\mathrm{S}}(W)$. In this case we know that the average and the maximal error capacity regions coincide because the capacities of the DMC with average and maximal error coincide [20, Problem 6.1 (a)]. Now in Lemma 2.16 we have seen that the complete full cooperation region is CONF-achieved even for smaller values of $C_1$ and $C_2$. But this is a result only concerning the performance under the average error criterion, not the structure of the channel. Thus from this we cannot conclude that at these smaller values of $C_1, C_2$ the maximal and the average error performance coincide as well.

## 2.3. Typical Sequences and More

In the proofs we will extensively use the method of random coding. This relies heavily on estimates for typical sets. Here, we give the definition plus the needed results. At the end of the paragraph we also include two lemmas which do not concern types but which are used very often in information theory.

Given a sequence $\mathbf{x} \in \mathscr{X}^n$ and an $x \in \mathscr{X}$, let $N(x|\mathbf{x})$ be the number of coordinates of $\mathbf{x}$ equal to $x$. This notation can also be applied to pairs of sequences $(\mathbf{x}, \mathbf{y}) \in \mathscr{X}^n \times \mathscr{Y}^n = (\mathscr{X} \times \mathscr{Y})^n$.

**Definition 2.17.** 1) The set of *$\delta$-typical sequences with respect to $X$*, denoted by $T^n_{X,\delta} \subset \mathscr{X}^n$, contains all $\mathbf{x} \in \mathscr{X}$ satisfying

i) $\left|\frac{1}{n}N(x|\mathbf{x}) - P_X(x)\right| \leq \delta \quad$ for all $x \in \mathscr{X}$,

ii) $N(x|\mathbf{x}) = 0$ if $P_X(x) = 0$.

2) The set of *conditionally $\delta$-typical sequences with respect to $P_{X|Y}$ given* $\mathbf{y}$, denoted by $T^n_{X|Y,\delta}(\mathbf{y}) \subset \mathscr{Y}^n$, contains all $\mathbf{x} \in \mathscr{X}$ satisfying

i) $\left|\frac{1}{n}N(x,y|\mathbf{x},\mathbf{y}) - P_{X|Y}(x|y)\right| \leq \delta \quad$ for all $(x,y) \in \mathscr{X} \times \mathscr{Y}$,

ii) $N(x,y|\mathbf{x},\mathbf{y}) = 0$ if $P_{X|Y}(x|y) = 0$.

**Lemma 2.18** ([20], Lemma 17.8)**.** *Let $(A,B)$ be a random pair on the finite Cartesian product $\mathscr{A} \times \mathscr{B}$. Let $\delta, \xi > 0$. Then there exists a $\tilde{c} = \tilde{c}(|\mathscr{A}||\mathscr{B}|) > 0$ such that for sufficiently large $n$*

$$P^{\otimes n}_{B|A}(T^n_{B|A,\delta}(\mathbf{a})^c|\mathbf{a}) \leq 2^{-n\tilde{c}\delta^2}. \tag{2.20}$$

*Further there is a $\zeta = \zeta(P_{AB}, \xi, \delta)$ with $\zeta \to 0$ as $\xi, \delta \to 0$ such that*

$$P^{\otimes n}_{B|A}(\mathbf{b}|\mathbf{a}) \leq 2^{-n(H(B|A)-\zeta)} \qquad \text{if } \mathbf{a} \in T^n_{A,\xi}, \mathbf{b} \in T^n_{B|A,\delta}, \tag{2.21}$$

*and that for $n$ sufficiently large,*

$$|T^n_{A,\xi}| \leq 2^{n(H(A)+\zeta)}, \tag{2.22}$$

$$|T^n_{B|A,\delta}(\mathbf{a})| \leq 2^{n(H(B|A)+\zeta)} \qquad \text{if } \mathbf{a} \in T^n_{A,\xi}. \tag{2.23}$$

**Lemma 2.19** ([20], Lemma 2.10)**.** *Let $(A, B)$ be a random pair on the Cartesian product $\mathscr{A} \times \mathscr{B}$ and $\delta > 0$. Then $(\mathbf{a}, \mathbf{b}) \in T_{AB,\delta}$ implies $\mathbf{a} \in \mathscr{T}_{A,|\mathscr{B}|\delta}$.*

**Lemma 2.20.** *Let $(A, B, C)$ be a random triple on the finite Cartesian product $\mathscr{A} \times \mathscr{B} \times \mathscr{C}$ and $\delta > 0$. Then*

*1) $(\mathbf{a}, \mathbf{b}, \mathbf{c}) \in T^n_{ABC,\delta}$ implies $\mathbf{a} \in T^n_{A|B,2|\mathscr{A}||\mathscr{C}|\delta}(\mathbf{b})$,*

*2) $\mathbf{c} \in T^n_{C|B,\delta}(\mathbf{b})$ implies $T^n_{A|BC,\delta}(\mathbf{b}, \mathbf{c}) \subset T_{A|B,2|\mathscr{C}|\delta}(\mathbf{b})$.*

*Proof.* First we prove 1). Assume that $(\mathbf{a}, \mathbf{b}, \mathbf{c}) \in T^n_{ABC,\delta}$ and let $(a, b) \in \mathscr{A} \times \mathscr{B}$. Then

$$\begin{aligned}
&\left|\frac{1}{n}N(a, b|\mathbf{a}, \mathbf{b}) - P_{A|B}(a|b)\frac{1}{n}N(b|\mathbf{b})\right| \\
&\leq \sum_c \left|\frac{1}{n}N(a, b, c|\mathbf{a}, \mathbf{b}, \mathbf{c}) - P_{ABC}(a, b, c)\right| + \sum_c P_{CA|B}(c, a|b)\left|\frac{1}{n}N(b|\mathbf{b}) - P_B(b)\right| \\
&\leq |\mathscr{C}|\delta + |\mathscr{A}||\mathscr{C}|\delta \leq 2|\mathscr{A}||\mathscr{C}|\delta,
\end{aligned}$$

the last inequality follows from 2.19.

Next we prove 2). Assume that $\mathbf{c} \in T^n_{C|B,\delta}(\mathbf{b})$ and $\mathbf{a} \in T^n_{A|BC,\delta}(\mathbf{b}, \mathbf{c})$. Let $(a, b) \in \mathscr{A} \times \mathscr{B}$. Then

$$\begin{aligned}
&\left|\frac{1}{n}N(a, b|\mathbf{a}, \mathbf{b}) - P_{A|B}(a|b)\frac{1}{n}N(b|\mathbf{b})\right| \\
&\leq \sum_c \left|\frac{1}{n}N(a, b, c|\mathbf{a}, \mathbf{b}, \mathbf{c}) - P_{A|BC}(a|b, c)\frac{1}{n}N(b, c|\mathbf{b}, \mathbf{c})\right| \\
&\qquad + \sum_c P_{A|BC}(a|b, c)\left|\frac{1}{n}N(b, c|\mathbf{b}, \mathbf{c}) - P_{C|B}(c|b)\frac{1}{n}N(b|\mathbf{b})\right| \\
&\leq |\mathscr{C}|\delta + |\mathscr{C}|\delta = 2|\mathscr{C}|\delta.
\end{aligned}$$

□

**Lemma 2.21.** *Let $\mathscr{A}$ and $\mathscr{B}$ be finite sets and let $\mathscr{S}$ be an arbitrary set. Let $A$ be a random variable on $\mathscr{A}$ and for every $s \in \mathscr{S}$ let $B_s$ be a random variable on $\mathscr{B}$. Further let $\delta > 0$. Then there exists a $\zeta = \zeta(|\mathscr{A}||\mathscr{B}|, \delta)$ with $\zeta \to 0$ as $\delta \to 0$ such that for any $\mathbf{b} \in \mathscr{B}^n$*

$$\frac{1}{n}\log\left|\left\{\mathbf{a} : (\mathbf{b}, \mathbf{a}) \in \bigcup_{s\in\mathscr{S}} T^n_{B_sA,\delta}\right\}\right| \leq \sup_{s\in\mathscr{S}} H(A|B_s) + \zeta. \tag{2.24}$$

*Proof.* Fix a $\mathbf{b} \in \mathscr{B}^n$. If the left-hand side of (2.24) equals 0, then nothing has to be shown. Otherwise, writing $\mathscr{S}^*$ for the nonempty subset of those $s \in \mathscr{S}$ satisfying $\mathbf{b} \in T^n_{B_s,|\mathscr{A}|\delta}$, we have by Lemma 2.19

$$\Big\{\mathbf{a} : (\mathbf{b},\mathbf{a}) \in \bigcup_{s\in\mathscr{S}} T^n_{B_sA,\delta}\Big\} \subset \Big\{\mathbf{a} : (\mathbf{b},\mathbf{a}) \in \bigcup_{s\in\mathscr{S}^*} T^n_{B_sA,\delta}\Big\}.$$

We call an element $Q$ of $\mathscr{P}(\mathscr{B}\times\mathscr{A})$ a *joint $n$-type on $\mathscr{B}\times\mathscr{A}$* if $Q(b,a)$ is a multiple of $n$ for all $(b,a)$. Denote by $T^n(s)$ the set of all those joint $n$-types $Q$ on $\mathscr{B}\times\mathscr{A}$ satisfying

$$|P_{B_sA}(b,a) - Q(b,a)| < \delta \qquad \text{for all } (b,a) \in \mathscr{B}\times\mathscr{A}$$

and such that $P_{B_sA}(b,a) = 0$ implies $Q(b,a) = 0$. We also write $T^n(\mathscr{S}^*) = \bigcup_{s\in\mathscr{S}^*} T^n(s)$. Every $T^n_{B_sA,\delta}$ equals the union of those $T^n_Q$ with $Q \in T^n(s)$. Hence

$$\Big|\Big\{\mathbf{a} : (\mathbf{b},\mathbf{a}) \in \bigcup_{s\in\mathscr{S}^*} T^n_{B_sA,\delta}\Big\}\Big| \le \Big|\Big\{\mathbf{a} : (\mathbf{b},\mathbf{a}) \in \bigcup_{Q\in T^n(\mathscr{S}^*)} T^n_Q\Big\}\Big|. \tag{2.25}$$

As there are at most $(n+1)^{|\mathscr{B}||\mathscr{A}|}$ different joint types on $\mathscr{B}^n\times\mathscr{A}^n$, this can be upper-bounded by

$$(n+1)^{|\mathscr{B}||\mathscr{A}|} \max_{Q\in T^n(\mathscr{S}^*)} \Big|\Big\{\mathbf{a} : (\mathbf{b},\mathbf{a}) \in T^n_Q\Big\}\Big|. \tag{2.26}$$

Now let $Q \in T^n(\mathscr{S}^*)$ be contained in $T^n(s)$ for some $s \in \mathscr{S}^*$. This implies by definition of $T^n(s)$ that $T^n_Q \subset T^n_{B_sA,\delta}$. Thus (2.26) is upper-bounded by

$$(n+1)^{|\mathscr{B}||\mathscr{A}|} \max_{s\in\mathscr{S}^*} \big|\big\{\mathbf{a} : (\mathbf{b},\mathbf{a}) \in T^n_{B_sA,\delta}\big\}\big| \le (n+1)^{|\mathscr{B}||\mathscr{A}|} \max_{s\in\mathscr{S}^*} \big|T_{A|B_s,2|\mathscr{A}|\delta}(\mathbf{b})\big|,$$

the inequality is due to Lemma 2.20. Applying (2.23), which is possible because of the definition of $\mathscr{S}^*$, we obtain that $1/n$ times the logarithm of the right-hand side is upper-bounded by

$$\sup_{s\in\mathscr{S}^*} H(A|B_s) + \zeta \le \sup_{s\in\mathscr{S}} H(A|B_s) + \zeta.$$

Thus the proof is complete. □

Finally, for completeness we note two lemmas which are frequently used in information theory. The first one quantifies the uniform continuity of entropy.

**Lemma 2.22** ([20], Lemma 2.7)**.** *Let $\mathscr{X}$ be a finite set and $P, Q \in \mathscr{P}(\mathscr{X})$. If $\delta := \|P - Q\| \le 1/2$, then*

$$|H(P) - H(Q)| \le -\delta \log \frac{\delta}{|\mathscr{X}|}.$$

We also cite the version from [20] (there Lemma 3.8) of Fano's inequality. The original version is generally attributed to Fano, but see the remarks on Chapter 3 of [20] for the story.

**Lemma 2.23.** *For random variables $X, Y$ with values in $\mathscr{X}$,*

$$H(X|Y) \le \mathbb{P}[X \ne Y] \log(|\mathscr{X}| - 1) + h(\mathbb{P}[X \ne Y]).$$

*Here, $h$ is the binary entropy, $h(x) = -x\log x - (1-x)\log(1-x)$.*

# 3. The Compound MAC with Common Message

## 3.1. Introduction

The compound MAC generalizes the discrete memoryless MAC. Compound channels model the situation that the channel is in one of several states and the encoders and the decoder only have limited Channel State Information (CSI). In comparison, there is only one channel state which is completely known to all users in the problem described in the previous section. Thus the encoding and decoding functions can be fitted exactly to the probability law governing the transmission of words.

A compound channel can be seen as a family of discrete memoryless channels. Transmission is done using one member of the family, but encoding and decoding must be performed in such a way that transmission is reliable no matter what the exact channel might be. This criterion is stricter than if the state were determined stochastically. We use the term "compound channel" only for channels where the state is not determined stochastically. (However, in [65], channels with stochastic state are treated in the chapter on compound channels.)

The compound channel we present below to our knowledge is the first in information-theoretic literature where channel knowledge is possible which is neither perfect nor completely unavailable. The receiver's CSI (CSIR) may be arbitrary between full and absent. The transmitters' CSI (CSIT) may be different from CSIR and asymmetric at the two encoders. It is restricted to a finite number of instances, even though the actual number of channel realizations may be infinite. We characterize the capacity region of the compound MAC with common message in this chapter, that of the compound MAC with conferencing encoders will be characterized in the next chapter. Preliminary work is due to Ahlswede [3], who found the capacity region of the 2-state compound MAC without common message, no CSIT and perfect CSIR. This was extended to the case with common message in [41].

There are several communication situations which are appropriately described by a compound MAC. One case is where information is to be sent from two transmitting terminals to one receiving terminal through a fading channel. If the channel remains constant during one transmission block, one obtains a compound channel. Usually, CSIT is not perfect. It might be, however, that the transmitters have access to partial CSI, e.g. by using feedback. This will not determine an exact channel state, but only an approximation. Coding must then be done in such a way that it is reliable for all those channel realizations which are possible according to CSIT.

Another situation to be modeled by compound channels occurs if there are two trans-

mitters each of which would like to send one message to several receivers at the same time. The channels to the different receivers differ from each other because all the terminals are at different locations. Then if CSIT is given as a certain subset of the set of all possible states, this describes that the messages are not intended for all receivers, but only for those corresponding to the given subset. Knowledge about the intended receivers may be asymmetric at the senders. If every receiver has its own decoding procedure, perfect CSIR would be a natural assumption. If the receivers must all use the same decoder, there is no CSIR. Non-trivial CSIR could mean that independently of the decision at the transmitters where data are to be sent (modeled by CSIT), a subset of receivers is chosen as the set which the data are intended for without informing the transmitters about this decision.

## 3.2. Compound MACs

Let $\mathscr{X}, \mathscr{Y}, \mathscr{T}$ be finite alphabets and $\mathscr{W}$ a family of stochastic matrices

$$W_s : \mathscr{X} \times \mathscr{Y} \to \mathscr{P}(\mathscr{T}), \qquad s \in \mathscr{S}. \tag{3.1}$$

**Definition 3.1.** The *compound MAC* $\mathrm{Cp}(\mathscr{W})$ is the MAC

$$W_s^{\otimes n} : \mathscr{X}^n \times \mathscr{Y}^n \to \mathscr{P}(\mathscr{T}^n), \qquad s \in \mathscr{S},\ n = 1, 2 \dots$$

$\mathscr{W}$, and thus its *state set* $\mathscr{S}$, may be finite or infinite. The compound channel model does not include a change of state in the middle of a transmission block. That situation is treated in Chapter 5.

We noted above that the encoders and the decoder may have some CSI. This means that every node has a partition of $\mathscr{S}$ and knows which element of this partition the actual channel state $s$ is contained in. The partitions of the different nodes may differ.

**Definition 3.2.** Let a triple $(T_1, T_2, R)$ of partitions of $\mathscr{S}$ be given. We call $T_1, T_2$ the *CSIT partitions* or the *partitions at encoder 1 or 2*, respectively, and we call $R$ the *CSIR partition* or the *partition at the receiver* if

1) $T_1, T_2$ are finite, i.e. they consist of finitely many disjoint sets covering $\mathscr{S}$, and

2) $W_s = W_{s'}$ implies that $s$ and $s'$ are contained in the same element of $T_1, T_2$ and $R$.

$T_1, T_2, R$ are also called *CSI partitions*. The sets $\tau_\nu \in T_\nu$ and $\rho \in R$ are called *blocks* (of their respective partitions) or *CSI instances*. We write

$$\mathscr{S}^{\rho}_{\tau_1\tau_2} := \tau_1 \cap \tau_2 \cap \rho \quad \text{and} \quad \mathscr{S}_{\tau_1\tau_2} := \bigcup_{\rho \in R} \mathscr{S}^{\rho}_{\tau_1\tau_2}.$$

The constraint that the encoders' partitions be finite is introduced for mathematical accessibility. The drawback of this constraint is that if $\mathscr{S}$ is infinite, the encoders' CSI

certainly cannot be perfect. However, for practical purposes, the restriction that the senders' CSI partitions be finite is no restriction, as this will always be the case. If, say, the receiver's channel state information is perfect, we will write $R = \mathscr{S}$ even though this is a slight abuse of notation (actually we mean $R = \{\{s\} : s \in \mathscr{S}\}$). $\mathscr{S}^{\rho}_{\tau_1\tau_2}$ represents the set of channel states that are possible if the encoders have CSI $\tau_\nu$ ($\nu = 1, 2$) and the receiver has CSI $\rho$. $\mathscr{S}_{\tau_1\tau_2}$ represents the set of channel states which are possible due to the encoders' joint CSI.

**Definition 3.3.** For the compound MAC $\mathrm{Cp}(\mathscr{W})$ with CSI partitions $T_1, T_2, R$ we write $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$.

In order to describe the rate regions of the various compound MAC coding problems we consider below, we now introduce two general sets of probability measures.

**Definition 3.4.** Let $\mathscr{U}, \mathscr{X}, \mathscr{Y}, T, T_1, T_2$ be finite sets. If we are given

- for every $\tau \in T$ a $P_{U_\tau} \in \mathscr{P}(\mathscr{U})$,
- for every $(\tau, \tau_1) \in T \times T_1$ a $P_{X_{\tau\tau_1}|U_\tau} : \mathscr{U} \to \mathscr{P}(\mathscr{X})$,
- for every $(\tau, \tau_2) \in T \times T_2$ a $P_{Y_{\tau\tau_2}|U_\tau} : \mathscr{U} \to \mathscr{P}(\mathscr{Y})$,

then we call the family

$$\pi := \{P_{U_\tau} \otimes (P_{X_{\tau\tau_1}|U_\tau} \otimes P_{Y_{\tau\tau_2}|U_\tau}) : (\tau, \tau_1, \tau_2) \in T \times T_1 \times T_2\} \tag{3.2}$$

a $(T, T_1, T_2)$*-input probability on* $\mathscr{U} \times \mathscr{X} \times \mathscr{Y}$.

Let $\mathscr{W}$ be a set of stochastic matrices as in (3.1) and $f : \mathscr{S} \to T \times T_1 \times T_2$. (As notation suggests, $T, T_1, T_2$ will correspond to certain CSIT constellations.) $f$ defines subsets $\mathscr{S}_{\tau\tau_1\tau_2} := f^{-1}(\tau, \tau_1, \tau_2)$ of $\mathscr{S}$. Assume we are given a $(T, T_1, T_2)$-input probability on $\mathscr{U} \times \mathscr{X} \times \mathscr{Y}$ as in (3.2), where $\mathscr{U}$ is an arbitrary finite subset of the integers. Then through $f$ every $s \in \mathscr{S}$ gives rise to a probability distribution $p_s$ on $\mathscr{U} \times \mathscr{X} \times \mathscr{Y} \times \mathscr{T}$, namely if $s \in \mathscr{S}_{\tau\tau_1\tau_2}$, then

$$p_s = P_{U_\tau} \otimes (P_{X_{\tau\tau_1}|U_\tau} \otimes P_{Y_{\tau\tau_2}|U_\tau}) \otimes W_s. \tag{3.3}$$

We collect these $p_s$ in the family $p = \{p_s : s \in \mathscr{S}\}$.

**Definition 3.5.** Let $\mathscr{W}$ be a family of stochastic matrices as in (3.1) and $f : \mathscr{S} \to T \times T_1 \times T_2$ a mapping as above. By $\Pi_f(\mathscr{W}, T, T_1, T_2)$ we denote the set of all families $p = \{p_s : s \in \mathscr{S}\}$ satisfying (3.3), where $\mathscr{U}$ ranges over the finite subsets of the integers.

## 3.3. The Compound MAC with Common Message

Here we assume that each transmitter has an individual message for the receiver and together they have a common message. The codes may depend on the respective CSI of the senders and the receiver.

**Definition 3.6.** Let $n$ be a positive integer. A *deterministic* $(n, T_1, T_2, R)$-*code*$_{\mathrm{CM}}$ *with alphabets* $\mathscr{X}, \mathscr{Y}, \mathscr{T}$ is a triple of mappings

$$\begin{aligned} f_1 &: ([K_0] \times [K_1]) \times T_1 \to \mathscr{X}^n, \\ f_2 &: ([K_0] \times [K_2]) \times T_2 \to \mathscr{Y}^n, \\ \varphi &: \mathscr{T}^n \times R \to [K_0] \times [K_1] \times [K_2], \end{aligned}$$

where $K_0, K_1, K_2$ are arbitrary positive integers. $f_1, f_2$ are the *encoding functions* and $\varphi$ is the *decoding function*. The triple $(K_0, K_1, K_2)$ is called the *codelength triple* and $n$ is called the *blocklength* of the code$_{\mathrm{CM}}$.

We denote the set of deterministic $(n, T_1, T_2, R)$-codes$_{\mathrm{CM}}$ with codelength triple $(K_0, K_1, K_2)$ by $\Gamma_{\mathrm{CM}}(n, K_0, K_1, K_2, T_1, T_2, R)$.

As usual, we will generally not mention the code alphabets as they will be clear from the context. If $\gamma = (f_1^\gamma, f_2^\gamma, \varphi^\gamma)$ is a code$_{\mathrm{CM}}$, in analogy with Section 2.1, we denote the codewords by $\mathbf{x}_{k_0 k_1}^{\tau_1}(\gamma)$ and $\mathbf{y}_{k_0 k_2}^{\tau_2}(\gamma)$, where $\tau_\nu \in T_\nu$ is an element of the respective CSI partitions. For any element $\rho$ of the receiver's CSI partition $R$, the corresponding decoding sets are called $D_{k_0 k_1 k_2}^{\rho}(\gamma)$. If the actual channel state is $s \in \mathscr{S}_{\tau_1 \tau_2}^{\rho}$, the combined CSI will be $(\tau_1, \tau_2, \rho) \in T_1 \times T_2 \times R$. If in addition the message triple is $(k_0, k_1, k_2)$, the encoders then use the codewords $\mathbf{x}_{k_0 k_1}^{\tau_1}(\gamma)$ and $\mathbf{y}_{k_0 k_2}^{\tau_2}(\gamma)$, respectively, whereas the decoder decides for $(k_0, k_1, k_2)$ if and only if the channel output is contained in $D_{k_0 k_1 k_2}^{\rho}(\gamma)$.

**Definition 3.7.** Let $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ be a compound MAC and $\gamma$ a deterministic $(n, T_1, T_2, R)$-code$_{\mathrm{CM}}$. Its *C-average error* is given by

$$\begin{aligned} &\bar{e}^{\mathrm{Cp}}(\gamma, \mathscr{W}, T_1, T_2, R) \\ &\qquad := \sup_{\tau_1, \tau_2, \rho} \sup_{s \in \mathscr{S}_{\tau_1 \tau_2}^{\rho}} \frac{1}{K_0 K_1 K_2} \sum_{k_0, k_1, k_2} W_s^{\otimes n}\big(D_{k_0 k_1 k_2}^{\rho}(\gamma)^c | \mathbf{x}_{k_0 k_1}^{\tau_1}(\gamma), \mathbf{y}_{k_0 k_2}^{\tau_2}(\gamma)\big). \end{aligned}$$

Its *C-maximal error* is given by

$$e^{\mathrm{Cp}}(\gamma, \mathscr{W}, T_1, T_2, R) := \sup_{\tau_1, \tau_2, \rho} \sup_{s \in \mathscr{S}_{\tau_1 \tau_2}^{\rho}} \max_{k_0, k_1, k_2} W_s^{\otimes n}\big(D_{k_0 k_1 k_2}^{\rho}(\gamma)^c | \mathbf{x}_{k_0 k_1}^{\tau_1}(\gamma), \mathbf{y}_{k_0 k_2}^{\tau_2}(\gamma)\big).$$

The definitions of average and maximal error formalize the requirement that a code$_{\mathrm{CM}}$ corresponding to any CSI triple $(\tau_1, \tau_2, \rho)$ be reliable for every channel state $s$ that is possible according to the joint CSI, i.e. for every $s \in \mathscr{S}_{\tau_1 \tau_2}^{\rho}$.

We also consider random codes$_{\mathrm{CM}}$. They are mainly used in the random coding proofs of the deterministic coding theorems. When we are treating arbitrarily varying MACs in Chapter 5, though, random codes$_{\mathrm{CM}}$ become even more important and play a fundamental role in the description of the deterministic coding region.

**Definition 3.8.** Let $n, K_0, K_1, K_2$ be positive integers. A random $(n, T_1, T_2, R)$-code$_{\mathrm{CM}}$ with alphabets $\mathscr{X}, \mathscr{Y}, \mathscr{T}$ is a random variable $G$ on $\Gamma_{\mathrm{CM}}(n, K_0, K_1, K_2, T_1, T_2, R)$. The *blocklength* of $G$ is $n$, its *codelength triple* is $(K_0, K_1, K_2)$.

In a real-life setting, the application of random codes would require performing a random experiment $G$ whose outcome is known to the encoders and the decoder. This would determine a deterministic code, which would then be used as described above.

**Definition 3.9.** Let $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ be a compound MAC and $G$ a random $(n, T_1, T_2, R)$-code$_{\mathrm{CM}}$. The *C-average error* of $G$ is defined as

$$\bar{e}^{\mathrm{Cp,r}}(G, \mathscr{W}, T_1, T_2, R) := \sup_{\tau_1,\tau_2,\rho} \sup_{s\in\mathscr{S}^{\rho}_{\tau_1\tau_2}} \frac{1}{K_0K_1K_2} \sum_{k_0,k_1,k_2} \sum_{\gamma} W_s^{\otimes n}\big(D^{\rho}_{k_0k_1k_2}(\gamma)^c|\mathbf{x}^{\tau_1}_{k_0k_1}(\gamma), \mathbf{y}^{\tau_2}_{k_0k_2}(\gamma)\big) P_G(\gamma).$$

Its *C-maximal error* is defined as

$$e^{\mathrm{Cp,r}}(G, \mathscr{W}, T_1, T_2, R) := \sup_{\tau_1,\tau_2,\rho} \sup_{s\in\mathscr{S}^{\rho}_{\tau_1\tau_2}} \max_{k_0,k_1,k_2} \sum_{\gamma} W_s^{\otimes n}\big(D^{\rho}_{k_0k_1k_2}(\gamma)^c|\mathbf{x}^{\tau_1}_{k_0k_1}(\gamma), \mathbf{y}^{\tau_2}_{k_0k_2}(\gamma)\big) P_G(\gamma).$$

**Definition 3.10.** 1) A triple $(R_0, R_1, R_2)$ of nonnegative real numbers is called a *deterministically CM-achievable rate triple for* $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ *under the average (maximal) error criterion* if for every $\lambda \in (0,1)$ and $\varepsilon > 0$ and $n \geq n_0(\lambda, \varepsilon)$ there exists a deterministic $(n, T_1, T_2, R)$-code$_{\mathrm{CM}}$ $\gamma$ with $\bar{e}^{\mathrm{Cp}}(\gamma, \mathscr{W}, T_1, T_2, R) \leq \lambda$ ($e^{\mathrm{Cp}}(\gamma, \mathscr{W}, T_1, T_2, R) \leq \lambda$) and

$$\frac{1}{n} \log K_\nu \geq R_\nu - \varepsilon \qquad (\nu = 0, 1, 2).$$

The set of deterministically CM-achievable rates under the average (maximal) error criterion is called the *deterministic CM-capacity region of* $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ *under the average (maximal) error criterion* and denoted by $\overline{\mathscr{C}}^{\mathrm{Cp}}_{\mathrm{CM}}(\mathscr{W}, T_1, T_2, R)$ ($\mathscr{C}^{\mathrm{Cp}}_{\mathrm{CM}}(\mathscr{W}, T_1, T_2, R)$).

2) A triple $(R_0, R_1, R_2)$ of nonnegative real numbers is called a *randomly CM-achievable rate triple for* $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ *under the average (maximal) error criterion* if for every $\lambda \in (0,1)$ and $\varepsilon > 0$ and $n \geq n_0(\lambda, \varepsilon)$ there exists a random $(n, T_1, T_2, R)$-code$_{\mathrm{CM}}$ $G$ with $\bar{e}^{\mathrm{Cp,r}}(G, \mathscr{W}, T_1, T_2, R) \leq \lambda$ ($e^{\mathrm{Cp,r}}(G, \mathscr{W}, T_1, T_2, R) \leq \lambda$) and which satisfies

$$\frac{1}{n} \log K_\nu \geq R_\nu - \varepsilon \qquad (\nu = 0, 1, 2).$$

The set of randomly CM-achievable rates under the average (maximal) error criterion is called the *random CM-capacity region of* $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ *under the average (maximal) error criterion* and denoted by $\overline{\mathscr{C}}^{\mathrm{Cp,r}}_{\mathrm{CM}}(\mathscr{W}, T_1, T_2, R)$ ($\mathscr{C}^{\mathrm{Cp,r}}_{\mathrm{CM}}(\mathscr{W}, T_1, T_2, R)$).

Now we define a function $f_1 : \mathscr{S} \to \{\mathscr{S}\} \times T_1 \times T_2$ mapping every $s$ to the $(\tau_1, \tau_2)$ with $s \in \mathscr{S}_{\tau_1\tau_2}$. This function $f$ has the form as in Definition 3.5 with $T = \{\mathscr{S}\}$, so we obtain a set $\Pi_1(\mathscr{W}, T_1, T_2) := \Pi_{f_1}(\mathscr{W}, \{\mathscr{S}\}, T_1, T_2)$. For every $s \in \mathscr{S}$, the distribution $p_s$ gives rise to a set $\mathscr{R}_{\mathrm{CM}}(p_s)$ as in (2.6)-(2.9). Any $p \in \Pi(\mathscr{W}, \{\mathscr{S}\}, T_1, T_2)$ thus leads to a set

$$\widehat{\mathscr{R}}_{\mathrm{CM}}(p) := \bigcap_{s\in\mathscr{S}} \mathscr{R}_{\mathrm{CM}}(p_s) = \bigcap_{\tau_1,\tau_2} \bigcap_{s\in\mathscr{S}_{\tau_1\tau_2}} \mathscr{R}_{\mathrm{CM}}(p_s).$$

Finally we define

$$\mathscr{C}_1(\mathscr{W}, T_1, T_2) := \mathrm{closure}\left( \bigcup_{p\in\Pi_1(\mathscr{W},T_1,T_2)} \widehat{\mathscr{R}}_{\mathrm{CM}}(p) \right),$$

where closure$(A)$ denotes the topological closure of the set $A$.

**Theorem 3.11.** *Let* $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ *be a compound MAC. We have*

$$\overline{\mathscr{C}}^{\mathrm{Cp}}_{\mathrm{CM}}(\mathscr{W}, T_1, T_2, R) = \overline{\mathscr{C}}^{\mathrm{Cp,r}}_{\mathrm{CM}}(\mathscr{W}, T_1, T_2, R) = \mathscr{C}^{\mathrm{Cp,r}}_{\mathrm{CM}}(\mathscr{W}, T_1, T_2, R) = \mathscr{C}_1(\mathscr{W}, T_1, T_2).$$

*The cardinality of* $\mathscr{U}$ *can be restricted to be at most* $\min\{|\mathscr{X}||\mathscr{Y}| + 2, |\mathscr{T}| + 3\}$. *There exists a weak converse for all three cases.*

The definition of weak converse is the same as in Definition 2.6 if one replaces $\mathscr{C}$ by $\overline{\mathscr{C}}^{\mathrm{Cp}}_{\mathrm{CM}}(\mathscr{W}, T_1, T_2, R)$ or $\overline{\mathscr{C}}^{\mathrm{Cp,r}}_{\mathrm{CM}}(\mathscr{W}, T_1, T_2, R)$ or $\mathscr{C}^{\mathrm{Cp,r}}_{\mathrm{CM}}(\mathscr{W}, T_1, T_2, R)$, respectively.

*Remark* 3.1. $\mathscr{C}_1(\mathscr{W}, T_1, T_2)$ is convex by the concavity of mutual information in the input distribution. The bounds on $|\mathscr{U}|$ follow in the same way as in [63].

*Remark* 3.2. The proof of Theorem 3.11 shows that in all three cases, the capacity regions can be achieved with codes whose error probability tends to zero at exponential speed. That means, e.g. for the case of deterministic coding, that for every rate triple $(R_0, R_1, R_2) \in \overline{\mathscr{C}}^{\mathrm{Cp}}_{\mathrm{CM}}(\mathscr{W}, T_1, T_2, R)$ and every $\varepsilon > 0$, there is a $\zeta > 0$ such that for sufficiently large $n$ one can find an $(n, T_1, T_2, R)$-code$_{\mathrm{CM}}$ $\gamma_n$ with codelength triple $(K_0(n), K_1(n), K_2(n))$ and

1) $\frac{1}{n}\log K_\nu(n) \geq R_\nu - \varepsilon$ for $\nu = 0, 1, 2$,

2) $\bar{e}^{\mathrm{Cp}}(\gamma_n, \mathscr{W}, T_1, T_2, R) \leq 2^{-n\zeta}$.

*Remark* 3.3. Note that $\mathscr{C}_1(\mathscr{W}, T_1, T_2)$ is independent of the receiver's CSI partition $R$. A heuristic explanation of this phenomenon is given in [65, Section 4.5]: the receiver can estimate the channel using a pilot sequence whose length is negligible compared to the blocklength.

*Remark* 3.4. First taking a union and then an intersection of sets in the definition of $\mathscr{C}_1(\mathscr{W}, T_1, T_2)$ is similar to the max-min capacity expression for the classical single-sender discrete memoryless compound channel [20, 65]. Due to the encoders' CSI, though, the analogy is not complete. More precisely, the analogy only works in the extreme case

$T_1 = T_2 = \{\mathscr{S}\}$, because the $p \in \Pi_1(\mathscr{W}, \{\mathscr{S}\}, \{\mathscr{S}\})$ are probability measures instead of families of probability measures. If both $T_1$ and $T_2$ are nontrivial and $p \in \Pi_1(\mathscr{W}, T_1, T_2)$ is the distribution of $(U, X_{\tau_1}, Y_{\tau_2}, T_s)$, only $P_U$ is independent of $(\tau_1, \tau_2)$ and $s$. For example, in the other extreme case that $T_1 = T_2 = \mathscr{S}$, we can write

$$\mathscr{C}_1(\mathscr{W}, \mathscr{S}, \mathscr{S}) = \text{closure}\left( \bigcup_{P_U} \bigcap_{s \in \mathscr{S}} \bigcup_{P_{X_s|U}, P_{Y_s|U}} \mathscr{R}_{\text{CM}}(p_s) \right),$$

where the unions are over the obvious sets of probability measures and stochastic matrices, the $p_s$ in $\mathscr{R}_{\text{CM}}(p_s)$ are built from these $P_U, P_{X_s|U}, P_{Y_s|U}$ and $\mathscr{W}$. As $U$ is independent of $s$, the outer union and the intersection do not commute. This is in contrast to the situation for compound MACs with conferencing encoders, see Remark 4.4.

*Remark* 3.5. For the deterministic capacity under the average error criterion, it has been shown by Ahlswede [1] that there is no strong converse for single-sender compound channels. There is, however, a strong converse if the maximal error criterion is applied. For the compound MAC, we do not consider the maximal error criterion in combination with deterministic coding. But clearly, the nonexistence of a strong converse carries over to the average error case considered above.

## 3.4. The Direct Part

In this section we prove that $\mathscr{C}_1(\mathscr{W}, T_1, T_2)$ is CM-achievable in all three scenarios considered in Theorem 3.11. Due to Remark 3.3, it is sufficient to assume $R = \{\mathscr{S}\}$, i.e. that the receiver does not have any CSI. The strategy is first to show

$$\mathscr{C}_1(\mathscr{W}, T_1, T_2) \subset \mathscr{C}_{\text{CM}}^{\text{Cp,r}}(\mathscr{W}, T_1, T_2, \{\mathscr{S}\}) \subset \overline{\mathscr{C}}_{\text{CM}}^{\text{Cp,r}}(\mathscr{W}, T_1, T_2, \{\mathscr{S}\}). \tag{3.4}$$

The second inclusion is clear, so we can concentrate on the first one. The core of its random coding proof is a general lemma which will be proved in the first part of this subsection. The lemma will also apply to the direct part of the coding theorem for the compound MAC with conferencing encoders treated in the next chapter. We will then specialize the lemma to the form needed for showing the first inclusion in (3.4). The third part of the section is devoted to derandomization, i.e. to derive

$$\mathscr{C}_1(\mathscr{W}, T_1, T_2, \{\mathscr{S}\}) \subset \overline{\mathscr{C}}_{\text{CM}}^{\text{Cp}}(\mathscr{W}, T_1, T_2, \{\mathscr{S}\}) \tag{3.5}$$

from (3.4).

### 3.4.1. A General Random Coding Lemma

The following definition generalizes the definition of half lattices from [32].

**Definition 3.12.** Let $\mathscr{U}, \mathscr{X}, \mathscr{Y}, T, T_1, T_2$ be finite sets and $\pi$ a $(T, T_1, T_2)$-input probability on $\mathscr{U} \times \mathscr{X} \times \mathscr{Y}$ as in (3.2). Let $J, K_1, K_2$ be positive integers. A *generalized random $(J, K_1, K_2)$-half lattice on $\mathscr{U} \times \mathscr{X} \times \mathscr{Y}$ based on $\pi$* is a family of random vectors

$$\left\{ (U_j^\tau, X_{jk_1}^{\tau\tau_1}, Y_{jk_2}^{\tau\tau_2}) : (\tau, \tau_1, \tau_2, j, k_1, k_2) \in T \times T_1 \times T_2 \times [J] \times [K_1] \times [K_2] \right\}$$

on $\mathscr{U} \times \mathscr{X} \times \mathscr{Y}$ distributed according to the probability measure

$$\bigotimes_{\tau} P_{U_\tau}^{\otimes J} \otimes \left( \bigotimes_{\tau_1} P_{X_{\tau\tau_1}|U_\tau}^{\otimes K_1} \otimes \bigotimes P_{Y_{\tau\tau_2}|U_\tau}^{\otimes K_2} \right).$$

This means that every $U_j^\tau$ is distributed according to $P_{U_\tau}$, the conditional distribution of every $X_{jk_1}^{\tau\tau_1}$ given $U_j^\tau$ is $P_{X_{\tau\tau_1}|U_\tau}$ and the conditional distribution of every $Y_{jk_2}^{\tau\tau_2}$ given $U_j^\tau$ is $P_{Y_{\tau\tau_2}|U_\tau}$.

**Definition 3.13.** Let $n$ be a positive integer and $\pi$ a $(T, T_1, T_2)$-input probability on $\mathscr{U} \times \mathscr{X} \times \mathscr{Y}$ as in (3.2). The *n-th memoryless extension of* $\pi$ is the $(T, T_1, T_2)$-input probability on $\mathscr{U}^n \times \mathscr{X}^n \times \mathscr{Y}^n$

$$\pi^{\otimes n} := \{P_{U_\tau}^{\otimes n} \otimes (P_{X_{\tau\tau_1}|U_\tau}^{\otimes n} \otimes P_{Y_{\tau\tau_2}|U_\tau}^{\otimes n}) : (\tau, \tau_1, \tau_2) \in T \times T_1 \times T_2\}$$

Let $\mathrm{Cp}(\mathscr{W})$ be a compound MAC, $n, J, K_1, K_2$ positive integers, and $T, T_1, T_2$ finite sets. Let $f : \mathscr{S} \to T \times T_1 \times T_2$. For any $p \in \Pi_f(\mathscr{W}, T, T_1, T_2)$ and any $p_s \in p$, let $(U_\tau, X_{\tau\tau_1}, Y_{\tau\tau_2}, T_s)$ be the corresponding random vector with values in $\mathscr{U} \times \mathscr{X} \times \mathscr{Y} \times \mathscr{T}$ and distribution $p_s$. $p$ gives rise to a $(T, T_1, T_2)$-input probability on $\mathscr{U} \times \mathscr{X} \times \mathscr{Y}$ and to its $n$-th memoryless extension $\pi^{\otimes n}$ on $\mathscr{U}^n \times \mathscr{X}^n \times \mathscr{Y}^n$. We define the family

$$\{(U_j^\tau, X_{vk_1}^{\tau\tau_1}, Y_{vk_2}^{\tau\tau_2}) : (\tau, \tau_1, \tau_2, j, k_1, k_2) \in T \times T_1 \times T_2 \times [J] \times [K_1] \times [K_2]\}.$$

of random variables to be a generalized random $(J, K_1, K_2)$-half lattice on $\mathscr{U}^n \times \mathscr{X}^n \times \mathscr{Y}^n$ based on $\pi^{\otimes n}$.

For every $(\tau, \tau_1, \tau_2)$, let

$$E_{\tau\tau_1\tau_2} := \bigcup_{s \in \mathscr{S}_{\tau\tau_1\tau_2}} T^n_{U_\tau X_{\tau\tau_1} Y_{\tau\tau_2} T_s, \delta}.$$

Further we define for every $(j, k_1, k_2) \in [J] \times [K_1] \times [K_2]$ the set $D_{jk_1k_2} \subset \mathscr{T}^n$ to contain exactly those $\mathbf{t}$ which satisfy

1) $(U_j^\tau, X_{jk_1}^{\tau\tau_1}, Y_{jk_2}^{\tau\tau_2}, \mathbf{t}) \in E_{\tau\tau_1\tau_2}$ for some $(\tau, \tau_1, \tau_2)$,

2) $(U_{j'}^{\tau'}, X_{j'k_1'}^{\tau'\tau_1'}, Y_{j'k_2'}^{\tau'\tau_2'}, \mathbf{t}) \notin E_{\tau'\tau_1'\tau_2'}$ for every $(\tau', \tau_1', \tau_2', j', k_1', k_2')$ with $(j', k_1', k_2') \neq (j, k_1, k_2)$.

This defines a disjoint family of sets.

**Lemma 3.14.** *If there is a* $\tilde{\zeta} > 0$ *such that*

$$\frac{1}{n} \log(JK_1K_2) < \min_{\tau,\tau_1,\tau_2} \inf_{s \in \mathscr{S}_{\tau\tau_1\tau_2}} I(T_s \wedge X_{\tau\tau_1} Y_{\tau\tau_2}) - \tilde{\zeta}, \tag{3.6}$$

$$\frac{1}{n} \log(K_1K_2) < \min_{\tau,\tau_1,\tau_2} \inf_{s \in \mathscr{S}_{\tau\tau_1\tau_2}} I(T_s \wedge X_{\tau\tau_1} Y_{\tau\tau_2} | U_\tau) - \tilde{\zeta}, \tag{3.7}$$

$$\frac{1}{n} \log K_1 < \min_{\tau,\tau_1,\tau_2} \inf_{s \in \mathscr{S}_{\tau\tau_1\tau_2}} I(T_s \wedge X_{\tau\tau_1} | Y_{\tau\tau_2} U_\tau) - \tilde{\zeta}, \tag{3.8}$$

$$\frac{1}{n} \log K_2 < \min_{\tau,\tau_1,\tau_2} \inf_{s \in \mathscr{S}_{\tau\tau_1\tau_2}} I(T_s \wedge Y_{\tau\tau_2} | X_{\tau\tau_1} U_\tau) - \tilde{\zeta}, \tag{3.9}$$

*then for sufficiently small $\delta$ there is a $\zeta = \zeta(\tilde{\zeta}, \delta)$ such that*

$$\mathbb{E}\big[W_s^{\otimes n}\big(D^c_{jk_1k_2}|X^{\tau\tau_1}_{jk_1}, Y^{\tau\tau_2}_{jk_2}\big)\big] \leq 2^{-n\zeta} \tag{3.10}$$

*for every $(\tau, \tau_1, \tau_2, j, k_1, k_2) \in T \times T_1 \times T_2 \times [J] \times [K_1] \times [K_2]$ and $s \in \mathscr{S}_{\tau\tau_1\tau_2}$.*

*Proof.* Without loss of generality we may assume that $T, T_1, T_2$ are subsets of the integers and write $[T], [T_1], [T_2]$ instead. The proof is similar to that of the Hit Lemmas in [32]. Due to symmetry, it is sufficient to bound

$$\mathbb{E}\big[W_s^{\otimes n}\big(D^c_{111}|X^{11}_{11}, Y^{11}_{11}\big)\big] \tag{3.11}$$

for any $s \in \mathscr{S}_{111}$ with a term independent of $(1,1,1,s) \in [T] \times [T_1] \times [T_2] \times \mathscr{S}_{\tau\tau_1\tau_2}$. (3.11) can be upper-bounded by

$$\mathbb{E}\big[W_s^{\otimes n}(\{\mathbf{t} : (U_1^1, X^{11}_{11}, Y^{11}_{11}, \mathbf{t}) \notin E_{111}\}|X^{11}_{11}, Y^{11}_{11})\big] \tag{3.12}$$

$$+ \sum_{(\tau,\tau_1\tau_2)} \sum_{\substack{(j,k_1,k_2)\\ \neq(1,1,1)}} \mathbb{E}\big[W_s^{\otimes n}(\{\mathbf{t} : (U_j^\tau, X^{\tau\tau_1}_{jk_1}, Y^{\tau\tau_2}_{jk_2}, \mathbf{t}) \in E_{\tau\tau_1\tau_2}\}|X^{11}_{11}, Y^{11}_{11})\big]. \tag{3.13}$$

Due to the definition of $(U_1, X_{11}, Y_{11}, T_s)$ and (2.20), (3.12) equals

$$P^{\otimes n}_{U_1X_{11}Y_{11}T_s}\Bigg(\bigcap_{s\in\mathscr{S}_{111}} \big(T^n_{U_1X_{11}Y_{11}T_s,\delta}\big)^c\Bigg) \leq 1 - P^{\otimes n}_{U_1X_{11}Y_{11}T_s}\big(T^n_{U_1X_{11}Y_{11}T_s,\delta}\big) \leq 2^{-nc\delta^2},$$

where $c = \tilde{c}(|\mathscr{U}||\mathscr{X}||\mathscr{Y}||\mathscr{T}|)$. Hence for (3.12), we obtain exponential convergence to zero independently of the choice of $J, K_1, K_2$. To bound (3.13), we need to distinguish four cases. If $(\tau, j) \neq (1,1)$, then the independence of $(U_1^1, X^{11}_{11}, Y^{11}_{11})$ and $(U_j^\tau, X^{\tau\tau_1}_{jk_1}, Y^{\tau\tau_2}_{jk_2})$ implies (recalling the notation for sections of subsets of Cartesian products in the notation section of Chapter 1)

$$\begin{aligned}
&\mathbb{E}\big[W_s^{\otimes n}(\{\mathbf{t} : (U_j^\tau, X^{\tau\tau_1}_{jk_1}, Y^{\tau\tau_2}_{jk_2}, \mathbf{t}) \in E_{\tau\tau_1\tau_2}\}|X^{11}_{11}, Y^{11}_{11})\big] \\
&= \sum_{\mathbf{t}} P^{\otimes n}_{T_s}(\mathbf{t})\mathbb{P}\big[(U_j^\tau, X^{\tau\tau_1}_{jk_1}, Y^{\tau\tau_2}_{jk_2}) \in E_{\tau\tau_1\tau_2|\mathbf{t}}\big] \\
&\leq \max_{\mathbf{t}} \mathbb{P}\big[(U_j^\tau, X^{\tau\tau_1}_{jk_1}, Y^{\tau\tau_2}_{jk_2}) \in E_{\tau\tau_1\tau_2|\mathbf{t}}\big].
\end{aligned} \tag{3.14}$$

If $(\tau, j) = (1,1)$, but $(\tau_1, k_1) \neq (1,1) \neq (\tau_2, k_2)$, then the conditional independence of the random vectors $(X^{11}_{11}, Y^{11}_{11})$ and $(X^{1\tau_1}_{1k_1}, Y^{1\tau_2}_{1k_2})$ given $U_1^1$ implies

$$\begin{aligned}
&\mathbb{E}\big[W_s^{\otimes n}(\{\mathbf{t} : (U_1^1, X^{1\tau_1}_{1k_1}, Y^{1\tau_2}_{1k_2}, \mathbf{t}) \in E_{1\tau_1\tau_2}\}|X^{11}_{11}, Y^{11}_{11})\big] \\
&= \sum_{\mathbf{u},\mathbf{t}} P^{\otimes n}_{U_1T_s}(\mathbf{u}, \mathbf{t})\mathbb{P}\big[(X^{1\tau_1}_{1k_1}, Y^{1\tau_2}_{1k_2}) \in E_{\tau\tau_1\tau_2|\mathbf{u},\mathbf{t}}|U_1^1 = \mathbf{u}\big] \\
&\leq \max_{\mathbf{u},\mathbf{t}} \mathbb{P}\big[(X^{1\tau_1}_{1k_1}, Y^{1\tau_2}_{1k_2}) \in E_{\tau\tau_1\tau_2|\mathbf{u},\mathbf{t}}|U_1^1 = \mathbf{u}\big].
\end{aligned} \tag{3.15}$$

Similar reasons lead to

$$\begin{aligned}
&\mathbb{E}\big[W_s^{\otimes n}(\{\mathbf{t}:(U_1^1,X_{1k_1}^{1\tau_1},Y_{11}^{11},\mathbf{t})\in E_{1\tau_1 1}\}|X_{11}^{11},Y_{11}^{11})\big]\\
&=\sum_{\mathbf{u},\mathbf{y},\mathbf{t}} P_{U_1Y_{11}T_s}^{\otimes n}(\mathbf{u},\mathbf{y},\mathbf{t})\mathbb{P}\big[X_{1k_1}^{1\tau_1}\in E_{1\tau_1 1|\mathbf{u},\mathbf{y},\mathbf{t}}|U_1^1=\mathbf{u},Y_{11}^{11}=\mathbf{y}\big]\\
&\le \max_{\mathbf{u},\mathbf{y},\mathbf{t}}\mathbb{P}\big[X_{1k_1}^{1\tau_1}\in E_{1\tau_1 1|\mathbf{u},\mathbf{y},\mathbf{t}}|U_1^1=\mathbf{u},Y_{11}^{11}=\mathbf{y}\big] \qquad (3.16)
\end{aligned}$$

if $(\tau,\tau_2,j,k_2)=(1,1,1,1)$ but $(\tau_1,k_1)\neq(1,1)$, and to

$$\begin{aligned}
&\mathbb{E}\big[W_s^{\otimes n}(\{\mathbf{t}:(U_1^1,X_{11}^{11},Y_{1k_2}^{1\tau_2},\mathbf{t})\in E_{11\tau_2}\}|X_{11}^{11},Y_{11}^{11})\big]\\
&=\sum_{\mathbf{u},\mathbf{x},\mathbf{t}} P_{U_1X_{11}T_s}^{\otimes n}(\mathbf{u},\mathbf{x},\mathbf{t})\mathbb{P}\big[Y_{1k_2}^{1\tau_2}\in E_{11\tau_2|\mathbf{u},\mathbf{x},\mathbf{t}}|U_1^1=\mathbf{u},X_{11}^{11}=\mathbf{x}\big]\\
&\le \max_{\mathbf{u},\mathbf{x},\mathbf{t}}\mathbb{P}\big[Y_{1k_2}^{1\tau_2}\in E_{11\tau_2|\mathbf{u},\mathbf{x},\mathbf{t}}|U_1^1=\mathbf{u},X_{11}^{11}=\mathbf{x}\big] \qquad (3.17)
\end{aligned}$$

if $(\tau,\tau_1,j,k_1)=(1,1,1,1)$ but $(\tau_2,k_2)\neq(1,1)$.

Next we derive upper bounds for (3.14)-(3.17). Assume that $(\tau,j)\neq(1,1)$. By Lemma 2.19, if $(\mathbf{u},\mathbf{x},\mathbf{y},\mathbf{t})\in T^n_{U_\tau X_{\tau\tau_1}Y_{\tau\tau_2}T_s,\delta}$, then $(\mathbf{u},\mathbf{x},\mathbf{y})\in T^n_{U_\tau X_{\tau\tau_1}Y_{\tau\tau_2},|\mathscr{T}|\delta}$. Thus (2.21) and Lemma 2.21 imply that (3.14) is upper-bounded by

$$\begin{aligned}
&\exp\big(-n(H(U_\tau X_{\tau\tau_1}Y_{\tau\tau_2})-\sup_{s'\in\mathscr{S}_{\tau\tau_1\tau_2}}H(U_\tau X_{\tau\tau_1}Y_{\tau\tau_2}|T_{s'})-\zeta_1)\big)\\
&=\exp\big(-n(\inf_{s'\in\mathscr{S}_{\tau\tau_1\tau_2}}I(T_{s'}\wedge U_\tau X_{\tau\tau_1}Y_{\tau\tau_2})-\zeta_1)\big)\\
&=\exp\big(-n(\inf_{s'\in\mathscr{S}_{\tau\tau_1\tau_2}}I(T_{s'}\wedge X_{\tau\tau_1}Y_{\tau\tau_2})-\zeta_1)\big), \qquad (3.18)
\end{aligned}$$

where $\zeta_1>0$ only depends on $\delta$ and the cardinalities of the alphabets and tends to zero as $\delta$ tends to zero, and the last equality is due to the fact that $T_{s'}$ is independent of $U_\tau$ given $(X_{\tau\tau_1},Y_{\tau\tau_2})$.

