Polarization and Channel Ordering: Characterizations and Topological Structures

THÈSE N^{O} 7912 (2017)

PRÉSENTÉE LE 20 OCTOBRE 2017
À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS
LABORATOIRE DE THÉORIE DE L'INFORMATION
PROGRAMME DOCTORAL EN INFORMATIQUE ET COMMUNICATIONS

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Rajai NASSER

acceptée sur proposition du jury:

Dr N. Macris, président du jury
Prof. E. Telatar, directeur de thèse
Prof. C. Nair, rapporteur
Prof. M. Raginsky, rapporteur
Prof. R. Renner, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Suisse
2017

Polarization and Channel Ordering: Characterizations and Topological Structures

Polarization and Channel Ordering: Characterizations and Topological Structures

Rajai Nasser

EPFL - Ecole Polytechnique Fédérale de Lausanne

Thesis presented to the faculty of computer and communication sciences for obtaining the degree of Docteur ès Sciences

Jury:

Dr. Nicolas Macris
President of the jury

Prof. Emre Telatar
Thesis director

Prof. Chandra Nair
Expert

Prof. Maxim Raginsky
Expert

Prof. Renato Renner
Expert

Ecole Polytechnique Fédérale de Lausanne, 2017

To all underpaid workers,

all undercredited researchers,

and everybody who is selflessly working

to make this universe a better place to live in.

In the beginning, there was nothing easy ...

and Arıkan said: "let there be polarization" ...

and there was polarization ...

and every channel was made easier ...

Abstract

Information theory is the field in which we study the fundamental limitations of communication. Shannon proved in 1948 that there exists a maximum rate, called capacity, at which we can reliably communicate information through a given channel. However, Shannon did not provide an explicit construction of a practical capacity-achieving coding scheme. Polar coding, invented by Arıkan, is the first low-complexity coding technique that achieves the capacity of binary-input memoryless symmetric channels. The construction of these codes is based on a phenomenon called polarization. The study of polar codes and their generalization to arbitrary channels is the subject of polarization theory, a subfield of information and coding theories.

This thesis consists of two parts. In the first part, we provide solutions to several open problems in polarization theory. The first open problem that we consider is to determine the binary operations that always lead to polarization when they are used in Arıkan-style constructions. In order to solve this problem, we develop an ergodic theory for binary operations. This theory is used to provide a necessary and sufficient condition that characterizes the polarizing binary operations, both in the single-user and the multiple-access settings. We prove that the exponent of a polarizing binary operation cannot exceed $\frac{1}{2}$. Furthermore, we show that the exponent of an arbitrary quasigroup operation is exactly $\frac{1}{2}$. This implies that quasigroup operations are among the best polarizing binary operations.

One drawback of polarization in the multiple-access setting is that it sometimes induces a loss in the symmetric capacity region of a given multiple-access channel (MAC). An open problem in MAC polarization theory is to determine all the MACs that do not lose any part of their capacity region by polarization. Using Fourier analysis, we solve this problem by providing a single-letter necessary and sufficient condition that characterizes all these MACs in the general setting where we have an arbitrary number of users, and each user uses an arbitrary Abelian group operation on his input alphabet.

We also study the polarization of classical-quantum (cq) channels. The input alphabet is endowed with an arbitrary Abelian group operation, and an Arıkan-style transformation is applied using this operation. We show that as the number of polarization steps becomes large, the synthetic cq-channels polarize to deterministic homomorphism channels that project their input to a quotient group of the input alphabet. This result is used to construct polar codes for arbitrary cq-channels and arbitrary classical-quantum multiple-access channels (cq-MAC).

In the second part of this thesis, we investigate several problems that are related

to three orderings of communication channels: degradedness, input-degradedness, and the Shannon ordering. We provide several characterizations for the input-degradedness and the Shannon ordering.

Two channels are said to be equivalent if they are degraded from each other. Input-equivalence and Shannon-equivalence between channels are similarly defined. We construct and study several topologies on the quotients of the spaces of discrete memoryless channels (DMC) by the equivalence, the input-equivalence and the Shannon-equivalence relations. Finally, we prove the continuity of several channel parameters and operations under various DMC topologies.

Keywords: Polar codes, ergodic theory, quasigroup, multiple-access channels, Fourier transform, classical-quantum channels, channel ordering, input degradedness, Shannon ordering, topology.

Résumé

La Théorie de l'Information est le domaine qui définit les restreintes théoriques sur la communication. En effet, en 1948, Shannon démontre l'existence d'un débit maximal de transmission fiable d'information: la capacité. Cependant, Shannon ne présente pas de construction explicite d'un système de codage pratique permettant d'atteindre celle-ci. Le code polaire, inventé par Arıkan, est le premier de ces codes atteignant la capacité des canaux symmétriques sans-mémoire à entré binaire. L'étude des codes polaires ainsi que leur généralization à des canaux arbitraires constitue ce qu'on nomme la théorie de la polarisation, un sous-domaine des théories des codes et de l'information.

Cette thèse se compose de deux axes. En un premier temps, nous présentons des solutions pour de plusieurs problèmes ouverts en théorie de la polarisation. Le premier de ces problèmes consiste à déterminer les lois de composition internes menant à une polarisation lorsqu'elles font parties de constructions similaires à celle d'Arıkan. Afin de résoudre ce problème, nous développons une théorie ergodique pour les lois de composition internes. Cette théorie nous donne une condition nécessaire et suffisante qui caractérise les lois de composition internes polarisantes dans les deux sytèmes d'accès: simple et multiple.

Toutefois, la polarisation d'un canal à accès multiple (CAM) induit une perte dans la région de capacité symmétrique. Un problème ouvert en théorie de la polarisation des CAMs consiste à déterminer les CAMs pour lesquels la polarisation n'aboutit à la perte d'aucune partie de leurs régions de capacité symmétrique. En utilisant l'analyse de Fourier, nous résolvons ce problème en introduisant une condition nécessaire et suffisante qui caractérise tous ces CAMs dans le cas général; où nous supposons un nombre quelconque d'utilisateurs et chaque utilisateur utilise une loi arbitraire d'un groupe abélien sur l'alphabet d'entrée.

Toujours dans le premier axe, nous étudions aussi la polarisation de canaux classiques quantiques. Dans ce cas, l'alphabet d'entrée est doté d'une loi arbitraire d'un groupe abélien et cette dernière est utilisée pour appliquer une transformation similaire à celle d'Arıkan. Nous démontrons que pour un grand nombre d'étapes de polarisation, les canaux classiques quantiques synthétiques se polarisent en des canaux déterministes qui ne sont que des homomorphismes projetant l'entrée du canal sur un groupe quotient de l'alphabet d'entrée. Nous utilisons ce résultat pour construire des codes polaires pour des canaux classiques quantiques et des canaux classiques quantiques à accès multiples quelconques.

En un deuxième temps, nous investiguons plusieurs problèmes reliés à trois classifications des canaux de communication: dégradation, dégradation d'entrée et la clas-

sification de Shannon. Nous proposons plusieurs caractérisations pour la dégradation d'entrée et la classification de Shannon.

De plus, deux canaux sont équivalents s'ils sont dégradés l'un de l'autre. De façon similaire, nous définissons l'équivalence d'entrée et l'équivalence de Shannon. Nous construisons et nous étudions plusieures topologies sur les quotients des espaces des canaux discrets sans mémoire par les relations d'équivalence, d'équivalence d'entrée et d'équivalence de Shannon. Finalement, nous démontrons la continuité de plusieurs paramètres et opérations des canaux sous divers topologies des quotients des espaces des canaux discrets sans mémoire.

Mot-clés: Codes polaires, théorie érgodique, quasigroupe, canaux à accès multiples, transformation de Fourier, canaux classiques quantiques, classification de canaux, dégradation d'entrée, classification de Shannon, topologie.

Acknowledgments

I am very lucky and privileged for having worked under the supervision of Prof. Emre Telatar. I am grateful for his guidance, support and contributions, without which this work could not have been possible. I am especially thankful for the freedom he gave me to work on whatever I find interesting and exciting. Besides the extremely helpful technical advice, I have learned a lot from Emre on the personal level. Despite being a very intelligent and successful researcher, he is an extremely humble person who never shows off or brags about his accomplishments. I will always be indebted to Emre for all his care, kindness and support, and for everything I learned from him.

I thank my thesis committee members, Prof. Chandra Nair, Prof. Maxim Raginsky, and Prof. Renato Renner, for reading my thesis and for providing helpful comments; and Dr. Nicolas Macris for presiding over the committee. I thank Elie Najm for his help in writing the French abstract of the thesis, and Jean Barbier for his comments on the French abstract.

I am grateful to Prof. Renato Renner for hosting me in his lab at ETH Zürich for one semester. During my visit to Zürich, I was fortunate to collaborate with Dr. Joseph Renes to whom I am thankful for his contributions to Chapter 10 of this thesis. I am also grateful to Prof. Maxim Raginsky for informing me about the work of Blackwell and Le Cam on the comparison of statistical experiments. Without this knowledge, Sections 10.3.3, 10.4.1 and 11.9.3 could not have been possible.

I extend my warmest gratitude to the Lebanese University for granting me the excellence scholarship for graduate studies. I am thankful to all the professors, from whom I learned a lot. Special thanks go to Prof. Amine El-Sahili and Prof. Ayman Mourad for being genuine friends and brothers, and for their support and care during the hard times.

I thank all past and present members in the information processing group (IPG): Prof. Michael Gastpar, Dr. Olivier Lévêque, Dr. Nicolas Macris, Prof. Bixio Rimoldi and Prof. Rüdiger Urbanke for all the interesting discussions that we had; Muriel Bardet and Françoise Behn for their administrative support; Damir Laurenzi for making sure that the computer network is running flawlessly; Alla Merzakreeva and Marc Desgroseilliers for sharing the office with me; Mine Alsan, Andrei Giurgiu, Young-Jun Ko, Karol Kruzelecki, Stefano Rosati, Adrian Tarniceriu and Marc Vuffray for the awesome ski trips; Jean Barbier, Marco Mondelli and Mani Bastani Parizi for the fun times at ISIT and in the lab's cafeteria; and Emmanuel Abbe, Vahid Aref, Saeid Haghighatsoar, Hamed Hassani and Eren Şaşoğlu for insightful discussions.

Many thanks to all my Lebanese friends in Switzerland – Elio Abi Karam, Ghofran Akil, Ali Baajour, Abbas Bazzi, Ali Beydoun, Amer Chamseddine, Farah Charab, Mohamad Dia, Raed El-Hage Ali, Marwa El-Halabi, Rafah El-Khatib, Hiba Faour, Abbass Hammoud, Serj Haddad, Sahar Hanna, Hamza Harkous, Rida Jichi, Ghid Maatouk, Elie Najm and Walaa Wehbi – for all the memorable moments that we have had together during the past five years.

Warm thanks go to my Lebanese friends in Europe – Mohammad Bazzi, Tarek Chehade, Mazen El-Ahmar, Ali Komaty, Bilal Komaty, Mokdad El-Mokdad, Abbas El-Mostrah and Yasser Fadlallah – for their true friendship and for always being there for me.

Finally, I would like to acknowledge and extend my heartfelt gratitude to my family for their love, sacrifice, help and support. I am grateful to my brother Hassan and my sisters Rajaa and Hasnaa for their continuous support and encouragement. I am forever indebted to my father Abdallah for emphasizing the importance of math and science when I was a little kid, and for teaching me the necessity of critical thinking. My deep and sincere gratitude to my mother Zahra for her unlimited and unconditional love and tenderness. This journey would not have been possible if not for my parents, brother and sisters.

Contents

1 Introduction

The digital revolution that the world has witnessed over the past few decades is the result of over a century of technological[1] and theoretical[2] developments. Claude Shannon is credited for laying out the foundations of digitization (at least in the areas of communication and storage) in his seminal paper *"A Mathematical Theory of Communication"* [1]. In his pioneering paper, Shannon formalized the problem of (digital) communication and provided clear answers to a number of questions about what is possible and what is not possible to achieve in communication.

The publication of Shannon's paper established a new field in applied mathematics, known as *information theory*. This field is the study of the fundamental limitations of communication. The channel coding theorem [1] shows that for every communication channel W, there exists a positive number $C(W) \geq 0$ that characterizes the highest rate of information[3] that can be reliably communicated through this channel. More precisely, for every $R < C(W)$ and every $\epsilon > 0$, there exists a channel coding scheme of a rate of at least R and whose probability of error is at most ϵ. Whereas, for every $R > C(W)$ there exists $\epsilon_{R,W} > 0$ such that every coding scheme of rate of at least R has a probability of error of at least $\epsilon_{R,W}$. $C(W)$ is called *the capacity of the channel* W.

The channel coding theorem means that the probability of error can be made arbitrarily small if and only if we communicate at a rate that is below the capacity of the channel. In order to show the existence of good codes for rates below capacity, Shannon used a non-constructive proof. Information and coding theorists needed sixty years to find an explicit construction of low-complexity capacity-achieving codes. This was possible due to the discovery of channel polarization by Arıkan [2]

[1] Technological advances that lead to the digital revolution include: the telegraph and Babbage's analytical engine (19th century), transistors (1947), microprocessors (late 1960s), digital mobile phones (1990s) and the internet.

[2] Theoretical advances that contributed to the digital revolution include: the sampling theorem, Turing's foundation of computer science (1936), and Shannon's foundation of communication and information theory (1948).

[3] The rate of information that is communicated through a channel is the average number of bits that is transmitted per channel use. The rigorous definition can be found in Section 1.1.

in 2008.

In this thesis, we provide answers to several questions in two areas of information theory: polarization and channel ordering[4]. In Section 1.1, we provide a brief description of the communication problem. The main purpose of Section 1.1 is to make this thesis accessible to readers who are not familiar with information theory. Readers already familiar with information theory may skip ahead to Section 1.2 where we discuss channel polarization and the construction of polar codes. We explain the channel orderings that we studied in this thesis in Section 1.3. We summarize the contributions of this thesis in Section 1.4.

1.1 The Communication Problem

Imagine that there is a *source of information*[5] that produces a sequence of symbols $U_1, \ldots, U_n, \ldots$ that take values in a set $\mathcal{U}$ that we call *the source alphabet*. Shannon modelled the source as a sequence of random variables[6] $(U_n)_{n\geq 1}$ taking values in $\mathcal{U}$. The probability distribution of the sequence $(U_n)_{n\geq 1}$ is assumed to be known.

A party has access to the source and wants to communicate the symbols $(U_n)_{n\geq 1}$ with another party. The former party is called a transmitter and the latter is called a receiver[7]. In order to achieve this communication, the transmitter and the receiver use *a channel*, which is a physical medium that they share. The channel can be a piece of paper, a magnetic tape, an electrical wire, an optical fiber, radio waves, or any other physical medium. We can think of the channel as a black box that takes symbols from the transmitter and produces symbols that are observed by the receiver. The symbols produced at the receiver's side depend on the symbols that were transmitted in a stochastic way. The set $\mathcal{X}$ of symbols that the transmitter can send is called the *input alphabet* of the channel, and the set $\mathcal{Y}$ of symbols that the receiver can observe is called the *output alphabet* of the channel.

For example, consider the case of an electrical wire. By using some electronic device, the transmitter can control the voltage at one end of the wire; and the receiver can measure (using another electronic device) the voltage at the other end of the wire. Assume that the transmitter's device can only produce voltages that are between $-V$ and V, and assume that the receiver's device can only read voltages that are between $-2V$ and $2V$. In this case, the input alphabet is the interval $[-V, V]$ and the output alphabet is the interval $[-2V, 2V]$. In practice, the output cannot be perfectly predicted from the input due to the interference with the ambient electromagnetic noise and due to the imperfections of the electronic devices. Therefore, for all practical purposes, we can assume that the output depends on the input in a stochastic way.

[4] A channel ordering is a partial order on the set of communication channels.

[5] The source can be an image, a video, a sound wave, the text of a book, the speech of a senator, the temperature measurements in a room, etc ...

[6] Even if the symbols $(U_n)_{n\geq 1}$ are generated according to a deterministic procedure, we do not usually have all the details of the generating procedure. Therefore, for all practical purposes, we can assume that $(U_n)_{n\geq 1}$ is a sequence of random variables following a probability distribution that we can measure by collecting data and studying their statistics.

[7] The transmitter and the receiver can be the same party but at two different instants of time, e.g., storage can be seen as a communication between a person and his older self.

One simple model of such a channel is "the additive noise" model: If $X \in [-V, V]$ is the input that is determined by the transmitter and if $Y \in [-2V, 2V]$ is the output that is observed by the receiver, we can model the relation between X and Y as follows:

$$Y = X + Z,$$

where Z is a random variable that might depend on X. Z represents the *random* noise that is added by the channel to the input.

In general, the channel is described by specifying the input alphabet $\mathcal{X}$, the output alphabet $\mathcal{Y}$, and the probabilistic relation between the input and the output, i.e., for every $x \in \mathcal{X}$, we have to specify a probability distribution $P_{Y|x}$ on the output alphabet $\mathcal{Y}$. Note that for every $y \in \mathcal{Y}$, $P_{Y|x}(y)$ represents the conditional probability of observing y at the output, given that x was the input. In the rest of this thesis, we consider only channels with finite input and output alphabets.

Formally, we can define a channel W as a 3-tuple $(\mathcal{X}, \mathcal{Y}, p_W)$, where $\mathcal{X}$ and $\mathcal{Y}$ are two finite sets that represent the input and output alphabets respectively, and $p_W : \mathcal{X} \times \mathcal{Y} \to [0, 1]$ is a mapping that satisfies $\sum_{y \in \mathcal{Y}} p_W(x, y) = 1$ for all $x \in \mathcal{X}$. For every $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we denote $p_W(x, y)$ as $W(y|x)$ and we interpret it as the conditional probability of receiving y at the output of the channel given that x was the input. We write $W : \mathcal{X} \longrightarrow \mathcal{Y}$ to denote that W is a channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. Note that we use the long arrow $(\longrightarrow)$ in the notation $W : \mathcal{X} \longrightarrow \mathcal{Y}$ and not the short arrow $(\to)$ that we only use to describe mappings. For example, $W : \mathcal{X} \longrightarrow \mathcal{Y}$ denotes a channel, and $V : \mathcal{X} \to \mathcal{Y}$ denotes a mapping from $\mathcal{X}$ to $\mathcal{Y}$.

Example 1.1. *The binary symmetric channel with crossover probability ϵ is the channel $W : \mathcal{X} \longrightarrow \mathcal{Y}$ satisfying $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, $W(0|0) = W(1|1) = 1 - \epsilon$ and $W(1|0) = W(0|1) = \epsilon$. In other words, there is a probability of ϵ that the input bit will be flipped by the channel, and there is a probability of $1 - \epsilon$ that the input bit will remain intact. This channel is denoted as* $\mathrm{BSC}(\epsilon)$.

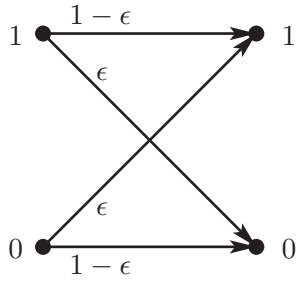

Figure 1.1 – Binary symmetric channel $\mathrm{BSC}(\epsilon)$.

The binary erasure channel with erasure probability ϵ is the channel $W : \mathcal{X} \longrightarrow \mathcal{Y}$ satisfying $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{0, 1, ?\}$, $W(0|0) = W(1|1) = 1 - \epsilon$ and $W(?|0) = W(?|1) = \epsilon$. If we observe 0 (respectively 1) at the output, then we are certain that the transmitted symbol was 0 (respectively 1). Whereas, if we observe the symbol ? at the output, then there is an equal probability that the transmitted symbol was 0 or 1 (we say that the transmitted bit was erased). This channel is denoted as $\mathrm{BEC}(\epsilon)$.

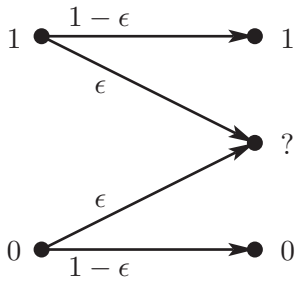

Figure 1.2 – Binary erasure channel BEC(ϵ).

As the source alphabet $\mathcal{U}$ might be different from the input alphabet $\mathcal{X}$ of the channel W, the transmitter has to transform the sequence $(U_n)_{n\geq 1}$ to a sequence of symbols chosen from $\mathcal{X}$ in order to be able to transmit over the channel W. On the other end, the receiver observes a sequence of symbols in $\mathcal{Y}$, and using these observations, the receiver has to estimate the sequence $(U_n)_{n\geq 1}$.

Now we are ready to mathematically formulate the communication problem: A communication scheme for transmitting the source symbols $(U_n)_{n\geq 1}$ through the channel W is a 4-tuple (K, N, f, g), where K and N are two positive integers, $f : \mathcal{U}^K \to \mathcal{X}^N$ is the transmitter's encoder, and $g : \mathcal{Y}^N \to \mathcal{U}^K$ is the receiver's decoder. The communication scheme is implemented as follows:

- The transmitter observes K source symbols $U_1, \ldots, U_K$.
- The transmitter computes $(X_1, \ldots, X_N) = f(U_1, \ldots, U_K)$.
- The transmitter sends the symbols $X_1, \ldots, X_N$ to the receiver by using the channel N times[8].
- The receiver observes the output of the channel W and receives N output symbols $Y_1, \ldots, Y_n$.
- The receiver computes $(\hat{U}_1, \ldots, \hat{U}_K) = g(Y_1, \ldots, Y_N)$.

This procedure can be repeated as many times as needed in order to transmit the subsequent source symbols.

The performance of the communication scheme can be assessed according to various performance parameters:

- The *speed of transmission*:
$$S = \frac{K}{N}.$$
S is the average number of source symbols that are transmitted per channel use. A higher speed corresponds to a more efficient use of the channel.

[8]We assume that the channel W is *memoryless*, in the sense that different uses of the channel are statistically independent. More precisely, for every $x_1, \ldots, x_N \in \mathcal{X}$ and every $y_1, \ldots, y_N \in \mathcal{Y}$, we have
$$P_{Y_1,\ldots,Y_N|X_1,\ldots,X_N}(y_1, \ldots, y_N|x_1, \ldots, x_N) = \prod_{i=1}^{N} W(y_i|x_i).$$

- The *probability of error*:

$$P_e = \mathbb{P}[\{(\hat{U}_1, \ldots, \hat{U}_K) \neq (U_1, \ldots, U_K)\}].$$

  A smaller probability of error corresponds to a more reliable communication scheme.

- The *blocklength* N of the communication scheme. A smaller blocklength corresponds to a smaller delay in the transmission.

- C_e and C_d which are *the computational complexity of the encoder* and *the decoder*, respectively. Obviously, lower computational complexities are preferred.

The study of the trade-off between all these performance parameters is one of the main goals of information theory. In [1], Shannon was interested in specifying the largest possible speed of transmission in a reliable communication scheme, regardless of the blocklength or the computational complexity of the encoder or the decoder.

A speed $S > 0$ is said to be *achievable* if for every $\delta, \epsilon > 0$, there exists a communication scheme of speed of at least $S - \delta$ and of probability of error of at most ϵ. The main question that Shannon answered in [1] was, what is the largest possible achievable speed of transmission?

Shannon solved this problem in two particular cases and then used his two solutions to provide an answer to the general question. The two scenarios that Shannon considered are as follows:

- The noiseless channel case: The distribution of the source is arbitrary but the channel is noiseless, i.e., $\mathcal{X} = \mathcal{Y}$ and $W(y|x) = \mathbb{1}_{\{y=x\}}$ for every $x, y \in \mathcal{X}$.

- The noisy channel with a uniformly distributed source: An arbitrary discrete memoryless channel (DMC) W is considered, but the source symbols are independent and uniformly distributed in $\mathcal{U}$.

1.1.1 The Noiseless Coding Theorem

We consider a source that is *memoryless*[9] in the sense that it produces independent and identically distributed random variables $(U_n)_{n \geq 1}$. We also assume that the channel is binary and noiseless, i.e., the channel can transmit bits without any error. In such a communication scheme, the encoder f transforms the source symbols into a sequence of bits, and the decoder g "reconstructs" the source symbols from the same sequence of bits. A higher speed of transmission $S = \frac{K}{N}$ corresponds to using fewer bits to represent the same number of source symbols.

This procedure is also known as *source coding*, because we are trying to represent the source symbols as efficiently as possible without any concern about the channel. We define the *source code rate* R as the average number of bits per source symbol, i.e.,

$$R = \frac{N}{K} = \frac{1}{S}.$$

We say that $R > 0$ is an achievable source code rate if the speed $\frac{1}{R}$ is achievable. The main question that we are trying to answer can now be reformulated as follows: What is the lowest possible achievable source code rate?

[9] Note that Shannon also studied sources that are not memoryless [1].

Theorem 1.1. *(*The noiseless coding theorem[10] *[1]) Let $(U_n)_{n\geq 1}$ be a sequence of independent and identically distributed random variables that take values in the source alphabet $\mathcal{U}$. The lowest achievable source code rate is equal to*

$$H(U) = -\sum_{u\in\mathcal{U}} P_U(u)\log_2 P_U(u),$$

where U is a random variable that has the same probability distribution as any of the random variables $U_1,\ldots,U_n,\ldots$ We adopt the convention that $0\log_2 0 = 0$.

The quantity $H(U)$ is known as *the entropy*[11] of the random variable U. The noiseless coding theorem (also known as *the source coding theorem*) provides an *operational interpretation* of the entropy of a random variable: It is equal to the lowest average number of bits that we need to describe one instance of the random variable reliably. Intuitively, this can be interpreted by saying that $H(U)$ represents the *amount of information contained in U*.

The entropy of U can also be interpreted as being *the amount of uncertainty* or *the amount of randomness that is contained in U*. This interpretation is reinforced by observing that the entropy is equal to zero when U is deterministic (i.e., no uncertainty nor randomness) and is maximal when U is uniformly distributed (i.e., maximum uncertainty and randomness). This "uncertainty interpretation" might seem to be inconsistent with the previous "information interpretation": How can information and uncertainty represent the same thing?

This apparent inconsistency disappears when we realize that the uncertainty about a random variable *before* observing it is the same as the amount of information that we gain *after* observing it. If there is no uncertainty about the random variable before observation, then we do not learn any new information by observing it[12].

1.1.2 Basic Information Theoretic Quantities

Let (X,Y) be a pair of random variables that might not be independent. Assume that X takes values in $\mathcal{X}$ and Y takes values in $\mathcal{Y}$. The *joint entropy* of X and Y is defined as

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{Y\in\mathcal{Y}} P_{X,Y}(x,y)\log_2 P_{X,Y}(x,y).$$

This is exactly equal to the entropy of the pair (X,Y) when it is seen as one random variable that takes values in $\mathcal{X}\times\mathcal{Y}$. $H(X,Y)$ represents the amount of information that is gained *after* observing both X and Y. The joint entropy of more than two random variables can be defined similarly.

For every $y\in\mathcal{Y}$, define

$$H(X|Y=y) = -\sum_{x\in\mathcal{X}} P_{X|Y}(x|y)\log_2 P_{X|Y}(x|y).$$

[10]The noiseless coding theorem that Shannon proved in [1] considered variable-length source coding. Variable-length source codes have the advantage that they can achieve the entropy without making any errors.

[11]Notice that the entropy is a function of the probability distribution of the random variable.

[12]Formalists might find these arguments informal, unnecessary, confusing and/or meaningless. We reassure the reader that such arguments are never used to prove theorems in information theory (which is as formal and rigorous as any other field of mathematics). These interpretations and arguments are used only to provide intuition.

This is equal to the amount of information that we gain by observing X, assuming that we already know that $Y = y$. The *conditional entropy* of X given Y is defined as

$$H(X|Y) = \sum_{y \in \mathcal{Y}} P_Y(y) H(X|Y = y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{X,Y}(x, y) \log_2 P_{X|Y}(x|y).$$

This is equal to the (average) amount of information that we gain *after* observing X, assuming that we already know the value of Y. $H(X|Y)$ is also equal to the amount of uncertainty about X which remains *after* observing Y (and before observing X).

The *mutual information* between X and Y is defined as

$$I(X;Y) = H(X) - H(X|Y).$$

If $H(X)$ is the amount of uncertainty about X *before* observing it, and $H(X|Y)$ is the amount of uncertainty about X which remains *after* observing Y, then $I(X;Y)$ is the amount of uncertainty about X which is removed by observing Y. Equivalently, $I(X;Y)$ represents the amount of information about X which we can infer from Y.

Now let X, Y and Z be three random variables taking values in $\mathcal{X}, \mathcal{Y}$ and $\mathcal{Z}$, respectively. The *conditional mutual information* between X and Y given Z is defined as

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z).$$

This is equal to the amount of information about X, which we can infer from Y, assuming that we already know Z.

The following properties are well-known [3]:

- If U is a random variable taking values in $\mathcal{U}$ then:
  (a) $0 \leq H(U) \leq \log_2 |\mathcal{U}|$.
  (b) $H(U) = 0$ if and only if U is deterministic.
  (c) $H(U) = \log_2 |\mathcal{U}|$ if and only if U is uniform in $\mathcal{U}$.
- Chain rule for entropy: $H(X,Y) = H(Y) + H(X|Y) = H(X) + H(Y|X)$.
- Conditioning reduces entropy: $H(X|Y) \leq H(X)$.
- $H(X|Y) = 0$ if and only if X can be written as a function of Y.
- $I(X;Y) = I(Y;X) = H(X) + H(Y) - H(X,Y) \geq 0$.
- $I(X;Y) = 0$ if and only if X and Y are independent.
- $I(X;Y|Z) = I(Y;X|Z) \geq 0$.
- Chain rule for mutual information: $I(X;YZ) = I(X;Z) + I(X;Y|Z)$.[13]

[13] $I(X;YZ)$ is the mutual information between X and (Y,Z). A clearer notation that is used for this quantity is $I(X;Y,Z)$. As products of random variables almost never appear in information theory, the notation $I(X;YZ)$ is much more common because it is simpler.

1.1.3 The Noisy-Channel Coding Theorem

The channel coding problem is about reliably communicating a random message through a noisy channel W. The message is assumed to be uniformly distributed in a set $\mathcal{M}$ that is called *the message set*.

A channel coding scheme for a channel $W : \mathcal{X} \longrightarrow \mathcal{Y}$ is a 4-tuple $(\mathcal{M}, N, f, g)$. $\mathcal{M}$ is the *message set*, N is the *blocklength*, $f : \mathcal{M} \to \mathcal{X}^N$ is the *(channel) encoder* and $g : \mathcal{Y}^N \to \mathcal{M}$ is the *(channel) decoder*. The scheme is implemented as follows:

- A random message M is uniformly chosen from $\mathcal{M}$.
- The transmitter computes $(X_1, \ldots, X_N) = f(M)$.
- The transmitter sends $X_1, \ldots, X_N$ to the receiver by using the channel N times.
- The receiver observes N output symbols $Y_1, \ldots, Y_N$.
- The receiver computes an estimate of the transmitted message as

$$\hat{M} = g(Y_1, \ldots, Y_N).$$

The probability of error of the coding scheme $\mathcal{C} = (\mathcal{M}, N, f, g)$ when it is used for the channel W is given by

$$P_e(\mathcal{C}, W) = \mathbb{P}[\hat{M} \neq M].$$

Remark 1.1. *If we have a memoryless source that produces symbols uniformly distributed in $\mathcal{U}$, then a (K, N, f, g) communication scheme can be seen as a $(\mathcal{U}^K, N, f, g)$ channel coding scheme.*

The *rate* of the channel coding scheme $(\mathcal{M}, N, f, g)$ is defined as $R = \frac{\log_2 |\mathcal{M}|}{N}$. This is equal to the number of bits that are transmitted per channel use. A higher rate corresponds to a higher speed of transmission.

A rate $R > 0$ is said to be *achievable* for a channel W if for every $\delta, \epsilon > 0$, there exists a channel coding scheme of rate of at least $R - \delta$ and whose probability of error is at most ϵ. The highest achievable rate is called *the capacity* of the channel W, and we denote it as $C(W)$.

Theorem 1.2. *(The noisy-channel coding theorem [1]) Let $W : \mathcal{X} \longrightarrow \mathcal{Y}$ be a discrete memoryless channel. The capacity of W is given by the following formula:*

$$C(W) = \sup_{P_X \in \Delta_{\mathcal{X}}} I(X;Y),$$

where $\Delta_{\mathcal{X}}$ is the set of probability distributions on $\mathcal{X}$, X is a random variable in $\mathcal{X}$ which is distributed as P_X, and Y is the output of the channel W when X is the input, i.e., for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we have $P_{X,Y}(x, y) = P_X(x)W(y|x)$.

The above characterization of the channel capacity is consistent with the intuitive interpretation of mutual information: If $I(X;Y)$ is the amount of information about X which we can infer from Y, then $\sup_{P_X \in \Delta_{\mathcal{X}}} I(X;Y)$ is the highest number of bits that can be transmitted through the channel W.

It is easy to see that for every channel W with input alphabet $\mathcal{X}$, we have $0 \leq C(W) \leq \log_2 |\mathcal{X}|$. If $C(W) = 0$, then the output of the channel W is always independent of the input. Whereas, if $C(W) = \log_2 |\mathcal{X}|$, then we can show that the input of W can be written as a function of the output[14]. In other words, if the capacity is maximal, then the channel is perfect; in the sense that we can determine the input from the output without errors.

1.1.4 Solution to the Communication Problem

The noiseless coding theorem and the noisy-channel coding theorem provide a solution to the communication problem that was formulated at the beginning of Section 1.1:

- Using the noiseless coding theorem, we can find a good source code whose rate is arbitrarily close to $H(U)$ bits per source symbol.
- Using the noisy channel coding theorem, we can find a good channel coding scheme whose rate is arbitrarily close to $C(W)$ bits per channel use.

By composing the source code with the channel code, we obtain a reliable communication scheme whose speed of transmission is arbitrarily close to $\frac{C(W)}{H(U)}$ source symbols per channel use. Conversely, Shannon showed that it is not possible to achieve a better speed of transmission.

This is known as the source-channel separation theorem: Any achievable speed of transmission can be realized by composing a source code with a channel code. The purpose of the source code is to represent the source symbols with as fewer bits as possible (i.e., combat the redundancy of the source), and the purpose of the channel code is to combat the noise of the channel.

1.2 Channel Polarization

Polar coding, invented by Arıkan [2], is the first low-complexity coding technique that achieves the *symmetric capacity* (defined below) of binary-input memoryless channels. Polar codes rely on a phenomenon that is called *polarization*: The process of converting a set of identical copies of a given binary-input channel into a set of "almost extremal channels", i.e., either "almost perfect channels", or "almost useless channels".

Definition 1.1. *Let $W : \mathcal{X} \longrightarrow \mathcal{Y}$ be a discrete memoryless channel of input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. The* symmetric capacity *of W, denoted as $I(W)$, is the quantity $I(X;Y)$ where X is a uniform random variable in $\mathcal{X}$ and Y is the output of W when X is the input. Clearly, $I(W) \leq C(W)$ for every discrete memoryless channel W.*

[14] Let P_X be the capacity-achieving input distribution. We have

$$\log_2 |\mathcal{X}| = C(W) = I(X;Y) \leq H(X) \leq \log_2 |\mathcal{X}|.$$

This shows that $H(X) = \log_2 |\mathcal{X}|$ (which means that X is uniform) and $H(X|Y) = H(X) - I(X;Y) = 0$, which implies that X can be written as a function of Y.

If a channel W satisfies some symmetry conditions, then the capacity of W can be shown to be equal to $I(W)$. An example of channels that satisfy $C(W) = I(W)$ is the well-known family of *binary-input memoryless symmetric channels*:

Definition 1.2. *Let $\mathbb{F}_2 := \{0,1\}$ be the binary field and let $W : \mathbb{F}_2 \longrightarrow \mathcal{Y}$ be a binary-input channel. We say that W is a* binary-input memoryless symmetric *(BMS) channel if there exists a bijection $\pi : \mathcal{Y} \to \mathcal{Y}$ satisfying $\pi^{-1} = \pi$ and $W(y|0) = W(\pi(y)|1)$ for every $y \in \mathcal{Y}$.* BSC(ϵ) *and* BEC(ϵ) *are examples of BMS channels.*

As $C(W) = I(W)$ for every BMS channel, we can see that polar codes achieve the capacity of all BMS channels.

1.2.1 Polarization of Binary-Input Channels

We start by an informal introduction to the polarization of binary-input channels. Formal and rigorous statements will be provided at the end of this subsection.

We can distinguish, among all binary-input channels, two that are extremal:

- Useless channels where the output is always independent of the input. Such channels satisfy $C(W) = I(W) = 0$.
- Perfect channels where the input can be determined from the output with probability 1. Such channels satisfy $C(W) = I(W) = 1$.

It is very easy to achieve the capacity of extremal channels: In the case of a useless channel, we can transmit a (frozen) bit that is already known to the receiver. Whereas, in the case of a perfect channel, we can transmit an information bit[15] and the receiver can decode it without error.

Now let $W : \mathbb{F}_2 \longrightarrow \mathcal{Y}$ be an arbitrary binary-input channel. If there is a way to transform a collection of independent and identical copies of the channel W into a collection of extremal channels while preserving the total symmetric capacity, then by transmitting frozen bits through the useless channels and information bits through the perfect channels, we can use this procedure to achieve the symmetric capacity. Arıkan proposed a method to do this by applying a basic transformation recursively.