Next we analyze (3.15). By Lemma 2.20, $(\mathbf{u},\mathbf{x},\mathbf{y},\mathbf{t})\in T^n_{U_\tau X_{\tau\tau_1}Y_{\tau\tau_2}T_s,\delta}$ implies $(\mathbf{x},\mathbf{y})\in T^n_{X_{\tau\tau_1}Y_{\tau\tau_2}|U_\tau,2|\mathscr{Y}||\mathscr{X}||\mathscr{T}|\delta}(\mathbf{u})$, so (2.21) and Lemma 2.21 imply the existence of a $\zeta_2$ with properties analogous to those of $\zeta_1$ yielding the upper bound

$$\begin{aligned}
&\exp\big(-n(H(X_{\tau\tau_1}Y_{\tau\tau_2}|U_\tau)-\sup_{s'\in\mathscr{S}_{\tau\tau_1\tau_2}}H(X_{\tau\tau_1}Y_{\tau\tau_2}|U_\tau T_{s'})-\zeta_2)\big)\\
&=\exp\big(-n(\inf_{s'\in\mathscr{S}_{\tau\tau_1\tau_2}}I(T_{s'}\wedge X_{\tau\tau_1}Y_{\tau\tau_2}|U_\tau)-\zeta_2)\big). \qquad (3.19)
\end{aligned}$$

For (3.16), we use that by Lemma 2.20, $(\mathbf{u},\mathbf{x},\mathbf{y},\mathbf{t})\in T^n_{U_\tau X_{\tau\tau_1}Y_{\tau\tau_2}T_s,\delta}$ implies $\mathbf{x}\in T^n_{X_{\tau\tau_1}|Y_{\tau\tau_2}U_\tau,2|\mathscr{X}||\mathscr{T}|\delta}(\mathbf{y},\mathbf{u})$. As above, we can thus conclude that there is a $\zeta_3>0$ with properties analogous to those of $\zeta_1$ and $\zeta_2$ such that (3.16) can be upper-bounded by

$$\begin{aligned}
&\exp\big(-n(H(X_{\tau\tau_1}|Y_{\tau\tau_2}U_\tau)-\sup_{s'\in\mathscr{S}_{\tau\tau_1\tau_2}}H(X_{\tau\tau_1}|U_\tau Y_{\tau\tau_2}T_{s'})-\zeta_3)\big)\\
&=\exp\big(-n(\inf_{s'\in\mathscr{S}_{\tau\tau_1\tau_2}}I(T_{s'}\wedge X_{\tau\tau_1}|Y_{\tau\tau_2}U_\tau)-\zeta_3)\big). \qquad (3.20)
\end{aligned}$$

In an analogous way, we upper-bound (3.17) by

$$\exp\bigl(-n(\inf_{s'\in\mathscr{S}_{\tau\tau_1\tau_2}} I(T_{s'}\wedge Y_{\tau\tau_2}|X_{\tau\tau_1}U_\tau)-\zeta_4)\bigr). \tag{3.21}$$

for an appropriate $\zeta_4>0$.

Using rough bounds on the numbers of $(\tau,\tau_1,\tau_2,j,k_1,k_2)$ corresponding to the above cases, we can now conclude with (3.18)-(3.21) that (3.13) is upper-bounded by

$$\begin{aligned}
&TT_1T_2JK_1K_2\exp\bigl(-n(\min_{\tau,\tau_1,\tau_2}\inf_{s'\in\mathscr{S}_{\tau\tau_1\tau_2}} I(T_{s'}\wedge X_{\tau\tau_1}Y_{\tau\tau_2})-\zeta_1)\bigr)\\
&+T_1T_2K_1K_2\exp\bigl(-n(\min_{\tau,\tau_1,\tau_2}\inf_{s'\in\mathscr{S}_{\tau\tau_1\tau_2}} I(T_{s'}\wedge X_{\tau\tau_1}Y_{\tau\tau_2}|U_\tau)-\zeta_2)\bigr)\\
&+T_1K_1\exp\bigl(-n(\min_{\tau,\tau_1,\tau_2}\inf_{s'\in\mathscr{S}_{\tau\tau_1\tau_2}} I(T_{s'}\wedge X_{\tau\tau_1}|Y_{\tau\tau_2}U_\tau)-\zeta_3)\bigr)\\
&+T_2K_2\exp\bigl(-n(\min_{\tau,\tau_1,\tau_2}\inf_{s'\in\mathscr{S}_{\tau\tau_1\tau_2}} I(T_{s'}\wedge Y_{\tau\tau_2}|X_{\tau\tau_1}U_\tau)-\zeta_4)\bigr).
\end{aligned}$$

If conditions (3.6)-(3.9) are satisfied, one can choose $\delta$ so small that $\max\{\zeta_1,\zeta_2,\zeta_3,\zeta_4\}<\tilde{\zeta}$, so the statement of the lemma holds with $\zeta:=\tilde{\zeta}-\max\{\zeta_1,\zeta_2,\zeta_3,\zeta_4\}$. □

### 3.4.2. Random Coding for the Compound MAC with Common Message

Recall that we may assume for the direct part of Theorem 3.11 that the receiver has no CSI, i.e. that $R=\{\mathscr{S}\}$. Under this assumption, we now prove (3.4) by specializing Lemma 3.14. Let a compound MAC $\mathrm{Cp}(\mathscr{W},T_1,T_2,\{\mathscr{S}\})$ be given. For every $p\in\Pi_1(\mathscr{W},T_1,T_2)$, we define a random $(n,T_1,T_2,\{\mathscr{S}\})$-code$_{\mathrm{CM}}$. Let

$$\bigl\{(U_{k_0},X^{\tau_1}_{k_0k_1},Y^{\tau_2}_{k_0k_2}):(\tau_1,\tau_2,k_0,k_1,k_2)\in T_1\times T_2\times[K_0]\times[K_1]\times[K_2]\bigr\} \tag{3.22}$$

be a generalized random $(K_0,K_1,K_2)$-half lattice on $\mathscr{U}^n\times\mathscr{X}^n\times\mathscr{Y}^n$ based on the $n$-th memoryless extension $\pi^{\otimes n}$ of the $(\{\mathscr{S}\},T_1,T_2)$-input probability $\pi$ induced by $p$. Given a message triple $(k_0,k_1,k_2)$ that is to be transmitted and a CSI instance $(\tau_1,\tau_2)$, the transmitters use the random codewords $X^{\tau_1}_{k_0k_1}$ and $Y^{\tau_2}_{k_0k_2}$.

The decoding sets are completely determined by the family (3.22) and a $\delta>0$ which is chosen later. For $s\in\mathscr{S}$ let $(U,X_{\tau_1},Y_{\tau_2},T_s)$ be the random vector corresponding to $p_s$. For every $\tau_1,\tau_2$, define

$$E_{\tau_1\tau_2}:=\bigcup_{s\in\mathscr{S}_{\tau_1\tau_2}} T^n_{UX_{\tau_1}Y_{\tau_2}T_s,\delta}.$$

This set does not depend on $s\in\mathscr{S}_{\tau_1\tau_2}$. The decoding sets $D_{k_0k_1k_2}$ consist exactly of those $\mathbf{t}\in\mathscr{T}^n$ which satisfy both of the following conditions:

1) there is a $(\tau_1,\tau_2)$ such that

$$(U_{k_0},X^{\tau_1}_{k_0k_1},Y^{\tau_2}_{k_0k_2},\mathbf{t})\in E_{\tau_1\tau_2},$$

2) for all $(k_0', k_1', k_2') \neq (k_0, k_1, k_2)$ and for all $\tau_1', \tau_2'$,

$$(U_{k_0'}, X^{\tau_1'}_{k_0'k_1'}, Y^{\tau_2'}_{k_0'k_2'}, \mathbf{t}) \notin E_{\tau_1'\tau_2'}.$$

Clearly the $D_{k_0k_1k_2}$ are disjoint and do not depend on $\tau_1, \tau_2$ nor on $s$.

Note that the family (3.22) together with the decoding sets has the structure treated in Lemma 3.14. We can thus conclude that if for some $\tilde{\zeta} > 0$

$$\frac{1}{n}\log(K_0K_1K_2) < \min_{\tau_1,\tau_2}\inf_{s\in\mathscr{S}_{\tau_1\tau_2}} I(T_s \wedge X_{\tau_1}Y_{\tau_2}) - \tilde{\zeta}, \tag{3.23}$$

$$\frac{1}{n}\log(K_1K_2) < \min_{\tau_1,\tau_2}\inf_{s\in\mathscr{S}_{\tau_1\tau_2}} I(T_s \wedge X_{\tau_1}Y_{\tau_2}|U) - \tilde{\zeta}, \tag{3.24}$$

$$\frac{1}{n}\log K_1 < \min_{\tau_1,\tau_2}\inf_{s\in\mathscr{S}_{\tau_1\tau_2}} I(T_s \wedge X_{\tau_1}|Y_{\tau_2}U) - \tilde{\zeta}, \tag{3.25}$$

$$\frac{1}{n}\log K_2 < \min_{\tau_1,\tau_2}\inf_{s\in\mathscr{S}_{\tau_1\tau_2}} I(T_s \wedge Y_{\tau_2}|X_{\tau_1}U) - \tilde{\zeta}, \tag{3.26}$$

the maximal error under random coding is bounded by $2^{-n\zeta}$ for some $\zeta > 0$. Thus $\mathscr{C}_1(\mathscr{W}, T_1, T_2)$ is randomly CM-achievable with a maximal error tending to zero in blocklength at exponential speed. This establishes (3.4).

### 3.4.3. Construction of Deterministic Codes

In order to show (3.5), we first assume $|\mathscr{W}| < \infty$. In this case we extract from every random code$_{\mathrm{CM}}$ with small error probability a deterministic code$_{\mathrm{CM}}$ with the same rate triple and with comparably small average error. When $|\mathscr{W}| = \infty$, we approximate $\mathscr{W}$ by finite-state compound MACs.

So let us first assume that $|\mathscr{W}| < \infty$. Let $G$ be a random $(n, T_1, T_2, \{\mathscr{S}\})$-code$_{\mathrm{CM}}$ with $\bar{e}^{\mathrm{Cp,r}}(G, \mathscr{W}, T_1, T_2, \{\mathscr{S}\}) \leq \exp(-n\tilde{\zeta})$ for some $\tilde{\zeta} > 0$. For fixed $\tau_1, \tau_2$ and $s \in \mathscr{S}_{\tau_1\tau_2}$, we define the random variable

$$\bar{e}_s(G, T_1, T_2) := \frac{1}{K_0K_1K_2}\sum_{k_0,k_1,k_2} W_s^{\otimes n}(D_{k_0k_1k_2}(G)^c|\mathbf{x}^{\tau_1}_{k_0k_1}(G), \mathbf{y}^{\tau_2}_{k_0k_2}(G))$$

which by assumption satisfies $\mathbb{E}[\bar{e}_s(G, T_1, T_2)] \leq \exp(-n\tilde{\zeta})$. For $0 < \zeta < \tilde{\zeta}$, define the event

$$B_s := \{\bar{e}_s(G, T_1, T_2) \leq 2^{-n\zeta}\}.$$

If the intersection of the $B_s$ is nonempty, we can infer the existence of a deterministic $(n, T_1, T_2, \{S\})$-code$_{\mathrm{CM}}$ $\gamma$ with $\bar{e}^{\mathrm{Cp}}(\gamma, T_1, T_2, \{\mathscr{S}\}) \leq 2^{-n\zeta}$. And indeed, Markov's inequality implies for large $n$

$$\mathbb{P}\Big[\bigcap_{s\in\mathscr{S}} B_s\Big] \geq 1 - \sum_{s\in\mathscr{S}}\mathbb{P}[B_s^c] \geq 1 - 2^{n\zeta}\sum_{s\in\mathscr{S}}\mathbb{E}[\bar{e}_s(G, T_1, T_2, R)] \geq 1 - |\mathscr{W}|2^{-n(\tilde{\zeta}-\zeta)} > 0.$$

Every rate triple contained in $\mathscr{C}_1(\mathscr{W}, T_1, T_2)$ is thus deterministically CM-achievable with an average error tending to zero exponentially in blocklength. Thus (3.4) implies (3.5) for the case that $|\mathscr{W}| < \infty$.

Now assume that $|\mathscr{W}| = \infty$. For a positive integer $N$ to be chosen later, we first define an approximating compound MAC $\mathscr{W}_N$ with a finite state set $\mathscr{S}_N$. It consists of all the stochastic matrices

$$W_{\tilde{s}} : \mathscr{X} \times \mathscr{Y} \to \mathscr{P}(\mathscr{T}), \qquad \tilde{s} \in \mathscr{S}_N,$$

where $W_{\tilde{s}}(t|x,y)$ is a multiple of $(2N|T_1||T_2|)^{-1}$ for all $x \in \mathscr{X}, y \in \mathscr{Y}, t \in \mathscr{T}$. Clearly, $|\mathscr{S}_N| \le (2N|T_1||T_2| + 1)^{|\mathscr{X}||\mathscr{Y}||\mathscr{T}|}$. The following is a slight variation of [13, Lemma 4].

**Lemma 3.15.** *For every $N > 2|\mathscr{T}|$, there is a function $a_N : \mathscr{S} \to \mathscr{S}_N$ satisfying $a_N(\mathscr{S}_{\tau_1\tau_2}) \cap a_N(\mathscr{S}_{\tau_1'\tau_2'}) = \varnothing$ if $(\tau_1, \tau_2) \ne (\tau_1', \tau_2')$ such that for every $s \in \mathscr{S}$,*

$$|W_s(t|x,y) - W_{a_N(s)}(t|x,y)| \le \frac{|\mathscr{T}|}{N}, \tag{3.27}$$

$$W_s(t|x,y) \le \exp\left(\frac{2|\mathscr{T}|^2}{N \ln 2}\right) W_{a_N(s)}(t|x,y). \tag{3.28}$$

Write $a_N(\mathscr{W})$ for the set of those members of $\mathscr{W}_N$ whose state is contained in $a_N(\mathscr{S})$. We define CSIT partitions $\tilde{T}_1, \tilde{T}_2$ of $a_N(\mathscr{S})$ by

$$\tilde{T}_\nu := \{a_N(\tau_\nu) : \tau_\nu \in T_\nu\} \qquad (\nu = 1, 2).$$

As the sets $a_N(\mathscr{S}_{\tau_1\tau_2})$ are pairwise disjoint, these partitions are well-defined. For $(\tilde{\tau}_1, \tilde{\tau}_2) \in \tilde{T}_1 \times \tilde{T}_2$, we write $\mathscr{S}_{N\tilde{\tau}_1\tilde{\tau}_2} := \tilde{\tau}_1 \cap \tilde{\tau}_2$. $T_\nu$ is in one-to-one correspondence with $\tilde{T}_\nu$ for both $\nu = 1, 2$. Thus we can uniquely identify every $(\{\mathscr{S}\}, T_1, T_2)$-input probability on $\mathscr{U} \times \mathscr{X} \times \mathscr{Y}$ with an $(\{\mathscr{S}\}, \tilde{T}_1, \tilde{T}_2)$-input probability on the same set. This identification induces a natural mapping from $\Pi_1(\mathscr{W}, T_1, T_2)$ to $\Pi_1(a_N(\mathscr{W}), \tilde{T}_1, \tilde{T}_2)$.

Let $p \in \Pi_1(\mathscr{W}, T_1, T_2)$ and $\tilde{p}$ the corresponding element of $\Pi_1(a_N(\mathscr{W}), \tilde{T}_1, \tilde{T}_2)$. Let $(\tau_1, \tau_2) \in T_1 \times T_2$, and $s \in \mathscr{S}_{\tau_1\tau_2}$. By (3.27) and Lemma 2.22, $p$ and $\tilde{p}$ satisfy the inequalities

$$\begin{aligned}
|I(T_s \wedge X_{\tau_1}Y_{\tau_2}) - I(T_{a_N(s)} \wedge X_{\tilde{\tau}_1}Y_{\tilde{\tau}_2})| &\le -2\frac{|\mathscr{T}|^3}{N}\log\frac{|\mathscr{T}|^2}{N},\\
|I(T_s \wedge X_{\tau_1}Y_{\tau_2}|U) - I(T_{a_N(s)} \wedge X_{\tilde{\tau}_1}Y_{\tilde{\tau}_2}|U)| &\le -2\frac{|\mathscr{T}|^3}{N}\log\frac{|\mathscr{T}|^2}{N},\\
|I(T_s \wedge X_{\tau_1}|Y_{\tau_2}U) - I(T_{a_N(s)} \wedge X_{\tilde{\tau}_1}|Y_{\tilde{\tau}_2}U)| &\le -2\frac{|\mathscr{T}|^3}{N}\log\frac{|\mathscr{T}|^2}{N},\\
|I(T_s \wedge Y_{\tau_2}|X_{\tau_1}U) - I(T_{a_N(s)} \wedge Y_{\tilde{\tau}_2}|X_{\tilde{\tau}_1}U)| &\le -2\frac{|\mathscr{T}|^3}{N}\log\frac{|\mathscr{T}|^2}{N}.
\end{aligned}$$

Now fix a triple $(R_0, R_1, R_2)$ contained in the interior of $\widehat{\mathscr{R}}_{\mathrm{CM}}(p)$. The above inequalities imply that for sufficiently large $N$ it is contained in the interior of $\widehat{\mathscr{R}}_{\mathrm{CM}}(\tilde{p})$. We have already established the validity of (3.5) for $\mathrm{Cp}(a_N(\mathscr{W}), \tilde{T}_1, \tilde{T}_2, \{a_N(\mathscr{S})\})$. In fact, $\widehat{\mathscr{R}}_{\mathrm{CM}}(\tilde{p})$

is CM-achievable for $\mathrm{Cp}(a_N(\mathscr{W}), \tilde{T}_1, \tilde{T}_2, \{a_N(\mathscr{S})\})$ by deterministic codes$_{\mathrm{CM}}$ with exponentially decreasing average error. Thus for any $\varepsilon > 0$ and $n$ sufficiently large, there is a deterministic $(n, \tilde{T}_1, \tilde{T}_2, \{a_N(\mathscr{S})\})$-code$_{\mathrm{CM}}$ $\gamma$ satisfying

$$\frac{1}{n} \log K_\nu \geq R_\nu - \varepsilon \qquad (\nu = 0, 1, 2) \tag{3.29}$$

and $\bar{e}^{\mathrm{Cp}}(\gamma, a_N(\mathscr{W}), \tilde{T}_1, \tilde{T}_2, \{a_N(\mathscr{S})\}) \leq 2^{-n\zeta}$ for some $\zeta > 0$. We now apply $\gamma$ for transmission over $\mathrm{Cp}(\mathscr{W}, T_1, T_2, \{\mathscr{S}\})$, which is possible due to the one-to-one correspondence of $T_\nu$ and $\tilde{T}_\nu$, and bound its average error. For any $s \in \mathscr{S}$, assume that $a_N(s) \in \mathscr{S}_{N\tilde{\tau}_1\tilde{\tau}_2}$. (3.28) implies

$$\begin{aligned}
&\frac{1}{K_0 K_1 K_2} \sum_{k_0,k_1,k_2} W_s^{\otimes n}(D_{k_0k_1k_2}(\gamma)^c | \mathbf{x}_{k_0k_1}^{\tilde{\tau}_1}(\gamma), \mathbf{y}_{k_0k_2}^{\tilde{\tau}_2}(\gamma)) \\
&\leq 2^{n \cdot 2|\mathscr{T}|^2/(N \ln 2)} \frac{1}{K_0 K_1 K_2} \sum_{k_0,k_1,k_2} W_{a_N(s)}^{\otimes n}(D_{k_0k_1k_2}(\gamma)^c | \mathbf{x}_{k_0k_1}^{\tilde{\tau}_1}(\gamma), \mathbf{y}_{k_0k_2}^{\tilde{\tau}_2}(\gamma)) \\
&\leq \exp\left(-n\left(\zeta - \frac{2|\mathscr{T}|^2}{N \ln 2}\right)\right).
\end{aligned} \tag{3.30}$$

By enlarging $N$ if necessary, this tends to zero exponentially as $n$ approaches infinity, so one obtains an exponentially small average probability of error when $\gamma$ is used for transmission over $\mathrm{Cp}(\mathscr{W}, T_1, T_2, \{S\})$. As the capacity region is closed, (3.29) and (3.30) imply the validity of (3.5) for compound MACs with common message which have infinitely many states.

*Remark* 3.6. Note that this method is independent of the exact form of the code. Only the family of codewords and decoding sets matters, so it will also be applicable for the compound MAC with conferencing encoders in the next chapter.

## 3.5. The Converse

In this section we again start with a general lemma which is the core of the converses both for the compound MAC with common message and the compound MAC with conferencing encoders. In the second part of the section, we specialize the lemma to the first case and show the weak converse for the compound MAC with common message.

### 3.5.1. A General Lemma

Assume we are given any compound MAC $\mathrm{Cp}(\mathscr{W})$ together with a function $f : \mathscr{S} \to T \times T_1 \times T_2$, where $T, T_1, T_2$ are finite sets. We write $\mathscr{S}_{\tau\tau_1\tau_2} := f^{-1}(\tau, \tau_1\tau_2)$. Let $n$ be a positive integer and let $G$ be a random $(n, T_1, T_2, \mathscr{S})$-code$_{\mathrm{CM}}$ with rate triple $(J, K_1, K_2)$. For every realization $\gamma$ of $G$, we write $\gamma = (f_1^\gamma, f_2^\gamma, \varphi^\gamma)$ for the triple of encoding and decoding functions. In addition to $G$, we define the following random variables:

1) for every $\tau \in T$ a random variable $M_0^\tau$ on $[J]$ which is independent of $G$,

2) random variables $M_1, M_2$ on $[K_1]$ and $[K_2]$, respectively, which are also independent of $G$ and which for every $\tau \in T$ are conditionally independent given $M_0^\tau$,

3) for each $(\tau, \tau_1, \tau_2) \in T \times T_1 \times T_2$

$$X^{\tau\tau_1} = f_1^G(M_0^\tau, M_1, \tau_1), \quad Y^{\tau\tau_2} = f_2^G(M_0^\tau, M_2, \tau_2),$$

4) for each $s \in \mathscr{S}_{\tau\tau_1\tau_2}$ a $T^s$ taking values in $\mathscr{T}^n$ such that for every $\mathbf{x} \in \mathscr{X}^n$, $\mathbf{y} \in \mathscr{Y}^n$, $\mathbf{t} \in \mathscr{T}^n$, $(j, k_1, k_2) \in [J] \times [K_1] \times [K_2]$, and $\gamma \in \Gamma_{\mathrm{CM}}(n, J, K_1, K_2, T_1, T_2, \mathscr{S})$

$$\mathbb{P}[T^s = \mathbf{t} | X^{\tau\tau_1} = \mathbf{x}, Y^{\tau\tau_2} = \mathbf{y}, M_0^\tau = j, M_1 = k_1, M_2 = k_2, G = \gamma] = W_s^{\otimes n}(\mathbf{t}|\mathbf{x}, \mathbf{y}),$$

5) for every $s \in \mathscr{S}$ a vector $(\hat{M}_0^s, \hat{M}_1^s, \hat{M}_2^s) = \varphi^G(T^s, s)$.

**Lemma 3.16.** *There is a $p = \{p_s : s \in \mathscr{S}\} \in \Pi_f(\mathscr{W}, T, T_1, T_2)$ such that for every $(\tau, \tau_1, \tau_2) \in T \times T_1 \times T_2$ and $s \in \mathscr{S}_{\tau\tau_1\tau_2}$, if $(U_\tau, X_{\tau\tau_1}, Y_{\tau\tau_2}, T_s)$ is the random vector corresponding to $p_s$,*

$$\begin{aligned}
\frac{1}{n}H(M_1|M_2) &\le I(X_{\tau\tau_1} \wedge T_s | Y_{\tau\tau_2} U_\tau) + \frac{1}{n}\big(I(M_1 \wedge M_0^\tau | M_2) + \Delta^s\big),\\
\frac{1}{n}H(M_2|M_1) &\le I(Y_{\tau\tau_2} \wedge T_s | X_{\tau\tau_1} U_\tau) + \frac{1}{n}\big(I(M_2 \wedge M_0^\tau | M_1) + \Delta^s\big),\\
\frac{1}{n}H(M_1M_2) &\le I(X_{\tau\tau_1}Y_{\tau\tau_2} \wedge T_s | U_\tau) + \frac{1}{n}\big(I(M_1M_2 \wedge M_0^\tau) + \Delta^s\big),\\
\frac{1}{n}H(M_0^\tau M_1 M_2) &\le I(X_{\tau\tau_1}Y_{\tau\tau_2} \wedge T_s) + \frac{1}{n}\Delta^s.
\end{aligned}$$

*Here, $\Delta^s := 1 + \mathbb{P}[(\hat{M}_0^s, \hat{M}_1^s, \hat{M}_2^s) \neq (M_0^\tau, M_1, M_2)] \cdot H(M_0^\tau M_1 M_2)$.*

*Proof of Lemma 3.16.* Let $s \in \mathscr{S}^\rho_{\tau\tau_1\tau_2}$. Set $\mathbb{P}[(\hat{M}_0^s, \hat{M}_1^s, \hat{M}_2^s) \neq (M_0^\tau, M_1, M_2) | G = \gamma] =: \lambda^{\gamma,s}$. Fano's inequality (Lemma 2.23) implies

$$H(M_0^\tau M_1 M_2 | T^s, \gamma) \le 1 + \lambda^{\gamma,s} H(M_0^\tau M_1 M_2) =: \Delta^{\gamma,s}. \tag{3.31}$$

The chain rule for entropy implies that also

$$\max\big\{H(M_1|M_2M_0^\tau T^s, \gamma), H(M_2|M_1M_0^\tau T^s, \gamma)\big\} \le H(M_1M_2|M_0^\tau T^s, \gamma) \le \Delta^{\gamma,s}. \tag{3.32}$$

Using (3.31),(3.32) and the independence of $(M_0^\tau, M_1, M_2)$ and $G$, we obtain the inequalities

$$H(M_1|M_2) \le I(M_1 \wedge T^s M_0^\tau | M_2, \gamma) + \Delta^{\gamma,s}, \tag{3.33}$$
$$H(M_2|M_1) \le I(M_2 \wedge T^s M_0^\tau | M_1, \gamma) + \Delta^{\gamma,s}, \tag{3.34}$$
$$H(M_1M_2) \le I(M_1M_2 \wedge T^s M_0^\tau | \gamma) + \Delta^{\gamma,s}, \tag{3.35}$$
$$H(M_0^\tau M_1 M_2) \le I(M_0^\tau M_1 M_2 \wedge T^s | \gamma) + \Delta^{\gamma,s}. \tag{3.36}$$

With the chain rule for mutual information, (3.33)-(3.35) can be transformed into

$$H(M_1|M_2) \leq I(M_1 \wedge T^s|M_2M_0^\tau,\gamma) + I(M_1 \wedge M_0^\tau|M_2) + \Delta^{\gamma,s}, \tag{3.37}$$
$$H(M_2|M_1) \leq I(M_2 \wedge T^s|M_1M_0^\tau,\gamma) + I(M_2 \wedge M_0^\tau|M_1) + \Delta^{\gamma,s}, \tag{3.38}$$
$$H(M_1M_2) \leq I(M_1M_2 \wedge T^s|M_0^\tau,\gamma) + I(M_1M_2 \wedge M_0^\tau) + \Delta^{\gamma,s}, \tag{3.39}$$

To further bound (3.36)-(3.39), we use Lemma 2 from [50] which is a generalized version of the Data Processing Inequality (see e.g. [20]). Its proof bases purely on the rules for calculating with mutual information and the structure of the random variables involved. Translated into our only notationally slightly more complicated setting, the lemma states that

$$I(M_1 \wedge T^s|M_2M_0^\tau,\gamma) \leq I(X^{\tau\tau_1} \wedge T^s|Y^{\tau\tau_2}M_0^\tau,\gamma), \tag{3.40}$$
$$I(M_2 \wedge T^s|M_1M_0^\tau,\gamma) \leq I(Y^{\tau\tau_2} \wedge T^s|X^{\tau\tau_1}M_0^\tau,\gamma), \tag{3.41}$$
$$I(M_1M_2 \wedge T^s|M_0^\tau,\gamma) \leq I(X^{\tau\tau_1}Y^{\tau\tau_2} \wedge T^s|M_0^\tau,\gamma), \tag{3.42}$$
$$I(M_0^\tau M_1M_2 \wedge T^s|\gamma) \leq I(X^{\tau\tau_1}Y^{\tau\tau_2} \wedge T^s|\gamma). \tag{3.43}$$

The next goal is a single-letter representation of the right-hand terms in (3.40)-(3.43). We start with (3.40). For $m = 1,\ldots,n$, set $T^s_{[m]} := (T^s_1,\ldots,T^s_m)$. As $T^s$ is linked to $(X^{\tau\tau_1},Y^{\tau\tau_2},M_0^\tau,G)$ through a memoryless channel,

$$I(X^{\tau\tau_1} \wedge T^s|Y^{\tau\tau_2}M_0^\tau,\gamma) = \sum_{m=1}^n \big\{H(T^s_m|Y^{\tau\tau_2}M_0^\tau T^s_{[m-1]},\gamma) - H(T^s_m|X^{\tau\tau_1}_m Y^{\tau\tau_2}_m M_0^\tau,\gamma)\big\}.$$

As $H(T^s_m|Y^{\tau\tau_2}M_0^\tau T^s_{[m-1]},\gamma) \leq H(T^s_m|Y^{\tau\tau_2}_m M_0^\tau,\gamma)$ this gives

$$\begin{aligned} I(X^{\tau\tau_1} \wedge T^s|Y^{\tau\tau_2}M_0^\tau,\gamma) &\leq \sum_{m=1}^n \big\{H(T^s_m|Y^{\tau\tau_2}_m M_0^\tau,\gamma) - H(T^s_m|X^{\tau\tau_1}_m Y^{\tau\tau_2}_m M_0^\tau,\gamma)\big\} \\ &= \sum_{m=1}^n I(T^s_m \wedge X^{\tau\tau_1}_m|Y^{\tau\tau_2}_m M_0^\tau,\gamma). \end{aligned} \tag{3.44}$$

In an analogous manner, one shows that

$$I(Y^{\tau\tau_2} \wedge T^s|X^{\tau\tau_1}M_0^\tau,\gamma) \leq \sum_{m=1}^n I(T^s_m \wedge Y^{\tau\tau_2}_m|X^{\tau\tau_1}_m M_0^\tau,\gamma), \tag{3.45}$$
$$I(T^s \wedge X^{\tau\tau_1}Y^{\tau\tau_2}|M_0^\tau,\gamma) \leq \sum_{m=1}^n I(T^s_m \wedge X^{\tau\tau_1}_m Y^{\tau\tau_2}_m|M_0^\tau,\gamma), \tag{3.46}$$
$$I(T^s \wedge X^{\tau\tau_1}Y^{\tau\tau_2}|\gamma) \leq \sum_{m=1}^n I(T^s_m \wedge X^{\tau\tau_1}_m Y^{\tau\tau_2}_m|\gamma). \tag{3.47}$$

Now we define the distribution $p_s$ of a random vector $(U_\tau, X_{\tau\tau_1}, Y_{\tau\tau_2}, T_s)$ with values in $([n] \times [J] \times \Gamma_{\mathrm{CM}}(n, K_0, K_1, K_2, T_1, T_2, \mathscr{S}) \times \mathscr{X} \times \mathscr{Y} \times \mathscr{T}$ by

$$\begin{aligned}
P_{U_\tau}(m, j, \gamma) &= \frac{1}{n} P_{M_0^\tau}(j) P_G(\gamma), \\
P_{X_{\tau\tau_1}|U_\tau}(x|m, j, \gamma) &= \mathbb{P}[X_m^{\tau\tau_1} = x | M_0^\tau = j, G = \gamma], \\
P_{Y_{\tau\tau_2}|U_\tau}(y|m, j, \gamma) &= \mathbb{P}[Y_m^{\tau\tau_2} = y | M_0^\tau = j, G = \gamma], \\
P_{T_s|X_{\tau\tau_1}Y_{\tau\tau_2}U_\tau}(t|x, y, (m, j, \gamma)) &= W_s(t|x, y).
\end{aligned}$$

The set $\{p_s : s \in \mathscr{S}\}$ is an element of $\Pi_f(\mathscr{W}, T, T_1, T_2)$. $U_\tau$ can be represented as $U_\tau = (\tilde{U}_\tau, G)$ with a $\tilde{U}_\tau$ independent of $G$. Then, following the estimates from (3.33)-(3.36) to (3.44)-(3.47) for every $s \in \mathscr{S}$ and dividing by $n$ gives

$$\begin{aligned}
\frac{1}{n} H(M_1|M_2) &\le I(X_{\tau\tau_1} \wedge T_s | Y_{\tau\tau_2} \tilde{U}_\tau, \gamma) + \frac{1}{n}\big(I(M_1 \wedge M_0^\tau | M_2) + \Delta^{\gamma,s}\big), \\
\frac{1}{n} H(M_2|M_1) &\le I(Y_{\tau\tau_2} \wedge T_s | X_{\tau\tau_1} \tilde{U}_\tau, \gamma) + \frac{1}{n}\big(I(M_2 \wedge M_0^\tau | M_1) + \Delta^{\gamma,s}\big), \\
\frac{1}{n} H(M_1 M_2) &\le I(X_{\tau\tau_1} Y_{\tau\tau_2} \wedge T_s | \tilde{U}_\tau, \gamma) + \frac{1}{n}\big(I(M_1 M_2 \wedge M_0^\tau) + \Delta^{\gamma,s}\big), \\
\frac{1}{n} H(M_0^\tau M_1 M_2) &\le I(X_{\tau\tau_1} Y_{\tau\tau_2} \wedge T_s | \gamma) + \frac{1}{n} \Delta^{\gamma,s}.
\end{aligned}$$

Taking the expectation with respect to $P_G$ on both sides yields

$$\begin{aligned}
\frac{1}{n} H(M_1|M_2) &\le I(X_{\tau\tau_1} \wedge T_s | Y_{\tau\tau_2} U_\tau) + \frac{1}{n}\big(I(M_1 \wedge M_0^\tau | M_2) + \Delta^{s}\big), \\
\frac{1}{n} H(M_2|M_1) &\le I(Y_{\tau\tau_2} \wedge T_s | X_{\tau\tau_1} U_\tau) + \frac{1}{n}\big(I(M_2 \wedge M_0^\tau | M_1) + \Delta^{s}\big), \\
\frac{1}{n} H(M_1 M_2) &\le I(X_{\tau\tau_1} Y_{\tau\tau_2} \wedge T_s | U_\tau) + \frac{1}{n}\big(I(M_1 M_2 \wedge M_0^\tau) + \Delta^{s}\big), \\
\frac{1}{n} H(M_0^\tau M_1 M_2) &\le I(X_{\tau\tau_1} Y_{\tau\tau_2} \wedge T_s | G) + \frac{1}{n} \Delta^{s}.
\end{aligned}$$

Finally, we note that $G$ and $T_s$ are independent given $(X_{\tau\tau_1}, Y_{\tau\tau_2})$, so

$$I(X_{\tau\tau_1} Y_{\tau\tau_2} \wedge T_s | G) \le I(G X_{\tau\tau_1} Y_{\tau\tau_2} \wedge T_s) = I(X_{\tau\tau_1} Y_{\tau\tau_2} \wedge T_s).$$

This proves the lemma. □

### 3.5.2. The Weak Converse

The weak converse for $\overline{\mathscr{C}}_{\mathrm{CM}}^{\mathrm{Cp,r}}(\mathscr{W}, T_1, T_2, R)$ implies the weak converses for $\overline{\mathscr{C}}_{\mathrm{CM}}^{\mathrm{Cp}}(\mathscr{W}, T_1, T_2, R)$ and $\mathscr{C}_{\mathrm{CM}}^{\mathrm{Cp,r}}(\mathscr{W}, T_1, T_2, R)$. Hence all we have to show is that any random $(n, T_1, T_2, R)$-code$_{\mathrm{CM}}$ $G$ whose codelength triple satisfies

$$\left\| \frac{1}{n} (\log K_0, \log K_1, \log K_2) - \mathscr{C}_1(\mathscr{W}, T_1, T_2) \right\| > \varepsilon \tag{3.48}$$

incurs an average error $\bar{e}^{\mathrm{Cp,r}}(G, \mathscr{W}, T_1, T_2, R) \geq \lambda(\varepsilon) > 0$ if $n$ is sufficiently large. Here, $\|\cdot\|$ is any norm on $\mathbb{R}^3$. In view of Remark 3.3, we may assume that $R = \mathscr{S}$.

Let $G$ be a random $(n, T_1, T_2, R)$-code$_{\mathrm{CM}}$ which satisfies (3.48). Let $(M_0, M_1, M_2)$ be a random vector independent of $G$ and uniformly distributed on $[K_0] \times [K_1] \times [K_2]$. For every $(\tau_1, \tau_2) \in T_1 \times T_2$ and $s \in \mathscr{S}_{\tau_1\tau_2}$, we further define random variables

1) $X^{\tau_1} = f_1^G(M_0, M_1, \tau_1), \quad Y^{\tau_2} = f_2^G(M_0, M_2, \tau_2)$,

2) $T^s$ satisfying for every $\mathbf{t} \in \mathscr{T}^n$

$$\mathbb{P}[T^s = \mathbf{t} | X^{\tau_1} = \mathbf{x}, Y^{\tau_2} = \mathbf{y}, M_0 = k_0, M_1 = k_1, M_2 = k_2, G = \gamma] = W_s^{\otimes n}(\mathbf{t}|\mathbf{x}, \mathbf{y}),$$

3) and finally $(\hat{M}_0^s, \hat{M}_1^s, \hat{M}_2^s) = \varphi^G(T^s, s)$.

Obviously, these random variables satisfy the conditions of Lemma 3.16 with $T = \{\mathscr{S}\}$ and $f = f_1$, so we can infer the existence of a $p \in \Pi_1(\mathscr{W}, T_1, T_2)$ such that for every $(\tau_1, \tau_2) \in T_1 \times T_2$ and $s \in \mathscr{S}_{\tau_1\tau_2}$

$$\frac{1}{n} \log K_1 \leq I(X_{\tau_1} \wedge T_s | Y_{\tau_2} U) + \frac{1}{n}\Delta, \tag{3.49}$$

$$\frac{1}{n} \log K_2 \leq I(Y_{\tau_2} \wedge T_s | X_{\tau_1} U) + \frac{1}{n}\Delta, \tag{3.50}$$

$$\frac{1}{n} \log K_1 K_2 \leq I(X_{\tau_1} Y_{\tau_2} \wedge T_s | U) + \frac{1}{n}\Delta, \tag{3.51}$$

$$\frac{1}{n} \log K_0 K_1 K_2 \leq I(X_{\tau_1} Y_{\tau_2} \wedge T_s) + \frac{1}{n}\Delta. \tag{3.52}$$

Here, $\Delta := 1 + \lambda \log K_0 K_1 K_2$, where $\lambda := \bar{e}^{\mathrm{Cp,r}}(G, \mathscr{W}, T_1, T_2, \mathscr{S})$.

On the other hand, (3.48) implies

$$\left\| \frac{1}{n}(\log K_0, \log K_1, \log K_2) - \bigcap_{s \in \mathscr{S}} \mathscr{R}_{\mathrm{CM}}(p_s) \right\| > \varepsilon$$

for the above $p$. From this we can infer the existence of an $\varepsilon' = \varepsilon'(\varepsilon) > 0$ such that

$$\frac{1}{n} \log K_1 \geq \inf_{s \in \mathscr{S}} I(X_{\tau_1} \wedge T_s | Y_{\tau_2} U) + \varepsilon', \tag{3.53}$$

or

$$\frac{1}{n} \log K_2 \geq \inf_{s \in \mathscr{S}} I(Y_{\tau_2} \wedge T_s | X_{\tau_1} U) + \varepsilon', \tag{3.54}$$

or

$$\frac{1}{n} \log K_1 K_2 \geq \inf_{s \in \mathscr{S}} I(X_{\tau_1} Y_{\tau_2} \wedge T_s | U) + \varepsilon', \tag{3.55}$$

or

$$\frac{1}{n} \log K_0 K_1 K_2 \geq \inf_{s \in \mathscr{S}} I(X_{\tau_1} Y_{\tau_2} \wedge T_s) + \varepsilon'. \tag{3.56}$$

Set

$$I_0 := \max_{p \in \Pi_1(\mathscr{W}, T_1, T_2)} \inf_{s \in \mathscr{S}} I(X_{\tau_1} Y_{\tau_2} \wedge T_s).$$

There are now four cases.

*Case 1:* (3.56) *holds.* Then we obtain from (3.52)

$$\inf_{s\in\mathscr{S}} I(X_{\tau_1}Y_{\tau_2}\wedge T_s)+\varepsilon' \le \frac{1}{1-\lambda}\left(\inf_{s\in\mathscr{S}} I(X_{\tau_1}Y_{\tau_2}\wedge T_s)+\frac{1}{n}\right)$$

which implies

$$\lambda \ge \frac{\varepsilon'-\frac{1}{n}}{I_0+\varepsilon'} > 0 \tag{3.57}$$

for sufficiently large $n$.

*Case 2:* (3.56) *does not hold, but* (3.55) *holds.* Then we obtain the inequalities

$$\begin{aligned}
&\inf_{s\in\mathscr{S}} I(X_{\tau_1}Y_{\tau_2}\wedge T_s|U)+\varepsilon' \\
&\le \inf_{s\in\mathscr{S}} I(X_{\tau_1}Y_{\tau_2}\wedge T_s|U)+\frac{1}{n}+\frac{\lambda}{n}\log K_0K_1K_2 \\
&\le \inf_{s\in\mathscr{S}} I(X_{\tau_1}Y_{\tau_2}\wedge T_s|U)+\frac{1}{n}+\lambda\Big(\inf_{s\in\mathscr{S}} I(X_{\tau_1}Y_{\tau_2}\wedge T_s)+\varepsilon'\Big).
\end{aligned}$$

This is equivalent to (3.57).

*Case 3:* (3.56) *does not hold, but* (3.53) *or* (3.54) *hold.* This can be treated like Case 2 and also gives (3.57).

The validity of (3.57) in the presence of (3.48) proves the weak converse for the compound MAC with common message.

# 4. The Compound MAC with Conferencing Encoders

## 4.1. Introduction

This chapter treats another communication model based on a compound MAC, namely the compound MAC with conferencing encoders. As in Section 2.2 the encoders may hold a Willems conference. However, this conference here may also concern CSIT. Related results can be found in [41], which characterizes the capacity region of compound MACs, both discrete and Gaussian, with two possible channel realizations and full CSI at the receiver. In the same paper, the connection with the interference channel was exploited by finding its capacity region if only one transmitter can send information to the other (unidirectional cooperation) and if the channel is in the strong interference regime. The Gaussian MAC with stochastic interference known non-causally at the encoders is an example of a channel whose state is determined stochastically. Its capacity region is derived in [16], the conference may also include information about the interference.

We characterize the capacity region of the general compound MAC with conferencing encoders with partial CSI. As in [63, 64], we show that every rate contained in the capacity region can be achieved using a one-shot Willems conference. We determine how large the conferencing capacities need to be in order to achieve the full-cooperation sum rate and the full-cooperation capacity region, respectively.

## 4.2. The Compound MAC with Conferencing Encoders

We now formalize the problem treated in this chapter.

**Definition 4.1.** Let $n$ be a positive integer and $C_1, C_2 \geq 0$. A *deterministic* $(n, C_1, C_2, T_1, T_2, R)$-*code*$_{\mathrm{CONF}}$ *with alphabets* $\mathscr{X}, \mathscr{Y}, \mathscr{T}$ is a quintuple $(c_1, c_2, f_1, f_2, \varphi)$ of functions, where

$$(c_1, c_2) : ([K_1] \times T_1) \times ([K_2] \times T_2) \to [J_1] \times [J_2]$$

is an $(n, C_1, C_2)$-Willems conference for positive integers $K_1, K_2$ and

$$\begin{aligned} f_1 &: [K_1] \times T_1 \times [J_2] \to \mathscr{X}^n, \\ f_2 &: [K_2] \times T_2 \times [J_1] \to \mathscr{Y}^n, \\ \varphi &: \mathscr{T}^n \times R \to [K_1] \times [K_2]. \end{aligned}$$

$f_1, f_2$ are the *encoding functions*, $\varphi$ is the *decoding function*, $n$ is the *blocklength*, and $(K_1, K_2)$ is the *codelength pair* of $(c_1, c_2, f_1, f_2, \varphi)$.

We denote the set of deterministic $(n, C_1, C_2, T_1, T_2, R)$-codes$_{\mathrm{CONF}}$ with codelength pair $(K_1, K_2)$ by $\Gamma_{\mathrm{CONF}}(n, K_1, K_2, T_1, T_2, R)$.

As usual, the alphabets will generally be clear from the context and do not have to be mentioned explicitly. The *codewords* of an $(n, C_1, C_2, T_1, T_2, R)$-code$_{\mathrm{CONF}}$ $\gamma = (c_1^\gamma, c_2^\gamma, f_1^\gamma, f_2^\gamma, \varphi^\gamma)$ are called $\mathbf{x}_{k_1k_2}^{\tau_1\tau_2}(\gamma), \mathbf{y}_{k_1k_2}^{\tau_1\tau_2}(\gamma)$ and are formed according to the rule

$$\begin{aligned}\mathbf{x}_{k_1k_2}^{\tau_1\tau_2}(\gamma) &:= f_1^\gamma(k_1, \tau_1, c_2(k_1, \tau_1, k_2, \tau_2)),\\ \mathbf{y}_{k_1k_2}^{\tau_1\tau_2}(\gamma) &:= f_2^\gamma(k_2, \tau_2, c_1(k_1, \tau_1, k_2, \tau_2)).\end{aligned}$$

The decoding sets are denoted by $D_{k_1k_2}^\rho(\gamma)$. The interpretation is the same as for the compound MAC with common message: the combined CSI $(\tau_1, \tau_2, \rho)$ together with a message pair $(k_1, k_2)$ determines the codewords $\mathbf{x}_{k_1k_2}^{\tau_1\tau_2}(\gamma), \mathbf{y}_{k_1k_2}^{\tau_1\tau_2}(\gamma)$ and the decoding set $D_{k_1k_2}^\rho(\gamma)$.

**Definition 4.2.** Let $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ be a compound MAC and $\gamma$ a deterministic $(n, C_1, C_2, T_1, T_2, R)$-code$_{\mathrm{CONF}}$. The *C-average error* of $\gamma$ is defined as

$$\bar{e}^{\mathrm{Cp}}(\gamma, \mathscr{W}, T_1, T_2, R) := \sup_{\tau_1,\tau_2,\rho} \sup_{s\in\mathscr{S}_{\tau_1\tau_2}^\rho} \frac{1}{K_1K_2} \sum_{k_1,k_2} W_s^{\otimes n}\big(D_{k_1k_2}^\rho(\gamma)^c | \mathbf{x}_{k_1k_2}^{\tau_1\tau_2}(\gamma), \mathbf{y}_{k_1k_2}^{\tau_1\tau_2}(\gamma)\big).$$

Its *C-maximal error* is defined as

$$e^{\mathrm{Cp}}(\gamma, \mathscr{W}, T_1, T_2, R) := \sup_{\tau_1,\tau_2,\rho} \sup_{s\in\mathscr{S}_{\tau_1\tau_2}^\rho} \max_{k_1,k_2} W_s^{\otimes n}\big(D_{k_1k_2}^\rho(\gamma)^c | \mathbf{x}_{k_1k_2}^{\tau_1\tau_2}(\gamma), \mathbf{y}_{k_1k_2}^{\tau_1\tau_2}(\gamma)\big).$$

As for the compound MAC with common message, we also consider random codes$_{\mathrm{CONF}}$.

**Definition 4.3.** Let $n, K_1, K_2$ be positive integers and let $C_1, C_2 \geq 0$. A random variable $G$ on $\Gamma_{\mathrm{CONF}}(n, K_1, K_2, C_1, C_2, T_1, T_2, R)$ is called a *random* $(n, C_1, C_2, T_1, T_2, R)$-*code*$_{\mathrm{CONF}}$. The *blocklength* and the *codelength pair* are defined analogous to the deterministic case.

**Definition 4.4.** Let $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ be a compound MAC and $G$ a random $(n, C_1, C_2, T_1, T_2, R)$-code$_{\mathrm{CM}}$. The *C-average error* of $G$ for $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ is defined as

$$\begin{aligned}&\bar{e}^{\mathrm{Cp,r}}(G, \mathscr{W}, T_1, T_2, R)\\ &\qquad := \sup_{\tau_1,\tau_2,\rho} \sup_{s\in\mathscr{S}_{\tau_1\tau_2}^\rho} \frac{1}{K_1K_2} \sum_{k_1,k_2} \sum_{\gamma} W_s^{\otimes n}\big(D_{k_1k_2}(\gamma)^c | \mathbf{x}_{k_1k_2}^{\tau_1\tau_2}(\gamma), \mathbf{y}_{k_1k_2}^{\tau_1\tau_2}(\gamma)\big) P_G(\gamma).\end{aligned}$$

Its *C-maximal error* for $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ is defined as

$$\begin{aligned}&e^{\mathrm{Cp,r}}(G, \mathscr{W}, T_1, T_2, R)\\ &\qquad := \sup_{\tau_1,\tau_2,\rho} \sup_{s\in\mathscr{S}_{\tau_1\tau_2}^\rho} \max_{k_1,k_2} \sum_{\gamma} W_s^{\otimes n}\big(D_{k_1k_2}(\gamma)^c | \mathbf{x}_{k_1k_2}^{\tau_1\tau_2}(\gamma), \mathbf{y}_{k_1k_2}^{\tau_1\tau_2}(\gamma)\big) P_G(\gamma).\end{aligned}$$

**Definition 4.5.** 1) A pair $(R_1, R_2)$ of nonnegative real numbers is called a *deterministically CONF-achievable rate pair for* $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ *with conferencing capacities* $C_1, C_2 \geq 0$ *under the average (maximal) error criterion* if for every $\varepsilon > 0$ and $\lambda \in (0,1)$ and for $n \geq n_0(\lambda, \varepsilon)$, there exists a deterministic $(n, C_1, C_2, T_1, T_2, R)$-$\mathrm{code}_{\mathrm{CONF}}$ $\gamma$ with $\bar{e}^{\mathrm{Cp}}(\gamma, \mathscr{W}, T_1, T_2, R) \leq \lambda$ ($e^{\mathrm{Cp}}(\gamma, \mathscr{W}, T_1, T_2, R) \leq \lambda$) and which satisfies

$$\frac{1}{n} \log K_\nu \geq R_\nu - \varepsilon \quad (\nu = 1, 2).$$

The set of deterministically CONF-achievable rates under the average (maximal) error criterion is called the *deterministic CONF-capacity region of* $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ *with conferencing capacities* $C_1, C_2$ *under the average (maximal) error criterion* and denoted by $\overline{\mathscr{C}}^{\mathrm{Cp}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, R)$($\mathscr{C}^{\mathrm{Cp}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, R)$).

2) A pair $(R_1, R_2)$ of nonnegative real numbers is called a *randomly CONF-achievable rate pair for* $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ *with conferencing capacities* $C_1, C_2 \geq 0$ *under the average (maximal) error criterion* if for every $\varepsilon > 0$ and $\lambda \in (0,1)$ and for $n \geq n_0(\lambda, \varepsilon)$, there is a random $(n, C_1, C_2, T_1, T_2, R)$-$\mathrm{code}_{\mathrm{CONF}}$ $G$ for $(\mathscr{W}, T_1, T_2, R)$ with $\bar{e}^{\mathrm{Cp,r}}(G, \mathscr{W}, T_1, T_2, R) \leq \lambda$ ($e^{\mathrm{Cp,r}}(G, \mathscr{W}, T_1, T_2, R) \leq \lambda$) and which satisfies

$$\frac{1}{n} \log K_\nu \geq R_\nu - \varepsilon \quad (\nu = 1, 2).$$

The set of randomly CONF-achievable rates under the average (maximal) error criterion is called the *random CONF-capacity region of* $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ *with conferencing capacities* $C_1, C_2$ *under the average (maximal) error criterion* and denoted by $\overline{\mathscr{C}}^{\mathrm{Cp,r}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, R)$ ($\mathscr{C}^{\mathrm{Cp,r}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, R)$).

The capacity regions have a different structure depending on whether the conferencing capacities are both positive or one equals zero. As $T_1, T_2$ are finite, if both are positive, the encoders can completely inform each other about their CSI partitions. If only, say, $C_1$ is positive, then the first encoder will remain ignorant of the second encoder's CSI.

We define the partition $T_1 \wedge T_2$ as the maximal common refinement of $T_1$ and $T_2$, i.e.

$$T_1 \wedge T_2 := \{\tau_1 \cap \tau_2 : \tau_1 \in T_1, \tau_2 \in T_2\}.$$

Then we set

$$(\tilde{T}, \tilde{T}_1, \tilde{T}_2) := \begin{cases} (T_1 \wedge T_2, \{\mathscr{S}\}, \{\mathscr{S}\}) & \text{if } C_1, C_2 > 0, \\ (T_1, \{\mathscr{S}\}, T_2) & \text{if } C_1 > 0, C_2 = 0, \\ (T_2, T_1, \{\mathscr{S}\}) & \text{if } C_1 = 0, C_2 > 0. \end{cases} \tag{4.1}$$

This definition induces a natural function $f_2 : \mathscr{S} \to \tilde{T} \times \tilde{T}_1 \times \tilde{T}_2$ and a corresponding partition of $\mathscr{S}$ into subsets $\mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2} := f_2^{-1}(\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2)$. $f_2$ also gives rise to a set of probability measures $\Pi_2(\mathscr{W}, \tilde{T}, \tilde{T}_1, \tilde{T}_2)$ as in Definition 3.5. For every $p_s \in p \in \Pi_2(\mathscr{W}, \tilde{T}, \tilde{T}_1, \tilde{T}_2)$ we define $\mathscr{R}_{\mathrm{CONF}}(p_s, C_1, C_2)$ analogous to (2.15)-(2.17) and set

$$\widehat{\mathscr{R}}_{\mathrm{CONF}}(p, C_1, C_2) := \bigcap_{s \in \mathscr{S}} \mathscr{R}_{\mathrm{CONF}}(p_s, C_1, C_2) = \bigcap_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \bigcap_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} \mathscr{R}_{\mathrm{CONF}}(p_s, C_1, C_2)$$

and

$$\mathscr{C}_2(\mathscr{W}, C_1, C_2, \tilde{T}, \tilde{T}_1, \tilde{T}_2) := \operatorname{closure}\Bigg( \bigcup_{p \in \Pi_2(\mathscr{W}, \tilde{T}, \tilde{T}_1, \tilde{T}_2)} \widehat{\mathscr{R}}_{\mathrm{CONF}}(p, C_1, C_2) \Bigg).$$

**Theorem 4.6.** *Let* $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ *be a compound MAC with conferencing capacities* $C_1, C_2 \geq 0$*. Then*

$$\begin{aligned} \overline{\mathscr{C}}^{\mathrm{Cp}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, R) &= \overline{\mathscr{C}}^{\mathrm{Cp,r}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, R) \\ &= \mathscr{C}^{\mathrm{Cp,r}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, R) = \mathscr{C}_2(\mathscr{W}, C_1, C_2, \tilde{T}, \tilde{T}_1, \tilde{T}_2), \end{aligned}$$

*where* $(\tilde{T}, \tilde{T}_1, \tilde{T}_2)$ *is defined as in* (4.1)*. In every case, the capacity region can be CONF-achieved using only one-shot Willems conferencing. The cardinality of* $\mathscr{U}$ *can be restricted to be at most* $\min\{|\mathscr{X}||\mathscr{Y}| + 2, |\mathscr{T}| + 3\}$*. For all cases there exists a weak converse.*

Again, the definition of weak converse is the same as in Definition 2.6 if one replaces $\mathscr{C}$ by $\overline{\mathscr{C}}^{\mathrm{Cp}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, R)$ or $\overline{\mathscr{C}}^{\mathrm{Cp,r}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, R)$ or $\mathscr{C}^{\mathrm{Cp,r}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, R)$, respectively, and adapts to the right dimension.

*Remark* 4.1. $\mathscr{C}_2(\mathscr{W}, C_1, C_2, \tilde{T}, \tilde{T}_1, \tilde{T}_2)$ is convex by the concavity of mutual information in the input distribution. The bounds on $|\mathscr{U}|$ follow in the same way as in [63].

*Remark* 4.2. Like the proof of Theorem 3.11, the proof of Theorem 4.6 shows that in all three cases, the capacity regions can be achieved with codes whose error probability tends to zero at exponential speed. See Remark 3.2. Moreover the random $\mathrm{codes}_{\mathrm{CONF}}$ can be chosen such that their deterministic component $\mathrm{codes}_{\mathrm{CONF}}$ share the same Willems conference, both to achieve $\overline{\mathscr{C}}^{\mathrm{Cp,r}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, R)$ and $\mathscr{C}^{\mathrm{Cp,r}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, R)$.

*Remark* 4.3. As for the compound MAC with common message, the capacity regions of all above cases are independent of the CSIR partition $R$. Further, it only depends on $(\tilde{T}, \tilde{T}_1, \tilde{T}_2)$, not on the single CSIT partitions $T_1$ or $T_2$. That means that all $T_1, T_2$ with the same $(\tilde{T}, \tilde{T}_1, \tilde{T}_2)$ lead to the same capacity region.

*Remark* 4.4. Assume that $\tilde{T} = T_1 \wedge T_2$. Then $\mathscr{C}_2(\mathscr{W}, C_1, C_2, T_1 \wedge T_2, \{\mathscr{S}\}, \{\mathscr{S}\})$ exhibits a different behavior than $C_1(\mathscr{W}, T_1, T_2)$ as far as the behavior under different CSI partitions is concerned. If $T_1 = T_2 = \mathscr{S}$ (and still assuming $C_1, C_2 > 0$), then

$$\mathscr{C}_2(\mathscr{W}, C_1, C_2, \mathscr{S}, \{\mathscr{S}\}, \{\mathscr{S}\}) = \operatorname{closure}\Bigg( \bigcap_{s \in \mathscr{S}} \bigcup_{p \in \Pi(W_s)} \mathscr{R}_{\mathrm{CONF}}(p, C_1, C_2) \Bigg).$$

By Definition 2.7, $\Pi(W_s)$ consists of probability measures on $\mathscr{U} \times \mathscr{X} \times \mathscr{Y} \times \mathscr{T}$ for auxiliary sets $\mathscr{U}$. Thus in this case, the compound MAC with conferencing encoders exhibits the same behavior as the single-sender compound channel with perfect channel state information at the encoder whose capacity equals the minimum of the capacities of the component discrete memoryless channels. Heuristically, this is due to the fact that

the common message generated by conferencing may depend on the encoders' joint CSI. In contrast, the common message in the previous chapter is independent of coding and thus of CSI.

In analogy to the discussion after Theorem 2.13 in Lemmas 2.15 and 2.16, we ask how large $C_1$ and $C_2$ need to be in order for infinite-cooperation performance to be achieved. We assume $C_1, C_2 > 0$, so the joint channel state information equals $T_1 \wedge T_2$. Denote the maximally achievable sum rate by $\mathscr{C}_+^{\mathrm{Cp}}(\mathscr{W}, T_1 \wedge T_2)$ (by Theorem 4.6 it is independent of $R$), so

$$\mathscr{C}_+^{\mathrm{Cp}}(\mathscr{W}, T_1 \wedge T_2) = \max_{p \in \Pi_2(\mathscr{W}, T_1 \wedge T_2, \{\mathscr{S}\}, \{\mathscr{S}\})} \inf_{s \in \mathscr{S}} I(T_s \wedge X_\tau Y_\tau), \tag{4.2}$$

where we denote the elements of $T_1 \wedge T_2$ by $\tau$. As for the discrete memoryless MAC with conferencing encoders, we consider both the case that the maximal sum rate

$$\max_{p \in \Pi_2(\mathscr{W}, T_1 \wedge T_2, \{\mathscr{S}\}, \{\mathscr{S}\})} \inf_{s \in \mathscr{S}} \min\big\{ I(T_s \wedge X_\tau Y_\tau), I(T_s \wedge X_\tau Y_\tau | U_\tau) + C_1 + C_2 \big\}$$

equals $\mathscr{C}_+^{\mathrm{Cp}}(\mathscr{W}, T_1 \wedge T_2)$ and that the complete triangular region only restricted by the coordinate axes and $\mathscr{C}_+^{\mathrm{Cp}}(\mathscr{W}, T_1 \wedge T_2)$ is CONF-achieved.

We denote the subset of $\mathscr{C}_+^{\mathrm{Cp}}(\mathscr{W}, T_1 \wedge T_2)$-achieving $p \in \Pi_2(\mathscr{W}, T_1 \wedge T_2, \{\mathscr{S}\}, \{\mathscr{S}\})$ by $\mathscr{M}$. We also assume that $C_1, C_2 > 0$, otherwise the comparability would be even worse.

**Lemma 4.7.** *$\mathscr{C}_+^{\mathrm{Cp}}(\mathscr{W}, T_1 \wedge T_2)$ is CONF-achieved by the maximal sum rate of $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ with conferencing capacities $C_1, C_2 > 0$ if and only if*

$$\begin{aligned} C_1 + C_2 &\geq \mathscr{C}_+^{\mathrm{Cp}}(\mathscr{W}, T_1 \wedge T_2) \\ &\quad - \max_{p \in \mathscr{M}} \min\big\{ \inf_{s \in \mathscr{S}} I(T_s \wedge X_\tau Y_\tau | U_\tau), \inf_{s \in \mathscr{S}} I(T_s \wedge X_\tau | Y_\tau U_\tau) + \inf_{s \in \mathscr{S}} I(T_s \wedge Y_\tau | X_\tau U_\tau) \big\}. \end{aligned} \tag{4.3}$$

*Proof.* We abbreviate $\mathscr{C} := \mathscr{C}_+^{\mathrm{Cp}}(\mathscr{W}, T_1 \wedge T_2)$. First assume that there is an $\varepsilon > 0$ such that

$$\begin{aligned} C_1 + C_2 &= \mathscr{C} \\ &\quad - \max_{p \in \mathscr{M}} \min\big\{ \inf_{s \in \mathscr{S}} I(T_s \wedge X_\tau Y_\tau | U_\tau), \inf_{s \in \mathscr{S}} I(T_s \wedge X_\tau | Y_\tau U_\tau) + \inf_{s \in \mathscr{S}} I(T_s \wedge Y_\tau | X_\tau U_\tau) \big\} - \varepsilon. \end{aligned} \tag{4.4}$$

We have to show that the maximal sum rate of $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ does not achieve $\mathscr{C}$. If $p \in \Pi_2(\mathscr{W}, T_1 \wedge T_2, \{\mathscr{S}\}, \{\mathscr{S}\}) \setminus \mathscr{M}$, then

$$\inf_{s \in \mathscr{S}} \min\big\{ I(T_s \wedge X_\tau Y_\tau), I(T_s \wedge X_\tau Y_\tau | U_\tau) + C_1 + C_2 \big\} \leq \inf_{s \in \mathscr{S}} I(T_s \wedge X_\tau Y_\tau) < \mathscr{C}. \tag{4.5}$$

Now fix a $p \in \mathscr{M}$. There are two cases.

*Case 1:*

$$\inf_{s \in \mathscr{S}} I(T_s \wedge X_\tau Y_\tau | U_\tau) \leq \inf_{s \in \mathscr{S}} I(T_s \wedge X_\tau | Y_\tau U_\tau) + \inf_{s \in \mathscr{S}} I(T_s \wedge Y_\tau | X_\tau U_\tau).$$

In this case, we have by (4.4)

$$C_1 + C_2 \leq \mathscr{C} - \inf_{s \in \mathscr{S}} I(T_s \wedge X_\tau Y_\tau | U_\tau) - \varepsilon.$$

Thus for any $s \in \mathscr{S}$,

$$\begin{aligned}
&\min\bigl\{I(T_s \wedge X_\tau Y_\tau), I(T_s \wedge X_\tau Y_\tau|U_\tau) + C_1 + C_2\bigr\} \\
&\leq I(T_s \wedge X_\tau Y_\tau|U_\tau) + C_1 + C_2 \\
&\leq I(T_s \wedge X_\tau Y_\tau|U_\tau) - \inf_{s\in\mathscr{S}} I(T_s \wedge X_\tau Y_\tau|U_\tau) + \mathscr{C} - \varepsilon.
\end{aligned}$$

Taking the infimum over $s \in \mathscr{S}$ on both sides gives

$$\inf_{s\in\mathscr{S}} \min\bigl\{I(T_s \wedge X_\tau Y_\tau), I(T_s \wedge X_\tau Y_\tau|U_\tau) + C_1 + C_2\bigr\} \leq \mathscr{C} - \varepsilon.$$

Thus there is no pair $(R_1, R_2) \in \mathscr{R}_{\mathrm{CONF}}(p, C_1, C_2)$ with $R_1 + R_2 = \mathscr{C}$.

*Case 2:*

$$\inf_{s\in\mathscr{S}} I(T_s \wedge X_\tau Y_\tau|U_\tau) > \inf_{s\in\mathscr{S}} I(T_s \wedge X_\tau|Y_\tau U_\tau) + \inf_{s\in\mathscr{S}} I(T_s \wedge Y_\tau|X_\tau U_\tau).$$

In this case let $(R_1, R_2) \in \mathscr{R}_{\mathrm{CONF}}(p)$. The single-rate bounds on $R_1$ and $R_2$ imply

$$\begin{aligned}
R_1 + R_2 &\leq \inf_{s\in\mathscr{S}} I(T_s \wedge X_\tau|Y_\tau U_\tau) + \inf_{s\in\mathscr{S}} I(T_s \wedge Y_\tau|X_\tau U_\tau) \\
&< \inf_{s\in\mathscr{S}} I(T_s \wedge X_\tau Y_\tau|U_\tau) \\
&\leq \mathscr{C}.
\end{aligned}$$

Thus if $p \notin \mathscr{M}$, $\mathscr{C}$ is not achieved by (4.5), and if $p \in \mathscr{M}$, then both cases above show that $\mathscr{C}$ is not obtained either.

It remains to prove the other direction, i.e. if $C_1 + C_2$ satisfies (4.3), then the maximal sum rate of $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ equals $\mathscr{C}$. Assume that $C_1 + C_2$ satisfies (4.3). Let $p$ maximize

$$\min\bigl\{\inf_{s\in\mathscr{S}} I(T_s \wedge X_\tau Y_\tau|U_\tau), \inf_{s\in\mathscr{S}} I(T_s \wedge X_\tau|Y_\tau U_\tau) + \inf_{s\in\mathscr{S}} I(T_s \wedge Y_\tau|X_\tau U_\tau)\bigr\}.$$

Due to the choice of $p$, we have for every $s \in \mathscr{S}$

$$\begin{aligned}
&I(T_s \wedge X_\tau Y_\tau|U_\tau) + C_1 + C_2 \\
&\geq I(T_s \wedge X_\tau Y_\tau|U_\tau) - \inf_{s\in\mathscr{S}} I(T_s \wedge X_\tau Y_\tau|U_\tau) + \mathscr{C} \\
&\geq \mathscr{C}.
\end{aligned}$$

It remains to show that the single-rate bounds do not exclude the existence of a pair $(R_1, R_2) \in \mathscr{R}_{\mathrm{CONF}}(p, C_1, C_2)$ with $R_1 + R_2 = \mathscr{C}$. But obviously by the assumption on $C_1 + C_2$,

$$\inf_{s\in\mathscr{S}} I(T_s \wedge X_\tau|Y_\tau U_\tau) + C_1 + \inf_{s\in\mathscr{S}} I(T_s \wedge Y_\tau|X_\tau U_\tau) + C_2 \geq \mathscr{C}.$$

This completes the proof. □

Next we would like to CONF-achieve the complete total cooperation region

$$\big\{(R_1,R_2): 0\le R_1+R_2\le \mathscr{C}_+^{\mathrm{Cp}}(\mathscr{W},T_1\wedge T_2,R)\big\}. \tag{4.6}$$

**Lemma 4.8.** *The full cooperation region* (4.6) *is CONF-achieved if and only if both*

$$C_1\ge \mathscr{C}_+^{\mathrm{Cp}}(\mathscr{W},T_1\wedge T_2,R)-\max_{p\in\Pi_2(T_1\wedge T_2,\{\mathscr{S}\},\{\mathscr{S}\})}\inf_{s\in\mathscr{S}} I(T_s\wedge X_\tau|Y_\tau),$$

*and*

$$C_2\ge \mathscr{C}_+^{\mathrm{Cp}}(\mathscr{W},T_1\wedge T_2,R)-\max_{p\in\Pi_2(T_1\wedge T_2,\{\mathscr{S}\},\{\mathscr{S}\})}\inf_{s\in\mathscr{S}} I(T_s\wedge Y_\tau|X_\tau).$$

The proof of this lemma is analogous to that of Lemma 2.16. Note that if, say, $C_1=0$, then $\mathscr{C}_+^{\mathrm{Cp}}(\mathscr{W},T_1,T_2,R)$ might not be CONF-achievable by the maximal sum rate because

$$\max_{p\in\Pi_2(\mathscr{W},T_2,T_1,\{\mathscr{S}\})} I(T_s\wedge X_{\tau_1}Y_{\tau_2})<\max_{p\in\Pi_2(\mathscr{W},T_1\wedge T_2,\{\mathscr{S}\},\{\mathscr{S}\})} I(T_s\wedge X_\tau Y_\tau)$$

might hold. But if $C_1,C_2>0$, as is the case for the Discrete Memoryless MAC with Conferencing Encoders, infinite-capacity cooperation is neither necessary in order to CONF-achieve the full-cooperation sum rate nor to CONF-achieve the full-cooperation rate region.

*Example* 1. In order to visualize the behavior of $\mathscr{C}_2(\mathscr{W},C_1,C_2,\tilde{T},\tilde{T}_1,\tilde{T}_2)$ under varying conferencing capacities $C_1,C_2$, we introduce a simple compound MAC. Assume $\mathscr{S}=\mathscr{X}=\mathscr{Y}=\mathscr{T}=\{0,1\}$. Let $\mathscr{W}$ consist of the stochastic matrices

$$W_0=\begin{pmatrix}0.9 & 0.1\\ 0.4 & 0.6\\ 0.6 & 0.4\\ 0 & 1\end{pmatrix}\quad\text{and}\quad W_1=\begin{pmatrix}0.9 & 0.1\\ 0.6 & 0.4\\ 0.4 & 0.6\\ 0 & 1\end{pmatrix},$$

where the output distribution corresponding to the input combination $(x,y)$ can be found in row $2x+y+1$.

In Figure 4.1, the capacity regions are pictured for different values of $C_1,C_2$. The regions denoted by $W_0$ and $W_1$ show the capacity regions of the MACs given by $W_0$ and $W_1$, respectively, without cooperation. Their intersection is the capacity region of the compound channel consisting of $W_0$ and $W_1$ where the exact channel state is known at both senders, i.e. equal to $\mathscr{C}_2(\mathscr{W},0,0,\mathscr{S},\{\mathscr{S}\},\{\mathscr{S}\})$. If the senders do not have any CSI, we have $T_1=T_2=\{\{0,1\}\}$. The capacity region denoted by "no coop." shows the case of no CSI and $C_1=C_2=0$, which is $\mathscr{C}_2(\mathscr{W},0,0,\{\mathscr{S}\},\{\mathscr{S}\},\{\mathscr{S}\})$. Note that absence of CSIT makes the region strictly smaller. The triangular region denoted by "full coop." shows the region obtained if $C_1,C_2$ exceed the thresholds derived in Lemma 4.8, so we have

$$C_1\ge 0.4154,\qquad C_2\ge 0.4123.$$

The intermediate case was chosen such that $C_1=C_2=C$, where $C$ equals one half times the threshold derived in Lemma 4.7. For $\mathscr{W}$ this means

$$C_1=C_2=C=0.2613=\frac{1}{2}\mathscr{C}_+^{\mathrm{Cp}}(\mathscr{W},\{\mathscr{S}\}).$$

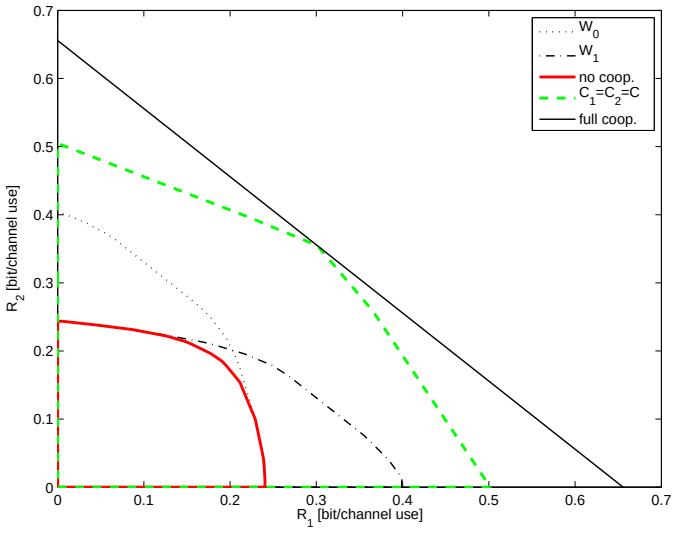


Figure 4.1.: Two discrete MAC capacity regions and those for the resulting compound MAC at no cooperation, maximal sum-rate achieving cooperation and full cooperation.

## 4.3. The Direct Part

As in the common message problem, we may without loss of generality assume that the receiver has no CSI. In contrast to Willems' approach to the discrete memoryless MAC with conferencing encoders, we cannot reduce the direct part of the proof of Theorem 4.6 to Theorem 3.11 for general $T_1, T_2$. Thus we will use Lemma 3.14 instead, so we start with proving

$$\mathscr{C}_2(\mathscr{W}, C_1, C_2, \tilde{T}, \tilde{T}_1, \tilde{T}_2) \subset \mathscr{C}^{\mathrm{Cp,r}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, \{\mathscr{S}\}) \tag{4.7}$$

Let $p \in \Pi_2(\mathscr{W}, \tilde{T}, \tilde{T}_1, \tilde{T}_2)$ (recall (4.1)) and $\tilde{J}_1, \tilde{J}_2, \tilde{K}_1, \tilde{K}_2$ positive integers. Setting $\tilde{J} := \tilde{J}_1\tilde{J}_2$, we do random coding with the generalized $(\tilde{J}, \tilde{K}_1, \tilde{K}_2)$-half lattice

$$\big\{(U^{\tilde{\tau}}_{\tilde{j}}, X^{\tilde{\tau}\tilde{\tau}_1}_{\tilde{j}\tilde{k}_1}, Y^{\tilde{\tau}\tilde{\tau}_2}_{\tilde{j}\tilde{k}_2}) : (\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2, \tilde{j}, \tilde{k}_1, \tilde{k}_2) \in \tilde{T} \times \tilde{T}_1 \times \tilde{T}_2 \times [\tilde{J}] \times [\tilde{K}_1] \times [\tilde{K}_2]\big\} \tag{4.8}$$

on $\mathscr{U}^n \times \mathscr{X}^n \times \mathscr{Y}^n$ based on the $n$-th memoryless extension $\pi^{\otimes n}$ of the $(\tilde{T}, \tilde{T}_1, \tilde{T}_2)$-input probability $\pi$ induced by $p$. With

$$\tilde{T}'_\nu := \begin{cases} \{\mathscr{S}\} & \text{if } \tilde{T}_\nu = T_\nu, \\ T_\nu & \text{if } \tilde{T}_\nu = \{\mathscr{S}\}, \end{cases} \qquad (\nu = 1, 2), \tag{4.9}$$

we define the following numbers:

$$J_\nu := \tilde{J}_\nu |\tilde{T}'_\nu| \qquad (\nu = 1, 2),$$
$$K_\nu := \tilde{J}_\nu \tilde{K}_\nu \qquad (\nu = 1, 2).$$

Now we assume that

$$\frac{1}{n} \log J_\nu \leq C_\nu \quad (\nu = 1, 2). \tag{4.10}$$

Since $\tilde{T}'_\nu$ is finite, this condition can be satisfied if $n$ is sufficiently large and $\tilde{J}_\nu$ chosen accordingly. Then (4.8) defines a random $(n, C_1, C_2, T_1, T_2, \{\mathscr{S}\})$-code$_{\mathrm{CONF}}$ as follows. For $\nu = 1, 2$, we identify $T_\nu$ with $\tilde{T}_\nu \times \tilde{T}'_\nu$. Also note that $\tilde{T} = \tilde{T}'_1 \wedge \tilde{T}'_2$. Further we can identify

- $[K_\nu]$ with $[\tilde{J}_\nu] \times [\tilde{K}_\nu]$, and
- $[J_\nu]$ with $[\tilde{J}_\nu] \times \tilde{T}'_\nu$.

The $[K_\nu]$ take the role of the message sets. If encoder $\nu \in \{1, 2\}$ is given the message $k_\nu = (\tilde{j}_\nu, \tilde{k}_\nu)$ and CSIT instance $\tau_\nu = (\tilde{\tau}_\nu, \tilde{\tau}'_\nu)$, it sends $c_\nu(k_\nu, \tau_\nu) = (\tilde{j}_\nu, \tilde{\tau}'_\nu) \in [\tilde{J}_\nu] \times \tilde{T}'_\nu = [J_\nu]$ to the other encoder. Due to (4.10) the one-shot conferencing functions $c_1, c_2$ thus defined are admissible and form a $(n, C_1, C_2)$-Willems conference

$$(c_1, c_2) : ([K_1] \times T_1) \times ([K_2] \times T_2) \to [J_1] \times [J_2].$$

Further given a message $k_1 = (\tilde{j}_1, \tilde{k}_1) \in [K_1]$, CSIT instance $\tau_1 = (\tilde{\tau}_1, \tilde{\tau}'_1) \in T_1$, and a conferencing result $j_2 = (\tilde{j}_2, \tilde{\tau}'_2) \in [J_2]$, the first encoder decides to use the random codeword $f_1(k_1, \tau_1, j_2) = X^{\tilde{\tau}\tilde{\tau}_1}_{\tilde{j}_1 \tilde{j}_2 \tilde{k}_1}$, where $\tilde{\tau} = \tilde{\tau}'_1 \cap \tilde{\tau}'_2$. The random encoding function $f_2$ of encoder 2 is defined in an analogous way. Altogether this defines the encoding process of a random code$_{\mathrm{CONF}}$. The decoding sets are defined as before Lemma 3.14 using $[\tilde{J}] \times [\tilde{K}_1] \times [\tilde{K}_2] = [K_1] \times [K_2]$. Note that the component codes$_{\mathrm{CONF}}$ of this random code$_{\mathrm{CONF}}$ share the same Willems conference $(c_1, c_2)$.

Lemma 3.14 now implies that the maximal error of the random code$_{\mathrm{CONF}}$ defined above tends to 0 at exponential speed if for some $\tilde{\zeta} > 0$

$$\frac{1}{n} \log(\tilde{J}\tilde{K}_1\tilde{K}_2) < \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(T_s \wedge X_{\tilde{\tau}\tilde{\tau}_1} Y_{\tilde{\tau}\tilde{\tau}_2}) - \tilde{\zeta},$$
$$\frac{1}{n} \log(\tilde{K}_1\tilde{K}_2) < \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(T_s \wedge X_{\tilde{\tau}\tilde{\tau}_1} Y_{\tilde{\tau}\tilde{\tau}_2} | U_{\tilde{\tau}}) - \tilde{\zeta},$$
$$\frac{1}{n} \log \tilde{K}_1 < \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(T_s \wedge X_{\tilde{\tau}\tilde{\tau}_1} | Y_{\tilde{\tau}\tilde{\tau}_2} U_{\tilde{\tau}}) - \tilde{\zeta},$$
$$\frac{1}{n} \log \tilde{K}_2 < \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(T_s \wedge Y_{\tilde{\tau}\tilde{\tau}_2} | X_{\tilde{\tau}\tilde{\tau}_1} U_{\tilde{\tau}}) - \tilde{\zeta},$$

where the $(\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2)$ range over $\tilde{T} \times \tilde{T}_1 \times \tilde{T}_2$. Due to (4.10) and the above identifications, if $\tilde{J}, \tilde{K}_1, \tilde{K}_2$ satisfy the above inequalities, they also satisfy

$$\begin{aligned}
\frac{1}{n} \log K_1 K_2 &< \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(T_s \wedge X_{\tilde{\tau}\tilde{\tau}_1} Y_{\tilde{\tau}\tilde{\tau}_2}) - \tilde{\zeta}, \\
\frac{1}{n} \log K_1 K_2 &< \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(T_s \wedge X_{\tilde{\tau}\tilde{\tau}_1} Y_{\tilde{\tau}\tilde{\tau}_2} | U_{\tilde{\tau}}) + C_1 + C_2 - \tilde{\zeta}, \\
\frac{1}{n} \log K_1 &< \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(T_s \wedge X_{\tilde{\tau}\tilde{\tau}_1} | Y_{\tilde{\tau}\tilde{\tau}_2} U_{\tilde{\tau}}) + C_1 - \tilde{\zeta}, \\
\frac{1}{n} \log K_2 &< \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(T_s \wedge Y_{\tilde{\tau}\tilde{\tau}_2} | X_{\tilde{\tau}\tilde{\tau}_1} U_{\tilde{\tau}}) + C_2 - \tilde{\zeta}.
\end{aligned}$$

This proves (4.7) and consequently also

$$\mathscr{C}_2(\mathscr{W}, C_1, C_2, \tilde{T}, \tilde{T}_1, \tilde{T}_2) \subset \overline{\mathscr{C}}^{\mathrm{Cp,r}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, \{\mathscr{S}\}).$$

Finally derandomization is done exactly as for the compound MAC with common message, so we also have

$$\mathscr{C}_2(\mathscr{W}, C_1, C_2, \tilde{T}, \tilde{T}_1, \tilde{T}_2) \subset \overline{\mathscr{C}}^{\mathrm{Cp}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, \{\mathscr{S}\}).$$

## 4.4. The Weak Converse

It is sufficient to show the weak converse for random codes and the average error criterion. Without loss of generality we may assume that the decoder has perfect CSIR, i.e. $R = \mathscr{S}$. We again use the sets $\tilde{T}_1', \tilde{T}_2'$ defined in (4.9) and have $\tilde{T} = \tilde{T}_1' \wedge \tilde{T}_2'$. Let $c = (c_1, c_2)$ be an $(n, C_1, C_2)$-Willems conference

$$(c_1, c_2) : ([K_1] \times T_1) \times ([K_2] \times T_2) \to [J_1] \times [J_2].$$

Fix a $(\tau_1, \tau_2) \in T_1 \times T_2$. Denote the mapping $(k_1, k_2) \mapsto c(k_1, \tau_1, k_2, \tau_2)$ by $\tilde{c}$. Then $\tilde{c}$ also is a $(n, C_1, C_2)$-Willems conference.

**Lemma 4.9.** *For every $j \in [J_1] \times [J_2]$ there are sets $\mathscr{K}_1^{(j)} \subset [K_1]$ and $\mathscr{K}_2^{(j)} \subset [K_2]$ such that $\tilde{c}^{-1}(j) = \mathscr{K}_1^{(j)} \times \mathscr{K}_2^{(j)}$.*

*Proof.* We first prove that $(\tilde{c}_1(k_1, k_2), \tilde{c}_2(k_1, k_2)) = (\tilde{c}_1(k_1', k_2'), \tilde{c}_2(k_1', k_2'))$ implies $(\tilde{c}_1(k_1', k_2), \tilde{c}_2(k_1', k_2)) = (\tilde{c}_1(k_1, k_2'), \tilde{c}_2(k_1, k_2'))$. In order to see this, for some $j := (j_1, j_2) \in [J_1] \times [J_2]$, let $k_1, k_1' \in [K_1]$ and $k_2, k_2' \in [K_2]$ satisfy

$$\big(\tilde{c}_1(k_1, k_2), \tilde{c}_2(k_1, k_2)\big) = \big(\tilde{c}_1(k_1', k_2'), \tilde{c}_2(k_1', k_2')\big) = j. \tag{4.11}$$

With the notation used in Chapter 2 for the definition of Willems conferencing, this is equivalent to saying that for every $\nu = 1, 2$ and $i = 1, \ldots, I$, if $j_\nu = (j_{\nu,1}, \ldots, j_{\nu,I})$

$$\tilde{c}^*_{\nu,i}(k_1, k_2) = \tilde{c}^*_{\nu,i}(k_1', k_2') = j_{\nu,i}. \tag{4.12}$$

We now show by induction over $i$ that (4.12) implies $\tilde{c}^*_{\nu,i}(k_1,k_2') = j_{\nu,i}$ for every $\nu$ and $i$. For $i=1$, we have

$$(\tilde{c}^*_{1,1}(k_1,k_2'),\tilde{c}^*_{2,1}(k_1,k_2')) = (\tilde{c}_{1,1}(k_1),\tilde{c}_{2,1}(k_2')) = (\tilde{c}^*_{1,1}(k_1,k_2),\tilde{c}^*_{2,1}(k_1',k_2')) = (j_{1,1},j_{2,1}).$$

Now assume that $\tilde{c}^*_{\nu,i'}(k_1,k_2') = j_{\nu,i'}$ for $\nu=1,2$ and $i'=1,\dots,i-1$. Then

$$\begin{aligned}
\big(\tilde{c}^*_{1,i}(k_1,k_2'),\tilde{c}^*_{2,i}(k_1,k_2')\big) &= \big(\tilde{c}_{1,i}(k_1,\tilde{c}^*_{2,i-1}(k_1,k_2')),\tilde{c}_{2,i}(k_2',\tilde{c}^*_{1,i-1}(k_1,k_2'))\big)\\
&= \big(\tilde{c}_{1,i}(k_1,j_{2,i-1}),\tilde{c}_{2,i}(k_2',j_{1,i-1})\big)\\
&= \big(\tilde{c}_{1,i}(k_1,\tilde{c}^*_{2,i-1}(k_1,k_2)),\tilde{c}_{2,i}(k_2',\tilde{c}^*_{1,i-1}(k_1',k_2'))\big)\\
&= \big(\tilde{c}^*_{1,i}(k_1,k_2),\tilde{c}^*_{2,i}(k_1',k_2')\big)\\
&= (j_{1,i},j_{2,i}).
\end{aligned}$$

Here, we used the induction hypothesis in the second equality and (4.12) in the third and the last equality. Altogether we have shown that (4.11) implies

$$(\tilde{c}_1(k_1,k_2'),\tilde{c}_2(k_1,k_2')) = (j_1,j_2),$$

so the first part of the proof is complete.

Now we define

$$\begin{aligned}
\mathscr{K}_1^{(j)} &:= \{k_1\in[K_1] : \tilde{c}(k_1,k_2') = j \text{ for some } k_2'\in[K_2]\},\\
\mathscr{K}_2^{(j)} &:= \{k_2\in[K_2] : \tilde{c}(k_1',k_2) = j \text{ for some } k_1'\in[K_1]\}.
\end{aligned}$$

Clearly, if $(k_1,k_2)\notin \mathscr{K}_1^{(j)}\times\mathscr{K}_2^{(j)}$, then $\tilde{c}(k_1,k_2)\neq j$. If $(k_1,k_2)\in\mathscr{K}_1^{(j)}\times\mathscr{K}_2^{(j)}$, choose $k_2'\in[K_2]$ and $k_1'\in[K_1]$ such that $\tilde{c}(k_1,k_2') = \tilde{c}(k_1',k_2) = j$. Then by the first part of the proof we immediately have $\tilde{c}(k_1,k_2)=j$. This completes the proof. □

Lemma 4.9 immediately implies the following corollary.

**Corollary 4.10** (Willems)**.** *Let $(M_1,M_2)$ be any random vector on $[K_1]\times[K_2]$ with $M_1$ independent of $M_2$ and let $(c_1,c_2)$ be a Willems conference*

$$(c_1,c_2) : ([K_1]\times T_1)\times([K_2]\times T_2)\to[J_1]\times[J_2].$$

*Then for any $(\tau_1,\tau_2)\in T_1\times T_2$, $M_1$ and $M_2$ are conditionally independent given $(c_1(M_1,\tau_1,M_2,\tau_2),c_2(M_1,\tau_1,M_2,\tau_2))$.*

We can now start with the proof of the weak converse. Let $G$ be a random $(n,C_1,C_2,T_1,T_2,\mathscr{S})$-code$_{\mathrm{CONF}}$ which satisfies

$$\left\|\frac{1}{n}(\log K_1,\log K_2) - \mathscr{C}_2(\mathscr{W},C_1,C_2,\tilde{T},\tilde{T}_1,\tilde{T}_2)\right\| > \varepsilon \tag{4.13}$$

for some norm $\|\cdot\|$. Every deterministic component code$_{\mathrm{CONF}}$ $\gamma$ is given as a quintuple $(c_1^\gamma,c_2^\gamma,f_1^\gamma,f_2^\gamma,\varphi^\gamma)$. Note that $f_\nu^\gamma$ can be considered as a function

$$\tilde{f}_\nu^\gamma : [K_\nu]\times([J_1]\times[J_2])\times T_\nu\to\mathscr{X}^n\quad(\mathscr{Y}^n),$$

because $(k_1, \tau_1)$ and $c_2(k_1, \tau_1, k_2, \tau_2)$ together uniquely determine $c_1(k_1, \tau_1, k_2, \tau_2)$. An analogous statement is true for $(k_2, \tau_2)$ and $c_1(k_1, \tau_1, k_2, \tau_2)$. Further, $\varphi$ can be considered as a function

$$\tilde{\varphi} : \mathscr{T}^n \times \mathscr{S} \to ([J_1] \times [J_2]) \times [K_1] \times [K_2],$$

because due to the decoder's perfect CSIR, every choice of $(k_1, k_2)$ together with $(\tau_1, \tau_2)$ also determines a unique $(j_1, j_2)$ via $(c_1, c_2)$. And the correct $(\tau_1, \tau_2)$ is known at the decoder due to its perfect CSIR. Thus, writing $J := J_1 J_2$, the triple $(\tilde{f}_1^\gamma, \tilde{f}_2^\gamma, \tilde{\varphi}^\gamma)$ is a deterministic $(n, T_1, T_2, \mathscr{S})$-code$_{\mathrm{CM}}$ with codelength triple $(J, K_1, K_2)$.

Now we define the random variables necessary for the application of Lemma 3.16. Fix a $(\tau_1, \tau_2) \in T_1 \times T_2$ and an $s \in \mathscr{S}_{\tau_1\tau_2}$. We identify $(\tau_1, \tau_2)$ with $(\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2)$ in the usual way. Let $(M_1, M_2)$ be uniformly distributed on $[K_1] \times [K_2]$ and independent of $G$. Further we define

1) $M_0^{\tilde{\tau}} := c^G(M_1, \tau_1, M_2, \tau_2)$ with $c^\gamma = (c_1^\gamma, c_2^\gamma)$,

2) $X^{\tilde{\tau}\tilde{\tau}_1} := \tilde{f}_1^G(M_1, M_0^{\tilde{\tau}}, \tau_1)$ and $Y^{\tilde{\tau}\tilde{\tau}_2} := \tilde{f}_2(M_2, M_0^{\tilde{\tau}}, \tau_2)$,

3) $T^s$ such that

$$\mathbb{P}[T^s = \mathbf{t} | X^{\tilde{\tau}\tilde{\tau}_1} = \mathbf{x}, Y^{\tilde{\tau}\tilde{\tau}_2} = \mathbf{y}, M_0^{\tilde{\tau}} = j, M_1 = k_1, M_2 = k_2, G = \gamma] = W_s^{\otimes n}(\mathbf{t}|\mathbf{x}, \mathbf{y}),$$

for every choice of $\mathbf{t}, \mathbf{x}, \mathbf{y}, j, k_1, k_2, \gamma$,

4) $(\hat{M}_0^s, \hat{M}_1^s, \hat{M}_2^s) := \tilde{\varphi}(T^s, s)$.

Note that by Corollary 4.10, $M_1$ is independent of $M_2$ conditional on $M_0^{\tilde{\tau}}$ for every $\tilde{\tau} \in \tilde{T}$. Further $\mathbb{P}[(\hat{M}_0^s, \hat{M}_1^s, \hat{M}_2^s) \neq (M_0^{\tilde{\tau}}, M_1, M_2)] = \mathbb{P}[(M_1, M_2) \neq (\hat{M}_1^s, \hat{M}_2^s)]$.

**Lemma 4.11.** *We have*

$$I(M_1 \wedge M_0^{\tilde{\tau}} | M_2) \leq nC_1, \quad I(M_2 \wedge M_0^{\tilde{\tau}} | M_1) \leq nC_2, \quad I(M_1 M_2 \wedge M_0^{\tilde{\tau}}) \leq n(C_1 + C_2).$$

*Proof.* This is clear for $I(M_1 M_2 \wedge M_0^{\tilde{\tau}})$. For $I(M_1 \wedge M_0^{\tilde{\tau}} | M_2)$ we prove by induction over $i = 1, \ldots, I$ that

$$I\big(M_1 \wedge (c_{1,i}^*(M_1, \tau_1, M_2, \tau_2), c_{2,i}^*(M_1, \tau_1, M_2, \tau_2)) | M_2\big) \leq \log J_{1,1} \cdots J_{1,i}. \tag{4.14}$$

(4.14) is immediate for $i = 1$. Assume we have already established (4.14) for $i' = 1, \ldots, i-1$. Then, using the recursive definition of the $c_{\nu,i}^*$,

$$\begin{aligned}
&I\big(M_1 \wedge (c_{1,i}^*(M_1, \tau_1, M_2, \tau_2), c_{2,i}^*(M_1, \tau_1, M_2, \tau_2)) | M_2\big) \\
&\leq H\big(c_{1,i}^*(M_1, \tau_1, M_2, \tau_2), c_{2,i}^*(M_1, \tau_1, M_2, \tau_2) | M_2\big) \\
&= H\big(c_{1,i}(M_1, \tau_1, c_{2,i-1}^*(M_1, \tau_1, M_2, \tau_2)), c_{2,i}(M_2, \tau_2, c_{1,i-1}^*(M_1, \tau_1, M_2, \tau_2)) | M_2\big) \\
&\leq H\big(c_{1,i}(M_1, \tau_1, c_{2,i-1}^*(M_1, \tau_1, M_2, \tau_2)) | c_{2,i}(M_2, \tau_2, c_{1,i-1}^*(M_1, \tau_1, M_2, \tau_2)), M_2\big) \\
&\quad + H\big(c_{2,i}(M_2, \tau_2, c_{1,i-1}^*(M_1, \tau_1, M_2, \tau_2)) | M_2\big) \\
&\leq \log J_{1,i} + H\big(c_{1,i-1}^*(M_1, \tau_1, M_2, \tau_2) | M_2\big) \\
&\leq \log J_{1,1} \cdots J_{1,i},
\end{aligned}$$

where we used the induction hypothesis in the last inequality. Thus we have established (4.14) and hence the lemma. □

We set $\bar{e}_{\mathrm{CONF}}(G, \mathscr{W}, T_1, T_2, R) =: \lambda$ and set $\Delta := 1 + \lambda \log K_1 K_2$. Using Lemma 3.16 and 4.11, we conclude that there is a $p \in \Pi_2(\tilde{T}, \tilde{T}_1, \tilde{T}_2)$ such that

$$\begin{aligned}
\frac{1}{n} \log K_1 &\leq \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(X_{\tilde{\tau}\tilde{\tau}_1} \wedge T_s | Y_{\tilde{\tau}\tilde{\tau}_2} U_{\tilde{\tau}}) + C_1 + \frac{\Delta}{n}, \\
\frac{1}{n} \log K_2 &\leq \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(Y_{\tilde{\tau}\tilde{\tau}_2} \wedge T_s | X_{\tilde{\tau}\tilde{\tau}_1} U_{\tilde{\tau}}) + C_2 + \frac{\Delta}{n}, \\
\frac{1}{n} \log K_1 K_2 &\leq \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(X_{\tilde{\tau}\tilde{\tau}_1} Y_{\tilde{\tau}\tilde{\tau}_2} \wedge T_s | U_{\tilde{\tau}}) + C_1 + C_2 + \frac{\Delta}{n}, \\
\frac{1}{n} \log K_1 K_2 &\leq \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(X_{\tilde{\tau}\tilde{\tau}_1} Y_{\tilde{\tau}\tilde{\tau}_2} \wedge T_s) + \frac{\Delta}{n}.
\end{aligned}$$

On the other hand it follows from (4.13) that at least one of the following inequalities has to be true for some $\varepsilon' = \varepsilon'(\varepsilon)$:

$$\begin{aligned}
\frac{1}{n} \log K_1 &\geq \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(X_{\tilde{\tau}\tilde{\tau}_1} \wedge T_s | Y_{\tilde{\tau}\tilde{\tau}_2} U_{\tilde{\tau}}) + C_1 + \varepsilon', \\
\frac{1}{n} \log K_2 &\geq \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(Y_{\tilde{\tau}\tilde{\tau}_2} \wedge T_s | X_{\tilde{\tau}\tilde{\tau}_1} U_{\tilde{\tau}}) + C_2 + \varepsilon', \\
\frac{1}{n} \log K_1 K_2 &\geq \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(X_{\tilde{\tau}\tilde{\tau}_1} Y_{\tilde{\tau}\tilde{\tau}_2} \wedge T_s | U_{\tilde{\tau}}) + C_1 + C_2 + \varepsilon', \\
\frac{1}{n} \log K_1 K_2 &\geq \min_{\tilde{\tau}, \tilde{\tau}_1, \tilde{\tau}_2} \inf_{s \in \mathscr{S}_{\tilde{\tau}\tilde{\tau}_1\tilde{\tau}_2}} I(X_{\tilde{\tau}\tilde{\tau}_1} Y_{\tilde{\tau}\tilde{\tau}_2} \wedge T_s) + \varepsilon'.
\end{aligned}$$

Proceeding as in the converse proof for the compound MAC with common message completes the proof of the weak converse for the compound MAC with conferencing encoders.

# 5. The Arbitrarily Varying MAC with Conferencing Encoders

## 5.1. Introduction

Arbitrarily Varying MACs (AV-MACs) model a very high degree of channel state uncertainty: the states may vary arbitrarily over time. The task is to use coding to enable reliable communication for every possible state sequence. The random coding capacity region of the AV-MAC without encoder cooperation was determined by Jahn in [32]. Jahn also showed that the deterministic coding capacity equals the random coding capacity if the former's interior is nonempty. A simple condition for this to hold was determined in [9]. There are also conditions for the deterministic region to equal $\{(0,0)\}$, regardless of the random coding region. However, all this does not yet give the complete picture, the full characterization of the deterministic AV-MAC coding region without encoder cooperation is still open.

We will use the "robustification" and "elimination of correlation" techniques developed by Ahlswede in [4, 5], and partly already used in [32] in a multi-user setting, in order to characterize both the deterministic and random coding capacity regions of any AV-MAC with conferencing encoders, i.e. of any AV-MAC where encoding is done using a Willems conference as in [63, 64]. Thus none of the techniques we apply for the AV-MAC is completely new, but in contrast to the non-conferencing situation, they allow for the complete solution of the problems considered here. The rather general "robustification" technique establishes the random coding capacity region of the AV-MAC with conferencing encoders by referring to the coding theorem for a related compound MAC. Both single- and multi-user arbitrarily varying channels are special in that random coding as commonly used in information theory does not yield the same results as deterministic coding. This shows that common randomness shared at the senders and the receiver is an important additional resource. As for single-sender AVCs, we find a dichotomy for the AV-MAC with conferencing encoders: either reliable communication at any non-zero rate pair is impossible with the application of deterministic codes, or the deterministic capacity region coincides with the random coding capacity region. In the latter case, we derandomize using the non-standard "elimination of correlation" [4]. It is a two-step protocol which achieves the random coding capacity region if this is possible.

The combination of the elimination technique with conferencing proves to be very fruitful. The main difference between the AV-MAC with and without conferencing lies in the different symmetrizability criteria that come into play. Symmetrizability can be interpreted in terms of an adversary knowing the channel input symbols and randomizing over the channel states. There are three kinds of symmetrizability for multiple-access

channels. The capacity region of the AV-MAC without conferencing equals $\{(0,0)\}$ if all three symmetrizability conditions are satisfied. In contrast, the elimination of correlation technique works if the AV-MAC with Willems conferencing encoders does not satisfy the conditions for the first of the three kinds of symmetrizabilities. The two others do not matter. By conferencing, the structure of the AV-MAC gets closer to that of a single-sender arbitrarily varying channel where only one symmetrizability condition exists [21]. This works even if the number of messages that may be exchanged during conferencing only grows subexponentially in code blocklength. The adversary interpretation of symmetrizability highlights the importance of the AV-MAC for the theory of information-theoretic secrecy: if a channel is symmetrizable, an adversary can completely prevent communication.

## 5.2. The Problem Setting

Like compound MACs, an AV-MAC with input alphabets $\mathscr{X}$ and $\mathscr{Y}$ and output alphabet $\mathscr{T}$ is also determined by a set of stochastic matrices $\mathscr{W}$ as in the previous chapters. The difference is that we here assume for simplicity that $\mathscr{S}$ is finite. In contrast to the compound MAC, though, the channel state varies arbitrarily from channel use to channel use.

**Definition 5.1.** The *Arbitrarily Varying MAC (AV-MAC)* $\mathrm{AV}(\mathscr{W})$ is the MAC

$$W^{\otimes n}(\,\cdot\,|\,\cdot\,,\,\cdot\,|\mathbf{s}) : \mathscr{X}^n \times \mathscr{Y}^n \to \mathscr{P}(\mathscr{T}^n), \quad \mathbf{s} \in \mathscr{S}^n,\ n = 1, 2, \ldots,$$

where

$$W^{\otimes n}(\mathbf{t}|\mathbf{x}, \mathbf{y}|\mathbf{s}) = \prod_{m=1}^{n} W_{s_m}(t_m|x_m, y_m).$$

We consider here AV-MACs whose senders can do Willems conferencing. We assume that the encoders do not have any CSI, so the conferencing protocols used in this chapter are the same as those for the original discrete memoryless MAC. For the same reason, the deterministic codes we use here are standard conferencing $\mathrm{codes}_{\mathrm{CONF}}$ from definition 2.10. Of course, the error criteria show the different channel characteristics.

**Definition 5.2.** Let $\mathrm{AV}(\mathscr{W})$ be an AV-MAC and let $\gamma$ be a deterministic $(n, C_1, C_2)$-$\mathrm{code}_{\mathrm{CONF}}$ (see Definition 2.10). Its *AV-average error* is given by

$$\bar{e}^{\mathrm{AV}}(\gamma, \mathscr{W}) := \max_{\mathbf{s}\in\mathscr{S}^n} \frac{1}{K_1 K_2} \sum_{k_1,k_2} W^{\otimes n}(D_{k_1k_2}(\gamma)^c|\mathbf{x}_{k_1k_2}(\gamma), \mathbf{y}_{k_1k_2}(\gamma)|\mathbf{s}).$$

Its *AV-maximal error* is given by

$$e^{\mathrm{AV}}(\gamma, \mathscr{W}) := \max_{\mathbf{s}\in\mathscr{S}^n} \max_{k_1,k_2} W^{\otimes n}(D_{k_1k_2}(\gamma)^c|\mathbf{x}_{k_1k_2}(\gamma), \mathbf{y}_{k_1k_2}(\gamma)|\mathbf{s}).$$

In the context of arbitrarily varying channels it is essential to also consider random codes, as these exhibit a different behavior from that of deterministic codes.

**Definition 5.3.** A random variable on $\Gamma_{\mathrm{CONF}}(n, K_1, K_2, C_1, C_2)$ is called a *random* $(n, C_1, C_2)$-*code*$_{\mathrm{CONF}}$. Its *blocklength* and *codelength* are defined analogous to the deterministic case.

**Definition 5.4.** Let $\mathrm{AV}(\mathscr{W})$ be an AV-MAC and let $G$ be a random $(n, C_1, C_2)$-code$_{\mathrm{CONF}}$. Its *AV-average error* is given by

$$\bar{e}^{\mathrm{AV,r}}(G, \mathscr{W}) := \max_{\mathbf{s}\in\mathscr{S}^n} \frac{1}{K_1 K_2} \sum_{k_1,k_2} \sum_{\gamma} W^{\otimes n}\big(D_{k_1k_2}(\gamma)^c | \mathbf{x}_{k_1k_2}(\gamma), \mathbf{y}_{k_1k_2}(\gamma) | \mathbf{s}\big) P_G(\gamma).$$

Its *AV-maximal error* is given by

$$e^{\mathrm{AV,r}}(G, \mathscr{W}) := \max_{\mathbf{s}\in\mathscr{S}^n} \max_{k_1,k_2} \sum_{\gamma} W^{\otimes n}\big(D_{k_1k_2}(\gamma)^c | \mathbf{x}_{k_1k_2}(\gamma), \mathbf{y}_{k_1k_2}(\gamma) | \mathbf{s}\big) P_G(\gamma).$$

The error criteria imply that uniformly for every state sequence, transmission using the given deterministic or random code should be reliable. The possible state sequences are not weighted by any probability measure. One can interpret this in a communication setting with an adversary who knows which words $\mathbf{x}, \mathbf{y}$ are input into the channel by the senders and then can choose any state sequence $\mathbf{s} \in \mathscr{S}^n$ in order to obstruct the transmission of $\mathbf{x}$ and $\mathbf{y}$. The goal of the encoders then is to enable reliable communication no matter what sequence $\mathbf{s}$ the bad guy might use.

We need to modify the concept of achievability here because the capacity regions may depend on the subexponential growth of the number of messages the encoders can exchange during the Willems conference. Thus we consider here general *conferencing capacity sequences* $(C_1(n), C_2(n))_{n=1}^{\infty}$ .

**Definition 5.5.** 1) A pair $(R_1, R_2)$ of nonnegative real numbers is called a *deterministically CONF-achievable rate pair for* $\mathrm{AV}(\mathscr{W})$ *with conferencing capacity sequence* $(C_1^{\infty}, C_2^{\infty}) := (C_1(n), C_2(n))_{n=1}^{\infty}$ *under the average (maximal) error criterion* if for every $\lambda \in (0,1)$ and $\varepsilon > 0$ and for $n \geq n_0(\lambda, \varepsilon)$ there exists a deterministic $(n, C_1(n), C_2(n))$-code$_{\mathrm{CONF}}$ $\gamma$ with $\bar{e}^{\mathrm{AV}}(\gamma, \mathscr{W}) \leq \lambda$ ($e^{\mathrm{AV}}(\gamma, \mathscr{W}) \leq \lambda$) and

$$\frac{1}{n} \log K_\nu \geq R_\nu - \varepsilon \quad (\nu = 1, 2).$$

The set of deterministically CONF-achievable rates under the average (maximal) error criterion is called the *deterministic CONF-capacity region of* $\mathrm{AV}(\mathscr{W})$ *with conferencing capacity sequence* $(C_1^{\infty}, C_2^{\infty})$ *under the average (maximal) error criterion* and denoted by $\overline{\mathscr{C}}_{\mathrm{CM}}^{\mathrm{AV}}(\mathscr{W}, C_1^{\infty}, C_2^{\infty})$ ($\mathscr{C}_{\mathrm{CONF}}^{\mathrm{AV}}(\mathscr{W}, C_1^{\infty}, C_2^{\infty})$).

2) A pair $(R_1, R_2)$ of nonnegative real numbers is called a *randomly CONF-achievable rate pair for* $\mathrm{AV}(\mathscr{W})$ *with conferencing capacity sequence* $(C_1^{\infty}, C_2^{\infty})$ :=

$(C_1(n), C_2(n))_{n=1}^{\infty}$ *under the average (maximal) error criterion* if for every $\lambda \in (0,1)$ and $\varepsilon > 0$ and for $n \geq n_0(\lambda, \varepsilon)$ there exists a random $(n, C_1(n), C_2(n))$-$\mathrm{code}_{\mathrm{CONF}}$ $G$ with $\bar{e}^{\mathrm{AV,r}}(G, \mathscr{W}) \leq \lambda$ ($e^{\mathrm{AV,r}}(G, \mathscr{W}) \leq \lambda$) and

$$\frac{1}{n} \log K_\nu \geq R_\nu - \varepsilon \quad (\nu = 1, 2).$$

The set of randomly CONF-achievable rates under the average (maximal) error criterion is called the *random CONF-capacity region of* $\mathrm{AV}(\mathscr{W})$ *with conferencing capacity sequence* $(C_1^\infty, C_2^\infty)$ *under the average (maximal) error criterion* and denoted by $\overline{\mathscr{C}}_{\mathrm{CM}}^{\mathrm{AV,r}}(\mathscr{W}, C_1^\infty, C_2^\infty)$ ($\mathscr{C}_{\mathrm{CONF}}^{\mathrm{AV,r}}(\mathscr{W}, C_1^\infty, C_2^\infty)$).

## 5.3. Main Results

To characterize the capacity regions we need to consider the convex hull $\overline{\mathscr{W}}$ of $\mathscr{W}$. It is parametrized by the set of probability distributions $\mathscr{P}(\mathscr{S})$ on $\mathscr{S}$, so one can regard $\mathscr{P}(\mathscr{S})$ as its state space. The stochastic matrix from $\overline{\mathscr{W}}$ assigned to the state $q \in \mathscr{P}(\mathscr{S})$ is the matrix with inputs from $\mathscr{X} \times \mathscr{Y}$ and outputs from $\mathscr{T}$ having the form

$$W_q(t|x,y) := \sum_{s \in \mathscr{S}} W_s(t|x,y) q(s), \quad (x,y,t) \in \mathscr{X} \times \mathscr{Y} \times \mathscr{T}.$$

We have $\mathscr{W} \subset \overline{\mathscr{W}}$ by identifying $s \in \mathscr{S}$ with the Dirac measure $\delta_s \in \mathscr{P}(\mathscr{S})$, so that $W_s = W_{\delta_s}$.

In the random coding theorem, only the exponential rates of the number of messages exchangeable between the encoders during conferencing matter, so we use the traditional conferencing capacities again. To state the theorem, we use the notation $\mathscr{C}_2(\overline{\mathscr{W}}, C_1, C_2)$ as there is no CSI.