Arıkan's basic transformation is illustrated in Figure 1.3. U_1 and U_2 are two independent and uniformly distributed bits. Let $X_1 = U_1 \oplus U_2$ and $X_2 = U_2$, where $\oplus$ denotes the XOR operation (i.e., addition modulo 2). It is easy to see that X_1 and X_2 are independent and uniform in $\mathbb{F}_2$. We transmit X_1 and X_2 through two independent copies of the channel W. Let Y_1 and Y_2 be the outputs corresponding to X_1 and X_2 respectively.

Consider applying a *successive cancellation decoder* to estimate (U_1, U_2) from (Y_1, Y_2): We first compute an estimate $\hat{U}_1$ of U_1, based on the output (Y_1, Y_2). After that, we compute an estimate $\hat{U}_2$ of U_2, based on $(Y_1, Y_2, \hat{U}_1)$. This procedure motivates us to study the following two *synthetic channels*:

[15] An information bit is a random variable that is uniformly distributed in $\mathbb{F}_2$ and not initially known to the receiver.

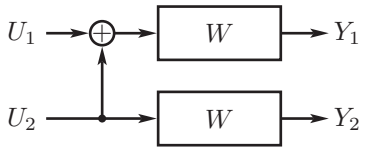

Figure 1.3 – Arıkan's basic transformation.

- The channel W^- whose input is U_1 and whose output is (Y_1, Y_2). U_2 is considered as noise.
- The channel W^+ whose input is U_2 and whose output is (Y_1, Y_2, U_1).

We have:

$$\begin{aligned} I(W^-) + I(W^+) &= I(U_1; Y_1Y_2) + I(U_2; Y_1Y_2U_1) \overset{(a)}{=} I(U_1; Y_1Y_2) + I(U_2; Y_1Y_2|U_1) \\ &= I(U_1U_2; Y_1Y_2) = I(X_1X_2; Y_1Y_2) = I(X_1; Y_1) + I(X_2; Y_2) \\ &= 2I(W), \end{aligned}$$

where (a) follows from the fact that $I(U_1; U_2) = 0$. This shows that the total symmetric capacity is preserved by Arıkan's basic transformation. Furthermore, we have

$$I(W^+) = I(U_2; Y_1Y_2U_1) \geq I(U_2; Y_2) = I(X_2; Y_2) = I(W).$$

This shows that $0 \leq I(W^-) \leq I(W) \leq I(W^+) \leq 1$. In other words, W^- is closer to the useless channel and W^+ is closer to the perfect channel. Therefore, Arıkan's basic transformation makes us closer to the desirable extremal channels. By applying this transformation recursively, we expect that we will get closer and closer to extremal channels. Figure 1.4 shows how we can implement two polarization steps:

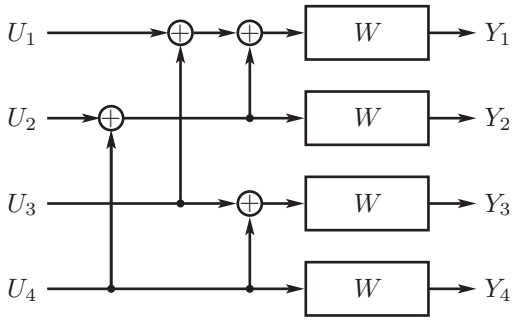

Figure 1.4 – Two polarization steps.

We apply the following successive cancellation decoder:

1. We compute an estimate $\hat{U}_1$ of U_1, based on the observation (Y_1, Y_2, Y_3, Y_4). This corresponds to decoding the synthetic channel whose input is U_1 and whose output is (Y_1, Y_2, Y_3, Y_4). It is easy to see that this is equivalent to the channel $W^{--} := (W^-)^-$.

2. We compute an estimate $\hat{U}_3$ of U_3, based on $(Y_1, Y_2, Y_3, Y_4, \hat{U}_1)$. This corresponds to decoding the synthetic channel whose input is U_3 and whose output is $(Y_1, Y_2, Y_3, Y_4, U_1)$. It is easy to see that this is equivalent to the channel $W^{-+} := (W^-)^+$.

3. We compute an estimate $\hat{U}_2$ of U_2, based on $(Y_1, Y_2, Y_3, Y_4, \hat{U}_1, \hat{U}_3)$. This corresponds to decoding a synthetic channel that is equivalent to $W^{+-} := (W^+)^-$.

4. Finally, we compute an estimate $\hat{U}_4$ of U_4, based on $(Y_1, Y_2, Y_3, Y_4, \hat{U}_1, \hat{U}_2, \hat{U}_3)$. This corresponds to decoding a synthetic channel that is equivalent to $W^{++} := (W^+)^+$.

It is easy to see that after n polarization steps, we obtain 2^n synthetic channels $\left\{W^s : \ s \in \{-,+\}^n\right\}$. Arıkan showed that as n becomes large, almost all the synthetic channels become either very close to a useless channel or very close to a perfect channel. In other words, for the vast majority of $s \in \{-,+\}^n$, we have either $I(W^s) \approx 0$ or $I(W^s) \approx 1$. Let I_G be the set of indices $s \in \{-,+\}^n$ satisfying $I(W^s) \approx 1$.

Polar codes are constructed as follows:

- For each $s \in I_G$, send an information bit over the channel W^s. Hence, we send a total of $|I_G|$ bits.

- For each $s \notin I_G$, send a frozen bit over the channel W^s. A frozen bit is a random symbol that is assumed to be known to the receiver. Hence, no information is being sent over W^s for $s \notin I_G$.

On one hand, as information bits are only sent through channels that are almost perfect, the polar coding scheme is reliable (i.e., it has a low probability of error). On the other hand, as we are sending a total of $|I_G|$ bits over 2^n uses of the channel W, we can see that the rate of the polar coding scheme is equal to $\frac{|I_G|}{2^n}$ bits per channel use.

As Arıkan's basic transformation preserves the total symmetric capacity, we have

$$2^n I(W) = \sum_{s \in \{-,+\}^n} I(W^s).$$

Therefore,

$$I(W) = \frac{1}{2^n} \sum_{s \in \{-,+\}^n} I(W^s) \overset{(a)}{\approx} \frac{1}{2^n} \sum_{s \in I_G} I(W^s) \approx \frac{|I_G|}{2^n},$$

where (a) follows from the fact that for almost all the indices $s \in \{-,+\}^n$, we either have $s \in I_G$ or $I(W^s) \approx 0$. We deduce that the rate of the aforementioned polar coding scheme is close to the symmetric capacity of the channel.

Arıkan showed that all the above approximations become arbitrarily good as n becomes large. This implies that we can construct polar codes with a probability of error that is arbitrarily small and a rate that is arbitrarily close to the symmetric capacity $I(W)$. Furthermore, this can be achieved using an encoder and a decoder of complexity $O(N \log N)$, where $N = 2^n$ is the blocklength of the code (see [2] for details). We conclude that polar codes can achieve the symmetric capacity of any binary-input channel using low-complexity encoder and decoder.

Formal Description of Channel Polarization

Definition 1.3. *Let $W : \mathbb{F}_2 \to \mathcal{Y}$ be a binary-input memoryless channel. We define the two channels $W^- : \mathbb{F}_2 \to \mathcal{Y} \times \mathcal{Y}$ and $W^+ : \mathbb{F}_2 \to \mathcal{Y} \times \mathcal{Y} \times \mathbb{F}_2$ as follows:*

$$W^-(y_1, y_2|u_1) = \frac{1}{2} \sum_{u_2 \in \mathbb{F}_2} W(y_1|u_1 \oplus u_2)W(y_2|u_2),$$

$$W^+(y_1, y_2, u_1|u_2) = \frac{1}{2} W(y_1|u_1 \oplus u_2)W(y_2|u_2).$$

Moreover, for all $s = (s_1, ..., s_n) \in \{-, +\}^n$, we define

$$W^s := ((W^{s_1})^{s_2}...)^{s_n}.$$

Theorem 1.3. *[2] Let $W : \mathbb{F}_2 \to \mathcal{Y}$ be a binary-input memoryless channel. For every $\delta > 0$, we have*

$$\lim_{n\to\infty} \frac{1}{2^n} \big|\{s \in \{-, +\}^n : \ \delta < I(W^s) < 1 - \delta\}\big| = 0.$$

Any construction that is similar to the one given in Definition 1.3 and Figure 1.3 is called *an Arıkan-style construction*. If such construction exhibits a *polarization phenomenon*, then the code obtained by this construction is called *a polar code* (the concepts of "polarization phenomena" and "Arıkan-style constructions" will be formally and rigorously defined in Chapter 3).

1.2.2 Polarization for Arbitrary Discrete Memoryless Channels

Any attempt to generalize Arıkan's technique to channels having a non-binary input alphabet $\mathcal{X}$ has to replace the XOR operation by a binary operation $*$ on the input alphabet $\mathcal{X}$. The first operation that was investigated is the addition modulo q, where $q = |\mathcal{X}|$ and $\mathcal{X}$ is endowed with the algebraic structure $\mathbb{Z}_q$. Şaşoğlu et al. [4] show that if q is prime, then the addition modulo q leads to the same polarization phenomenon as in the binary input case.

Park and Barg [5] show that if $q = 2^r$ with $r > 0$, then the addition modulo q leads to a polarization phenomenon which is different from the polarization in the binary input case, but it can still be used to construct capacity-achieving polar codes. They show that we have a multilevel polarization: Although we do not always have polarization to "almost perfect" or "almost useless" channels, we always have polarization to channels that are easy to use for communication. Sahebi and Pradhan [6] show that multilevel polarization also happens if an arbitrary Abelian group operation on the alphabet $\mathcal{X}$ is used. This enables the construction of polar codes for arbitrary discrete memoryless channels (DMC) since any alphabet can be endowed with an Abelian group structure.

Polar codes for arbitrary DMCs were also constructed by Şaşoğlu [7] using a special quasigroup operation that ensures two-level polarization.

1.2.3 Polarization for Multiple-Access Channels

So far we have only considered discrete memoryless channels. These channels have exactly one transmitter and one receiver. There exists another kind of channels

that allow more than one user to transmit information to a single receiver. Such channels are called multiple-access channels[16] (MAC). The polarization phenomenon can be generalized to MACs: If W is an m-user MAC, we can apply an Arıkan-style construction on W by using a binary operation on the input alphabet of each user.

Şaşoğlu et al. constructed MAC-polar codes for a two-user MAC with an input alphabet of prime size [8]. Abbe and Telatar used matroid theory to construct MAC-polar codes for an m-user MAC with binary inputs [9].

1.3 Channel Ordering

The ordering of communication channels was first introduced by Shannon [10]. A channel W' is said to contain another channel W if W can be simulated from W' by randomization at the input and the output using a shared randomness between the transmitter and the receiver. More precisely, $W' : \mathcal{X}' \longrightarrow \mathcal{Y}'$ contains $W : \mathcal{X} \longrightarrow \mathcal{Y}$ if there exist an integer n and three sequences $(\alpha_l)_{1\leq l\leq n}$, $(T_l)_{1\leq l\leq n}$ and $(R_l)_{1\leq l\leq n}$ such that:

- α_l is a positive number for every $1 \leq l \leq n$, and

$$\sum_{l=1}^{n} \alpha_l = 1.$$

  In other words, $(\alpha_l)_{1\leq l\leq n}$ forms a probability distribution on $\{1, \ldots, n\}$.

- For every $1 \leq l \leq n$, T_l is a channel of input alphabet $\mathcal{X}$ and output alphabet $\mathcal{X}'$.

- For every $1 \leq l \leq n$, R_l is a channel of input alphabet $\mathcal{Y}'$ and output alphabet $\mathcal{Y}$.

- For every $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we have

$$W(y|x) = \sum_{l=1}^{n} \alpha_l \sum_{x'\in\mathcal{X}'} \sum_{y'\in\mathcal{Y}'} T_l(x'|x) W'(y'|x') R_l(y|y').$$

Assume that a transmitter and a receiver share a random variable L taking values in $\{1, \ldots, n\}$, and assume that L is distributed as $(\alpha_l)_{1\leq l\leq n}$. If the transmitter and the receiver have access to the channel W', they can use the random variable L in order to simulate the channel W as follows: In order to transmit a symbol $X \in \mathcal{X}$ through the simulated channel W, the transmitter first observes the random variable L and then applies the random mapping[17] T_L on X. Let $X' \in \mathcal{X}'$ be the (random) output of T_L. The transmitter sends X' through the channel W'. Let $Y' \in \mathcal{Y}'$ be the output of the channel W'. The receiver observes the random variable L and applies the random mapping R_L on Y'. Let Y be the output of R_L. It is easy to see that the channel from X to Y is equivalent to W.

[16] See Chapter 4 for the formal definition of multiple-access channels.

[17] A discrete memoryless channel can be seen as a random mapping from the input alphabet to the output alphabet.

Shannon showed in [10] that if W' contains W, then the existence of a coding scheme of blocklength N, rate R and probability of error ϵ for the channel W implies the existence of a coding scheme of blocklength N, rate R and probability of error of at most ϵ for the channel W'. This shows that $C(W) \leq C(W')$ and $P_e(N, R, W') \leq P_e(N, R, W)$ for every integer N and every positive real number $R > 0$, where $P_e(N, R, W)$ is the smallest probability of error among all coding schemes of blocklength N and of rate of at least R, assuming that the schemes are used for the channel W.

Another ordering that has been well studied is the degradedness between channels. A channel W is said to be degraded from another channel W' if W can be obtained from W' by composing it with another channel. In other words, W is degraded with respect to W' if W can be simulated from W' by a randomization at the output. In Part II of this thesis, we will refer to degradedness as *output-degradedness* in order to distinguish it from the notion of input-degradedness that we introduce in Chapter 10. It is easy to see that output-degradedness is a special case of Shannon's ordering. We can trace the roots of the notion of output-degradedness to the seminal work of Blackwell, in the 1950s, about comparing statistical experiments [11]. Note that in the Shannon ordering, the input and output alphabets need not be the same, whereas in the output-degradedness definition, we have to assume that W and W' share the same input alphabet $\mathcal{X}$ but they can have different output alphabets. A characterization of output-degradedness is given by the famous Blackwell-Sherman-Stein (BSS) theorem [11, 12, 13].

1.4 Outline and Contributions of this Thesis

This thesis consists of two parts. In the first part (Chapters 2–9), we provide solutions to several problems related to channel polarization. We summarize the main results of Part I in Section 1.4.1. Part I is concluded in Chapter 9. In the second part (Chapters 10–13), we investigate several problems related to channel orderings. We present the main results of Part II in Section 1.4.2. Part II is concluded in Chapter 13.

1.4.1 Part I: Channel polarization

An Ergodic Theory for Binary Operations

In Section 1.2.2, we saw that Abelian group operations are polarizing in the sense that they always lead to a (multilevel) polarization phenomenon when they are used in Arıkan-style constructions. An open problem in polarization theory is to characterize all the polarizing binary operations (in the general multilevel sense). Chapters 2 and 3 solve this problem by providing a necessary and sufficient condition for a binary operation to be polarizing. In Chapter 2, we develop an ergodic theory for binary operations. This theory will be used in Chapter 3 to characterize the polarizing operations.

In **Chapter 2**, we define uniformity preserving, irreducible, ergodic and strongly ergodic operations and we study their properties. We introduce the concepts of a stable partition and the residue of a stable partition. We show that an ergodic operation is strongly ergodic if and only if all its stable partitions are their own

residues. We also study the products of binary operations and the structure of their stable partitions. We show that the product of a sequence of binary operations is strongly ergodic if and only if all the operations in the sequence are strongly ergodic.

Polarizing Binary Operations

Let $*$ be a binary operation on a finite set $\mathcal{X}$. We say that $*$ is polarizing if for every discrete memoryless channel W with input alphabet $\mathcal{X}$, the recursive application of the Arıkan-style construction that is based on $*$ transforms a collection of independent and identical copies of W into a collection of "easy channels". In **Chapter 3**, we provide rigorous definitions for the concepts of easy channels and polarizing binary operations. We show that a binary operation is polarizing if and only if it is uniformity preserving and its right-inverse is strongly ergodic.

We define the exponent E_* of a polarizing binary operation $*$. We show that if $*$ is a polarizing operation on a finite set $\mathcal{X}$, then for every channel W with input alphabet $\mathcal{X}$, every $\beta < E_*$ and every $\delta > 0$, there exists $n_0 = n_0(W, \beta, \delta, *) > 0$ such that for every $n \geq n_0$, there exists a polar code of blocklength $N = 2^n$ and of rate of at least $I(W) - \delta$ such that the probability of error of the successive cancellation decoder is at most 2^{-N^β}. In other words, the probability of error of polar codes that are constructed using $*$ decays faster than $2^{-N^{E_* - \epsilon}}$ for any $\epsilon > 0$.

MAC Polarization Theory

Let $\mathcal{X}_1, \ldots, \mathcal{X}_m$ be m finite sets and let $*_1, \ldots, *_m$ be m binary operations defined on $\mathcal{X}_1, \ldots, \mathcal{X}_m$ respectively. We say that the sequence $(*_1, \ldots, *_m)$ is MAC-polarizing if every MAC of input alphabets $\mathcal{X}_1, \ldots, \mathcal{X}_m$ can be polarized by applying an Arıkan-style transformation that is based on the binary operations $*_1, \ldots, *_m$. In **Chapter 4**, we show that a sequence of binary operations is MAC-polarizing if and only if every binary operation in the sequence is uniformity preserving and its right inverse is strongly ergodic.

We define the exponent $E_{*_1,\ldots,*_m}$ of a MAC-polarizing sequence $(*_1, \ldots, *_m)$. We show that if $*_1, \ldots, *_m$ are binary operations on $\mathcal{X}_1, \ldots, \mathcal{X}_m$ respectively, and if $(*_1, \ldots, *_m)$ is MAC-polarizing, then for every MAC W of input alphabets $\mathcal{X}_1, \ldots, \mathcal{X}_m$, every $\beta < E_{*_1,\ldots,*_m}$ and every $\delta > 0$, there exists

$$n_0 = n_0(W, \beta, \delta, *_1, \ldots, *_m) > 0$$

such that for every $n \geq n_0$, there exists a MAC-polar code of blocklength $N = 2^n$ and of sum-rate of at least $I(W) - \delta$ such that the probability of error of the successive cancellation decoder is at most 2^{-N^β}. In other words, the probability of error of MAC-polar codes that are constructed using $*_1, \ldots, *_m$ decays faster than $2^{-N^{E_{*_1,\ldots,*_m} - \epsilon}}$ for any $\epsilon > 0$.

We also show that if we use special binary operations (namely, the addition modulo the size of the input alphabets), the MAC-polar code construction becomes simpler.

Error Exponents

In **Chapter 5**, we study the exponents of polarizing binary operations and the exponents of MAC-polarizing sequences of binary operations. We show that the

exponent of a polarizing binary operation cannot exceed $\frac{1}{2}$. We provide a sufficient condition for a polarizing operation to have a zero exponent. We prove that the exponent of a quasigroup operation is exactly $\frac{1}{2}$. This implies that quasigroup operations are among the best polarizing binary operations.

We show that the exponent of a MAC-polarizing sequence of binary operations is upper bounded by the exponent of the product of all the binary operations that are present in the sequence, which in turn is upper bounded by the exponent of every binary operation in the sequence. Furthermore, we prove that the exponent of a sequence of quasigroup operations is exactly $\frac{1}{2}$.

Fourier Analysis of MAC Polarization

One drawback of MAC-polar codes (i.e., codes that are based on MAC polarization) is that they might not achieve the entire symmetric-capacity region[18]. The reason behind this problem is that MAC polarization sometimes induces a loss in the symmetric-capacity region.

Chapter 6 provides a single-letter necessary and sufficient condition that characterizes the set of MACs that do not lose any part of their symmetric-capacity region by polarization. The characterization that we provide relies on Fourier analysis, and works in the general setting where we have an arbitrary number of users and each user uses an arbitrary Abelian group operation on his input alphabet. We show that the reason why a given MAC W loses parts of its symmetric-capacity region by polarization is because its transition probabilities are not "aligned", which makes W "incompatible" with polarization. The "alignment" condition is expressed in terms of the Fourier transforms of the transition probabilities of W.

Erasure Schemes Using Generalized Polar Codes

One possible way to enhance the performance of polar codes is through decoding with erasure; it is sometimes desirable to allow the receiver not to decide which message was transmitted, especially when there is a feedback from the receiver to the transmitter: If a confusing string of symbols was received (in the sense that there is a high probability of a decoding error to occur, no matter which message the receiver chooses as the decoded message), the receiver can ask the transmitter to resend the message, in the hope that the received string will not be confusing in the next transmission.

There are two types of error when we allow decoding with erasure:

- If the receiver decides on the transmitted message and makes an error, we say that an undetected error occurs.
- If the receiver does not decide, we say that an erasure occurs.

In **Chapter 7**, we study the tradeoff between the probability of undetected error and the erasure probability for generalized polar (GP) codes[19]. We derive a closed-form formula for the zero-undetected-error capacity $I_0^{\mathrm{GP}}(W)$ of GP codes for a given

[18] The definition of the symmetric-capacity region can be found in Chapter 4.

[19] Generalized polar codes are a family of codes that contains, among others, the standard polar codes of Arıkan and Reed-Muller codes.

binary-input memoryless symmetric channel W under the low-complexity successive cancellation decoder with erasure. We show that for every $\epsilon > 0$ and every $R < I_0^{\mathrm{GP}}(W)$, there exists a generalized polar code of blocklength N and of rate of at least R where the undetected-error probability is zero and the erasure probability is less than $2^{-N^{\frac{1}{2}-\epsilon}}$. Conversely, we show that for any $\epsilon > 0$ and any GP code of rate $I_0^{\mathrm{GP}}(W) < R < I(W)$ and blocklength N, the undetected error probability cannot be made less than $2^{-N^{\frac{1}{2}+\epsilon}}$ unless the erasure probability is close to 1.

Polar Codes for Arbitrary Classical-Quantum Channels

The polarization phenomenon can be generalized to the setting where the input of the channel is classical and the output is a quantum state. In **Chapter 8**, we prove polarization theorems for arbitrary classical-quantum channels (cq-channel). The input alphabet is endowed with an arbitrary Abelian group operation and an Arıkan-style transformation is applied using this operation. We show that as the number of polarization steps becomes large, the synthetic cq-channels polarize to deterministic homomorphism cq-channels that project their input to a quotient group of the input alphabet. This result is used to construct polar codes for arbitrary cq-channels and arbitrary classical-quantum multiple-access channels (cq-MAC). The encoder can be implemented in $O(N \log N)$ operations, where N is the blocklength of the code. We propose a quantum successive cancellation decoder for the constructed codes. Furthermore, we show that the probability of error of this decoder decays faster than $2^{-N^{\beta}}$ for any $\beta < \frac{1}{2}$.

1.4.2 Part II: Channel ordering

Characterizations of Various Channel Orderings

In **Chapter 10**, we introduce the input-degradedness as a novel channel ordering. A channel W is said to be input-degraded from another channel W' if W can be simulated from W' by randomization at the input. We provide a necessary and sufficient condition for a channel to be input-degraded from another one. We show that any decoder that is good for W' is also good for W. We provide two characterizations for input-degradedness, one of which is similar to the Blackwell-Sherman-Stein (BSS) theorem.

We also study the Shannon ordering of communication channels in **Chapter 10**. We show that W' contains W (in the Shannon ordering sense) if and only if W is the skew-composition of W' with a convex-product channel. We use this fact to derive a characterization of the Shannon ordering that is similar to the BSS theorem. The characterization that we provide is given in terms of blind randomized in the middle (BRM) games[20].

Topological Structures on DMC Spaces

A topology on a given set is a mathematical structure that enables us to formally talk about the neighborhood of a given point of the set. This makes it possible to define continuous mappings and converging sequences. Topological spaces generalize metric

[20]The definition of BRM games is given in Chapter 10.

spaces which are mathematical structures that specify distances between the points of the space. Links between information theory and topology were investigated in [14].

Two channels are said to be output-equivalent if they are output-degraded from each other. Input-equivalence and Shannon-equivalence between channels are similarly defined. In **Chapter 11**, we construct and study several topologies on the quotients of the spaces of discrete memoryless channels (DMC) by the output-equivalence, the input-equivalence and the Shannon-equivalence relations. In Chapter 12, we show that many channel parameters and operations are continuous under the constructed topologies.

The space of output-equivalent channels with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ can be naturally endowed with the quotient of the Euclidean topology by the output-equivalence relation. We show that this topology is compact, path-connected and metrizable. A topology on the space of output-equivalent channels with fixed input alphabet $\mathcal{X}$ and arbitrary but finite output alphabet is said to be natural if and only if it induces the quotient topology on the subspaces of output-equivalent channels sharing the same output alphabet. We show that every natural topology is σ-compact, separable and path-connected. Whereas, if $|\mathcal{X}| \geq 2$, we prove that a Hausdorff natural topology is not Baire and it is not locally compact anywhere. This implies that no natural topology can be completely metrized if $|\mathcal{X}| \geq 2$. We show that the finest natural topology, which we call the strong topology, is compactly generated, sequential and T_4. However, if $|\mathcal{X}| \geq 2$, we prove that the strong topology is not first-countable anywhere, hence it is not metrizable. We show that in the strong topology, a subspace is compact if and only if it is rank-bounded and strongly-closed. We provide a necessary and sufficient condition for a sequence of channels to converge in the strong topology.

We introduce a metric distance on the space of output-equivalent channels which compares the noise levels between channels. We show that the induced metric topology, which we call the noisiness topology, is natural. We also study topologies that are inherited from the space of meta-probability measures by identifying channels with their Blackwell measures. We show that the weak-$*$ topology is exactly the same as the noisiness topology and hence it is natural. We prove that if $|\mathcal{X}| \geq 2$, the total-variation topology is not natural nor Baire, hence it is not completely metrizable. Furthermore, we show that it is not locally compact anywhere. Finally, we prove that the Borel σ-algebra is the same for all Hausdorff natural topologies on the space of output-equivalent channels.

We then study the topologies that can be constructed on the spaces of input-equivalent channels. The space of input-equivalent channels with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ can be naturally endowed with the quotient of the Euclidean topology by the input-equivalence relation. We show that this topology is compact, path-connected and metrizable. A topology on the space of input-equivalent channels with a fixed output alphabet $\mathcal{Y}$ and arbitrary but finite input alphabet is said to be natural if and only if it induces the quotient topology on the subspaces of input-equivalent channels sharing the same input alphabet. We show that every natural topology is σ-compact, separable and path-connected. Whereas, if $|\mathcal{Y}| \geq 3$, we prove that a Hausdorff natural topology is not Baire and it is not locally compact anywhere. We show that the finest natural topology, which we call the strong topology, is compactly generated, sequential and T_4. However, if $|\mathcal{Y}| \geq 3$, we prove

that the strong topology is not first-countable anywhere, hence it is not metrizable. We introduce the similarity metric on the space of input-equivalent channels, and we prove that its induced topology is natural.

Some of the above results can also be shown for the spaces of Shannon-equivalent channels. The space of Shannon-equivalent channels with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ can be naturally endowed with the quotient of the Euclidean topology by the Shannon-equivalence relation. We show that this topology is compact, path-connected and metrizable. A topology on the space of Shannon-equivalent channels with arbitrary but finite input and output alphabets is said to be natural if and only if it induces the quotient topology on the subspaces of Shannon-equivalent channels sharing the same input and output alphabets. We show that every natural topology is σ-compact, separable and path-connected. We show that the finest natural topology, which we call the strong topology, is compactly generated, sequential and T_4. We introduce the BRM metric on the space of Shannon-equivalent channels, and we prove that its induced topology is natural. The definition of the BRM metric relies on the characterization of the Shannon ordering in terms of BRM games.

Continuity of Channel Parameters and Operations

In **Chapter 12**, we study the continuity of many channel parameters and operations under various topologies on the space of output-equivalent channels, the space of input-equivalent channels, and the space of Shannon-equivalent channels. The continuity of channel parameters and operations might be helpful in the following two problems:

- If a parameter (such as the optimal probability of error of a given code) is difficult to compute for a channel W, one can approximate it by computing the same parameter for a sequence of channels $(W_n)_{n\geq 0}$ that converges to W in some topology where the parameter is continuous.

- The study of robustness of a communication system against the imperfect specification of the channel.

We show that mutual information, channel capacity, Bhattacharyya parameter, the probability of error of a fixed code, and the optimal probability of error for a given code rate and blocklength, are continuous under various topologies on the space of output-equivalent channels. We also show that channel operations such as sums, products, interpolations, and Arıkan-style transformations are continuous under these topologies.

As for the space of input-equivalent channels, we show that the channel capacity, the probability of error of a given decoder, and the optimal probability of error for a given code rate and blocklength, are continuous under the strong topology. We also prove that channel sums and products are continuous under both the strong and similarity topologies.

Finally, we study the continuity of channel parameters and operations on the space of Shannon-equivalent channels. We show that the channel capacity and the optimal probability of error for a given code rate and blocklength are continuous under the strong topology. We also prove that channel sums and products are continuous under the strong topology.

Part I

Polarization

An Ergodic Theory of Binary Operations

2

In this chapter[1], we develop an ergodic theory for binary operations. This theory will be used in Chapter 3 to provide a necessary and sufficient condition for a binary operation to be polarizing.

In Section 2.1 we introduce the notion of uniformity-preserving operations. A *uniformity-preserving operation* $*$ on $\mathcal{X}$ is a binary operation for which the mapping $f_* : \mathcal{X}^2 \to \mathcal{X}^2$ defined as $f_*(x, y) = (x * y, y)$ is bijective. It is called uniformity-preserving since for any pair of random variables (X_1, X_2) in $\mathcal{X}^2$, $(X_1 * X_2, X_2)$ is uniform in $\mathcal{X}^2$ if and only if (X_1, X_2) is uniform in $\mathcal{X}^2$. As we will see in Chapter 3, if $*$ is not uniformity-preserving, then the Arıkan style construction that is based on $*$ does not conserve the symmetric capacity. Hence being uniformity-preserving is a necessary condition to be polarizing. On the other hand, being a quasigroup operation is a sufficient condition [17]. Therefore, a necessary and sufficient condition must be a property that is stronger than uniformity-preserving and weaker than quasigroup. A reasonable strategy to search for a necessary and sufficient condition is to relax the quasigroup property while keeping the uniformity-preserving property.

The difference between a quasigroup operation and a uniformity-preserving operation is that in the case of a quasigroup operation, any element is reachable from any other element by one multiplication on the right. This property does not always hold for a uniformity-preserving operation.

One possible relaxation of the quasigroup property is to consider uniformity-preserving operations where all the elements are reachable from each other by multiple multiplications on the right. Irreducible and ergodic operations — which are defined and studied in Section 2.2 — satisfy this property. The concepts of irreducible and ergodic operations are very similar to the concepts of irreducible and ergodic Markov chains. The reason why we consider such binary operations is because of their good connectability properties: If the elements of $\mathcal{X}$ are well connected under $*$, this will create strong correlations between the inputs of the synthetic channels, which should ultimately lead to a polarization phenomenon.

Although ergodic operations seem to have good connectability properties, this

[1]The material of this chapter is based on [15, 16].

is not enough to ensure polarization as we will see in Chapter 3. It turns out that we need a stronger notion of ergodicity. In order to define this stronger notion of ergodicity, we first need to define stable partitions. Section 2.3 introduces balanced, periodic and stable partitions and investigates their properties. Stable partitions are a generalization of the concept of quotient groups. In Section 2.4, we introduce and study the notion of the residue of a stable partition and in Section 2.5 we define and investigate strongly ergodic operations. We show that an ergodic operation is strongly ergodic if and only if each stable partition is its own residue. Strong ergodicity is a novel concept and has no analog in the ergodic theory of Markov chains. We will show in Chapter 3 that a binary operation is polarizing if and only if it is uniformity-preserving and its right-inverse is strongly ergodic.

Generated stable partitions are introduced and studied in Section 2.6. This concept is needed to show that the strong ergodicity of the right-inverse operation is a sufficient condition for polarization.

The products of binary operations are defined in Section 2.7 and the structure of their stable partitions is studied. We show that the product of a sequence of binary operations is strongly ergodic if and only if every operation in the sequence is strongly ergodic. As we will see in Chapter 4, the products of binary operations and their stable partitions are important for the study of MAC polarization theory.

2.1 Uniformity-Preserving Operations

All the sets that are considered in this chapter are finite.

Definition 2.1. *A* uniformity-preserving operation $*$ *on* $\mathcal{X}$ *is a binary operation such that the mapping* $f_* : \mathcal{X}^2 \to \mathcal{X}^2$ *defined by* $f_*(x, y) = (x * y, y)$ *is bijective. It is called uniformity-preserving since for any pair of random variables* (X_1, X_2) *in* $\mathcal{X}^2$, $(X_1 * X_2, X_2)$ *is uniform in* $\mathcal{X}^2$ *if and only if* (X_1, X_2) *is uniform in* $\mathcal{X}^2$.

Remark 2.1. *It is easy to see that* $*$ *is uniformity-preserving if and only if it satisfies the following condition:*

- *The multiplication-on-the-right mappings* $\pi_b : \mathcal{X} \to \mathcal{X}$ *defined by* $\pi_b(x) = x * b$ *are bijective for all* $b \in \mathcal{X}$. *We denote* $\pi_b^{-1}(a)$ *as* $a/^*b$. *The binary operation* $/^*$ *is called the* right-inverse *of* $*$.

It is easy to see that if $*$ *is uniformity-preserving then* $/^*$ *is uniformity-preserving as well.*

Definition 2.2. *A uniformity-preserving operation is said to be a quasigroup operation if it also satisfies the following:*

- *The multiplication-on-the-left mappings* $\eta_b : \mathcal{X} \to \mathcal{X}$ *defined by* $\eta_b(x) = b * x$ *are bijective for all* $b \in \mathcal{X}$. *We denote* $\eta_b^{-1}(a)$ *as* $b\backslash_* a$. *The binary operation* $\backslash_*$ *is called the* left-inverse *of* $*$.

It is easy to see that if $*$ *is a quasigroup operation then* $/^*$ *and* $\backslash_*$ *are quasigroup operations as well.*

Note that for a general quasigroup operation $*$, *we may find* $a, b \in \mathcal{X}$ *such that* $\pi_b^{-1}(a) = a/^*b \neq b\backslash_* a = \eta_b^{-1}(a)$. *This is why we use different notations for left and right inverses.*

Notation 2.1. *Let A and B be two subsets of $\mathcal{X}$. We define the set:*

$$A * B := \{a * b : \ a \in A, b \in B\}.$$

For $a, b \in \mathcal{X}$, we denote $\{a\} * B$ and $A * \{b\}$ by $a * B$ and $A * b$ respectively.

It is easy to see that if $*$ is uniformity-preserving and B is non-empty, then $|A * B| \geq |A|$. On the other hand, the relation $|A * B| \geq |B|$ does not hold in general unless $*$ is a quasigroup operation and A is non-empty.

2.2 Irreducible and Ergodic Operations

In this section and throughout the chapter, $*$ is always a uniformity-preserving operation.

Definition 2.3. *Let $*$ be a uniformity-preserving operation on a set $\mathcal{X}$. We say that $a \in \mathcal{X}$ is $*$-connectable to $b \in \mathcal{X}$ in l-steps if there exist l elements $x_0, \ldots, x_{l-1} \in \mathcal{X}$ satisfying $(\ldots((a * x_0) * x_1) \ldots * x_{l-1}) = b$. We denote this relation by $a \xrightarrow{*,l} b$.*

We say that a is $*$-connectable to b if there exists $l > 0$ such that $a \xrightarrow{*,l} b$. We denote this relation by $a \xrightarrow{*} b$.

Definition 2.4. *A uniformity-preserving operation $*$ is said to be* irreducible *if all the elements of $\mathcal{X}$ are $*$-connectable to each other. If $*$ is irreducible, we define the period of an element $a \in \mathcal{X}$ as $\mathrm{per}(*, a) := \gcd\{l > 0 : \ a \xrightarrow{*,l} a\}$, and we define the period of $*$ as:*

$$\mathrm{per}(*) := \gcd\{\mathrm{per}(*, a) : \ a \in \mathcal{X}\} = \gcd\left\{l > 0 : \ \exists a \in \mathcal{X}, a \xrightarrow{*,l} a\right\}.$$

Definition 2.5. *If there exists $l > 0$ such that all the elements of $\mathcal{X}$ are $*$-connectable to each other in l steps, we say that the operation $*$ is* ergodic. *In this case, we call the minimum integer $l > 0$ which satisfies this property the* connectability *of the operation $*$, and we denote it by $\mathrm{con}(*)$, i.e.,*

$$\mathrm{con}(*) = \min\left\{l > 0 : \ \forall a, b \in \mathcal{X}, \ a \xrightarrow{*,l} b\right\}.$$

Remark 2.2. *In order to justify our choice of terminology in the previous definition, consider a sequence $(X'_n)_{n\geq 0}$ of independent and uniformly distributed random variables in $\mathcal{X}$. Define $(X_n)_{n\geq 0}$ recursively as follows: $X_0 = X'_0$ and $X_n = X_{n-1} * X'_n$ for $n > 0$. It is easy to see that $(X_n)_{n\geq 0}$ is a stationary Markov chain. We have the following:*

- *$*$ is irreducible if and only if $(X_n)_{n\geq 0}$ is irreducible.*
- *$*$ is ergodic if and only if $(X_n)_{n\geq 0}$ is ergodic.*

The following proposition shows the important properties of irreducible and ergodic operations. These properties will be used in Chapter 3 to show that every polarizing operation is ergodic.

Proposition 2.1. *We have the following:*

1. *Every quasigroup operation is ergodic, and every ergodic operation is irreducible.*

2. *If $*$ is uniformity-preserving but not irreducible, there exists two disjoint nonempty subsets A_1 and A_2 of $\mathcal{X}$ such that $A_1 \cup A_2 = \mathcal{X}$, $A_1 * \mathcal{X} = A_1$ and $A_2 * \mathcal{X} = A_2$.*

3. *If $*$ is irreducible, we have $\mathrm{per}(*, a) = \mathrm{per}(*)$ for all $a \in \mathcal{X}$.*

4. *If $*$ is irreducible, there exists a partition $\mathcal{E}_*$ of $\mathcal{X}$ containing $n = \mathrm{per}(*)$ subsets $H_0, \ldots, H_{n-1}$ such that $H_i * \mathcal{X} = H_{i+1 \bmod n}$ for all $0 \leq i < n$. Moreover, we have $|H_0| = \ldots = |H_{n-1}|$.*

5. *If $*$ is irreducible, there exists an integer $d > 0$ such that for every $0 \leq i < n = \mathrm{per}(*)$, every element of H_i is $*$-connectable to every element of $H_{i+d \bmod n}$ in d steps. We call the least integer $d > 0$ satisfying this property the* connectability *of the irreducible operation $*$ and we denote it $\mathrm{con}(*)$ (This definition is consistent with the definition of the connectability of ergodic operations. I.e., the connectability of an ergodic operation when it is seen as an irreducible operation is the same as its connectability when it is seen as an ergodic operation).*

6. *If $*$ is irreducible, then for every $s \geq \mathrm{con}(*)$ and every $0 \leq i < n = \mathrm{per}(*)$, any element of H_i is $*$-connectable to any element of $H_{i+s \bmod n}$ in s steps.*

7. *If $*$ is irreducible, $\mathrm{per}(*) = 1$ if and only if $*$ is ergodic.*

8. *If $*$ is ergodic, all the elements of $\mathcal{X}$ are $*$-connectable to each other in s steps for any $s \geq \mathrm{con}(*)$.*

9. *If $*$ is ergodic, then $\mathrm{con}(*) = 1$ if and only if $*$ is a quasigroup operation.*

10. *If $*$ is irreducible (resp. ergodic), then $/^*$ is irreducible (resp. ergodic) as well.*

Proof. See Appendix 2.8.1. □

2.3 Balanced, Periodic and Stable Partitions

Notation 2.2. *Let $\mathcal{H}$ be a set of subsets of a set $\mathcal{X}$, we define the following:*

- $\|\mathcal{H}\|_\wedge = \min_{H \in \mathcal{H}} |H|$.
- $\|\mathcal{H}\|_\vee = \max_{H \in \mathcal{H}} |H|$.

Definition 2.6. *A partition $\mathcal{H}$ of a set $\mathcal{X}$ is said to be a* balanced partition *if all the elements of $\mathcal{H}$ have the same size. We denote the common size of its elements by $\|\mathcal{H}\|$. The number of elements in $\mathcal{H}$ is denoted by $|\mathcal{H}|$. Clearly, $|\mathcal{X}| = |\mathcal{H}| \cdot \|\mathcal{H}\|$ and $\|\mathcal{H}\| = \|\mathcal{H}\|_\wedge = \|\mathcal{H}\|_\vee$ for such a partition.*

Definition 2.7. *Let $\mathcal{H}$ be a partition of a set $\mathcal{X}$. We define the* projection onto $\mathcal{H}$ *as the mapping $\mathrm{Proj}_\mathcal{H} : \mathcal{X} \to \mathcal{H}$, where $\mathrm{Proj}_\mathcal{H}(x)$ is the unique element $H \in \mathcal{H}$ such that $x \in H$.*

Notation 2.3. *Let $\mathcal{A}$ and $\mathcal{B}$ be two sets of subsets of $\mathcal{X}$. We define $\mathcal{A}*\mathcal{B}$ as follows:*

$$\mathcal{A} * \mathcal{B} = \{A * B : \ A \in \mathcal{A}, B \in \mathcal{B}\}.$$

Definition 2.8. *Let $\mathcal{H}$ be a set of subsets of $\mathcal{X}$, and let $*$ be a uniformity-preserving operation on $\mathcal{X}$. We define the set $\mathcal{H}^* = \mathcal{H} * \mathcal{H} = \{A * B : \ A, B \in \mathcal{H}\}$, and we define the sequence $(\mathcal{H}^{n*})_{n\geq 0}$ recursively as follows:*

- $\mathcal{H}^{0*} = \mathcal{H}$.
- $\mathcal{H}^{n*} := (\mathcal{H}^{(n-1)*})^* = \mathcal{H}^{(n-1)*} * \mathcal{H}^{(n-1)*}$ *for all* $n > 0$.

Definition 2.9. *A partition $\mathcal{H}$ of $\mathcal{X}$ is said to be a* periodic partition *of $(\mathcal{X}, *)$ if there exists $n > 0$ such that $\mathcal{H}^{n*} = \mathcal{H}$. In this case, the minimum integer $n > 0$ which satisfies $\mathcal{H}^{n*} = \mathcal{H}$ is called the* period *of $\mathcal{H}$, and it is denoted by* $\mathrm{per}(\mathcal{H})$.

A partition $\mathcal{H}$ of $\mathcal{X}$ is said to be a stable partition *of $(\mathcal{X}, *)$ if $\mathcal{H}$ is both balanced and periodic.*

Throughout the chapter, we write that $\mathcal{H}$ is a periodic (resp. stable) partition of $\mathcal{X}$ if the binary operation $*$ is clear from the context.

Example 2.1. *Let $Q = \mathbb{Z}_n \times \mathbb{Z}_n$, define $(x_1, y_1)*(x_2, y_2) = (x_1+y_1+x_2+y_2, y_1+y_2)$ which is a quasigroup operation. For each $j \in \mathbb{Z}_n$ and each $0 \leq i < n$, define $H_{i,j} = \{(j + ik, k) : \ k \in \mathbb{Z}_n\}$. Let $\mathcal{H}_i = \{H_{i,j} : \ j \in \mathbb{Z}_n\}$ for $0 \leq i < n$. It is easy to see that $\mathcal{H}_i^* = \mathcal{H}_{i+1}$ for $0 \leq i < n-1$ and $\mathcal{H}_{n-1}^* = \mathcal{H}_0$. Therefore, $\mathcal{H} := \mathcal{H}_0$ is a periodic partition of $(Q, *)$ and $\mathrm{per}(\mathcal{H}) = n$. Moreover, $\mathcal{H}$ is balanced with $\|\mathcal{H}\| = n$, hence $\mathcal{H}$ is a stable partition.*

Proposition 2.2. *Let $\mathcal{H}$ be a periodic partition of $(\mathcal{X}, *)$. For every $n > 0$, we have:*

1. *$\mathcal{H}^{n*}$ is a periodic partition and has the same period as $\mathcal{H}$, i.e.,* $\mathrm{per}(\mathcal{H}^{n*}) = \mathrm{per}(\mathcal{H})$.
2. $|\mathcal{H}^{n*}| = |\mathcal{H}|$.

Proof. see Appendix 2.8.2. □

Lemma 2.1. $\|\mathcal{H}^*\|_\vee \geq \|\mathcal{H}\|_\vee$ *and* $\|\mathcal{H}^*\|_\wedge \geq \|\mathcal{H}\|_\wedge$.

Proof. Let $A \in \mathcal{H}$ be such that $A = \|\mathcal{H}\|_\vee$, then $A * A \in \mathcal{H}^*$. Thus, $\|\mathcal{H}^*\|_\vee \geq |A * A| \geq |A| = \|\mathcal{H}\|_\vee$.

Now let B and C be two elements of $\mathcal{H}$ such that $|B * C| = \|\mathcal{H}^*\|_\wedge$. We have $|B * C| \geq |B| \geq \|\mathcal{H}\|_\wedge$. This implies that $\|\mathcal{H}^*\|_\wedge \geq \|\mathcal{H}\|_\wedge$. □

Proposition 2.3. *Let $\mathcal{H}$ be a stable partition of $(\mathcal{X}, *)$. For every $n > 0$, $\mathcal{H}^{n*}$ is a stable partition satisfying* $\mathrm{per}(\mathcal{H}^{n*}) = \mathrm{per}(\mathcal{H})$ *and* $\|\mathcal{H}^{n*}\| = \|\mathcal{H}\|$.

Proof. Proposition 2.2 shows that $\mathcal{H}^{n*}$ is a periodic partition of period $\mathrm{per}(\mathcal{H}^{n*}) = \mathrm{per}(\mathcal{H})$. It remains to show that $\mathcal{H}^{n*}$ is balanced and that $\|\mathcal{H}^{n*}\| = \|\mathcal{H}\|$. Let $p > 0$ be the smallest multiple of $\mathrm{per}(\mathcal{H})$ which is greater than n, i.e.,

$$p = \min\{k \cdot \mathrm{per}(\mathcal{H}) : \ k > 0, \ k \cdot \mathrm{per}(\mathcal{H}) > n\}.$$

We have $\mathcal{H}^{p*} = \mathcal{H}$ since $\mathrm{per}(\mathcal{H})$ divides p. By Lemma 2.1 we have:

- $\|\mathcal{H}\| = \|\mathcal{H}\|_\wedge \leq \|\mathcal{H}^*\|_\wedge \leq \ldots \leq \|\mathcal{H}^{n*}\|_\wedge \leq \ldots \leq \|\mathcal{H}^{p*}\|_\wedge = \|\mathcal{H}\|_\wedge = \|\mathcal{H}\|$.
- $\|\mathcal{H}\| = \|\mathcal{H}\|_\vee \leq \|\mathcal{H}^*\|_\vee \leq \ldots \leq \|\mathcal{H}^{n*}\|_\vee \leq \ldots \leq \|\mathcal{H}^{p*}\|_\vee = \|\mathcal{H}\|_\vee = \|\mathcal{H}\|$.

Therefore, $\|\mathcal{H}^{n*}\|_\wedge = \|\mathcal{H}^{n*}\|_\vee = \|\mathcal{H}\|$, which means that for every $A \in \mathcal{H}^{n*}$ we have $|A| = \|\mathcal{H}\|$. We conclude that $\mathcal{H}^{n*}$ is balanced and $\|\mathcal{H}^{n*}\| = \|\mathcal{H}\|$. □

Lemma 2.2. *If $*$ is ergodic then every periodic partition is stable.*

Proof. Let $\mathcal{H}$ be a periodic partition of $\mathcal{X}$. We only need to show that $\mathcal{H}$ is balanced.

Let $n = \text{per}(\mathcal{H})$ and $m = \min\{kn : k > 0 \text{ and } kn > \text{con}(*)\}$. Clearly, $\mathcal{H}^{m*} = \mathcal{H}$. Moreover, statement 8 of Proposition 2.1 shows that all the elements of $\mathcal{X}$ are $*$-connectable to each other in m steps. Let $H \in \mathcal{H}$ be chosen such that $|H|$ is maximal and let H' be any element of $\mathcal{H}$. Let $h \in H$ and $h' \in H'$. We have $h \xrightarrow{*,m} h'$ so there exist m elements $x_0, \ldots, x_{m-1} \in \mathcal{X}$ satisfying $(\ldots((h * x_0) * x_1)\ldots * x_{m-1}) = h'$.

Since $\mathcal{H}$ covers $\mathcal{X}$, then each of $\mathcal{H}^*$, $\mathcal{H}^{2*}$, …, and $\mathcal{H}^{(m-1)*}$ covers $\mathcal{X}$ as well. And so there exist $X_0 \in \mathcal{H}$, $X_1 \in \mathcal{H}^*$, …, and $X_{m-1} \in \mathcal{H}^{(m-1)*}$ such that $x_0 \in X_0$, $x_1 \in X_1$, …, and $x_{m-1} \in X_{m-1}$. Now since $(\ldots((h * x_0) * x_1)\ldots * x_{m-1}) = h'$ and since $h \in H$, we have $h' \in H'' := (\ldots((H * X_0) * X_1)\ldots * X_{m-1})$. From the definition of H'', we have $H'' \in \mathcal{H}^{m*} = \mathcal{H}$. Moreover, $h' \in H' \cap H''$, so $H' = H''$ since $\mathcal{H}$ is a partition. We conclude that $H' = (\ldots((H * X_0) * X_1)\ldots * X_{m-1})$ which implies that $|H'| \geq |H|$. On the other hand, we have $|H| \geq |H'|$ since H was chosen so that $|H|$ is maximal. We conclude that $|H'| = |H|$ for all $H' \in \mathcal{H}$, hence $\mathcal{H}$ is balanced. □

Remark 2.3. *The ergodicity condition in the previous lemma cannot be replaced by irreducibility. Consider the following irreducible (but not ergodic) operation:*

$*$	0	1	2	3
0	2	3	2	2
1	3	2	3	3
2	0	0	0	1
3	1	1	1	0

Although the partition $\mathcal{H} = \{\{0,1\},\{2\},\{3\}\}$ is not balanced, we have $\mathcal{H}^{2*} = \mathcal{H}$.

The following proposition shows that the concept of periodic partitions generalizes the concept of quotient groups:

Proposition 2.4. *Let $(G,*)$ be a finite group, and let $\mathcal{H}$ be a periodic partition of $(G,*)$. There exists a normal subgroup H of G such that $\mathcal{H}$ is the quotient group of G by H (denoted by G/H).*

Proof. Since every group operation is ergodic, Lemma 2.2 implies that $\mathcal{H}$ is stable, i.e., it is also balanced.

Let H be the element of $\mathcal{H}$ containing the neutral element e of G. For every $H' \in \mathcal{H}$, we have $|H'| = |H * H'| = |H' * H| = \|\mathcal{H}\|$ since $H * H' \in \mathcal{H}^*$, $H' * H \in \mathcal{H}^*$ and $\|\mathcal{H}^*\| = \|\mathcal{H}\|$. On the other hand, we have $H' = e * H' \subset H * H'$ and $H' = H' * e \subset H' * H$. We conclude that $H * H' = H' * H = H'$. Therefore,

- $H * H = H$, hence $x * y \in H$ for every $x, y \in H$.

- For every $x \in H$, we have $|H*x| = |H|$. On the other hand, $H*x \subset H*H = H$. Therefore, $H * x = H$ which implies that $e \in H * x$ and so there exists $x' \in H$ such that $x' * x = e$. We conclude that the inverse of every element of H is also in H.
- For every $x \in G$ let $H_x \in \mathcal{H}$ be such that $x \in H_x$. We have $x*H \subset H_x*H = H_x$ and $|x * H| \overset{(a)}{=} |H| = |H_x|$, where (a) follows from the fact that $*$ is a group operation. Therefore, $x * H = H_x$. Similarly, we can show that $H * x = H_x$. Hence $x * H = H * x = H_x$ for every $x \in G$.

We conclude that H is a normal subgroup of G, and $\mathcal{H}$ is the quotient group of G by H. □

Definition 2.10. *A periodic partition $\mathcal{H}_1$ is said to be a sub-periodic partition of another periodic partition $\mathcal{H}_2$ if for every $H_1 \in \mathcal{H}_1$, there exists $H_2 \in \mathcal{H}_2$ such that $H_1 \subset H_2$. We denote this relation by $\mathcal{H}_1 \preceq \mathcal{H}_2$, and we say that $\mathcal{H}_1$ is finer than $\mathcal{H}_2$.*

If $\mathcal{H}_1$ and $\mathcal{H}_2$ are two stable partitions satisfying $\mathcal{H}_1 \preceq \mathcal{H}_2$, we say that $\mathcal{H}_1$ is a sub-stable partition of $\mathcal{H}_2$ (in such case, we clearly have $\|\mathcal{H}_1\|$ divides $\|\mathcal{H}_2\|$).

Remark 2.4. *Let $(G, *)$ be a group and let $\mathcal{H}_1$ be a sub-periodic partition of a periodic partition $\mathcal{H}_2$. If $H_{\mathcal{H}_1}$ and $H_{\mathcal{H}_2}$ are the normal subgroups associated with $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively, then $H_{\mathcal{H}_1}$ is a normal subgroup of $H_{\mathcal{H}_2}$.*

Definition 2.11. *For any two partitions $\mathcal{H}_1$ and $\mathcal{H}_2$ of a set $\mathcal{X}$, we define:*

$$\mathcal{H}_1 \wedge \mathcal{H}_2 = \{H_1 \cap H_2 :\ H_1 \in \mathcal{H}_1,\ H_2 \in \mathcal{H}_2,\ H_1 \cap H_2 \neq \emptyset\}.$$

Proposition 2.5. *If $\mathcal{H}_1$ and $\mathcal{H}_2$ are periodic partitions then $\mathcal{H}_1 \wedge \mathcal{H}_2$ is a periodic partition of period of at most* $\mathrm{lcm}\{\mathrm{per}(\mathcal{H}_1), \mathrm{per}(\mathcal{H}_2)\}$. *Moreover, we have $(\mathcal{H}_1 \wedge \mathcal{H}_2)^{n*} = \mathcal{H}_1^{n*} \wedge \mathcal{H}_2^{n*}$ for every $n \geq 0$.*

Proof. See Appendix 2.8.2. □

Corollary 2.1. *Let $*$ be an ergodic operation. If $\mathcal{H}_1$ and $\mathcal{H}_2$ are two stable partitions then $\mathcal{H}_1 \wedge \mathcal{H}_2$ is a stable partition of period of at most* $\mathrm{lcm}\{\mathrm{per}(\mathcal{H}_1), \mathrm{per}(\mathcal{H}_2)\}$.

Proof. The corollary follows from Proposition 2.5 and Lemma 2.2. □

Remark 2.5. *Let $(G, *)$ be a group. If $\mathcal{H}_1$ and $\mathcal{H}_2$ are two periodic partitions of $(G, *)$, then $H_{\mathcal{H}_1 \wedge \mathcal{H}_2} = H_{\mathcal{H}_1} \cap H_{\mathcal{H}_2}$.*

Remark 2.6. *The ergodicity condition in Corollary 2.1 cannot be replaced by irreducibility. Consider the following irreducible (but not ergodic) operation:*

$*$	0	1	2	3	4	5	6	7
0	4	5	6	7	4	4	4	4
1	5	4	7	6	5	5	5	5
2	6	7	4	5	6	6	6	6
3	7	6	5	4	7	7	7	7
4	0	0	0	0	0	1	2	3
5	1	1	1	1	1	0	3	2
6	2	2	2	2	2	3	0	1
7	3	3	3	3	3	2	1	0

Define:

$$\mathcal{H}_1 = \big\{\{0,1\},\{2,3\},\{4,5\},\{6,7\}\big\},$$
$$\mathcal{H}_2 = \big\{\{0,2\},\{1,3\},\{4,5\},\{6,7\}\big\}.$$

While both $\mathcal{H}_1$ and $\mathcal{H}_2$ are stable partitions of periods 1 and 2 respectively, the partition $\mathcal{H}_1 \wedge \mathcal{H}_2 = \big\{\{0\},\{1\},\{2\},\{3\},\{4,5\},\{6,7\}\big\}$ is periodic but it is not stable as it is not balanced.

2.4 The Residue of a Stable Partition

Let $\mathcal{H}$ be a stable partition. Let $H \in \mathcal{H}$ and $x \in H$. For any sequence $(X_n)_{n\geq 0}$ satisfying $X_n \in \mathcal{H}^{n*}$ for all $n \geq 0$, define the sequences $(A_n)_{n\geq 0}$ and $(H_n)_{n\geq 0}$ recursively as follows:

- $A_0 = \{x\}$ and $H_0 = H$.
- $A_n = A_{n-1} * X_{n-1} = (\ldots((x * X_0) * X_1)\ldots * X_{n-1})$.
- $H_n = H_{n-1} * X_{n-1} = (\ldots((H * X_0) * X_1)\ldots * X_{n-1})$.

Since $x \in H$, we can show by induction on n that $A_n \subset H_n \in \mathcal{H}^{n*}$ and so $|A_n| \leq |H_n| = \|\mathcal{H}^{n*}\| = \|\mathcal{H}\|$ for all $n \geq 0$. Therefore, $|H_n|$ is constant. On the other hand, $|A_n| \geq |A_{n-1}|$ since $A_n = A_{n-1} * X_{n-1}$. Hence, $|A_n|$ is increasing and it is upper bounded by $\|\mathcal{H}\|$.

Does $|A_n|$ reach $\|\mathcal{H}\|$ or does $|A_n|$ remain strictly less than $\|\mathcal{H}\|$ for all $n \geq 0$? In other words, do we have $A_n = H_n$ for some $n > 0$ or does A_n remain a strict subset of H_n for all $n \geq 0$? The answer depends of course on the sequence $(X_n)_{n\geq 0}$, so one can ask: Is it possible to choose at least one sequence $(X_n)_{n\geq 0}$ for which $|A_n| = \|\mathcal{H}\|$ and $A_n = H_n$ for some $n > 0$?

What are the stable partitions $\mathcal{H}$ for which it is always possible to reach a set in $\mathcal{H}^{n*}$ for some $n > 0$ starting from an arbitrary singleton in $\mathcal{X}$ and then recursively multiplying on the right by sets chosen from $\mathcal{H}^{i*}$ $(0 \leq i < n)$?

It is easy to see that for the trivial stable partition $\mathcal{H} = \{\mathcal{X}\}$, the above condition is equivalent to ergodicity. Therefore, satisfying the above condition for every stable partition is a stronger notion of ergodicity. Strong ergodicity turns out to be important for polarization theory as we will see in Chapter 3. In this section, we introduce the notions and concepts that are necessary to understand strong ergodicity.

Notation 2.4. *Let $\mathfrak{X} = (X_i)_{0\leq i<k}$ be a sequence of subsets X_i of $\mathcal{X}$. We denote the length k of the sequence $\mathfrak{X}$ by $|\mathfrak{X}|$.*

For every $A \subset \mathcal{X}$, we denote $(\ldots((A*X_0)*X_1)\ldots)*X_{k-1})$ by $A*\mathfrak{X}$. If $A = \{a\}$, we write $a * \mathfrak{X}$ to denote $\{a\} * \mathfrak{X}$.

The n^{th} power of the sequence $\mathfrak{X} = (X_i)_{0\leq i<k}$ is the sequence $\mathfrak{X}^n = (X'_i)_{0\leq i<kn}$, where $X'_i = X_{i \bmod k}$ for $0 \leq i < kn$. I.e., $\mathfrak{X}^n$ is obtained by concatenating n copies of $\mathfrak{X}$.

Definition 2.12. *Let $\mathcal{H}$ be a stable partition of $(\mathcal{X}, *)$ where $*$ is uniformity-preserving. A sequence $\mathfrak{X} = (X_i)_{0\leq i<k}$ is said to be* $\mathcal{H}$-sequence *if $X_0 \in \mathcal{H}$, $X_1 \in \mathcal{H}^*$, $\ldots$, $X_{k-1} \in \mathcal{H}^{(k-1)*}$. If we also have that $\mathrm{per}(\mathcal{H})$ divides $|\mathfrak{X}| = k$, we say that the sequence is* $\mathcal{H}$-repeatable.

An $\mathcal{H}$-repeatable sequence $\mathfrak{X}$ is said to be $\mathcal{H}$-augmenting *if $A \subset A * \mathfrak{X}$ for all $A \subset \mathcal{X}$.*

Remark 2.7. *If $\mathfrak{X}$ is $\mathcal{H}$-repeatable, then $\mathfrak{X}^l$ is an $\mathcal{H}$-sequence for every $l > 0$. This is not necessarily true if $\mathfrak{X}$ is an $\mathcal{H}$-sequence which is not repeatable.*

If a sequence is $\mathcal{H}$-augmenting then it is also $\mathcal{H}$-repeatable by definition. Therefore, whenever we need to show that a sequence is $\mathcal{H}$-augmenting, we have to show first that it is $\mathcal{H}$-repeatable.

If $\mathfrak{X}$ is $\mathcal{H}$-augmenting then $\mathfrak{X}^l$ is $\mathcal{H}$-augmenting for every $l > 0$.

Theorem 2.1. *Let $\mathcal{H}$ be a stable partition of $(\mathcal{X}, *)$ where $*$ is ergodic. There exists a unique sub-stable partition $\mathcal{K}_\mathcal{H}$ of $\mathcal{H}$ such that:*

- *For every $K \in \mathcal{K}_\mathcal{H}$ and every $\mathcal{H}$-sequence $\mathfrak{X}$, we have $K * \mathfrak{X} \in {\mathcal{K}_\mathcal{H}}^{|\mathfrak{X}|*}$.*
- *For every $K \in \mathcal{K}_\mathcal{H}$ and every $x \in K$, there exists an $\mathcal{H}$-augmenting sequence $\mathfrak{X}$ such that $x * \mathfrak{X} = K$.*
- *For every $K \in \mathcal{K}_\mathcal{H}$, every $x \in K$, and every $\mathcal{H}$-augmenting sequence $\mathfrak{X}'$, we have $x * \mathfrak{X}' \subset K$.*

$\mathcal{K}_\mathcal{H}$ is called the first residue *of the stable partition $\mathcal{H}$. We also have ${\mathcal{K}_\mathcal{H}}^{l*} = \mathcal{K}_{\mathcal{H}^{l*}}$ for all $l \geq 0$.*

Proof. See Appendix 2.8.3. □

Remark 2.8. *Theorem 2.1 implies that an ergodic operation is strongly ergodic if and only if $\mathcal{K}_\mathcal{H} = \mathcal{H}$ for every stable partition $\mathcal{H}$ of $\mathcal{X}$. This will be explained and proven in detail in Section 2.5.*

Remark 2.9. *It is possible to prove a more general theorem for the periodic partitions of an arbitrary uniformity-preserving operation:*

Let $\mathcal{H}$ be a periodic partition of $(\mathcal{X}, *)$ where $*$ is an arbitrary uniformity-preserving operation. There exists a unique sub-periodic partition $\mathcal{K}_\mathcal{H}$ of $\mathcal{H}$ such that:

- *For every $K \in \mathcal{K}_\mathcal{H}$ and every $\mathcal{H}$-sequence $\mathfrak{X}$, we have $K * \mathfrak{X} \in {\mathcal{K}_\mathcal{H}}^{|\mathcal{X}|*}$.*
- *For every $K \in \mathcal{K}_\mathcal{H}$ and every $x \in K$, there exists an $\mathcal{H}$-augmenting sequence $\mathfrak{X}$ such that $x * \mathfrak{X} = K$.*
- *For every $K \in \mathcal{K}_\mathcal{H}$, every $x \in K$, and every $\mathcal{H}$-augmenting sequence $\mathfrak{X}'$, we have $x * \mathfrak{X}' \subset K$.*

$\mathcal{K}_\mathcal{H}$ is called the first residue *of the periodic partition $\mathcal{H}$. We also have ${\mathcal{K}_\mathcal{H}}^{l*} = \mathcal{K}_{\mathcal{H}^{l*}}$ for all $l \geq 0$.*

We will not prove this general theorem here since Theorem 2.1 is sufficient for our purposes. The proof of the general theorem is more complicated but follows similar steps as the proof of Theorem 2.1.

Note that if the operation $*$ is not ergodic, $\mathcal{K}_\mathcal{H}$ may not be a stable partition even if $\mathcal{H}$ is a stable partition. Consider the following irreducible (but not ergodic) operation:

$*$	0	1	2	3	4	5	6	7
0	4	5	4	5	4	4	4	4
1	5	4	5	4	5	5	5	5
2	6	7	6	7	6	6	6	6
3	7	6	7	6	7	7	7	7
4	2	2	2	2	2	3	2	3
5	3	3	3	3	3	2	3	2
6	0	0	0	0	0	1	0	1
7	1	1	1	1	1	0	1	0

Let $\mathcal{H} = \big\{\{0,2\},\{1,3\},\{4,5\},\{6,7\}\big\}$*, which is a stable partition of period 2. The reader can check that* $\mathcal{K}_{\mathcal{H}} = \big\{\{0\},\{1\},\{2\},\{3\},\{4,5\},\{6,7\}\big\}$ *which is periodic but not stable as it is not balanced.*

Definition 2.13. *Let* $\mathcal{H}$ *be a stable partition of* $(\mathcal{X},*)$ *where* $*$ *is ergodic. For every* $n \geq 0$*, we define the* n^{th} *residue* $\mathcal{R}_n(\mathcal{H})$ *of* $\mathcal{H}$ *recursively as follows:*

- $\mathcal{R}_0(\mathcal{H}) = \mathcal{H}$.
- $\mathcal{R}_1(\mathcal{H}) = \mathcal{K}_{\mathcal{H}}$.
- $\mathcal{R}_{n+1}(\mathcal{H}) = \mathcal{R}_1(\mathcal{R}_n(\mathcal{H})) = \mathcal{K}_{\mathcal{R}_n(\mathcal{H})}$ *for every* $n \geq 1$.

The residual degree $\deg_{\mathcal{R}}(\mathcal{H})$ *of* $\mathcal{H}$ *is the smallest integer* $n \geq 0$ *that satisfies* $\mathcal{R}_{n+1}(\mathcal{H}) = \mathcal{R}_n(\mathcal{H})$*. The residue of* $\mathcal{H}$ *is defined as* $\mathcal{R}(\mathcal{H}) := \mathcal{R}_{\deg_{\mathcal{R}}(\mathcal{H})}(\mathcal{H})$*. Clearly* $\mathcal{R}_1(\mathcal{R}(\mathcal{H})) = \mathcal{K}_{\mathcal{R}(\mathcal{H})} = \mathcal{R}(\mathcal{H})$ *and* $\mathcal{R}(\mathcal{R}(\mathcal{H})) = \mathcal{R}(\mathcal{H})$.

Remark 2.10. *In the application to polarization theory, we will only need the first residue. We just note here that for every* $n \geq 0$*, there exists an ergodic operation and a stable partition* $\mathcal{H}$ *of residual degree* n*. In other words, there are stable partitions of arbitrary residual degrees.*

2.5 Strongly Ergodic Operations

Definition 2.14. *A uniformity-preserving operation* $*$ *is said to be* strongly ergodic *if for every stable partition* $\mathcal{H}$ *and for every* $x \in \mathcal{X}$*, there exists an integer* $n = n(x,\mathcal{H})$ *such that for every* $H \in \mathcal{H}^{n*}$*, there exists an* $\mathcal{H}$*-sequence* $\mathfrak{X}_{x,H}$ *of length* n *such that* $x * \mathfrak{X}_{x,H} = H$.

Theorem 2.2. *We have the following:*

1. *If* $*$ *is strongly ergodic then it is ergodic.*
2. *If* $*$ *is strongly ergodic, there exists an integer* $d > 0$ *such that for every* $s \geq d$*, every stable partition* $\mathcal{H}$*, every* $x \in \mathcal{X}$ *and every* $H \in \mathcal{H}^{s*}$*, there exists an* $\mathcal{H}$*-sequence* $\mathcal{X}_{x,H}$ *of length* s *satisfying* $x * \mathcal{X}_{x,H} = H$*. If* d *is minimal with this property, we call it the* strong connectability *of* $*$*, and we denote it by* $\mathrm{scon}(*)$.
3. *If* $*$ *is ergodic, then* $*$ *is strongly ergodic if and only if* $\mathcal{K}_{\mathcal{H}} = \mathcal{H}$ *for every stable partition* $\mathcal{H}$ *(i.e., every stable partition* $\mathcal{H}$ *is its own residue, and so the residual degree is zero).*

4. If $*$ is a quasigroup operation then it is strongly ergodic.

Proof. 1) Suppose that $*$ is strongly ergodic and consider the trivial stable partition $\{\mathcal{X}\}$. For every $x \in \mathcal{X}$, there exists $n_x > 0$ such that $x * (\mathcal{X})^{n_x} = \mathcal{X}$. This shows that for every $y \in \mathcal{X}$, $x \xrightarrow{*,n_x} y$ which shows that $*$ is irreducible. Let $n = \text{per}(*)$ and let $H_0, \ldots, H_{n-1}$ be the equally sized subsets of $\mathcal{X}$ given by the fourth point of Proposition 2.1.

Let $x \in H_0$. We have $\mathcal{X} = x * (\mathcal{X})^{n_x} \subset H_0 * (\mathcal{X})^{n_x} = H_{n_x \bmod n}$, where the last equality follows from the fourth point of Proposition 2.1. Therefore, $H_{n_x \bmod n} = \mathcal{X}$ which implies that $n = 1$ since $\{H_0, \ldots, H_{n-1}\}$ is a partition. Therefore, $\text{per}(*) = 1$ and so $*$ is ergodic by the seventh point of Proposition 2.1.

2) Let $*$ be strongly ergodic, and define $d = \max_{x,\mathcal{H}} n(x, \mathcal{H})$, where $n(x, \mathcal{H})$ is as in Definition 2.14. Now fix $x \in \mathcal{X}$ and fix a stable partition $\mathcal{H}$. Let $s \geq d$ and fix $H \in \mathcal{H}^{s*}$. If $s = n(x, \mathcal{H})$, there is nothing to prove. Now suppose that $s > n := n(x, H)$, then there exists $H' \in \mathcal{H}^{n*}$ and an $\mathcal{H}^{n*}$-sequence $\mathfrak{X}$ of length $s - n$ such that $H' * \mathfrak{X} = H$. Moreover, there exists an $\mathcal{H}$-sequence $\mathfrak{X}_{x,H'}$ of length n such that $x * \mathfrak{X}_{x,H'} = H'$. We conclude that $x * (\mathfrak{X}_{x,H'}, \mathfrak{X}) = H$.

3) Let $\mathcal{H}$ be a stable partition of $(\mathcal{X}, *)$ where $*$ is strongly ergodic, and let $x \in \mathcal{X}$, $K \in \mathcal{K}_\mathcal{H}$ and $H \in \mathcal{H}$ be chosen so that $x \in K \subset H$. Let $s = \text{scon}(*) \cdot \text{per}(\mathcal{H})$. We have $\mathcal{H}^{s*} = \mathcal{H}$ since $\text{per}(\mathcal{H})$ divides s. Now since $s \geq \text{scon}(*)$ and $H \in \mathcal{H} = \mathcal{H}^{s*}$, there exists an $\mathcal{H}$-sequence $\mathfrak{X}_{x,H}$ of length s such that $x * \mathfrak{X}_{x,H} = H$. We have $x \in H = x * \mathfrak{X}_{x,H} \subset K * \mathfrak{X}_{x,H}$, so $x \in K * \mathfrak{X}_{x,H}$ which implies that $K \cap (K * \mathfrak{X}_{x,H}) \neq \emptyset$ (since we also have $x \in K$). On the other hand, Theorem 2.1 implies that $K * \mathfrak{X}_{x,H} \in \mathcal{K}_\mathcal{H}{}^{s*} = \mathcal{K}_\mathcal{H}$. Therefore, $K * \mathfrak{X}_{x,H} = K$ since $\mathcal{K}_\mathcal{H}$ is a partition. We conclude that $H = x * \mathfrak{X}_{x,H} \subset K * \mathfrak{X}_{x,H} = K$ which implies that $H = K$ since we also have $K \subset H$. Therefore, $\|\mathcal{K}_\mathcal{H}\| = \|\mathcal{H}\|$ and so $\mathcal{K}_\mathcal{H} = \mathcal{H}$.

Now suppose that $*$ is an ergodic operation which satisfies $\mathcal{K}_\mathcal{H} = \mathcal{H}$ for every stable partition $\mathcal{H}$. Let $x \in \mathcal{X}$ and let $\mathcal{H}$ be a stable partition. Let $k = \text{con}(*) \cdot \text{per}(\mathcal{H}) \geq \text{con}(*)$, and for each $H \in \mathcal{H}$ fix $x_H \in H$ and let $\mathfrak{X}_H$ be an $\mathcal{H}$-augmenting sequence such that $x_H * \mathfrak{X}_H = H$ (such $\mathfrak{X}_H$ exists due to Theorem 2.1). Define $n(x, H) = k + \sum_{H \in \mathcal{H}} |\mathfrak{X}_H|$ and define $\mathfrak{X}'$ to be the $\mathcal{H}$-augmenting sequence obtained by concatenating all the $\mathfrak{X}_H$ sequences (the order of the concatenation is not important). It is easy to see that $x_H * \mathfrak{X}' = H$ for all $H \in \mathcal{H}$: We have $x_H * \mathfrak{X}' \subset H$ from Theorem 2.1. On the other hand, $H \subset x_H * \mathfrak{X}'$ follows from the fact that $\mathfrak{X}'$ is the concatenation of a collection of $\mathcal{H}$-augmenting sequences containing $\mathfrak{X}_H$ and that $x_H * \mathfrak{X}_H = H$. We also have $|\mathfrak{X}'| = \sum_{H \in \mathcal{H}} |\mathfrak{X}_H|$. Now since $k \geq \text{con}(*)$, it follows from Proposition 2.1 that for every $H \in \mathcal{H}$ there exists a sequence $x_0, \ldots, x_{k-1}$ satisfying $(\ldots((x * x_0) * x_1) \ldots * x_{k-1}) = x_H$. Let $\mathfrak{X}'_H = (X_0, \ldots, X_{k-1})$ be an $\mathcal{H}$-sequence of length k such that $x_i \in X_i$ for all $0 \leq i < k$. Clearly, $x_H \in x * \mathfrak{X}'_H$. It is easy to see that the sequence $\mathfrak{X}''_H = (\mathfrak{X}'_H, \mathfrak{X}')$ is of length $n(x, \mathcal{H})$ and satisfies $H \subset x * \mathfrak{X}''_H$. Now let $H_x \in \mathcal{H}$ be chosen so that $x \in H_x$. Since $H_x \in \mathcal{H} = \mathcal{K}_\mathcal{H}$, Theorem 2.1 implies that we have $H_x * \mathfrak{X}''_H \in \mathcal{K}_{\mathcal{H}^{n(x,\mathcal{H})*}} = \mathcal{K}_\mathcal{H} = \mathcal{H}$ (note that $\mathcal{H}^{n(x,\mathcal{H})*} = \mathcal{H}$ since $\text{per}(\mathcal{H})$ divides $n(x, \mathcal{H})$). We conclude that $H \subset x * \mathfrak{X}''_H \subset H_x * \mathfrak{X}''_H \in \mathcal{H}$, which implies that $H = x * \mathfrak{X}''_H = H_x * \mathfrak{X}''_H$ since we have $H \in \mathcal{H}$ and $\mathcal{H}$ is a partition. Therefore,

for every $H \in \mathcal{H} = \mathcal{H}^{n(x,\mathcal{H})*}$, there exists an $\mathcal{H}$-sequence $\mathfrak{X}''_H$ of length $n(x,\mathcal{H})$ such that $x * \mathfrak{X}''_H = H$. Thus, $*$ is a strongly ergodic operation.