**Theorem 5.6.** *For* $\mathrm{AV}(\mathscr{W})$ *with conferencing capacities* $C_1, C_2 \geq 0$ *we have*

$$\overline{\mathscr{C}}_{\mathrm{CONF}}^{\mathrm{AV,r}}(\mathscr{W}, C_1, C_2) = \mathscr{C}_2(\overline{\mathscr{W}}, C_1, C_2).$$

$\overline{\mathscr{C}}_{\mathrm{CONF}}^{\mathrm{AV,r}}(\mathscr{W}, C_1, C_2)$ *can be achieved using one-shot Willems conferencing. There exists a weak converse.*

*Remark* 5.1. The proof of Theorem 5.6 shows that $\overline{\mathscr{C}}_{\mathrm{CONF}}^{\mathrm{AV,r}}(\mathscr{W}, C_1, C_2)$ can be achieved using random $(n, C_1, C_2)$-codes whose deterministic component $\mathrm{codes}_{\mathrm{CONF}}$ share the same one-shot Willems conferencing protocol and whose average error tends to zero exponentially in blocklength (cf. Remark 4.2).

Next we use the general conferencing capacity sequences $C_1^\infty, C_2^\infty$ to get a detailed picture of deterministic coding. The structure of $\overline{\mathscr{C}}_{\mathrm{CONF}}^{\mathrm{AV}}(\mathscr{W}, C_1^\infty, C_2^\infty)$ is complicated compared to the capacity regions encountered so far. It depends on an additional property $\mathscr{W}$ might or might not have.

**Definition 5.7** ([30]). 1) $\mathscr{W}$ is called $(\mathscr{X},\mathscr{Y})$-*symmetrizable* if there is a stochastic matrix $\sigma : \mathscr{X} \times \mathscr{Y} \to \mathscr{P}(\mathscr{S})$ such that for every $t \in \mathscr{T}$ and $x, x' \in \mathscr{X}$ and $y, y' \in \mathscr{Y}$,

$$\sum_s W_s(t|x,y)\sigma(s|x',y') = \sum_s W_s(t|x',y')\sigma(s|x,y).$$

2) $\mathscr{W}$ is called $\mathscr{X}$-*symmetrizable* if there is a stochastic matrix $\sigma_1 : \mathscr{X} \to \mathscr{S}$ such that for every $t \in \mathscr{T}$ and $x, x' \in \mathscr{X}$ and $y \in \mathscr{Y}$,

$$\sum_s W_s(t|x,y)\sigma_1(s|x') = \sum_s W_s(t|x',y)\sigma_1(s|x).$$

3) $\mathscr{W}$ is called $\mathscr{Y}$-*symmetrizable* if there is a stochastic matrix $\sigma_2 : \mathscr{Y} \to \mathscr{S}$ such that for every $t \in \mathscr{T}$ and $x \in \mathscr{X}$ and $y, y' \in \mathscr{Y}$,

$$\sum_s W_s(t|x,y)\sigma_2(s|y') = \sum_s W_s(t|x,y')\sigma_2(s|y).$$

**Theorem 5.8.** *1) If $\mathscr{W}$ is symmetrizable, then $\overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\mathscr{W}, C_1^\infty, C_2^\infty) = \{(0,0)\}$ for every pair of conferencing capacity sequences $C_1^\infty, C_2^\infty$. There exists an "almost strong" converse: every deterministic code$_{\mathrm{CONF}}$ that encodes at least two messages incurs an average error at least $1/4$.*

*2) If $\mathscr{W}$ is not symmetrizable and there is an $\eta > 0$ such that*

$$\lim_{n\to\infty} n \max\bigl\{C_1(n), C_2(n)\bigr\} - (1+\eta)\log n = \infty, \tag{5.1}$$

*then*

$$\overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\mathscr{W}, C_1^\infty, C_2^\infty) = \mathscr{C}_2(\overline{\mathscr{W}}, \liminf_{n\to\infty} C_1(n), \liminf_{n\to\infty} C_2(n)).$$

*There exists a weak converse. The Willems conferencing protocols can again be assumed to be one-shot.*

*Remark* 5.2. Using Landau symbols, one can write (5.1) as

$$\max\bigl\{C_1(n), C_2(n)\bigr\} - \frac{(1+\eta)\log n}{n} \in \omega\left(\frac{1}{n}\right).$$

This condition is satisfied if there is an $\eta' > \eta$ with $\max\{C_1(n), C_2(n)\} \geq ((1+\eta')\log n)/n$ for sufficiently large $n$. In particular, $\liminf C_1(n) = \liminf C_2(n) = 0$ is possible.

*Remark* 5.3. If $\mathscr{W}$ is not $(\mathscr{X},\mathscr{Y})$-symmetrizable and $\max\{\liminf C_1(n), \liminf C_2(n)\} > 0$, then $\overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\overline{\mathscr{W}}, \liminf C_1(n), \liminf C_2(n))$ is at least one-dimensional. In order to show this it clearly suffices to check that

$$\max_{p\in\Pi(\overline{\mathscr{W}})} \min_{q\in\mathscr{P}(\mathscr{S})} I(T_q \wedge XY) > 0 \tag{5.2}$$

if $\mathscr{W}$ is not $(\mathscr{X},\mathscr{Y})$-symmetrizable, where we set

$$\Pi(\overline{\mathscr{W}}) := \Pi_1(\overline{\mathscr{W}},\{\mathscr{S}\},\{\mathscr{S}\}) = \Pi_2(\overline{\mathscr{W}},\{\mathscr{S}\},\{\mathscr{S}\},\{\mathscr{S}\}).$$

If (5.2) were violated, then by [20, Lemma 1.3.2] there would be a $q \in \mathscr{P}(\mathscr{S})$ such that

$$W_q(z|x,y) = W_q(z|x',y') \qquad \text{for all } x,x' \in \mathscr{X}, y,y' \in \mathscr{Y}, z \in \mathscr{Z}.$$

Thus $\mathscr{W}$ would be $(\mathscr{X},\mathscr{Y})$-symmetrizable using the stochastic matrix

$$\sigma(s|x,y) = q(s), \quad (x,y,s) \in \mathscr{X}\times\mathscr{Y}\times\mathscr{S}.$$

But this would contradict our assumption, so (5.2) must hold.

*Remark* 5.4. One can regard symmetrizability as the single-letterization of the adversary interpretation of the AV-MAC given above. In this interpretation, a complete input word pair has to be known to the adversary who can then choose the state sequence. In the definition of $(\mathscr{X},\mathscr{Y})$-symmetrizability, the stochastic matrix $\sigma : \mathscr{X} \to \mathscr{S}$ means that given a *letter* $x \in \mathscr{X}$, the adversary chooses a *random* state $s \in \mathscr{S}$. If $\mathscr{W}$ is $(\mathscr{X},\mathscr{Y})$-symmetrizable, the adversary can thus produce a useless single-state MAC $\tilde{W} : (\mathscr{X}\times\mathscr{Y})^2 \to \mathscr{Z}$ defined by

$$\tilde{W}(z|x,y,x',y') = \sum_{s\in\mathscr{S}} W(z|x,y|s)\sigma(s|x',y').$$

DMAC($\tilde{W}$) is useless because it is symmetric in $(x,y)$ and $(x',y')$. Thus for word pairs $(\mathbf{x},\mathbf{y})$ and $(\mathbf{x}',\mathbf{y}')$, the receiver cannot decide which of the pairs was input into the channel by the senders and which was induced by the adversary's random state choice.

*Remark* 5.5. The above adversary interpretation of symmetrizability makes AV-MACs relevant for information-theoretic secrecy. Clearly, we do not say anything about the decodability of communication taking place in an AV-MAC for non-legitimate listeners. However, reliable communication can be completely prevented in the case the AV-MAC is symmetrizable. A discussion of the single-sender arbitrarily varying wiretap channel can be found in [11].

Theorem 6.11 does not carry over to the case $C_1(n) = C_2(n) = 0$ for all $n$, which is the traditional AV-MAC with non-cooperative coding. To our knowledge, the full characterization of the deterministic capacity region $\overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\mathscr{W},0,0)$ of AV($\mathscr{W}$) without cooperation is still an open problem. We summarize here what has been found out in [9], [29], [30], and [32]. For notation, observe that

$$\begin{aligned}\max_{p\in\Pi(\overline{\mathscr{W}})}\inf_{q\in\mathscr{P}(\mathscr{S})} I(T_q\wedge X|YU) &= \max_{p\in\Pi(\overline{\mathscr{W}})}\inf_{q\in\mathscr{P}(\mathscr{S})} I(T_q\wedge X|Y)\\ &= \max_{y\in\mathscr{Y}}\max_{r\in\mathscr{P}(\mathscr{X})}\inf_{q\in\mathscr{P}(\mathscr{S})} I(T_q\wedge X|Y=y),\end{aligned}$$

where in the last term, the random vector $(X,T_q)$ has the distribution $r(x)W_q(z|x,y)$.

**Theorem 5.9.** *1) If $\mathscr{W}$ is neither $(\mathscr{X},\mathscr{Y})$- nor $\mathscr{X}$- nor $\mathscr{Y}$-symmetrizable, then $\overline{\mathscr{C}}_{\mathrm{CONF}}^{\mathrm{AV}}(\mathscr{W},0,0)=\mathscr{C}_2(\overline{\mathscr{W}},0,0)$ and $\mathscr{C}_2(\overline{\mathscr{W}},0,0)$ has nonempty interior.*

*2) If $\mathscr{W}$ is neither $(\mathscr{X},\mathscr{Y})$- nor $\mathscr{X}$-symmetrizable, but $\mathscr{Y}$-symmetrizable, then*

$$\overline{\mathscr{C}}_{\mathrm{CONF}}^{\mathrm{AV}}(\mathscr{W},0,0)\subset[0,\max_{y\in\mathscr{Y}}\max_{r\in\mathscr{P}(\mathscr{X})}\inf_{q\in\mathscr{P}(\mathscr{S})}I(T_q;X|Y=y)]\times\{\mathscr{S}\}.$$

*3) If $\mathscr{W}$ is neither $(\mathscr{X},\mathscr{Y})$- nor $\mathscr{Y}$-symmetrizable, but $\mathscr{X}$-symmetrizable, then*

$$\overline{\mathscr{C}}_{\mathrm{CONF}}^{\mathrm{AV}}(\mathscr{W},0,0)\subset\{\mathscr{S}\}\times[0,\max_{x\in\mathscr{X}}\max_{r\in\mathscr{P}(\mathscr{Y})}\inf_{q\in\mathscr{P}(\mathscr{S})}I(T_q;Y|X=x)].$$

*4) If $\mathscr{W}$ is $(\mathscr{X},\mathscr{Y})$-symmetrizable, then $\overline{\mathscr{C}}_{\mathrm{CONF}}^{\mathrm{AV}}(\mathscr{W},0,0)=\{(0,0)\}$.*

*In particular if $\mathscr{W}$ is both $\mathscr{X}$- and $\mathscr{Y}$-symmetrizable, then $\overline{\mathscr{C}}_{\mathrm{CONF}}^{\mathrm{AV}}(\mathscr{W},0,0)=\{(0,0)\}$.*

*Remark* 5.6. 1) from Theorem 5.9 is due to [9] and [32]. The other points are due to [29, 30]. The precise characterization of $\overline{\mathscr{C}}_{\mathrm{CONF}}^{\mathrm{AV}}(\mathscr{W},0,0)$ in points 2) and 3) is still open.

*Remark* 5.7. The relation between the three kinds of symmetrizability from Definition 5.7 is treated in Section 5.6. There we provide the example of an AV-MAC which is both $\mathscr{X}$- and $\mathscr{Y}$-symmetrizable but not $(\mathscr{X},\mathscr{Y})$-symmetrizable.

## 5.4. The Direct Parts

We derive the direct part of Theorem 5.6 from Theorem 4.6 in Subsection 5.4.1. Then, if $\mathscr{W}$ is not $(\mathscr{X},\mathscr{Y})$-symmetrizable, we derandomize in Subsections 5.4.2-5.4.4 to obtain the direct part of Theorem 5.8.

### 5.4.1. From Compound to Arbitrarily Varying

Let $\mathscr{W}$ be an AV-MAC and $C_1,C_2\geq$ conferencing capacities. Here we prove

$$\mathscr{C}_2(\overline{\mathscr{W}},C_1,C_2)\subset\overline{\mathscr{C}}_{\mathrm{CONF}}^{\mathrm{AV,r}}(\mathscr{W},C_1,C_2). \tag{5.3}$$

We use Ahlswede's "robustification technique", in particular the "robustification lemma". Let $S_n$ be the symmetric group (the group of permutations) on the set $[1,n]$. $S_n$ operates on $\mathscr{S}^n$ by $\pi(\mathbf{s}):=(s_{\pi(1)},\ldots,s_{\pi(n)})$ for any $\pi\in S_n$ and $\mathbf{s}=(s_1,\ldots,s_n)\in\mathscr{S}^n$.

**Lemma 5.10** ([7], Lemma RT)**.** *If $h:\mathscr{S}^n\to[0,1]$ satisfies for a $\lambda\in(0,1)$ and for all $q\in\mathscr{P}(\mathscr{S})$ the inequality*

$$\sum_{\mathbf{s}\in\mathscr{S}^n}h(\mathbf{s})q^{\otimes n}(\mathbf{s})\geq 1-\lambda, \tag{5.4}$$

*then it also satisfies the inequality*

$$\frac{1}{n!}\sum_{\pi\in S_n}h(\pi(\mathbf{s}))\geq 1-(n+1)^{|\mathscr{S}|}\lambda\qquad\text{for all }\mathbf{s}\in\mathscr{S}^n.$$

Now let $(R_1, R_2) \in \mathscr{C}_2(\overline{\mathscr{W}}, C_1, C_2)$. Theorem 4.6 and Remark 4.2 state that for any $\varepsilon > 0$ there is a $\zeta > 0$ such that for sufficiently large $n$ there is a deterministic $(n, C_1, C_2)$-code$_{\mathrm{CONF}}$ $\gamma$ with $\bar{e}^{\mathrm{Cp}}(\gamma, \mathscr{W}) \leq 2^{-n\zeta}$ (omitting the notation for the nonexistent CSI) and a codelength pair $(K_1, K_2)$ that satisfies

$$\frac{1}{n} \log K_\nu \geq R_\nu - \varepsilon \quad (\nu = 1, 2).$$

This means for every $q \in \mathscr{P}(\mathscr{S})$ that

$$\frac{1}{K_1 K_2} \sum_{k_1, k_2} W_q^{\otimes n}\big(D_{k_1 k_2}(\gamma) | \mathbf{x}_{k_1 k_2}(\gamma), \mathbf{y}_{k_1 k_2}(\gamma)\big) \geq 1 - 2^{-n\zeta}. \tag{5.5}$$

We would like to apply Lemma 5.10 with $\lambda = 2^{-n\zeta}$ to the function $h : \mathscr{S}^n \to [0, 1]$ defined by

$$h(\mathbf{s}) := \frac{1}{K_1 K_2} \sum_{k_1, k_2} W^{\otimes n}(D_{k_1 k_2}(\gamma) | \mathbf{x}_{k_1 k_2}(\gamma), \mathbf{y}_{k_1 k_2}(\gamma) | \mathbf{s}).$$

Thus we need to show that $h$ satisfies (5.4). Let $q \in \mathscr{P}(\mathscr{S})$. By (5.5), one obtains

$$\begin{aligned} \sum_{\mathbf{s} \in \mathscr{S}^n} h(\mathbf{s}) q^{\otimes n}(\mathbf{s}) &= \frac{1}{K_1 K_2} \sum_{k_1, k_2} \sum_{\mathbf{t} \in D_{k_1 k_2}(\gamma)} \sum_{\mathbf{s} \in \mathscr{S}^n} W^{\otimes n}(\mathbf{t} | \mathbf{x}_{k_1 k_2}(\gamma), \mathbf{y}_{k_1 k_2}(\gamma) | \mathbf{s}) q^{\otimes n}(\mathbf{s}) \\ &= \frac{1}{K_1 K_2} \sum_{k_1, k_2} \sum_{\mathbf{t} \in D_{k_1 k_2}(\gamma)} W_q^{\otimes n}(\mathbf{t} | \mathbf{x}_{k_1 k_2}(\gamma), \mathbf{y}_{k_1 k_2}(\gamma)) \\ &\geq 1 - 2^{-n\zeta}, \end{aligned}$$

and (5.4) is satisfied. Applying Lemma 5.10, one obtains

$$\frac{1}{n!} \sum_{\pi \in S_n} h(\pi(\mathbf{s})) \geq 1 - (n+1)^{|\mathscr{S}|} 2^{-n\zeta} \qquad \text{for all } \mathbf{s} \in \mathscr{S}^n. \tag{5.6}$$

Recall that $\pi^{-1}$ also is an element of $S_n$. Writing $\pi^{-1}(D_{k_1 k_2}(\gamma)) = \{\pi^{-1}(\mathbf{t}) : \mathbf{t} \in D_{k_1 k_2}(\gamma)\}$, the left side of (5.6) equals

$$\begin{aligned} &\frac{1}{n!} \sum_{\pi \in S_n} \left( \frac{1}{K_1 K_2} \sum_{k_1, k_2} W^{\otimes n}(D_{k_1 k_2}(\gamma) | \mathbf{x}_{k_1 k_2}(\gamma), \mathbf{y}_{k_1 k_2}(\gamma) | \pi(\mathbf{s})) \right) \\ &= \frac{1}{n!} \sum_{\pi \in S_n} \left( \frac{1}{K_1 K_2} \sum_{k_1, k_2} W^{\otimes n}\left(\pi^{-1}(D_{k_1 k_2}(\gamma)) | \pi^{-1}(\mathbf{x}_{k_1 k_2}(\gamma)), \pi^{-1}(\mathbf{y}_{k_1 k_2}(\gamma)) | \mathbf{s}\right) \right). \end{aligned} \tag{5.7}$$

Because of the bijectivity of $\pi^{-1}$, the family of sets $\{\pi^{-1}(D_{k_1 k_2}(\gamma)) : (k_1, k_2) \in [K_1] \times [K_2]\}$ is disjoint. Thus one obtains for every $\pi \in S_n$ a deterministic $(n, C_1, C_2)$-code$_{\mathrm{CONF}}$ $(c_1, c_2, f_1^\pi, f_2^\pi, \varphi^\pi)$ through the set

$$\{\big(\pi^{-1}(\mathbf{x}_{k_1 k_2}(\gamma)), \pi^{-1}(\mathbf{y}_{k_1 k_2}(\gamma)), \pi^{-1}(D_{k_1 k_2}(\gamma))\big) : (k_1, k_2) \in [K_1] \times [K_2]\}.$$

Note that all these deterministic codes$_{\mathrm{CONF}}$ share the Willems conference of $\gamma$. A random variable $G$ uniformly distributed on $S_n$ thus induces a random $(n, C_1, C_2)$-code$_{\mathrm{CONF}}$, and (5.7) equals $1-\bar{e}^{\mathrm{AV,r}}(G, \mathscr{W})$. By (5.6) we have $\bar{e}^{\mathrm{AV,r}}(G, \mathscr{W}) \leq (n+1)^{|\mathscr{S}|} 2^{-n\zeta}$, in particular we obtain an exponential decrease of the average error towards zero. This proves (5.3) and thus the direct part of Theorem 5.6.

### 5.4.2. Bounding the amount of correlation

As a first derandomization step to proving the direct part of Theorem 5.8, we have to show the following lemma. Recall that, for a random variable $G$, we have defined $\mathrm{supp}(G)$ as the set of those values $\gamma$ of $G$ with $P_G(\gamma) > 0$.

**Lemma 5.11.** *Let $\eta > 0$. To every random $(n, C_1, C_2)$-code$_{\mathrm{CONF}}$ $G$ with $\bar{e}^{\mathrm{AV,r}}(G, \mathscr{W}) \leq \lambda$ there exists a random $(n, C_1, C_2)$-code$_{\mathrm{CONF}}$ $G'$ with the same codelength pair and*

*1) $\bar{e}^{\mathrm{AV,r}}(G', \mathscr{W}) \leq 3\lambda$,*

*2) $\mathrm{supp}(G') \subset \mathrm{supp}(G)$,*

*3) $|\mathrm{supp}(G')| \leq n^{1+\eta}$.*

For the proof of Lemma 5.11, we need a simple result from [32, Section IV].

**Lemma 5.12.** *Let $N$ i.i.d. random variables $T_1, \ldots, T_N$ with values in $[0, 1]$ and underlying probability measure $\mathbb{P}$ be given. Let $\bar{\lambda} > 0$. Denote by $\mathbb{E}$ the expectation corresponding to $\mathbb{P}$. Then*

$$\mathbb{P}\left[\frac{1}{N}\sum_{m=1}^{N} T_m > \bar{\lambda}\right] \leq \exp\left(-(\bar{\lambda} - e\mathbb{E}[T_1])N\right).$$

*Proof of Lemma 5.11.* Let $G$ be a random $(n, C_1, C_2)$-code$_{\mathrm{CONF}}$ with $\bar{e}^{\mathrm{AV,r}}(G, \mathscr{W}) \leq \lambda$. With

$$\bar{e}_{\mathbf{s}}(G) := \frac{1}{K_1 K_2} \sum_{k_1, k_2} W^{\otimes n}(D_{k_1 k_2}(G)^c | \mathbf{x}_{k_1 k_2}(G), \mathbf{y}_{k_1 k_2}(G) | \mathbf{s}), \tag{5.8}$$

the fact that $\bar{e}^{\mathrm{AV,r}}(G, \mathscr{W}) \leq \lambda$ can be stated as

$$\max_{\mathbf{s} \in \mathscr{S}^n} \mathbb{E}[\bar{e}_{\mathbf{s}}(G)] \leq \lambda.$$

Define $N := \lfloor n^{1+\eta} \rfloor$ and let $G_1, \ldots, G_N$ be independent copies of $G$. The goal is to show

$$\mathbb{P}\left[\frac{1}{N}\sum_{m=1}^{N} \bar{e}_{\mathbf{s}}(G_m) \leq 3\lambda \text{ for all } \mathbf{s} \in \mathscr{S}^n\right] > 0. \tag{5.9}$$

Given (5.9), there is a realization $(\gamma_1, \ldots, \gamma_N)$ of $(G_1, \ldots, G_N)$ such that

$$\frac{1}{N}\sum_{m=1}^{N} \bar{e}_{\mathbf{s}}(\gamma_m) \leq 3\lambda \tag{5.10}$$

for every $\mathbf{s} \in \mathscr{S}^n$. Then one defines a random $(n, C_1, C_2)$-code as a random variable $G'$ uniformly distributed on $\{\gamma_1, \ldots, \gamma_N\}$. The expression (5.10) then is nothing but the statement that $\bar{e}^{\mathrm{AV,r}}(G', \mathscr{W}) \leq 3\lambda$, and we are done.

It remains to prove (5.9). $\mathscr{S}$ is finite by assumption, so $|\mathscr{S}^n|$ grows exponentially with blocklength. Hence it suffices to show that

$$\mathbb{P}\left[ \frac{1}{N} \sum_m \bar{e}_{\mathbf{s}}(G_m) > 3\lambda \right] \tag{5.11}$$

is superexponentially small uniformly in $\mathbf{s} \in \mathscr{S}^n$. Let us fix an $\mathbf{s} \in \mathscr{S}^n$. The $G_m$ are i.i.d. copies of $G$, so by Lemma 5.12, the term (5.11) is smaller than

$$\exp\bigl(-\left(3\lambda - e\,\mathbb{E}[\bar{e}_{\mathbf{s}}(G)]\right) N\bigr). \tag{5.12}$$

By assumption $\mathbb{E}[\bar{e}_{\mathbf{s}}(G)] \leq \lambda$, so the exponent in (5.12) is negative. This gives the desired superexponential bound on (5.11). □

*Remark* 5.8. Note that we cannot require the codes$_{\mathrm{CONF}}$ $G'$ with at most $\lfloor n^{1+\eta} \rfloor$ deterministic values to have an exponentially small probability of error. This is due to the fact that the exponent in (5.12) must not decrease exponentially in order for the proof to work. Thus there is a trade-off between the error probability and the number of deterministic component codes$_{\mathrm{CONF}}$ of the random codes$_{\mathrm{CONF}}$ used to achieve the random capacity region of the AV-MAC with conferencing encoders.

### 5.4.3. A Positive Rate

We use Lemma A.7 from the theory of single-sender Arbitrarily Varying Channels (AVCs) to show that sufficiently many messages can be transmitted through AV($\mathscr{W}$) for Ahlswede's elimination technique to work. Observe that $(\mathscr{X}, \mathscr{Y})$-symmetrizability of $\mathscr{W}$ is equivalent to symmetrizability of $\mathscr{W}$ when considered as a set of single-sender stochastic matrices with input alphabet $\mathscr{X} \times \mathscr{Y}$, see Appendix A.

Assume that $\mathscr{W}$ is not $(\mathscr{X}, \mathscr{Y})$-symmetrizable and let $C_1^\infty, C_2^\infty$ be sequences of conferencing capacities satisfying (5.1) for some $\eta > 0$. Then the single-sender AVC-capacity $\overline{\mathscr{C}}^{\mathrm{AVC}}(\mathscr{W})$ defined in Appendix A is positive by Lemma A.7.

**Lemma 5.13.** *Let $\varepsilon, \lambda \in (0,1)$. Then there is a $n_0(\varepsilon, \lambda)$ such that for all $n \geq n_0(\varepsilon, \lambda)$, defining*

$$m := \left\lfloor \frac{n\varepsilon}{1-\varepsilon} \right\rfloor,$$

*there exists an $(m, C_1(m), C_2(m))$-code$_{\mathrm{CONF}}$ with a codelength pair $(K_1, K_2)$ satisfying $K_1 K_2 = \lfloor n^{1+\eta} \rfloor$ and with the identities on the message sets as conferencing functions.*

*Proof.* Choose $n$ so large that there exists an $m$-code$_{\mathrm{1S}}$ $(f, \varphi)$ (see Appendix 5) with $\bar{e}^{\mathrm{AVC}}(f, \varphi, \mathscr{W}) \leq \lambda/2$ and a code rate $L$ satisfying

$$2m\overline{\mathscr{C}}^{\mathrm{AVC}}(\mathscr{W}) \geq 2 \log L \geq m\overline{\mathscr{C}}^{\mathrm{AVC}}(\mathscr{W}). \tag{5.13}$$

Choose a sub-$m$-code$_{1S}$ $(f', \varphi')$ of $(f, \varphi)$ with codelength $\lfloor n^{1+\eta} \rfloor$. This is possible because

$$\frac{1}{m} \log n^{1+\eta} \leq \frac{1+\eta}{m} \log(m+1) + \frac{1+\eta}{m} \log \frac{1-\varepsilon}{\varepsilon} \leq \frac{\overline{\mathscr{C}}^{\mathrm{AVC}}(\mathscr{W})}{2}$$

for $n$ sufficiently large. Lemma 5.14 below implies $\bar{e}^{\mathrm{AVC}}(f', \varphi', \mathscr{W}) \leq 2\bar{e}^{\mathrm{AVC}}(f, \varphi, \mathscr{W}) \leq \lambda$.

Due to (5.1), we also have

$$\frac{1}{m} \log n^{1+\eta} \leq \frac{1+\eta}{m} \log(m+1) + \frac{1+\eta}{m} \log \frac{1-\varepsilon}{\varepsilon} \leq \max\bigl\{C_1(m), C_2(m)\bigr\}.$$

Without loss of generality assuming $C_1(m) \geq C_2(m)$ and setting $K_1 := \lfloor n^{1+\eta} \rfloor$, $K_2 = 1$, we can thus interpret $(f', \varphi')$ as an $(m, C_1(m), C_2(m))$-code$_{\mathrm{CONF}}$ $\gamma^*$ with codelength pair $K_1, K_2$ using the identities on $[K_1]$ and $[K_2]$ as conferencing functions. In this way, each sender can completely inform the other sender about its message and then each encoder applies the codeword corresponding to the joint message pair from the $m$-code$_{1S}$ $(f', \varphi')$ constructed above. We also have $\bar{e}^{\mathrm{AV}}(\gamma^*, \mathscr{W}) = \bar{e}^{\mathrm{AVC}}(f', \varphi', \mathscr{W}) \leq \lambda$. This completes the proof of the lemma. □

**Lemma 5.14.** *Let $K < L$ be positive integers, $\lambda > 0$ and let $e : [L] \to [0, \infty)$. If*

$$\frac{1}{L} \sum_{l=1}^{L} e(l) \leq \lambda,$$

*then there is an $A \subset [L]$ with $|A| = K$ satisfying*

$$\frac{1}{K} \sum_{a \in A} e(a) \leq \frac{L}{L-K} \lambda.$$

*Proof.* Assume there were no such $A$. Write $L = bK + c$ for nonnegative integers $b, c$ with $c < K$. Then

$$\begin{aligned}
\lambda &\geq \frac{1}{L} \sum_{l=1}^{L} e(l) \\
&= \frac{K}{L} \sum_{i=0}^{b-1} \left( \frac{1}{K} \sum_{l=i+1}^{i+K} e(l) \right) + \frac{1}{L} \sum_{l=bK+1}^{L} e(l) \\
&> \frac{K}{L} \cdot b \cdot \frac{L}{L-K} \lambda \\
&= \frac{L-c}{L} \cdot \frac{L}{L-K} \lambda \\
&\geq \frac{L-K}{L} \cdot \frac{L}{L-K} \lambda \\
&= \lambda.
\end{aligned}$$

This is a contradiction, so there must be a set $A$ as claimed. □

### 5.4.4. From Random to Deterministic

Here we perform the final step of derandomization by showing that if $\mathscr{W}$ is not $(\mathscr{X},\mathscr{Y})$-symmetrizable and $C_1^\infty, C_2^\infty$ satisfy (5.1), then

$$\mathscr{C}_2(\overline{\mathscr{W}}, C_1(\infty), C_2(\infty)) \subset \overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\mathscr{W}, C_1^\infty, C_2^\infty),$$

where we set

$$C_\nu(\infty) := \liminf_{n\to\infty} C_\nu(n) \quad (\nu = 1,2).$$

To do so we follow Ahlswede's "Elimination Technique" [4], whose idea is to use random codes and to replace the randomness needed there by a prefix code with small blocklength which encodes the set of constituent deterministic codes.

Fix $\varepsilon, \lambda \in (0,1)$. For $0 < \delta < \min\{C_\nu(\infty) : \nu = 1,2,\, C_\nu(\infty) > 0\}$, let $(R_1,R_2) \in \mathscr{R}_{\mathrm{CONF}}(p, [C_1(\infty)-\delta]_+, [C_2(\infty)-\delta]_+)$ for some $p \in \Pi(\overline{\mathscr{W}})$. Set $\varepsilon' := \varepsilon/(2\max\{R_1,R_2\})$. Choose $n$ so large that the following conditions hold:

(i) Lemma 5.13 holds true with $\varepsilon$ replaced by $\varepsilon'$ and $\lambda$ by $\lambda/2$, consequently we set $m = \lfloor n\varepsilon'/(1-\varepsilon')\rfloor$,

(ii) there is a random $(n, [C_1(\infty)-\delta]_+, [C_2(\infty)-\delta]_+)$-code$_{\mathrm{CONF}}$ $G$ with $\bar{e}^{\mathrm{AV}}(G,\mathscr{W}) \le \lambda/2$ and

$$\frac{1}{n}\log K_\nu \ge R_\nu - \frac{\varepsilon}{2},$$

$G$ being uniformly distributed on $[1, \lfloor n^{1+\eta}\rfloor]$ and the deterministic component codes$_{\mathrm{CONF}}$ all sharing the same one-shot Willems conference

$$(c_1,c_2) : [K_1]\times[K_2] \to [J_1]\times[J_2].$$

The second assumption can be made because of Theorem 5.6 and Lemma 5.11.

We write $\lfloor n^{1+\eta}\rfloor =: N$. Let the deterministic $(m, C_1(m), C_2(m))$-code$_{\mathrm{CONF}}$ $\gamma^*$ from assumption (i) have codelength pair $(n_1,n_2)$. We have $\gamma^* = (c_1^{\gamma^*}, c_2^{\gamma^*}, f_1^{\gamma^*}, f_2^{\gamma^*}, \varphi^{\gamma^*})$. Recall that $c_\nu^{\gamma^*}$ is the identity on $[n_\nu]$ for $\nu = 1,2$ and that $n_1 n_2 = N$. We may further assume that

$$\mathrm{supp}(G) = \{\gamma_1,\dots,\gamma_N\} \subset \Gamma_{\mathrm{CONF}}(n, K_1, K_2, [C_1(\infty)-\delta]_+, [C_2(\infty)-\delta]_+),$$

recall that the set on the right-hand side was defined in Definition 4.1, we omit mentioning the trivial CSI.

We now construct a deterministic $(m+n, C_1(m+n), C_2(m+n))$-code$_{\mathrm{CONF}}$ $\tilde{\gamma}$ with message sets $[n_1]\times[K_1]$ and $[n_2]\times[K_2]$ and $\bar{e}^{\mathrm{AV}}(\tilde{\gamma},\mathscr{W}) \le \lambda$. It is defined via concatenation. We define the one-shot Willems conferencing functions by

$$\tilde{c}_\nu(\xi_\nu, k_\nu) := (\xi_\nu, c_\nu(k_\nu)) \in [n_\nu]\times[J_\nu] \qquad (\nu = 1,2).$$

We have to check that $(\tilde{c}_1, \tilde{c}_2)$ is an admissible Willems conferencing protocol. For $\nu = 1, 2$, we have to distinguish the cases $C_\nu(\infty) = 0$ and $C_\nu(\infty) > 0$. In the former case,

$$\frac{n_\nu J_\nu}{m+n} \leq \frac{(1+\eta)\log n}{m+n} + \frac{n}{m+n} C_\nu(\infty) \leq \frac{(1+\eta)\log(m+n)}{m+n} \leq C_\nu(m+n),$$

where the last inequality holds for sufficiently large $n$. In the case $C_\nu(\infty) > 0$, choose $n$ so large that

$$C_\nu(m+n) \geq C_\nu(\infty) - \frac{\delta}{2}.$$

Then

$$\begin{aligned}
\frac{n_\nu J_\nu}{m+n} &\leq \frac{(1+\eta)\log n}{m+n} + \frac{n}{m+n}\big(C_\nu(\infty) - \delta\big) \\
&\leq C_\nu(m+n) - \frac{\delta}{2} - \frac{1}{m+n}\big(m\big(C_\nu(\infty) - \delta\big) - (1+\eta)\log n\big) \\
&\leq C_\nu(m+n) - \frac{\delta}{2} - \frac{1}{m+n}\left(\frac{n\varepsilon'}{1-\varepsilon'}\big(C_\nu(\infty) - \delta\big) - (1+\eta)\log n\right) \\
&\leq C_\nu(m+n),
\end{aligned}$$

where the last inequality holds for sufficiently large $n$.

Thus $(\tilde{c}_1, \tilde{c}_2)$ is an $(m+n, C_1(m+n), C_2(m+n))$-conference. The codewords used by the encoders are concatenations of codewords from $\gamma^*$ and the elements of $\mathrm{supp}(G)$. If a message pair $((\xi_1, k_1), (\xi_2, k_2))$ is given, identifying $[N]$ with $[n_1] \times [n_2]$ and writing

$$\mathrm{supp}(G) = \{\gamma_{\xi_1\xi_2} : (\xi_1, \xi_2) \in [n_1] \times [n_2]\},$$

the encoders use the codewords

$$\big(\mathbf{x}_{\xi_1\xi_2}(\gamma^*), \mathbf{x}_{k_1k_2}(\gamma_{\xi_1\xi_2})\big) \in \mathscr{X}^{m+n} \quad \text{and} \quad \big(\mathbf{y}_{\xi_1\xi_2}(\gamma^*), \mathbf{y}_{k_1k_2}(\gamma_{\xi_1\xi_2})\big) \in \mathscr{Y}^{m+n}.$$

Together with the conferencing protocol $(\tilde{c}_1, \tilde{c}_2)$ defined above, this fixes encoding functions $\tilde{f}_1$ and $\tilde{f}_2$. The decoding set of the $\mathrm{code}_{\mathrm{CONF}}$ deciding for the pair $\big((\xi_1, k_1), (\xi_2, k_2)\big)$ is defined to be $D_{\xi_1\xi_2}(\gamma^*) \times D_{k_1k_2}(\gamma_{\xi_1\xi_2}) \subset \mathscr{T}^{m+n}$. Thus $\gamma^*$ is used as a prefix code which distinguishes the deterministic component $\mathrm{codes}_{\mathrm{CONF}}$ of $G$. In this way, derandomization can be seen as a two-step protocol.

The rates achieved with $\tilde{\gamma}$ are for $\nu = 1, 2$

$$\frac{1}{m+n}\log(n_\nu K_\nu) \geq \frac{\log n_\nu}{m+n} + \frac{n}{m+n} \cdot \frac{1}{n}\log K_\nu \geq (1-\varepsilon')\left(R_\nu - \frac{\varepsilon}{2}\right) \geq R_2 - \varepsilon.$$

The randomness of $G$ is needed in the estimation of the average error incurred by this coding procedure. Recall Ahlswede's Innerproduct Lemma [4].

**Lemma 5.15.** *Let $(\alpha_1, \dots, \alpha_N)$ and $(\beta_1, \dots, \beta_N)$ be two vectors with $0 \le \alpha_m, \beta_m \le 1$ for $m = 1, \dots, N$ which for some $\lambda \in (0,1)$ satisfy*

$$\frac{1}{N}\sum_{m=1}^{N} \beta_m \ge 1-\lambda, \qquad \frac{1}{N}\sum_{m=1}^{N} \alpha_m \ge 1-\lambda, \tag{5.14}$$

*then*

$$\frac{1}{N}\sum_{m=1}^{N} \alpha_m\beta_m \ge 1-2\lambda.$$

Now fix an $\mathbf{s} \in \mathscr{S}^n$ and set for every $(\xi_1, \xi_2) \in [n_1] \times [n_2]$

$$\begin{aligned}\alpha_{\xi_1\xi_2} &= W^{\otimes m}(D_{\xi_1\xi_2}(\gamma^*)|\mathbf{x}_{\xi_1\xi_2}(\gamma^*), \mathbf{y}_{\xi_1\xi_2}(\gamma^*)|\mathbf{s}),\\ \beta_{\xi_1\xi_2} &= \frac{1}{K_1K_2}\sum_{k_1,k_2} W^{\otimes n}\big(D_{k_1k_2}(\gamma_{\xi_1\xi_2})|\mathbf{x}_{k_1k_2}(\gamma_{\xi_1\xi_2}), \mathbf{y}_{k_1k_2}(\gamma_{\xi_1\xi_2})|\mathbf{s}\big).\end{aligned}$$

Replacing $N$ by $n_1n_2$, Lemma 5.15 now implies by the assumptions on $\gamma^*$ and $G$

$$1-\bar{e}^{\mathrm{AV}}(\tilde{\gamma}, \mathscr{W}) = \frac{1}{n_1n_2}\sum_{\xi_1,\xi_2} \alpha_{\xi_1\xi_2}\beta_{\xi_1\xi_2} \ge 1-\lambda$$

This shows that the rate pair $(R_1, R_2)$ is deterministically CONF-achievable for $\mathrm{AV}(\mathscr{W})$ with conferencing capacity sequences $C_1^\infty, C_2^\infty$ under the average error criterion. Consequently one obtains

$$\begin{aligned}&\mathscr{C}_2(\overline{\mathscr{W}}, C_1(\infty), C_2(\infty))\\ &\qquad = \bigcup_{\delta>0} \mathscr{C}_2(\overline{\mathscr{W}}, [C_1(\infty)-\delta]_+, [C_2(\infty)-\delta]_+) \subset \overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\mathscr{W}, C_1^\infty, C_2^\infty),\end{aligned}$$

proving the direct part of Theorem 5.8.

## 5.5. Converses for the AV-MAC with Conferencing Encoders

### 5.5.1. Random Coding

Here we prove the weak converse for Theorem 5.6. The idea of the proof is to reduce it to the weak converse for $\overline{\mathscr{C}}^{\mathrm{Cp,r}}_{\mathrm{CONF}}(\overline{\mathscr{W}}, C_1, C_2)$.

Let $\mathbf{q} \in \mathscr{P}(\mathscr{S})^n$ and $G$ a random $\mathrm{code}_{\mathrm{CONF}}$ with blocklength $n$. We generalize the notation from (5.8) to $\overline{\mathscr{W}}$,

$$\bar{e}_{\mathbf{q}}(G) := \frac{1}{K_1K_2}\sum_{k_1,k_2} W^{\otimes n}(D_{k_1k_2}(G)^c|\mathbf{x}_{k_1k_2}(G), \mathbf{y}_{k_1k_2}(G)|\mathbf{q}).$$

The following lemma is a generalized version of Lemma 2.6.3 in [20].

**Lemma 5.16.** *For any random code*$_{\mathrm{CONF}}$ $G$ *with blocklength* $n$ *one has*

$$\sup_{\mathbf{s}\in\mathscr{S}^n} \mathbb{E}\big[\bar{e}_{\mathbf{s}}(G)\big] = \sup_{\mathbf{q}\in\mathscr{P}(\mathscr{S})^n} \mathbb{E}\big[\bar{e}_{\mathbf{q}}(G)\big].$$

*Proof.* The direction "$\leq$" is clear. In order to prove "$\geq$", let $\mathbf{q} \in \mathscr{P}(\mathscr{S})^n$. Observe that

$$\mathbb{E}[\bar{e}_{\mathbf{q}}(G)] = \sum_{\mathbf{s}\in\mathscr{S}^n} \mathbb{E}[\bar{e}_{\mathbf{s}}(G)] q_1(s_1)\cdots q_n(s_n) \leq \sup_{\mathbf{s}\in\mathscr{S}^n} \mathbb{E}[\bar{e}_{\mathbf{s}}(G)].$$

Upon taking the supremum over $\mathbf{q} \in \mathscr{P}(\mathscr{S})^{\otimes n}$ on the left-hand side, the lemma is proved. □

Now let $G$ be a random $(n, C_1, C_2)$-code$_{\mathrm{CONF}}$ with $\bar{e}^{\mathrm{AV,r}}(G, \mathscr{W}) = \lambda$. Assume that the pair $((1/n)\log K_1, (1/n)\log K_2)$ is at distance at least $\varepsilon$ from $\mathscr{C}_2(\overline{\mathscr{W}}, C_1, C_2)$. For $q \in \mathscr{P}(\mathscr{S})$, set $q^n := (q, \ldots, q) \in \mathscr{P}(\mathscr{S})^n$. Because of Lemma 5.16,

$$\lambda_0 := \bar{e}^{\mathrm{Cp,r}}(G, \overline{\mathscr{W}}) = \sup_{q\in\mathscr{P}(\mathscr{S})} \mathbb{E}[\bar{e}_{q^n}(G)] \leq \lambda. \tag{5.15}$$

But the weak converse for $\overline{\mathscr{C}}^{\mathrm{Cp,r}}_{\mathrm{CONF}}(\overline{\mathscr{W}}, C_1, C_2)$ from Theorem 4.6 implies that (5.15) can only hold if $\lambda_0 \geq \lambda(\varepsilon) > 0$, in particular, $\lambda \geq \lambda(\varepsilon)$. This concludes the weak converse for AV($\mathscr{W}$) with conferencing encoders using random codes$_{\mathrm{CONF}}$ under the average error criterion, and Theorem 5.6 is proved.

## 5.5.2. Deterministic Coding

### If $\mathscr{W}$ is $(\mathscr{X}, \mathscr{Y})$-symmetrizable

If $\mathscr{W}$ is $(\mathscr{X}, \mathscr{Y})$-symmetrizable, then as remarked at the beginning of Paragraph 5.4.3 it is also symmetrizable if considered as a set of stochastic matrices with single-sender input alphabet $\mathscr{X} \times \mathscr{Y}$. Thus Theorem A.7 implies that any single-user code$_{\mathrm{1S}}$ with at least two codewords incurs an average error greater than $1/4$. Finally, note that every code$_{\mathrm{CONF}}$ can be interpreted as a code$_{\mathrm{1S}}$, so this carries over to the multi-user situation. This proves Theorem 5.8 if $\mathscr{W}$ is $(\mathscr{X}, \mathscr{Y})$-symmetrizable.

### If $\mathscr{W}$ is not $(\mathscr{X}, \mathscr{Y})$-symmetrizable

We show that the weak converse for $\overline{\mathscr{C}}^{\mathrm{Cp}}_{\mathrm{CONF}}(\overline{\mathscr{W}}, C_1(\infty), C_2(\infty))$ implies the weak converse for $\overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\mathscr{W}, C_1^{\infty}, C_2^{\infty})$. Let $\gamma$ be a deterministic $(n, C_1(n), C_2(n))$-code$_{\mathrm{CONF}}$. Assume that $n$ is so large that the pair $((1/n)K_1, (1/n)K_2)$ is at least distance $\varepsilon$ away from $\mathscr{C}_2(\overline{\mathscr{W}}, C_1(\infty), C_2(\infty))$ and that for both $\nu = 1, 2$

$$C_\nu(n) \geq [C_\nu(\infty) - \varepsilon]_+ .$$

Thus $((1/n)\log K_1, (1/n)\log K_2)$ is also at least distance $\varepsilon$ away from

$$\mathscr{C}_2(\overline{\mathscr{W}}, [C_1(\infty) - \varepsilon]_+, [C_2(\infty) - \varepsilon]_+) = \overline{\mathscr{C}}^{\mathrm{Cp}}_{\mathrm{CONF}}(\overline{\mathscr{W}}, [C_1(\infty) - \varepsilon]_+, [C_2(\infty) - \varepsilon]_+).$$

By enlarging $n$ if necessary, the weak converse for

$$\overline{\mathscr{C}}^{\mathrm{Cp}}_{\mathrm{CONF}}(\overline{\mathscr{W}}, [C_1(\infty) - \varepsilon]_+, [C_2(\infty) - \varepsilon]_+)$$

ensures that there is a $q \in \mathscr{P}(\mathscr{S})$ such that for some $\lambda(\varepsilon) > 0$

$$\frac{1}{K_1 K_2} \sum_{k_1,k_2} W^{\otimes n}(D_{k_1k_2}(\gamma)^c | \mathbf{x}_{k_1k_2}(\gamma), \mathbf{y}_{k_1k_2}(\gamma) | q^n) \geq \lambda(\varepsilon),$$

recall the notation $q^n$ from the random coding weak converse. Lemma 5.16 now implies that

$$\sup_{\mathbf{s} \in \mathscr{S}^n} \frac{1}{K_1 K_2} \sum_{k_1,k_2} W^{\otimes n}(D_{k_1k_2}(\gamma)^c | \mathbf{x}_{k_1k_2}(\gamma), \mathbf{y}_{k_1k_2}(\gamma) | \mathbf{s}) \geq \lambda(\varepsilon)$$

must hold. Thus the proof of Theorem 5.8 is complete.

## 5.6. Discussion of Conferencing for AV-MACs

For both compound and AV-MACs, conferencing may help to achieve positive rates where only the rate pair $(0, 0)$ is achievable without transmitter cooperation. This effect is similar to the "superactivation" of quantum channels as observed in [51], where it was shown that there are pairs of quantum channels with zero quantum capacity each which achieve positive rates when used together.

Willems conferencing plays two roles in AV-MACs. The "traditional" role already exploited for discrete memoryless MACs is to generate a common message and to use the coding result for the MAC with common message to enlarge the capacity region. The role special to AV-MACs is that conferencing can change the channel structure. To achieve this, it is not necessary to have positive conferencing capacities, the pair of conferencing capacity sequences $(C_1^\infty, C_2^\infty)$ just has to satisfy (5.1). Under the conditions that $\mathscr{W}$ is not $(\mathscr{X}, \mathscr{Y})$-symmetrizable and $\overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\mathscr{W}, 0, 0) \neq \overline{\mathscr{C}}^{\mathrm{AV,r}}_{\mathrm{CONF}}(\mathscr{W}, 0, 0)$, we can strictly enlarge the capacity region of the AV-MAC with this kind of conferencing. Actually in the case of inequality of the random and deterministic capacity regions at $C_1 = C_2 = 0$, $R_1$ or $R_2$ must equal 0 (see Theorem 5.9), so $\mathrm{AV}(\mathscr{W})$ is useless for at least one sender without conferencing. Just a little bit of conferencing (as quantified by eqrefeq:growthcond) suffices to make reliable transmission of that sender's messages over $\mathrm{AV}(\mathscr{W})$ possible. As (5.1) does not exclude $\liminf C_1(n) = \liminf C_2(n) = 0$, it is clear that the change must lie in the channel structure.

The question arises when (5.1) produces such a change of the channel structure. General conditions for $\overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\mathscr{W}, 0, 0) \neq \overline{\mathscr{C}}^{\mathrm{AV,r}}_{\mathrm{CONF}}(\mathscr{W}, 0, 0)$ to hold cannot be given because an exact characterization of $\overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\mathscr{W}, 0, 0)$ is generally unavailable, see Theorem 5.9 and Remark 5.6. We certainly know that if $\overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\mathscr{W}, 0, 0)$ is two-dimensional, then $\overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\mathscr{W}, 0, 0) = \overline{\mathscr{C}}^{\mathrm{AV,r}}_{\mathrm{CONF}}(\mathscr{W}, 0, 0)$. This is due to the fact that two-dimensionality of $\overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\mathscr{W}, 0, 0)$ implies that $\mathscr{W}$ is neither $(\mathscr{X}, \mathscr{Y})$- nor $\mathscr{X}$- nor $\mathscr{Y}$-symmetrizable by Theorem 5.9, which then implies $\overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\mathscr{W}, 0, 0) = \overline{\mathscr{C}}^{\mathrm{AV,r}}_{\mathrm{CONF}}(\mathscr{W}, 0, 0)$.

However, if in addition to not being $(\mathscr{X},\mathscr{Y})$-symmetrizable, $\mathscr{W}$ is both $\mathscr{X}$- and $\mathscr{Y}$-symmetrizable, then $\overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\mathscr{W},0,0) = \{(0,0)\}$. In this situation, Willems conferencing with conferencing sequences $C_1^\infty, C_2^\infty$ satisfying (5.1) helps. Assume that $\liminf C_1(n) = \liminf C_2(n) = 0$. As seen in the proof of Theorem 5.8, coding to obtain $\overline{\mathscr{C}}^{\mathrm{AV}}_{\mathrm{CONF}}(\mathscr{W},C_1^\infty,C_2^\infty) = \overline{\mathscr{C}}^{\mathrm{AV,r}}_{\mathrm{CONF}}(\mathscr{W},0,0)$ can be regarded as a two-step protocol. In this protocol, conferencing is not used to transmit additional messages, but to establish transmission of subexponentially many auxiliary messages in a deterministic prefix $\mathrm{code}_{\mathrm{CONF}}$. The actual message transmission is performed using a random $\mathrm{code}_{\mathrm{CONF}}$, but the underlying random experiment with subexponentially many possible outcomes is only done at the senders and the receiver is informed about the outcome through the prefix $\mathrm{code}_{\mathrm{CONF}}$.

Gubner [30] has found the example of a $\mathscr{W}$ which is both $\mathscr{X}$- and $\mathscr{Y}$-symmetrizable, but not $(\mathscr{X},\mathscr{Y})$-symmetrizable.

*Example* 2. Let $\mathscr{X} = \mathscr{Y} = \mathscr{S} = \{0,1\}$ and $\mathscr{T} = \{0,1,2,3\}$. For $s \in \mathscr{S}$ set

$$W_s(z|x,y) = \delta(z-x-y-s),$$

where $\delta(t) = 1$ if $t = 0$ and $\delta(t) = 0$ else. An equivalent description of this is

$$z = x + y + s.$$

Gubner shows that $\mathscr{W}$ is not $(\mathscr{X},\mathscr{Y})$-symmetrizable, but that it is both $\mathscr{X}$- and $\mathscr{Y}$-symmetrizable. Thus this channel is useless if coding is done without conferencing, even though the interfering signal is only added to the sum of the transmitters' signals – the reliable transmission of messages through the channel is completely prevented. This shows that even the structure of rather simple AV-MACs can be changed by conferencing.

# 6. The Wiretap MAC

## 6.1. The Wiretap MAC

In the previous chapter, we mentioned that an AV-MAC can be seen as a MAC which is under attack: an adversary disturbs the transmission of codewords and might be able to completely prevent communication. A discrete memoryless MAC could also be under a different kind of attack: the adversary could be a wiretapper, i.e. instead of disturbing communication, he might want to overhear communication within the network. Can Willems conferencing help in this situation, too? The traditional approach to making communication secure is cryptography. Information theory provides an alternative, "physical-layer" approach to secrecy. This method exploits the noise inherent in the channel, especially the fact that the noise in channels to legitimate receivers differs from the noise in channels to non-legitimate receivers.

It was noted by Wyner, who introduced the single-sender wiretap channel [67], that a secret key shared at both legitimate terminals is not necessary to establish secret transmission – if the channel statistics are taken into consideration, it is sufficient that the sender randomizes his inputs in order to secure transmission. Since this discovery, information-theoretic secrecy for message transmission without a key shared between sender and legitimate receiver has been generalized in various directions. The first paper on multi-user information-theoretic security is due to Csiszár and Körner [19]. In that article, the second receiver only is a partial eavesdropper: there is a common message intended for both receivers, but as in the original wiretap channel, an additional private message intended for the first receiver must be kept secret from the second. We come to multiple-access models below. An overview over the area is given in [37].

The original secrecy criterion used in [67] and [19] and in most of the subsequent literature until today has become known as the "weak secrecy criterion". Given a code, it measures the mutual information normalized by the code blocklength between the randomly chosen message and the eavesdropper's corresponding output. Maurer introduced the "strong secrecy criterion" in [42] by omitting the normalization. The advantage of this criterion was revealed in [12]: it can be given an operational meaning, i.e. one can specify the attacks it withstands. It is possible to show that if transmission obeys the strong secrecy criterion, then the eavesdropper's average error tends to one for any decoder it might apply. Translated into practical secrecy schemes, this means that no matter how large the computing power of a possible eavesdropper might be, it will not succeed in breaking the security of this scheme. For the weak criterion, there are still only heuristic argumentations as to why it should be secret. Further secrecy metrics are presented in [14], but without giving them an operational meaning, and strong secrecy remains the strongest of these metrics. To our knowledge, there are three different ap-

proaches to establishing strong secrecy in a wiretap channel so far [43, 18, 23]. In fact, the last of these approaches also applies to classical-quantum wiretap channels [23] and also was used to derive an achievable rate for the classical compound wiretap channel [12].

There exist many MAC models where secrecy is an issue. This may even be the case when there is no eavesdropper, as each encoder might have access to noisy observations of the other sender's codeword but wants to protect its own message from decoding at the other sender [40, 36, 27]. The case where the encoders have access to generalized feedback but only keep their messages secret from an external eavesdropper is considered in [52]. In the cognitive MAC, only one encoder has a private message, and together, the encoders have a common message. There are again two cases: In the case without an eavesdropper, the encoder without a private message has access to the codeword sent by the other encoder through a noisy channel and must be kept ignorant of the other encoder's private message [39]. In [49], the cognitive MAC without feedback was investigated where the messages must be kept secret from an eavesdropper and the encoders have unrestricted access to common randomness. All of these papers use the weak secrecy criterion.

We first generalize and strengthen the achievability result from [28] where multi-letter characterizations of an achievable region and of an outer bound on the capacity region of a MAC without common message and with an external eavesdropper under the weak secrecy criterion are given. The channel there needs to satisfy certain relatively strong conditions for the bounds to work. Extensions to the Gaussian case can be found in [28, 53, 31].

We call the two senders $\text{Alice}_1$ and $\text{Alice}_2$. In the common message setting, a message triple must be transmitted to Bob over a discrete memoryless MAC in such a way that Eve who obtains a version of the sent codewords through another discrete memoryless MAC cannot decode the messages. We apply the strong secrecy criterion. In order to find a code which satisfies this criterion, we use Devetak's approach [23], which is similar to the approach taken in [17]. In the quantum case it builds on the Ahlswede-Winter lemma [10] and classically on a Chernoff bound. As the senders have a common message and as the second part of the chapter deals with the wiretap MAC with conferencing encoders, we assume that the encoders have access to a restricted amount of common randomness. Common randomness for encoding has so far only been used in [49], but without setting any limitations on its amount. We only obtain an achievable region. In this achievable region it is not possible to transmit a common message if no common randomness is available. Further it is notable that we use random coding and have to apply time-sharing *before* derandomizing.

The wiretap MAC with common message and common randomness is also needed when treating the wiretap MAC with conferencing encoders. We assume that no common randomness is available. However, conferencing is used to produce both a common message and common randomness, which allows the reduction to the wiretap MAC with common message. A consequence of the fact that no common message can be transmitted by the wiretap MAC with common message if there is no common randomness is that one has to use conferencing to establish some common randomness if this is supposed

to enable the transmission of a common message. Note that this consequence presumes that the achievable region equals the capacity region even though we cannot prove this.

Extensions of information-theoretic security to channels where different messages require different degrees of secrecy can be found in [19] and [68].

## 6.2. The Communication Problems

Let $\mathscr{X}, \mathscr{Y}, \mathscr{T}, \mathscr{Z}$ be finite alphabets. The (discrete memoryless) wiretap MAC is determined by a stochastic matrix

$$W : \mathscr{X} \times \mathscr{Y} \to \mathscr{P}(\mathscr{T} \times \mathscr{Z}).$$

**Definition 6.1.** The *wiretap MAC* $\mathrm{WMAC}(W)$ is the channel

$$W^{\otimes n} : \mathscr{X}^n \times \mathscr{Y}^n \to \mathscr{P}(\mathscr{T}^n \times \mathscr{Z}^n), \quad n = 1, 2, \ldots$$

Of course, a wiretap MAC with output alphabets $\mathscr{T}$ and $\mathscr{Z}$ is the same as a discrete memoryless MAC with output alphabet $\mathscr{T} \times \mathscr{Z}$. However, the wiretap coding problem assigns different roles to $\mathscr{T}$ and $\mathscr{Z}$, which justifies the existence of Definition 6.1. We write $W_b$ and $W_e$ for the stochastic matrices determining the discrete memoryless marginal MACs to $\mathscr{T}$ and $\mathscr{Z}$, so e.g.

$$W_b(t|x, y) := \sum_{z \in \mathscr{Z}} W(t, z|x, y).$$

It is usual in the wiretap setting to give names to the four channel nodes: $\mathscr{X}$ and $\mathscr{Y}$ are the finite alphabets of $\mathrm{Alice}_1$ and $\mathrm{Alice}_2$, respectively. $\mathscr{T}$ is the finite alphabet of the receiver called Bob and the outputs received by the eavesdropper Eve are elements of the finite alphabet $\mathscr{Z}$.

### 6.2.1. With Common Message

**Definition 6.2.** Let $H_C$ be a nonnegative real number. An $(n, H_C)$-$\mathrm{code}_{\mathrm{WCM}}$ is a pair $(G, \varphi)$, where $G$ is a stochastic matrix

$$G : [K_0] \times [K_1] \times [K_2] \to \mathscr{P}(\mathscr{X}^n \times \mathscr{Y}^n)$$

with positive integers $K_0, K_1, K_2$, and $\varphi$ a decoding function

$$\varphi : \mathscr{T}^n \to [K_0] \times [K_1] \times [K_2].$$

$G$ is required to have the form

$$G(\mathbf{x}, \mathbf{y}|k_0, k_1, k_2) = \sum_{j \in [J]} G_0(j|k_0) G_1(\mathbf{x}|k_0, k_1, j) G_2(\mathbf{y}|k_0, k_2, j),$$

where $J$ is some positive integer and

$$\begin{aligned} G_0 &: [K_0] \to \mathscr{P}([J]), \\ G_1 &: [K_0] \times [K_1] \times [J] \to \mathscr{P}(\mathscr{X}), \\ G_2 &: [K_0] \times [K_2] \times [J] \to \mathscr{P}(\mathscr{Y}). \end{aligned}$$

Further, $G_0$ has to satisfy that $H(\Xi|M_0) \leq nH_C$ for $M_0$ uniformly distributed on $[K_0]$ and $P_{\Xi|M_0} = G_0$.

$n$ is called the *blocklength* and $H_C$ the *common randomness bound* of $(G, \varphi)$, the triple $(K_0, K_1, K_2)$ is called its *codelength triple*.

A code$_{\mathrm{WCM}}$ $(G, \varphi)$ is deterministic at the decoder side, but stochastic at the encoder side. Given a message triple $(k_0, k_1, k_2)$, the encoders perform a random experiment described by $G(\,\cdot\,|k_0, k_1, k_2)$. $G_0$ is the common part, an experiment whose outcome both encoders have access to. Then they individually perform the random experiments described by $G_1$ and $G_2$. Altogether, $(k_0, k_1, k_2)$ is encoded into the codeword pair $(\mathbf{x}, \mathbf{y})$ with probability $G(\mathbf{x}, \mathbf{y}|k_0, k_1, k_2)$.

Every $(n, H_C)$-code$_{\mathrm{WCM}}$ is a random $n$-code$_{\mathrm{CM}}$ with deterministic decoder. Even though we denoted complete random codes including the decoders in previous chapters by $G$, the notation introduced above for the wiretap MAC should not be confusing as the random codes$_{\mathrm{CM}}$ from previous chapters do not appear here.

Every wiretap MAC WMAC$(W)$ together with an $(n, H_C)$-code$_{\mathrm{WCM}}$ $(G, \varphi)$ with codelength triple $(K_0, K_1, K_2)$ gives rise to a sequence of random variables. Assume that $M_0, M_1, M_2$ are independent random variables uniformly distributed on $[K_0]$, $[K_1]$ and $[K_2]$, respectively. Further, let $X^n, Y^n, T^n, Z^n$ be random variables such that for $(\mathbf{x}, \mathbf{y}, \mathbf{t}, \mathbf{z}) \in \mathscr{X}^n \times \mathscr{Y}^n \times \mathscr{T}^n \times \mathscr{Z}^n$

$$\begin{aligned} P_{X^nY^n|M_0M_1M_2}(\mathbf{x}, \mathbf{y}|k_0, k_1, k_2) &= G(\mathbf{x}, \mathbf{y}|k_0, k_1, k_2), \\ P_{T^nZ^n|X^nY^nM_0M_1M_2}(\mathbf{t}, \mathbf{z}|\mathbf{x}, \mathbf{y}, k_0, k_1, k_2) &= W^{\otimes n}(\mathbf{t}, \mathbf{z}|\mathbf{x}, \mathbf{y}). \end{aligned}$$

**Definition 6.3.** For a wiretap MAC WMAC$(W)$ and an $(n, H_C)$-code$_{\mathrm{WCM}}$ $(G, \varphi)$ we define the *average error* of $(G, \varphi)$ by

$$\bar{e}^{\mathrm{WT}}(G, \varphi, W) := \mathbb{P}[\varphi(T^n) \neq (M_0, M_1, M_2)].$$

$\bar{e}^{\mathrm{WT}}(G, \varphi, W)$ is the obvious generalization of the average error $\bar{e}^{\mathrm{DM}}(f, \varphi, \tilde{W})$ of deterministic $n$-codes$_{\mathrm{CM}}$ $(f, \varphi)$ for discrete memoryless MACs DMAC$(\tilde{W})$ to $n$-codes$_{\mathrm{CM}}$ with stochastic encoding for DMAC$(W_b)$.

**Definition 6.4.** A triple $(R_0, R_1, R_2)$ of nonnegative real numbers is called a *WCM-achievable rate triple for* WMAC$(W)$ *under the common randomness bound* $H_C \geq 0$ if for every $\eta > 0$ and every $\varepsilon \in (0, 1)$ and $n \geq n_0(\eta, \varepsilon)$ there exists an $(n, H_C)$-code$_{\mathrm{WCM}}$

$(G, \varphi)$ satisfying

$$\begin{aligned} \frac{1}{n} \log K_\nu &\geq R_\nu - \eta \qquad (\nu = 0, 1, 2), \\ \bar{e}^{\mathrm{WT}}(G, \varphi, W) &\leq \varepsilon, \\ I(Z^n \wedge M_0 M_1 M_2) &\leq \varepsilon. \end{aligned}$$

We denote the set of $_{\mathrm{WCM}}$-achievable rate triples under the common randomness bound $H_C$ by $\overline{\mathscr{C}}_{\mathrm{WCM}}(W, H_C)$.

*Remark* 6.1. It was shown in [12] that no matter how Eve tries to decode the messages from the Alices, the average error must tend to one. More precisely, assume that we are given a wiretap MAC WMAC($W$) and for every $n$ an $(n, H_C)$-code$_{\mathrm{WCM}}$ with codelength triple $(K_0(n), K_1(n), K_2(n))$ satisfying

$$\varepsilon_n := I(Z^n \wedge M_0(n) M_1(n) M_2(n)) \longrightarrow 0,$$

where $M_\nu(n)$ is uniformly distributed on $K_\nu(n)$, $\nu = 0, 1, 2$. In this case we say that the code sequences satisfies the *strong secrecy criterion*.

Further assume that for every $n$ Eve has a decoding function

$$\chi_n : \mathscr{Z}^n \to [K_0(n)] \times [K_1(n)] \times [K_2(n)].$$

Then

$$\mathbb{P}[\chi_n(Z^n) \neq (M_0(n), M_1(n), M_2(n))] \geq 1 - \varepsilon'_n$$

for a sequence $\varepsilon'_n$ with $\varepsilon'_n \to 0$ as $\varepsilon \to 0$. If $\varepsilon_n$ tends to zero exponentially fast and $K_0(n), K_1(n), K_2(n)$ grow exponentially, then $\varepsilon'_n$ tends to zero at exponential speed.

More generally assume that $f_n : [K_0(n)] \times [K_1(n)] \times [K_2(n)] \to [K(n)']$ is a function satisfying $\mathbb{P}[f_n(M_0(n), M_1(n), M_2(n)) = k'] = 1/K(n)'$ for all $k' \in [K(n)']$. Then with the same argument as in [12] one can show that for every function $g_n : \mathscr{Z}^n \to [K(n)']$, one has $\mathbb{P}[f_n(M_0(n), M_1(n), M_2(n)) \neq g(Z^n)] \geq 1 - 1/K(n)' - \varepsilon'$ for the same $\varepsilon'_n$ as above. That is, even for $K(n)' = 2$, blind guessing is the best way for Eve to estimate $f(M_0(n), M_1(n), M_2(n))$. In particular, no subset of the message random variables, like $M_0(n)$ or $(M_1(n), M_2(n))$, can be reliably decoded by Eve.

### 6.2.2. With Conferencing Encoders

As wiretap encoding is stochastic in general, this carries over to the conferencing protocols employed – the encoders may share randomness in addition to information about their messages. This generalization is straightforward. Assume that the respective message sets are $[K_1]$ and $[K_2]$. Let $J_1$ and $J_2$ be positive integers which can be written as products

$$J_\nu = J_{\nu,1} \cdots J_{\nu,I} \qquad (\nu = 1, 2)$$

for some positive integer $I$ which does not depend on $\nu$. A *Willems conferencing stochastic matrix* $c$ completely describing such a conference is determined in an iterative manner via sequences of stochastic matrices $c_{1,1}, \ldots, c_{1,I}$ and $c_{2,1}, \ldots, c_{2,I}$. $c_{1,i}$ describes the probability distribution of what $\text{Alice}_1$ tells $\text{Alice}_2$ in the $i$-th conferencing iteration given the knowledge accumulated so far at $\text{Alice}_1$. Thus in general, using the notation (2.10), these stochastic matrices satisfy for $\nu = 1, 2$ and $i = 2, \ldots, I$,

$$\begin{aligned} c_{\nu,1} &: [K_\nu] \to \mathscr{P}([J_{\nu,1}]), \\ c_{\nu,i} &: [K_\nu] \times [J_{\bar{\nu},1}] \times \ldots \times [J_{\bar{\nu},i-1}] \to \mathscr{P}([J_{\nu,i}]). \end{aligned}$$

The conferencing stochastic matrix $c : [K_1] \times [K_2] \to \mathscr{P}([J_1] \times [J_2])$ is obtained by setting

$$\begin{aligned} &c(j_{1,1}, \ldots, j_{1,I}, j_{2,1}, \ldots, j_{2,I} | k_1, k_2) \\ &:= \big(c_{1,1}(j_{1,1}|k_1)\, c_{2,1}(j_{2,1}|k_2)\big)\big(c_{1,2}(j_{1,2}|k_1, j_{2,1})\, c_{2,2}(j_{2,2}|k_2, j_{1,1})\big) \cdots \\ &\quad \cdots \big(c_{1,I}(j_{1,I}|k_1, j_{2,1}, \ldots, j_{2,I-1})\, c_{2,I}(j_{2,I}|k_2, j_{1,1}, \ldots, j_{1,I-1})\big). \end{aligned}$$

We denote the $[J_1]$- and $[J_2]$-marginals of this stochastic matrix by $c_1$ and $c_2$, so one obtains $c_1(j_{1,1}, \ldots, j_{1,I}|k_1, k_2)$ by summing over $j_{2,1}, \ldots, j_{2,I}$ and $c_2$ is obtained analogously.

**Definition 6.5.** Let $n$ be a positive integer and $C_1, C_2$ nonnegative real numbers. A stochastic matrix

$$c : [K_1] \times [K_2] \to \mathscr{P}([J_1] \times [J_2])$$

as described above which satisfies (2.11) is called a *stochastic* $(n, C_1, C_2)$*-Willems conference*, $C_1, C_2$ are called the *conferencing capacities*.

**Definition 6.6.** Let $n$ be a positive integer and $C_1, C_2 \geq 0$. An $(n, C_1, C_2)$-$\text{code}_{\text{WCONF}}$ with alphabets $\mathscr{X}, \mathscr{Y}, \mathscr{T}, \mathscr{Z}$ is a quadruple $(c, G_1, G_2, \varphi)$, where

$$c : [K_1] \times [K_2] \to \mathscr{P}([J_1] \times [J_2])$$

is a stochastic $(n, C_1, C_2)$-Willems conference,

$$\begin{aligned} G_1 &: [K_1] \times [J_2] \to \mathscr{X}^n, \\ G_2 &: [K_2] \times [J_1] \to \mathscr{Y}^n \end{aligned}$$

are stochastic matrices and

$$\varphi : \mathscr{T}^n \to [K_1] \times [K_2].$$

$n$ is called the *blocklength* and $(K_1, K_2)$ the *codelength pair* of $(c, G_1, G_2, \varphi)$.

A pair $(k_1, k_2) \in [K_1] \times [K_2]$ is encoded into the codeword pair $(\mathbf{x}, \mathbf{y}) \in \mathscr{X}^n \times \mathscr{Y}^n$ with probability

$$\sum_{(j_1, j_2) \in [J_1] \times [J_2]} c(j_1, j_2|k_1, k_2)\, G_1(\mathbf{x}|k_1, j_2)\, G_2(\mathbf{y}|k_2, j_1). \tag{6.1}$$

A wiretap MAC WMAC($W$) together with an $(n, C_1, C_2)$-code$_{\mathrm{WCONF}}$ $(c, G_1, G_2, \varphi)$ gives rise to a sequence of random variables. Assume that $M_1, M_2$ are independent random variables uniformly distributed on $[K_1]$ and $[K_2]$, respectively. Let $X^n, Y^n, T^n, Z^n$ be random variables such that conditional on $(M_1, M_2)$, the distribution of $(X^n, Y^n)$ is given by (6.1) and such that

$$P_{T^n Z^n | X^n Y^n M_1 M_2} = W^{\otimes n}.$$

**Definition 6.7.** For a wiretap MAC WMAC($W$) and an $(n, C_1, C_2)$-code$_{\mathrm{WCONF}}$ $(c, G_1, G_2, \varphi)$, we define the *average error* of $(c, G_1, G_2, \varphi)$ as

$$\bar{e}^{\mathrm{WT}}(c, G_1, G_2, \varphi, W) := \mathbb{P}[\varphi(T^n) \neq (M_1, M_2)].$$

**Definition 6.8.** A pair $(R_1, R_2)$ of nonnegative real numbers is called a *WCONF-achievable rate pair for* WMAC($W$) *with conferencing capacities* $C_1, C_2 \geq 0$ if for every $\eta > 0$ and every $\varepsilon \in (0, 1)$ and $n \geq n_0(\eta, \varepsilon)$ there exists an $(n, C_1, C_2)$-code$_{\mathrm{WCONF}}$ satisfying

$$\begin{aligned} \frac{1}{n} \log K_\nu &\geq R_\nu - \eta \qquad (\nu = 1, 2), \\ \bar{e}^{\mathrm{WT}}(c, G_1, G_2, \varphi, W) &\leq \varepsilon, \\ I(Z^n \wedge M_1 M_2) &\leq \varepsilon. \end{aligned}$$

We denote the set of $_{\mathrm{WCONF}}$-achievable rate pairs with conferencing capacities $C_1, C_2$ by $\overline{\mathscr{C}}_{\mathrm{WCONF}}(W, C_1, C_2)$.

*Remark* 6.2. Here again, as in Remark 6.1, the average decoding error for any decoder Eve might apply tends to 1 if the strong secrecy criterion is satisfied.

## 6.3. Coding Theorems

### 6.3.1. For the Wiretap MAC with Common Message

For the description of the WCM-achievable regions we will derive we need the following definition.

**Definition 6.9.** For $W : \mathscr{X} \times \mathscr{Y} \to \mathscr{P}(\mathscr{T} \times \mathscr{Z})$, we set

$$\begin{aligned} \Psi(W) := \{p \in \mathscr{P}(U \times \mathscr{V}_1 \times \mathscr{V}_2 \times \mathscr{X} \times \mathscr{Y} \times \mathscr{T} \times \mathscr{Z} : \\ \mathscr{U}, \mathscr{V}_1, \mathscr{V}_2 \text{ finite subsets of the integers,} \\ p = P_U \otimes (P_{V_1|U} \otimes P_{V_2|U}) \otimes (P_{X|V_1} \otimes P_{Y|V_2}) \otimes W\}. \end{aligned}$$

Let $H_C \geq 0$ be the common randomness bound. We are going to prove the WCM-achievability of a region which be written as the closure of the convex hull of the union of certain rate sets which are parametrized by the elements of a subset $\Psi_{H_C}(W)$ of the set $\Psi(W)$. $\Psi_{H_C}(W)$ is defined as follows.

There are four cases altogether, numbered Case 0 to Case 3. Case 0 corresponds to $H_C = 0$ and if $H_C > 0$, then $\Psi_{H_C}(W)$ has the form $\Psi_{H_C}(W) = \Psi^{(1)}_{H_C}(W) \cup \Psi^{(2)}_{H_C}(W) \cup \Psi^{(3)}_{H_C}(W)$, and each of these subsets corresponds to one of these cases. The one condition all cases have in common is that $I(Z \wedge V_1V_2) \leq I(T \wedge V_1V_2)$.

**Case 0:** If $H_C = 0$ define the set $\Psi^{(0)}(W)$ as the set of those $p \in \Psi(W)$ where $V_1$ and $V_2$ are independent of $U$ and where $p$ satisfies the inequalities

$$I(Z \wedge V_1) \leq I(T \wedge V_1|V_2), \tag{6.2}$$
$$I(Z \wedge V_2) \leq I(T \wedge V_2|V_1). \tag{6.3}$$

Thus in this case, we can omit $U$ and just assume that $V_1$ and $V_2$ are independent. For $p \in \Psi^{(0)}(W)$ define the set $\mathscr{R}^{(0)}(p)$ to be the set of nonnegative triples $(R_0, R_1, R_2)$ satisfying

$$\begin{aligned} R_0 &= 0, \\ R_1 &\leq I(T \wedge V_1|V_2) - I(Z \wedge V_1) - [I(Z \wedge V_2|V_1) - I(T \wedge V_2|V_1)]_+, \\ R_2 &\leq I(T \wedge V_2|V_1) - I(Z \wedge V_2) - [I(Z \wedge V_1|V_2) - I(T \wedge V_1|V_2)]_+, \\ R_1 + R_2 &\leq I(T \wedge V_1V_2) - I(Z \wedge V_1V_2). \end{aligned}$$

**Case 1:** $\Psi^{(1)}_{H_C}(W)$ is the set of those $p \in \Psi(W)$ which satisfy $I(Z \wedge U) < H_C$ and

$$I(Z \wedge V_1|U) \leq I(T \wedge V_1|V_2U), \tag{6.4}$$
$$I(Z \wedge V_2|U) \leq I(T \wedge V_2|V_1U), \tag{6.5}$$
$$I(Z \wedge V_1V_2|U) \leq I(T \wedge V_1V_2|U). \tag{6.6}$$

Then we denote by $\mathscr{R}^{(1)}(p)$ the set of nonnegative real triples $(R_0, R_1, R_2)$ satisfying

$$\begin{aligned} R_1 &\leq I(T \wedge V_1|V_2U) - I(Z \wedge V_1|U) - [I(Z \wedge V_2|V_1U) - I(T \wedge V_2|V_1U)]_+, \\ R_2 &\leq I(T \wedge V_2|V_1U) - I(Z \wedge V_2|U) - [I(Z \wedge V_1|V_2U) - I(T \wedge V_1|V_2U)]_+, \\ R_1 + R_2 &\leq I(T \wedge V_1V_2|U) - I(Z \wedge V_1V_2|U), \\ R_0 + R_1 + R_2 &\leq I(T \wedge V_1V_2) - I(Z \wedge V_1V_2). \end{aligned}$$

**Case 2:** The conditions for $p$ to be contained in $\Psi^{(2)}_{H_C}(W)$ cannot be phrased as simply as for $\Psi^{(1)}_{H_C}(W)$. Generally, if $p \in \Psi^{(2)}_{H_C}(W)$ then

$$\min\{I(Z \wedge V_1U), I(Z \wedge V_2U)\} < H_C \leq I(Z \wedge V_1V_2).$$

This is sufficient if $I(Z \wedge V_1|V_2U) = I(Z \wedge V_2|V_1U)$. If $I(Z \wedge V_1|V_2U) > I(Z \wedge V_2|V_1U)$ then we additionally require that

$$\begin{aligned}\alpha_0^{(2)} &:= \max\left(\frac{I(Z \wedge V_1U) - H_C}{I(Z \wedge V_1|V_2U) - I(Z \wedge V_2|V_1U)}, 1 - \frac{I(T \wedge V_2|V_1U)}{I(Z \wedge V_2|V_1U)}, 0\right)\\ \leq \alpha_1^{(2)} &:= \min\left(\frac{I(T \wedge V_1|V_2U)}{I(Z \wedge V_1|V_2U)}, \frac{I(T \wedge V_1V_2|U) - I(Z \wedge V_2|V_1U)}{I(Z \wedge V_1|V_2U) - I(Z \wedge V_2|V_1U)}, 1\right)\end{aligned}$$

whereas if $I(Z \wedge V_1|V_2U) < I(Z \wedge V_2|V_1U)$ then we need

$$\begin{aligned}\alpha_0^{(2)} &:= \max\left(1 - \frac{I(T \wedge V_2|V_1U)}{I(Z \wedge V_2|V_1U)}, \frac{I(T \wedge V_1V_2|U) - I(Z \wedge V_2|V_1U)}{I(Z \wedge V_1|V_2U) - I(Z \wedge V_2|V_1U)}, 0\right)\\ \leq \alpha_1^{(2)} &:= \min\left(\frac{H_C - I(Z \wedge V_1U)}{I(Z \wedge V_2|V_1U) - I(Z \wedge V_1|V_2U)}, \frac{I(T \wedge V_1|V_2U)}{I(Z \wedge V_1|V_2U)}, 1\right).\end{aligned}$$

In the case of equality, i.e. if $I(Z \wedge V_1|V_2U) = I(Z \wedge V_2|V_1U)$, we define $\mathscr{R}^{(2)}(p)$ as

$$\begin{aligned}R_1 &\leq I(T \wedge V_1|V_2U),\\ R_2 &\leq I(T \wedge V_2|V_1U),\\ R_1 + R_2 &\leq I(T \wedge V_1V_2|U) - I(Z \wedge V_1|V_2U),\\ R_0 + R_1 + R_2 &\leq I(T \wedge V_1V_2) - I(Z \wedge V_1V_2).\end{aligned}$$

If $I(Z \wedge V_1|V_2U) > I(Z \wedge V_2|V_1U)$, we define $\mathscr{R}^{(2)}(p)$ by

$$\begin{aligned}R_1 &\leq I(T \wedge V_1|V_2U) - \alpha_0^{(2)} I(Z \wedge V_1|V_2U),\\ R_2 &\leq I(T \wedge V_2|V_1U) - (1 - \alpha_1^{(2)}) I(Z \wedge V_2|V_1U),\\ R_1 + R_2 &\leq I(T \wedge V_1V_2|U) - \alpha_0^{(2)} I(Z \wedge V_1|V_2U)\\ &\qquad - (1 - \alpha_0^{(2)}) I(Z \wedge V_2|V_1U), \qquad (6.7)\\ R_1 + \frac{I(Z \wedge V_2|V_1U)}{I(Z \wedge V_1|V_2U)} R_2 &\leq I(T \wedge V_2|V_1U)\\ &\qquad + \left(\frac{I(T \wedge V_1|U)}{I(Z \wedge V_1|V_2U)} - 1\right) I(Z \wedge V_2|V_1U), \qquad (6.8)\\ R_0 + R_1 + R_2 &\leq I(T \wedge V_1V_2) - I(Z \wedge V_1V_2).\end{aligned}$$

The bound (6.7) on $R_1 + R_2$ can be reformulated as

$$\begin{aligned}R_1 + R_2 \leq\ & I(T \wedge V_1V_2|U) - I(Z \wedge V_1V_2|U)\\ &+ \min\Big\{H_C - I(Z \wedge U), I(Z \wedge V_1|U),\\ &\quad I(T \wedge V_1|V_2U)\left(\frac{I(Z \wedge V_2|V_1U)}{I(Z \wedge V_1|V_2U)} - 1\right) + I(Z \wedge V_1|U)\Big\},\end{aligned}$$

and if $I(Z \wedge V_2|V_1U) > 0$, we can give the weighted sum bound (6.8) the almost symmetric form

$$\frac{R_1}{I(Z \wedge V_1|V_2U)} + \frac{R_2}{I(Z \wedge V_2|V_1U)} \leq \frac{I(T \wedge V_1|U)}{I(Z \wedge V_1|V_2U)} + \frac{I(T \wedge V_2|V_1U)}{I(Z \wedge V_2|V_1U)} - 1.$$

For the case that $I(Z \wedge V_1|V_2U) < I(Z \wedge V_2|V_1U)$, we define $\mathscr{R}^{(2)}(p)$ by exchanging the roles of $V_1$ and $V_2$.

**Case 3:** We define $\Psi^{(3)}_{H_C}(W)$ to be the set of those $p \in \Psi(W)$ with $I(Z \wedge V_1V_2) < H_C$. For such a $p$ let $\mathscr{R}^{(3)}(p)$ equal

$$\begin{aligned} R_1 &\leq I(T \wedge V_1|V_2U), \\ R_2 &\leq I(T \wedge V_2|V_1U), \\ R_1 + R_2 &\leq I(T \wedge V_1V_2|U), \\ R_0 + R_1 + R_2 &\leq I(T \wedge V_1V_2) - I(Z \wedge V_1V_2). \end{aligned}$$

**Theorem 6.10.** *For* WMAC$(W)$ *with the common randomness bound* $H_C = 0$*,*

$$\text{closure}\left(\text{conv}\left(\bigcup_{p \in \Psi^{(0)}(W)} \mathscr{R}^{(0)}(p)\right)\right) \subset \overline{\mathscr{C}}_{\text{WCM}}(W, 0). \tag{6.9}$$

*If* $H_C > 0$*, then the closure of the convex hull of the set*

$$\bigcup_{p \in \Psi^{(1)}_{H_C}(W)} \mathscr{R}^{(1)}(p) \;\cup \bigcup_{p \in \Psi^{(2)}_{H_C}(W)} \mathscr{R}^{(2)}(p) \;\cup \bigcup_{p \in \Psi^{(3)}_{H_C}(W)} \mathscr{R}^{(3)}(p)$$

*is contained in* $\overline{\mathscr{C}}_{\text{WCM}}(W, H_C)$*.*

*Remark* 6.3. Using the standard Carathéodory-Fenchel technique as in [63], one can show that one may without loss of generality assume $|\mathscr{U}| \leq |\mathscr{X}||\mathscr{Y}| + 5$. However, $|\mathscr{V}_1|$ and $|\mathscr{V}_2|$ cannot be bounded in this way, as the application of the Carathéodory-Fenchel theorem does not preserve the conditional independence of $V_1$ and $V_2$. Thus a characterization of the above WCM-achievable region involving auxiliary sets with upper-bounded cardinality is currently not available. As it would be important for an efficient calculation of the WCM-achievable region, it still requires further consideration.

*Remark* 6.4. If no common randomness is available, then we cannot show the achievability of any rate triple $(R_0, R_1, R_2)$ with $R_0 > 0$. As we do not have a converse for $\overline{\mathscr{C}}_{\text{WCM}}(W, 0)$, however, this does not mean that the secret transmission of a common message without common randomness is impossible.

*Remark* 6.5. We have $\mathscr{R}^{(1)}(p) \subset \mathscr{R}^{(2)}(p) \subset \mathscr{R}^{(3)}(p)$. This can be seen directly at the beginning of the proof in Subsection 6.4.1 where we decompose the regions $\mathscr{R}^{(\nu)}(p)$ for $\nu = 1, 2$ into a union of simpler regions.

In particular, if $H_C$ is larger than the capacity of $W_e$ considered as a single-sender discrete memoryless channel with input alphabet $\mathscr{X} \times \mathscr{Y}$ and output alphabet $\mathscr{Z}$, i.e.

if $H_C \geq \overline{\mathscr{C}}_{1S}(W_e)$ (see Appendix A), then $\Psi^{(3)}_{H_C}(W) = \Psi(W)$ and the WCM-achievable set equals

$$\mathrm{closure}\left(\mathrm{conv}\left(\bigcup_{p\in\Psi(W)} \mathscr{R}^{(3)}(p)\right)\right).$$

In this case the maximal sum rate equals

$$\overline{\mathscr{C}}^{\mathrm{WT}}_{1S}(W) := \max_{p\in\Psi(W)} \big(I(T\wedge V_1V_2) - I(Z\wedge V_1V_2)\big). \tag{6.10}$$

$\overline{\mathscr{C}}^{\mathrm{WT}}_{1S}(W)$ equals the secrecy capacity of the single-sender wiretap channel $W$ when $\mathrm{Alice}_1$ and $\mathrm{Alice}_2$ together are considered as one single sender. This can be seen as in the proof of Lemma 2.14. The remaining conditions on $R_1$ and $R_2$ formulated in the definition of $\mathscr{R}^{(3)}(p)$ are not concerned with $W_e$, they are required by the non-wiretap MAC coding theorem applied to $W_b$.

### 6.3.2. For the Wiretap MAC with Conferencing Encoders

For conferencing capacities $C_1, C_2 > 0$, the rate region whose WCONF-achievability we are going to prove is parametrized by the members of $\Psi_{C_1+C_2}(W)$. We have Cases 1-3 from the common message part.

**Case 1:** For $p \in \Psi^{(1)}_{C_1+C_2}(W)$ we define $\mathscr{R}^{(1)}(p, C_1, C_2)$ by

$$\begin{aligned}
R_1 &\leq I(T\wedge V_1|V_2U) - I(Z\wedge V_1|U)\\
&\quad -[I(Z\wedge V_2|V_1U) - I(T\wedge V_2|V_1U)]_+ + C_1 - [I(Z\wedge U) - C_2]_+,\\
R_2 &\leq I(T\wedge V_2|V_1U) - I(Z\wedge V_2|U)\\
&\quad -[I(Z\wedge V_1|V_2U) - I(T\wedge V_1|V_2U)]_+ + C_2 - [I(Z\wedge U) - C_1]_+,\\
R_1+R_2 &\leq \min\{I(T\wedge V_1V_2|U) + C_1 + C_2, I(T\wedge V_1V_2))\} - I(Z\wedge V_1V_2)).
\end{aligned}$$

**Case 2:** For $p \in \Psi^{(2)}_{C_1+C_2}(W)$, we set $J_0^{(\alpha)} := \alpha I(Z\wedge V_2U) + (1-\alpha)I(Z\wedge V_1U)$. For $\alpha \in [\alpha_0^{(2)}, \alpha_1^{(2)}]$ define the set $\mathscr{R}^{(2)}_\alpha(p, C_1, C_2)$ by

$$\begin{aligned}
R_1 &\leq I(T\wedge V_1|V_2U) - \alpha I(Z\wedge V_1|V_2U) + C_1 - [J_0^{(\alpha)} - C_2]_+,\\
R_2 &\leq I(T\wedge V_2|V_1U) - (1-\alpha) I(Z\wedge V_2|V_1U) + C_2 - [J_0^{(\alpha)} - C_1]_+,\\
R_1+R_2 &\leq \min\{I(T\wedge V_1V_2|U) + C_1 + C_2, I(T\wedge V_1V_2)\} - I(Z\wedge V_1V_2).
\end{aligned}$$

Then we set

$$\mathscr{R}^{(2)}(p, C_1, C_2) := \bigcup_{\alpha_0^{(2)}\leq\alpha\leq\alpha_1^{(2)}} \mathscr{R}^{(2)}_\alpha(p, C_1, C_2).$$

**Case 3:** For $p \in \Psi^{(3)}_{C_1+C_2}(W)$ we define $\mathscr{R}^{(3)}(p, C_1, C_2)$ by

$$\begin{aligned}
R_1 &\leq I(T\wedge V_1|V_2U) + C_1 - [I(Z\wedge V_1V_2) - C_2]_+,\\
R_2 &\leq I(T\wedge V_2|V_1U) + C_2 - [I(Z\wedge V_1V_2) - C_1]_+,\\
R_1+R_2 &\leq \min\{I(T\wedge V_1V_2|U) + C_1 + C_2, I(T\wedge V_1V_2)\} - I(Z\wedge V_1V_2).
\end{aligned}$$

**Theorem 6.11.** *For the wiretap MAC* $\mathrm{WMAC}(W)$ *with conferencing capacities* $C_1, C_2$ *satisfying* $\max\{C_1, C_2\} \geq 0$*, the set*

$$\bigcup_{p\in\Psi^{(1)}_{C_1+C_2}(W)} \mathscr{R}^{(1)}(p, C_1, C_2) \cup \bigcup_{p\in\Psi^{(2)}_{C_1+C_2}(W)} \mathscr{R}^{(2)}(p, C_1, C_2) \cup \bigcup_{p\in\Psi^{(3)}_{C_1+C_2}(W)} \mathscr{R}^{(3)}(p, C_1, C_2)$$

*is contained in* $\overline{\mathscr{C}}_{\mathrm{WCONF}}(W, C_1, C_2)$*.*

*Remark* 6.6. Remark 6.3 applies here, too.

*Remark* 6.7. The stochastic conferencing protocols employed to achieve the sets in Theorem 6.11 are non-iterative. That means that the $c$ we use in the proof have the form

$$c(j_1, j_2|k_1, k_2) = c_1(j_1|k_1)c_2(j_2|k_2).$$

*Remark* 6.8. If $C_1 = C_2 = 0$, then the maximal rate set whose WCONF-achievability we can show is the left-hand side of (6.9). Conferencing only enlarges this set in the presence of a wiretapper if it is used to establish common randomness between the encoders. At least this is true for the WCONF-achievable region we can show, it cannot be verified in general as long as one does not have a converse. The reason for this effect is that conferencing generates a common message shared by $\mathrm{Alice}_1$ and $\mathrm{Alice}_2$. The proof of Theorem 6.11 relies on Theorem 6.10, and as noted in Remark 6.4, we can only establish the secret transmission of a common message if common randomness is available. As the Alices do not have common randomness a priori, this also has to be generated by conferencing, so the Willems conferencing protocol has to be stochastic.

*Remark* 6.9. Judging from our achievable regions, conferencing may enable secure transmission if this is not possible without. That means that there are wiretap MACs where the WCONF-achievable region with $C_1 = C_2 = 0$ on the left-hand side of (6.9) only contains the rate pair $(0, 0)$, whereas the achievable regions for $\max\{C_1, C_2\} > 0$ contain non-trivial rate pairs. Of course, we again need to keep in mind that we do not have a converse for $\overline{\mathscr{C}}_{\mathrm{WCONF}}(W, 0, 0)$. See Section 6.6 for an example.

*Remark* 6.10. If $C_1, C_2$ are sufficiently large, then the maximal WCONF-achievable sum rate equals $\overline{\mathscr{C}}^{\mathrm{WT}}_{1\mathrm{S}}(W)$, see (6.10). In fact, this happens if

1) $C_1 + C_2 > \overline{\mathscr{C}}_{1\mathrm{S}}(W_e)$, the capacity of the single-sender discrete memoryless channel $W_e$ with input alphabet $\mathscr{X} \times \mathscr{Y}$ and output alphabet $\mathscr{Z}$ (see Appendix A),

2) $C_1 + C_2 \geq \min_{p\in\Psi^*(W)} I(T \wedge U)$, where $\Psi^*(W)$ contains those $p \in \Psi(W)$ which achieve $\overline{\mathscr{C}}^{\mathrm{WT}}_{1\mathrm{S}}(W)$.

Condition 1) is sufficient to guarantee that $\overline{\mathscr{C}}^{\mathrm{WT}}_{1\mathrm{S}}(W)$ is attainable by an element of $\Psi^{(3)}_{C_1+C_2}(W)$ which then equals $\Psi(W)$, see Remark 6.5. In particular $\Psi^*(W)$ is nonempty, and 2) ensures that the maximum over $\Psi(W)$ of the sum rate bounds from $\mathscr{R}^{(3)}(p, C_1, C_2)$ equals $\overline{\mathscr{C}}^{\mathrm{WT}}_{1\mathrm{S}}(W)$.

## 6.4. Proof of Theorem 6.10

### 6.4.1. Elementary Rate Regions

For Cases 0, 1 and 2 we first represent the sets $\mathscr{R}^{(0)}(p), \mathscr{R}^{(1)}(p), \mathscr{R}^{(2)}(p)$ as unions of convex combinations of sets whose WCM-achievability will be shown using random coding later.

#### For Case 0 and 1:

We only consider Case 1, Case 0 is analogous. Let $p \in \Psi^{(1)}_{H_C}(W)$ for some $H_C > 0$. The considerations hold for $I(Z \wedge V_1|U) < I(Z \wedge V_1|V_2U)$ which is equivalent to $I(Z \wedge V_2|U) < I(Z \wedge V_2|V_1U)$. In the case of equality we can prove the WCM-achievability of $\mathscr{R}(p)$ directly. Define

$$\alpha_0^{(1)} := \left[\frac{I(T \wedge V_2|V_1U) - I(Z \wedge V_2|V_1U)}{I(Z \wedge V_2|U) - I(Z \wedge V_2|V_1U)}\right]_+,$$
$$\alpha_1^{(1)} := \min\left\{\frac{I(T \wedge V_1|V_2U) - I(Z \wedge V_1|U)}{I(Z \wedge V_1|V_2U) - I(Z \wedge V_1|U)}, 1\right\}.$$

Note that conditions (6.4)-(6.6) are equivalent to $\alpha_0^{(1)} \le \alpha_1^{(1)}$. For $\alpha \in [\alpha_0^{(1)}, \alpha_1^{(1)}]$ we define a rate region $\mathscr{R}_\alpha^{(1)}(p)$ by the bounds

$$\begin{aligned}
R_1 &\le I(T \wedge V_1|V_2U) - \alpha I(Z \wedge V_1|V_2U) - (1-\alpha) I(Z \wedge V_1|U),\\
R_2 &\le I(T \wedge V_2|V_1U) - \alpha I(Z \wedge V_2|U) - (1-\alpha) I(Z \wedge V_2|V_1U),\\
R_1 + R_2 &\le I(T \wedge V_1V_2|U) - I(Z \wedge V_1V_2|U),\\
R_0 + R_1 + R_2 &\le I(T \wedge V_1V_2) - I(Z \wedge V_1V_2).
\end{aligned}$$

**Lemma 6.12.** *We have*

$$\mathscr{R}^{(1)}(p) = \bigcup_{\alpha_0^{(1)} \le \alpha \le \alpha_1^{(1)}} \mathscr{R}_\alpha^{(1)}(p).$$

*Thus if $\mathscr{R}_\alpha^{(1)}(p)$ is an WCM-achievable rate region for every $\alpha \in [\alpha_0^{(1)}, \alpha_1^{(1)}]$, then $\mathscr{R}^{(1)}(p)$ is WCM-achievable.*

For the proof we use the following lemma which is proved in Appendix B.

**Lemma 6.13.** *Assume that $a_1, a_2, b_1, b_2, c, d, r_1, r_2, r_{12}, r_{012}$ are nonnegative reals satisfying*

$$a_1 > b_1, \quad a_2 < b_2, \quad a_1 + a_2 = b_1 + b_2 = c, \quad r_1 + r_2 \ge r_{12}.$$

*Let $0 \le \alpha_0 \le \alpha_1 \le 1$. For every $\alpha \in [\alpha_0, \alpha_1]$, let a three-dimensional convex subset $\mathscr{K}_\alpha$ of $\mathbb{R}^3_{\ge 0}$ be defined by*

$$\begin{aligned} R_1 &\le r_1 - \alpha a_1 - (1-\alpha) b_1, \\ R_2 &\le r_2 - \alpha a_2 - (1-\alpha) b_2, \\ R_1 + R_2 &\le r_{12} - c, \\ R_0 + R_1 + R_2 &\le r_{012} - d \end{aligned}$$

*and assume that $\mathscr{K}_\alpha \neq \varnothing$ for every $\alpha$. Then*

$$\bigcup_{\alpha_0 \le \alpha \le \alpha_1} \mathscr{K}_\alpha = \mathscr{K}, \tag{6.11}$$

*where $\mathscr{K}$ is defined by*

$$R_1 \le r_1 - \alpha_0 a_1 - (1-\alpha_0) b_1, \tag{6.12}$$
$$R_2 \le r_2 - \alpha_1 a_2 - (1-\alpha_1) b_2, \tag{6.13}$$
$$R_1 + R_2 \le r_{12} - c, \tag{6.14}$$
$$R_0 + R_1 + R_2 \le r_{012} - d. \tag{6.15}$$

*Proof of Lemma 6.12.* The proof is a direct application of Lemma 6.13 by setting

$$\begin{aligned} r_1 &= I(T \wedge V_1 | V_2 U), & r_2 &= I(T \wedge V_2 | V_1 U), \\ r_{12} &= I(T \wedge V_1 V_2 | U), & r_{012} &= I(T \wedge V_1 V_2), \\ a_1 &= I(Z \wedge V_1 | V_2 U), & a_2 &= I(Z \wedge V_2 | U), \\ b_1 &= I(Z \wedge V_1 | U), & b_2 &= I(Z \wedge V_2 | V_1 U), \\ \alpha_0 &= \alpha_0^{(1)}, & \alpha_1 &= \alpha_1^{(1)}. \end{aligned}$$

We just need to show that the bounds (6.12) and (6.13) coincide with those from the definition of $\mathscr{R}^{(1)}(p)$. This is easy for the case $\alpha_0^{(1)} = 0$ because in that case we have $I(T \wedge V_2 | V_1 U) \ge I(Z \wedge V_2 | V_1 U)$ and the positive part in the bound on $R_1$ in the definition of $\mathscr{R}^{(1)}(p)$ vanishes. Similarly $\alpha_1^{(1)} = 1$ implies $I(T \wedge V_1 | V_2 U) \ge I(Z \wedge V_1 | V_2 U)$ and the positive part in the bound on $R_2$ in the definition of $\mathscr{R}^{(1)}(p)$ vanishes. Now assume that $\alpha_0^{(1)} > 0$. This assumption implies $I(Z \wedge V_2 | V_1 U) > I(T \wedge V_2 | V_1 U)$. Thus we obtain for the equivalent of (6.12)

$$\begin{aligned} & I(T \wedge V_1 | V_2 U) - I(Z \wedge V_1 | U) \\ & \quad - \frac{I(T \wedge V_2 | V_1 U) - I(Z \wedge V_2 | V_1 U)}{I(Z \wedge V_2 | U) - I(Z \wedge V_2 | V_1 U)} (I(Z \wedge V_1 | V_2 U) - I(Z \wedge V_1 | U)) \\ &= I(T \wedge V_1 | V_2 U) - I(Z \wedge V_1 | U) \\ & \quad - \frac{I(T \wedge V_2 | V_1 U) - I(Z \wedge V_2 | V_1 U)}{I(Z \wedge V_2 | U) - I(Z \wedge V_2 | V_1 U)} (I(Z \wedge V_2 | V_1 U) - I(Z \wedge V_2 | U)) \\ &= I(T \wedge V_1 | V_2 U) + I(T \wedge V_2 | V_1 U) - I(Z \wedge V_1 V_2 | U) \\ &= I(T \wedge V_1 | V_2 U) - I(Z \wedge V_1 | U) - [I(Z \wedge V_2 | V_1 U) - I(T \wedge V_2 | V_1 U)]_+. \end{aligned}$$

If $\alpha_1^{(1)} < 1$, we obtain the analog for the bound on $R_2$. This shows with Lemma 6.13 that $\mathscr{R}^{(1)}(p)$ can be represented as the union of the sets $\mathscr{R}_\alpha^{(1)}(p)$ for $\alpha_0^{(1)} \le \alpha \le \alpha_1^{(1)}$. $\square$

**For Case 2:**

Let $p \in \Psi_{H_C}^{(2)}(W)$ for some $H_C > 0$. Here we assume that $I(Z \wedge V_1|V_2U) \neq I(Z \wedge V_2|V_1U)$ which is equivalent to $I(Z \wedge V_1U) \neq I(Z \wedge V_2U)$. In the case of equality, the WCM-achievability of $\mathscr{R}^{(2)}(p)$ can be shown directly. Define for $\alpha \in [\alpha_0^{(2)}, \alpha_1^{(2)}]$ the rate set $\mathscr{R}_\alpha^{(2)}(p)$ by the conditions

$$\begin{aligned} R_1 &\le I(T \wedge V_1|V_2U) - \alpha I(Z \wedge V_1|V_2U), \\ R_2 &\le I(T \wedge V_2|V_1U) - (1-\alpha) I(Z \wedge V_2|V_1U), \\ R_1 + R_2 &\le I(T \wedge V_1V_2|U) - \alpha I(Z \wedge V_1|V_2U) - (1-\alpha) I(Z \wedge V_2|V_1U), \\ R_0 + R_1 + R_2 &\le I(T \wedge V_1V_2) - I(Z \wedge V_1V_2). \end{aligned}$$

**Lemma 6.14.** *We have that*

$$\mathscr{R}^{(2)}(p) = \bigcup_{\alpha_0^{(2)} \le \alpha \le \alpha_1^{(2)}} \mathscr{R}_\alpha^{(2)}(p).$$

*In particular, if $\mathscr{R}_\alpha^{(2)}(p)$ is WCM-achievable for every $\alpha \in [\alpha_0^{(2)}, \alpha_1^{(2)}]$, then so is $\mathscr{R}^{(2)}(p)$.*

*Remark* 6.11. The similarity between the rate regions for Case 1 and Case 2 becomes clear in these decompositions. The description for Case 2 is more complex because $\alpha_0^{(2)}$ and $\alpha_1^{(2)}$ are defined through three minima/maxima. This is due to the fact that the sum $\alpha I(Z \wedge V_1|V_2U) + (1-\alpha) I(Z \wedge V_2|V_1U)$ is not constant in $\alpha$. Hence the conditions for $\alpha_0^{(2)} \le \alpha_1^{(2)}$ cannot be reformulated into simple conditions for the corresponding $p$.

One obtains Lemma 6.14 from the next lemma by making the following replacements:

$$\begin{aligned} r_1 &= I(T \wedge V_1|V_2U), & r_2 &= I(T \wedge V_2|V_1U), \\ r_{12} &= I(T \wedge V_1V_2|U), & r_{012} &= I(T \wedge V_1V_2), \\ a &= I(Z \wedge V_1|V_2U), & b &= I(Z \wedge V_2|V_1U), \\ c &= I(Z \wedge V_1V_2), & & \\ \alpha_0 &= \alpha_0^{(2)}, & \alpha_1 &= \alpha_1^{(2)}. \end{aligned}$$

**Lemma 6.15.** *Let $r_1, r_2, r_{12}, r_{012}, a, b, c$ be nonnegative reals with $\max(r_1, r_2) \le r_{12} \le r_1 + r_2$. Let $\alpha_0, \alpha_1 \in [0,1]$ be given such that for every $\alpha \in [\alpha_0, \alpha_1]$ the set $\mathscr{K}_\alpha$ defined by*

$$\begin{aligned} R_1 &\le r_1 - \alpha a, \\ R_2 &\le r_2 - (1-\alpha) b, \\ R_1 + R_2 &\le r_{12} - \alpha a - (1-\alpha) b, \\ R_0 + R_1 + R_2 &\le r_{012} - c \end{aligned}$$

*is nonempty. If $a \leq b$, the convex hull of the union of these sets is given by the set $\mathscr{K}$ which is characterized by*

$$0 \leq R_1 \leq r_1 - \alpha_0 a, \tag{6.16}$$
$$0 \leq R_2 \leq r_2 - (1-\alpha_1) b, \tag{6.17}$$
$$R_1 + R_2 \leq r_{12} - \alpha_1 a - (1-\alpha_1) b, \tag{6.18}$$
$$bR_1 + aR_2 \leq r_{12} a + r_1 (b-a) - ab, \tag{6.19}$$
$$R_0 + R_1 + R_2 \leq r_{012} - c. \tag{6.20}$$

*If $a > b$, the convex hull of the union of the sets $\mathscr{K}_\alpha$ is given by analogous bounds where $a$ and $b$ are exchanged in* (6.19)*.*

The proof of Lemma 6.15 can be found in Appendix B.

### 6.4.2. How to Prove Secrecy

Proving secrecy using Chernoff-type concentration inequalities (see Subsection 6.4.3) is the core of Devetak's approach to the wiretap channel [23]. Due to the multi-user structure of the inputs of the wiretap MAC, we need several such Chernoff-type inequalities basing on each other. (Devetak only needs one, but as he treats quantum channels, we has to apply the Ahlswede-Winter lemma.) However, once the bounds are established, the way of obtaining secrecy is exactly the same as presented by Devetak. With the help of the inequalities one obtains an $n$-code$_{\mathrm{WCM}}$ and a measure $\vartheta$ (not necessarily a probability measure!) such that for all message triples $(k_0, k_1, k_2)$

$$\| P_{Z^n | M_0 = k_0, M_1 = k_1, M_2 = k_2} - \vartheta \| \leq 2^{-n\beta}. \tag{6.21}$$

Given this, we now derive an upper bound on $I(Z^n \wedge M_0 M_1 M_2)$, where the random triple $(M_0, M_1, M_2)$ is uniformly distributed on the possible input message triples and $Z^n$ represents the output received by Eve. Observe that

$$\begin{aligned} & I(Z^n \wedge M_0 M_1 M_2) \\ &= \frac{1}{K_0 K_1 K_2} \sum_{k_0, k_1, k_2} (H(Z^n) - H(Z^n | M_0 = k_0, M_1 = k_1, M_2 = k_2)). \end{aligned} \tag{6.22}$$

Due to (6.21),

$$\begin{aligned} & \| P_{Z^n} - P_{Z^n | M_0 = k_0, M_1 = k_1, M_2 = k_2} \| \\ &\leq \| P_{Z^n} - \vartheta \| + \| \vartheta - P_{Z^n | M_0 = k_0, M_1 = k_1, M_2 = k_2} \| \\ &\leq \frac{1}{K_0 K_1 K_2} \sum_{\tilde{k}_0, \tilde{k}_1, \tilde{k}_2} \| P_{Z^n | M_0 = \tilde{k}_0, M_1 = \tilde{k}_1, M_2 = \tilde{k}_2} - \vartheta \| + 2^{-n\beta} \\ &\leq 2^{-n\beta/2}. \end{aligned}$$

Lemma 2.22 now implies that (6.22) is upper-bounded by $n(|\mathscr{Z}| + \beta/2) 2^{-n\beta/2}$, which converges to 0 at exponential speed. This also means by Remark 6.1 that the average error of the wiretapper approaches zero at exponential speed.

### 6.4.3. Probabilistic Bounds for Secrecy

In this subsection we define the random variables from which we will build a stochastic wiretap code in Subsection 6.4.5. For this family of random variables we prove several Chernoff-type estimates which will serve to find a code satisfying (6.21). For Case 3, two such estimates are sufficient, Case 0 and 2 require three each and Case 1 requires four. Within each case, the first estimate deals with the joint typicality of the inputs at $\text{Alice}_1$ and $\text{Alice}_2$. The other estimates base on each other. This is due to the complex structure of our family of random variables. Still, all the cases are nothing but a classical multi-user generalization of Devetak's approach taken in [23]. For each case, we first show the probabilistic bounds in one paragraph and then in another paragraph how to obtain (6.21) from those bounds.