4) Let $\mathcal{H}$ be a stable partition of a quasigroup operation $*$. For every $K \in \mathcal{K}_\mathcal{H}$ and every $x \in K$, there exists an $\mathcal{H}$-augmenting sequence $\mathfrak{X} = (X_i)_{0\leq i<k}$ such that $K = x * \mathfrak{X}$, which implies that $|K| = |x * \mathfrak{X}| = \left|\left(x * (X_i)_{0\leq i<k-1}\right) * X_{k-1}\right| \overset{(a)}{\geq} |X_{k-1}| = \|\mathcal{H}\|$, where (a) is true because $*$ is a quasigroup operation. We conclude that $\|\mathcal{K}_\mathcal{H}\| = \|\mathcal{H}\|$ which implies that $\mathcal{K}_\mathcal{H} = \mathcal{H}$. □

Remark 2.11. *While every strongly ergodic operation $*$ is ergodic, the converse is not true. Consider the following operation:*

$*$	0	1	2	3
0	2	2	0	0
1	3	3	1	1
2	1	1	3	3
3	0	0	2	2

The first residue of the stable partition $\mathcal{H} = \{\{0,1\},\{2,3\}\}$ is

$$\mathcal{K}_\mathcal{H} = \{\{0\},\{1\},\{2\},\{3\}\} \neq \mathcal{H}.$$

Also, a strongly ergodic operation need not be a quasigroup operation, here is an example:

$*$	0	1	2	3
0	3	3	3	3
1	0	1	0	0
2	1	0	1	1
3	2	2	2	2

2.6 Generated Stable Partitions

Definition 2.15. *Let $\mathcal{A}$ and $\mathcal{B}$ be two sets of subsets of $\mathcal{X}$. We say that $\mathcal{A}$ is* finer *than $\mathcal{B}$ (or $\mathcal{B}$ is* coarser *than $\mathcal{A}$) if for every $A \in \mathcal{A}$ there exists $B \in \mathcal{B}$ such that $A \subset B$. We write $\mathcal{A} \preceq \mathcal{B}$ to denote the relation "$\mathcal{A}$ is finer than $\mathcal{B}$".*

Let $\mathcal{A}$ be a set of subsets of $\mathcal{X}$. Is it possible to find a periodic partition of $(\mathcal{X},*)$ which is coarser than $\mathcal{A}$ and finer than every other periodic partition that is coarser than $\mathcal{A}$? Similarly, is it possible to find a stable partition of $(\mathcal{X},*)$ which is coarser than $\mathcal{A}$ and finer than every other stable partition that is coarser than $\mathcal{A}$? The following answer these two questions.

Proposition 2.6. *Let $*$ be a uniformity-preserving operation on $\mathcal{X}$, and let $\mathcal{A}$ be a set of subsets of $\mathcal{X}$. There exists a unique periodic partition $\langle\mathcal{A}\rangle$ which satisfies the following:*

- $\mathcal{A} \preceq \langle\mathcal{A}\rangle$.
- *For every periodic partition $\mathcal{H}$ of $\mathcal{X}$, if $\mathcal{A} \preceq \mathcal{H}$ then $\langle\mathcal{A}\rangle \preceq \mathcal{H}$.*

In other words, $\langle\mathcal{A}\rangle$ is the finest periodic partition that is coarser than $\mathcal{A}$. $\langle\mathcal{A}\rangle$ is called the periodic partition generated *by $\mathcal{A}$.*

Proof. Define

$$\langle\mathcal{A}\rangle = \bigwedge_{\substack{\mathcal{H} \text{ is a periodic partition} \\ \mathcal{A}\preceq\mathcal{H}}} \mathcal{H}. \tag{2.1}$$

Proposition 2.5 implies that $\langle\mathcal{A}\rangle$ is a periodic partition. Moreover, it follows from (2.1) and from the definition of the wedge operator (Definition 2.11) that for every periodic partition $\mathcal{H}$ satisfying $\mathcal{A}\preceq\mathcal{H}$, we have $\langle\mathcal{A}\rangle\preceq\mathcal{H}$.

Now let $A\in\mathcal{A}$. We have:

- If $A=\emptyset$, then $A\subset B$ for every $B\in\langle\mathcal{A}\rangle$.
- If $A\neq\emptyset$, then for every periodic partition $\mathcal{H}$ satisfying $\mathcal{A}\preceq\mathcal{H}$, choose $B_{\mathcal{H}}\in\mathcal{H}$ such that $A\subset B_{\mathcal{H}}$. Define

$$B = \bigcap_{\substack{\mathcal{H} \text{ is a periodic partition} \\ \mathcal{A}\preceq\mathcal{H}}} B_{\mathcal{H}}.$$

  Clearly, $A\subset B$ which implies that $B\neq\emptyset$ and so $B\in\langle\mathcal{A}\rangle$ (see Definition 2.11).

We conclude that for every $A\in\mathcal{A}$, there exists $B\in\langle\mathcal{A}\rangle$ such that $A\subset B$. Therefore, $\mathcal{A}\preceq\langle\mathcal{A}\rangle$.

Now let $\mathcal{H}'$ be a periodic partition satisfying the conditions of the proposition. I.e.,

- $\mathcal{A}\preceq\mathcal{H}'$.
- For every periodic partition $\mathcal{H}$ of $\mathcal{X}$, if $\mathcal{A}\preceq\mathcal{H}$ then $\mathcal{H}'\preceq\mathcal{H}$.

Since $\mathcal{A}\preceq\langle\mathcal{A}\rangle$, we have $\mathcal{H}'\preceq\langle\mathcal{A}\rangle$. Similarly, since $\mathcal{A}\preceq\mathcal{H}'$ we have $\langle\mathcal{A}\rangle\preceq\mathcal{H}'$. Therefore, $\mathcal{H}'=\langle\mathcal{A}\rangle$ and so $\langle\mathcal{A}\rangle$ is unique. □

Remark 2.12. *It is possible to show that $\langle\mathcal{A}\rangle^{n*}=\langle\mathcal{A}^{n*}\rangle$ for every $n>0$, but we will not prove this here since we do not need this property for our purposes.*

Corollary 2.2. *Let $*$ be an ergodic operation on $\mathcal{X}$, and let $\mathcal{A}$ be a set of subsets of $\mathcal{X}$. There exists a unique stable partition $\langle\mathcal{A}\rangle$ which satisfies the following:*

- *$\mathcal{A}\preceq\langle\mathcal{A}\rangle$.*
- *For every stable partition $\mathcal{H}$ of $\mathcal{X}$, if $\mathcal{A}\preceq\mathcal{H}$ then $\langle\mathcal{A}\rangle\preceq\mathcal{H}$.*

In other words, $\langle\mathcal{A}\rangle$ is the finest stable partition that is coarser than $\mathcal{A}$. $\langle\mathcal{A}\rangle$ is called the stable partition generated *by $\mathcal{A}$.*

Proof. The corollary follows from Proposition 2.6 and from the fact that if $*$ is an ergodic operation on $\mathcal{X}$ then every periodic partition is stable (see Lemma 2.2). □

Remark 2.13. *The ergodicity condition in Corollary 2.2 cannot be replaced by irreducibility. Consider the irreducible (but not ergodic) operation $*$ of Remark 2.6, and let $\mathcal{A} = \big\{\{0,1\},\{2,3\}\big\}$. Notice that there is no stable partition that is both coarser than $\mathcal{A}$ and finer than every stable partition that is coarser than $\mathcal{A}$. Therefore, if $*$ is not ergodic, the concept of "generated stable partitions" is not always well defined.*

Let $\mathcal{A}$ be a set of subsets of $\mathcal{X}$ which covers $\mathcal{X}$ and does not contain the empty set as an element. We have $\mathcal{A} \preceq \langle\mathcal{A}\rangle$ which implies that $\mathcal{A}^{n*} \preceq \langle\mathcal{A}\rangle^{n*}$ for every $n > 0$. Can we find $n > 0$ for which $\mathcal{A}^{n*} = \langle\mathcal{A}\rangle^{n*}$? The rest of this section is dedicated to show that the answer to this question is affirmative if $*$ is strongly ergodic. This property of strongly ergodic operations turns out to be important for polarization theory as we will see in Chapter 3.

Definition 2.16. *Let $\mathcal{A}$ be a set of subsets of $\mathcal{X}$. We say that $\mathcal{A}$ is an* $\mathcal{X}$-cover *if $\emptyset \notin \mathcal{A}$ and $\mathcal{X} = \bigcup_{A\in\mathcal{A}} A$.*

We say that an $\mathcal{X}$-cover $\mathcal{A}$ is periodic *if $\mathcal{A}^{n*} = \mathcal{A}$ for some $n > 0$. The least integer $n > 0$ satisfying $\mathcal{A}^{n*} = \mathcal{A}$ is called the* period *of $\mathcal{A}$, and it is denoted by* $\mathrm{per}(\mathcal{A})$.

We say that an $\mathcal{X}$-cover $\mathcal{A}$ is balanced *if for every $A_1, A_2 \in \mathcal{A}$ we have $|A_1| = |A_2|$. An $\mathcal{X}$-cover $\mathcal{A}$ is said to be* stable *if it is both periodic and balanced.*

Proposition 2.7. *If $*$ is a strongly ergodic operation on a set $\mathcal{X}$, then every stable $\mathcal{X}$-cover is a stable partition.*

Proof. See Appendix 2.8.4. □

Remark 2.14. *The strong ergodicity condition in Proposition 2.7 cannot be replaced by ergodicity. Consider the following ergodic (but not strongly ergodic) operation:*

$*$	0	1	2	3	4	5
0	3	3	3	0	0	0
1	4	4	4	1	1	1
2	5	5	5	2	2	2
3	1	1	1	5	5	5
4	2	2	2	3	3	3
5	0	0	0	4	4	4

The set $\big\{\{0,1\},\{0,2\},\{1,2\},\{3,4\},\{3,5\},\{4,5\}\big\}$ is a stable $\mathcal{X}$-cover of period 1, but it is not a partition.

Definition 2.17. *Let $\mathcal{A}$ be a set of subsets of $\mathcal{X}$. The* core *of $\mathcal{A}$ is defined as*

$$\mathrm{core}(\mathcal{A}) = \{A \in \mathcal{A} :\ |A| = \|\mathcal{A}\|_\vee\} = \Big\{A \in \mathcal{A} :\ |A| = \max_{B\in\mathcal{A}} |B|\Big\}.$$

Lemma 2.3. *Let $*$ be a uniformity-preserving operation on $\mathcal{X}$ and let $\mathcal{A}$ be a periodic $\mathcal{X}$-cover. We have $\|\mathcal{A}^{n*}\|_\vee = \|\mathcal{A}\|_\vee$ for every $n \geq 1$.*

Proof. Let $p = \min\{k \cdot \mathrm{per}(\mathcal{A}) :\ k \cdot \mathrm{per}(\mathcal{A}) > n\}$. Lemma 2.1 implies that

$$\|\mathcal{A}\|_\vee \leq \|\mathcal{A}^*\|_\vee \leq \ldots \leq \|\mathcal{A}^{n*}\|_\vee \leq \ldots \leq \|\mathcal{A}^{p*}\|_\vee = \|\mathcal{A}\|_\vee.$$

Therefore, $\|\mathcal{A}^{n*}\|_\vee = \|\mathcal{A}\|_\vee$. □

Proposition 2.8. *Let $*$ be an ergodic operation on $\mathcal{X}$. If $\mathcal{A}$ is a periodic $\mathcal{X}$-cover, then* $\mathrm{core}(\mathcal{A})$ *is a stable $\mathcal{X}$-cover and* $\mathrm{per}(\mathrm{core}(\mathcal{A}))$ *divides* $\mathrm{per}(\mathcal{A})$*. Moreover, we have* $\mathrm{core}(\mathcal{A})^{n*} = \mathrm{core}(\mathcal{A}^{n*})$ *for every $n \geq 1$.*

Proof. See Appendix 2.8.5. □

Proposition 2.9. *Let $*$ be a strongly ergodic operation on $\mathcal{X}$. If $\mathcal{A}$ is a periodic $\mathcal{X}$-cover, then $\langle \mathcal{A} \rangle = \mathrm{core}(\mathcal{A})$.*

Proof. Proposition 2.8 implies that $\mathrm{core}(\mathcal{A})$ is a stable $\mathcal{X}$-cover and $\mathrm{per}(\mathrm{core}(\mathcal{A}))$ divides $\mathrm{per}(\mathcal{A})$. On the other hand, Proposition 2.7 implies that $\mathrm{core}(\mathcal{A})$ is a stable partition.

Fix $a \in A \in \mathcal{A}$ and let $B \in \mathrm{core}(\mathcal{A})$ be such that $a \in B$. Theorem 2.1 implies the existence of a $\mathrm{core}(\mathcal{A})$-augmenting sequence $\mathfrak{X}$ such that $a * \mathfrak{X} = B$. Since $a \in A$, we have $B = a * \mathfrak{X} \subset A * \mathfrak{X}$. On the other hand, we have $A * \mathfrak{X} \in \mathcal{A}^{n*}$, where $n = |\mathfrak{X}|$. This means that $|A * \mathfrak{X}| \leq \|\mathcal{A}^{n*}\|_\vee \overset{(a)}{=} \|\mathcal{A}\|_\vee = |B|$, where (a) follows from Lemma 2.3.

Now since $B \subset A * \mathfrak{X}$ and $|A * \mathfrak{X}| \leq |B|$, we must have $A * \mathfrak{X} = B$. On the other hand, since $\mathfrak{X}$ is $\mathrm{core}(\mathcal{A})$-augmenting, we have $A \subset A * \mathfrak{X} = B$.

We have just shown that for every $A \in \mathcal{A}$, there exists $B \in \mathrm{core}(\mathcal{A})$ such that $A \subset B$. Therefore, $\mathcal{A} \preceq \mathrm{core}(\mathcal{A})$, which implies that $\langle \mathcal{A} \rangle \preceq \mathrm{core}(\mathcal{A})$. On the other hand, since $\mathrm{core}(\mathcal{A}) \subset \mathcal{A}$, we have $\mathrm{core}(\mathcal{A}) \preceq \mathcal{A}$, which implies that $\mathrm{core}(\mathcal{A}) \preceq \langle \mathcal{A} \rangle$. We conclude that $\langle \mathcal{A} \rangle = \mathrm{core}(\mathcal{A})$. □

Remark 2.15. *The strong ergodicity condition in Proposition 2.9 cannot be replaced by ergodicity. Consider the ergodic operation $*$ of Remark 2.14, and consider the the $\mathcal{X}$-cover*

$$\mathcal{A} = \big\{\{0,1\},\{0,2\},\{1,2\},\{3,4\},\{3,5\},\{4,5\}\big\}.$$

$\mathrm{core}(\mathcal{A}) = \mathcal{A}$ *is not a partition, hence* $\mathrm{core}(\mathcal{A}) \neq \langle \mathcal{A} \rangle$.

Theorem 2.3. *Let $*$ be a strongly ergodic operation on a set $\mathcal{X}$. For every $\mathcal{X}$-cover $\mathcal{A}$, there exists an integer $n < 2^{2^{|\mathcal{X}|}}$ such that* $\langle \mathcal{A} \rangle = \mathrm{core}(\mathcal{A}^{n*})$ *and* $\mathrm{per}(\langle \mathcal{A} \rangle)$ *divides n, i.e.,* $\langle \mathcal{A} \rangle = \langle \mathcal{A} \rangle^{n*} = \mathrm{core}(\mathcal{A}^{n*}) \subset \mathcal{A}^{n*}$.

Proof. $2^{|\mathcal{X}|}$ is the number of subsets of $\mathcal{X}$, and $2^{2^{|\mathcal{X}|}}$ is the number of sets of subsets of $\mathcal{X}$. Thus, the sets $\mathcal{A}^{i*}$ for $0 \leq i \leq 2^{2^{|\mathcal{X}|}}$ cannot be pairwise different. Therefore, there exist at least two integers $0 \leq n_1 < n_2 \leq 2^{2^{|\mathcal{X}|}}$ such that $\mathcal{A}^{n_1*} = \mathcal{A}^{n_2*}$. Define $p = n_2 - n_1$ and let $0 \leq n_3 < p$ be such that $n_3 \equiv -n_1 \bmod p$. Define $n = n_1 + n_3$. We have $n < n_1 + p = n_2 \leq 2^{2^{|\mathcal{X}|}}$. On the other hand, since $n \equiv 0 \bmod p$, it follows that p divides n.

We have

$$(\mathcal{A}^{n*})^{p*} = \mathcal{A}^{(n_1+n_3+p)*} = \mathcal{A}^{(n_2+n_3)*} = (\mathcal{A}^{n_2*})^{n_3*} = (\mathcal{A}^{n_1*})^{n_3*} = \mathcal{A}^{n*},$$

which shows that $\mathcal{A}^{n*}$ is a periodic $\mathcal{X}$-cover and $\mathrm{per}(\mathcal{A}^{n*})$ divides p. But p divides n, so $\mathrm{per}(\mathcal{A}^{n*})$ divides n.

Proposition 2.8 shows that $\mathrm{core}(\mathcal{A}^{n*})$ is a stable $\mathcal{X}$-cover and $\mathrm{per}(\mathrm{core}(\mathcal{A}^{n*}))$ divides $\mathrm{per}(\mathcal{A}^{n*})$. This implies that $\mathrm{per}(\mathrm{core}(\mathcal{A}^{n*}))$ divides n. On the other hand, Proposition 2.7 implies that $\mathrm{core}(\mathcal{A}^{n*})$ is a stable partition.

Now let $A \in \mathcal{A}$ and let a be an arbitrary element of $\mathcal{X}$. Define the mapping $\pi : \mathcal{X} \to \mathcal{X}$ as $\pi(x) = x * a$. Since π is a permutation, there exists $k > 0$ such that $\pi^k(x) = x$ for every $x \in \mathcal{X}$. Now for every $0 \leq i < kn$, let $X_i \in \mathcal{A}^{i*}$ be such that $a \in X_i$ and let $\mathfrak{X} = (X_i)_{0 \leq i < kn}$. We have:

- $A * \mathfrak{X} \in \mathcal{A}^{kn*}$.
- $A \subset A * \mathfrak{X}$ since $\pi^{kn}(x) = x$ for every $x \in \mathcal{X}$.
- $\mathcal{A}^{kn*} = (\mathcal{A}^{n*})^{(k-1)n*} = \mathcal{A}^{n*}$ since $\mathrm{per}(\mathcal{A}^{n*})$ divides n.

We conclude that $A \subset A * \mathfrak{X} \in \mathcal{A}^{n*}$. Therefore, $\mathcal{A} \preceq \mathcal{A}^{n*}$. On the other hand, Proposition 2.9 implies that $\mathcal{A}^{n*} \preceq \mathrm{core}(\mathcal{A}^{n*})$. Therefore, $\mathcal{A} \preceq \mathrm{core}(\mathcal{A}^{n*})$.

Now since $\mathrm{core}(\mathcal{A}^{n*})$ is a stable partition (hence it is also periodic), we must have $\langle \mathcal{A} \rangle \preceq \mathrm{core}(\mathcal{A}^{n*})$ by Proposition 2.6. On the other hand, we have:

- Since $\mathcal{A} \preceq \langle \mathcal{A} \rangle$ then $\mathcal{A}^{np*} \preceq \langle \mathcal{A} \rangle^{np*}$, where $p = \mathrm{per}(\langle \mathcal{A} \rangle)$.
- $\mathcal{A}^{np*} = (\mathcal{A}^{n*})^{(p-1)n*} \overset{(a)}{=} \mathcal{A}^{n*}$, where (a) follows from the fact that $\mathrm{per}(\mathcal{A}^{n*})$ divides n.
- $\langle \mathcal{A} \rangle^{np*} = \langle \mathcal{A} \rangle$ since $p = \mathrm{per}(\langle \mathcal{A} \rangle)$.

Therefore, $\mathcal{A}^{n*} \preceq \langle \mathcal{A} \rangle$. But $\mathrm{core}(\mathcal{A}^{n*}) \subset \mathcal{A}^{n*}$, which implies that $\mathrm{core}(\mathcal{A}^{n*}) \preceq \mathcal{A}^{n*}$, hence $\mathrm{core}(\mathcal{A}^{n*}) \preceq \langle \mathcal{A} \rangle$. We conclude that $\mathrm{core}(\mathcal{A}^{n*}) = \langle \mathcal{A} \rangle$ as we have already shown that $\langle \mathcal{A} \rangle \preceq \mathrm{core}(\mathcal{A}^{n*})$. □

Remark 2.16. *The strong ergodicity condition in Theorem 2.3 cannot be replaced by ergodicity. Consider the ergodic operation $*$ of Remark 2.14, and consider the the $\mathcal{X}$-cover*

$$\mathcal{A} = \big\{\{0,1\},\{0,2\},\{1,2\},\{3,4\},\{3,5\},\{4,5\}\big\},$$

which is not a partition. We have the following:

- *It is easy to see that* $\mathrm{core}(\mathcal{A}^{n*}) = \mathcal{A}^{n*} = \mathcal{A}$ *for every* $n \geq 0$.
- *Since $\mathcal{A}$ is not a partition,* $\mathrm{core}(\mathcal{A}^{n*}) = \mathcal{A}$ *is not a partition for any* $n \geq 0$.

Therefore, $\mathrm{core}(\mathcal{A}^{n*}) \neq \langle \mathcal{A} \rangle^{n*}$ *for every* $n \geq 0$.

2.7 Product of Binary Operations

Definition 2.18. *Let $\mathcal{X}_1, \ldots, \mathcal{X}_m$ be m sets, and let $*_1, \ldots, *_m$ be m binary operations on $\mathcal{X}_1, \ldots, \mathcal{X}_m$ respectively. We define the product of $*_1, \ldots, *_m$, denoted $* = *_1 \otimes \ldots \otimes *_m$, as the binary operation $*$ on $\mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ defined by:*

$$(x_1, x_2, \ldots, x_m) * (x'_1, x'_2, \ldots, x'_m) = (x_1 *_1 x'_1, x_2 *_2 x'_2, \ldots, x_m *_m x'_m).$$

Proposition 2.10. *Let $*_1, \ldots, *_m$ be m binary operations on $\mathcal{X}_1, \ldots, \mathcal{X}_m$ respectively. Let $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ and $* = *_1 \otimes \ldots \otimes *_m$. We have:*

1. *$*$ is uniformity-preserving if and only if $*_1, \ldots, *_m$ are uniformity-preserving.*

2. *If $*$ is irreducible then $*_1, \ldots, *_m$ are irreducible. The converse is not necessarily true.*

3. *$*$ is ergodic if and only if $*_1, \ldots, *_m$ are ergodic. Moreover,*

$$\operatorname{con}(*) = \max\{\operatorname{con}(*_1), \ldots, \operatorname{con}(*_m)\}.$$

Proof. 1) Suppose that $*_1, \ldots, *_m$ are uniformity-preserving. Fix

$$b = (b_1, \ldots, b_m) \in \mathcal{X}$$

and define the mapping $\pi_b : \mathcal{X} \to \mathcal{X}$ as $\pi_b(x) = x * b$ for all $x \in \mathcal{X}$. Now let $y = (y_1, \ldots, y_m) \in \mathcal{X}$. For every $1 \leq i \leq m$, $*_i$ is uniformity-preserving and so there exists $x_i \in \mathcal{X}_i$ such that $x_i *_i b_i = y_i$. Define $x = (x_1, \ldots, x_m)$. We have $\pi_b(x) = x * b = y$. Therefore, π_b is surjective which implies that it is bijective. Since this is true for every $b \in \mathcal{X}$, $*$ is uniformity-preserving.

Conversely, suppose that $*$ is uniformity-preserving and let $1 \leq i \leq m$. Fix $b_i \in \mathcal{X}_i$ and define the mapping $\pi_{b_i} : \mathcal{X}_i \to \mathcal{X}_i$ as $\pi_{b_i}(x_i) = x_i *_i b_i$ for all $x_i \in \mathcal{X}_i$. Now let $y_i \in \mathcal{X}_i$ and choose arbitrarily $y_j \in \mathcal{X}_j$ for each $j \neq i$. Define $y = (y_1, \ldots, y_m) \in \mathcal{X}$. Since $*$ is uniformity-preserving, there exists $x = (x_1, \ldots, x_m) \in \mathcal{X}$ such that $y = x * b$ which implies that $y_i = x_i *_i b_i$. Therefore, π_{b_i} is surjective which implies that it is bijective. Since this is true for every $b_i \in \mathcal{X}_i$, $*_i$ is uniformity-preserving.

2) Suppose that $*$ is irreducible and fix $1 \leq i \leq m$. Let $a_i, b_i \in \mathcal{X}_i$ and choose arbitrarily $a_j, b_j \in \mathcal{X}_j$ for each $j \neq i$. Define $a = (a_1, \ldots, a_m) \in \mathcal{X}$ and $b = (b_1, \ldots, b_m) \in \mathcal{X}$. Since $*$ is irreducible, a is $*$-connectable to b and so there exists $l > 0$ and $x_0, \ldots, x_{l-1} \in \mathcal{X}$ such that $b = (\ldots((a * x_0) * x_1) \ldots * x_{l-1})$. For each $0 \leq k < l$, let $x_k = (x_{1,k}, \ldots, x_{m,k})$ and so $x_{i,k} \in \mathcal{X}_i$. It is easy to see that we have $b_i = (\ldots((a_i *_i x_{i,0}) *_i x_{i,1}) \ldots *_i x_{i,l-1})$. Therefore, a_i is $*_i$-connectable to b_i for all $a_i, b_i \in \mathcal{X}_i$, hence $*_i$ is irreducible.

In order to see that the converse is not necessarily true, let $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$ and define $x *_1 y = x *_2 y = x \oplus 1$ for every $x, y \in \{0, 1\}$. It is easy to see that $*_1$ and $*_2$ are irreducible and $\operatorname{per}(*_1) = \operatorname{per}(*_2) = 2$. Let $* = *_1 \otimes *_2$. It is easy to see that $(0, 0)$ is not $*$-connectable to $(0, 1)$. Therefore, $*$ is not irreducible.

3) Suppose that $*_1, \ldots, *_m$ are ergodic and let

$$d = \max\{\operatorname{con}(*_1), \ldots, \operatorname{con}(*_m)\}.$$

Let $a = (a_1, \ldots, a_m) \in \mathcal{X}$ and $b = (b_1, \ldots, b_m) \in \mathcal{X}$. For each $1 \leq i \leq m$, since $d \geq \operatorname{con}(*_i)$ there exist $x_{i,0}, \ldots, x_{i,d-1} \in \mathcal{X}_i$ such that $b_i = (\ldots((a_i *_i x_{i,0}) *_i x_{i,1}) \ldots *_i x_{i,d-1})$. For each $0 \leq k < d$ define $x_k = (x_{1,k}, \ldots, x_{m,k}) \in \mathcal{X}$. It is easy to see that $b = (\ldots((a * x_0) * x_1) \ldots * x_{d-1})$. Therefore, all the elements of $\mathcal{X}$ are $*$-connectable to each other in d steps. We conclude that $*$ is ergodic and $\operatorname{con}(*) \leq d = \max\{\operatorname{con}(*_1), \ldots, \operatorname{con}(*_m)\}$.

Conversely, suppose that $*$ is ergodic and let $1 \leq i \leq m$. Let $a_i, b_i \in \mathcal{X}_i$ and choose arbitrarily $a_j, b_j \in \mathcal{X}_j$ for each $j \neq i$. Define $a = (a_1, \ldots, a_m) \in \mathcal{X}$ and $b = (b_1, \ldots, b_m) \in \mathcal{X}$. Since $*$ is ergodic, a is $*$-connectable to b in $\operatorname{con}(*)$ steps. It follows that a_i is $*_i$-connectable to b_i in $\operatorname{con}(*)$ steps (we use the same argument that we used for the irreducible case). Since this is true for every $a_i, b_i \in \mathcal{X}_i$, we conclude that

$*_i$ is ergodic and $\text{con}(*_i) \leq \text{con}(*)$. We conclude that $\max\{\text{con}(*_1), \ldots, \text{con}(*_m)\} \leq \text{con}(*)$ which implies that

$$\text{con}(*) = \max\{\text{con}(*_1), \ldots, \text{con}(*_m)\}$$

since we have $\text{con}(*) \leq \max\{\text{con}(*_1), \ldots, \text{con}(*_m)\}$ from the previous paragraph. □

Definition 2.19. *Let $\mathcal{H}_1, \ldots, \mathcal{H}_m$ be m stable partitions of $(\mathcal{X}_1, *_1), \ldots, (\mathcal{X}_m, *_m)$ respectively. The product of $\mathcal{H}_1, \ldots, \mathcal{H}_m$, denoted $\mathcal{H} = \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_m$ is defined as*

$$\mathcal{H} = \{A_1 \times \ldots \times A_m : \ A_1 \in \mathcal{H}_1, \ldots, A_m \in \mathcal{H}_m\}.$$

It is easy to see that $\mathcal{H}$ is a stable partition of $(\mathcal{X}_1 \times \cdots \times \mathcal{X}_m, *_1 \otimes \ldots \otimes *_m)$ of period $\text{per}(\mathcal{H}) = \text{lcm}\{\text{per}(\mathcal{H}_1), \ldots, \text{per}(\mathcal{H}_m)\}$.

Theorem 2.4. *Let $*_1$ and $*_2$ be two ergodic operations on $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively. Let $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ and $* = *_1 \otimes *_2$ (thus, $*$ is ergodic). Let $\mathcal{H}$ be a stable partition of $\mathcal{X}$. There exist two unique stable partitions $\mathcal{L}_1 := \mathcal{L}_1(\mathcal{H})$ and $\mathcal{U}_1 := \mathcal{U}_1(\mathcal{H})$ of $\mathcal{X}_1$ and two unique stable partitions $\mathcal{L}_2 := \mathcal{L}_2(\mathcal{H})$ and $\mathcal{U}_2 := \mathcal{U}_2(\mathcal{H})$ of $\mathcal{X}_2$ such that:*

- *$\mathcal{L}_1 \preceq \mathcal{U}_1$, $\mathcal{L}_2 \preceq \mathcal{U}_2$ and $\frac{\|\mathcal{U}_1\|}{\|\mathcal{L}_1\|} = \frac{\|\mathcal{U}_2\|}{\|\mathcal{L}_2\|} = n$ for some integer $n > 0$.*
- *$\mathcal{L}_1 \otimes \mathcal{L}_2 \preceq \mathcal{H} \preceq \mathcal{U}_1 \otimes \mathcal{U}_2$.*
- *For every $H \in \mathcal{H}$, there exist n disjoint sets $H_{1,1}, \ldots, H_{1,n} \in \mathcal{L}_1$ and n disjoint sets $H_{2,1}, \ldots, H_{2,n} \in \mathcal{L}_2$ such that:*
  - *$H_{1,1} \cup \ldots \cup H_{1,n} \in \mathcal{U}_1$.*
  - *$H_{2,1} \cup \ldots \cup H_{2,n} \in \mathcal{U}_2$.*
  - *$H = (H_{1,1} \times H_{2,1}) \cup \ldots \cup (H_{1,n} \times H_{2,n})$.*

  Therefore, $\|\mathcal{H}\| = n \cdot \|\mathcal{L}_1\| \cdot \|\mathcal{L}_2\| = \|\mathcal{L}_1\| \cdot \|\mathcal{U}_2\| = \|\mathcal{U}_1\| \cdot \|\mathcal{L}_2\|$.

The integer n is called the correlation *of $\mathcal{H}$ and is denoted by $\text{cor}_{*_1,*_2}(\mathcal{H})$.*

We also have $\mathcal{L}_1(\mathcal{H})^{i*_1} = \mathcal{L}_1(\mathcal{H}^{i*})$, $\mathcal{L}_2(\mathcal{H})^{i*_2} = \mathcal{L}_2(\mathcal{H}^{i*})$, $\mathcal{U}_1(\mathcal{H})^{i*_1} = \mathcal{U}_1(\mathcal{H}^{i*})$ and $\mathcal{U}_2(\mathcal{H})^{i*_2} = \mathcal{U}_2(\mathcal{H}^{i*})$ for every $i \geq 0$.

Proof. See Appendix 2.8.6. □

Remark 2.17. *If $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$, then $\mathcal{L}_1(\mathcal{H}) = \mathcal{U}_1(\mathcal{H}) = \mathcal{H}_1$, $\mathcal{L}_2(\mathcal{H}) = \mathcal{U}_2(\mathcal{H}) = \mathcal{H}_2$ and $\text{cor}_{*_1,*_2}(\mathcal{H}) = 1$.*

Example 2.2. *Figure 2.1 shows an element H of a stable partition $\mathcal{H}$ of correlation $n = \text{cor}_{*_1,*_2}(\mathcal{H}) = 3$. H is represented by the regions that are enclosed in thick lines.*

Example 2.3. *Let $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$ and define $*_1$ and $*_2$ as $x *_1 y = x *_2 y = x \oplus y$ for every $x, y \in \{0, 1\}$. Let $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$, $* = *_1 \otimes *_2$ and $\mathcal{H} = \big\{\{(0,0),(1,1)\}, \{(0,1),(1,0)\}\big\}$. It is easy to see that $\mathcal{H}$ is a stable partition of $(\mathcal{X}, *)$. We have:*

- *$\mathcal{L}_1(\mathcal{H}) = \mathcal{L}_2(\mathcal{H}) = \big\{\{0\}, \{1\}\big\}$.*

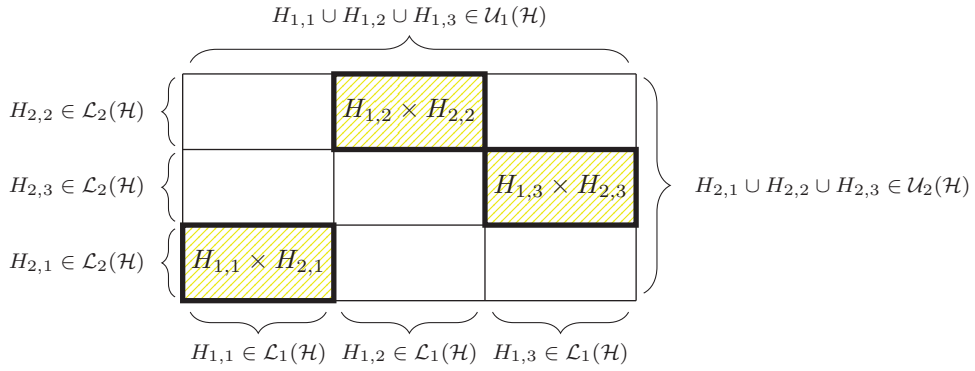

Figure 2.1 – $H = (H_{1,1} \times H_{2,1}) \cup (H_{1,2} \times H_{2,2}) \cup (H_{1,3} \times H_{2,3}) \in \mathcal{H}$.

- $\mathcal{U}_1(\mathcal{H}) = \mathcal{U}_2(\mathcal{H}) = \big\{\{0,1\}\big\}$.
- $n = \mathrm{cor}_{*_1,*_2}(\mathcal{H}) = 2$.

For $H = \{(0,1),(1,0)\} \in \mathcal{H}$, we have:

- $H_{1,1} = \{0\}$, $H_{1,2} = \{1\}$ *and* $H_{1,1} \cup H_{1,2} = \{0,1\} \in \mathcal{U}_1(\mathcal{H})$.
- $H_{2,1} = \{1\}$, $H_{2,2} = \{0\}$ *and* $H_{2,1} \cup H_{2,2} = \{0,1\} \in \mathcal{U}_2(\mathcal{H})$.
- $(H_{1,1} \times H_{2,1}) \cup (H_{1,2} \times H_{2,2}) = \{(0,1),(1,0)\} = H$.