Let $p = P_U \otimes P_{X|U} \otimes P_{Y|U} \otimes W \in \Pi(W) \subset \Psi(W)$ be the distribution of a random vector $(U, X, Y, T, Z)$. The auxiliary random variables $V_1$ and $V_2$ will be introduced later in the usual way of prefixing a channel as a means of additional randomization. Let $\delta > 0$ and define for any $n$

$$\begin{aligned} P_U^n(\mathbf{u}) &:= \frac{P_U^{\otimes n}(\mathbf{u})}{P_U^{\otimes n}(T_{U,\delta}^n)} && (\mathbf{u} \in T_{U,\delta}^n), \\ P_{X|U}^n(\mathbf{x}|\mathbf{u}) &:= \frac{P_{X|U}^{\otimes n}(\mathbf{x}|\mathbf{u})}{P_{X|U}^{\otimes n}(T_{X|U,\delta}^n(\mathbf{u})|\mathbf{u})} && (\mathbf{x} \in T_{X|U,\delta}^n(\mathbf{u}), \mathbf{u} \in T_{U,\delta}^n), \\ P_{Y|U}^n(\mathbf{y}|\mathbf{u}) &:= \frac{P_{Y|U}^{\otimes n}(\mathbf{y}|\mathbf{u})}{P_{Y|U}^{\otimes n}(T_{Y|U,\delta}^n(\mathbf{u})|\mathbf{u})} && (\mathbf{y} \in T_{Y|U,\delta}^n(\mathbf{u}), \mathbf{u} \in T_{U,\delta}^n). \end{aligned}$$

Let $L_0, L_1, L_2$ be positive integers. We define $L_0$ independent families of random variables $(U^{l_0}, \mathscr{F}_{l_0})$ as follows. $U^{l_0}$ is distributed according to $P_U^n$. We let $\mathscr{F}_{l_0} := \{X^{l_0 l_1}, Y^{l_0 l_2} : l_1 \in [L_1], l_2 \in [L_2]\}$ be a set of random variables which are independent given $U^{l_0}$ and which satisfy $X^{l_0 l_1} \sim P_{X|U}^n(\,\cdot\,|U^{l_0})$ and $Y^{l_0 l_2} \sim P_{Y|U}^n(\,\cdot\,|U^{l_0})$. Finally we define

$$\mathscr{F} := \bigcup_{l_0 \in [L_0]} (U^{l_0}, \mathscr{F}_{l_0}). \tag{6.23}$$

Thus $\mathscr{F}$ is similar to a generalized half lattice as used in Chapter 3. The notation used here is more convenient for the proof of Theorem 6.10.

Throughout the section, let a small $\varepsilon > 0$ be fixed. The core of the proofs of all the lemmas of this subsection is the following Chernoff bound, see e.g. [8] for a slightly less general version requiring i.i.d. random variables.

**Lemma 6.16.** *Let $b > 0$ and $0 < \varepsilon < 1/2$. For an independent sequence of random variables $Z_1, \dots, Z_L$ with values in $[0, b]$ with $\mu_l := \mathbb{E}[Z_l]$ and with $\mu := \frac{1}{L}\sum_l \mu_l$ one has*

$$\mathbb{P}\left[\frac{1}{L}\sum_{l=1}^{L} Z_l > (1+\varepsilon)\mu\right] \leq \exp\left(-L \cdot \frac{\varepsilon^2 \mu}{2b \ln 2}\right)$$

*and*

$$\mathbb{P}\left[\frac{1}{L}\sum_{l=1}^{L} Z_l < (1-\varepsilon)\mu\right] \leq \exp\left(-L\cdot\frac{\varepsilon^2\mu}{2b\ln 2}\right).$$

We set

$$c := \tilde{c}(|\mathscr{U}||\mathscr{X}||\mathscr{Y}||\mathscr{Z}|),$$

where $\tilde{c}$ is from (2.20). This is the minimal $\tilde{c}$ we will need in the following.

**Bounds for Case 0 and 1:**

Let $L_0, L_1, L_2$ be arbitrary. Due to their conditional independence, the $X^{l_0l_1}$ and $Y^{l_0l_2}$ cannot be required to be jointly conditionally typical given $U^{l_0}$. However, the next lemma shows that most of them are jointly conditionally typical with high probability. Obviously, it is not needed for a single sender.

**Lemma 6.17.** *For $(l_0,l_2)\in[L_0]\times[L_2]$, let the event $A_*^{(1)}(l_0,l_2)$ be defined by*

$$A_*^{(1)}(l_0,l_2) := \left\{\left|\{l_1\in[L_1] : X^{l_0l_1}\in T^n_{X|YU,\delta}(Y^{l_0l_2},U^{l_0})\}\right| \geq (1-\varepsilon)(1-2\cdot 2^{-nc\delta^2})L_1\right\}.$$

*Then*

$$\mathbb{P}[A_*^{(1)}(l_0,l_2)^c] \leq \exp\left(-L_1\cdot\frac{\varepsilon^2(1-2\cdot 2^{-nc\delta^2})}{2\ln 2}\right).$$

*Proof.* Let $\mathbf{u}\in T^n_{U,\delta}$ and $\mathbf{y}\in T^n_{Y|U,\delta}(\mathbf{u})$. We first condition on the event $\{Y^{l_0l_2}=\mathbf{y}, U^{l_0}=\mathbf{u}\}$. Due to (2.20), we have

$$\begin{aligned}
&\mathbb{P}[X^{11}\notin T^n_{X|YU,\delta}(\mathbf{y},\mathbf{u})|Y^{11}=\mathbf{y},U^1=\mathbf{u}]\\
&= \frac{1}{P^{\otimes n}_{X|U}(T^n_{X|U,\delta}(\mathbf{u})|\mathbf{u})}\sum_{\mathbf{x}\in T^n_{X|U,\delta}(\mathbf{u})\setminus T^n_{X|YU,\delta}(\mathbf{y},\mathbf{u})} P^{\otimes n}_{X|U}(\mathbf{x}|\mathbf{u})\\
&\leq \frac{1}{P^{\otimes n}_{X|U}(T^n_{X|U,\delta}(\mathbf{u})|\mathbf{u})}\sum_{\mathbf{x}\notin T^n_{X|YU,\delta}(\mathbf{y},\mathbf{u})} P^{\otimes n}_{X|YU}(\mathbf{x}|\mathbf{y},\mathbf{u})\\
&\leq \frac{2^{-nc\delta^2}}{1-2^{-nc\delta^2}}.
\end{aligned}$$

In particular,

$$\mu := \mathbb{P}[X^{11}\in T^n_{X|YU,\delta}(\mathbf{y},\mathbf{u})|Y^{11}=\mathbf{y},U^1=\mathbf{u}] \geq 1-2\cdot 2^{-nc\delta^2}.$$

Therefore

$$\begin{aligned}
&\mathbb{P}[A_*^{(1)}(l_0,l_2)^c|Y^{l_0l_2}=\mathbf{y},U^{l_0}=\mathbf{u}]\\
&\qquad\leq \mathbb{P}\left[\sum_{l_1} 1_{T^n_{X|YU,\delta}(\mathbf{y},\mathbf{u})}(X^{l_0l_1}) \leq (1-\varepsilon)\mu\, L_1 \,\middle|\, Y^{l_0l_2}=\mathbf{y},U^{l_0}=\mathbf{u}\right],
\end{aligned}$$

which by Lemma 6.16 can be bounded by

$$\exp\left(-L_1 \cdot \frac{\varepsilon^2 \mu}{2\ln 2}\right) \leq \exp\left(-L_1 \cdot \frac{\varepsilon^2(1-2\cdot 2^{-nc\delta^2})}{2\ln 2}\right).$$

This completes the proof as this bound is independent of $(\mathbf{y}, \mathbf{u})$. □

As we cannot guarantee the joint conditional typicality of both senders' inputs, we need to introduce an explicit bound on the channel transition probabilities. This is done in the set $E_1^{(1)}$. Then we prove three lemmas each of which exploits one of the three types of independence contained in $\mathscr{F}$. Altogether these lemmas provide lower bounds on $L_0, L_1, L_2$ which if satisfied allow the construction of a wiretap code satisfying (6.21). Let

$$E_1^{(1)}(\mathbf{u},\mathbf{x},\mathbf{y}) := \{\mathbf{z} \in T^n_{Z|YU,2|\mathscr{X}|\delta}(\mathbf{y},\mathbf{u}) : W_e^{\otimes n}(\mathbf{z}|\mathbf{x},\mathbf{y}) \leq 2^{-n(H(Z|XY)-f_2(\delta))}\},$$

where $f_2(\delta) = \tau(P_{UXYZ}, 3\delta, \delta)$ (see (2.21)). Let

$$\vartheta^{(1)}_{\mathbf{uy}}(\mathbf{z}) := \mathbb{E}[W_e^{\otimes n}(\mathbf{z}|X^{11},\mathbf{y}) 1_{E_1^{(1)}(\mathbf{u},X^{11},\mathbf{y})}(\mathbf{z}) | U^1 = \mathbf{u}]$$

and for

$$F_1^{(1)}(\mathbf{u},\mathbf{y}) := \{\mathbf{z} \in T^n_{Z|YU,2|\mathscr{X}|\delta}(\mathbf{y},\mathbf{u}) : \vartheta^{(1)}_{\mathbf{uy}}(\mathbf{z}) \geq \varepsilon |T^n_{Z|YU,2|\mathscr{X}|\delta}(\mathbf{y},\mathbf{u})|^{-1}\}$$

define

$$\hat{\vartheta}^{(1)}_{\mathbf{uy}} := \vartheta^{(1)}_{\mathbf{uy}} \cdot 1_{F_1^{(1)}(\mathbf{u},\mathbf{y})}, \quad E_2^{(1)}(\mathbf{u},\mathbf{x},\mathbf{y}) := E_1^{(1)}(\mathbf{u},\mathbf{x},\mathbf{y}) \cap F_1^{(1)}(\mathbf{u},\mathbf{y}).$$

**Lemma 6.18.** *For every $\mathbf{z} \in \mathscr{Z}^n$ and $(l_0, l_2) \in [L_0] \times [L_2]$, let $A_1^{(1)}(l_0, l_2, \mathbf{z})$ be the event that*

$$\frac{1}{L_1}\sum_{l_1} W_e^{\otimes n}(\mathbf{z}|X^{l_0 l_1}, Y^{l_0 l_2}) 1_{E_2^{(1)}(U^{l_0}, X^{l_0 l_1}, Y^{l_0 l_2})}(\mathbf{z}) \in [(1 \pm \varepsilon)\hat{\vartheta}^{(1)}_{U^{l_0} Y^{l_0 l_2}}(\mathbf{z})].$$

*Then*

$$\mathbb{P}[A_1^{(1)}(l_0, l_2, \mathbf{z})^c] \leq 2\exp\left(-L_1 \cdot \frac{\varepsilon^3 2^{-n(I(Z\wedge X|YU)+f_1(\delta)+f_2(\delta))}}{2\ln 2}\right)$$

*for $f_1(\delta) = \tau(P_{UYZ}, 2\delta, 2|\mathscr{X}|\delta)$ and $n$ sufficiently large.*

*Proof.* For $\mathbf{u} \in T^n_{U,\delta}$ and $\mathbf{y} \in T^n_{Y|U,\delta}(\mathbf{u})$ we condition on the event $\{Y^{l_0 l_2} = \mathbf{y}, U^{l_0} = \mathbf{u}\}$. The conditional expectation of the bounded conditionally i.i.d. random variables

$$W_e^{\otimes n}(\mathbf{z}|X^{l_0 l_1}, \mathbf{y}) 1_{E_2^{(1)}(\mathbf{u}, X^{l_0 l_1}, \mathbf{y})}(\mathbf{z}) \leq 2^{-n(H(Z|XY)-f_2(\delta))} \qquad (l_1 \in [L_1])$$

is $\hat{\vartheta}^{(1)}_{\mathbf{uy}}(\mathbf{z})$. We use Lemma 6.16, the definition of $F_1^{(1)}(\mathbf{u},\mathbf{y})$, and (2.23) to obtain for $n$ sufficiently large

$$\begin{aligned}
&\mathbb{P}[A_1^{(1)}(l_0,l_2,\mathbf{z})^c|Y^{l_0l_2}=\mathbf{y},U^{l_0}=\mathbf{u}]\\
&\leq 2\exp\left(-L_1\cdot\frac{\varepsilon^2\hat{\vartheta}^{(1)}_{\mathbf{uy}}(\mathbf{z})2^{n(H(Z|XY)-f_2(\delta))}}{2\ln 2}\right)\\
&\leq 2\exp\left(-L_1\cdot\frac{\varepsilon^3 2^{-n(I(Z\wedge X|YU)+f_1(\delta)+f_2(\delta))}}{2\ln 2}\right).
\end{aligned}$$

This bound is uniform in $\mathbf{u}$ and $\mathbf{y}$, so the proof is complete. □

For the next lemma, define

$$\vartheta^{(1)}_{\mathbf{u}}(\mathbf{z}) := \mathbb{E}[W_e^{\otimes n}(\mathbf{z}|X^{11},Y^{11})1_{E_2^{(1)}(\mathbf{u},X^{11},Y^{11})}(\mathbf{z})|U^1=\mathbf{u}].$$

Further let

$$F_2^{(1)}(\mathbf{u}) := \{\mathbf{z}\in T^n_{Z|U,3|\mathscr{Y}||\mathscr{X}|\delta}(\mathbf{u}) : \vartheta^{(1)}_{\mathbf{u}}(\mathbf{z})\geq\varepsilon|T^n_{Z|U,3|\mathscr{Y}||\mathscr{X}|\delta}(\mathbf{u})|^{-1}\}$$

and

$$\hat{\vartheta}^{(1)}_{\mathbf{u}} = \vartheta^{(1)}_{\mathbf{u}}\cdot 1_{F_2^{(1)}(\mathbf{u})},\quad E_0^{(1)}(\mathbf{u},\mathbf{x},\mathbf{y}) := E_2^{(1)}(\mathbf{u},\mathbf{x},\mathbf{y})\cap F_2^{(1)}(\mathbf{u}).$$

**Lemma 6.19.** *For every $\mathbf{z}\in\mathscr{Z}^n$ and $l_0\in[L_0]$, let $A_2^{(1)}(l_0,\mathbf{z})$ be the event*

$$\frac{1}{L_1L_2}\sum_{l_1l_2}W_e^{\otimes n}(\mathbf{z}|X^{l_0l_1},Y^{l_0l_2})1_{E_0^{(1)}(U^{l_0},X^{l_0l_1},Y^{l_0l_2})}(\mathbf{z})\in[(1\pm 3\varepsilon)\hat{\vartheta}^{(1)}_{U^{l_0}}(\mathbf{z})].$$

*Then for $\varepsilon$ sufficiently small and $n$ sufficiently large,*

$$\begin{aligned}
\mathbb{P}[A_2^{(1)}(l_0,\mathbf{z})^c] \leq{}& 2|\mathscr{Y}|^n\exp\left(-L_1\cdot\frac{\varepsilon^3 2^{-n(I(Z\wedge X|YU)+f_1(\delta)+f_2(\delta))}}{2\ln 2}\right)\\
&+2\exp\left(-L_2\cdot\frac{\varepsilon^3 2^{-n(I(Z\wedge Y|U)+f_1(\delta)+f_4(\delta))}}{4\ln 2}\right),
\end{aligned}$$

*where $f_4(\delta)=\tau(P_{UZ},\delta,3|\mathscr{Y}||\mathscr{X}|\delta)$.*

*Proof.* We have

$$\mathbb{P}[A_2^{(1)}(l_0,\mathbf{z})^c] = \sum_{\mathbf{u}\in T^n_{U,\delta}}\mathbb{P}[U^{l_0}=\mathbf{u}]\mathbb{P}[A_2^{(1)}(l_0,\mathbf{z})^c|U^{l_0}=\mathbf{u}].$$

If $\mathbf{z} \notin F_2^{(1)}(\mathbf{u})$, then $\mathbb{P}[A_2^{(1)}(l_0,\mathbf{z})^c | U^{l_0} = \mathbf{u}] = 0$. Thus let $\mathbf{u} \in T^n_{U,\delta}$ and assume $\mathbf{z} \in F_2^{(1)}(\mathbf{u})$. We define the set $B_{\mathbf{u}} \subset (T^n_{X|U,\delta}(\mathbf{u}))^{L_1}$ as

$$\bigcap_{\mathbf{y}\in T^n_{Y|U,\delta}(\mathbf{u})} \Big\{(\mathbf{x}^1,\dots,\mathbf{x}^{L_1}) \in (T^n_{X|U,\delta}(\mathbf{u}))^{L_1} : \\ \frac{1}{L_1}\sum_{l_1} W_e^{\otimes n}(\mathbf{z}|\mathbf{x}^{l_1},\mathbf{y}) 1_{E_2^{(1)}(\mathbf{u},x^{l_1},\mathbf{y})}(\mathbf{z}) \in [(1\pm\varepsilon)\hat{\vartheta}^{(1)}_{\mathbf{uy}}(\mathbf{z})]\Big\}.$$

One has

$$\begin{aligned}
&\mathbb{P}[A_2^{(1)}(l_0,\mathbf{z})^c | U^{l_0} = \mathbf{u}] \\
&\le \mathbb{P}\big[(X^{l_0 1},\dots,X^{l_0 L_1}) \notin B_{\mathbf{u}} | U^{l_0} = \mathbf{u}\big] \\
&\quad + \sum_{(\mathbf{x}^1,\dots,\mathbf{x}^{L_1})\in B_{\mathbf{u}}} \mathbb{P}\big[A_2^{(1)}(l_0,\mathbf{z})^c | X^{l_0 1} = \mathbf{x}^1,\dots,X^{l_0 L_1} = \mathbf{x}^{L_1}, U^{l_0} = \mathbf{u}\big]\cdot \\
&\qquad\qquad \cdot \mathbb{P}[X^{l_0 1} = \mathbf{x}^1,\dots,X^{l_0 L_1} = \mathbf{x}^{L_1} | U^{l_0} = \mathbf{u}\big].
\end{aligned}$$

From the proof of Lemma 6.18 it follows that

$$\mathbb{P}\big[(X^{l_0 1},\dots,X^{l_0 L_1}) \notin B_{\mathbf{u}} | U^{l_0} = \mathbf{u}\big] \\ \le 2|\mathscr{Y}|^n \exp\left(-L_1 \cdot \frac{\varepsilon^3 2^{-n(I(Z\wedge X|YU)+f_1(\delta)+f_2(\delta))}}{2\ln 2}\right), \quad (6.24)$$

which gives a bound independent of $\mathbf{u}$. Now let $(\mathbf{x}^1,\dots,\mathbf{x}^{L_1}) \in B_{\mathbf{u}}$. By (2.20) and (2.21),

$$\begin{aligned}
\hat{\vartheta}^{(1)}_{\mathbf{uy}}(\mathbf{z}) &= \mathbb{E}[W_e^{\otimes n}(\mathbf{z}|X^{11},\mathbf{y}) 1_{E_2^{(1)}(\mathbf{u},X^{11},\mathbf{y})}(\mathbf{z}) | U^1 = \mathbf{u}] \\
&\le \mathbb{E}[W_e^{\otimes n}(\mathbf{z}|X^{11},\mathbf{y}) | U^1 = \mathbf{u}] \\
&\le \frac{1}{P^{\otimes n}_{X|U}(T^n_{X|U,\delta}(\mathbf{u})|\mathbf{u})} (P_{Z|YU})^{\otimes n}(\mathbf{z}|\mathbf{y},\mathbf{u}) \\
&\le (1-2^{-nc\delta^2})^{-1} 2^{-n(H(Z|YU)-f_1(\delta))}.
\end{aligned}$$

Hence the random variables

$$\tilde{W}^{(1)}_{\mathbf{uz}}(l_0,l_2) := \frac{1}{L_1}\sum_{l_1} W_e^{\otimes n}(\mathbf{z}|\mathbf{x}^{l_1},Y^{l_0 l_2}) 1_{E_2^{(1)}(\mathbf{u},\mathbf{x}^{l_1},Y^{l_0 l_2})}(\mathbf{z}) \qquad (l_2 \in [L_2]),$$

which are independent conditional on $\{U^{l_0} = \mathbf{u}\}$, are upper-bounded by

$$\frac{(1+\varepsilon)}{(1-2^{-nc\delta^2})} \cdot 2^{-n(H(Z|YU)-f_1(\delta))}.$$

For their conditional expectation we have

$$\mu_{l_0l_2} := \mathbb{E}[\tilde{W}^{(1)}_{\mathbf{uz}}(l_0,l_2)|U^{l_0}=\mathbf{u}] \in [(1\pm\varepsilon)\mathbb{E}[\hat{\vartheta}^{(1)}_{\mathbf{u}Y^{l_0l_2}}(\mathbf{z})|U^1=\mathbf{u}]] = [(1\pm\varepsilon)\hat{\vartheta}^{(1)}_{\mathbf{u}}(\mathbf{z})].$$

Thus their arithmetic mean $\bar{\mu} = (1/L_2)\sum_{l_2}\mu_{l_0l_2}$ must also be contained in $[(1\pm\varepsilon)\hat{\vartheta}^{(1)}_{\mathbf{u}}(\mathbf{z})]$. Applying Lemma 6.16, we conclude

$$\begin{aligned}
&\mathbb{P}\big[A_2^{(1)}(l_0,\mathbf{z})^c|X^{l_01}=\mathbf{x}^1,\dots,X^{l_0L_1}=\mathbf{x}^{L_1},U^{l_0}=\mathbf{u}\big]\\
&=\mathbb{P}\left[\frac{1}{L_2}\sum_{l_2}\tilde{W}^{(1)}_{\mathbf{uz}}(l_0,l_2)\notin[(1\pm3\varepsilon)\hat{\vartheta}^{(1)}_{\mathbf{u}}(\mathbf{z})]\middle|U^{l_0}=\mathbf{u}\right]\\
&\le\mathbb{P}\left[\frac{1}{L_2}\sum_{l_2}\tilde{W}^{(1)}_{\mathbf{uz}}(l_0,l_2)\notin[(1\pm\varepsilon)\bar{\mu}]\middle|U^{l_0}=\mathbf{u}\right]\\
&\le 2\exp\left(-L_2\cdot\frac{\varepsilon^2(1-2^{-nc\delta^2})2^{n(H(Z|YU)-f_1(\delta))}(1-\varepsilon)\hat{\vartheta}^{(1)}_{\mathbf{u}}(\mathbf{z})}{2(1+\varepsilon)\ln 2}\right).
\end{aligned}$$

Due to the definition of $F_2^{(1)}(\mathbf{u})$ and to (2.23), this is smaller than

$$2\exp\left(-L_2\cdot\frac{\varepsilon^3 2^{-n(I(Z\wedge Y|U)+f_1(\delta)+f_4(\delta))}}{4\ln 2}\right) \tag{6.25}$$

if $\varepsilon$ is sufficiently small and $n$ is sufficiently large, giving a bound independent of $\mathbf{u}$ and $\mathbf{x}^1,\dots,\mathbf{x}^{L_1}$. Adding the bounds (6.24) and (6.25) concludes the proof. □

The next lemma is only needed in Case 1. Let $A_2^{(1)}(\mathbf{z}) := A_2^{(1)}(1,\mathbf{z})\cap\ldots\cap A_2^{(1)}(L_0,\mathbf{z})$. For every $\mathbf{z}$, we then define a new probability measure by $\hat{\mathbb{P}}^{(1)}_{\mathbf{z}} := \mathbb{P}[\cdot|A_2^{(1)}(\mathbf{z})]$. With $\vartheta^{(1)}(\mathbf{z}) := \hat{\mathbb{E}}^{(1)}_{\mathbf{z}}[\hat{\vartheta}^{(1)}_{U^1}(\mathbf{z})]$ define

$$F_0^{(1)} := \{\mathbf{z}\in T^n_{Z,4|\mathscr{Y}||\mathscr{X}||\mathscr{U}|\delta} : \vartheta^{(1)}(\mathbf{z}) \ge |T^n_{Z,4|\mathscr{Y}||\mathscr{X}||\mathscr{U}|\delta}|^{-1}\}$$

and $\hat{\vartheta}^{(1)} := \vartheta^{(1)}\cdot 1_{F_0^{(1)}}$.

**Lemma 6.20.** *Let $\mathbf{z}\in F_0^{(1)}$ and let $A_0^{(1)}(\mathbf{z})$ be the event that*

$$\frac{1}{L_0L_1L_2}\sum_{l_0,l_1,l_2} W_e^{\otimes n}(\mathbf{z}|X^{l_0l_1},Y^{l_0l_2})1_{E_0^{(1)}(U^{l_0},X^{l_0l_1},Y^{l_0l_2})}(\mathbf{z}) \in [(1\pm5\varepsilon)\hat{\vartheta}^{(1)}(\mathbf{z})].$$

*Then for $f_6(\delta) = \tau(P_Z, 4|\mathscr{Y}||\mathscr{X}||\mathscr{U}|\delta, \delta)$, sufficiently small $\varepsilon$ and $n$ sufficiently large,*

$$\begin{aligned}
&\mathbb{P}[A_0^{(1)}(\mathbf{z})^c] \\
&\leq 2L_0|\mathscr{Y}|^n \exp\left(-L_1 \cdot \frac{\varepsilon^3 2^{-n(I(Z\wedge X|YU)+f_1(\delta)+f_2(\delta))}}{2\ln 2}\right) \\
&\quad +2L_0 \exp\left(-L_2 \cdot \frac{\varepsilon^3 2^{-n(I(Z\wedge Y|U)+f_1(\delta)+f_4(\delta))}}{4\ln 2}\right) \\
&\quad +2\exp\left(-L_0 \cdot \frac{\varepsilon^3 2^{-n(I(Z\wedge U)+f_4(\delta)+f_6(\delta))}}{4\ln 2}\right).
\end{aligned}$$

*Proof.* We have

$$\mathbb{P}[A_0^{(1)}(\mathbf{z})^c] \leq \hat{\mathbb{P}}_{\mathbf{z}}^{(1)}[A_0^{(1)}(\mathbf{z})^c] + \mathbb{P}[A_2^{(1)}(\mathbf{z})^c]. \tag{6.26}$$

By Lemma 6.19, for $\varepsilon$ sufficiently small and $n$ sufficiently large,

$$\begin{aligned}
\mathbb{P}[A_2^{(1)}(\mathbf{z})^c] \leq\ & 2L_0|\mathscr{Y}|^n \exp\left(-L_1 \cdot \frac{\varepsilon^3 2^{-n(I(Z\wedge X|YU)+f_1(\delta)+f_2(\delta))}}{2\ln 2}\right) \\
& +2L_0 \exp\left(-L_2 \cdot \frac{\varepsilon^3 2^{-n(I(Z\wedge Y|U)+f_1(\delta)+f_4(\delta))}}{4\ln 2}\right).
\end{aligned} \tag{6.27}$$

In order to bound $\hat{\mathbb{P}}_{\mathbf{z}}^{(1)}[A_0^{(1)}(\mathbf{z})^c]$, note that the sets $A_2^{(1)}(1, \mathbf{z}), \ldots, A_2^{(1)}(L_0, \mathbf{z})$ are independent with respect to $\mathbb{P}$. Thus under $\hat{\mathbb{P}}_{\mathbf{z}}^{(1)}$, the random variables

$$\tilde{W}_{\mathbf{z}}^{(1)}(l_0) := \frac{1}{L_1 L_2} \sum_{l_1, l_2} W_e^{\otimes n}(\mathbf{z}|X^{l_0 l_1}, Y^{l_0 l_2}) 1_{E_0^{(1)}(U^{l_0}, X^{l_0 l_1}, Y^{l_0 l_2})}(\mathbf{z}) \qquad (l_0 \in [L_0])$$

retain their independence and are upper-bounded by

$$(1+3\varepsilon) \max_{\mathbf{u} \in T_{U,\delta}^n} \hat{\vartheta}_{\mathbf{u}}^{(1)}(\mathbf{z}).$$

We can further bound this last term as follows: for $\mathbf{u} \in T_{U,\delta}^n$, applying (2.20) and (2.21),

$$\begin{aligned}
\hat{\vartheta}_{\mathbf{u}}^{(1)}(\mathbf{z}) &= \mathbb{E}[W_e^{\otimes n}(\mathbf{z}|X^{11}, Y^{11}) 1_{E_0^{(1)}(\mathbf{u}, X^{11}, Y^{11})}(\mathbf{z}) | U^1 = \mathbf{u}] \\
&\leq \mathbb{E}[W_e^{\otimes n}(\mathbf{z}|X^{11}, Y^{11}) | U^1 = \mathbf{u}] \\
&\leq \frac{1}{P_1^{\otimes n}(T_{X|U,\delta}^n(\mathbf{u})|\mathbf{u}) P_2^{\otimes n}(T_{Y|U,\delta}^n(\mathbf{u})|\mathbf{u})} P_{Z|U}^{\otimes n}(\mathbf{z}|\mathbf{u}) \\
&\leq (1 - 2^{-nc_1\delta^2})^{-2} 2^{-n(H(Z|U)-f_4(\delta))}.
\end{aligned}$$

Observing that $\hat{\mathbb{E}}_{\mathbf{z}}^{(1)}[\tilde{W}_{\mathbf{z}}^{(1)}(1)] \in [(1 \pm 3\varepsilon)\hat{\vartheta}^{(1)}(\mathbf{z})]$ and applying Lemma 6.16 and (2.23) in the usual way yields

$$\begin{aligned}\hat{\mathbb{P}}_{\mathbf{z}}^{(1)}[A_0^{(1)}(\mathbf{z})^c] &\leq 2\exp\left(-L_0 \cdot \frac{\varepsilon^2(1-2^{-nc\delta^2})^2\, 2^{n(H(Z|U)-f_4(\delta))}(1-3\varepsilon)\,\hat{\vartheta}^{(1)}(\mathbf{z})}{2(1+3\varepsilon)\ln 2}\right)\\ &\leq 2\exp\left(-L_0 \cdot \frac{\varepsilon^3 2^{-n(I(Z\wedge U)+f_4(\delta)+f_6(\delta))}}{4\ln 2}\right)\end{aligned}$$

if $\varepsilon$ is sufficiently small and $n$ sufficiently large. Inserting this and (6.27) in (6.26) completes the proof. □

We finally note that results analogous to Lemma 6.17-6.20 hold where the roles of $X$ and $Y$ are exchanged. We denote the corresponding events by $A_*^{(1)}(l_0,l_2)'$ and $A_1^{(1)}(l_0,l_2,\mathbf{z})', A_2^{(1)}(l_0,\mathbf{z})', A_0^{(1)}(\mathbf{z})'$.

**Secrecy for Case 0 and 1:**

Lemma 6.21 below links the above probabilistic bounds to secrecy. In Paragraph 6.4.4, roughly speaking, we will associate a family $\mathscr{F}$ to every message triple $(k_0,k_1,k_2)$. If $L_0, L_1, L_2$ are large enough, the conditions of Lemma 6.21 are satisfied for every such $\mathscr{F}$ with very high probability. Hence there is a joint realization of all the $\mathscr{F}$ such that the statement of the lemma is satisfied for every message triple. This implies that this realization determines an $(n, H_C)$-code$_{\mathrm{WCM}}$ satisfying (6.21) for all $(k_0,k_1,k_2)$.

**Lemma 6.21.** *Denote by $p^{(1)}$ the bound on $\mathbb{P}[A_2^{(1)}(l_0,\mathbf{z})^c]$ derived in Lemma 6.19. Let $\{\mathbf{u}^{l_0}, \mathbf{x}^{l_0l_1}, \mathbf{y}^{l_0l_2} : (l_0,l_1,l_2) \in [L_0]\times[L_1]\times[L_2]\}$ be a realization of $\mathscr{F}$ satisfying the conditions of*

$$\bigcap_{l_0,l_2} A_*^{(1)}(l_0,l_2), \tag{6.28}$$

$$\bigcap_{l_0,l_2}\bigcap_{\mathbf{z}\in\mathscr{Z}^n} A_1^{(1)}(l_0,l_2,\mathbf{z}), \tag{6.29}$$

$$\bigcap_{l_0}\bigcap_{\mathbf{z}\in\mathscr{Z}^n} A_2^{(1)}(l_0,\mathbf{z}), \tag{6.30}$$

$$\bigcap_{\mathbf{z}\in F_0^{(1)}} A_0^{(1)}(\mathbf{z}). \tag{6.31}$$

*Then*

$$\|\hat{\vartheta}^{(1)} - \frac{1}{L_0L_1L_2}\sum_{l_0,l_1,l_2} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})\| \leq 20\varepsilon + 9\cdot 2^{-nc\delta^2} + L_0|\mathscr{Z}|^n p^{(1)}.$$

*The same inequality is true if we require conditions* (6.28′)-(6.31′) *which contain the primed equivalents of* (6.28)-(6.31) *defined at the end of the previous paragraph. If $L_0 = 1$, then* (6.31) *and* (6.31′) *do not have to hold.*

We now prove the above lemma. We have

$$\begin{aligned}
&\|\hat{\vartheta}^{(1)} - \frac{1}{L_0L_1L_2}\sum_{l_0,l_1,l_2} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})\| \\
&\le \|\hat{\vartheta}^{(1)} - \frac{1}{L_0L_1L_2}\sum_{l_0,l_1,l_2} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})1_{E_0^{(1)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})}1_{F_0^{(1)}}\| && (6.32)\\
&+ \|\frac{1}{L_0L_1L_2}\sum_{l_0,l_1,l_2} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})1_{E_0^{(1)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})}(1-1_{F_0^{(1)}})\| && (6.33)\\
&+ \|\frac{1}{L_0L_1L_2}\sum_{l_0,l_1,l_2} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})1_{E_2^{(1)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})}(1-1_{F_2^{(1)}(\mathbf{u}^{l_0})})\| && (6.34)\\
&+ \|\frac{1}{L_0L_1L_2}\sum_{l_0,l_1,l_2} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})1_{E_1^{(1)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})}(1-1_{F_1^{(1)}(\mathbf{u}^{l_0},\mathbf{y}^{l_0l_2})})\| && (6.35)\\
&+ \|\frac{1}{L_0L_1L_2}\sum_{l_0,l_1,l_2} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})(1-1_{E_1^{(1)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})})\|. && (6.36)
\end{aligned}$$

Due to (6.31), we know that (6.32) $\le 5\varepsilon$.

Next we consider (6.35). Due to (6.29) we have

$$\begin{aligned}
&(6.35)\\
&\le 1 - \frac{1}{L_0L_1L_2}\sum_{l_0,l_1,l_2} W_e^{\otimes n}(E_2^{(1)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})\\
&\le 1 - \frac{1-\varepsilon}{L_0L_2}\sum_{l_0,l_2} \hat{\vartheta}^{(1)}_{\mathbf{u}^{l_0}\mathbf{y}^{l_0l_2}}(\mathscr{Z}^n).
\end{aligned}$$

The support of $\vartheta^{(1)}_{\mathbf{u}^{l_0}\mathbf{y}^{l_0l_2}}$ is contained in $T^n_{Z|YU,2|\mathscr{X}|\delta}(\mathbf{y}^{l_0l_2},\mathbf{u}^{l_0})$, so by the definition of $F_1^{(1)}(\mathbf{u}^{l_0},\mathbf{y}^{l_0l_2})$ we obtain

$$\hat{\vartheta}^{(1)}_{\mathbf{u}^{l_0}\mathbf{y}^{l_0l_2}}(\mathscr{Z}^n) \ge \vartheta^{(1)}_{\mathbf{u}^{l_0}\mathbf{y}^{l_0l_2}}(\mathscr{Z}^n) - \varepsilon. \tag{6.37}$$

**Lemma 6.22.** *If* $\mathbf{u} \in T^n_{U,\delta}$ *and* $\mathbf{y} \in T^n_{Y|U,\delta}(\mathbf{u})$*, then*

$$\vartheta^{(1)}_{\mathbf{uy}}(\mathscr{Z}^n) \ge 1 - 2\cdot 2^{-nc\delta^2}.$$

*Proof.* Recall the notation $\mathbb{E}[X;A] = \mathbb{E}[X1_A]$ with $X$ a random variable with values in $\mathscr{X}$ and $A \subset \mathscr{X}$. Note that

$$\begin{aligned}
&\vartheta^{(1)}_{\mathbf{uy}}(\mathscr{Z}^n)\\
&= \mathbb{E}[W_e^{\otimes n}(E_1^{(1)}(\mathbf{u},X^{11},\mathbf{y})|X^{11},\mathbf{y})|U^1=\mathbf{u}]\\
&\ge \mathbb{E}[W_e^{\otimes n}(E_1^{(1)}(\mathbf{u},X^{11},\mathbf{y})|X^{11},\mathbf{y}); X^{11}\in T^n_{X|YU,\delta}(\mathbf{y},\mathbf{u})|U^1=\mathbf{u}]. && (6.38)
\end{aligned}$$

By Lemma 2.20, we have for $\mathbf{x} \in T^n_{X|YU,\delta}(\mathbf{y},\mathbf{u})$

$$T^n_{Z|YXU,\delta}(\mathbf{y},\mathbf{x},\mathbf{u}) \subset T^n_{Z|YU,2|\mathscr{X}|\delta}(\mathbf{y},\mathbf{u}). \tag{6.39}$$

Due to the choice of $f_2(\delta)$ and to (2.21), we thus see that $T^n_{Z|YXU,\delta}(\mathbf{y},\mathbf{x},\mathbf{u})$ is contained in $E_1^{(1)}(\mathbf{u},\mathbf{x},\mathbf{y})$ for $\mathbf{x} \in T^n_{X|YU,\delta}(\mathbf{y},\mathbf{u})$, and we have that (6.38) is lower-bounded by

$$\mathbb{E}[W_e^{\otimes n}(T^n_{Z|YXU,\delta}(\mathbf{y},X^{11},\mathbf{u})|X^{11},\mathbf{y}); X^{11} \in T^n_{X|YU,\delta}(\mathbf{y},\mathbf{u})|U^1=\mathbf{u}]. \tag{6.40}$$

Further, as in the proof of Lemma 6.17 one sees that

$$\mathbb{P}[X^{11} \in T^n_{X|YU,\delta}(\mathbf{y},\mathbf{u})|U^1=\mathbf{u}] \geq 1 - \frac{2^{-nc\delta^2}}{1-2^{-nc\delta^2}}. \tag{6.41}$$

Due to (6.41) and (2.20), we can lower-bound (6.40) for sufficiently large $n$ by

$$(1-2^{-nc\delta^2}) \cdot \left(1 - \frac{2^{-nc\delta^2}}{1-2^{-nc\delta^2}}\right) \geq 1 - 2 \cdot 2^{-nc\delta^2},$$

which proves Lemma 6.22. □

Using (6.37) and Lemma 6.22 we can conclude that

$$(6.35) \leq 2(\varepsilon + 2^{-nc\delta^2}).$$

One starts similarly for (6.34). We have by (6.30)

$$\begin{aligned}(6.34) &\leq 1 - \frac{1}{L_0L_1L_2} \sum_{l_0,l_1,l_2} W_e^{\otimes n}(E_0^{(1)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2}) \\ &\leq 1 - \frac{(1-3\varepsilon)}{L_0} \sum_{l_0} \hat{\vartheta}^{(1)}_{\mathbf{u}^{l_0}}(\mathscr{Z}^n).\end{aligned}$$

As the support of $\vartheta^{(1)}_{\mathbf{u}^{l_0}}$ is contained in $T^n_{Z|U,3|\mathscr{Y}||\mathscr{X}|\delta}(\mathbf{u}^{l_0})$, we can lower-bound $\hat{\vartheta}^{(1)}_{\mathbf{u}^{l_0}}(\mathscr{Z}^n)$ by $\vartheta^{(1)}_{\mathbf{u}^{l_0}}(\mathscr{Z}^n) - \varepsilon$. Using (6.37) and Lemma 6.22, we have

$$\vartheta^{(1)}_{\mathbf{u}^{l_0}}(\mathscr{Z}^n) = \mathbb{E}[\hat{\vartheta}^{(1)}_{\mathbf{u}^{l_0}Y^{11}}(\mathscr{Z}^n)|U^1=\mathbf{u}^{l_0}] \geq 1 - 2 \cdot 2^{-nc\delta^2} - \varepsilon, \tag{6.42}$$

so we conclude

$$(6.34) \leq 5\varepsilon + 2 \cdot 2^{-nc\delta^2}.$$

For (6.33), one has by (6.31)

$$\begin{aligned}(6.33) &\leq 1 - \frac{1}{L_0L_1L_2} \sum_{l_0,l_1,l_2} W_e^{\otimes n}(E_0^{(1)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2}) \cap F_0^{(1)}|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2}) \\ &\leq 1 - (1-5\varepsilon)\hat{\vartheta}^{(1)}(F_0^{(1)}).\end{aligned}$$

It remains to lower-bound $\hat{\vartheta}^{(1)}(F_0^{(1)})$. Observe that the support of $\vartheta^{(1)}$ is restricted to $T^n_{Z,4|\mathscr{Y}||\mathscr{X}||\mathscr{U}|\delta}$, so due to the definition of $F_0^{(1)}$, one has $\hat{\vartheta}^{(1)}(F_0^{(1)}) = \vartheta^{(1)}(F_0^{(1)}) \geq \vartheta^{(1)}(\mathscr{Z}^n) - \varepsilon$. Further,

$$\begin{aligned}\vartheta^{(1)}(\mathscr{Z}^n) &= \sum_{\mathbf{z}\in\mathscr{Z}^n} \hat{\mathbb{E}}^{(1)}_{\mathbf{z}}[\hat{\vartheta}^{(1)}_{U^1}(\mathbf{z})] \\ &\geq \mathbb{E}[\hat{\vartheta}^{(1)}_{U^1}(\mathscr{Z}^n)] - \sum_{\mathbf{z}\in\mathscr{Z}^n} \mathbb{P}[A_2^{(1)}(\mathbf{z})^c] \\ &\geq \mathbb{E}[\hat{\vartheta}^{(1)}_{U^1}(\mathscr{Z}^n)] - L_0|\mathscr{Z}|^n p^{(1)}.\end{aligned}$$

In (6.42), the integrand of $\mathbb{E}[\vartheta^{(1)}_{U^1}(\mathscr{Z}^n)]$ was lower-bounded by $1 - 2\cdot 2^{-nc\delta^2} - \varepsilon$. We conclude

$$(6.33) \leq 7\varepsilon + 2\cdot 2^{-nc\delta^2} + L_0|\mathscr{Z}|^n p^{(1)}.$$

Finally, we use condition (6.28) to bound (6.36). We have

$$\begin{aligned}&(6.36) && (6.43)\\ &= \frac{1}{L_0L_1L_2}\sum_{l_0,l_1,l_2} W_e^{\otimes n}(E_1^{(1)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})^c|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2}) \\ &= \frac{1}{L_0L_2}\sum_{l_0,l_2}\Bigg( \\ &\quad \frac{1}{L_1}\sum_{l_1:\mathbf{x}^{l_0l_1}\in T^n_{X|YU,\delta}(\mathbf{y}^{l_0l_2},\mathbf{u}^{l_0})} W_e^{\otimes n}(E_1^{(1)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})^c|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2}) && (6.44)\\ &\quad + \frac{1}{L_1}\sum_{l_1:\mathbf{x}^{l_0l_1}\notin T^n_{X|YU,\delta}(\mathbf{y}^{l_0l_2},\mathbf{u}^{l_0})} W_e^{\otimes n}(E_1^{(1)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})^c|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0l_2})\Bigg). && (6.45)\end{aligned}$$

For every $(l_0,l_2)$, we use $T^n_{Z|YXU,\delta}(\mathbf{y},\mathbf{x},\mathbf{u}) \subset E_1^{(1)}(\mathbf{u},\mathbf{x},\mathbf{y})$ for $(\mathbf{u},\mathbf{x},\mathbf{y}) \in T^n_{U,\delta}\times T^n_{Y|U,\delta}(\mathbf{u})\times T^n_{X|YU,\delta}(\mathbf{y},\mathbf{u})$ as shown in the proof of Lemma 6.22 to upper-bound the term in (6.44) by $2^{-nc\delta^2}$. For (6.45), we know from assumption (6.28) that it is at most $1-(1-\varepsilon)(1-2\cdot 2^{-nc\delta^2})$. Thus

$$(6.36) \leq 2^{-nc\delta^2} + (1-\varepsilon)(1-2\cdot 2^{-nc\delta^2}) \leq \varepsilon + 3\cdot 2^{-nc\delta^2}.$$

Collecting the bounds on (6.32)-(6.36), we obtain a total upper bound of

$$20\varepsilon + 9\cdot 2^{-nc\delta^2} + L_0|\mathscr{Z}|^n p^{(1)}.$$

This finishes the proof of Lemma 6.21.

**Bounds for Case 2:**

Now we specialize to the case that $L_2 = 1$, but $L_0$ and $L_1$ arbitrary. This reduces the number of Chernoff-type estimates needed by one. Lemma 6.18 carries over, Lemma 6.19 is not needed, but Lemma 6.20 changes. We write $Y^{l_0 1} =: Y^{l_0}$. The definitions of $E_1^{(1)}(\mathbf{u},\mathbf{x},\mathbf{y}), F_1^{(1)}(\mathbf{u},\mathbf{y})$ and $\vartheta_{\mathbf{uy}}^{(1)}$ carry over to this case, we just call them $E_1^{(2)}(\mathbf{u},\mathbf{x},\mathbf{y}), F_1^{(2)}(\mathbf{u},\mathbf{y})$ and $\vartheta_{\mathbf{uy}}^{(2)}$. Further we define

$$E_0^{(2)}(\mathbf{u},\mathbf{x},\mathbf{y}) := E_1^{(2)}(\mathbf{u},\mathbf{x},\mathbf{y}) \cap F_1^{(2)}(\mathbf{u},\mathbf{y}).$$

For every $l_0$, let $A_1^{(2)}(l_0,\mathbf{z}) := A_1^{(1)}(l_0,1,\mathbf{z})$ and we set $A_1^{(2)}(\mathbf{z}) := A_1^{(2)}(1,\mathbf{z}) \cap \ldots \cap A_1^{(2)}(L_0,\mathbf{z})$. We define for every $\mathbf{z}$ a new probability measure by $\hat{\mathbb{P}}_{\mathbf{z}}^{(2)} := \mathbb{P}[\cdot|A_1^{(2)}(\mathbf{z})]$. Let

$$\vartheta^{(2)}(\mathbf{z}) := \hat{\mathbb{E}}_{\mathbf{z}}^{(2)}[\hat{\vartheta}_{U^1Y^1}^{(2)}(\mathbf{z})].$$

Further let

$$F_0^{(2)} := \{\mathbf{z} \in T_{Z,4|\mathscr{Y}||\mathscr{X}||\mathscr{U}|\delta}^n : \vartheta^{(2)}(\mathbf{z}) \geq \varepsilon |T_{Z,\delta}^n|^{-1}\}$$

and

$$\hat{\vartheta}^{(2)} = \vartheta^{(2)} \cdot 1_{F_0^{(2)}}.$$

**Lemma 6.23.** *Let $\mathbf{z} \in F_0^{(2)}$. Let $A_0^{(2)}(\mathbf{z})$ be the event*

$$\frac{1}{L_0 L_1} \sum_{l_0,l_1} W_e^{\otimes n}(\mathbf{z}|X^{l_0 l_1}, Y^{l_0}) 1_{E_0^{(2)}(U^{l_0}, X^{l_0 l_1}, Y^{l_0})}(\mathbf{z}) \in [(1 \pm 3\varepsilon)\hat{\vartheta}^{(2)}(\mathbf{z})].$$

*Then for $\varepsilon$ sufficiently small and $n$ sufficiently large,*

$$\mathbb{P}[A_0^{(2)}(\mathbf{z})^c] \leq 2L_0 \exp\left(-L_1 \cdot \frac{\varepsilon^3 2^{-n(I(Z\wedge X|YU)+f_1(\delta)+f_2(\delta))}}{2\ln 2}\right) + 2\exp\left(-L_0 \cdot \frac{\varepsilon^3 2^{-n(I(Z\wedge YU)+f_1(\delta)+f_6(\delta))}}{4\ln 2}\right).$$

*Proof.* We have

$$\mathbb{P}[A_0^{(2)}(\mathbf{z})^c] \leq \hat{\mathbb{P}}_{\mathbf{z}}^{(2)}[A_0^{(2)}(\mathbf{z})^c] + \mathbb{P}[A_1^{(2)}(\mathbf{z})^c]. \tag{6.46}$$

By Lemma 6.18, we know that

$$\mathbb{P}[A_1^{(2)}(\mathbf{z})^c] \leq 2L_0 \exp\left(-L_1 \cdot \frac{\varepsilon^3 2^{-n(I(Z\wedge X|YU)+f_1(\delta)+f_2(\delta))}}{2\ln 2}\right). \tag{6.47}$$

In order to bound $\mathbb{P}_{\mathbf{z}}^{(2)}[A_0^{(2)}(\mathbf{z})]$, note that the sets $A_1^{(2)}(1,\mathbf{z}),\ldots,A_1^{(2)}(L_0,\mathbf{z})$ are independent with respect to $\mathbb{P}$. Thus under $\hat{\mathbb{P}}_{\mathbf{z}}^{(2)}$, the random variables

$$\tilde{W}_{\mathbf{z}}^{(2)}(l_0) := \frac{1}{L_1} \sum_{l_1} W_e^{\otimes n}(\mathbf{z}|X^{l_0 l_1}, Y^{l_0}) 1_{E_0^{(2)}(U^{l_0}, X^{l_0 l_1}, Y^{l_0})}(\mathbf{z}) \qquad (l_0 \in [L_0])$$

retain their independence and are upper-bounded by

$$(1+\varepsilon)\max_{\mathbf{u}\in T^n_{U,\delta}}\max_{\mathbf{y}\in T^n_{Y|U,\delta}(\mathbf{u})}\hat{\vartheta}^{(2)}_{\mathbf{uy}}(\mathbf{z}).$$

We can further bound this last term as follows: for $\mathbf{u}\in T^n_{U,\delta}$ and $\mathbf{y}\in T^n_{Y|U,\delta}(\mathbf{u})$ one obtains by (2.20) and (2.21)

$$\begin{aligned}\hat{\vartheta}^{(2)}_{\mathbf{uy}}(\mathbf{z}) &\le \mathbb{E}[W_e^{\otimes n}(\mathbf{z}|X^{11},\mathbf{y})|U^{l_0}=\mathbf{u}]\\ &\le \frac{1}{1-2^{-nc\delta^2}}P^{\otimes n}_{Z|YU}(\mathbf{z}|\mathbf{y},\mathbf{u})\\ &\le \frac{1}{1-2^{-nc\delta^2}}2^{-n(H(Z|YU)-f_1(\delta))}.\end{aligned}$$

Observing that $\hat{\mathbb{E}}_{\mathbf{z}}[\tilde{W}^{(2)}_{\mathbf{z}}(1)]\in[(1\pm\varepsilon)\hat{\vartheta}^{(2)}(\mathbf{z})]$ and applying Lemma 6.16 in the usual way yields

$$\begin{aligned}\hat{\mathbb{P}}_{\mathbf{z}}[A^{(2)}_0(\mathbf{z})] &\le 2\exp\left(-L_0\cdot\frac{\varepsilon^2(1-2^{-nc\delta^2})2^{n(H(Z|YU)-f_1(\delta))}(1-\varepsilon)\hat{\vartheta}^{(2)}(\mathbf{z})}{2(1+\varepsilon)\ln 2}\right)\\ &\le 2\exp\left(-L_0\cdot\frac{\varepsilon^3 2^{-n(I(Z\wedge YU)+f_1(\delta)+f_6(\delta))}}{4\ln 2}\right)\end{aligned}$$

if $\varepsilon$ is sufficiently small and $n$ sufficiently large. Inserting this and (6.47) in (6.46) completes the proof. □

Again we note that a result analogous to Lemma 6.23 holds where the roles of $X$ and $Y$ are exchanged. Setting $A^{(2)}_*(l_0):=A^{(1)}_*(l_0,1)$, we denote the events corresponding to such an exchange by $A^{(2)}_*(l_0)'$ and $A^{(2)}_1(l_0,\mathbf{z})', A^{(2)}_0(\mathbf{z})'$.

### Secrecy for Case 2:

**Lemma 6.24.** *Denote by $p^{(2)}$ the bound on $\mathbb{P}[A^{(2)}_1(l_0,\mathbf{z})^c]$ derived in Lemma 6.18. Let $\{\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0}:(l_0,l_1,l_2)\in[L_0]\times[L_1]\times[L_2]\}$ be a realization of $\mathscr{F}$ satisfying the conditions of*

$$\bigcap_{l_0}A^{(2)}_*(l_0), \tag{6.48}$$

$$\bigcap_{l_0}\bigcap_{\mathbf{z}\in\mathscr{Z}^n}A^{(2)}_1(l_0,\mathbf{z}), \tag{6.49}$$

$$\bigcap_{\mathbf{z}\in F^{(2)}_0}A^{(2)}_0(\mathbf{z}). \tag{6.50}$$

*Then*

$$\|\hat{\vartheta}^{(2)}-\frac{1}{L_0L_1}\sum_{l_0,l_1}W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0})\|\le 9\varepsilon+7\cdot 2^{-nc\delta^2}+L_0|\mathscr{Z}|^n p^{(2)}.$$

*The same inequality is true if we require conditions* (6.48′)-(6.50′) *which contain the primed equivalents of* (6.48)-(6.50) *defined at the end of the previous paragraph.*

We now prove the above lemma. We have

$$\begin{aligned}
&\|\hat{\vartheta}^{(2)} - \frac{1}{L_0L_1}\sum_{l_0,l_1} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0})\| \\
&\leq \|\hat{\vartheta}^{(2)} - \frac{1}{L_0L_1}\sum_{l_0,l_1} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0})1_{E_0^{(2)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0})}1_{F_0^{(2)}}\| && (6.51)\\
&+ \|\frac{1}{L_0L_1}\sum_{l_0,l_1} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0})1_{E_0^{(2)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0})}(1-1_{F_0^{(2)}})\| && (6.52)\\
&+ \|\frac{1}{L_0L_1}\sum_{l_0,l_1} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0})1_{E_1^{(2)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0})}(1-1_{F_1^{(2)}(\mathbf{u}^{l_0},\mathbf{y}^{l_0})})\| && (6.53)\\
&+ \|\frac{1}{L_0L_1}\sum_{l_0,l_1} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0})(1-1_{E_1^{(2)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0})})\|. && (6.54)
\end{aligned}$$

Due to (6.50), we know that (6.51) $\leq \varepsilon$.

Next we consider (6.53). Due to (6.49), we have

$$\begin{aligned}
(6.53) &\leq 1 - \frac{1}{L_0L_1}\sum_{l_0,l_1} W_e^{\otimes n}(E_0^{(2)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0})|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0}) \\
&\leq 1 - \frac{1-\varepsilon}{L_0}\sum_{l_0} \hat{\vartheta}^{(2)}_{\mathbf{u}^{l_0}\mathbf{y}^{l_0}}(\mathscr{Z}^n).
\end{aligned}$$

As done in Lemma 6.22 for Case 1, one lower-bounds $\hat{\vartheta}^{(2)}_{\mathbf{u}^{l_0}\mathbf{y}^{l_0}}(\mathscr{Z}^n) \geq \vartheta^{(2)}_{\mathbf{u}^{l_0}\mathbf{y}^{l_0}}(\mathscr{Z}^n) - \varepsilon$ by $1 - 2\cdot 2^{-nc\delta^2} - \varepsilon$. Thus we can conclude that

$$(6.53) \leq 2(\varepsilon + 2^{-nc\delta^2}).$$

For (6.52), we have by (6.49)

$$\begin{aligned}
(6.52) &\leq 1 - \frac{1}{L_0L_1}\sum_{l_0,l_1} W_e^{\otimes n}(E_0^{(2)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0}) \cap F_0^{(2)}|\mathbf{x}^{l_0l_1},\mathbf{y}^{l_0}) \\
&\leq 1 - (1-3\varepsilon)\hat{\vartheta}^{(2)}(F_0^{(2)}).
\end{aligned}$$

It remains to lower-bound $\hat{\vartheta}^{(2)}(F_0^{(2)}) \geq \vartheta^{(2)}(\mathscr{Z}^n) - \varepsilon$. As in the lower bound on $\vartheta^{(1)}(\mathscr{Z}^n)$ above, one obtains the bound

$$\vartheta^{(2)}(\mathscr{Z}^n) \geq 1 - 2\cdot 2^{-nc\delta^2} - \varepsilon - L_0|\mathscr{Z}|^n p^{(2)}.$$

Thus we conclude

$$(6.52) \leq 5\varepsilon + 2\cdot 2^{-nc\delta^2} + L_0|\mathscr{Z}|^n p^{(2)}.$$

Finally, we use condition (6.48) to bound (6.54). We have

$$\begin{aligned}(6.54) &= \frac{1}{L_0 L_1}\sum_{l_0,l_1} W_e^{\otimes n}(E_1^{(2)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0 l_1},\mathbf{y}^{l_0})|\mathbf{x}^{l_0 l_1},\mathbf{y}^{l_0}) \\ &= \frac{1}{L_0}\sum_{l_0}\Bigg(\frac{1}{L_1}\sum_{l_1:\mathbf{x}^{l_0 l_1}\in T^n_{X|YU,\delta}(\mathbf{y}^{l_0},\mathbf{u}^{l_0})} W_e^{\otimes n}(E_1^{(2)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0 l_1},\mathbf{y}^{l_0})|\mathbf{x}^{l_0 l_1},\mathbf{y}^{l_0}) \qquad (6.55)\\ &\quad + \frac{1}{L_1}\sum_{l_1:\mathbf{x}^{l_0 l_1}\notin T^n_{X|YU,\delta}(\mathbf{y}^{l_0},\mathbf{u}^{l_0})} W_e^{\otimes n}(E_1^{(2)}(\mathbf{u}^{l_0},\mathbf{x}^{l_0 l_1},\mathbf{y}^{l_0})|\mathbf{x}^{l_0 l_1},\mathbf{y}^{l_0})\Bigg). \qquad (6.56)\end{aligned}$$

For every $l_0$, the summand appearing in (6.55) can be upper-bounded by $2^{-nc\delta^2}$. By assumption (6.48), (6.56) is upper-bounded by $1-(1-\varepsilon)(1-2\cdot 2^{-nc\delta^2})$. Thus

$$(6.54) \leq \varepsilon + 3\cdot 2^{-nc\delta^2}.$$

Collecting the bounds for (6.51)-(6.54), we obtain a total upper bound of

$$9\varepsilon + 7\cdot 2^{-nc\delta^2} + L_0|\mathscr{Z}|^n p^{(2)}.$$

This finishes the proof of Lemma 6.24.

**Bounds for Case 3:**

Now we treat the case $L_1 = L_2 = 1$. Lemma 6.25 is the analog of Lemma 6.17, the proof is analogous.

**Lemma 6.25.** *Let the event $A_*^{(3)}$ be defined by*

$$A_*^{(3)} := \big\{|\{l_0 \in [L_0] : X^{l_0} \in T^n_{X|YU,\delta}(Y^{l_0},U^{l_0})\}| \geq (1-\varepsilon)(1-2\cdot 2^{-nc_1\delta^2})L_0\big\}.$$

*Then*

$$\mathbb{P}[(A_*^{(3)})^c] \leq \exp\left(-L_0\cdot\frac{\varepsilon^2(1-2\cdot 2^{-nc_1\delta^2})}{2\ln 2}\right).$$

Let

$$E^{(3)}(\mathbf{x},\mathbf{y}) := \{\mathbf{z}\in T^n_{Z,4|\mathscr{Y}||\mathscr{X}||\mathscr{U}|\delta} : W_e^{\otimes n}(\mathbf{z}|\mathbf{x},\mathbf{y}) \leq 2^{-n(H(Z|XY)-f_2(\delta))}\},$$

where $f_2(\delta) = \tau(P_{XYZ}, 3\delta, \delta)$. Let

$$\vartheta^{(3)}(\mathbf{z}) := \mathbb{E}[W_e^{\otimes n}(\mathbf{z}|X^1,Y^1)1_{E_1(X^1,Y^1)}(\mathbf{z})]$$

and for

$$F^{(3)} := \{\mathbf{z}\in T^n_{Z,4|\mathscr{Y}||\mathscr{X}||\mathscr{U}|\delta} : \vartheta(\mathbf{z}) \geq \varepsilon|T^n_{Z,\delta}|^{-1}\}$$

define the measure

$$\hat{\vartheta}^{(3)} := \hat{\vartheta}^{(3)}\cdot 1_{F^{(3)}}.$$

**Lemma 6.26.** *Let $\mathbf{z} \in F^{(3)}$. Let $A^{(3)}(\mathbf{z})$ be the event that*

$$\frac{1}{L_0}\sum_{l_0} W_e^{\otimes n}(\mathbf{z}|X^{l_0}, Y^{l_0}) 1_{E^{(3)}(X^{l_0},Y^{l_0})}(\mathbf{z}) \in [(1 \pm \varepsilon)\hat{\vartheta}^{(3)}(\mathbf{z})].$$

*Then for $f_1(\delta) = \tau(P_{UYZ}, 4|\mathscr{Y}||\mathscr{X}||\mathscr{U}|\delta, \delta)$,*

$$\mathbb{P}[A^{(3)}(\mathbf{z})^c] \leq 2\exp\left(-L_0 \cdot \frac{\varepsilon^3 2^{-n(I(Z\wedge XY)+f_1(\delta)+f_2(\delta))}}{2\ln 2}\right).$$

The proof of this lemma consists of the usual application of Lemma 6.16.

**Secrecy for Case 3:**

**Lemma 6.27.** *Let $\{(\mathbf{u}^{l_0}, \mathbf{x}^{l_0}, \mathbf{y}^{l_0})\}$ be a realization of $\mathscr{F}$ satisfying the conditions of*

$$A_*^{(3)}, \tag{6.57}$$

$$\bigcap_{\mathbf{z}\in F^{(3)}} A^{(3)}(\mathbf{z}). \tag{6.58}$$

*Then for sufficiently large $n$,*

$$\|\hat{\vartheta}^{(3)} - \frac{1}{L_0}\sum_{l_0} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0}, \mathbf{y}^{l_0})\| \leq 4\varepsilon + 5\cdot 2^{-nc\delta^2}. \tag{6.59}$$

We now prove the above lemma. We have

$$\begin{aligned}
&\|\hat{\vartheta}^{(3)} - \frac{1}{L_0}\sum_{l_0} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0}, \mathbf{y}^{l_0})\| \\
&\leq \|\hat{\vartheta}^{(3)} - \frac{1}{L_0}\sum_{l_0} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0}, \mathbf{y}^{l_0}) 1_{E^{(3)}(\mathbf{x}^{l_0},\mathbf{y}^{l_0})} 1_{F^{(3)}}\| && (6.60)\\
&+ \|\frac{1}{L_0}\sum_{l_0} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0}, \mathbf{y}^{l_0}) 1_{E^{(3)}(\mathbf{x}^{l_0},\mathbf{y}^{l_0})}(1 - 1_{F^{(3)}})\| && (6.61)\\
&+ \|\frac{1}{L_0}\sum_{l_0} W_e^{\otimes n}(\cdot|\mathbf{x}^{l_0}, \mathbf{y}^{l_0})(1 - 1_{E^{(3)}(\mathbf{x}^{l_0},\mathbf{y}^{l_0})})\|. && (6.62)
\end{aligned}$$

Due to (6.58) we have (6.60) $\leq \varepsilon$.

Next we bound (6.61). Again using (6.58),

$$\begin{aligned}
(6.61) &\leq 1 - \frac{1}{L_0}\sum_{l_0} W_e^{\otimes n}(E^{(3)}(\mathbf{x}^{l_0}, \mathbf{y}^{l_0}) \cap F^{(3)}|\mathbf{x}^{l_0}, \mathbf{y}^{l_0}) \\
&\leq 1 - (1-\varepsilon)\hat{\vartheta}^{(3)}(F^{(3)}). && (6.63)
\end{aligned}$$

As in Case 1 and 2, $\hat{\vartheta}^{(3)}(F^{(3)})$ can be lower-bounded by $1-2\cdot 2^{-nc\delta^2}-\varepsilon$, so

$$(6.61) \leq 1-(1-\varepsilon)(1-2\cdot 2^{-nc\delta^2}-\varepsilon) \leq 2(\varepsilon+2^{-nc\delta^2}).$$

Finally, the third term (6.62) equals

$$\begin{aligned}
&\frac{1}{L_0}\sum_{l_0} W_e^{\otimes n}(E^{(3)}(\mathbf{x}^{l_0},\mathbf{y}^{l_0})^c|\mathbf{x}^{l_0},\mathbf{y}^{l_0}) \\
&= \frac{1}{L_0}\sum_{l_0:\mathbf{x}^{l_0}\in T_{X|YU,\delta}(\mathbf{y}^{l_0},\mathbf{u}^{l_0})} W_e^{\otimes n}(E^{(3)}(\mathbf{x}^{l_0},\mathbf{y}^{l_0})^c|\mathbf{x}^{l_0},\mathbf{y}^{l_0}) \qquad (6.64)\\
&+ \frac{1}{L_0}\sum_{l_0:\mathbf{x}^{l_0}\notin T_{X|YU,\delta}(\mathbf{y}^{l_0},\mathbf{u}^{l_0})} W_e^{\otimes n}(E^{(3)}(\mathbf{x}^{l_0},\mathbf{y}^{l_0})^c|\mathbf{x}^{l_0},\mathbf{y}^{l_0}). \qquad (6.65)
\end{aligned}$$

and is lower-bounded by

$$(6.62) \leq 2^{-nc\delta^2}+(1-\varepsilon)(1-2\cdot 2^{-nc\delta^2}) \leq \varepsilon+3\cdot 2^{-nc\delta^2}.$$

Combining the above bounds, we can conclude that

$$(6.60)+(6.61)+(6.62) \leq 4\varepsilon+5\cdot 2^{-nc\delta^2},$$

which completes the proof of Lemma 6.27.

### 6.4.4. Random Coding for the Non-Wiretap MAC with Common Message

Assume we are given another family of random variables

$$\mathscr{F}' := \bigcup_{l_0\in[L_0]} (U^{l_0}, \mathscr{F}'_{l_0})$$

with $\mathscr{F}'_{l_0} = \{X^{l_0 l_1'}, Y^{l_0 l_2'} : l_1', l_2' \in [L_1']\times[L_2']\}$ for other positive integers $L_0', L_1', L_2'$ with blocklength $n'$ which is independent of $\mathscr{F}$, but which has the same structure as $\mathscr{F}$ and whose distribution is defined according to the same $p$ as $\mathscr{F}$. Assume that for some $\eta>0$

$$\begin{aligned}
\frac{n\log L_1 + n'\log L_1'}{n+n'} &\leq [I(T\wedge X|YU)-\eta]_+,\\
\frac{n\log L_2 + n'\log L_2'}{n+n'} &\leq [I(T\wedge Y|XU)-\eta]_+,\\
\frac{n\log(L_1L_2) + n'\log(L_1'L_2')}{n+n'} &\leq [I(T\wedge XY|U)-\eta]_+,\\
\frac{n\log(L_0L_1L_2) + n'\log(L_0'L_1'L_2')}{n+n'} &\leq [I(T\wedge XY)-\eta]_+.
\end{aligned}$$

Note that this means in particular

$$\left(\frac{n\log L_0 + n'\log L_0'}{n+n'}, \frac{n\log L_1 + n'\log L_1'}{n+n'}, \frac{n\log L_2 + n'\log L_2'}{n+n'}\right) \in \mathscr{R}_{\mathrm{CM}}(p),$$

$\mathscr{R}_{\mathrm{CM}}(p)$ was defined in Section 2.1. Define a new family of random vectors

$$\mathscr{F} \circ \mathscr{F}' := \{\tilde{U}^{l_0 l_0'}, \tilde{X}^{l_0 l_0' l_1 l_1'}, \tilde{Y}^{l_0 l_0' l_2 l_2'} : l_0 \ldots, l_2'\} \tag{6.66}$$

by concatenating the corresponding elements of $\mathscr{F}$ and $\mathscr{F}'$, so e.g. $\tilde{U}^{l_0 l_0'} = (U^{l_0}, U^{l_0'}) \in \mathscr{U}^{n+n'}$, $\tilde{X}^{l_0 l_0' l_1 l_1'} = (X^{l_0 l_1}, X^{l_0' l_1'}) \in \mathscr{X}^{n+n'}$.

**Lemma 6.28.** *For any $\delta, \eta > 0$ there are $\zeta_1, \zeta_2 = \zeta_1(\eta, \delta), \zeta_2(\eta, \delta) > 0$ such that the probability of the event $A_{\mathrm{MAC}}$ that the family*

$$\{\tilde{X}^{l_0 l_0' l_1 l_1'}, \tilde{Y}^{l_0 l_0' l_2 l_2'} : l_0, l_0', l_1, l_1', l_2, l_2'\}$$

*is the codeword set of a deterministic $(n+n')$-code$_{\mathrm{CM}}$ $\gamma$ with $\bar{e}^{\mathrm{DM}}(\gamma, W_b) \leq \exp(-(n+n')\zeta_1)$ is lower-bounded by $1-\exp(-(n+n')\zeta_2)$. The same result is true if it is formulated only for $\mathscr{F}$ or $\mathscr{F}'$ without concatenation.*

*Proof.* The difference to standard random coding proofs is that the random variables from $\mathscr{F}$ and $\mathscr{F}'$ are conditioned on typicality. Using the random sets

$$E^{l_0 l_0' l_1 l_1' l_2 l_2'} := \{\mathbf{t} \in \mathscr{T}^{n+n'} : (\tilde{U}^{l_0 l_0'}, \tilde{X}^{l_0 l_0' l_1 l_1'}, \tilde{Y}^{l_0 l_0' l_2 l_2'}, \mathbf{t}) \in T^{n+n'}_{UXYT,\delta}\},$$

we define the decoding sets $F^{l_0 l_0' l_1 l_1' l_2 l_2'}$ by deciding for $(l_0, l_0', l_1, l_1', l_2, l_2')$ if the output is contained in $E^{l_0 l_0' l_1 l_1' l_2 l_2'}$ and if at the same time it is not contained in any $E^{\tilde{l}_0 \tilde{l}_0' \tilde{l}_1 \tilde{l}_1' \tilde{l}_2 \tilde{l}_2'}$ for a different message vector $(\tilde{l}_0, \tilde{l}_0', \tilde{l}_1, \tilde{l}_1', \tilde{l}_2, \tilde{l}_2')$. This decoder is known to be the right decoder in the case where the codewords have the standard i.i.d. structure, i.e. for a family of random variables

$$\{\hat{U}^{l_0 l_0'}, \hat{X}^{l_0 l_0' l_1 l_1'}, Y^{l_0 l_0' l_2 l_2'}\}$$

where $\hat{U}^{l_0 l_0'} \sim P_U^{\otimes(n+n')}$ and where conditional on $\hat{U}^{l_0 l_0'}$, the $\hat{X}^{l_0 l_0' l_1 l_1'}$ and $\hat{Y}^{l_0 l_0' l_2 l_2'}$ are independent with $\hat{X}^{l_0 l_0' l_1 l_1'} \sim P_{X|U}^{\otimes(n+n')}$ and $\hat{Y}^{l_0 l_0' l_2 l_2'} \sim P_{Y|U}^{\otimes(n+n')}$. It is easily seen that

$$\begin{aligned}\mathbb{E}[W_b^{\otimes n}((F^{l_0 l_0' l_1 l_1' l_2 l_2'})^c | \tilde{X}^{l_0 l_0' l_1 l_1'}, \tilde{Y}^{l_0 l_0' l_2 l_2'})] \\ \leq (1 - 2^{-nc\delta^2})^3 (1 - 2^{-n'c\delta^2})^3 \mathbb{E}[W_b^{\otimes n}((F^{l_0 l_0' l_1 l_1' l_2 l_2'})^c | \hat{X}^{l_0 l_0' l_1 l_1'}, \hat{Y}^{l_0 l_0' l_2 l_2'})].\end{aligned}$$

Then the standard random coding proof technique yields the result. The specialization for the case that only $\mathscr{F}$ or $\mathscr{F}'$ is treated is obvious. □

### 6.4.5. Coding

In this subsection we show the WCM-achievability of the rate sets $\mathscr{R}^{(\nu)}(p)$ for $\nu = 0, 1, 2, 3$ and appropriate $p$. For the cases where we showed that $\mathscr{R}^{(\nu)}(p)$ can be written as the union over certain $\alpha$ of rate sets $\mathscr{R}^{(\nu)}_\alpha(p)$, we show the WCM-achievability of the latter for every $\alpha$.

Throughout this section fix a common randomness bound $H_C \geq 0$. Let $\delta > 0$ which will be specified later and $n$ a blocklength which will have to be large enough. Every $p$ considered in this section is from $\Pi(W)$, i.e. has the form $p = P_U \otimes (P_{X|U} \otimes P_{Y|U}) \otimes W$.

Without loss of generality we may assume that $I(Z \wedge XY) < I(T \wedge XY)$, in particular, $I(T \wedge XY) > 0$. Let $n, n'$ be nonnegative integers and

$$K_0, K_1, K_2, L_0, L_1, L_2, \qquad K_0', K_1', K_2', L_0', L_1', L_2' \tag{6.67}$$

be arbitrary positive integers such that $K_0, \dots, L_2$ all equal 1 if $n = 0$ and analogously for the primed parameters. We define two independent families $\mathscr{G}, \mathscr{G}'$ of random vectors. $\mathscr{G}$ has the same form as $\mathscr{F}$ with the parameters $L_0, L_1, L_2$ replaced by $K_0L_0, K_1L_1, K_2L_2$. $\mathscr{G}'$ is defined analogously with the parameters on the left-hand side of (6.67) replaced by those on its right-hand side. Every choice of $(k_0, k_1, k_2)$ induces a subfamily $\mathscr{F}$ of $\mathscr{G}$ which has the same parameters $l_0, l_1, l_2, n$ as the $\mathscr{F}$ treated above, every subfamily of $\mathscr{G}'$ corresponding to any $(k_0', k_1', k_2')$ induces an $\mathscr{F}'$ with parameters $l_0', l_1', l_2', n'$. Further recall the notation $\mathscr{G} \circ \mathscr{G}'$ as the family of concatenated words from $\mathscr{G}$ and $\mathscr{G}'$ as in (6.66).

**Case 0 and 1:**

Let $p \in \Psi^{(0)}(W) \cap \Pi(W)$ or $p \in \Psi_{H_C}^{(1)}(W) \cap \Pi(W)$. Note that $\alpha_0^{(1)} \le \alpha_1^{(1)}$ if and only if the vector $(J_0^{(\alpha)}, J_1^{(\alpha)}, J_2^{(\alpha)})$ whose components are given by

$$\begin{aligned} J_0^{(\alpha)} &= I(Z \wedge U), \\ J_1^{(\alpha)} &= \alpha I(Z \wedge X|YU) + (1-\alpha) I(Z \wedge X|U), \\ J_2^{(\alpha)} &= \alpha I(Z \wedge Y|U) + (1-\alpha) I(Z \wedge Y|XU) \end{aligned}$$

is contained in $\mathscr{R}_{\mathrm{CM}}(p)$. We first consider Case 1. Let a rate vector $(R_0, R_1, R_2)$ with positive components be given such that $(\tilde{R}_0, \tilde{R}_1, \tilde{R}_2) := (R_0, R_1, R_2) + (J_0^{(\alpha)}, J_1^{(\alpha)}, J_2^{(\alpha)}) \in \mathscr{R}_{\mathrm{CM}}(p)$, which means that $(R_0, R_1, R_2) \in \mathscr{R}_\alpha^{(1)}(p)$. We now define a wiretap code whose rates approximate $(R_0, R_1, R_2)$. If $\alpha = 0$, we only need $\mathscr{G}'$ and set $n = 0$, if $\alpha = 1$, we only need $\mathscr{G}$ and set $n' = 0$. Otherwise we do time-sharing in the following way: choose for a small $0 < \xi < \min\{\alpha, 1-\alpha\}$ blocklengths $n$ and $n'$ with $n/(n+n') \in (\alpha - \xi, \alpha + \xi)$. For some $0 < 2\eta < \min\{R_0, R_1, R_2\}$ and every $\nu = 0, 1, 2$ let

$$\tilde{R}_\nu - \eta \le \frac{\log(K_\nu L_\nu) + \log(K_\nu' L_\nu')}{n + n'} \le \tilde{R}_\nu - \frac{\eta}{2}$$

(and this modifies accordingly for $\alpha \in \{0, 1\}$). By Lemma 6.28 we know that with probability exponentially close to 1, the random variables $\tilde{X}_{k_0 k_0' k_1 k_1'}^{l_0 l_0' l_1 l_1'}$ and $\tilde{Y}_{k_0 k_0' k_2 k_2'}^{l_0 l_0' l_2 l_2'}$ form the codewords of an $(n+n)$-code$_{\mathrm{CM}}$ $\gamma$ for DMAC$(W_b)$ with $\bar{e}^{\mathrm{DM}}(\gamma, W_b) \le \exp(-(n + n')\zeta_1)$ for some $\zeta_1 > 0$, i.e. satisfy the conditions of $A_{\mathrm{MAC}}$. Choosing $\delta$ so small that $4(f_1(\delta) + f_2(\delta) + f_4(\delta) + f_6(\delta)) \le \min\{\eta, H_C - J_0^{(\alpha)}\}$, we can achieve

$$\begin{aligned} \frac{\log L_1 + \log L_1'}{n + n'} &\in J_1^{(\alpha)} + \big(f_1(\delta) + (\alpha f_2(\delta) + (1-\alpha) f_4(\delta))\big) \cdot [2, 3], \\ \frac{\log L_2 + \log L_2'}{n + n'} &\in J_2^{(\alpha)} + \big(f_1(\delta) + (\alpha f_4(\delta) + (1-\alpha) f_2(\delta))\big) \cdot [2, 3], \\ \frac{\log L_0 + \log L_0'}{n + n'} &\in J_0^{(\alpha)} + \big(f_4(\delta) + f_6(\delta)\big) \cdot [2, 3]. \end{aligned}$$

If additionally $\varepsilon$ is chosen according to

$$-\frac{1}{n}\log\varepsilon = \frac{1}{4}\min\{4\zeta_1, f_1(\delta)+f_2(\delta)+f_4(\delta)+f_6(\delta)\},$$

then for every $(k_0,k_1,k_2)\in[K_0]\times[K_1]\times[K_2]$, the corresponding subfamily $\mathscr{F}$ of $\mathscr{G}$ satisfies (6.28)-(6.31) with probability exponentially close to 1, and for every $(k_0',k_1',k_2')\in[K_0']\times[K_1']\times[K_2']$, the corresponding subfamily $\mathscr{F}'$ of $\mathscr{G}'$ satisfies (6.28′)-(6.31′) with probability exponentially close to 1. Thus we can choose a realization of $\mathscr{G}\circ\mathscr{G}'$ which simultaneously has all these properties plus those of $A_{\mathrm{MAC}}$ and use it to define an $(n+n',H_C)$- code$_{\mathrm{WCM}}$. We define independent encoders $G$ and $G'$ by setting

$$G_0(l_0|k_0) = \frac{1}{L_0}, \qquad (k_0\in[K_0], l_0\in[L_0]),$$

$$G_1(\mathbf{x}|k_0,k_1,l_0) = \frac{1}{L_1}\sum_{l_1}\delta_{\mathbf{x}_{k_0k_1}^{l_0l_1}}(\mathbf{x}), \qquad (\mathbf{x}\in\mathscr{X}^n, k_1\in[K_1], k_0\in[K_0], l_0\in[L_0]),$$

$$G_2(\mathbf{y}|k_0,k_2,l_0) = \frac{1}{L_2}\sum_{l_2}\delta_{\mathbf{y}_{k_0k_2}^{l_0l_2}}(\mathbf{y}), \qquad (\mathbf{y}\in\mathscr{Y}^n, k_2\in[K_2], k_0\in[K_0], l_0\in[L_0]),$$

and defining $G'$ analogously.

A message triple $((k_0,k_0'),(k_1,k_1'),(k_2,k_2'))$ is encoded into the pair of codewords $((\mathbf{x},\mathbf{x}'),(\mathbf{y},\mathbf{y}'))$ with probability

$$(G\otimes G')\big((\mathbf{x},\mathbf{x}'),(\mathbf{y},\mathbf{y}')|(k_0,k_0'),(k_1,k_1'),(k_2,k_2')\big) := G(\mathbf{x}|k_0,k_1,k_2)G'(\mathbf{x}'|k_0',k_1',k_2').$$

By choice of $\delta$, the common randomness constraint is satisfied. We choose the decoder as $\varphi$, the decoder from the $(n+n')$-code$_{\mathrm{CM}}$ $\gamma$ determined by the chosen realization of $\mathscr{G}\circ\mathscr{G}'$.

We have $\bar{e}^{\mathrm{WT}}(G\otimes G',\varphi,W) = \bar{e}^{\mathrm{DM}}(\gamma,W_b)$, recall that $\gamma$ is the deterministic $(n+n')$-code$_{\mathrm{CM}}$ for $W_b$ determined by the realization of $\mathscr{G}\circ\mathscr{G}'$. In particular $\bar{e}^{\mathrm{WT}}(G\otimes G',\varphi,W)\le\varepsilon$. Due to the choice of $\delta$ the rates of this code satisfy

$$\frac{\log K_\nu+\log K_\nu'}{n+n'} \ge R_\nu - 2\eta \qquad (\nu=0,1,2).$$

Finally if we let $M_\nu$ be uniformly distributed on $[K_\nu]$ and $M_\nu'$ on $[K_\nu']$, then it follows from Lemma 6.21 and (6.21) together with the fact that $\varepsilon$ is exponentially small that the strong secrecy criterion is satisfied. Thus the rate triple $(R_0,R_1,R_2)$ is WCM-achievable. So far, this excludes $(R_0,R_1,R_2)$ with some components equal to zero, but as $\delta$ and $\eta$ may be arbitrarily close to 0 and the WCM-achievable region of $W$ is closed by definition, we can conclude that the whole region $\mathscr{R}_\alpha^{(1)}(p)$ is WCM-achievable.

For Case 0, everything goes through if one sets $K_0=K_0'=L_0=L_0'=1$ and $R_0=0$. The crucial difference to Case 1 is that even if $J_0^{(\alpha)}=0$, one needs a little bit more common randomness than that in order to protect a common message, as can be seen in the choice of $L_0$ and $L_0'$ above. Thus the secret transmission of a common message is impossible if common randomness is not available.

**Case 2:**

Let $p \in \Psi_{H_C}^{(2)}(W) \cap \Pi(W)$. In this case we apply $\mathscr{G}$ with $L_2 = 1$ and $\mathscr{G}'$ with $L_1 = 1$. We define the vector $(J_0^{(\alpha)}, J_1^{(\alpha)}, J_2^{(\alpha)})$ by

$$J_0^{(\alpha)} = \alpha I(Z \wedge YU) + (1-\alpha) I(Z \wedge XU), \tag{6.68}$$

$$J_1^{(\alpha)} = \alpha I(Z \wedge X|YU), \tag{6.69}$$

$$J_2^{(\alpha)} = (1-\alpha) I(Z \wedge Y|XU) \tag{6.70}$$

As it should always be clear which case we are treating, this should not lead to confusion with case 1. Note that $\alpha_0^{(2)} \le \alpha \le \alpha_1^{(2)}$ if and only if $(J_0^{(\alpha)}, J_1^{(\alpha)}, J_2^{(\alpha)})$ is contained in $\mathscr{R}_{\mathrm{CM}}(p)$ and satisfies $J_0^{(\alpha)} < H_C$. Let a rate vector $(R_0, R_1, R_2)$ be given whose $\nu$-th component may only vanish if $L_\nu = L'_\nu = 1$. Further we require that $(\tilde{R}_0, \tilde{R}_1, \tilde{R}_2) = (R_0, R_1, R_2) + (J_0^{(\alpha)}, J_1^{(\alpha)}, J_2^{(\alpha)})$ is contained in $\mathscr{R}_{\mathrm{CM}}(p)$. If $\alpha = 0$, we only need $\mathscr{G}'$, if $\alpha = 1$, we only need $\mathscr{G}$. Otherwise, let $0 < \xi < \min\{\alpha, 1-\alpha\}$ be small and let $n$ and $n'$ be large enough such that $n/(n+n') \in (\alpha - \xi, \alpha + \xi)$. Further for some $0 < 2\eta < \min\{R_\nu : \nu = 0,1,2, R_\nu > 0\}$ let

$$[\tilde{R}_\nu - \eta]_+ \le \frac{\log(K_\nu L_\nu) + \log(K'_\nu L'_\nu)}{n+n'} \le [\tilde{R}_\nu - \frac{\eta}{2}]_+,$$

and modify this accordingly for $\alpha \in \{0,1\}$. By Lemma 6.28 we know that with probability exponentially close to 1, the random variables $\tilde{X}_{k_0 k'_0 k_1 k'_1}^{l_0 l'_0 l_1 l'_1}$ and $\tilde{Y}_{k_0 k'_0 k_2 k'_2}^{l_0 l'_0 l_2 l'_2}$ form the codewords of a deterministic $(n+n')$-code$_{\mathrm{CM}}$ for DMAC$(W_b)$ with $\bar{e}^{\mathrm{DM}}(\gamma, W_b) \le \exp(-(n+n')\zeta_1)$ for some $\zeta_1 > 0$. We define $(j_1^1, j_1^2) = (j_2^1, j_2^2) = (1,2)$ and $(j_0^1, j_0^2) = (1,6)$ and choose $\delta$ so small that $4(f_{j_\nu^1}(\delta) + f_{j_\nu^2}(\delta)) \le \min\{\eta, H_C - J_0^{(\alpha)}\}$ for all $\nu$. Then let for $\nu = 0,1,2$

$$J_\nu^{(\alpha)} + 2(f_{j_\nu^1}(\delta) + f_{j_\nu^2}(\delta)) \le \frac{\log L_\nu + \log L'_\nu}{n+n'} \le J_\nu^{(\alpha)} + 3(f_{j_\nu^1}(\delta) + f_{j_\nu^2}(\delta)),$$

If additionally $\varepsilon$ is chosen according to

$$-\frac{1}{n}\log\varepsilon = \frac{1}{4}\min\{4\zeta_1, f_1(\delta) + f_2(\delta), f_1(\delta) + f_6(\delta)\},$$

then for every $(k_0, k_1, k_2) \in [K_0] \times [K_1] \times [K_2]$, the corresponding subfamily $\mathscr{F}$ of $\mathscr{G}$ satisfies (6.48)-(6.50) with probability exponentially close to 1, and for every $(k'_0, k'_1, k'_2) \in [K'_0] \times [K'_1] \times [K'_2]$, the corresponding subfamily $\mathscr{F}'$ of $\mathscr{G}'$ satisfies (6.48′)-(6.50′) with probability exponentially close to 1. Thus we can choose a realization of $\mathscr{G} \circ \mathscr{G}'$ which has all these properties plus those defining $A_{\mathrm{MAC}}$ and use it to define an $(n+n', H_C)$-

$\text{code}_{\text{WCM}}$. We define independent encoders $G$ and $G'$ by setting

$$G_0(l_0|k_0) = \frac{1}{L_0}, \qquad (l_0 \in [L_0], k_0 \in [K_0]),$$
$$G_1(\mathbf{x}|k_0,k_1,l_0) = \frac{1}{L_1}\sum_{l_1} \delta_{\mathbf{x}_{k_0k_1}^{l_0l_1}}(\mathbf{x}), \qquad (\mathbf{x} \in \mathscr{X}^n, k_1 \in [K_1], k_0 \in [K_0], l_0 \in [L_0]),$$
$$G_2(\mathbf{y}|k_0,k_2,l_0) = \delta_{\mathbf{y}_{k_0k_2}^{l_0}}(\mathbf{y}), \qquad (\mathbf{y} \in \mathscr{Y}^n, k_2 \in [K_2], k_0 \in [K_0], l_0 \in [L_0]),$$

and defining $G'$ analogously. The encoder of the desired $\text{code}_{\text{WCM}}$ then is $G \otimes G'$ as in Case 1. The decoder $\varphi$ is the decoder of $\gamma$, the $n+n'$-$\text{code}_{\text{CM}}$ corresponding to the chosen realization of $\mathscr{G} \circ \mathscr{G}'$. $G \otimes G'$ satisfies the common randomness constraint. Due to the simple form of $G \otimes G'$, we have

$$\bar{e}^{\text{WT}}(G \otimes G', \varphi, W) = \bar{e}^{\text{DM}}(\gamma, W_b) \leq \varepsilon.$$

Due to the choice of $\delta$, the rates of this code satisfy

$$\frac{\log K_\nu + \log K'_\nu}{n+n'} \geq R_\nu - 2\eta \qquad (\nu = 0,1,2,).$$

Finally if we let $M_\nu$ be uniformly distributed on $[K_\nu]$ and $M'_\nu$ on $[K'_\nu]$, then it follows from Lemma 6.21 and (6.21) together with the fact that $\varepsilon$ is exponentially small that the strong secrecy criterion is satisfied. Thus the rate triple $(R_0, R_1, R_2)$ is WCM-achievable. So far, this may exclude rate triples $(R_0, R_1, R_2)$ where one component equals zero, but as $\delta$ and $\eta$ may be arbitrarily close to 0 and the WCM-achievable region of $W$ is closed by definition, we can conclude that the whole region $\mathscr{R}_\alpha^{(2)}(p)$ is WCM-achievable.

**Case 3:**

Let $p \in \Psi_{H_C}^{(3)}(W) \cap \Pi(W)$. We only need $\mathscr{G}$ with $L_1 = L_2 = 1$. Let $R_0 > 0$ and assume that the rate vector $(\tilde{R}_0, \tilde{R}_1, \tilde{R}_2) := (R_0 + I(Z \wedge XY), R_1, R_2)$ is contained in $\mathscr{R}_{\text{CM}}(p)$. Further for some $0 < 2\eta < \min\{R_\nu : \nu = 0,1,2, R_\nu > 0\}$ let

$$[\tilde{R}_\nu - \eta]_+ \leq \frac{1}{n}\log(K_\nu L_\nu) \leq [\tilde{R}_\nu - \frac{\eta}{2}]_+.$$

$\mathscr{G}$ satisfies $A_{\text{MAC}}$ with probability exponentially close to 1, so the $X_{k_0k_1}^{l_0l_1}$ and $Y_{k_0k_2}^{l_0l_2}$ form the codewords of a deterministic $n$-$\text{code}_{\text{CM}}$ $\gamma$ with $\bar{e}^{\text{DM}}(\gamma, W_b) \leq \exp(-n\zeta_1)$ for some $\zeta_1 > 0$. Now let

$$I(Z \wedge XY) + 2(f_1(\delta) + f_2(\delta)) \leq \frac{1}{n}\log L_0 \leq I(Z \wedge XY) + 3(f_1(\delta) + f_2(\delta))$$

for $\delta$ so small that $4(f_1(\delta) + f_2(\delta)) \leq \min(\eta, H_C - I(Z \wedge XY))$ and choose $\varepsilon$ such that

$$-\frac{1}{n}\log\varepsilon = \frac{1}{4}\min\{4\zeta_1, f_1(\delta) + f_2(\delta)\}.$$

Then for every $(k_0, k_1, k_2)$ the corresponding family $\mathscr{F}$ satisfies the conditions (6.57) and (6.58) with probability exponentially close to 1. We can thus choose a realization $\{(\mathbf{u}_{k_0}^{l_0}, \mathbf{x}_{k_0k_1}^{l_0}, \mathbf{y}_{k_0k_2}^{l_0})\}$ which satisfies the conditions of (6.57) and (6.58) and which determines a deterministic $n$-code$_{\mathrm{CM}}$ for DMAC$(W_b)$ with decoder $\varphi$. Now we can define an $(n, H_C)$-code$_{\mathrm{WCM}}$ whose decoder is $\varphi$ and whose stochastic encoder $G$ is given by

$$\begin{aligned} G_0(l_0|k_0) &= \frac{1}{L_0}, && (k_0 \in [K_0], l_0 \in [L_0]), \\ G_1(\mathbf{x}|k_0, k_1, l_0) &= \delta_{\mathbf{x}_{k_0k_1}^{l_0}}(\mathbf{x}), && (\mathbf{x} \in \mathscr{X}^n, k_1 \in [K_1], k_0 \in [K_0], l_0 \in [L_0]), \\ G_2(\mathbf{y}|k_0, k_2, l_0) &= \delta_{\mathbf{y}_{k_0k_2}^{l_0}}(\mathbf{y}), && (\mathbf{y} \in \mathscr{Y}^n, k_2 \in [K_2], k_0 \in [K_0], l_0 \in [L_0]). \end{aligned}$$

Note that $G_0$ satisfies the common randomness constraint. Due to the uniform distribution of $G_0$, we have $\bar{e}^{\mathrm{WT}}(G, \varphi, W) = \bar{e}^{\mathrm{DM}}(\gamma, W_b) \le \varepsilon$. We have for $\nu = 0, 1, 2$

$$\frac{1}{n} \log K_\nu \ge R_\nu - 2\eta.$$

due to the choice of $\delta$. Finally if we let $M_\nu$ be uniformly distributed on $[K_\nu]$, then it follows from Lemma 6.27 and (6.21) together with the fact that $\varepsilon$ is exponentially small that the strong secrecy criterion is satisfied. Thus the rate triple $(R_0, R_1, R_2)$, and hence $\mathscr{R}^{(3)}(p)$, is WCM-achievable.

### 6.4.6. Concluding Steps

We can reduce coding for a general $p \in \Psi(W)$ which is the distribution of a random vector $(U, V_1, V_2, X, Y, T, Z)$ to the case treated above by constructing a new wiretap MAC as follows: its input alphabets are $\mathscr{V}_1$ and $\mathscr{V}_2$, its output alphabets still are $\mathscr{T}$ and $\mathscr{Z}$. The transition probability for inputs $(v_1, v_2)$ and outputs $(t, z)$ is given by

$$\tilde{W}(t, z|v_1, v_2) := \sum_{(x,y) \in \mathscr{X} \times \mathscr{Y}} W(t, z|x, y) P_{X|V_1}(x|v_1) P_{Y|V_2}(y|v_2).$$

For this channel we do the same construction as above considering the joint distribution $\tilde{p} \in \Pi(\tilde{W})$ of random variables $(U, V_1, V_2, T, Z)$. In this way we also construct a code$_{\mathrm{WCM}}$ for the original channel $W$ because the additional randomness $P_{V_1V_2|U}$ can be integrated into the stochastic encoders $G_1$ and $G_2$. $G_0$ remains unchanged, so the additional randomness in the encoders does not increase the common randomness needed to do the encoding.

On the other hand, we need to show that the rate regions thus obtained are those appearing in the statement of Theorem 6.10. Note that $\tilde{p}$ is contained in $\Psi^{(0)}(\tilde{W})$ or $\Psi_{H_C}^{(\nu)}(\tilde{W})$ for some $\nu = 1, 2, 3$ if and only if $p$ is contained in the corresponding $\Psi^{(0)}(W)$ or $\Psi_{H_C}^{(\nu)}(W)$. This immediately implies that the rate regions also coincide.

## 6.5. Proof of Theorem 6.11

### 6.5.1. Elementary Rate Regions

As for the wiretap MAC with common message we show that we can write the claimed WCONF-achievable regions as unions of simpler sets whose WCONF-achievability will be shown in the next step.

#### For Case 1:

Define

$$\beta_0^{(1)} := [1 - \frac{C_2}{I(Z \wedge U)}]_+, \qquad \beta_1^{(1)} := \min\{\frac{C_1}{I(Z \wedge U)}, 1\}.$$

We have $\beta_0^{(1)} \leq \beta_1^{(1)}$ because $I(Z \wedge U) < C_1 + C_2$.

**Lemma 6.29.** *For $\beta_0^{(1)} \leq \beta \leq \beta_1^{(1)}$, let $\mathscr{R}_\beta^{(1)}(p, C_1, C_2)$ be the set of those real pairs $(R_1, R_2)$ satisfying*

$$\begin{aligned}
R_1 &\leq I(T \wedge V_1|V_2U) - I(Z \wedge V_1|U) \\
&\quad - [I(Z \wedge V_2|V_1U) - I(T \wedge V_2|V_1U)]_+ - \beta I(Z \wedge U) + C_1, \\
R_2 &\leq I(T \wedge V_2|V_1U) - I(Z \wedge V_2|U) \\
&\quad - [I(Z \wedge V_1|V_2U) - I(T \wedge V_1|V_2U)]_+ - (1-\beta) I(Z \wedge U) + C_2, \\
R_1 + R_2 &\leq \min\big\{I(T \wedge V_1V_2|U) - I(Z \wedge V_1V_2|U) - I(Z \wedge U) + C_1 + C_2, \\
&\qquad I(T \wedge V_1V_2) - I(Z \wedge V_1V_2)\big\}.
\end{aligned}$$

*Then*

$$\mathscr{R}^{(1)}(p, C_1, C_2) = \bigcup_{\beta_0^{(1)} \leq \beta \leq \beta_1^{(1)}} \mathscr{R}_\beta^{(1)}(p, C_1, C_2).$$

Thus it is sufficient to show the WCONF-achievability of $\mathscr{R}_\beta^{(1)}(p, C_1, C_2)$ for every $\beta$. For the proof one uses Lemma 6.13.

#### For Case 2:

Recall the vector $(J_0^{(\alpha)}, J_1^{(\alpha)}, J_2^{(\alpha)})$ defined in (6.68)-(6.70). Define

$$\beta_0^{(2,\alpha)} := [1 - \frac{C_2}{J_0^{(\alpha)}}]_+, \qquad \beta_1^{(2,\alpha)} := \min\{\frac{C_1}{J_0^{(\alpha)}}, 1\}.$$

We show that every $\mathscr{R}_\alpha^{(2)}(p, C_1, C_2)$ with $\alpha_0^{(2)} \le \alpha \le \alpha_1^{(2)}$ can be represented as the union of sets $\mathscr{R}_{\alpha,\beta}^{(2)}(p, C_1, C_2)$ for $\beta_0^{(2,\alpha)} \le \beta \le \beta_1^{(2,\alpha)}$. Define $\mathscr{R}_{\alpha,\beta}^{(2)}(p, C_1, C_2)$ by

$$\begin{aligned}
R_1 &\le I(T \wedge V_1|V_2U) - \alpha I(Z \wedge V_1|V_2U) + C_1 - \beta J_0^{(\alpha)},\\
R_2 &\le I(T \wedge V_2|V_1U) - (1-\alpha) I(Z \wedge V_2|V_1U) + C_2 - (1-\beta) J_0^{(\alpha)},\\
R_1 + R_2 &\le I(T \wedge V_1V_2|U) - \alpha I(Z \wedge V_1|V_2U) - (1-\alpha) I(Z \wedge V_2|V_1U)\\
&\qquad + C_1 + C_2 - J_0^{(\alpha)},\\
R_1 + R_2 &\le I(T \wedge V_1V_2) - I(Z \wedge V_1V_2).
\end{aligned}$$

**Lemma 6.30.** *We have for every* $\alpha \in [\alpha_0^{(2)}, \alpha_1^{(2)}]$

$$\mathscr{R}_\alpha^{(2)}(p, C_1, C_2) = \bigcup_{\beta_0^{(2,\alpha)} \le \beta \le \beta_1^{(2,\alpha)}} \mathscr{R}_{\alpha,\beta}^{(2)}(p, C_1, C_2).$$

This is seen immediately using Lemma 6.13.

**For Case 3:**

Define

$$\beta_0^{(1)} := [1 - \frac{C_2}{I(Z \wedge V_1V_2)}]_+, \qquad \beta_1^{(1)} := \min\{\frac{C_1}{I(Z \wedge V_1V_2)}, 1\}.$$

We have $\beta_0^{(1)} \le \beta_1^{(1)}$ because $I(Z \wedge V_1V_2) < C_1 + C_2$.

**Lemma 6.31.** *For* $\beta_0^{(3)} \le \beta \le \beta_1^{(3)}$, *let* $\mathscr{R}_\beta^{(3)}(p, C_1, C_2)$ *be the set of those real pairs* $(R_1, R_2)$ *satisfying*

$$\begin{aligned}
R_1 &\le I(T \wedge V_1|V_2U_0) + C_1 - \beta I(Z \wedge V_1V_2),\\
R_2 &\le I(T \wedge V_2|V_1U_0) + C_2 - (1-\beta) I(Z \wedge V_1V_2),\\
R_1 + R_2 &\le \min\big\{I(T \wedge V_1V_2|U) + C_1 + C_2 - I(Z \wedge V_1V_2),\\
&\qquad I(T \wedge V_1V_2) - I(Z \wedge V_1V_2)\big\}.
\end{aligned}$$

*Then*

$$\mathscr{R}^{(1)}(p, C_1, C_2) = \bigcup_{\beta_0^{(1)} \le \beta \le \beta_1^{(1)}} \mathscr{R}_\beta^{(1)}(p, C_1, C_2).$$

Thus it is sufficient to show the WCONF-achievability of $\mathscr{R}_\beta^{(3)}(p, C_1, C_2)$ for every $\beta$. For the proof one uses Lemma 6.13.

### 6.5.2. Coding

Let $C_1, C_2 > 0$ and let $p \in \Psi_{C_1+C_2}(W)$. Further let $(R_1, R_2) \in \mathscr{R}(p, C_1, C_2)$. In Case 1 we then know that there is a $\beta \in [\beta_0^{(1)}, \beta_1^{(1)}]$ such that $(R_1, R_2) \in \mathscr{R}_\beta^{(1)}(p, C_1, C_2)$, in Case

2 we have an $\alpha \in [\alpha_0^{(2)}, \alpha_1^{(2)}]$ and a $\beta \in [\beta_0^{(2,\alpha)}, \beta_1^{(2,\alpha)}]$ with $(R_1, R_2) \in \mathscr{R}_{\alpha,\beta}^{(2)}(p, C_1, C_2)$. For Case 3, there is a $\beta \in [\beta_0^{(3)}, \beta_1^{(3)}]$ with $(R_1, R_2) \in \mathscr{R}_{\beta}^{(3)}(p, C_1, C_2)$. Recall the notation

$$J_0^{(\alpha)} = \begin{cases} I(Z \wedge U) & \text{in Case 1,} \\ \alpha I(Z \wedge V_2 U) + (1-\alpha) I(Z \wedge V_1 U) & \text{in Case 2,} \\ I(Z \wedge V_1 V_2) & \text{in Case 3.} \end{cases}$$

We set

$$\tilde{R}_0^{(1)} := R_1 \wedge (C_1 - \beta J_0^{(\alpha)}), \quad \tilde{R}_0^{(2)} := R_2 \wedge (C_2 - (1-\beta) J_0^{(\alpha)})$$

and

$$\tilde{R}_\nu := R_\nu - \tilde{R}_0^{(\nu)} \qquad (\nu = 1, 2).$$

Then setting

$$\tilde{R}_0 := \tilde{R}_0^{(1)} + \tilde{R}_0^{(2)},$$

we conclude that $(\tilde{R}_0, \tilde{R}_1, \tilde{R}_2) \in \mathscr{R}_\alpha^{(\nu)}(p)$ in Case $\nu \in \{1, 2, 3\}$. In particular, $(\tilde{R}_0, \tilde{R}_1, \tilde{R}_2)$ is WCM-achievable by the wiretap MAC $W$ with common message under the common randomness bound $C_1 + C_2$. That means that for any $\eta, \varepsilon > 0$ and for sufficiently large $n$, there is an $(n, C_1 + C_2)$-code$_{\mathrm{WCM}}$ $(\tilde{G}, \varphi)$ with codelength triple $(\tilde{K}_0, \tilde{K}_1, \tilde{K}_2)$. The proof of Theorem 6.10 shows that we may assume that $\tilde{G}$ has the form

$$\tilde{G}(\mathbf{x}, \mathbf{y} | \tilde{k}_0, \tilde{k}_1, \tilde{k}_2) = \frac{1}{\tilde{L}_0} \sum_{l_0=1}^{\tilde{L}_0} \tilde{G}_1(\mathbf{x} | \tilde{k}_0, \tilde{k}_1, l_0) \tilde{G}_2(\mathbf{y} | \tilde{k}_0, \tilde{k}_2, l_0)$$

for two stochastic matrices $\tilde{G}_1, \tilde{G}_2$. For $\tilde{L}_0$ we have the bounds

$$J_0^{(\alpha)} + \frac{\eta}{4} \le \frac{1}{n} \log \tilde{L}_0 \le J_0^{(\alpha)} + \frac{\eta}{2}.$$

Without loss of generality we may additionally assume that $\tilde{L}_0^{(1)} := \tilde{L}_0^{\beta}$ and $\tilde{L}_0^{(2)} := \tilde{L}_0^{(1-\beta)}$ are integers. If $0 < 2\eta < \min\{\tilde{R}_\nu : \nu = 0, 1, 2, \tilde{R}_\nu > 0\}$, the codelength triple $(\tilde{K}_0, \tilde{K}_1, \tilde{K}_2)$ may be assumed to satisfy

$$[\tilde{R}_\nu - 2\eta]_+ \le \frac{1}{n} \log \tilde{K}_\nu \le [\tilde{R}_\nu - \eta]_+, \qquad (\nu = 0, 1, 2), \tag{6.71}$$

and both $\bar{e}^{\mathrm{WT}}(\tilde{G}, \varphi, W)$ as well as $I(\tilde{M}_0 \tilde{M}_1 \tilde{M}_2 \wedge Z^n)$ are upper-bounded by $\varepsilon$, where $(\tilde{M}_0, \tilde{M}_1, \tilde{M}_2)$ is distributed uniformly on $[\tilde{K}_0] \times [\tilde{K}_1] \times [\tilde{K}_2]$ and $Z^n$ is Eve's corresponding output random variable. The definitions imply that

$$\frac{1}{n} \log \tilde{K}_0 \tilde{L}_0 \le C_1 + C_2.$$

We can find $\tilde{K}_0', \tilde{K}_0^{(1)}, \tilde{K}_0^{(2)}$ such that $\tilde{K}_0' = \tilde{K}_0^{(1)} \tilde{K}_0^{(2)}$ and $\tilde{K}_0' \le \tilde{K}_0$ and satisfying

$$[\tilde{R}_0^{(\nu)} - 2\eta]_+ \le \frac{1}{n} \log \tilde{K}_0^{(\nu)} \le [\tilde{R}_0^{(\nu)} - \frac{\eta}{2}]_+, \tag{6.72}$$

$$[\tilde{R}_0 - 2\eta]_+ \le \frac{1}{n} \log \tilde{K}_0'. \tag{6.73}$$

Thus one obtains a natural embedding

$$[\tilde{K}_0^{(\nu)}] \times [L_0^{(\nu)}] \subset [\lfloor 2^{nC_\nu} \rfloor] \qquad (\nu = 1, 2). \tag{6.74}$$

We now construct an $(n, C_1, C_2)$-code$_{\mathrm{WCONF}}$. Let

$$K_\nu := \tilde{K}_0^{(\nu)} \tilde{K}_\nu \qquad (\nu = 1, 2).$$

Thus every $k_\nu \in [K_\nu]$ has the form $(a_\nu(k_\nu), b_\nu(k_\nu))$ with $a_\nu(k_\nu) \in [\tilde{K}_0^{(\nu)}]$ and $b_\nu(k_\nu) \in [\tilde{K}_\nu]$. We then define a stochastic one-shot Willems conferencing protocol

$$c_1 : [K_1] \to \mathscr{P}([\lfloor 2^{nC_1} \rfloor]), \qquad c_2 : [K_2] \to \mathscr{P}([\lfloor 2^{nC_2} \rfloor])$$

which is used to generate both a common message as well as common randomness. Given a message $k_\nu \in [K_\nu]$, Alice$_\nu$ chooses an $l_\nu$ uniformly at random from the set $[L_0^{(\nu)}]$ and then maps the pair $(k_\nu, l_\nu)$ to $(a_\nu(k_\nu), l_\nu)$, so $c_\nu(k_\nu, l_\nu) = (a_\nu(k_\nu), l_\nu)$.

Next we define stochastic encoders $G_1, G_2$ as in required for a code$_{\mathrm{WCONF}}$ by setting

$$\mathscr{J} := [\lfloor 2^{nC_1} \rfloor] \times [\lfloor 2^{nC_2} \rfloor]$$

and, using the embedding (6.74),

$$G_1(\mathbf{x}|k_1, j) = \tilde{G}_1(\mathbf{x}|(a_1(k_1), k_0^{(2)}), b_1(k_1), (l_1, l_2))$$

if $j = ((a_1(k_1), l_1), (k_0^{(2)}, l_2))$ and letting $G_1(\mathbf{x}|k_1, j)$ be arbitrary else; $G_2$ is defined analogously. For decoding, one takes the decoder $\varphi$ from the code$_{\mathrm{WCM}}$ $(\tilde{G}, \varphi)$ and lets it combine the messages it receives into elements of $[K_1]$ and $[K_2]$. By (6.71), the numbers $K_1$ and $K_2$ satisfy

$$\frac{1}{n} \log K_1 \geq R_1 - 3\eta,$$
$$\frac{1}{n} \log K_2 \geq R_2 - 3\eta.$$

Thus depending on the case, every rate pair $(R_1, R_2)$ contained in $\mathscr{R}_\beta^{(1)}(p, C_1, C_2)$ or $\mathscr{R}_{\alpha,\beta}^{(2)}(p, C_1, C_2)$ or $\mathscr{R}_\beta^{(3)}(p, C_1, C_2)$ is WCONF-achievable.

## 6.6. Discussion

### 6.6.1. Conferencing and Secret Transmission

This subsection is devoted to the comparison of the wiretap MAC without conferencing nor common randomness and the wiretap MAC if conferencing is allowed. As our focus is on conferencing, we assume that there is no external source of common randomness, i.e. that common randomness can only be established by conferencing. We show that

there exists a wiretap MAC where the only rate pair contained in the region on the left-hand side of (6.9) which is WCONF-achievable without conferencing is $(0,0)$, whereas if conferencing is enabled with arbitrarily small $C_1, C_2 > 0$, then the corresponding WCONF-achievable region contains positive rates. Note that this does not mean that there are cases where conferencing is necessary to establish secret transmission as we do not have a converse for $\overline{\mathscr{C}}_{\mathrm{WCONF}}(W,0,0)$. This restriction limits the use of this discussion and should be kept in mind.