Theorem 2.4 shows that the stable partitions of the product of two ergodic operations have a very particular structure. This structure will be useful to prove the following theorem:

Theorem 2.5. *Let $*_1, \ldots, *_m$ be $m \geq 2$ binary operations on $\mathcal{X}_1, \ldots, \mathcal{X}_m$ respectively. Let $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ and $* = *_1 \otimes \ldots \otimes *_m$. Then $*$ is strongly ergodic if and only if $*_1, \ldots, *_m$ are strongly ergodic.*

Proof. See Appendix 2.8.6. □

Notation 2.5. *Let $*_1, \ldots, *_m$ be $m \geq 2$ ergodic operations on $\mathcal{X}_1, \ldots, \mathcal{X}_m$ respectively. Define $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ and $* = *_1 \otimes \ldots \otimes *_m$. Let A and B be two non-empty subsets of $I_m := \{1, \ldots, m\}$ which form a non-trivial partition (i.e., $A \cup B = I_m$, $A \cap B = \emptyset$, $A \neq \emptyset$ and $B \neq \emptyset$). Let $i_1 < \ldots < i_{|A|}$ and $j_1 < \ldots < j_{|B|}$ be such that $A = \{i_1, \ldots, i_{|A|}\}$ and $B = \{j_1, \ldots, j_{|B|}\}$. Define $\mathcal{X}_A = \mathcal{X}_{i_1} \times \cdots \times \mathcal{X}_{i_{|A|}}$, $\mathcal{X}_B = \mathcal{X}_{j_1} \times \cdots \times \mathcal{X}_{j_{|B|}}$, $*_A = *_{i_1} \otimes \ldots \otimes *_{i_{|A|}}$ and $*_B = *_{j_1} \otimes \ldots \otimes *_{j_{|B|}}$. Define the mapping $f_{A,B} : \mathcal{X} \to \mathcal{X}_A \times \mathcal{X}_B$ as*

$$f_{A,B}(x_1, \ldots, x_m) = \big((x_{i_1}, \ldots, x_{i_{|A|}}), (x_{j_1}, \ldots, x_{j_{|B|}})\big).$$

Clearly, $f_{A,B}$ is a bijection. We call $f_{A,B}$ the canonical bijection *between $\mathcal{X}$ and $\mathcal{X}_A \times \mathcal{X}_B$. Throughout this chapter, we identify $(\mathcal{X}, *)$ with $(\mathcal{X}_A \times \mathcal{X}_B, *_A \otimes *_B)$ through the canonical bijection $f_{A,B}$.*

Let $\mathcal{H}$ be a stable partition of $(\mathcal{X}, *)$. Since $*_A$ and $*_B$ are ergodic, there are two unique stable partitions $\mathcal{L}_A(\mathcal{H}) \preceq \mathcal{U}_A(\mathcal{H})$ of $(\mathcal{X}_A, *_A)$ and two unique stable partitions

$\mathcal{L}_B(\mathcal{H}) \preceq \mathcal{U}_B(\mathcal{H})$ *of* $(\mathcal{X}_B, *_B)$ *and* $n_A = \mathrm{cor}_{*_A,*_B}(\mathcal{H}) = \mathrm{cor}_{*_B,*_A}(\mathcal{H}) = n_B > 0$ *satisfying the conditions of Theorem 2.4. We adopt the convention that* $\mathcal{U}_{I_m}(\mathcal{H}) = \mathcal{H}$.

If $A = \{i\}$ *contains only one element* i, *we denote* $\mathcal{L}_{\{i\}}(\mathcal{H})$ *and* $\mathcal{U}_{\{i\}}(\mathcal{H})$ *as* $\mathcal{L}_i(\mathcal{H})$ *and* $\mathcal{U}_i(\mathcal{H})$ *respectively.*

Notation 2.6. *For each* $A \subset B \subset I_m = \{1, \ldots, m\}$ *we define the mapping* $P_{B\to A} : \mathcal{X}_B \to \mathcal{X}_A$ *as* $P_{B\to A}(x_{j_1}, \ldots, x_{j_{|B|}}) = (x_{i_1}, \ldots, x_{i_{|A|}})$, *where* $A = \{i_1, \ldots, i_{|A|}\} \subset \{j_1, \ldots, j_{|B|}\} = B$, $i_1 < \ldots < i_{|A|}$ *and* $j_1 < \ldots < j_{|B|}$. *If* A *contains only one element* i, *we denote* $P_{B\to\{i\}}$ *by* $P_{B\to i}$.

Now for each $A \subsetneq B \subset I_m = \{1, \ldots, m\}$, *each* $x_{B\setminus A} \in \mathcal{X}_{B\setminus A}$ *and each* $X_B \subset \mathcal{X}_B$, *we define the set* $P_{B\to A|x_{B\setminus A}}(X_B) := \{x_A \in \mathcal{X}_A : \ (x_A, x_{B\setminus A}) \in X_B\} \subset \mathcal{X}_A$. *If* A *contains only one element* i, *we denote* $P_{B\to\{i\}|x_{B\setminus\{i\}}}(X_B)$ *by* $P_{B\to i|x_{B\setminus i}}(X_B)$.

It is easy to see that if $A \subset B \subset C \subset \{1, \ldots, m\}$ *then we have* $P_{B\to A} \circ P_{C\to B} = P_{C\to A}$. *Similarly, if* $A \subsetneq B \subsetneq C \subset \{1, \ldots, m\}$, *then for each* $X_C \subset \mathcal{X}_C$, *each* $x_{C\setminus B} \in \mathcal{X}_{C\setminus B}$ *and each* $x_{B\setminus A} \in \mathcal{X}_{B\setminus A}$, *we have*

$$P_{B\to A|x_{B\setminus A}}\big(P_{C\to B|x_{C\setminus B}}(X_C)\big) = P_{C\to A|(x_{C\setminus B}, x_{B\setminus A})}(X_C).$$

Here we have $(x_{C\setminus B}, x_{B\setminus A}) \in \mathcal{X}_{C\setminus A}$ *since we are identifying* $\mathcal{X}_{C\setminus A}$ *with* $\mathcal{X}_{C\setminus B} \times \mathcal{X}_{B\setminus A}$ *through the canonical bijection.*

Remark 2.18. *Let* $\mathcal{H}$ *be a stable partition of* $(\mathcal{X}, *) = (\mathcal{X}_1 \times \cdots \times \mathcal{X}_m, *_1 \otimes \ldots \otimes *_m)$, *where* $*$ *is ergodic. If* $A \subset I_m = \{1, \ldots, m\}$, *we have from Definition 2.25:*

$$\mathcal{U}_A(\mathcal{H}) = \{P_{I_m\to A}(H) : \ H \in \mathcal{H}\}.$$

Furthermore, if $A \subsetneq I_m = \{1, \ldots, m\}$, *we have from Definition 2.26:*

$$\begin{aligned}\mathcal{L}_A(\mathcal{H}) &= \{P_{I_m\to A|x_{I_m\setminus A}}(H) : \ H \in \mathcal{H}, \ x_{I_m\setminus A} \in \mathcal{X}_{I_m\setminus A}, \ P_{I_m\to A|x_{I_m\setminus A}}(H) \neq \emptyset\} \\ &\overset{(a)}{=} \{P_{I_m\to A|x_{I_m\setminus A}}(H) : \ H \in \mathcal{H}, \ x_{I_m\setminus A} \in P_{I_m\to I_m\setminus A}(H)\}.\end{aligned}$$

(a) follows from the fact that $P_{I_m\to A|x_{I_m\setminus A}}(H) \neq \emptyset$ *if and only if*

$$x_{I_m\setminus A} \in P_{I_m\to I_m\setminus A}(H).$$

Proposition 2.11. *Let* $*_1, \ldots, *_m$ *be* $m \geq 2$ *ergodic operations on* $\mathcal{X}_1, \ldots, \mathcal{X}_m$ *respectively. Define* $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ *and* $* = *_1 \otimes \ldots \otimes *_m$. *Let* $\mathcal{H}$ *be a stable partition of* $(\mathcal{X}, *)$ *and* $A \subsetneq B \subsetneq I_m = \{1, \ldots, m\}$. *Then* $\mathcal{U}_A\big(\mathcal{U}_B(\mathcal{H})\big) = \mathcal{U}_A(\mathcal{H})$ *and* $\mathcal{L}_A\big(\mathcal{L}_B(\mathcal{H})\big) = \mathcal{L}_A(\mathcal{H})$.

Proof. From Remark 2.18 we have:

$$\begin{aligned}\mathcal{U}_A\big(\mathcal{U}_B(\mathcal{H})\big) &= \{P_{B\to A}(H_B) : \ H_B \in \mathcal{U}_B(\mathcal{H})\} \\ &= \big\{P_{B\to A}\big(P_{I_m\to B}(H)\big) : \ H \in \mathcal{H}\big\} \\ &= \{P_{I_m\to A}(H) : \ H \in \mathcal{H}\} = \mathcal{U}_A(\mathcal{H}).\end{aligned}$$

On the other hand, we have:

$$
\begin{aligned}
&\mathcal{L}_A\big(\mathcal{L}_B(\mathcal{H})\big)\\
&\overset{(a)}{=} \{P_{B\to A|x_{B\setminus A}}(H_B):\ H_B\in\mathcal{L}_B(\mathcal{H}),\ x_{B\setminus A}\in\mathcal{X}_{B\setminus A},\ P_{B\to A|x_{B\setminus A}}(H_B)\neq o\}\\
&\overset{(b)}{=} \Big\{P_{B\to A|x_{B\setminus A}}\big(P_{I_m\to B|x_{I_m\setminus B}}(H)\big):\\
&\qquad\qquad H\in\mathcal{H},\ x_{I_m\setminus B}\in\mathcal{X}_{I_m\setminus B},\ P_{I_m\to B|x_{I_m\setminus B}}(H)\neq\emptyset,\\
&\qquad\qquad x_{B\setminus A}\in\mathcal{X}_{B\setminus A},\ P_{B\to A|x_{B\setminus A}}\big(P_{I_m\to B|x_{I_m\setminus B}}(H)\big)\neq o\Big\}\\
&\overset{(c)}{=} \Big\{P_{B\to A|x_{B\setminus A}}\big(P_{I_m\to B|x_{I_m\setminus B}}(H)\big):\ H\in\mathcal{H},\ x_{I_m\setminus B}\in\mathcal{X}_{I_m\setminus B},\\
&\qquad\qquad x_{B\setminus A}\in\mathcal{X}_{B\setminus A},\ P_{B\to A|x_{B\setminus A}}\big(P_{I_m\to B|x_{I_m\setminus B}}(H)\big)\neq o\Big\}\\
&\overset{(d)}{=} \Big\{P_{I_m\to A|(x_{I_m\setminus B},x_{B\setminus A})}(H):\ H\in\mathcal{H},\ x_{I_m\setminus B}\in\mathcal{X}_{I_m\setminus B},\ x_{B\setminus A}\in\mathcal{X}_{B\setminus A},\\
&\qquad\qquad P_{I_m\to A|(x_{I_m\setminus B},x_{B\setminus A})}(H)\neq\emptyset\Big\}\\
&\overset{(e)}{=} \{P_{I_m\to A|x_{I_m\setminus A}}(H):\ H\in\mathcal{H},\ x_{I_m\setminus A}\in\mathcal{X}_{I_m\setminus A},\ P_{I_m\to A|x_{I_m\setminus A}}(H)\neq o\}\\
&= \mathcal{L}_A(\mathcal{H}).
\end{aligned}
$$

(a) and (b) follow from Remark 2.18. (c) follows from the fact that

$$P_{B\to A|x_{B\setminus A}}\big(P_{I_m\to B|x_{I_m\setminus B}}(H)\big)\neq\emptyset$$

entails $P_{I_m\to B|x_{I_m\setminus B}}(H)\neq\emptyset$. (d) follows from the fact that

$$P_{B\to A|x_{B\setminus A}}\big(P_{I_m\to B|x_{I_m\setminus B}}(H)\big)=P_{I_m\to A|(x_{I_m\setminus B},x_{B\setminus A})}(H).$$

(e) follows from the fact that $I_m\setminus A=(I_m\setminus B)\cup(B\setminus A)$ and so $\mathcal{X}_{I_m\setminus A}$ is identified to $\mathcal{X}_{I_m\setminus B}\times\mathcal{X}_{B\setminus A}$. □

Definition 2.20. *Let $*_1,\ldots,*_m$ be $m\geq 2$ ergodic operations on $\mathcal{X}_1,\ldots,\mathcal{X}_m$ respectively. Let $\mathcal{X}=\mathcal{X}_1\times\cdots\times\mathcal{X}_m$ and $*=*_1\otimes\ldots\otimes *_m$. Let $\mathcal{H}$ be a stable partition of $(\mathcal{X},*)$. The* canonical factorization *of $\mathcal{H}$ is the sequence $(\mathcal{H}_i)_{1\leq i\leq m}$ defined as:*

- $\mathcal{H}_m=\mathcal{U}_m(\mathcal{H})$.
- *For each $1\leq i<m$, $\mathcal{H}_i=\mathcal{U}_i\big(\mathcal{L}_{I_i}(\mathcal{H})\big)$, where $I_i=\{1,\ldots,i\}$.*

Lemma 2.4. *Let $*_1,\ldots,*_m$ be $m\geq 2$ ergodic operations on $\mathcal{X}_1,\ldots,\mathcal{X}_m$ respectively. Let $\mathcal{X}=\mathcal{X}_1\times\cdots\times\mathcal{X}_m$ and $*=*_1\otimes\ldots\otimes *_m$. Let $\mathcal{H}$ be a stable partition of $(\mathcal{X},*)$. If $(\mathcal{H}_i)_{1\leq i\leq m}$ is the canonical factorization of $\mathcal{H}$, then $(\mathcal{H}_i)_{1\leq i\leq m-1}$ is the canonical factorization of $\mathcal{L}_{I_{m-1}}(\mathcal{H})$, where $I_{m-1}=\{1,\ldots,m-1\}$.*

Proof. For each $1\leq i\leq m$, define $I_i=\{1,\ldots,i\}$. Let $\{\mathcal{H}'_i\}_{1\leq i\leq m-1}$ be the canonical factorization of $\mathcal{L}_{I_{m-1}}(\mathcal{H})$. We have:

- $\mathcal{H}'_{m-1}=\mathcal{U}_{m-1}\big(\mathcal{L}_{I_{m-1}}(\mathcal{H})\big)=\mathcal{H}_{m-1}$.

- For each $1 \leq i < m-1$, we have

$$\mathcal{H}'_i = \mathcal{U}_i\big(\mathcal{L}_{I_i}\big(\mathcal{L}_{I_{m-1}}(\mathcal{H})\big)\big) \overset{(a)}{=} \mathcal{U}_i\big(\mathcal{L}_{I_i}(\mathcal{H})\big) = \mathcal{H}_i,$$

  where (a) follows from Proposition 2.11.

□

Definition 2.21. *Let $\mathcal{H}$ be a partition of a set $\mathcal{X}$. A section of $\mathcal{H}$ is a subset $C \subset \mathcal{X}$ such that:*

- $|C| = \mathcal{H}$.
- *For each $H \in \mathcal{H}$, there exists a unique $x \in C$ such that $x \in H$. In other words, the mapping $\mathrm{Proj}_{\mathcal{H}}$, restricted to C, is a bijection between C and $\mathcal{H}$.*

Lemma 2.5. *Let $*_1$ and $*_2$ be two ergodic operations on $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively. Let $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ and $* = *_1 \otimes *_2$ (thus, $*$ is ergodic). Let $\mathcal{H}$ be a stable partition of $\mathcal{X}$. If C_1 and C_2 are sections of $\mathcal{L}_1(\mathcal{H})$ and $\mathcal{U}_2(\mathcal{H})$ respectively, then $C = C_1 \times C_2$ is a section of $\mathcal{H}$.*

Proof. Let $f_{C,\mathcal{H}} : C \to \mathcal{H}$ be the mapping $\mathrm{Proj}_{\mathcal{H}}$ restricted to C, i.e., $f_{C,\mathcal{H}}(x) = \mathrm{Proj}_{\mathcal{H}}(x)$ for every $x \in C$.

Let $H \in \mathcal{H}$ and $I_2 = \{1, 2\}$. We have $P_{I_2 \to 2}(H) \in \mathcal{U}_2(\mathcal{H})$ by Remark 2.18. Now since C_2 is a section of $\mathcal{U}_2(\mathcal{H})$, there exists a unique $x_2 \in C_2$ such that $x_2 \in P_{I_2 \to 2}(H)$.

Since $x_2 \in P_{I_2 \to 2}(H)$, we have $P_{I_2 \to 1|x_2}(H) \in \mathcal{L}_1(\mathcal{H})$ by Remark 2.18. But C_1 is a section of $\mathcal{L}_1(\mathcal{H})$, so there exists a unique $x_1 \in C_1$ such that $x_1 \in P_{I_2 \to 1|x_2}(H)$, which means that $(x_1, x_2) \in H$. Therefore, there exists $(x_1, x_2) \in C_1 \times C_2 = C$ such that $f_{C,\mathcal{H}}(x_1, x_2) = \mathrm{Proj}_{\mathcal{H}}(x_1, x_2) = H$. We conclude that $f_{C,\mathcal{H}}$ is surjective.

On the other hand, we have $|C| = |C_1 \times C_2| = |C_1| \cdot |C_2| = |\mathcal{L}_1(\mathcal{H})| \cdot |\mathcal{U}_2(\mathcal{H})| = |\mathcal{H}|$, where the last equality follows from Theorem 2.4. Therefore, $f_{C,\mathcal{H}}$ is bijective since $f_{C,\mathcal{H}} : C \to \mathcal{H}$ is surjective and $|C| = |\mathcal{H}|$. Hence, $C = C_1 \times C_2$ is a section of $\mathcal{H}$. □

Proposition 2.12. *Let $*_1, \ldots, *_m$ be $m \geq 2$ ergodic operations on $\mathcal{X}_1, \ldots, \mathcal{X}_m$ respectively. Let $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ and $* = *_1 \otimes \ldots \otimes *_m$. Let $\mathcal{H}$ be a stable partition of $(\mathcal{X}, *)$ and $(\mathcal{H}_i)_{1 \leq i \leq m}$ be the canonical factorization of $\mathcal{H}$. We have:*

- $|\mathcal{H}| = |\mathcal{H}_1| \times \cdots \times |\mathcal{H}_m|$.
- *If C_i is a section of $\mathcal{H}_i$ for every $1 \leq i \leq m$, then $C = C_1 \times \cdots \times C_m$ is a section of $\mathcal{H}$.*

Proof. For each $1 \leq i \leq m$, we define $I_i = \{1, \ldots, i\}$. We will prove the proposition by induction on m. If $m = 2$, we have:

- $\mathcal{H}_1 = \mathcal{U}_1\big(\mathcal{L}_{I_1}(\mathcal{H})\big) = \mathcal{U}_1\big(\mathcal{L}_1(\mathcal{H})\big) = \mathcal{L}_1(\mathcal{H})$ and $\mathcal{H}_2 = \mathcal{U}_2(\mathcal{H})$.
- By Theorem 2.4, we have $|\mathcal{H}| = |\mathcal{L}_1(\mathcal{H})| \cdot |\mathcal{U}_2(\mathcal{H})| = |\mathcal{H}_1| \cdot |\mathcal{H}_2|$.
- If C_1 and C_2 are sections of $\mathcal{H}_1 = \mathcal{L}_1(\mathcal{H})$ and $\mathcal{H}_2 = \mathcal{U}_2(\mathcal{H})$ respectively, then Lemma 2.5 shows that $C = C_1 \times C_2$ is a section of $\mathcal{H}$

Therefore the proposition is true for $m = 2$.

Now let $m > 2$ and suppose that the proposition is true for $m - 1$. By Lemma 2.4, $(\mathcal{H}_i)_{1 \leq i \leq m-1}$ is the canonical factorization of $\mathcal{L}_{I_{m-1}}(\mathcal{H})$. We have:

- $|\mathcal{H}| = |\mathcal{L}_{I_{m-1}}(\mathcal{H})| \cdot |\mathcal{U}_m(\mathcal{H})| = |\mathcal{L}_{I_{m-1}}(\mathcal{H})| \cdot |\mathcal{H}_m|$ by Theorem 2.4. On the other hand, we have $|\mathcal{L}_{I_{m-1}}(\mathcal{H})| = |\mathcal{H}_1| \times \cdots \times |\mathcal{H}_{m-1}|$ from the induction hypothesis. Therefore, $|\mathcal{H}| = |\mathcal{H}_1| \times \cdots \times |\mathcal{H}_m|$.

- For every $1 \leq i \leq m$, let C_i be a section of $\mathcal{H}_i$. From the induction hypothesis we get that $C_1 \times \cdots \times C_{m-1}$ is a section of $\mathcal{L}_{I_{m-1}}(\mathcal{H})$. Now since $C_1 \times \cdots \times C_{m-1}$ and C_m are sections of $\mathcal{L}_{I_{m-1}}(\mathcal{H})$ and $\mathcal{U}_m(\mathcal{H})$ respectively, Lemma 2.5 implies that $C = C_1 \times \cdots \times C_m$ is a section of $\mathcal{H}$.

Therefore, the proposition is also true for m. We conclude that the proposition is true for every $m \geq 2$. □

2.8 Appendix

2.8.1 Proof of Proposition 2.1

1) Trivial: For a quasigroup operation, all the elements of $\mathcal{X}$ are $*$-connectable to each other in one step.

2) Suppose that $*$ is uniformity-preserving but not irreducible. There exist two elements a_1 and a_2 of $\mathcal{X}$ such that a_1 is not $*$-connectable to a_2. Let $A_1 = \{x \in \mathcal{X} : \ a_1 \xrightarrow{*} x\}$ and $A_2 = \mathcal{X} \setminus A_1$. Clearly, $a_1 * a_1 \in A_1$ and $a_2 \in A_2$. Therefore, A_1 and A_2 are two disjoint non-empty sets such that $A_1 \cup A_2 = \mathcal{X}$. Moreover, we have $A_1 * \mathcal{X} \subset A_1$ from the definition of A_1. Now since $|A_1 * \mathcal{X}| \geq |A_1|$, we must have $A_1 * \mathcal{X} = A_1$.

For every $x \in \mathcal{X}$, define $\pi_x : \mathcal{X} \to \mathcal{X}$ as $\pi_x(a) = a * x$ for all $a \in \mathcal{X}$. Since $*$ is uniformity-preserving, π_x is bijective for all $x \in \mathcal{X}$. Therefore, $|\pi_x(A_1)| = |A_1|$. On the other hand, $\pi_x(A_1) = A_1 * x \subset A_1 * \mathcal{X} = A_1$. This means that $\pi_x(A_1) = A_1$, which implies that $\pi_x(A_2) = \pi_x(\mathcal{X} \setminus A_1) = \mathcal{X} \setminus \pi_x(A_1) = \mathcal{X} \setminus A_1 = A_2$ since π_x is bijective. Therefore, $A_2 * x = A_2$ for every $x \in \mathcal{X}$, hence $A_2 * \mathcal{X} = A_2$.

3) Suppose that $*$ is irreducible, and let $a, b \in \mathcal{X}$. Since $*$ is irreducible, there exist $l_1, l_2 \geq 0$ such that $a \xrightarrow{*,l_1} b$ and $b \xrightarrow{*,l_2} a$, so $a \xrightarrow{*,l_1+l_2} a$ which means that $\text{per}(*, a)$ divides $l_1 + l_2$. Now for any integer $l > 0$ satisfying $b \xrightarrow{*,l} b$, we have that $a \xrightarrow{*,l_1+l+l_2} a$. This shows that $\text{per}(*, a)$ divides $l_1 + l_2 + l$, which implies that $\text{per}(*, a)$ divides l since we have just shown that $\text{per}(*, a)$ divides $l_1 + l_2$. But this is true for every $l > 0$ that satisfies $b \xrightarrow{*,l} b$. We conclude that $\text{per}(*, a)$ divides $\text{per}(*, b)$. Similarly, we can show that $\text{per}(*, b)$ divides $\text{per}(*, a)$. Therefore, $\text{per}(*, a)$ is the same for all $a \in \mathcal{X}$. Now since $\text{per}(*) = \gcd\{\text{per}(*, a) : \ a \in \mathcal{X}\}$, we have $\text{per}(*) = \text{per}(*, a)$ for all $a \in \mathcal{X}$.

4) Suppose that $*$ is irreducible and let $n = \text{per}(*)$. Fix $a \in \mathcal{X}$ and for every $0 \leq i < n$, define $H_i = \left\{x \in \mathcal{X} : \ \exists l > 0, \ a \xrightarrow{*,l} x \text{ and } l \equiv i \bmod n\right\}$. We have the following:

- If $x \in \mathcal{X}$, then $a \overset{*,l_{a,x}}{\longrightarrow} x$ for some integer $l_{a,x} > 0$ because of irreducibility. This shows that for every $x \in \mathcal{X}$, we have $x \in H_{l_{a,x} \bmod n} \subset \bigcup_{i=0}^{n-1} H_i$. Therefore, $\mathcal{X} \subset \bigcup_{i=0}^{n-1} H_i \subset \mathcal{X}$, hence $\bigcup_{i=0}^{n-1} H_i = \mathcal{X}$.

- Let $x \in H_i$ and $y \in H_j$. We have $a \overset{*,l_{a,x}}{\longrightarrow} x$ for some $l_{a,x} > 0$ satisfying $l_{a,x} \equiv i \bmod n$. Moreover, $x \overset{*,l_{x,a}}{\longrightarrow} a$ for some $l_{x,a} > 0$, and so $a \overset{*,l_{a,x}+l_{x,a}}{\longrightarrow} a$. The definition of per$(*)$ implies that n divides $l_{a,x} + l_{x,a}$ and so $l_{x,a} \equiv -i \bmod n$. Now since $y \in H_j$, we have $a \overset{*,l_{a,y}}{\longrightarrow} y$ for some $l_{a,y} > 0$ satisfying $l_{a,y} \equiv j \bmod n$. We conclude that $x \overset{*,l_{x,y}}{\longrightarrow} y$, where $l_{x,y} = l_{x,a} + l_{a,y} \equiv j - i \bmod n$.

- Suppose there exist $i \neq j$ such that $H_i \cap H_j \neq \emptyset$ and let $x \in H_i \cap H_j$. From the previous paragraph we have $x \overset{*,l_{x,x}}{\longrightarrow} x$, where $l_{x,x} \equiv j - i \not\equiv 0 \bmod n$. The definition of per$(*)$ implies that n divides $l_{x,x}$ which is a contradiction since $l_{x,x} \not\equiv 0 \bmod n$. We conclude that $H_i \cap H_j = \emptyset$ for all $i \neq j$.

- For every $0 \leq i < n$ and every $y \in H_i * \mathcal{X}$, there exist $x \in H_i$ and $z \in \mathcal{X}$ such that $y = x * z$, which implies that $y \in H_{i+1 \bmod n}$. Therefore $H_i * \mathcal{X} \subset H_{i+1 \bmod n}$, and so $|H_{i+1 \bmod n}| \geq |H_i * \mathcal{X}| \geq |H_i|$. Thus, $|H_0| \geq |H_{n-1}| \geq \ldots \geq |H_1| \geq |H_0|$, which implies that $|H_0| = |H_1| = \ldots = |H_{n-1}|$.

Therefore, $\{H_0, \ldots, H_{n-1}\}$ is a partition of $\mathcal{X}$ satisfying $|H_0| = |H_1| = \ldots = |H_{n-1}|$.

Now let $0 \leq i < n$. We have shown that $H_i * \mathcal{X} \subset H_{i+1 \bmod n}$. On the other hand, we have $|H_i * \mathcal{X}| \geq |H_i| = |H_{i+1 \bmod n}|$. Therefore, $H_i * \mathcal{X} = H_{i+1 \bmod n}$.

5) For every $x \in \mathcal{X}$ and every $j > 0$ define

$$K_{x,j} = \left\{ y \in \mathcal{X} : \; x \overset{*,j}{\longrightarrow} y \right\}.$$

Since $K_{x,j+1} = K_{x,j} * \mathcal{X}$ and since the number of subsets of $\mathcal{X}$ is finite, there exists $d_x > 0$ such that the sequence $(K_{x,j})_{j \geq d_x}$ is periodic. Let per$_x$ be the period of $(K_{x,j})_{j \geq d_x}$. Now since $K_{x,j+1} = K_{x,j} * \mathcal{X}$, we have $|K_{x,j+1}| \geq |K_{x,j}|$. Therefore, the sequence $(|K_{x,j}|)_{j \geq d_x}$ is both periodic and non-decreasing, which implies that it is constant.

Fix $j \geq d_x$, and let $l > 0$ be such that $x \overset{*,l}{\longrightarrow} x$. For every $x' \in K_{x,j}$ we have $x \overset{*,j}{\longrightarrow} x'$ which implies that $x \overset{*,l+j}{\longrightarrow} x'$ (since $x \overset{*,l}{\longrightarrow} x$) and so $x' \in K_{x,j+l}$. Therefore, $K_{x,j} \subset K_{x,j+l}$, which implies that $K_{x,j} = K_{x,j+l}$ (since we know that $|K_{x,j}| = |K_{x,j+l}|$). Now since this is true for every $j \geq d_x$, we conclude that per$_x$ divides every $l > 0$ satisfying $x \overset{*,l}{\longrightarrow} x$. Therefore, per$_x$ divides $\gcd\{l > 0 : \; x \overset{*,l}{\longrightarrow} x\} = \text{per}(*, x) = n$. Hence,

$$K_{x,j} = K_{x,j+kn} \text{ for all } j \geq d_x \text{ and all } k \geq 0. \tag{2.2}$$

For every $x \in \mathcal{X}$, let i_x be the unique index $0 \leq i_x < n$ satisfying $x \in H_{i_x}$. Clearly, $K_{x,j} \subset H_{i_x+j \bmod n}$. Now let $x' \in K_{x,j}$ and $x'' \in H_{i_x+j \bmod n}$, where $j \geq d_x$. Since both x' and x'' are in $H_{i_x+j \bmod n}$, we know from the discussion of the fourth

point that we have $x' \overset{*,l_{x',x''}}{\longrightarrow} x''$ for some $l_{x',x''} \equiv 0 \bmod n$. Since n divides $l_{x',x''}$, we have $K_{x,j+l_{x',x''}} = K_{x,j}$ from (2.2). Now since $x' \in K_{x,j}$ and $x' \overset{*,l_{x',x''}}{\longrightarrow} x''$, we have $x'' \in K_{x,j+l_{x',x''}} = K_{x,j}$. But this is true for every $x'' \in H_{i_x+j \bmod n}$. Therefore, $H_{i_x+j \bmod n} \subset K_{x,j}$, which implies that $K_{x,j} = H_{i_x+j \bmod n}$ as we already have $K_{x,j} \subset H_{i_x+j \bmod n}$.

Define $d = \max_{x \in \mathcal{X}} d_x$. Let $0 \leq i < n$ and $x \in H_i$. We have $i_x = i$ (since $x \in H_i$) and $d \geq d_x$. Therefore, from the above discussion we have $H_{i+d \bmod n} = H_{i_x+d \bmod n} = K_{x,d}$. Hence, for every $y \in H_{i+d \bmod n}$, we have $y \in K_{x,d}$ and so $x \overset{*,d}{\longrightarrow} y$.

6) We will prove the claim by induction on $s \geq \mathrm{con}(*)$. If $s = \mathrm{con}(*)$, the claim follows from 5). Now let $s > \mathrm{con}(*)$ and suppose that the claim is true for $s-1$. Let $0 \leq i < n$, $x \in H_i$ and $y \in H_{i+s \bmod n}$. Since $H_{i+s \bmod n} = H_{i+s-1 \bmod n} * \mathcal{X}$, there exists $y' \in H_{i+s-1 \bmod n}$ such that $y' \overset{*,1}{\longrightarrow} y$. Now since $y' \in H_{i+s-1 \bmod n}$, it follows from the induction hypothesis that $x \overset{*,s-1}{\longrightarrow} y'$. Therefore, $x \overset{*,s}{\longrightarrow} y$.

7) Let $*$ be an irreducible operation of period $\mathrm{per}(*) = 1$. Let $\mathcal{E}_*$ be the partition defined in 4). Since $\mathrm{per}(*) = 1$, the partition $\mathcal{E}_*$ contains only one element H_0 which must be $\mathcal{X}$. Now 5) implies that there exists $d > 0$ such that any element of $\mathcal{X} = H_0$ is $*$-connectable to any element of $H_{0+d \bmod 1} = H_0 = \mathcal{X}$ in d steps. Therefore, $*$ is ergodic.

Conversely, if $*$ is ergodic, let $d = \mathrm{con}(*)$ and $n = \mathrm{per}(*)$. Define $\mathcal{E}_* = \{H_0, \ldots, H_{n-1}\}$ as in 4) and let $a \in H_0$. Since $a \overset{*,d}{\longrightarrow} x$ for all $x \in \mathcal{X}$, then $\mathcal{X} \subset H_{d \bmod n}$ which implies that $\mathcal{X} = H_{d \bmod n}$. Now since $|H_0| = \ldots = |H_{n-1}| = |H_{d \bmod n}| = |\mathcal{X}|$, then $H_0 = \ldots = H_{n-1} = \mathcal{X}$ and $\mathcal{E}_* = \{\mathcal{X}\}$. Therefore, $\mathrm{per}(*) = n = |\mathcal{E}_*| = 1$.

8) If $*$ is ergodic, then $\mathrm{per}(*) = 1$ by 7). Therefore, $\mathcal{E}_*$ contains only one element H_0 which must be $\mathcal{X}$. Now 6) implies that for every $s \geq \mathrm{con}(*)$, any element of $\mathcal{X} = H_0$ is $*$-connectable to any element of $H_{0+s \bmod 1} = H_0 = \mathcal{X}$ in s steps.

9) and 10) are trivial.

2.8.2 Proofs for Section 2.3

Proof of Proposition 2.2 (1). For every $k > 0$ and every sequence $H_0 \in \mathcal{H}$, $H_1 \in \mathcal{H}^*$, $\ldots$, $H_{k-1} \in \mathcal{H}^{(k-1)*}$, define

$$\mathcal{H}_{H_0,\ldots,H_{k-1}} := \big\{(\ldots((H * H_0) * H_1)\ldots * H_{k-1}) : \ H \in \mathcal{H}\big\}. \tag{2.3}$$

We have:

$$\begin{aligned}\bigcup_{X \in \mathcal{H}_{H_0,\ldots,H_{k-1}}} X &= \bigcup_{H \in \mathcal{H}} (\ldots((H * H_0) * H_1)\ldots * H_{k-1}) \\ &= \Big(\ldots\Big(\Big(\Big(\bigcup_{H \in \mathcal{H}} H\Big) * H_0\Big) * H_1\Big)\ldots * H_{k-1}\Big) \\ &= (\ldots((\mathcal{X} * H_0) * H_1)\ldots * H_{k-1}) = \mathcal{X}.\end{aligned}$$

Therefore, $\mathcal{H}_{H_0,\ldots,H_{k-1}}$ covers $\mathcal{X}$ for any sequence $H_0 \in \mathcal{H}$, $H_1 \in \mathcal{H}^*$, $\ldots$, $H_{k-1} \in \mathcal{H}^{(k-1)*}$. Moreover, it is easy to see from (2.3) that $\mathcal{H}_{H_0,\ldots,H_{k-1}} \subset \mathcal{H}^{k*}$, which implies that $\mathcal{H}^{k*}$ covers $\mathcal{X}$.

Fix $n > 0$ and suppose that $\mathcal{H}^{n*}$ is not a partition. Since we have shown that $\mathcal{H}^{n*}$ covers $\mathcal{X}$, there must exist $X_1, X_1' \in \mathcal{H}^{n*}$ such that $X_1 \cap X_1' \neq \emptyset$ and $X_1 \neq X_1'$. We may assume without loss of generality that $|X_1| \leq |X_1'|$. If $X_1' \setminus X_1 = \emptyset$ then $X_1' \subset X_1$ which implies that $X_1' = X_1$ (because $|X_1| \leq |X_1'|$) which is a contradiction. Therefore, we must have $X_1' \setminus X_1 \neq \emptyset$.

Since $X_1 \in \mathcal{H}^{n*}$, there exists $H \in \mathcal{H}$ and a sequence $H_0 \in \mathcal{H}$, $H_1 \in \mathcal{H}^*$, ..., $H_{n-1} \in \mathcal{H}^{(n-1)*}$ such that $X_1 = (\ldots((H * H_0) * H_1) \ldots * H_{n-1})$ which implies that $X_1 \in \mathcal{H}_{H_0,\ldots,H_{n-1}}$. Now since we have shown that $\mathcal{H}_{H_0,\ldots,H_{n-1}}$ covers $\mathcal{X}$ and since $X_1' \setminus X_1 \neq \emptyset$, there must exist $X_2 \in \mathcal{H}_{H_0,\ldots,H_{n-1}}$ such that $X_2 \cap (X_1' \setminus X_1) \neq \emptyset$. Clearly, $X_1 \neq X_2$ since $X_1 \cap (X_1' \setminus X_1) = \emptyset$ and $X_2 \cap (X_1' \setminus X_1) \neq \emptyset$.