Our goal is to find multiple access channels $W_b$ and $W_e$ such that for every $W$ with these marginals and $p = P_{V_1V_2XYTZ} \in \Psi(W)$ (i.e. with constant $U$, meaning that $V_1, V_2$ are independent) one has

$$I(T \wedge V_1V_2) \leq I(Z \wedge V_1V_2). \tag{6.75}$$

We noted in Remark 6.8 that the left-hand side of (6.9) is WCONF-achievable without conferencing and it is easy to see that condition (6.75) is an equivalent condition for this region to equal $\{(0,0)\}$. At the same time, there should be a $p = P_{UXYTZ} \in \Pi(W)$ for the same $W$ as above such that

$$I(T \wedge XY) > I(Z \wedge XY).$$

This would prove the existence of a rate pair $(R_1, R_2)$ with positive components for sufficiently large $C_1, C_2 > 0$.

We recall one concept of comparison for single-sender discrete memoryless channels (DMCs) introduced by Körner and Marton [35].

**Definition 6.32.** A DMC $W_e : \mathscr{X} \to \mathscr{P}(\mathscr{Z})$ is *less noisy* than a DMC $W_b : \mathscr{X} \to \mathscr{P}(\mathscr{T})$ if for every Markov chain $(U, X, (T, Z))$ with $P_{T|X} = W_b$ and $P_{Z|X} = W_e$ one has

$$I(Z \wedge U) \geq I(T \wedge U).$$

It was observed by van Dijk [54] that this is nothing but saying that the function

$$P_X \mapsto I(Z \wedge X) - I(T \wedge X), \qquad P_X \in \mathscr{P}(\mathscr{X})$$

is concave. Now we generalize this to the MAC case to obtain an equivalent condition for (6.75). We closely follow van Dijk's proof for the single-sender situation.

**Lemma 6.33.** (6.75) *holds for every Markov chain* $((V_1, V_2), (X, Y), (T, Z))$ *with independent* $V_1, V_2$ *and* $X$ *independent of* $V_2$ *and* $Y$ *independent of* $V_1$ *and* $P_{T|XY} = W_b$ *and* $P_{Z|XY} = W_e$ *if and only if the function*

$$(P_X, P_Y) \mapsto I(Z \wedge XY) - I(T \wedge XY), \qquad X, Y \text{ independent r.v.s on } \mathscr{X} \times \mathscr{Y}$$

*is concave in each of its components.*

*Proof.* Let a Markov chain be given as required in the lemma. One has

$$\begin{aligned} &I(Z \wedge V_1V_2) - I(T \wedge V_1V_2) \\ &= \big(I(Z \wedge XY) - I(T \wedge XY)\big) - \big(I(Z \wedge XY|V_1V_2) - I(T \wedge XY|V_1V_2)\big). \end{aligned} \tag{6.76}$$

Now note that the rightmost bracket equals

$$\sum_{v_1}\sum_{v_2} P_{V_1}(v_1)P_{V_2}(v_2)\big(I(Z\wedge XY|V_1=v_1,V_2=v_2)-I(T\wedge XY|V_1=v_1,V_2=v_2)\big),$$

so it is clear that the nonnegativity of (6.76) is equivalent to the concavity in each component of the function from the lemma statement. □

We now define the channels $W_b$ and $W_e$ which will provide the desired example. Let $N_1, N_2$ be i.i.d. random variables uniformly distributed on $\{0,1\}$. The input alphabets are $\mathscr{X}=\mathscr{Y}=\{0,1\}$. The output alphabet of $W_b$ is $GF(3)$ and the output alphabet of $W_e$ is $\{-2,\dots,3\}$. The outputs $t$ of $W_b$ are given by

$$t = x + y + N_1,$$

those of $W_e$ by

$$z = 2x - 2y + N_2.$$

Let $W$ be any stochastic matrix $W : \mathscr{X}\times\mathscr{Y}\to\mathscr{P}(\mathscr{T}\times\mathscr{Z})$ whose marginals are $W_b$ and $W_e$. The intuition is that in $W_e$, one can exactly determine through the output whether or not the inputs were equal and if they were unequal, which input was 0 and which was 1. For $W_b$, however, there are for every output at least two input possibilities, so it is reasonable that an independent choice of the inputs makes $W_e$ better than $W_b$. However, if one may choose the inputs with some correlation, one may choose the inputs to be equal. Then the output of $W_e$ is only noise, whereas one can still extract some information about the input from $W_b$.

As the entries of the corresponding stochastic matrices of both channels are only $1/2$ or 0, the conditional output entropy is independent of the input distribution and equals 1. Further any pair of independent random variables on $\mathscr{X}$ and $\mathscr{Y}$ is given by parameters $q, r\in[0,1]$ such that

$$\mathbb{P}[X^{(q)}=0]=q, \qquad \mathbb{P}[Y^{(r)}=0]=r.$$

Thus in order to determine whether (6.75) holds, it is enough to consider the function $H(Z^{(q,r)})-H(T^{(q,r)})$ for $T^{(q,r)}, Z^{(q,r)}$ being the outputs of $W_b$ and $W_e$, respectively, corresponding to the pair $(X^{(q)}, Y^{(r)})$. One has

$$\begin{aligned} f_Z(q,r) := H(Z^{(q,r)}) = &-q(1-r)\log(q(1-r)/2)\\ &-(qr+(1-q)(1-r))\log((qr+(1-q)(1-r))/2)\\ &-(1-q)r\log((1-q)r/2)\end{aligned}$$

and

$$\begin{aligned} f_T(q,r) &:= H(T^{(q,r)}) \\ &= -\frac{1}{2}(qr+(1-q)(1-r))\log((qr+(1-q)(1-r))/2) \\ &\quad -\frac{1}{2}(qr+q(1-r)+(1-q)r)\log((qr+q(1-r)+(1-q)r)/2) \\ &\quad -\frac{1}{2}(q(1-r)+(1-q)r+(1-q)(1-r))\cdot \\ &\qquad\qquad \cdot\log((q(1-r)+(1-q)r+(1-q)(1-r))/2). \end{aligned}$$

Both entropies are symmetric in $q$ and $r$ and continuous on $[0,1]^2$ and differentiable on $(0,1)^2$, so by Lemma 6.33 it suffices to find the second derivatives in $q$ of both of them and to compare.

We have

$$\begin{aligned} \frac{\partial f_Z}{\partial q}(q,r) = &-(1-r)\log(q(1-r)/2) \\ &-(2r-1)\log((qr+(1-q)(1-r))/2) \\ &+r\log((1-q)r/2) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial f_T}{\partial q}(q,r) = &-\frac{1}{2}(2r-1)\log((qr+(1-q)(1-r))/2) \\ &-\frac{1}{2}(1-r)\log((qr+q(1-r)+(1-q)r)/2) \\ &+\frac{r}{2}\log((q(1-r)+(1-q)r+(1-q)(1-r))/2). \end{aligned}$$

Thus

$$\frac{\partial^2 f_Z}{\partial q^2}(q,r) = -\frac{1-r}{q} - \frac{(2r-1)^2}{qr+(1-q)(1-r)} - \frac{r}{1-q}$$

and

$$\begin{aligned} \frac{\partial^2 f_T}{\partial q^2}(q,r) = &-\frac{(2r-1)^2}{2(qr+(1-q)(1-r))} \\ &-\frac{(1-r)^2}{2(qr+q(1-r)+(1-q)r)} \\ &-\frac{r^2}{2(q(1-r)+(1-q)r+(1-q)(1-r))}. \end{aligned}$$

After some algebra, it turns out that for $q, r \in (0,1)$,

$$\begin{aligned}\frac{\partial^2 f_Z}{\partial q^2}(q,r) - \frac{\partial^2 f_T}{\partial q^2}(q,r) = &-\frac{1-r}{2q} \cdot \frac{q+2r-qr}{q+r-qr} \\ &-\frac{(2r-1)^2}{2(qr+(1-q)(1-r))} \\ &-\frac{r}{2(1-q)} \cdot \frac{2-r-qr}{1-qr} \\ &< 0.\end{aligned}$$

Thus $f_Z - f_T$ is concave and (6.75) is true for $W_b, W_e$.

Now we show that there exists an input distribution with $I(T \wedge \tilde{X}\tilde{Y}) > I(Z \wedge \tilde{X}\tilde{Y})$. Of course, $\tilde{X}$ and $\tilde{Y}$ cannot be independent any more in this case. Every probability distribution $\tilde{p}$ on $\{0,1\}$ induces a probability distribution $\tilde{p}^2$ on $\{0,1\}^2$ via $\tilde{p}^2(x,x) = \tilde{p}(x)$. Let the pair $(\tilde{X}, \tilde{Y})$ be distributed according to $\tilde{p}^2$. It is immediate from the definition of $W_e$ that $I(Z \wedge \tilde{X}\tilde{Y}) = 0$. On the other hand, $P_T$ can be described by the vector $(1/2)(1, \tilde{p}(0), \tilde{p}(1))$. One sees easily that this is maximized for $\tilde{p}(0) = \tilde{p}(1) = 1/2$, resulting in

$$I(T \wedge \tilde{X}\tilde{Y}) = \frac{1}{2}.$$

As in the proof of Lemma 2.14 we can find a $p = P_{UXYTZ} \in \Pi(W)$ with $P_{XY} = \tilde{p}^2$. Note that $I(Z \wedge U) = 0$, so secret transmission is possible with arbitrarily small conferencing capacities $C_1, C_2 > 0$.

### 6.6.2. Necessity of Time-Sharing in Random Coding

We show here that doing time-sharing during random coding is necessary in our proof of Theorem 6.10. This only serves to justify the effort we had to make in coding using two independent families $\mathscr{G}$ and $\mathscr{G}'$. We concentrate on Case 0 and 1. We have to show that it may happen that $\alpha_0^{(1)} > 0$ or $\alpha_1^{(1)} < 1$. Let $\mathscr{X} = \mathscr{Y} = \mathscr{T} = \mathscr{Z} = \{0,1\}$ and let $W_b, W_e : \{0,1\}^2 \to \mathscr{P}(\{0,1\})$ be defined by

$$W_b = \begin{pmatrix} 0.6178 & 0.3822 \\ 0.0624 & 0.9376 \\ 0.9350 & 0.0650 \\ 0.2353 & 0.7647 \end{pmatrix}, \qquad W_e = \begin{pmatrix} 0.0729 & 0.9271 \\ 0.7264 & 0.2736 \\ 0.3662 & 0.6338 \\ 0.4643 & 0.5357 \end{pmatrix},$$

where the output distribution for the input pair $(x,y)$ is given in row number $2x+y+1$ for each matrix. With $q = 0.6933$ and $r = 0.3151$, let $p = p^{(q)} \otimes p^{(r)} \in \mathscr{P}(\mathscr{X} \times \mathscr{Y})$ be the product measure with the marginals

$$p^{(q)} = (q, 1-q), \qquad p^{(r)} = (r, 1-r).$$

Note that $p \in \Psi^{(0)}(W)$. One obtains the following entropies:

$$\begin{aligned} H(T|XY) &\approx 0.5685, & H(Z|XY) &\approx 0.7851, \\ H(T|X) &\approx 0.8532, & H(Z|X) &\approx 0.9952, \\ H(T|Y) &\approx 0.6251, & H(Z|Y) &\approx 0.8442, \\ H(T) &\approx 0.8866, & H(Z) &\approx 0.9999. \end{aligned}$$

Calculating with the above values returns

$$\begin{aligned} I(T \wedge XY) &= 0.3181, & I(Z \wedge XY) &= 0.2147, \\ I(T \wedge X|Y) &= 0.0566, & I(Z \wedge X|Y) &= 0.0590, \\ I(T \wedge Y|X) &= 0.2847, & I(Z \wedge Y|X) &= 0.2101, \\ & & I(Z \wedge X) &= 0.0047, \\ & & I(Z \wedge Y) &= 0.1557. \end{aligned}$$

Thus the conditions (6.2) and (6.3) are satisfied. If $H_C < \min\{I(Z \wedge X|Y), I(Z \wedge Y|X)\} = 0.0590$, then we can only show that $\mathscr{R}^{(0)}(p)$ or $\mathscr{R}^{(1)}(p)$ is WCM-achievable and might have to use time-sharing during random coding to do so. In fact, this is necessary as

$$I(Z \wedge X|Y) > I(T \wedge X|Y),$$

whereas

$$I(Z \wedge Y|X) < I(T \wedge Y|X).$$

Hence $\alpha_0^{(1)} > 0$, but $\alpha_1^{(1)} = 1$. This example was found by a brute-force search using the computer.

# A. Single-Sender Channels

In this appendix we define notation and collect some results from single-sender single-receiver information theory. First we define the classic discrete memoryless channel. Let a stochastic matrix $H : \mathscr{A} \to \mathscr{P}(\mathscr{B})$ be given which has inputs and output in the finite alphabets $\mathscr{A}$ and $\mathscr{B}$, respectively.

**Definition A.1.** The *Discrete Memoryless Channel (DMC)* $\mathrm{DMC}(H)$ is the channel

$$H^{\otimes n} : \mathscr{A}^n \to \mathscr{P}(\mathscr{B}^n), \quad n = 1, 2, \ldots$$

**Definition A.2.** A *deterministic $n$-code$_{1\mathrm{S}}$* ("1S" for "one-sender") with alphabets $\mathscr{X}$ and $\mathscr{T}$ is a pair of functions

$$f : [L] \to \mathscr{X}^n, \quad \varphi : \mathscr{T}^n \to [L]$$

for some positive integers $L$.

A deterministic $n$-code$_{1\mathrm{S}}$ can alternatively be described as a set

$$\{(\mathbf{a}_l, D_l) : l \in [L]\}, \tag{A.1}$$

where $\mathbf{a}_l \in \mathscr{A}^n$ and the sets $D_l$ are disjoint subsets of $\mathscr{B}^n$.

**Definition A.3.** Let $H : \mathscr{A} \to \mathscr{P}(\mathscr{B})$ be a stochastic matrix and let $(f, \varphi)$ be a deterministic $n$-code$_{1\mathrm{S}}$ given by a family (A.1). Its *DMC-average error* is defined as

$$\frac{1}{L} \sum_l H^{\otimes n}(D_l^c | \mathbf{a}_l),$$

its *maximal error* is defined as

$$\max_l H^{\otimes n}(D_l^c | \mathbf{a}_l).$$

**Definition A.4.** A nonnegative real number $R$ is called a *deterministically achievable rate for* $\mathrm{DMC}(H)$ *under the average (maximal) error criterion* if for every $\lambda \in (0, 1)$ and $\varepsilon > 0$ and $n \geq n_0(\lambda, \varepsilon)$ there is an deterministic $n$-code with average (maximal) error at most $\lambda$ and

$$\frac{1}{n} \log L \geq R - \varepsilon.$$

The maximal achievable rate is called the *deterministic capacity of $H$ under the average (maximal) error criterion* and denoted by $\overline{\mathscr{C}}_{1\mathrm{S}}(H)$ ($\mathscr{C}_{1\mathrm{S}}(H)$).

The formulation of the next theorem is due to Shannon [46] a proof can be found e.g. in [20].

**Theorem A.5** (Shannon)**.** *For a stochastic matrix $H : \mathscr{A} \to \mathscr{P}(B)$, we have*

$$\overline{\mathscr{C}}_{1\mathrm{S}}(H) = \mathscr{C}_{1\mathrm{S}}(H) = \max_{\substack{(A,B):\\ P_{B|A}=H}} I(B \wedge A).$$

*The maximum is over pairs of random variables $(A, B)$ with values in $\mathscr{A} \times \mathscr{B}$. The distribution of $A$ is unrestricted, but $P_{B|A}$ must equal $H$. There exists a strong converse.*

Next we consider Arbitrarily Varying Channels (AVCs). An AVC with input alphabet $\mathscr{A}$ and output alphabet $\mathscr{B}$ is determined by a set $\mathscr{H}$ of stochastic matrices

$$H_s : \mathscr{A} \to \mathscr{P}(\mathscr{B}), \quad s \in \mathscr{S},$$

the set of transition probabilities is

$$H^{\otimes n}(\,\cdot\,|\,\cdot\,|\mathbf{s}) : \mathscr{A}^n \to \mathscr{P}(\mathscr{B}^n), \quad \mathbf{s} \in \mathscr{S}^n,\ n = 1, 2 \ldots,$$

where

$$H^{\otimes n}(\mathbf{b}|\mathbf{a}|\mathbf{s}) = \prod_{m=1}^{n} H_{s_m}(b_m|a_m).$$

Its *AVC-average error* is defined as

$$\bar{e}^{\mathrm{AVC}}(f, \varphi, \mathscr{H}) := \max_{\mathbf{s}\in\mathscr{S}^n} \frac{1}{L} \sum_{\ell} H^{\otimes n}(D_\ell^c|\mathbf{a}_\ell|\mathbf{s}),$$

and a nonnegative real number $R$ is called an *achievable rate for the AVC $\mathscr{W}$ under the average error criterion* if for every $\lambda \in (0, 1)$ and every $\varepsilon > 0$ there is an $n_0 = n_0(\lambda, \varepsilon)$ such that for every $n \geq n_0$ there is a deterministic $n$-code$_{1\mathrm{S}}$ with $\bar{e}^{\mathrm{AVC}}(f, \varphi, \mathscr{W}) \leq \lambda$ and

$$\frac{1}{n} \log L \geq R - \varepsilon.$$

The maximum of the set of achievable rates for the AVC $\mathscr{H}$ exists and is called its *deterministic capacity* and denoted by $\overline{\mathscr{C}}^{\mathrm{AVC}}(\mathscr{H})$. The deterministic coding theorem for the AVC $\mathscr{H}$ exhibits a dichotomy analogous to that claimed in Theorem 5.8 and also depending on whether or not $\mathscr{H}$ is symmetrizable.

**Definition A.6.** The AVC $\mathscr{H}$ is called *symmetrizable* if there is a stochastic matrix $\sigma : \mathscr{A} \to \mathscr{P}(\mathscr{S})$ such that for every $b \in \mathscr{B}$ and $a, a' \in \mathscr{A}$

$$\sum_s H(b|a|s)\sigma(s|a') = \sum_s H(b|a'|s)\sigma(s|a).$$

**Theorem A.7** (Csiszár, Narayan)**.** $\overline{\mathscr{C}}^{\mathrm{AVC}}(\mathscr{H})$ *is positive if and only if $\mathscr{H}$ is not symmetrizable. If $\mathscr{H}$ is symmetrizable, then every code with at least two codewords incurs an average error at least* $1/4$.

# B. Two Proofs

*Proof of Lemma 6.13.* The direction "$\subset$" in (6.11) is obvious. For the other direction, let $(R_0, R_1, R_2) \in \mathscr{K}$. We may assume that for some $0 \le \beta \le 1$,

$$\begin{aligned} R_1 &= r_1 - \beta(\alpha_1 a_1 + (1-\alpha_1)b_1) - (1-\beta)(\alpha_0 a_1 + (1-\alpha_0)b_1) \\ &= r_1 - (\beta\alpha_1 + (1-\beta)\alpha_0)a_1 - (\beta(1-\alpha_1) + (1-\beta)(1-\alpha_0))b_1 \end{aligned}$$

because the claim is obvious for $R_1 \le r_1 - \alpha_1 a_1 - (1-\alpha_1)b_1$. We show that $(R_0, R_1, R_2) \in \mathscr{K}_{\beta\alpha_1+(1-\beta)\alpha_0}$. The $R_1$-bound is satisfied due to our assumption. Further due to the bound on $R_1 + R_2$,

$$\begin{aligned} & R_2 \\ &\le r_{12} - c - r_1 + (\beta\alpha_1 + (1-\beta)\alpha_0)a_1 + (\beta(1-\alpha_1) + (1-\beta)(1-\alpha_0))b_1 \\ &\le r_2 - (\beta\alpha_1 + (1-\beta)\alpha_0)a_2 - (\beta(1-\alpha_1) + (1-\beta)(1-\alpha_0))b_2, \end{aligned}$$

so $R_2$ also satisfies the necessary upper bound. The sum constraints are independent of $\alpha$. Hence all upper bounds in the definition of $\mathscr{K}_{\beta\alpha_1+(1-\beta)\alpha_0}$ are satisfied, and Lemma 6.13 is proved. □

*Proof of Lemma 6.15.* For $\alpha \in [\alpha_0, \alpha_1]$, the set $\mathscr{K}_\alpha$ is contained in the convex hull of $\mathscr{K}_{\alpha_0} \cup \mathscr{K}_{\alpha_1}$. Thus we only have to prove that $\mathscr{K} = \mathrm{conv}(\mathscr{K}_{\alpha_0} \cup \mathscr{K}_{\alpha_1})$. Without loss of generality we assume that $b > a$.

We first prove $\mathrm{conv}(\mathscr{K}_{\alpha_0} \cup \mathscr{K}_{\alpha_1}) \subset \mathscr{K}$. Let $(R_0, R_1, R_2) \in \mathrm{conv}(\mathscr{K}_{\alpha_0} \cup \mathscr{K}_{\alpha_1})$. Using the convexity of $\mathscr{K}_{\alpha_0}$ and $\mathscr{K}_{\alpha_1}$ we infer that there is a $(R_0^{(0)}, R_1^{(0)}, R_2^{(0)}) \in \mathscr{K}_{\alpha_0}$ and a $(R_0^{(1)}, R_1^{(1)}, R_2^{(1)}) \in \mathscr{K}_{\alpha_1}$ and a $\beta \in [0,1]$ such that

$$(R_0, R_1, R_2) = \beta(R_0^{(0)}, R_1^{(0)}, R_2^{(0)}) + (1-\beta)(R_0^{(1)} R_1^{(1)}, R_2^{(1)}).$$

One sees immediately that $(R_0, R_1, R_2)$ satisfies the bounds (6.16)-(6.18) and (6.20). It is sufficient to check that (6.19) is satisfied by the triples $(R_0^{(0)}, R_1^{(0)}, R_2^{(0)})$ and $(R_0^{(1)}, R_1^{(1)}, R_2^{(1)})$. For $(R_0^{(0)}, R_1^{(0)}, R_2^{(0)})$ we assume that

$$R_1^{(0)} = \xi(r_1 - \alpha_0 a)$$

for some $\xi \in [0,1]$. After some calculations this yields

$$\begin{aligned} bR_1^{(0)} + aR_2^{(0)} &\le (b-a)r_1 + ar_{12} - ab - (1-\xi)(b-a)(r_1 - \alpha_0 a) \\ &\le (b-a)r_1 + ar_{12} - ab. \end{aligned}$$

One proceeds analogously for $(R_0^{(1)}, R_1^{(1)}, R_2^{(1)})$.

Next we have to check that $\mathscr{K} \subset \mathrm{conv}(\mathscr{K}_{\alpha_0} \cup \mathscr{K}_{\alpha_1})$. It is sufficient to check whether those points $(R_0, R_1, R_2)$ are contained in $\mathrm{conv}(\mathscr{K}_{\alpha_0} \cup \mathscr{K}_{\alpha_1})$ that satisfy both (6.19) and one of (6.16)-(6.18) with equality. So assume that

$$bR_1 + aR_2 = r_{12}a + r_1(b-a) - ab. \tag{B.1}$$

First we also assume that

$$R_1 + R_2 = r_{12} - \alpha_0 a - (1-\alpha_1)b.$$

Then

$$R_2 = r_{12} - \alpha_0 a - (1-\alpha_1)b - R_1$$

and using (B.1) we obtain

$$R_1 = r_1 - \frac{\alpha_1 b - \alpha_0 a}{b-a} a \leq r_1 - \alpha_1 a.$$

For $R_2$ this gives

$$R_2 = r_{12} - r_1 - \left(\alpha_0 + \frac{\alpha_1 b - \alpha_0 a}{b-a}\right) a - (1-\alpha_1)b \leq r_2 - (1-\alpha_1)b,$$

so $(R_1, R_2) \in \mathscr{K}_{\alpha_1}$.

Now we assume

$$R_1 = r_1 - \alpha_0 a.$$

Then inserting this in (B.1) one obtains

$$R_2 \leq r_2 - (1-\alpha_0)b,$$

so $(R_1, R_2) \in \mathscr{K}_{\alpha_0}$.

Finally for

$$R_2 = r_2 - (1-\alpha_1)b$$

we obtain

$$R_1 \leq r_1 - \alpha_1 a,$$

so $(R_1, R_2) \in \mathscr{K}_{\alpha_1}$. This proves the lemma. □

# C. Publication List

This list collects the author's publications on the topics of this thesis which have appeared in conference proceedings or journals. They are also included in the Bibliography.

- M. Wiese and H. Boche. Strong secrecy for multiple access channels. Aydinian, Harout (ed.) et al., Information theory, combinatorics, and search theory. In memory of Rudolf Ahlswede. Berlin: Springer. Lecture Notes in Computer Science 7777, 71-122, 2013.
- M. Wiese and H. Boche. The arbitrarily varying multiple-access channel with conferencing encoders. In *Proc. 2011 IEEE International Symposium on Information Theory (ISIT 2011)*, pages 993 –997, St. Petersburg, Russia, July/August 2011.
- M. Wiese and H. Boche. An achievable region for the wiretap multiple-access channel with common message. In *Proc. 2012 IEEE International Symposium on Information Theory (ISIT 2012)*, pages 249–253, Cambridge, MA, USA, July 2012.
- M. Wiese and H. Boche. The arbitrarily varying multiple-access channel with conferencing encoders. *IEEE Trans. Inf. Theory*, 59(3):1405-1416, 2013.
- M. Wiese, H. Boche, and I. Bjelaković. The compound MAC with common message and partial channel state information. In *Proc. 2010 Intern. Symp. on Inf. Theory and Applications (ISITA 2010)*, Taichung, Taiwan, 2010.
- M. Wiese, H. Boche, I. Bjelakovic, and V. Jungnickel. Downlink with partially cooperating base stations. In *The 11th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2010)*, Marrakech, Morocco, June 2010.
- M. Wiese, H. Boche, I. Bjelaković, and V. Jungnickel. The compound multiple access channel with partially cooperating encoders. *IEEE Trans. Inf. Theory*, 57(5):3045–3066, 2011.

# List of Symbols

| | |
|---|---|
| $\bar{e}^{\mathrm{AVC}}(f, \varphi, \mathscr{H})$ | average error of $(f, \varphi)$ for transmission over the AVC determined by $\mathscr{H}$, page 124 |
| $\overline{\mathscr{C}}^{\mathrm{AVC}}(\mathscr{H})$ | deterministic capacity of the AVC determined by $\mathscr{H}$ under the average error criterion, page 124 |
| $A^c$ | complement of $A$, page 9 |
| $A_{\|y}$ | $\{x \in \mathscr{X} : (x, y) \in A\}$ for $A \subset \mathscr{X} \times \mathscr{Y}$, page 9 |
| $T_1 \wedge T_2$ | maximal common refinement of $T_1$ and $T_2$, page 45 |
| $X \sim P$ | the distribution of $X$ is $P$, page 9 |
| $[M]$ | the set $\{1, \ldots, M\}$, page 8 |
| $[x]_+$ | $\max\{x, 0\}$, page 8 |
| $\\| \cdot \\|$ | total variation distance, page 9 |
| $\lfloor \cdot \rfloor$ | floor function, page 8 |
| $P \otimes Q$ | product of the probability measures $P$ and $Q$, page 9 |
| $P \otimes W$ | joint distribution with input distribution $P$ and conditional output distribution $W$, page 9 |
| $P^{\otimes n}$ | $n$-fold product of $P$ with itself, page 9 |
| $W^{\otimes n}$ | $n$-fold memoryless extension of $W$, page 9 |
| $1_A$ | indicator function of $A$, page 9 |
| $\mathrm{AV}(\mathscr{W})$ | AV-MAC determined by $\mathscr{W}$, page 58 |
| $\mathrm{closure}(A)$ | closure of $A$, page 8 |
| $\mathrm{conv}(A)$ | convex hull of $A$, page 8 |
| $\mathrm{Cp}(\mathscr{W})$ | compound MAC determined by $\mathscr{W}$, page 24 |
| $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$ | $\mathrm{Cp}(\mathscr{W})$ with CSI partitions $T_1, T_2, R)$, page 25 |
| $(n, C_1, C_2)$-code$_{\mathrm{CONF}}$ | MAC code with conferencing encoders, page 16 |

| | |
|---|---|
| $(n, C_1, C_2)$-code$_{\mathrm{WCM}}$ | wiretap MAC code with conferencing encoders, page 80 |
| $(n, C_1, C_2, T_1, T_2, R)$-code$_{\mathrm{CONF}}$ | deterministic/random MAC code with conferencing encoders with CSI partitions $(T_1, T_2, R)$, page 43 |
| $(n, H_C)$-code$_{\mathrm{WCM}}$ | wiretap MAC code with common message, page 77 |
| $(n, T_1, T_2, R)$-code$_{\mathrm{CM}}$ | deterministic or random MAC code with common message for CSI partitions $(T_1, T_2, R)$, page 26 |
| $n$-code$_{\mathrm{CM}}$ | MAC code with common message, page 12 |
| $n$-code$_{\mathrm{1S}}$ | single-sender code, page 123 |
| $\mathscr{C}_1(\mathscr{W}, T_1, T_2)$ | rate set for common message transmission, page 28 |
| $\mathscr{C}_2(\mathscr{W}, C_1, C_2, \tilde{T}, \tilde{T}_1, \tilde{T}_2)$ | rate set for transmission with conferencing encoders, page 46 |
| $\mathscr{C}_{\mathrm{CM}}^{\mathrm{AV}}(\mathscr{W}, C_1^{\infty}, C_2^{\infty})$ | deterministic capacity region of AV$(\mathscr{W})$ with conferencing encoders under the maximal error criterion, page 59 |
| $\mathscr{C}_{\mathrm{CM}}^{\mathrm{AV,r}}(\mathscr{W}, C_1^{\infty}, C_2^{\infty})$ | random capacity region of AV$(\mathscr{W})$ with conferencing encoders under the maximal error criterion, page 60 |
| $\mathscr{C}_{\mathrm{CM}}^{\mathrm{Cp}}(\mathscr{W}, T_1, T_2, R)$ | deterministic capacity region of Cp$(\mathscr{W}, T_1, T_2, R)$ with common message under the maximal error criterion, page 27 |
| $\mathscr{C}_{\mathrm{CM}}^{\mathrm{Cp,r}}(\mathscr{W}, T_1, T_2, R)$ | random capacity region of Cp$(\mathscr{W}, T_1, T_2, R)$ with common message under the average error criterion, page 27 |
| $\mathscr{C}_{+}^{\mathrm{Cp}}(\mathscr{W}, T_1 \wedge T_2)$ | maximal achievable sum rate for infinite conferencing capacities, page 47 |
| $\mathscr{C}_{\mathrm{CONF}}^{\mathrm{Cp}}(\mathscr{W}, C_1, C_2, T_1, T_2, R)$ | deterministic capacity region of Cp$(\mathscr{W}, T_1, T_2, R)$ with conferencing encoders under the maximal error criterion, page 45 |
| $\mathscr{C}_{\mathrm{CONF}}^{\mathrm{DM}}(W, C_1, C_2)$ | deterministic capacity region of DMAC$(W)$ with conferencing encoders under the maximal error criterion, page 17 |
| $\overline{\mathscr{C}}_{\mathrm{CM}}^{\mathrm{AV}}(\mathscr{W}, C_1^{\infty}, C_2^{\infty})$ | deterministic capacity region of AV$(\mathscr{W})$ with conferencing encoders under the average error criterion, page 59 |
| $\overline{\mathscr{C}}_{\mathrm{CM}}^{\mathrm{AV,r}}(\mathscr{W}, C_1^{\infty}, C_2^{\infty})$ | random capacity region of AV$(\mathscr{W})$ with conferencing encoders under the average error criterion, page 60 |
| $\overline{\mathscr{C}}_{\mathrm{CM}}^{\mathrm{Cp}}(\mathscr{W}, T_1, T_2, R)$ | deterministic capacity region of Cp$(\mathscr{W}, T_1, T_2, R)$ with common message under the average error criterion, page 27 |
| $\overline{\mathscr{C}}_{\mathrm{CM}}^{\mathrm{Cp,r}}(\mathscr{W}, T_1, T_2, R)$ | random capacity region of Cp$(\mathscr{W}, T_1, T_2, R)$ with common message under the average error criterion, page 27 |

$\overline{\mathscr{C}}^{\mathrm{DM}}_{\mathrm{CM}}(W)$ deterministic capacity region of DMAC($W$) with common message under average error criterion, page 13

$\overline{\mathscr{C}}^{\mathrm{DM}}_{\mathrm{CM}}(W)$ deterministic capacity region of DMAC($W$) with common message under maximal error criterion, page 13

$\overline{\mathscr{C}}^{\mathrm{Cp}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, R)$ deterministic capacity region of Cp($\mathscr{W}, T_1, T_2, R$) with conferencing encoders under the average error criterion, page 45

$\overline{\mathscr{C}}^{\mathrm{Cp}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, R)$ random capacity region of Cp($\mathscr{W}, T_1, T_2, R$) with conferencing encoders under the average error criterion, page 45

$\overline{\mathscr{C}}^{\mathrm{Cp}}_{\mathrm{CONF}}(\mathscr{W}, C_1, C_2, T_1, T_2, R)$ random capacity region of Cp($\mathscr{W}, T_1, T_2, R$) with conferencing encoders under the maximal error criterion, page 45

$\overline{\mathscr{C}}^{\mathrm{DM}}_{\mathrm{CONF}}(W, C_1, C_2)$ deterministic capacity region of DMAC($W$) with conferencing encoders under the average error criterion, page 17

$\overline{\mathscr{C}}_{\mathrm{1S}}(H)$ deterministic capacity of DMC($H$) under the average error criterion, page 123

$\overline{\mathscr{C}}_{\mathrm{1S}}(H)$ deterministic capacity of DMC($H$) under the maximal error criterion, page 123

$\overline{\mathscr{C}}_{\mathrm{WCM}}(W, C_1, C_2)$ capacity region of WMAC($W$) with conferencing encoders, page 81

$\overline{\mathscr{C}}_{\mathrm{WCM}}(W, H_C)$ capacity region of WMAC($W$) with common message, page 79

$\overline{\mathscr{C}}^{\mathrm{WT}}_{\mathrm{1S}}(W)$ secrecy capacity of single-sender wiretap channel determined by $W$, page 85

$\delta_x$ Dirac measure with mass on $\mathbf{x}$, page 9

DMAC($W$) discrete memoryless MAC determined by $W$, page 12

DMC($H$) DMC determined by $H$, page 123

$\mathbb{E}$ expectation corresponding to $\mathbb{P}$, page 9

$\mathbb{E}[X; A]$ $E[X1_A]$, page 9

$\exp(x)$ $2^x$, page 9

$\bar{e}(\gamma, W)$ average error of $\gamma$ for transmission over DMAC($W$), page 16

$\bar{e}^{\mathrm{AV}}(\gamma, \mathscr{W})$ average error of $\gamma$ for transmission over AV($\mathscr{W}$), page 58

$\bar{e}^{\mathrm{AV,r}}(G, \mathscr{W})$ average error of $G$ for transmission over AV($\mathscr{W}$), page 59

$\bar{e}^{\mathrm{Cp}}(\gamma, \mathscr{W}, T_1, T_2, R)$ average error of $G$ for transmission over Cp($\mathscr{W}, T_1, T_2, R$), page 27

| | |
|---|---|
| $\bar{e}^{\mathrm{Cp}}(\gamma, \mathscr{W}, T_1, T_2, R)$ | average error of $\gamma$ for transmission over $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$, page 26 |
| $\bar{e}^{\mathrm{DM}}(\gamma, W)$ | average error of $\gamma$ for transmission over the DMAC($W$), page 13 |
| $\bar{e}^{\mathrm{WT}}(c, G_1, G_2, \varphi, W)$ | average error of $(c, G_1, G_2, \varphi)$ for transmission over WMAC($W$), page 81 |
| $\bar{e}^{\mathrm{WT}}(G, \varphi, W)$ | average error of $(G, \varphi)$ for transmission over WMAC($W$), page 78 |
| $e(\gamma, W)$ | maximal error of $\gamma$ for transmission over DMAC($W$), page 16 |
| $e^{\mathrm{AV}}(\gamma, \mathscr{W})$ | maximal error of $\gamma$ for transmission over AV($\mathscr{W}$), page 58 |
| $e^{\mathrm{AV,r}}(G, \mathscr{W})$ | maximal error of $G$ for transmission over AV($\mathscr{W}$), page 59 |
| $e^{\mathrm{Cp}}(\gamma, \mathscr{W}, T_1, T_2, R)$ | maximal error of $G$ for transmission over $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$, page 27 |
| $e^{\mathrm{Cp}}(\gamma, \mathscr{W}, T_1, T_2, R)$ | maximum error of $\gamma$ for transmission over $\mathrm{Cp}(\mathscr{W}, T_1, T_2, R)$, page 26 |
| $e^{\mathrm{DM}}(\gamma, W)$ | maximal error of $\gamma$ for transmission over DMAC($W$), page 13 |
| $\Gamma_{\mathrm{CM}}(n, K_0, K_1, K_2)$ | set of MAC codes with common message, page 12 |
| $\Gamma_{\mathrm{CM}}(n, K_0, K_1, K_2, T_1, T_2, R)$ | set of deterministic MAC codes with conferencing encoders and CSI partitions $(T_1, T_2, R)$, page 26 |
| $\Gamma_{\mathrm{CONF}}(n, K_1, K_2, C_1, C_2)$ | set of MAC codes with conferencing encoders, page 16 |
| $\Gamma_{\mathrm{CONF}}(n, K_1, K_2, T_1, T_2, R)$ | set of deterministic MAC codes with conferencing encoders and CSI partitions $(T_1, T_2, R)$, page 44 |
| $h$ | binary entropy, page 21 |
| $H(X)$ | entropy of $X$, page 9 |
| $H(X\|Y)$ | conditional entropy of $X$ given $Y$, page 9 |
| $I(X \wedge Y)$ | mutual information of $X$ and $Y$, page 9 |
| $I(X \wedge Y\|Z)$ | conditional mutual information of $X$ and $Y$ given $Z$, page 9 |
| $I(X \wedge Y\|z)$ | conditional mutual information of $X$ and $Y$ conditional on $\{Z = z\}$, page 9 |
| $\ln x$ | natural logarithm of $x$, page 9 |
| $\log x$ | logarithm of $x$ to base 2, page 9 |
| $N(x\|\mathbf{x})$ | number of times $x$ appears in $\mathbf{x}$, page 19 |
| $\mathscr{P}(X)$ | probability measures on $\mathscr{X}$, page 9 |

| | |
|---|---|
| $\mathbb{P}$ | underlying probability measure, page 9 |
| $P_X$ | the distribution of $X$, page 9 |
| $P_{X\|Y}$ | the conditional distribution of $X$ given $Y$, page 9 |
| $\pi$ | $(T, T_1, T_2)$-input probability, page 25 |
| $\pi^{\otimes n}$ | $n$-th memoryless extension of input probability $\pi$, page 30 |
| $\Pi(\overline{\mathscr{W}})$ | $\Pi_1(\overline{\mathscr{W}}, \{\mathscr{S}\}, \{\mathscr{S}\})$, page 62 |
| $\Pi(W)$ | a set of joint input-output probabilities for $W$, page 14 |
| $\Pi_1(\mathscr{W}, T_1, T_2)$ | $\Pi_{f_1}(\mathscr{W}, \{\mathscr{S}\}, T_1, T_2)$, page 28 |
| $\Pi_2(\mathscr{W}, \tilde{T}, \tilde{T}_1, \tilde{T}_2)$ | $\Pi_{f_2}(\mathscr{W}, \tilde{T}, \tilde{T}_1, \tilde{T}_2)$, page 45 |
| $\Pi_f(\mathscr{W}, T, T_1, T_2)$ | set of joint probabilities for $\mathscr{W}$ with $f$ mapping to $T \times T_1 \times T_2$, page 25 |
| $\Psi(W)$ | a set of joint input-output probabilities for $W$, page 81 |
| $\Psi^{(0)}(W)$ | a subset of $\Psi(W)$, page 82 |
| $\Psi^{(1)}_{H_C}(W)$ | a subset of $\Psi(W)$, page 82 |
| $\Psi^{(2)}_{H_C}(W)$ | a subset of $\Psi(W)$, page 82 |
| $\Psi^{(3)}_{H_C}(W)$ | a subset of $\Psi(W)$, page 84 |
| $\mathscr{R}^{(0)}(p)$ | elementary rate set for secret common message transmission, page 82 |
| $\mathscr{R}^{(1)}(p)$ | elementary rate set for secret common message transmission, page 82 |
| $\mathscr{R}^{(1)}(p, C_1, C_2)$ | elementary rate set for secret transmission with conferencing encoders, page 85 |
| $\mathscr{R}^{(2)}(p)$ | elementary rate set for secret common message transmission, page 83 |
| $\mathscr{R}^{(2)}(p, C_1, C_2)$ | elementary rate set for secret transmission with conferencing encoders, page 85 |
| $\mathscr{R}^{(3)}(p)$ | elementary rate set for secret common message transmission, page 84 |
| $\mathscr{R}^{(3)}(p, C_1, C_2)$ | elementary rate set for secret transmission with conferencing encoders, page 85 |

| | |
|---|---|
| $\mathscr{R}_{\mathrm{CM}}(p)$ | elementary rate set for common message transmission, page 14 |
| $\mathscr{R}_{\mathrm{CONF}}(p, C_1, C_2)$ | elementary rate set for transmission with conferencing encoders, page 17 |
| $\widehat{\mathscr{R}}_{\mathrm{CM}}(p)$ | elementary rate set for common message transmission, page 28 |
| $\widehat{\mathscr{R}}_{\mathrm{CONF}}(p, C_1, C_2)$ | elementary rate set for transmission with conferencing encoders, page 46 |
| $R$ | CSIR partition of the receiver, page 24 |
| $\mathscr{S}_{\tau_1\tau_2}$ | set of channel states possible under CSIT $(\tau_1, \tau_2)$, page 24 |
| $\mathscr{S}^{\rho}_{\tau_1\tau_2}$ | set of channel states possible under joint CSI $(\tau_1, \tau_2, \rho)$, page 24 |
| $\mathrm{supp}(\vartheta)$ | support of $\vartheta$, page 9 |
| $\mathrm{supp}(X)$ | the support of $X$, page 9 |
| $T_\nu$ | CSIT partition of sender $\nu$, page 24 |
| $T^n_{X,\delta}$ | set of $\delta$-typical sequences, page 19 |
| $T^n_{X\|Y,\delta}(\mathbf{y})$ | set of conditionally $\delta$-typical sequences, page 19 |
| $\mathrm{WMAC}(W)$ | the wiretap MAC determined by $W$, page 77 |
| $W_b, W_e$ | marginals of $\mathrm{WMAC}(W)$, page 77 |

# Bibliography


[1] R. Ahlswede. Certain results in coding theory for compound channels I. In *Proc. Colloquium Inf. Th.*, pages 35–60, Debrecen (Hungary), 1967.

[2] R. Ahlswede. Multi-way communication channels. In *Proceedings of 2nd International Symposium on Information Theory*, pages 23–52, Tsahkadsor, Armenian SSR, 1971. Akadémiai Kiadó, Budapest.

[3] R. Ahlswede. The capacity of a channel with two senders and two receivers. *Ann. Probab.*, 2:805–814, 1974.

[4] R. Ahlswede. Elimination of correlation in random codes for arbitrarily varying channels. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 44:159–175, 1978.

[5] R. Ahlswede. Coloring hypergraphs: A new approach to multi-user source coding–II. *J. Comb. Inform. Syst. Sci.*, 5(3):220–268, 1980.

[6] R. Ahlswede. An elementary proof of the strong converse theorem for the multiple-access channel. *J. Comb. Inf. Syst. Sci.*, 7:216–230, 1982.

[7] R. Ahlswede. Arbitrarily varying channels with states sequence known to the sender. *IEEE Trans. Inf. Theory*, IT-32(5):621–629, 1986.

[8] R. Ahlswede. On concepts of performance parameters for channels. In R. Ahlswede, L. Bumer, N. Cai, H. Aydinian, V. Blinovsky, C. Deppe, and H. Mashurian, editors, *General Theory of Information Transfer and Combinatorics*, volume 4123 of *Lecture Notes in Computer Science*, pages 639–663. Springer Berlin Heidelberg, 2006.

[9] R. Ahlswede and N. Cai. Arbitrarily varying multiple-access channels part I–Ericson's symmetrizability is adequate, Gubner's conjecture is true. *IEEE Trans. Inf. Theory*, 45(2):742–749, 1999.

[10] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Trans. Inf. Theory*, 48(3):569–579, 2002.

[11] I. Bjelaković, H. Boche, and J. Sommerfeld. Capacity results for arbitrarily varying wiretap channels. Accepted for publication in *Lecture Notes in Computer Science*, available online at http://arxiv.org/abs/1209.6325, 2012.

[12] I. Bjelaković, H. Boche, and J. Sommerfeld. Secrecy results for compound wiretap channels. Accepted for publication in *Problems of Information Transmission*, available online at http://arxiv.org/abs/1106.2013, 2012.

[13] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacity of a class of channels. *Ann. Math. Statist.*, 30(4):1229–1241, 1959.

[14] M. R. Bloch and J. N. Laneman. Secrecy from resolvability. Submitted to *IEEE Trans. Inf. Theory*, May 2011.

[15] S. Bross, A. Lapidoth, and M. Wigger. The Gaussian MAC with conferencing encoders. In *Proc. IEEE International Symposium on Information Theory (ISIT 2008)*, pages 2702–2706, July 2008.

[16] S. I. Bross, A. Lapidoth, and M. A. Wigger. Dirty-paper coding for the gaussian multiaccess channel with conferencing. *IEEE Trans. Inf. Theory*, 58(9):5640–5668, 2012.

[17] N. Cai, A. Winter, and R. W. Yeung. Quantum privacy and quantum wiretap channels. *Problems of Information Transmission*, 40(4):318–336, 2004.

[18] I. Csiszár. Almost independence and secrecy capacity. *Problems of Information Transmission*, 32(1):40–47, 1996.

[19] I. Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory*, IT-24(3):339–348, 1978.

[20] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, Cambridge, second edition, 2011.

[21] I. Csiszar and P. Narayan. The capacity of the arbitrarily varying channel revisited: positivity, constraints. *IEEE Trans. Inf. Theory*, 34(2):181–193, mar 1988.

[22] R. Dabora and S. Servetto. Broadcast channels with cooperating decoders. *IEEE Trans. Inf. Theory*, 52(12):5438–5454, 2006.

[23] I. Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Trans. Inf. Theory*, 51(1):44–55, 2005.

[24] H. T. Do, T. J. Oechtering, and M. Skoglund. The gaussian Z-interference channel with rate-constrained conferencing decoders. In *Proc. IEEE International Conference on Communications (ICC)*, Cape Town, South Africa, May 2010.

[25] G. Dueck. Maximal error capacity regions are smaller than average error capacity regions for multi-user channels. *Probl. Control Inform. Theory*, 7:11–19, 1978.

[26] G. Dueck. The strong converse of the coding theorem for the multiple-access channel. *J. Comb. Inf. Syst. Sci.*, 6:187–196, 1981.

[27] E. Ekrem and S. Ulukus. Effects of cooperation on the secrecy of multiple access channels with generalized feedback. In *Proc. Conf. on Inf. Sciences and Systems (CISS)*, pages 791–796, Princeton, NJ, March 2008.

[28] E. Ekrem and S. Ulukus. On the secrecy of multiple access wiretap channel. In *Proc. Allerton Conference*, pages 1014–1021, Allerton House, UIUC, IL, USA, September 2008.

[29] J. Gubner. *Deterministic Codes for Arbitrarily Varying Multiple-Access Channels*. PhD thesis, University of Maryland, 1988.

[30] J. Gubner. On the deterministic-code capacity of the multiple-access arbitrarily varying channel. *IEEE Trans. Inf. Theory*, 36(2):262–275, 1990.

[31] X. He and A. Yener. MIMO wiretap channel with arbitrarily varying eavesdropper channel states. Submitted to *IEEE Trans. Inf. Theory*, available at http://arxiv.org/abs/1007.4801, 2010.

[32] J.-H. Jahn. Coding of arbitrarily varying multiuser channels. *IEEE Trans. Inf. Theory*, 27(2):212–226, 1981.

[33] V. Jungnickel, L. Thiele, T. Wirth, T. Haustein, S. Schiffermuller, A. Forck, S. Wahls, S. Jaeckel, S. Schubert, H. Gabler, C. Juchems, F. Luhn, R. Zavrtak, H. Droste, G. Kadel, W. Kreher, J. Mueller, W. Stoermer, and G. Wannemacher. Coordinated multipoint trials in the downlink. In *GLOBECOM Workshops, 2009 IEEE*, pages 1–7, November/December 2009.

[34] M. Karakayali, G. Foschini, and R. Valenzuela. Network coordination for spectrally efficient communications in cellular systems. *IEEE Wireless Communications*, 13(4):56–61, 2006.

[35] J. Körner and K. Marton. The comparison of two noisy channels. In I. Csiszár and P. Elias, editors, *Topics in Information Theory*, number 16 in Coll. Math. Soc. J. Bolyai. North Holland, Amsterdam, 1977.

[36] Y. Liang and H. V. Poor. Multiple-access channels with confidential messages. *IEEE Trans. Inf. Theory*, 54(3):976–1002, 2008.

[37] Y. Liang, H. V. Poor, and S. Shamai. Information theoretic security. *Found. Trends Commun. Inf. Theory*, 5(4-5):355–580, 2008.

[38] H. J. Liao. *Multiple Access Channels*. PhD thesis, Dept. of Electrical Engineering, University of Hawaii, Honolulu, 1972.

[39] R. Liu, Y. Liang, and H. V. Poor. Fading cognitive multiple-access channels with confidential messages. Submitted to *IEEE Trans. Inf. Theory*, available online at http://arxiv.org/abs/0910.4613, 2009.

[40] R. Liu, I. Marić, R. Yates, and P. Spasojević. The discrete memoryless multiple-access channel with confidential messages. In *Proc. Int. Symp. Inf. Theory*, pages 957–961, Seattle, USA, July 2006.

[41] I. Maric, R. Yates, and G. Kramer. Capacity of interference channels with partial transmitter cooperation. *IEEE Trans. Inf. Theory*, 53(10):3536–3548, 2007.

[42] U. Maurer. The strong secret key rate of discrete random triples. In R. Blahut, editor, *Communication and Cryptography – Two Sides of One Tapestry*, pages 271–285. Kluwer Academic Publishers, 1994.

[43] U. M. Maurer. Secret key agreement by public discussion from common information. *IEEE Trans. Inf. Theory*, 39(3):733–742, 1993.

[44] T. Mayer, H. Jenkac, and J. Hagenauer. Turbo base-station cooperation for intercell interference cancellation. In *IEEE International Conference on Communications (ICC)*, volume 11, pages 4977–4982, June 2006.

[45] C. T. K. Ng, I. Maric, A. J. Goldsmith, S. Shamai (Shitz), and R. D. Yates. Iterative and one-shot conferencing in relay channels. In *Proc. IEEE Information Theory Workshop*, Punta del Este, Uruguay, March 2006.

[46] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423, 623–656, 1948.

[47] O. Simeone, D. Gunduz, H. V. Poor, A. J. Goldsmith, and S. Shamai. Compound multiple-access channels with partial cooperation. *IEEE Trans. Inf. Theory*, 55(6):2425–2441, june 2009.

[48] O. Simeone, O. Somekh, G. Kramer, H. V. Poor, and S. Shamai (Shitz). Three-user gaussian multiple access channel with partially cooperating encoders. In *Proc. Asilomar Conference on Signals, Systems and Computers*, 2008.

[49] O. Simeone and A. Yener. The cognitive multiple access wire-tap channel. In *Proc. Conf. on Inf. Sciences and Systems (CISS)*, Baltimore, NJ, USA, March 2009.

[50] D. Slepian and K. Wolf. A coding theorem for multiple access channels with correlated sources. *Bell System Techn. J.*, 52(7):1037–1076, 1973.

[51] G. Smith and J. Yard. Quantum communication with zero-capacity channels. *Science*, 321(5897):1812–1815, 2008.

[52] X. Tang, R. Liu, P. Spasojević, and H. V. Poor. Multiple acess channels with generalized feedback and confidential messages. In *Proc. Inf. Theory Workshop*, pages 608–613, Lake Tahoe, CA, USA, September 2007.

[53] E. Tekin and A. Yener. The gaussian multiple access wire-tap channel. *IEEE Trans. Inf. Theory*, 54(12):5747–5755, 2008.

[54] M. van Dijk. On a special class of broadcast channels with confidential messages. *IEEE Trans. Inf. Theory*, 43(2):712–714, 1997.

[55] M. Wiese and H. Boche. The arbitrarily varying multiple-access channel with conferencing encoders. In *Proc. 2011 IEEE International Symposium on Information Theory (ISIT 2011)*, pages 993 –997, St. Petersburg, Russia, July/August 2011.

[56] M. Wiese and H. Boche. An achievable region for the wiretap multiple-access channel with common message. In *Proc. 2012 IEEE International Symposium on Information Theory (ISIT 2012)*, pages 249–253, Cambridge, MA, USA, July 2012.

[57] M. Wiese and H. Boche. The arbitrarily varying multiple-access channel with conferencing encoders. *IEEE Trans. Inf. Theory*, 59(3):1405–1416, 2013.

[58] M. Wiese and H. Boche. Strong secrecy for multiple access channels. Aydinian, Harout (ed.) et al., Information theory, combinatorics, and search theory. In memory of Rudolf Ahlswede. Berlin: Springer. Lecture Notes in Computer Science 7777, 71-122, 2013.

[59] M. Wiese, H. Boche, and I. Bjelaković. The compound MAC with common message and partial channel state information. In *Proc. 2010 Intern. Symp. on Inf. Theory and Applications (ISITA 2010)*, Taichung, Taiwan, 2010.

[60] M. Wiese, H. Boche, I. Bjelakovic, and V. Jungnickel. Downlink with partially cooperating base stations. In *The 11th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2010)*, Marrakech, Morocco, June 2010.

[61] M. Wiese, H. Boche, I. Bjelaković, and V. Jungnickel. The compound multiple access channel with partially cooperating encoders. *IEEE Trans. Inf. Theory*, 57(5):3045–3066, 2011.

[62] M. A. Wigger. *Cooperation on the Multiple-Access Channel*. PhD thesis, ETH Zürich, Switzerland, 2008.

[63] F. M. J. Willems. *Informationtheoretical Results for the Discrete Memoryless Multiple Access Channel*. PhD thesis, Katholieke Universiteit Leuven, Belgium, 1982.

[64] F. M. J. Willems. The discrete memoryless multiple access channel with partially cooperating encoders. *IEEE Trans. Inf. Theory*, IT-29(3):441–445, 1983.

[65] J. Wolfowitz. *Coding theorems of information theory*. Ergebnisse der Mathematik und ihrer Grenzgebiete, 31. Springer, 1978.

[66] A. Wyner. Recent results in the Shannon theory. *IEEE Trans. Inf. Theory*, 20(1):2–10, 1974.

[67] A. Wyner. The wire-tap channel. *The Bell System Tech. J.*, 54(8):1355–1387, 1975.

[68] R. F. Wyrembelski, M. Wiese, and H. Boche. Strong secrecy in bidirectional broadcast channels with confidential messages. Accepted for publication in *IEEE Transactions on Information Forensics and Security*, 2012.