Let $p > 0$ be the smallest multiple of $\text{per}(\mathcal{H})$ which is greater than n, i.e.,

$$p = \min\{k \cdot \text{per}(\mathcal{H}) : \ k > 0, \ k \cdot \text{per}(\mathcal{H}) > n\}.$$

We have $\mathcal{H}^{p*} = \mathcal{H}$ since $\text{per}(\mathcal{H})$ divides p. Fix $H_n \in \mathcal{H}^{n*}, H_{n+1} \in \mathcal{H}^{(n+1)*}, \ldots,$ $H_{p-1} \in \mathcal{H}^{(p-1)*}$ and define:

- $A = (\ldots((X_1 * H_n) * H_{n+1}) \ldots * H_{p-1}) \in \mathcal{H}^{p*} = \mathcal{H}$.
- $B = (\ldots((X_2 * H_n) * H_{n+1}) \ldots * H_{p-1}) \in \mathcal{H}^{p*} = \mathcal{H}$.
- $C = (\ldots((X_1' * H_n) * H_{n+1}) \ldots * H_{p-1}) \in \mathcal{H}^{p*} = \mathcal{H}$.

We have $X_1 \cap X_1' \neq \emptyset$ and $X_2 \cap X_1' \neq \emptyset$, which imply that $A \cap C \neq \emptyset$ and $B \cap C \neq \emptyset$. Now since A, B, C are members of $\mathcal{H}$ which is a partition (i.e., the elements of $\mathcal{H}$ are non-empty, disjoint and cover $\mathcal{X}$), we must have $A = B = C$. We conclude that

$$(\ldots((X_1 * H_n) * H_{n+1}) \ldots * H_{p-1}) \ = \ (\ldots((X_2 * H_n) * H_{n+1}) \ldots * H_{p-1}). \quad (2.4)$$

We have:

- $\mathcal{H}_{H_0,\ldots,H_{p-1}} \subset \mathcal{H}^{p*}$ from the definition of $\mathcal{H}_{H_0,\ldots,H_{p-1}}$ (see (2.3)). We have shown that $\mathcal{H}_{H_0,\ldots,H_{p-1}}$ covers $\mathcal{X}$ and we know that $\mathcal{H}^{p*} = \mathcal{H}$ is a partition. Therefore, we must have $\mathcal{H}_{H_0,\ldots,H_{p-1}} = \mathcal{H}^{p*} = \mathcal{H}$.
- The mapping $\mathcal{H}_{H_0,\ldots,H_{n-1}} \to \mathcal{H}_{H_0,\ldots,H_{p-1}}$ defined by $X \to (\ldots((X * H_n) * H_{n+1}) \ldots * H_{p-1})$ is surjective but not injective because of (2.4). This implies that $|\mathcal{H}_{H_0,\ldots,H_{p-1}}| < |\mathcal{H}_{H_0,\ldots,H_{n-1}}|$.
- The mapping $\mathcal{H} \to \mathcal{H}_{H_0,\ldots,H_{n-1}}$ defined by $H \to (\ldots((H * H_0) * H_1) \ldots * H_{n-1})$ is surjective. Therefore, $|\mathcal{H}_{H_0,\ldots,H_{n-1}}| \leq |\mathcal{H}|$.

We conclude that $|\mathcal{H}| = |\mathcal{H}_{H_0,\ldots,H_{p-1}}| < |\mathcal{H}_{H_0,\ldots,H_{n-1}}| \leq |\mathcal{H}|$ which is a contradiction. Therefore, $\mathcal{H}^{n*}$ must be a partition. On the other hand, we have, $(\mathcal{H}^{n*})^{\text{per}(\mathcal{H})*} = (\mathcal{H}^{\text{per}(\mathcal{H})*})^{n*} = \mathcal{H}^{n*}$ which implies that $\mathcal{H}^{n*}$ is a periodic partition of period $\text{per}(\mathcal{H}^{n*})$ $\leq \text{per}(\mathcal{H})$. But since $\mathcal{H} = \mathcal{H}^{p*} = (\mathcal{H}^{n*})^{(p-n)*}$, we must also have $\text{per}(\mathcal{H}) = \text{per}(\mathcal{H}^{p*}) \leq \text{per}(\mathcal{H}^{n*})$. Therefore, $\text{per}(\mathcal{H}^{n*}) = \text{per}(\mathcal{H})$ for every $n > 0$. □

Lemma 2.6. *Let $\mathcal{H}$ be a periodic partition of $(\mathcal{X}, *)$. For every $H_2 \in \mathcal{H}$, we have*

$$\mathcal{H}^* = \mathcal{H} * \{H_2\} = \{H_1 * H_2 : \ H_1 \in \mathcal{H}\}.$$

Proof. For every $H_2 \in \mathcal{H}$, we have:

$$\mathcal{X} = \mathcal{X} * H_2 = \Big(\bigcup_{H_1 \in \mathcal{H}} H_1 \Big) * H_2 = \bigcup_{H_1 \in \mathcal{H}} (H_1 * H_2).$$

Therefore, the set $\{H_1 * H_2 : \ H_1 \in \mathcal{H}\}$ covers $\mathcal{X}$ and it is a subset of $\mathcal{H}^*$ which is a partition of $\mathcal{X}$ by Proposition 2.2 (1). Therefore, we must have $\mathcal{H}^* = \{H_1 * H_2 : H_1 \in \mathcal{H}\}$. □

Proof of Proposition 2.2 (2). For every $l \geq 0$, Proposition 2.2 (1) shows that $\mathcal{H}^{l*}$ is a periodic partition. If we fix $H_2 \in \mathcal{H}^{l*}$, then we have $\mathcal{H}^{(l+1)*} = \{H_1 * H_2 : \ H_1 \in \mathcal{H}^{l*}\}$ by Lemma 2.6. Therefore,

$$|\mathcal{H}^{(l+1)*}| = \Big|\{H_1 * H_2 : \ H_1 \in \mathcal{H}^{l*}\}\Big| \leq \Big|\{H_1 : \ H_1 \in \mathcal{H}^{l*}\}\Big| = |\mathcal{H}^{l*}|. \tag{2.5}$$

Now fix $n > 0$ and let $p > 0$ be the smallest multiple of $\mathrm{per}(\mathcal{H})$ which is greater than n, i.e., $p = \min\{k \cdot \mathrm{per}(\mathcal{H}) : \ k > 0, \ k \cdot \mathrm{per}(\mathcal{H}) > n\}$. From (2.5) we have

$$|\mathcal{H}| = |\mathcal{H}^{p*}| \leq |\mathcal{H}^{(p-1)*}| \leq \ldots \leq |\mathcal{H}^{n*}| \leq \ldots \leq |\mathcal{H}|.$$

Therefore, $|\mathcal{H}^{n*}| = |\mathcal{H}|$ for every $n > 0$. □

Proof of Proposition 2.5. Since $\mathcal{H}_1$ and $\mathcal{H}_2$ are two partitions of $\mathcal{X}$, it is easy to see that $\mathcal{H}_1 \wedge \mathcal{H}_2$ is also a partition of $\mathcal{X}$. Now let $H_1, H_1' \in \mathcal{H}_1$ and $H_2, H_2' \in \mathcal{H}_2$. If $H_1 \cap H_2 \neq \emptyset$ and $H_1' \cap H_2' \neq \emptyset$, we have:

$$(H_1 \cap H_2) * (H_1' \cap H_2') \ \subset \ (H_1 * H_1') \cap (H_2 * H_2') \in \mathcal{H}_1^* \wedge \mathcal{H}_2^*. \tag{2.6}$$

Fix $H_1' \in \mathcal{H}_1$ and $H_2' \in \mathcal{H}_2$ such that $H_1' \cap H_2' \neq \emptyset$. Lemma 2.6 implies that $\mathcal{H}_1^* = \{H_1 * H_1' : \ H_1 \in \mathcal{H}_1\}$ and $\mathcal{H}_2^* = \{H_2 * H_2' : \ H_2 \in \mathcal{H}_2\}$. Since $\mathcal{H}_1^*$ and $\mathcal{H}_2^*$ are partitions of $\mathcal{X}$, we have:

$$|\mathcal{X}| = \sum_{A_1 \in \mathcal{H}_1^*, A_2 \in \mathcal{H}_2^*} |A_1 \cap A_2| = \sum_{H_1 \in \mathcal{H}_1, H_2 \in \mathcal{H}_2} |(H_1 * H_1') \cap (H_2 * H_2')|,$$

which implies that

$$|\mathcal{X}| \geq \sum_{\substack{H_1 \in \mathcal{H}_1, H_2 \in \mathcal{H}_2: \\ H_1 \cap H_2 \neq \emptyset}} |(H_1 * H_1') \cap (H_2 * H_2')| \tag{2.7}$$

$$\geq \sum_{\substack{H_1 \in \mathcal{H}_1, H_2 \in \mathcal{H}_2: \\ H_1 \cap H_2 \neq \emptyset}} |(H_1 \cap H_2) * (H_1' \cap H_2')|, \tag{2.8}$$

where (2.8) follows from (2.6). Now since $H_1' \cap H_2' \neq \emptyset$, we have

$$|(H_1 \cap H_2) * (H_1' \cap H_2')| \geq |H_1 \cap H_2|. \tag{2.9}$$

Therefore,

$$\sum_{\substack{H_1\in\mathcal{H}_1,H_2\in\mathcal{H}_2:\\ H_1\cap H_2\neq\emptyset}} |(H_1\cap H_2)*(H_1'\cap H_2')| \geq \sum_{\substack{H_1\in\mathcal{H}_1,H_2\in\mathcal{H}_2:\\ H_1\cap H_2\neq\emptyset}} |H_1\cap H_2|. \tag{2.10}$$

Now since $\mathcal{H}_1$ and $\mathcal{H}_2$ are two partitions of $\mathcal{X}$, we have

$$\sum_{\substack{H_1\in\mathcal{H}_1,H_2\in\mathcal{H}_2:\\ H_1\cap H_2\neq\emptyset}} |H_1\cap H_2| = |\mathcal{X}|. \tag{2.11}$$

We conclude that all the inequalities in (2.7), (2.8), (2.9) and (2.10) are in fact equalities because if one of them were a strict inequality, we would have a contradiction with (2.11). Therefore, for all $H_1 \in \mathcal{H}_1$ and $H_2 \in \mathcal{H}_2$ satisfying $H_1 \cap H_2 \neq \emptyset$, we have $|(H_1 \cap H_2) * (H_1' \cap H_2')| = |(H_1 * H_1') \cap (H_2 * H_2')|$. Equation (2.6) now implies that $(H_1 \cap H_2) * (H_1' \cap H_2') = (H_1 * H_1') \cap (H_2 * H_2')$. We conclude that for every $H_1, H_1' \in \mathcal{H}_1$ and $H_2, H_2' \in \mathcal{H}_2$ satisfying $H_1 \cap H_2 \neq \emptyset$ and $H_1' \cap H_2' \neq \emptyset$, we have $(H_1\cap H_2)*(H_1'\cap H_2') = (H_1*H_1')\cap(H_2*H_2') \in \mathcal{H}_1^*\wedge\mathcal{H}_2^*$. Hence $(\mathcal{H}_1\wedge\mathcal{H}_2)^* \subset \mathcal{H}_1^*\wedge\mathcal{H}_2^*$. We have the following:

- $(\mathcal{H}_1 \wedge \mathcal{H}_2)^*$ covers $\mathcal{X}$ since $\mathcal{H}_1 \wedge \mathcal{H}_2$ covers $\mathcal{X}$.
- $\mathcal{H}_1^* \wedge \mathcal{H}_2^*$ is a partition of $\mathcal{X}$.
- $(\mathcal{H}_1 \wedge \mathcal{H}_2)^* \subset \mathcal{H}_1^* \wedge \mathcal{H}_2^*$.

Therefore, we must have $(\mathcal{H}_1 \wedge \mathcal{H}_2)^* = \mathcal{H}_1^* \wedge \mathcal{H}_2^*$.

It follows by induction that $(\mathcal{H}_1 \wedge \mathcal{H}_2)^{n*} = \mathcal{H}_1^{n*} \wedge \mathcal{H}_2^{n*}$ for all $n \geq 0$. In particular, for $l = \mathrm{lcm}(\mathrm{per}(\mathcal{H}_1), \mathrm{per}(\mathcal{H}_2))$, we have $(\mathcal{H}_1 \wedge \mathcal{H}_2)^{l*} = \mathcal{H}_1^{l*} \wedge \mathcal{H}_2^{l*} = \mathcal{H}_1 \wedge \mathcal{H}_2$, which implies that $\mathcal{H}_1 \wedge \mathcal{H}_2$ is a periodic partition of period of at most $\mathrm{lcm}(\mathrm{per}(\mathcal{H}_1), \mathrm{per}(\mathcal{H}_2))$. □

2.8.3 Proof of Theorem 2.1

In order to prove Theorem 2.1, we need several lemmas:

Lemma 2.7. *For every stable partition $\mathcal{H}$, and for every $\mathcal{H}$-repeatable sequence $\mathfrak{X}$, there exists an integer $l > 0$ such that $\mathfrak{X}^l$ is $\mathcal{H}$-augmenting.*

Proof. Let $\mathfrak{X} = (X_i)_{0\leq i<k}$ and let $x_i \in X_i$ for $0 \leq i < k$. Consider the mapping $\pi : \mathcal{X} \to \mathcal{X}$ defined by $\pi(x) = (\ldots((x*x_0)*x_1)\ldots)*x_{k-1})$. Since π is a permutation, there exists an integer $l > 0$ such that $\pi^l(x) = x$ for all $x \in \mathcal{X}$. For every $A \subset \mathcal{X}$, we have $A = \pi^l(A) \subset A * \mathfrak{X}^l$. Therefore, $\mathfrak{X}^l$ is $\mathcal{H}$-augmenting. □

Definition 2.22. *Let $A \subset \mathcal{X}$. We say that an $\mathcal{H}$-augmenting sequence $\mathfrak{X}$* connects *A if for every $a \in A$ we have $A \subset a * \mathfrak{X}$.*

Lemma 2.8. *If there exists an $\mathcal{H}$-augmenting sequence that connects a set $A \subset \mathcal{X}$, then there exists $H \in \mathcal{H}$ such that $A \subset H$.*

Proof. Let $\mathfrak{X}$ be such an $\mathcal{H}$-augmenting sequence. Let $a \in A$ and $H' \in \mathcal{H}$ be such that $a \in H'$. Define $H = H' * \mathfrak{X} \in \mathcal{H}^{|\mathfrak{X}|*}$. Since $\mathfrak{X}$ is $\mathcal{H}$-augmenting, $|\mathfrak{X}|$ divides $\mathrm{per}(\mathcal{H})$ and so $\mathcal{H}^{|\mathfrak{X}|*} = \mathcal{H}$. Therefore, $H \in \mathcal{H}$. On the other hand, $\mathfrak{X}$ connects A, so we have $A \subset a * \mathfrak{X} \subset H' * \mathfrak{X} = H$. □

Lemma 2.9. *Let $x \in \mathcal{X}$ and let $\mathfrak{X}$ be an $\mathcal{H}$-augmenting sequence. For every $y \in x*\mathfrak{X}$, there exists an $\mathcal{H}$-augmenting sequence $\mathfrak{X}'$ which connects $\{x, y\}$.*

Proof. Let $y \in x*\mathfrak{X} = (\ldots((x*X_0)*X_1)\ldots)*X_{k-1})$. There exist $x_i \in X_i$ $(0 \leq i < k)$ such that $y = (\ldots((x * x_0) * x_1)\ldots) * x_{k-1})$. Define the mapping $\pi : \mathcal{X} \to \mathcal{X}$ as $\pi(a) = (\ldots((a*x_0)*x_1)\ldots)*x_{k-1})$ for every $a \in \mathcal{X}$. Clearly, π is a permutation. The fact that $y = \pi(x)$ implies that x and y belong to the same cycle of the permutation π. Therefore, there exists $s > 0$ such that $x = \pi^s(y)$. Let $\mathfrak{X}' = \mathfrak{X}^s$. It is easy to see that $\mathfrak{X}'$ is $\mathcal{H}$-augmenting. Moreover, we have:

- $x \in y * \mathfrak{X}'$ because $x = \pi^s(y)$, and $y \in y * \mathfrak{X}'$ because $\mathfrak{X}'$ is $\mathcal{H}$-augmenting. Therefore, $\{x, y\} \subset y * \mathfrak{X}'$.
- $y \in x * \mathfrak{X}$ by assumption and $x \in x * \mathfrak{X}$ since $\mathfrak{X}$ is $\mathcal{H}$-augmenting. Therefore, $\{x, y\} \subset x * \mathfrak{X}$. On the other hand, $x * \mathfrak{X} \subset (x * \mathfrak{X}) * \mathfrak{X}^{s-1}$ since $\mathfrak{X}^{s-1}$ is $\mathcal{H}$-augmenting. Hence $\{x, y\} \subset (x * \mathfrak{X}) * \mathfrak{X}^{s-1} = x * \mathfrak{X}'$.

We conclude that $\mathfrak{X}'$ connects $\{x, y\}$. □

Lemma 2.10. *If there exists an $\mathcal{H}$-augmenting sequence that connects a set $A \subset \mathcal{X}$, and if there exists an $\mathcal{H}$-augmenting sequence that connects another set $B \subset \mathcal{X}$ such that $A \cap B \neq \emptyset$, then there exists an $\mathcal{H}$-augmenting sequence that connects $A \cup B$.*

Proof. Let $\mathfrak{X}$ be an $\mathcal{H}$-augmenting sequence that connects A, and let $\mathfrak{X}'$ be an $\mathcal{H}$-augmenting sequence that connects B. Let $\mathfrak{X}'' = (\mathfrak{X}, \mathfrak{X}', \mathfrak{X})$ be the $\mathcal{H}$-repeatable sequence that is obtained by concatenating $\mathfrak{X}$, $\mathfrak{X}'$ and $\mathfrak{X}$. Clearly, $\mathfrak{X}''$ is $\mathcal{H}$-augmenting. Fix $x \in A \cap B$ and let $y \in A \cup B$. We have the following:

- If $y \in A$, then $A \subset y * \mathfrak{X}$. In particular, $x \in y * \mathfrak{X}$. Now since $x \in B$ and since $\mathfrak{X}'$ connects B, we have $B \subset x * \mathfrak{X}'$. Therefore, $B \subset (y * \mathfrak{X}) * \mathfrak{X}'$.
- If $y \in B$, then $y \in y * \mathfrak{X}$ since $\mathfrak{X}$ is $\mathcal{H}$-augmenting. Now since $y \in B$ and since $\mathfrak{X}'$ connects B, we have $B \subset y * \mathfrak{X}'$. Therefore, $B \subset (y * \mathfrak{X}) * \mathfrak{X}'$.

We conclude that for every $y \in A \cup B$, we have $B \subset (y * \mathfrak{X}) * \mathfrak{X}'$. This implies that:

- $B \subset ((y * \mathfrak{X}) * \mathfrak{X}') * \mathfrak{X} = y * \mathfrak{X}''$ since $\mathfrak{X}$ is $\mathcal{H}$-augmenting.
- Since $B \subset (y * \mathfrak{X}) * \mathfrak{X}'$, we have $x \in (y * \mathfrak{X}) * \mathfrak{X}'$. Now since $x \in A$ and since $\mathfrak{X}$ connects A, we have $A \subset x * \mathfrak{X}$. Therefore, $A \subset ((y * \mathfrak{X}) * \mathfrak{X}') * \mathfrak{X} = y * \mathfrak{X}''$.

We conclude that $A \cup B \subset y * \mathfrak{X}''$ for every $y \in A \cup B$. Hence $\mathfrak{X}''$ connects $A \cup B$. □

Definition 2.23. *For every stable partition $\mathcal{H}$ of $(\mathcal{X}, *)$, define the* connectivity relation *$R_{\mathcal{H}}$ of $\mathcal{H}$ on $\mathcal{X}$ as follows: $aR_{\mathcal{H}}b$ if and only if there exists an $\mathcal{H}$-augmenting sequence that connects $\{a, b\}$.*

Lemma 2.11. *For every stable partition $\mathcal{H}$, $R_{\mathcal{H}}$ is an equivalence relation.*

Proof. Clearly, $R_{\mathcal{H}}$ is symmetric. Lemma 2.10 shows that $R_{\mathcal{H}}$ is transitive. In order to show that $R_{\mathcal{H}}$ is reflexive, let $x \in \mathcal{X}$, and let $\mathfrak{X}$ be an arbitrary $\mathcal{H}$-repeatable sequence. Lemma 2.7 implies that there exists $l > 0$ such that $\mathfrak{X}^l$ is $\mathcal{H}$-augmenting. We have $x \in x*\mathfrak{X}^l$ and so $\mathfrak{X}^l$ connects $\{x\}$. Therefore, $xR_{\mathcal{H}}x$ for every $x \in \mathcal{X}$, hence $R_{\mathcal{H}}$ is reflexive. We conclude that $R_{\mathcal{H}}$ is an equivalence relation. □

Notation 2.7. *For every stable partition $\mathcal{H}$, we denote the set of equivalence classes of its connectivity relation $R_{\mathcal{H}}$ by $\mathcal{K}_{\mathcal{H}}$.*

Lemma 2.12. *Let $\mathcal{H}$ be a stable partition and let $K \in \mathcal{K}_{\mathcal{H}}$. We have:*

- *For every $x \in K$ and every $\mathcal{H}$-augmenting sequence $\mathfrak{X}'$, $x * \mathfrak{X}' \subset K$.*
- *There exists an $\mathcal{H}$-augmenting sequence $\mathfrak{X}$ satisfying $x * \mathfrak{X} = K$ for all $x \in K$.*

Proof. For every $K \in \mathcal{K}_{\mathcal{H}}$, every $x \in K$, every $\mathcal{H}$-augmenting sequence $\mathfrak{X}'$, and every $y \in x * \mathfrak{X}'$, we have $xR_{\mathcal{H}}y$ because of Lemma 2.9, so $y \in K$. This shows that $x * \mathfrak{X}' \subset K$.

Now fix $K \in \mathcal{K}_{\mathcal{H}}$ and let $K = \{a_1, \dots, a_r\}$ where $r = |K|$. For each $1 \leq i \leq r$, define $K_i := \{a_1, \dots, a_i\}$. Since $a_1 R_{\mathcal{H}} a_1$ there exists an $\mathcal{H}$-augmenting sequence that connects K_1. Now let $1 < i \leq r$ and suppose that there exists an $\mathcal{H}$-augmenting sequence that connects K_{i-1}. Since $a_{i-1} R_{\mathcal{H}} a_i$, there exists an $\mathcal{H}$-augmenting sequence that connects $\{a_{i-1}, a_i\}$. Now since $K_{i-1} \cap \{a_{i-1}, a_i\} = \{a_{i-1}\} \neq \emptyset$, Lemma 2.10 implies that there exists an $\mathcal{H}$-augmenting sequence that connects $K_{i-1} \cup \{a_{i-1}, a_i\} = K_i$, and so the claim is true for i. By induction we conclude that the claim is true for every $1 \leq i \leq r$. In particular, there exists an $\mathcal{H}$-augmenting sequence $\mathfrak{X}$ that connects $K_r = K$.

Let $x \in K$. Since $\mathfrak{X}$ connects K, we have $K \subset x * \mathfrak{X}$, which implies that $x * \mathfrak{X} = K$ as we already have $x * \mathfrak{X} \subset K$. □

Lemma 2.13. *If $*$ is an ergodic operation on $\mathcal{X}$, then for every stable partition $\mathcal{H}$, we have the following:*

- *$\mathcal{K}_{\mathcal{H}^{l*}}$ is a balanced partition and $\|\mathcal{K}_{\mathcal{H}^{l*}}\| = \|\mathcal{K}_{\mathcal{H}}\|$ for all $l \geq 0$.*
- *For every $l \geq 0$, $K_1 \in \mathcal{K}_{\mathcal{H}}$, $K_2 \in \mathcal{K}_{\mathcal{H}^{l*}}$, and every $a \in K_1$, there exists an $\mathcal{H}$-sequence $\mathfrak{X}_{a,K_2}$ such that $|\mathfrak{X}_{a,K_2}| \equiv l \bmod n$ and $K_2 = a * \mathfrak{X}_{a,K_2} = K_1 * \mathfrak{X}_{a,K_2}$.*

Proof. Let $K_1 \in \mathcal{K}_{\mathcal{H}}$, $l \geq 0$ and $K_2 \in \mathcal{K}_{\mathcal{H}^{l*}}$. Let $n = \mathrm{per}(\mathcal{H})$, $k_1 = \mathrm{con}(*)n + l$ and $k_2 = \mathrm{con}(*)n + (-l \bmod n)$. Choose $a \in K_1$ and $b \in K_2$. Since $*$ is ergodic and since $k_1 \geq \mathrm{con}(*)$ and $k_2 \geq \mathrm{con}(*)$, it follows from Proposition 2.1 that there exist $x_0, \dots, x_{k_1-1} \in \mathcal{X}$ such that $b = (\dots((a * x_0) * x_1) \dots * x_{k_1-1})$ and there exist $y_0, \dots, y_{k_2-1} \in \mathcal{X}$ such that $a = (\dots((b * y_0) * y_1) \dots * y_{k_2-1})$. Let $\mathfrak{X}_1 = (X_i)_{0 \leq i < k_1}$ and $\mathfrak{X}_2 = (Y_i)_{0 \leq i < k_2}$ be such that $x_i \in X_i \in \mathcal{H}^{i*}$ for $0 \leq i < k_1$ and $y_i \in Y_i \in \mathcal{H}^{(l+i)*}$ for $0 \leq i < k_2$. Clearly, $b \in a * \mathfrak{X}_1$ and $a \in b * \mathfrak{X}_2$. The concatenation $\mathfrak{X} = (\mathfrak{X}_1, \mathfrak{X}_2)$ is an $\mathcal{H}$-repeatable sequence since n divides $k_1 + k_2$. Lemma 2.7 implies that there exists an integer $s > 0$ such that $\mathfrak{X}^s$ is $\mathcal{H}$-augmenting. Lemma 2.12, applied to $\mathcal{K}_{\mathcal{H}^{l*}}$, implies the existence of an $\mathcal{H}^{l*}$-augmenting sequence $\mathfrak{X}'$ such that $b * \mathfrak{X}' = K_2$.

Consider the sequence $\mathfrak{X}'' = (\mathfrak{X}_1, \mathfrak{X}', \mathfrak{X}_2, \mathfrak{X}^{s-1})$. It is easy to see that $\mathfrak{X}''$ is $\mathcal{H}$-augmenting and so $K_1 \subset K_1 * \mathfrak{X}''$. On the other hand, since $\mathfrak{X}''$ is $\mathcal{H}$-augmenting, Lemma 2.12 shows that for every $x \in K_1$ we have $x * \mathfrak{X}'' \subset K_1$, which means that $K_1 * \mathfrak{X}'' \subset K_1$. Therefore, $K_1 = K_1 * \mathfrak{X}''$. Moreover, since $b \in a * \mathfrak{X}_1$ and $b * \mathfrak{X}' = K_2$, we have

$$K_2 \subset (a * \mathfrak{X}_1) * \mathfrak{X}' \subset (K_1 * \mathfrak{X}_1) * \mathfrak{X}', \tag{2.12}$$

which implies that $|K_2| \leq |(K_1 * \mathfrak{X}_1) * \mathfrak{X}'| \leq |(((K_1 * \mathfrak{X}_1) * \mathfrak{X}') * \mathfrak{X}_2) * \mathfrak{X}^{s-1}| = |K_1 * \mathfrak{X}''| = |K_1|$. By exchanging the roles of K_1 and K_2, we get $|K_1| \leq |K_2|$.

Therefore, $|K_2| = |K_1|$ for every $K_1 \in \mathcal{K}_{\mathcal{H}}$ and every $K_2 \in \mathcal{K}_{\mathcal{H}^{l*}}$. We conclude that both $\mathcal{K}_{\mathcal{H}}$ and $\mathcal{K}_{\mathcal{H}^{l*}}$ are balanced partitions and $\|\mathcal{K}_{\mathcal{H}}\| = \|\mathcal{K}_{\mathcal{H}^{l*}}\|$.

Now define $\mathfrak{X}_{a,K_2} = (\mathfrak{X}_1, \mathfrak{X}')$. Since $\mathfrak{X}_{a,K_2}$ is an initial segment of $\mathfrak{X}''$, we have $|K_1 * \mathfrak{X}_{a,K_2}| \leq |K_1 * \mathfrak{X}''|$. But we have shown that $K_1 * \mathfrak{X}'' = K_1$ and $|K_1| = |K_2|$, so we must have $|K_1 * \mathfrak{X}_{a,K_2}| \leq |K_2|$. Moreover, we have $K_2 \subset a * \mathfrak{X}_{a,K_2} \subset K_1 * \mathfrak{X}_{a,K_2}$ from (2.12). We conclude that $K_2 = a * \mathfrak{X}_{a,K_2} = K_1 * \mathfrak{X}_{a,K_2}$. □

Lemma 2.14. *Let $\mathcal{H}$ be a stable partition of $(\mathcal{X}, *)$ where $*$ is ergodic. For every $K \in \mathcal{K}_{\mathcal{H}}$ and every $\mathcal{H}$-sequence $\mathfrak{X}$, we have $|K * \mathfrak{X}| = |K| = \|\mathcal{K}_{\mathcal{H}}\|$.*

Proof. Let $K' = K * \mathfrak{X}$ and $l = |\mathfrak{X}|$, and let $\mathfrak{X}' = (X'_i)_{0 \leq i < (-l \bmod n)}$ be an arbitrary $\mathcal{H}^{l*}$-sequence of length $(-l \bmod n)$. Clearly, $(\mathfrak{X}, \mathfrak{X}')$ is $\mathcal{H}$-repeatable. Lemma 2.7 implies that there exists an integer $s > 0$ such that $(\mathfrak{X}, \mathfrak{X}')^s$ is $\mathcal{H}$-augmenting. We have $K \subset K * (\mathfrak{X}, \mathfrak{X}')^s$. On the other hand, Lemma 2.12 implies that $K * (\mathfrak{X}, \mathfrak{X}')^s \subset K$. Therefore, $K = K * (\mathfrak{X}, \mathfrak{X}')^s = K' * (\mathfrak{X}', (\mathfrak{X}, \mathfrak{X}')^{s-1})$ which implies that $|K'| \leq |K|$. We also have $|K| \leq |K'|$ since $K' = K * \mathfrak{X}$. Thus, $|K'| = |K| = \|\mathcal{K}_{\mathcal{H}}\|$. □

Lemma 2.15. *Let $\mathcal{H}$ be a stable partition of $(\mathcal{X}, *)$ where $*$ is ergodic. Let $K \in \mathcal{K}_{\mathcal{H}}$ and $l > 0$. If $\mathfrak{X} = (X_i)_{0 \leq i < l}$ is an $\mathcal{H}$-sequence, then $K * \mathfrak{X} \in \mathcal{K}_{\mathcal{H}^{l*}}$.*

Proof. Let $K' = K * \mathfrak{X}$. Fix $x \in K'$ and let $K'' \in \mathcal{K}_{\mathcal{H}^{l*}}$ be chosen so that $x \in K''$. Lemma 2.12 implies the existence of an $\mathcal{H}^{l*}$-augmenting sequence $\mathfrak{X}''$ such that $x * \mathfrak{X}'' = K''$. We have $K'' \subset K' * \mathfrak{X}''$ since $x \in K'$, and $K' \subset K' * \mathfrak{X}''$ since $\mathfrak{X}''$ is $\mathcal{H}^{l*}$-augmenting. Therefore, $K' \cup K'' \subset K' * \mathfrak{X}''$. On the other hand, we have the following:

- $|K'| = |K * \mathfrak{X}| = |K| = \|\mathcal{K}_{\mathcal{H}}\|$ from Lemma 2.14.
- $(\mathfrak{X}, \mathfrak{X}'')$ is an $\mathcal{H}$-sequence, so Lemma 2.14 implies that $|K * (\mathfrak{X}, \mathfrak{X}'')| = |K| = \|\mathcal{K}_{\mathcal{H}}\|$. Now since $K' * \mathfrak{X}'' = K * (\mathfrak{X}, \mathfrak{X}'')$, we deduce that $|K' * \mathfrak{X}''| = \|\mathcal{K}_{\mathcal{H}}\|$.
- Lemma 2.13 implies that $\|\mathcal{K}_{\mathcal{H}}\| = \|\mathcal{K}_{\mathcal{H}^{l*}}\|$, so $|K''| = \|\mathcal{K}_{\mathcal{H}^{l*}}\| = \|\mathcal{K}_{\mathcal{H}}\|$.

Therefore, $|K''| = |K'| = |K' * \mathfrak{X}''| = \|\mathcal{K}_{\mathcal{H}}\|$ and $K' \cup K'' \subset K' * \mathfrak{X}''$, hence $K' = K''$ and $K' \in \mathcal{K}_{\mathcal{H}^{l*}}$. □

Lemma 2.16. *Let $\mathcal{H}$ be a stable partition of $(\mathcal{X}, *)$ where $*$ is ergodic. $\mathcal{K}_{\mathcal{H}}$ is a sub-stable partition of $\mathcal{H}$ and $\mathcal{K}_{\mathcal{H}^{l*}} = \mathcal{K}_{\mathcal{H}}{}^{l*}$ for all $l \geq 0$.*

Proof. We will prove that $\mathcal{K}_{\mathcal{H}^{l*}} = \mathcal{K}_{\mathcal{H}}{}^{l*}$ by induction on $l \geq 0$. The statement is trivial for $l = 0$. Now let $l > 0$ and suppose that $\mathcal{K}_{\mathcal{H}^{(l-1)*}} = \mathcal{K}_{\mathcal{H}}{}^{(l-1)*}$. Let $K \in \mathcal{K}_{\mathcal{H}}{}^{l*} = (\mathcal{K}_{\mathcal{H}}{}^{(l-1)*})^* = (\mathcal{K}_{\mathcal{H}^{(l-1)*}})^*$. There exist $K_1, K_2 \in \mathcal{K}_{\mathcal{H}^{(l-1)*}} = \mathcal{K}_{\mathcal{H}}{}^{(l-1)*}$ such that $K = K_1 * K_2$. Let $H_2 \in \mathcal{H}^{(l-1)*}$ be chosen such that $K_2 \subset H_2$ (Lemma 2.8 guarantees the existence of H_2). From Lemma 2.15, we have $K_1 * H_2 \in \mathcal{K}_{\mathcal{H}^{l*}}$ and so $|K_1 * H_2| = \|\mathcal{K}_{\mathcal{H}^{l*}}\| \overset{(a)}{=} \|\mathcal{K}_{\mathcal{H}^{(l-1)*}}\| = |K_1|$, where (a) follows from Lemma 2.13. We have $K_1 * K_2 \subset K_1 * H_2$ and $|K_1| \leq |K_1 * K_2| \leq |K_1 * H_2| = |K_1|$. Therefore, $K = K_1 * K_2 = K_1 * H_2$ which implies that $K \in \mathcal{K}_{\mathcal{H}^{l*}}$. This shows that $\mathcal{K}_{\mathcal{H}}{}^{l*} \subset \mathcal{K}_{\mathcal{H}^{l*}}$, which implies that $\mathcal{K}_{\mathcal{H}}{}^{l*} = \mathcal{K}_{\mathcal{H}^{l*}}$ since $\mathcal{K}_{\mathcal{H}}{}^{l*}$ covers $\mathcal{X}$ and $\mathcal{K}_{\mathcal{H}^{l*}}$ is a partition of $\mathcal{X}$.

We conclude that $\mathcal{K}_{\mathcal{H}}{}^{l*} = \mathcal{K}_{\mathcal{H}^{l*}}$ for all $l \geq 0$. In particular, $\mathcal{K}_{\mathcal{H}}{}^{n*} = \mathcal{K}_{\mathcal{H}^{n*}} = \mathcal{K}_{\mathcal{H}}$, where $n = \mathrm{per}(\mathcal{H})$ so $\mathcal{K}_{\mathcal{H}}$ is periodic. Moreover, Lemma 2.13 shows that $\mathcal{K}_{\mathcal{H}}$ is balanced. Therefore, $\mathcal{K}_{\mathcal{H}}$ is a stable partition. Lemma 2.8 now implies that $\mathcal{K}_{\mathcal{H}}$ is a sub-stable partition of $\mathcal{H}$. □

Proposition 2.13. *Let $\mathcal{H}$ be a stable partition of $(\mathcal{X}, *)$ where $*$ is ergodic, and let $\mathcal{K}$ be a partition of $\mathcal{X}$ which satisfies the following two conditions:*

- *For every $K \in \mathcal{K}$ and every $x \in K$, there exists an $\mathcal{H}$-augmenting sequence $\mathfrak{X}$ such that $x * \mathfrak{X} = K$.*
- *For every $K \in \mathcal{K}$, every $x \in K$, and every $\mathcal{H}$-augmenting sequence $\mathfrak{X}'$, we have $x * \mathfrak{X}' \subset K$.*

Then $\mathcal{K} = \mathcal{K}_{\mathcal{H}}$.

Proof. Fix $x \in \mathcal{X}$ and let $K_{1,x} \in \mathcal{K}_{\mathcal{H}}$ and $K_{2,x} \in \mathcal{K}$ be chosen such that $x \in K_{1,x}$ and $x \in K_{2,x}$. Lemma 2.12 implies the existence of an $\mathcal{H}$-augmenting sequence $\mathfrak{X}_1$ such that $x * \mathfrak{X}_1 = K_{1,x}$, and the first condition of the proposition implies the existence of an $\mathcal{H}$-augmenting sequence $\mathfrak{X}_2$ such that $x * \mathfrak{X}_2 = K_{2,x}$. The second condition of the proposition implies that $x * \mathfrak{X}_1 \subset K_{2,x}$, and Lemma 2.12 implies that $x * \mathfrak{X}_2 \subset K_{1,x}$. Therefore, $K_{1,x} \subset K_{2,x}$ and $K_{2,x} \subset K_{1,x}$ which implies that $K_{1,x} = K_{2,x}$. Since this is true for all $x \in \mathcal{X}$, we conclude that $\mathcal{K} = \mathcal{K}_{\mathcal{H}}$. □

Now we are ready to prove Theorem 2.1:

Proof of Theorem 2.1. Lemma 2.16 shows that $\mathcal{K}_{\mathcal{H}}$ is a sub-stable partition of $\mathcal{H}$ satisfying ${\mathcal{K}_{\mathcal{H}}}^{l*} = \mathcal{K}_{\mathcal{H}^{l*}}$ for all $l \geq 0$. Moreover, we have:

- For every $K \in \mathcal{K}_{\mathcal{H}}$ and every $\mathcal{H}$-sequence $\mathfrak{X}$, we have $K * \mathfrak{X} \in \mathcal{K}_{\mathcal{H}^{|\mathfrak{X}|*}} = {\mathcal{K}_{\mathcal{H}}}^{|\mathfrak{X}|*}$ by Lemma 2.15.
- For every $K \in \mathcal{K}_{\mathcal{H}}$ and every $x \in K$, Lemma 2.12 shows that there exists an $\mathcal{H}$-augmenting sequence $\mathfrak{X}$ such that $x * \mathfrak{X} = K$.
- For every $K \in \mathcal{K}_{\mathcal{H}}$, every $x \in K$, and every $\mathcal{H}$-augmenting sequence $\mathfrak{X}'$, we have $x * \mathfrak{X}' \subset K$ by Lemma 2.12.

This shows the existence part of Theorem 2.1. The uniqueness follows from Proposition 2.13. □

2.8.4 Proof of Proposition 2.7

Definition 2.24. *Let $\mathcal{A}$ be an $\mathcal{X}$-cover. Define the relation $P_{\mathcal{A}}$ on $\mathcal{X}$ as follows: $xP_{\mathcal{A}}y$ if and only if there exists a finite sequence $(A_i)_{1 \leq i \leq n}$ such that $x \in A_1$, $y \in A_n$, $A_i \in \mathcal{A}$ for all $1 \leq i \leq n$, and $A_i \cap A_{i+1} \neq \emptyset$ for all $1 \leq i < n$. Clearly, $P_{\mathcal{A}}$ is an equivalence relation on $\mathcal{X}$. The set of equivalence classes of $P_{\mathcal{A}}$ (denoted by $\mathcal{P}(\mathcal{A})$) is called the partition of $\mathcal{X}$* generated by *$\mathcal{A}$.*

Lemma 2.17. *Let $\mathcal{A}$ be an $\mathcal{X}$-cover. For every $B \in \mathcal{P}(\mathcal{A})$, there exists a finite sequence $(A_i)_{1 \leq i \leq n}$ such that $B = \bigcup_{i=1}^{n} A_i$, $A_i \in \mathcal{A}$ for all $1 \leq i \leq n$, and $A_i \cap A_{i+1} \neq \emptyset$ for all $1 \leq i < n$.*

Proof. Let $B \in \mathcal{P}(\mathcal{A})$ and let $x \in B$. We say that a sequence $(A_i)_{1\leq i\leq n}$ is $(x,\mathcal{A})$-connected if $x \in A_1$, $A_i \in \mathcal{A}$ for all $1 \leq i \leq n$, and $A_i \cap A_{i+1} \neq \emptyset$ for all $1 \leq i < n$. If $(A_i)_{1\leq i\leq n}$ is such a sequence, we clearly have $xP_{\mathcal{A}}y$ for every $y \in \bigcup_{i=1}^{n} A_i$. Therefore, $\bigcup_{i=1}^{n} A_i \subset B$.

Let $A_1 \in \mathcal{A}$ be such that $x \in A_1$. The sequence (A_1) of length 1 is $(x,\mathcal{A})$-connected. Therefore, there exists at least one $(x,\mathcal{A})$-connected sequence. Now consider an $(x,\mathcal{A})$-connected sequence $(A_i)_{1\leq i\leq n}$ such that $\bigcup_{i=1}^{n} A_i$ is maximal. If $\bigcup_{i=1}^{n} A_i \neq B$, there exists $y \in B$ such that $y \notin \bigcup_{i=1}^{n} A_i$. Let $x' \in A_n$. Since $x', y \in B$, $x'P_{\mathcal{A}}y$ and so there exists a sequence $(A'_i)_{1\leq i\leq m}$ such that $x' \in A'_1$, $y \in A'_m$, $A'_i \in \mathcal{A}$ for all $1 \leq i \leq m$, and $A'_i \cap A'_{i+1} \neq \emptyset$ for all $1 \leq i < m$. Consider the sequence $(A''_i)_{1\leq i\leq n+m}$ defined by $A''_i = A_i$ for $1 \leq i \leq n$ and $A''_i = A'_{i-n}$ for $n+1 \leq i \leq n+m$. Since $x' \in A_n \cap A'_1 = A''_n \cap A''_{n+1}$, $(A''_i)_{1\leq i\leq n+m}$ is $(x,\mathcal{A})$-connected. We have $\bigcup_{i=1}^{n} A_i \subsetneq \bigcup_{i=1}^{n+m} A''_i$ since $y \in \bigcup_{i=1}^{n+m} A''_i$ and $y \notin \bigcup_{i=1}^{n} A_i$. This contradicts the maximality of $\bigcup_{i=1}^{n} A_i$. Therefore, we must have $\bigcup_{i=1}^{n} A_i = B$. □

Lemma 2.18. *Let $*$ be a uniformity-preserving operation on a set $\mathcal{X}$, and let $\mathcal{A}$ be an $\mathcal{X}$-cover. For every $n > 0$ and every $A \in \mathcal{A}^{n*}$, there exists $B \in \mathcal{P}(\mathcal{A})^{n*}$ such that $A \subset B$.*

Proof. We will show the lemma by induction on n. The lemma is trivial for $n = 0$.

Now let $n > 0$ and suppose that the lemma is true for $n-1$. Let $A \in \mathcal{A}^{n*}$, there exists $A_1, A_2 \in \mathcal{A}^{(n-1)*}$ such that $A = A_1 * A_2$. The induction hypothesis implies the existence of two sets $B_1, B_2 \in \mathcal{P}(\mathcal{A})^{(n-1)*}$ such that $A_1 \subset B_1$ and $A_2 \subset B_2$. We have $A = A_1 * A_2 \subset B_1 * B_2$ and $B_1 * B_2 \in \mathcal{P}(\mathcal{A})^{n*}$. □

Lemma 2.19. *Let $*$ be a uniformity-preserving operation on a set $\mathcal{X}$, and let $\mathcal{A}$ be an $\mathcal{X}$-cover. For every $n \geq 0$, we have $\mathcal{P}\big(\mathcal{P}(\mathcal{A})^{n*}\big) = \mathcal{P}(\mathcal{A}^{n*})$.*

Proof. We will show the lemma by induction on n. The lemma is trivial for $n = 0$.

Now let $n > 0$ and suppose that $\mathcal{P}\big(\mathcal{P}(\mathcal{A})^{(n-1)*}\big) = \mathcal{P}(\mathcal{A}^{(n-1)*})$, which means that for every $x, y \in \mathcal{X}$, we have $xP_{\mathcal{A}^{(n-1)*}}y$ if and only if $xP_{\mathcal{P}(\mathcal{A})^{(n-1)*}}y$.

Let $x, y \in \mathcal{X}$ be such that $xP_{\mathcal{P}(\mathcal{A})^{n*}}y$. There exists a sequence $(D_j)_{1\leq j\leq m}$ such that: $x \in D_1$, $y \in D_m$, $D_j \in \mathcal{P}(\mathcal{A})^{n*}$ for $1 \leq j \leq m$, and $D_j \cap D_{j+1} \neq \emptyset$ for $1 \leq j < m$. Define $x_1 = x$ and $x_{m+1} = y$, and for each $2 \leq j \leq m$, choose $x_j \in D_{j-1} \cap D_j$. For every $1 \leq j \leq m$, we have $x_j, x_{j+1} \in D_j$ and $D_j \in \mathcal{P}(\mathcal{A})^{n*}$. We are going to show that $x_jP_{\mathcal{A}^{n*}}x_{j+1}$ for every $1 \leq j \leq m$ which will imply that $xP_{\mathcal{A}^{n*}}y$.

Fix $j \in \{1, \ldots, m\}$. Since $D_j \in \mathcal{P}(\mathcal{A})^{n*}$, there exist $D'_j, D''_j \in \mathcal{P}(\mathcal{A})^{(n-1)*}$ such that $D_j = D'_j * D''_j$. Moreover, since $x_j, x_{j+1} \in D_j$ there exist $a'_j, b'_{j+1} \in D'_j$ and $a''_j, b''_{j+1} \in D''_j$ such that $x_j = a'_j * a''_j$ and $x_{j+1} = b'_{j+1} * b''_{j+1}$. We have $a'_jP_{\mathcal{P}(\mathcal{A})^{(n-1)*}}b'_{j+1}$ and $a''_jP_{\mathcal{P}(\mathcal{A})^{(n-1)*}}b''_{j+1}$. Therefore, from the induction hypothesis

we have $a'_j P_{\mathcal{A}^{(n-1)*}} b'_{j+1}$ and $a''_j P_{\mathcal{A}^{(n-1)*}} b''_{j+1}$. There exist two sequences $(A'_i)_{1\leq i\leq m'_j}$ and $(A''_i)_{1\leq i\leq m''_j}$ such that:

- $a'_j \in A'_1$, $b'_{j+1} \in A'_{m'_j}$, $A'_i \in \mathcal{A}^{(n-1)*}$ for $1 \leq i \leq m'_j$, and $A'_i \cap A'_{i+1} \neq \emptyset$ for $1 \leq i < m'_j$.
- $a''_j \in A''_1$, $b''_{j+1} \in A''_{m''_j}$, $A''_i \in \mathcal{A}^{(n-1)*}$ for $1 \leq i \leq m''_j$, and $A''_i \cap A''_{i+1} \neq \emptyset$ for $1 \leq i < m''_j$.

Now consider the sequence $(A_i)_{1\leq i\leq m'_j+m''_j}$ defined as $A_i = A'_i * A''_1$ for $1 \leq i \leq m'_j$, and $A_i = A'_{m'_j} * A''_{i-m'_j}$ for $m'_j + 1 \leq i \leq m'_j + m''_j$. The sequence $(A_i)_{1\leq i\leq m'_j+m''_j}$ satisfies the following: $x_j = a'_j * a''_j \in A_1$, $x_{j+1} = b'_{j+1} * b''_{j+1} \in A_{m'_j+m''_j}$ and $A_i \in \mathcal{A}^{n*}$ for $1 \leq i \leq m'_j + m''_j$. Moreover, it is easy to see that $A_i \cap A_{i+1} \neq \emptyset$ for $1 \leq i < m'_j + m''_j$. Therefore, $x_j P_{\mathcal{A}^{n*}} x_{j+1}$. Now since this is true for all $1 \leq j \leq m$, we have $x_1 P_{\mathcal{A}^{n*}} x_{m+1}$ and so $x P_{\mathcal{A}^{n*}} y$. We conclude that for every $x, y \in \mathcal{X}$, $x P_{\mathcal{P}(\mathcal{A})^{n*}} y$ implies $x P_{\mathcal{A}^{n*}} y$.

Now let $x, y \in \mathcal{X}$ be such that $x P_{\mathcal{A}^{n*}} y$. There exists a sequence $(E_i)_{1\leq i\leq k}$ such that: $x \in E_1$, $y \in E_k$, $E_i \in \mathcal{A}^{n*}$ for $1 \leq i \leq k$, and $E_i \cap E_{i+1} \neq \emptyset$ for $1 \leq i < k$. Now for every $1 \leq i \leq k$, we can apply Lemma 2.18 to get a set $F_i \in \mathcal{P}(\mathcal{A})^{n*}$ such that $E_i \subset F_i$. Clearly, we have $x \in F_1$, $y \in F_k$, $F_i \in \mathcal{P}(\mathcal{A})^{n*}$ for $1 \leq i \leq k$, and $F_i \cap F_{i+1} \neq \emptyset$ for $1 \leq i < k$. Thus, $x P_{\mathcal{P}(\mathcal{A})^{n*}} y$.

We conclude that for every $x, y \in \mathcal{X}$, $x P_{\mathcal{P}(\mathcal{A})^{n*}} y$ if and only if $x P_{\mathcal{A}^{n*}} y$. Therefore, $\mathcal{P}\big(\mathcal{P}(\mathcal{A})^{n*}\big) = \mathcal{P}(\mathcal{A}^{n*})$. □

Lemma 2.20. *Let $*$ be an ergodic operation on a set $\mathcal{X}$. If $\mathcal{A}$ is a periodic $\mathcal{X}$-cover, then $\mathcal{P}(\mathcal{A})$ is a stable partition.*

Proof. Let $n = \text{per}(\mathcal{A}) \cdot \text{con}(*)$. Since $\text{per}(\mathcal{A})$ divides n, we have $\mathcal{A}^{n*} = \mathcal{A}$. Let $A \in \mathcal{P}(\mathcal{A})$ be chosen so that $|A|$ is maximal, and let $B \in \mathcal{P}(\mathcal{A})$. We clearly have $|B| \leq |A|$. We also have $B \in \mathcal{P}(\mathcal{A}^{n*})$ since $\mathcal{A}^{n*} = \mathcal{A}$. From Lemma 2.19 we have $\mathcal{P}\big(\mathcal{P}(\mathcal{A})^{n*}\big) = \mathcal{P}(\mathcal{A}^{n*})$, and so $B \in \mathcal{P}\big(\mathcal{P}(\mathcal{A})^{n*}\big) = \mathcal{P}(\mathcal{A}^{n*}) = \mathcal{P}(\mathcal{A})$.

Fix $x \in A$ and $y \in B$. Since $n \geq \text{con}(*)$, there exists a sequence $x_0, \ldots, x_{n-1} \in \mathcal{X}$ such that $y = (\ldots((x * x_0) * x_1) \ldots * x_{n-1})$. Now choose $X_0, \ldots, X_{n-1}$ such that $x_i \in X_i \in \mathcal{P}(\mathcal{A})^{i*}$ for $0 \leq i < n$. Define $C := (\ldots((A * X_0) * X_1) \ldots * X_{n-1})$. Clearly, $y \in C \in \mathcal{P}(\mathcal{A})^{n*}$. Now since $y \in B \in \mathcal{P}\big(\mathcal{P}(\mathcal{A})^{n*}\big)$ and $y \in C \in \mathcal{P}(\mathcal{A})^{n*}$, we must have $C = (\ldots((A * X_0) * X_1) \ldots * X_{n-1}) \subset B$ and so $|A| \leq |C| \leq |B|$, which implies that $|A| = |B| = |C|$ since we already have $|B| \leq |A|$. Therefore, $C = B$ and so $B \in \mathcal{P}(\mathcal{A})^{n*}$ for every $B \in \mathcal{P}(\mathcal{A})$, from which we conclude that $\mathcal{P}(\mathcal{A}) \subset \mathcal{P}(\mathcal{A})^{n*}$. On the other hand, since $|A| = |B|$ for every $B \in \mathcal{P}(\mathcal{A})$, $\mathcal{P}(\mathcal{A})$ is a balanced partition.

Now for every $C \in \mathcal{P}(\mathcal{A})^{n*}$, there exists a set $D \in \mathcal{P}(\mathcal{A})$ and a sequence $X_0, \ldots, X_{n-1}$ such that $X_i \in \mathcal{P}(\mathcal{A})^{i*}$ and $C = (\ldots((D * X_0) * X_1) \ldots * X_{n-1})$. We have $|D| \leq |C|$. On the other hand, Lemma 2.18 (applied to the $\mathcal{X}$-cover $\mathcal{P}(\mathcal{A})^{n*}$) implies the existence of a set $B \in \mathcal{P}\big(\mathcal{P}(\mathcal{A})^{n*}\big)$ such that $C \subset B$. Therefore, $|D| \leq |C| \leq |B|$. Now since $\mathcal{P}\big(\mathcal{P}(\mathcal{A})^{n*}\big) = \mathcal{P}(\mathcal{A}^{n*})$ (by Lemma 2.19) and $\mathcal{A}^{n*} = \mathcal{A}$, we have $B \in \mathcal{P}\big(\mathcal{P}(\mathcal{A})^{n*}\big) = \mathcal{P}(\mathcal{A}^{n*}) = \mathcal{P}(\mathcal{A})$. Therefore, $|D| = |B|$ since $D, B \in \mathcal{P}(\mathcal{A})$ and since $\mathcal{P}(\mathcal{A})$ was shown to be a balanced partition. Thus, $|B| = |C| = |D|$ which implies that $C = B \in \mathcal{P}(A)$ since $C \subset B$. We conclude that $C \in \mathcal{P}(A)$ for every $C \in \mathcal{P}(\mathcal{A})^{n*}$. Therefore, $\mathcal{P}(\mathcal{A})^{n*} \subset \mathcal{P}(\mathcal{A})$. This means that

$\mathcal{P}(\mathcal{A})^{n*} = \mathcal{P}(\mathcal{A})$ since we already have $\mathcal{P}(\mathcal{A}) \subset \mathcal{P}(\mathcal{A})^{n*}$. We conclude that $\mathcal{P}(\mathcal{A})$ is a stable partition. □

Lemma 2.21. *Let $*$ be a uniformity-preserving operation on a set $\mathcal{X}$. If $\mathcal{A}$ is a stable $\mathcal{X}$-cover, then for every $i \geq 0$, every $A \in \mathcal{A}$ and every $B \in \mathcal{A}^{i*}$, we have $|A| = |B|$.*

Proof. Fix $i \geq 0$, and let $p = \min\{k \cdot \mathrm{per}(\mathcal{A}) : \ k \cdot \mathrm{per}(\mathcal{A}) > i\}$. Clearly, $\mathcal{A}^{p*} = \mathcal{A}$. Let $A \in \mathcal{A}$ and $B \in \mathcal{A}^{i*}$. We have

$$|A| \overset{(a)}{=} \|\mathcal{A}\|_\wedge \overset{(b)}{\leq} \|\mathcal{A}^{i*}\|_\wedge \leq |B| \leq \|\mathcal{A}^{i*}\|_\vee \overset{(c)}{\leq} \|\mathcal{A}^{p*}\|_\vee = \|\mathcal{A}\|_\vee \overset{(d)}{=} |A|,$$

where (a) and (d) follow from the fact that $\mathcal{A}$ is a balanced $\mathcal{X}$-cover. (b) and (c) follow from Lemma 2.1. This shows that $|A| = |B|$. □

Lemma 2.22. *Let $*$ be a uniformity-preserving operation on a set $\mathcal{X}$, and let $\mathcal{A}$ be a stable $\mathcal{X}$-cover. For every $A, B, C \in \mathcal{A}$, if $B \cap C \neq \emptyset$ then $A * B = A * C$.*

Proof. We have $A * B \in \mathcal{A}^*$, and from Lemma 2.21 we get $|A * B| = |A|$. On the other hand, since $*$ is uniformity-preserving, we have $|A * x| = |A|$ for every $x \in \mathcal{X}$. Now since $A * B = \bigcup_{b \in B} A * b$, and since $|A * b| = |A| = |A * B|$ for every $b \in B$, we must have $A * B = A * b$ for every $b \in B$. Similarly, $A * C = A * c$ for every $c \in C$. We conclude that $A * B = A * C$ since $B \cap C \neq \emptyset$ (for any $x \in B \cap C$, we have $A * B = A * x = A * C$). □

Lemma 2.23. *Let $*$ be a uniformity-preserving operation on a set $\mathcal{X}$, and let $\mathcal{A}$ be a stable $\mathcal{X}$-cover. For every $A \in \mathcal{A}$ and every $B \in \mathcal{P}(\mathcal{A})$, we have $A * B \in \mathcal{A}^*$.*

Proof. According to Lemma 2.17 there exists a finite sequence $(A_i)_{1 \leq i \leq l}$ such that $B = \bigcup_{i=1}^{l} A_i$, $A_i \in \mathcal{A}$ for all $1 \leq i \leq l$, and $A_i \cap A_{i+1} \neq \emptyset$ for all $1 \leq i < l$. Lemma 2.22 shows that $A * A_1 = A * A_2 = \ldots = A * A_l$. Therefore, $A * B = A * A_1 \in \mathcal{A}^*$. □

Lemma 2.24. *Let $*$ be an ergodic operation on a set $\mathcal{X}$, and let $\mathcal{A}$ be a stable $\mathcal{X}$-cover. For every $A \in \mathcal{A}$ and every $\mathcal{P}(\mathcal{A})$-sequence $\mathfrak{X}$, we have $A * \mathfrak{X} \in \mathcal{A}^{|\mathfrak{X}|*}$.*

Proof. We will prove the lemma by induction on $k = |\mathfrak{X}| > 0$. Lemma 2.23 implies that the statement is true for $k = 1$. Now let $k > 1$ and suppose that the lemma is true for $|\mathfrak{X}| = k - 1$. Now let $\mathfrak{X} = (X_i)_{0 \leq i < k}$ be a $\mathcal{P}(\mathcal{A})$-sequence of length k. Define $\mathfrak{X}' = (X_i)_{0 \leq i < k-1}$. We have:

- $A' = A * \mathfrak{X}' \in \mathcal{A}^{(k-1)*}$ from the induction hypothesis.
- Lemma 2.20 shows that $\mathcal{P}(\mathcal{A})$ is a stable partition, and so $\mathcal{P}(\mathcal{A})^{(k-1)*}$ is also a stable partition. In particular, $\mathcal{P}(\mathcal{A})^{(k-1)*}$ is a partition and so $\mathcal{P}(\mathcal{A})^{(k-1)*} = \mathcal{P}\big(\mathcal{P}(\mathcal{A})^{(k-1)*}\big)$. On the other hand, Lemma 2.19 shows that $\mathcal{P}\big(\mathcal{P}(\mathcal{A})^{(k-1)*}\big) = \mathcal{P}(\mathcal{A}^{(k-1)*})$. Therefore, $\mathcal{P}(\mathcal{A})^{(k-1)*} = \mathcal{P}(\mathcal{A}^{(k-1)*})$. We conclude that $X_{k-1} \in \mathcal{P}(\mathcal{A}^{(k-1)*})$ since we have $X_{k-1} \in \mathcal{P}(\mathcal{A})^{(k-1)*}$.

- Since $(\mathcal{A}^{(k-1)*})^{n*} = (\mathcal{A}^{n*})^{(k-1)*} = \mathcal{A}^{(k-1)*}$ (where $n = \mathrm{per}(\mathcal{A})$), $\mathcal{A}^{(k-1)*}$ is a periodic $\mathcal{X}$-cover. On the other hand, Lemma 2.21 implies that $\mathcal{A}^{(k-1)*}$ is balanced. Therefore, $\mathcal{A}^{(k-1)*}$ is a stable $\mathcal{X}$-cover.

Now since $A' \in \mathcal{A}^{(k-1)*}$ and $X_{k-1} \in \mathcal{P}(\mathcal{A}^{(k-1)*})$, and since $\mathcal{A}^{(k-1)*}$ is a stable $\mathcal{X}$-cover, we can apply Lemma 2.23 to obtain $A' * X_{k-1} \in (\mathcal{A}^{(k-1)*})^* = \mathcal{A}^{k*}$. We conclude that $A * \mathfrak{X} = A' * X_{k-1} \in \mathcal{A}^{k*}$ which completes the induction argument. □

Now we are ready to prove Proposition 2.7:

Proof of Proposition 2.7. Let $*$ be a strongly ergodic operation on $\mathcal{X}$ and let $\mathcal{A}$ be a stable $\mathcal{X}$-cover. Lemma 2.20 shows that $\mathcal{P}(\mathcal{A})$ is a stable partition. Let $n = \mathrm{per}(\mathcal{A}).\,\mathrm{scon}(*)$. We have the following:

- $\mathcal{P}(\mathcal{A})^{n*} = \mathcal{P}\big(\mathcal{P}(\mathcal{A})^{n*}\big)$ since $\mathcal{P}(\mathcal{A})$ is a stable partition.
- $\mathcal{P}\big(\mathcal{P}(\mathcal{A})^{n*}\big) = \mathcal{P}(\mathcal{A}^{n*})$ by Lemma 2.19.
- $\mathcal{A}^{n*} = \mathcal{A}$ since $\mathrm{per}(\mathcal{A})$ divides n.

Therefore, $\mathcal{P}(\mathcal{A})^{n*} = \mathcal{P}(\mathcal{A}^{n*}) = \mathcal{P}(\mathcal{A})$.

Fix $A \in \mathcal{A}$. From Lemma 2.18 there exists $B \in \mathcal{P}(\mathcal{A})$ such that $A \subset B$. Fix $a \in A$. Since $a \in B \in \mathcal{P}(\mathcal{A}) = \mathcal{P}(\mathcal{A})^{n*}$ and since $n \geq \mathrm{scon}(*)$, we can apply Theorem 2.2 to get a $\mathcal{P}(\mathcal{A})$-sequence of length n such that $a * \mathfrak{X} = B * \mathfrak{X} = B$. Since $B = a * \mathfrak{X} \subset A * \mathfrak{X} \subset B * \mathfrak{X} = B$, we have $A * \mathfrak{X} = B$. Now from Lemma 2.24, we have $B = A * \mathfrak{X} \in \mathcal{A}^{n*} = \mathcal{A}$. This means that $|A| = |B|$ because $A, B \in \mathcal{A}$ and $\mathcal{A}$ is stable. Therefore, $A = B$ since we have $A \subset B$ and $|A| = |B|$.

We conclude that $A \in \mathcal{P}(\mathcal{A})$ for every $A \in \mathcal{A}$. Now since $\mathcal{P}(\mathcal{A})$ is a partition, we have $A \cap A' = \emptyset$ for every $A, A' \in \mathcal{A}$ satisfying $A \neq A'$. On the other hand, $\mathcal{A}$ is an $\mathcal{X}$-cover. This shows that $\mathcal{A}$ itself is a partition, hence $\mathcal{A} = \mathcal{P}(\mathcal{A})$. Therefore, $\mathcal{A}$ is a stable partition. □

2.8.5 Proof of Proposition 2.8

Lemma 2.25. *Let $*$ be a uniformity-preserving operation on $\mathcal{X}$ and let $\mathcal{A}$ be a periodic $\mathcal{X}$-cover. We have* $\mathrm{core}(\mathcal{A})^{n*} \subset \mathrm{core}(\mathcal{A}^{n*})$ *for every $n \geq 1$.*

Proof. Let $A \in \mathrm{core}(\mathcal{A})^{n*}$. There exist $A_0, A_0' \in \mathrm{core}(\mathcal{A}), A_1' \in \mathrm{core}(\mathcal{A})^*, \ldots, A_{n-1}' \in \mathrm{core}(\mathcal{A})^{(n-1)*}$ such that $A = (\ldots((A_0 * A_0') * A_1') \ldots * A_{n-1}')$. We have

$$\|\mathcal{A}^{n*}\|_\vee \geq |A| = |(\ldots((A_0 * A_0') * A_1') \ldots * A_{n-1}')| \geq |A_0| = \|\mathcal{A}\|_\vee \overset{(a)}{=} \|\mathcal{A}^{n*}\|_\vee,$$

where (a) follows from Lemma 2.3. Therefore, $|A| = \|\mathcal{A}^{n*}\|_\vee$ and so $A \in \mathrm{core}(\mathcal{A}^{n*})$. We conclude that $\mathrm{core}(\mathcal{A})^{n*} \subset \mathrm{core}(\mathcal{A}^{n*})$. □

Lemma 2.26. *Let $*$ be a uniformity-preserving operation on $\mathcal{X}$ and let $\mathcal{A}$ be a periodic $\mathcal{X}$-cover. We have* $|\mathrm{core}(\mathcal{A})^{n*}| \geq |\mathrm{core}(\mathcal{A})|$ *for every $n \geq 1$.*

Proof. Fix $b_0 \in B_0 \in \operatorname{core}(\mathcal{A}), b_1 \in B_1 \in \operatorname{core}(\mathcal{A})^*, \dots, b_{n-1} \in B_{n-1} \in \operatorname{core}(\mathcal{A})^{(n-1)*}$. Let $\pi : \mathcal{X} \to \mathcal{X}$ be defined as $\pi(x) = (\dots((x * b_0) * b_1) \dots * b_{n-1})$. Clearly, π is a bijection because $*$ is uniformity-preserving.

For every $A \in \operatorname{core}(\mathcal{A})$, we have $(\dots((A * B_0) * B_1) \dots * B_{n-1}) \in \operatorname{core}(\mathcal{A})^{n*}$. Lemma 2.25 now implies that $(\dots((A * B_0) * B_1) \dots * B_{n-1}) \in \operatorname{core}(\mathcal{A}^{n*})$ and so

$$|(\dots((A * B_0) * B_1) \dots * B_{n-1})| = \|\mathcal{A}^{n*}\|_\vee \overset{(a)}{=} \|\mathcal{A}\| = |A|,$$

where (a) follows from Lemma 2.3. Now since $\pi(A) = (\dots((A*b_0)*b_1)\dots*b_{n-1}) \subset (\dots((A * B_0) * B_1) \dots * B_{n-1})$ and $|(\dots((A * B_0) * B_1) \dots * B_{n-1})| = |A| = |\pi(A)|$, we have $(\dots((A * B_0) * B_1) \dots * B_{n-1}) = \pi(A)$. Therefore, $\pi(A) \in \operatorname{core}(\mathcal{A})^{n*}$ for every $A \in \operatorname{core}(\mathcal{A})$. We conclude that

$$|\operatorname{core}(\mathcal{A})^{n*}| \geq |\{\pi(A) : \ A \in \operatorname{core}(\mathcal{A})\}| \overset{(a)}{=} |\{A : \ A \in \operatorname{core}(\mathcal{A})\}| = |\operatorname{core}(\mathcal{A})|,$$

where (a) follows from the fact that π is a bijection. □

Lemma 2.27. *Let $*$ be a uniformity-preserving operation on $\mathcal{X}$ and let $\mathcal{A}$ be a periodic $\mathcal{X}$-cover. We have* $\operatorname{core}(\mathcal{A})^{n*} = \operatorname{core}(\mathcal{A}^{n*})$ *for every $n \geq 1$.*

Proof. Let $p = \min\{k \cdot \operatorname{per}(\mathcal{A}) : \ k \cdot \operatorname{per}(\mathcal{A}) > n\}$. Lemmas 2.25 and 2.26 imply that $|\operatorname{core}(\mathcal{A}^*)| \geq |\operatorname{core}(\mathcal{A})^*| \geq |\operatorname{core}(\mathcal{A})|$. Therefore, we have

$$|\operatorname{core}(\mathcal{A})| = |\operatorname{core}(\mathcal{A}^{p*})| \geq |\operatorname{core}(\mathcal{A}^{(p-1)*})| \geq \dots \geq |\operatorname{core}(\mathcal{A}^{n*})| \geq \dots \geq |\operatorname{core}(\mathcal{A})|,$$

hence $|\operatorname{core}(\mathcal{A}^{n*})| = |\operatorname{core}(\mathcal{A})|$. Lemma 2.26 now implies that

$$|\operatorname{core}(\mathcal{A})^{n*}| \geq |\operatorname{core}(\mathcal{A})| = |\operatorname{core}(\mathcal{A}^{n*})|,$$

and from Lemma 2.25 we have $\operatorname{core}(\mathcal{A})^{n*} \subset \operatorname{core}(\mathcal{A}^{n*})$. We conclude that we have $\operatorname{core}(\mathcal{A})^{n*} = \operatorname{core}(\mathcal{A}^{n*})$. □

Lemma 2.28. *Let $*$ be an ergodic operation on $\mathcal{X}$. If $\mathcal{A}$ is a periodic $\mathcal{X}$-cover, then* $\operatorname{core}(\mathcal{A})$ *is an $\mathcal{X}$-cover.*

Proof. Let $n = \operatorname{per}(\mathcal{A}) \cdot \operatorname{con}(*)$. Fix $A \in \operatorname{core}(\mathcal{A})$ and $a \in A$. Now let $x \in \mathcal{X}$. Since $n \geq \operatorname{con}(*)$, the eighth point of Proposition 2.1 implies that $a \xrightarrow{*,n} x$. Therefore, there exist $x_0, \dots, x_{n-1}$ such that $(\dots((a * x_0) * x_1) \dots * x_{n-1}) = x$.

Now since $\mathcal{A}$ is an $\mathcal{X}$-cover, $\mathcal{A}^{i*}$ is an $\mathcal{X}$-cover for every $i \geq 0$. Therefore, for every $0 \leq i < n$, there exists $A_i \in \mathcal{A}^{i*}$ such that $x_i \in A_i$. Let

$$B := (\dots((A * A_0) * A_1) \dots * A_{n-1}) \in \mathcal{A}^{n*}.$$

We have $\mathcal{A}^{n*} = \mathcal{A}$ since $\operatorname{per}(\mathcal{A})$ divides n, hence $B \in \mathcal{A}$. We also have

$$\|\mathcal{A}\|_\vee \overset{(a)}{\geq} |B| = |(\dots((A * A_0) * A_1) \dots * A_{n-1})| \overset{(b)}{\geq} |A| = \|\mathcal{A}\|_\vee,$$

where (a) follows from the fact that $B \in \mathcal{A}$, and (b) follows from the fact that $*$ is uniformity-preserving. Therefore, $|B| = \|\mathcal{A}\|_\vee$, which implies that $B \in \operatorname{core}(\mathcal{A})$. Now since

$$x = (\dots((a * x_0) * x_1) \dots * x_{n-1}) \in (\dots((A * A_0) * A_1) \dots * A_{n-1}) = B \in \operatorname{core}(\mathcal{A}),$$

we have

$$x \in \bigcup_{C \in \mathrm{core}(\mathcal{A})} C.$$

But this is true for every $x \in \mathcal{X}$. We conclude that $\mathrm{core}(\mathcal{A})$ is an $\mathcal{X}$-cover. □

Now we are ready to prove Proposition 2.8:

Proof of Proposition 2.8. Let $*$ be an ergodic operation on $\mathcal{X}$ and let $\mathcal{A}$ be a periodic $\mathcal{X}$-cover. Lemma 2.28 implies that $\mathrm{core}(\mathcal{A})$ is an $\mathcal{X}$-cover.

Let $p = \mathrm{per}(\mathcal{A})$. Lemma 2.27 implies that $\mathrm{core}(\mathcal{A})^{n*} = \mathrm{core}(\mathcal{A}^{n*})$ for every $n \geq 1$. In particular, we have $\mathrm{core}(\mathcal{A})^{p*} = \mathrm{core}(\mathcal{A}^{p*}) = \mathrm{core}(\mathcal{A})$, which implies that $\mathrm{core}(\mathcal{A})$ is periodic and $\mathrm{per}(\mathrm{core}(\mathcal{A}))$ divides p. Now since $\mathrm{core}(\mathcal{A})$ is clearly balanced, we conclude that $\mathrm{core}(\mathcal{A})$ is a stable $\mathcal{X}$-cover. □

2.8.6 Proofs for Section 2.7

Proof of Theorem 2.4

In order to prove Theorem 2.4, we need a few definitions and lemmas:

Definition 2.25. *Define the two projection mappings $P_1 : \mathcal{X} \to \mathcal{X}_1$ and $P_2 : \mathcal{X} \to \mathcal{X}_2$ as $P_1(x_1, x_2) = x_1$ and $P_2(x_1, x_2) = x_2$ for all $(x_1, x_2) \in \mathcal{X}$. Define the following:*

- $\mathcal{U}_1(\mathcal{H}) = \{P_1(H) : H \in \mathcal{H}\}$.
- $\mathcal{U}_2(\mathcal{H}) = \{P_2(H) : H \in \mathcal{H}\}$.

Lemma 2.29. *For every $x_2, x_2' \in \mathcal{X}_2$, there exists an $\mathcal{H}$-repeatable sequence $\mathfrak{X}$ such that:*

- *For every $x_1 \in \mathcal{X}_1$, we have $(x_1, x_2') \in (x_1, x_2) * \mathfrak{X}$.*
- *For every $X \subset \mathcal{X}$, we have $P_1(X) \subset P_1(X * \mathfrak{X})$.*

We say that the sequence $\mathfrak{X}$ can take the second coordinate from x_2 to x_2' while keeping the first coordinate unchanged.

Proof. Let $k = \mathrm{per}(\mathcal{H})\,\mathrm{con}(*_2) \geq \mathrm{con}(*_2)$. Choose arbitrarily a sequence of k elements $x_{1,0}, \ldots, x_{1,k-1}$ in $\mathcal{X}_1$ and define the mapping $\pi : \mathcal{X}_1 \to \mathcal{X}_1$ as $\pi(x_1) = (\ldots((x_1 *_1 x_{1,0}) *_1 x_{1,1}) \ldots *_1 x_{1,k-1})$. Since π is a permutation of $\mathcal{X}_1$, there exists an integer $s > 0$ such that $\pi^s(x_1) = x_1$ for all $x_1 \in \mathcal{X}_1$. Let $l = ks$ and define the sequence $x_{1,i}$ for $k \leq i < l$ as $x_{1,i} = x_{1,i \bmod k}$. Clearly,

$$(\ldots((x_1 *_1 x_{1,0}) *_1 x_{1,1}) \ldots *_1 x_{1,l-1}) = \pi^s(x_1) = x_1 \text{ for all } x_1 \in \mathcal{X}_1. \tag{2.13}$$

Now since $l \geq k \geq \mathrm{con}(*_2)$ and since $*_2$ is ergodic, there exists a sequence $(x_{2,i})_{0 \leq i < l}$ in $\mathcal{X}_2$ such that

$$x_2' = (\ldots((x_2 *_2 x_{2,0}) *_2 x_{2,1}) \ldots *_2 x_{2,l-1}). \tag{2.14}$$

Define the $\mathcal{H}$-repeatable sequence $\mathfrak{X} = (X_i)_{0 \leq i < l}$ such that $(x_{1,i}, x_{2,i}) \in X_i \in \mathcal{H}^{i*}$ for all $0 \leq i < l$. For every $x_1 \in \mathcal{X}_1$, we have:

$$(x_1, x_2') \overset{(a)}{=} (x_1, x_2) * \Big((x_{1,i}, x_{2,i})_{0 \leq i < l}\Big) \overset{(b)}{\in} (x_1, x_2) * \mathfrak{X},$$

where (a) follows from (2.13) and (2.14), and (b) follows from the fact that $(x_{1,i}, x_{2,i}) \in X_i$ for all $0 \leq i < l$.

Now let $X \subset \mathcal{X}$. We have:

$$P_1(X) \overset{(a)}{=} (\dots((P_1(X) *_1 x_{1,0}) *_1 x_{1,1}) \dots *_1 x_{1,l-1}) \\ \overset{(b)}{=} P_1(X * (x_{1,i}, x_{2,i})_{0 \leq i < l}) \overset{(c)}{\subset} P_1(X * \mathfrak{X}),$$

where (a) follows from (2.13), (b) follows from the definition of $*$ and P_1, and (c) follows from the fact that $(x_{1,i}, x_{2,i}) \in X_i$ for all $0 \leq i < l$. □

Lemma 2.30. *Let $\mathfrak{X}$ be an $\mathcal{H}$-repeatable sequence which takes the second coordinate from x_2 to x_2' while keeping the first coordinate unchanged as in Lemma 2.29. If there exist $H, H' \in \mathcal{H}$ and $x_1 \in \mathcal{X}_1$ such that $(x_1, x_2) \in H$ and $(x_1, x_2') \in H'$, then $H' = H * \mathfrak{X}$.*

Proof. From Lemma 2.29 we have $(x_1, x_2') \in (x_1, x_2) * \mathfrak{X} \subset H * \mathfrak{X}$. Therefore, $H' \cap (H * \mathfrak{X}) \neq \emptyset$. On the other hand, we have $H' \in \mathcal{H}$ and $H * \mathfrak{X} \in \mathcal{H}^{|\mathfrak{X}|*} = \mathcal{H}$. Therefore, $H' = H * \mathfrak{X}$ since $\mathcal{H}$ is a partition. □

Lemma 2.31. *$\mathcal{U}_1(\mathcal{H})$ (resp. $\mathcal{U}_2(\mathcal{H})$) is a partition of $\mathcal{X}_1$ (resp. $\mathcal{X}_2$).*

Proof. Clearly, $\mathcal{U}_1(\mathcal{H})$ covers $\mathcal{X}_1$. Now suppose that there exist $A, B \in \mathcal{U}_1(\mathcal{H})$ such that $A \cap B \neq \emptyset$ and let $x_1 \in A \cap B$. Let $H_A, H_B \in \mathcal{H}$ be such that $P_1(H_A) = A$ and $P_1(H_B) = B$. There exist $x_{2,A} \in \mathcal{X}_2$ and $x_{2,B} \in \mathcal{X}_2$ such that $(x_1, x_{2,A}) \in H_A$ and $(x_1, x_{2,B}) \in H_B$. Using Lemma 2.29, choose an $\mathcal{H}$-repeatable sequence $\mathfrak{X}$ which can take the second coordinate from $x_{2,A}$ to $x_{2,B}$ while keeping the first coordinate unchanged.

Lemma 2.30 shows that $H_B = H_A * \mathfrak{X}$ and Lemma 2.29 implies that $P_1(H_A) \subset P_1(H_A * \mathfrak{X})$. We conclude that $A = P_1(H_A) \subset P_1(H_A * \mathfrak{X}) = P_1(H_B) = B$. By exchanging the roles of A and B, we can also get $B \subset A$. Therefore, $A = B$. We conclude that $\mathcal{U}_1(\mathcal{H})$ is a partition of $\mathcal{X}_1$. A similar argument shows that $\mathcal{U}_2(\mathcal{H})$ is a partition of $\mathcal{X}_2$. □

Lemma 2.32. *$\mathcal{U}_1(\mathcal{H})$ (resp. $\mathcal{U}_2(\mathcal{H})$) is a stable partition of $\mathcal{X}_1$ (resp. $\mathcal{X}_2$) of period of at most $\mathrm{per}(\mathcal{H})$. Moreover, for every $i \geq 0$, we have $\mathcal{U}_1(\mathcal{H})^{i*_1} = \mathcal{U}_1(\mathcal{H}^{i*})$ and $\mathcal{U}_2(\mathcal{H})^{i*_2} = \mathcal{U}_2(\mathcal{H}^{i*})$.*

Proof. We will only prove the lemma for $\mathcal{U}_1(\mathcal{H})$ since the proof for $\mathcal{U}_2(\mathcal{H})$ is similar. We will start by showing by induction on $i \geq 0$ that $\mathcal{U}_1(\mathcal{H})^{i*_1} = \mathcal{U}_1(\mathcal{H}^{i*})$. The claim is trivial for $i = 0$. Now let $i > 0$ and suppose that the claim is true for $i - 1$. We have:

$$\begin{aligned} \mathcal{U}_1(\mathcal{H})^{i*_1} &= \left(\mathcal{U}_1(\mathcal{H})^{(i-1)*_1}\right)^{*_1} \overset{(a)}{=} \left(\mathcal{U}_1(\mathcal{H}^{(i-1)*})\right)^{*_1} \\ &= \{H_1' *_1 H_1'' : \ H_1', H_1'' \in \mathcal{U}_1(\mathcal{H}^{(i-1)*})\} \\ &= \{P_1(H') *_1 P_1(H'') : \ H', H'' \in \mathcal{H}^{(i-1)*}\} \\ &\overset{(b)}{=} \{P_1(H' * H'') : \ H', H'' \in \mathcal{H}^{(i-1)*}\} \\ &= \{P_1(H) : \ H \in \mathcal{H}^{i*}\} = \mathcal{U}_1(\mathcal{H}^{i*}), \end{aligned}$$

where (a) follows from the induction hypothesis and (b) follows from the identity $P_1(H') *_1 P_1(H'') = P_1(H' * H'')$ which is very easy to check. We conclude that we have $\mathcal{U}_1(\mathcal{H})^{i*_1} = \mathcal{U}_1(\mathcal{H}^{i*})$ for all $i \geq 0$. In particular, for $p = \text{per}(\mathcal{H})$, we have $\mathcal{U}_1(\mathcal{H})^{p*_1} = \mathcal{U}_1(\mathcal{H}^{p*}) = \mathcal{U}_1(\mathcal{H})$.

Lemma 2.31 shows that $\mathcal{U}_1(\mathcal{H})$ is a partition, and we have just shown that $\mathcal{U}_1(\mathcal{H})^{p*_1} = \mathcal{U}_1(\mathcal{H})$. Therefore, $\mathcal{U}_1(\mathcal{H})$ is periodic of period of at most p. Lemma 2.2 now implies that $\mathcal{U}_1(\mathcal{H})$ is a stable partition of $\mathcal{X}_1$. □

Definition 2.26. *Let $X \subset \mathcal{X}$, $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$. Define the sets $P_{1|x_2}(X) \subset \mathcal{X}_1$ and $P_{2|x_1}(X) \subset \mathcal{X}_2$ as:*

- $P_{1|x_2}(X) = \{x_1 \in \mathcal{X}_1 : \ (x_1, x_2) \in X\} = P_1\big(X \cap (\mathcal{X}_1 \times \{x_2\})\big)$.
- $P_{2|x_1}(X) = \{x_2 \in \mathcal{X}_2 : \ (x_1, x_2) \in X\} = P_2\big(X \cap (\{x_1\} \times \mathcal{X}_2)\big)$.

Define the following:

- $\mathcal{L}_1(\mathcal{H}) = \{P_{1|x_2}(H) : \ H \in \mathcal{H}, \ x_2 \in \mathcal{X}_2, \ P_{1|x_2}(H) \neq \emptyset\}$.
- $\mathcal{L}_2(\mathcal{H}) = \{P_{2|x_1}(H) : \ H \in \mathcal{H}, \ x_1 \in \mathcal{X}_1, \ P_{2|x_1}(H) \neq \emptyset\}$.

Lemma 2.33. *$\mathcal{L}_1(\mathcal{H})$ (resp. $\mathcal{L}_2(\mathcal{H})$) is a partition of $\mathcal{X}_1$ (resp. $\mathcal{X}_2$).*

Proof. Clearly, $\mathcal{L}_1(\mathcal{H})$ covers $\mathcal{X}_1$. Suppose that there exist $A, B \in \mathcal{L}_1(\mathcal{H})$ such that $A \cap B \neq \emptyset$ and let $x_1 \in A \cap B$. Let $H_A, H_B \in \mathcal{H}$ and $x_{2,A}, x_{2,B} \in \mathcal{X}_2$ be such that $A = P_{1|x_{2,A}}(H_A)$ and $B = P_{2|x_{2,B}}(H_B)$. Using Lemma 2.29, choose an $\mathcal{H}$-repeatable sequence $\mathfrak{X}$ which can take the second coordinate from $x_{2,A}$ to $x_{2,B}$ while keeping the first coordinate unchanged.

Since $x_1 \in A = P_{1|x_{2,A}}(H_A)$ and $x_1 \in B = P_{1|x_{2,B}}(H_B)$, we have $(x_1, x_{2,A}) \in H_A$ and $(x_1, x_{2,B}) \in H_B$. It follows from Lemma 2.30 that $H_B = H_A * \mathfrak{X}$.

Now for every $x_1' \in A = P_{1|x_{2,A}}(H_A)$, we have $(x_1', x_{2,A}) \in H_A$ and so by Lemma 2.29 we have $(x_1', x_{2,B}) \in (x_1', x_{2,A}) * \mathfrak{X} \subset H_A * \mathfrak{X} = H_B$. We conclude that $x_1' \in P_{1|x_{2,B}}(H_B) = B$ for every $x_1' \in A$. Therefore, $A \subset B$. By exchanging the roles of A and B we can also get $B \subset A$ which implies that $A = B$. We conclude that $\mathcal{L}_1(\mathcal{H})$ is a partition of $\mathcal{X}_1$. A similar argument shows that $\mathcal{L}_2(\mathcal{H})$ is a partition of $\mathcal{X}_2$. □

Lemma 2.34. *$\mathcal{L}_1(\mathcal{H})$ (resp. $\mathcal{L}_2(\mathcal{H})$) is a balanced partition of $\mathcal{X}_1$ (resp. $\mathcal{X}_2$).*

Proof. Let $A, B \in \mathcal{L}_1(\mathcal{H})$. There exist $H_A, H_B \in \mathcal{H}$ and $x_{2,A}, x_{2,B} \in \mathcal{X}_2$ such that $A = P_{1|x_{2,A}}(H_A)$ and $B = P_{2|x_{2,B}}(H_B)$. Fix $x_{1,A} \in A$ and $x_{1,B} \in B$ and define $k = \text{per}(\mathcal{H}) \cdot \max\{\text{con}(*_1), \text{con}(*_2)\}$. Clearly, $(x_{1,A}, x_{2,A}) \in H_A$ and $(x_{1,B}, x_{2,B}) \in H_B$.

Since $k \geq \text{con}(*_1)$ and $k \geq \text{con}(*_2)$, and since $*_1$ and $*_2$ are ergodic, there exist a sequence $(x_{1,i})_{0 \leq i < k}$ in $\mathcal{X}_1$ and a sequence $(x_{2,i})_{0 \leq i < k}$ in $\mathcal{X}_2$ such that:

$$\begin{aligned} (\dots((x_{1,A} *_1 x_{1,0}) *_1 x_{1,1}) \dots *_1 x_{1,k-1}) &= x_{1,B}, \\ (\dots((x_{2,A} *_2 x_{2,0}) *_2 x_{2,1}) \dots *_2 x_{2,k-1}) &= x_{2,B}. \end{aligned} \tag{2.15}$$

Now define the $\mathcal{H}$-repeatable sequence $\mathfrak{X} = (X_i)_{0 \leq i < k}$ such that $(x_{1,i}, x_{2,i}) \in X_i \in \mathcal{H}^{i*}$ for all $0 \leq i < k$. We have:

$$(x_{1,B}, x_{2,B}) \overset{(a)}{=} (x_{1,A}, x_{2,A}) * \Big((x_{1,i}, x_{2,i})_{0 \leq i < k}\Big) \overset{(b)}{\in} H_A * \mathfrak{X},$$

where (a) follows from (2.15) and (b) follows from the fact that $(x_{1,A}, x_{2,A}) \in H_A$ and $(x_{1,i}, x_{2,i}) \in X_i$ for every $0 \leq i < k$. We conclude that $H_B \cap (H_A * \mathfrak{X}) \neq \emptyset$. On the other hand, we have $H_B \in \mathcal{H}$ and $H_A * \mathfrak{X} \in \mathcal{H}^{k*} = \mathcal{H}$. Therefore, $H_B = H_A * \mathfrak{X}$ since $\mathcal{H}$ is a partition.

Define the mapping $\pi_1 : \mathcal{X}_1 \to \mathcal{X}_1$ as $\pi_1(x_1) = (\dots((x_1 *_1 x_{1,0}) *_1 x_{1,1}) \dots *_1 x_{1,k-1})$ for every $x_1 \in \mathcal{X}_1$ and the mapping $\pi_2 : \mathcal{X}_2 \to \mathcal{X}_2$ as $\pi_2(x_2) = (\dots((x_2 *_2 x_{2,0}) *_2 x_{2,1}) \dots *_2 x_{2,k-1})$ for every $x_2 \in \mathcal{X}_2$.

Now let $x_1 \in A = P_{1|x_{2,A}}(H_A)$, we have:

$$(\pi_1(x_1), x_{2,B}) \overset{(a)}{=} (\pi_1(x_1), \pi_2(x_{2,A})) \overset{(b)}{=} (x_1, x_{2,A}) * \Big((x_{1,i}, x_{2,i})_{0 \leq i < k}\Big) \overset{(c)}{\in} H_A * \mathfrak{X} = H_B,$$

where (a) follows from (2.15), (b) follows from the definition of π_1 and π_2 and (c) follows from the fact that $(x_1, x_{2,A}) \in H_A$ and $(x_{1,i}, x_{2,i}) \in X_i$ for every $0 \leq i < k$.

We conclude that $\pi_1(x_1) \in P_{1|x_{2,B}}(H_B) = B$ for every $x_1 \in A$. Therefore, $\pi_1(A) \subset B$, which implies that $|A| \overset{(a)}{=} |\pi_1(A)| \leq |B|$, where (a) follows from the fact that π_1 is a permutation. By exchanging the roles of A and B we can also get $|B| \leq |A|$ which implies that $|A| = |B|$. We conclude that $\mathcal{L}_1(\mathcal{H})$ is a balanced partition of $\mathcal{X}_1$ as Lemma 2.33 already showed that $\mathcal{L}_1(\mathcal{H})$ is a partition. A similar argument shows that $\mathcal{L}_2(\mathcal{H})$ is a balanced partition of $\mathcal{X}_2$. □

Lemma 2.35. *For every $i \geq 0$ and every $A \in \mathcal{L}_1(\mathcal{H})^{i*_1}$, there exists $B \in \mathcal{L}_1(\mathcal{H}^{i*})$ such that $A \subset B$.*

Proof. We will prove the lemma by induction on $i \geq 0$. The lemma is trivial for $i = 0$.

Now let $i > 0$ and suppose that the lemma is true for $i - 1$. Let $A \in \mathcal{L}_1(\mathcal{H})^{i*_1}$, there exist $A', A'' \in \mathcal{L}_1(\mathcal{H})^{(i-1)*_1}$ such that $A = A' *_1 A''$. From the induction hypothesis, there exist $B', B'' \in \mathcal{L}_1(\mathcal{H}^{(i-1)*})$ such that $A' \subset B'$ and $A'' \subset B''$. This means that there exist $H', H'' \in \mathcal{H}^{(i-1)*}$ and $x_2', x_2'' \in \mathcal{X}_2$ such that $B' = P_{1|x_2'}(H')$ and $B'' = P_{1|x_2''}(H'')$. We have:

$$A = A' *_1 A'' \subset B' *_1 B'' = P_{1|x_2'}(H') *_1 P_{1|x_2''}(H'') \overset{(a)}{\subset} P_{1|x_2' *_2 x_2''}(H' * H''),$$

where (a) follows from the fact that for every $x_1' \in P_{1|x_2'}(H')$ and $x_1'' \in P_{1|x_2''}(H'')$, we have $(x_1', x_2') \in H'$ and $(x_1'', x_2'') \in H''$, and so $(x_1' *_1 x_1'', x_2' *_2 x_2'') = (x_1', x_2') * (x_1'', x_2'') \in H' * H''$, which implies that $x_1' *_1 x_1'' \in P_{1|x_2' *_2 x_2''}(H' * H'')$.

If we define $B = P_{1|x_2' *_2 x_2''}(H' * H'') \in \mathcal{L}_1(\mathcal{H}^{i*})$, we get $A \subset B$. We conclude that the lemma is true for all $i \geq 0$. □

Lemma 2.36. *$\mathcal{L}_1(\mathcal{H})$ (resp. $\mathcal{L}_2(\mathcal{H})$) is a stable partition of $\mathcal{X}_1$ (resp. $\mathcal{X}_2$) of period of at most $\mathrm{per}(\mathcal{H})$. Moreover, for every $i \geq 0$, we have $\mathcal{L}_1(\mathcal{H})^{i*_1} = \mathcal{L}_1(\mathcal{H}^{i*})$ and $\mathcal{L}_2(\mathcal{H})^{i*_2} = \mathcal{L}_2(\mathcal{H}^{i*})$.*

Proof. We will only prove the lemma for $\mathcal{L}_1(\mathcal{H})$ since the proof for $\mathcal{L}_2(\mathcal{H})$ is similar.

Let $p = \mathrm{per}(\mathcal{H})$. According to Lemma 2.35, for every $A \in \mathcal{L}_1(\mathcal{H})^{p*_1}$, there exists $B \in \mathcal{L}_1(\mathcal{H}^{p*}) = \mathcal{L}_1(\mathcal{H})$ such that $A \subset B$. On the other hand, we have:

$$|A| \geq \|\mathcal{L}_1(\mathcal{H})^{p*_1}\|_\wedge \overset{(a)}{\geq} \|\mathcal{L}_1(\mathcal{H})\|_\wedge \overset{(b)}{=} \|\mathcal{L}_1(\mathcal{H})\| = |B|,$$

where (a) follows from Lemma 2.1 and (b) follows from the fact that $\mathcal{L}_1(\mathcal{H})$ is a balanced partition (Lemma 2.34). We conclude that $A = B \in \mathcal{L}_1(\mathcal{H})$ since $|A| \geq |B|$ and $A \subset B$. Now since this is true for every $A \in \mathcal{L}_1(\mathcal{H})^{p*_1}$, we have $\mathcal{L}_1(\mathcal{H})^{p*_1} \subset \mathcal{L}_1(\mathcal{H})$ which implies that $\mathcal{L}_1(\mathcal{H})^{p*_1} = \mathcal{L}_1(\mathcal{H})$ since $\mathcal{L}_1(\mathcal{H})$ is a partition of $\mathcal{X}_1$ and $\mathcal{L}_1(\mathcal{H})^{p*_1}$ is an $\mathcal{X}_1$-cover. We conclude that $\mathcal{L}_1(\mathcal{H})$ is a stable partition of period of at most $p = \text{per}(\mathcal{H})$. Now since this is true for every stable partition and since $\mathcal{H}^{i*}$ is a stable partition for every $i \geq 0$, we conclude that $\mathcal{L}_1(\mathcal{H}^{i*})$ is a stable partition for every $i \geq 0$. This implies that $\mathcal{L}_1(\mathcal{H}^{i*})^{j*_1}$ is a stable partition for every $i \geq 0$ and every $j \geq 0$.

For every $i > 0$, Lemma 2.35 (applied to $\mathcal{H}^{(i-1)*}$) implies that $\mathcal{L}_1(\mathcal{H}^{(i-1)*})^{*_1}$ is a sub-stable partition of $\mathcal{L}_1(\mathcal{H}^{i*})$ and so $\|\mathcal{L}_1(\mathcal{H}^{(i-1)*})\| = \|\mathcal{L}_1(\mathcal{H}^{(i-1)*})^{*_1}\| \leq \|\mathcal{L}_1(\mathcal{H}^{i*})\|$. Therefore,

$$\|\mathcal{L}_1(\mathcal{H})\| \leq \|\mathcal{L}_1(\mathcal{H}^*)\| \leq \ldots \leq \|\mathcal{L}_1(\mathcal{H}^{p*})\| = \|\mathcal{L}_1(\mathcal{H})\|.$$

We conclude that $\|\mathcal{L}_1(\mathcal{H}^{i*})\| = \|\mathcal{L}_1(\mathcal{H}^{(i \bmod p)*})\| = \|\mathcal{L}_1(\mathcal{H})\|$ for every $i \geq 0$. Moreover, since $\mathcal{L}_1(\mathcal{H})$ is stable, we have $\|\mathcal{L}_1(\mathcal{H})^{i*_1}\| = \|\mathcal{L}_1(\mathcal{H})\|$, which implies that $\|\mathcal{L}_1(\mathcal{H})^{i*_1}\| = \|\mathcal{L}_1(\mathcal{H}^{i*})\|$ for every $i \geq 0$.

Now for every $i \geq 0$, $\mathcal{L}_1(\mathcal{H})^{i*_1}$ is a sub-stable partition of $\mathcal{L}_1(\mathcal{H}^{i*})$ (by Lemma 2.35) and we have just shown that $\|\mathcal{L}_1(\mathcal{H})^{i*_1}\| = \|\mathcal{L}_1(\mathcal{H}^{i*})\|$. We conclude that $\mathcal{L}_1(\mathcal{H})^{i*_1} = \mathcal{L}_1(\mathcal{H}^{i*})$ for every $i \geq 0$. □

Now we are ready to prove Theorem 2.4:

Proof of Theorem 2.4. Lemma 2.36 shows that $\mathcal{L}_1(\mathcal{H})$ and $\mathcal{L}_2(\mathcal{H})$ are stable partitions of $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively, and Lemma 2.32 shows that $\mathcal{U}_1(\mathcal{H})$ and $\mathcal{U}_2(\mathcal{H})$ are stable partitions of $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively. Moreover, Lemma 2.36 shows that $\mathcal{L}_1(\mathcal{H})^{i*_1} = \mathcal{L}_1(\mathcal{H}^{i*})$ and $\mathcal{L}_2(\mathcal{H})^{i*_2} = \mathcal{L}_2(\mathcal{H}^{i*})$ for every $i > 0$, and Lemma 2.32 shows that $\mathcal{U}_1(\mathcal{H})^{i*_1} = \mathcal{U}_1(\mathcal{H}^{i*})$ and $\mathcal{U}_2(\mathcal{H})^{i*_2} = \mathcal{U}_2(\mathcal{H}^{i*})$ for every $i > 0$.

It is easy to see that $\mathcal{L}_1(\mathcal{H}) \preceq \mathcal{U}_1(\mathcal{H})$ and $\mathcal{L}_2(\mathcal{H}) \preceq \mathcal{U}_2(\mathcal{H})$. Now we turn to show that $\mathcal{L}_1(\mathcal{H}) \otimes \mathcal{L}_2(\mathcal{H}) \preceq \mathcal{H} \preceq \mathcal{U}_1(\mathcal{H}) \otimes \mathcal{U}_2(\mathcal{H})$. Let $A \times B \in \mathcal{L}_1(\mathcal{H}) \otimes \mathcal{L}_2(\mathcal{H})$ (i.e., $A \in \mathcal{L}_1(\mathcal{H})$ and $B \in \mathcal{L}_2(\mathcal{H})$), and fix $x_1 \in A$ and $x_2 \in B$. Let $H \in \mathcal{H}$ be such that $(x_1, x_2) \in H$. We have $x_1 \in P_{1|x_2}(H)$ as $(x_1, x_2) \in H$. Therefore, $P_{1|x_2}(H) \cap A \neq \emptyset$ which implies that $A = P_{1|x_2}(H)$ since both A and $P_{1|x_2}(H)$ are in $\mathcal{L}_1(\mathcal{H})$ which was shown to be a stable partition.

Now fix $(x_A, x_B) \in A \times B$. Since $x_A \in A = P_{1|x_2}(H)$, we have $(x_A, x_2) \in H$ which means that $x_2 \in P_{2|x_A}(H)$. Therefore, $B \cap P_{2|x_A}(H) \neq \emptyset$ which implies that $B = P_{2|x_A}(H)$ since both B and $P_{2|x_A}(H)$ are in $\mathcal{L}_2(\mathcal{H})$ which was shown to be a stable partition. Now since $x_B \in B = P_{2|x_A}(H)$, we conclude that $(x_A, x_B) \in H$. But this is true for all $(x_A, x_B) \in A \times B$, hence $A \times B \subset H$. Therefore, $\mathcal{L}_1(\mathcal{H}) \otimes \mathcal{L}_2(\mathcal{H}) \preceq \mathcal{H}$.

In order to prove that $\mathcal{H} \preceq \mathcal{U}_1(\mathcal{H}) \otimes \mathcal{U}_2(\mathcal{H})$, let $H \in \mathcal{H}$, $A' = P_1(H) \in \mathcal{U}_1(\mathcal{H})$ and $B' = P_2(H) \in \mathcal{U}_2(\mathcal{H})$. Clearly, $H \subset A' \times B'$, hence $\mathcal{H} \preceq \mathcal{U}_1(\mathcal{H}) \otimes \mathcal{U}_2(\mathcal{H})$.

Now let $H \in \mathcal{H}$. Since $\mathcal{L}_1(\mathcal{H}) \otimes \mathcal{L}_2(\mathcal{H}) \preceq \mathcal{H}$, there exist an integer $n_H > 0$ and n_H sets $H_1, \ldots, H_{n_H} \in \mathcal{L}_1(\mathcal{H}) \otimes \mathcal{L}_2(\mathcal{H})$ such that $H_1, \ldots, H_{n_H}$ are disjoint and $H = H_1 \cup \ldots \cup H_{n_H}$. Since $H_1, \ldots, H_{n_H} \in \mathcal{L}_1(\mathcal{H}) \otimes \mathcal{L}_2(\mathcal{H})$, there exist n_H sets $H_{1,1}, \ldots, H_{1,n_H} \in \mathcal{L}_1(\mathcal{H})$ and n_H sets $H_{2,1}, \ldots, H_{2,n_H} \in \mathcal{L}_2(\mathcal{H})$ such that $H_1 = H_{1,1} \times H_{2,1}$, ..., and $H_{n_H} = H_{1,n_H} \times H_{2,n_H}$. Clearly, $H_{1,i} = P_1(H_i)$ and $H_{2,i} = P_2(H_i)$ for every $1 \leq i \leq n_H$. We have:

- $H_{1,1} \cup \ldots \cup H_{1,n_H} = P_1(H_1) \cup \ldots \cup P_1(H_{n_H}) = P_1(H_1 \cup \ldots \cup H_{n_H}) = P_1(H) \in \mathcal{U}_1(\mathcal{H})$.

- $H_{2,1} \cup \ldots \cup H_{2,n_H} = P_2(H_1) \cup \ldots \cup P_2(H_{n_H}) = P_2(H_1 \cup \ldots \cup H_{n_H}) = P_2(H) \in \mathcal{U}_2(\mathcal{H})$.

- Suppose that $H_{1,i} = H_{1,j}$ for some $i \neq j$ and let $x_1 \in H_{1,i} = H_{1,j}$, then $H_{2,i} \cup H_{2,j} \subset P_{2|x_1}(H) \in \mathcal{L}_2(\mathcal{H})$ which cannot happen unless $H_{2,i} = H_{2,j} = P_{2|x_1}(H)$. This is a contradiction since $(H_{1,i} \times H_{2,i})$ and $(H_{1,j} \times H_{2,j})$ are disjoint. We conclude that $H_{1,1}, \ldots, H_{1,n_H}$ are disjoint. Similarly, $H_{2,1}, \ldots, H_{2,n_H}$ are also disjoint.

Now since $H_{1,1}, \ldots, H_{1,n_H}$ are disjoint, we have $\|\mathcal{U}_1(\mathcal{H})\| = |P_1(H)| = |H_{1,1}| + \cdots + |H_{1,n_H}| = n_H \|\mathcal{L}_1(\mathcal{H})\|$. Therefore, $n_H = \frac{\|\mathcal{U}_1(\mathcal{H})\|}{\|\mathcal{L}_1(\mathcal{H})\|}$. Similarly, $n_H = \frac{\|\mathcal{U}_2(\mathcal{H})\|}{\|\mathcal{L}_2(\mathcal{H})\|}$. We conclude that n_H is the same for all $H \in \mathcal{H}$. Let us denote this common integer as n. It is now easy to see that $\|\mathcal{H}\| = n \cdot \|\mathcal{L}_1(\mathcal{H})\| \cdot \|\mathcal{L}_2(\mathcal{H})\| = \|\mathcal{L}_1(\mathcal{H})\| \cdot \|\mathcal{U}_2(\mathcal{H})\| = \|\mathcal{U}_1(\mathcal{H})\| \cdot \|\mathcal{L}_2(\mathcal{H})\|$.

Now in order to prove the uniqueness of $\mathcal{L}_1(\mathcal{H})$, $\mathcal{L}_2(\mathcal{H})$, $\mathcal{U}_1(\mathcal{H})$ and $\mathcal{U}_2(\mathcal{H})$, suppose that $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}'_1$, $\mathcal{H}'_2$, and $n' > 0$ satisfy the conditions of the theorem (i.e. $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}'_1$, $\mathcal{H}'_2$ and n' play the roles of $\mathcal{L}_1(\mathcal{H})$, $\mathcal{L}_2(\mathcal{H})$, $\mathcal{U}_1(\mathcal{H})$, $\mathcal{U}_2(\mathcal{H})$ and n respectively). Let $H \in \mathcal{H}$, then there exist n' disjoint sets $H'_{1,1}, \ldots, H'_{1,n'} \in \mathcal{H}_1$ and n' disjoint sets $H'_{2,1}, \ldots, H'_{2,n'} \in \mathcal{H}_2$ such that:

- $H'_{1,1} \cup \ldots \cup H'_{1,n'} \in \mathcal{H}'_1$.

- $H'_{2,1} \cup \ldots \cup H'_{2,n'} \in \mathcal{H}'_2$.

- $H = (H'_{1,1} \times H'_{2,1}) \cup \ldots \cup (H'_{1,n'} \times H'_{2,n'})$.

Since $H = (H'_{1,1} \times H'_{2,1}) \cup \ldots \cup (H'_{1,n'} \times H'_{2,n'})$, we have $P_1(H) = H'_{1,1} \cup \ldots \cup H'_{1,n'} \in \mathcal{H}'_1$. But this is true for every $H \in \mathcal{H}$. Therefore, $\mathcal{U}_1(\mathcal{H}) \subset \mathcal{H}'_1$ which implies that $\mathcal{H}'_1 = \mathcal{U}_1(\mathcal{H})$ since $\mathcal{H}'_1$ and $\mathcal{U}_1(\mathcal{H})$ are partitions. Similarly, $\mathcal{H}'_2 = \mathcal{U}_2(\mathcal{H})$.

Now let $x_2 \in \mathcal{X}_2$ be such that $P_{1|x_2}(H) \neq \emptyset$. Clearly, $x_2 \in H'_{2,i}$ for some $1 \leq i \leq n'$ and so $P_{1|x_2}(H) = H'_{1,i} \in \mathcal{H}_1$ since $H = (H'_{1,1} \times H'_{2,1}) \cup \ldots \cup (H'_{1,n'} \times H'_{2,n'})$ and since $H'_{2,1}, \ldots, H'_{2,n'}$ are disjoint. Therefore, for every $x_2 \in \mathcal{X}_2$ satisfying $P_{1|x_2}(H) \neq \emptyset$, we have $P_{1|x_2}(H) \in \mathcal{H}_1$. We conclude that $\mathcal{L}_1(\mathcal{H}) \subset \mathcal{H}_1$ which implies that $\mathcal{H}_1 = \mathcal{L}_1(\mathcal{H})$ since $\mathcal{H}_1$ and $\mathcal{L}_1(\mathcal{H})$ are partitions. Similarly, $\mathcal{H}_2 = \mathcal{L}_2(\mathcal{H})$. Moreover, $n' = \frac{\|\mathcal{H}'_1\|}{\|\mathcal{H}_1\|} = \frac{\|\mathcal{U}_1(\mathcal{H})\|}{\|\mathcal{L}_1(\mathcal{H})\|} = n$.

We conclude that the stable partitions $\mathcal{L}_1(\mathcal{H})$, $\mathcal{L}_2(\mathcal{H})$, $\mathcal{U}_1(\mathcal{H})$, $\mathcal{U}_2(\mathcal{H})$ are unique. □

Proof of Theorem 2.5

For Theorem 2.5, we will first prove it for $m = 2$ using two lemmas. The general result can then be proven by induction on $m \geq 2$.

Lemma 2.37. *If $* = *_1 \otimes *_2$ is a strongly ergodic operation on $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$, then $*_1$ and $*_2$ are strongly ergodic.*

Proof. Let $\mathcal{H}_1$ be a stable partition of $\mathcal{X}_1$, then $\mathcal{H} = \mathcal{H}_1 \otimes \{\mathcal{X}_2\}$ is a stable partition of $\mathcal{X}_1 \times \mathcal{X}_2$. Fix $x_2 \in \mathcal{X}_2$ and let $x_1 \in \mathcal{X}_1$. Since $*$ is strongly ergodic, then by Definition 2.14 there exists $n = n(x_1, x_2, \mathcal{H}) > 0$ such that for every $H \in \mathcal{H}^{n*}$, there exists an $\mathcal{H}$-sequence $\mathfrak{X} = (X_i)_{0 \le i < n}$ satisfying $(x_1, x_2) * \mathfrak{X} = H$. Let $H_1 \in \mathcal{H}_1^{n*_1}$. Clearly, $H_1 \times \mathcal{X}_2 \in \mathcal{H}_1^{n*_1} \otimes \{\mathcal{X}_2\} = (\mathcal{H}_1 \otimes \{\mathcal{X}_2\})^{n*} = \mathcal{H}^{n*}$.

Since $H_1 \times \mathcal{X}_2 \in \mathcal{H}^{n*}$, there exists an $\mathcal{H}$-sequence $\mathfrak{X} = (X_i)_{0 \le i < n}$ such that $(x_1, x_2) * \mathfrak{X} = H_1 \times \mathcal{X}_2$. For every $0 \le i < n$, $X_i \in (\mathcal{H}_1 \otimes \{\mathcal{X}_2\})^{i*} = \mathcal{H}_1^{i*_1} \otimes \{\mathcal{X}_2\}$ and so there exists $X_{1,i} \in \mathcal{H}_1^{i*_1}$ such that $X_i = X_{1,i} \times \mathcal{X}_2$. By projecting the equation $(x_1, x_2) * \mathfrak{X} = H_1 \times \mathcal{X}_2$ on the first coordinate, we get $x_1 *_1 \mathfrak{X}_1 = H_1$, where $\mathfrak{X}_1$ is the $\mathcal{H}_1$-sequence $(X_{1,i})_{0 \le i < n}$. By fixing $x_2 \in \mathcal{X}_2$, n will depend only on x_1 and $\mathcal{H}_1$ as required in the definition of strong ergodicity. This proves that $*_1$ is strongly ergodic. A similar argument shows that $*_2$ is also strongly ergodic. □

Lemma 2.38. *If $*_1$ and $*_2$ are two strongly ergodic operations on $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively, then $* = *_1 \otimes *_2$ is a strongly ergodic operation on $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$.*

Proof. Fix a stable partition $\mathcal{H}$ of $\mathcal{X}$. Since $*_1$ and $*_2$ are strongly ergodic, they are ergodic and so Theorem 2.4 can be applied. Let $\mathcal{L}_1(\mathcal{H})$, $\mathcal{L}_2(\mathcal{H})$, $\mathcal{U}_1(\mathcal{H})$ and $\mathcal{U}_2(\mathcal{H})$ be defined as in Theorem 2.4, and let P_1 and P_2 be the projection onto the first and second coordinate respectively as in Definition 2.25.

Let $(x_1, x_2) \in H \in \mathcal{H}$. We will construct an $\mathcal{H}$-augmenting sequence $\mathfrak{X}$ satisfying $H \subset (x_1, x_2) * \mathfrak{X}$ in two steps: We first construct an $\mathcal{H}$-augmenting sequence $\mathfrak{X}_U$ such that $P_1(H) \subset P_1\big((x_1, x_2) * \mathfrak{X}_U\big)$, i.e., $\mathfrak{X}_U$ stretches $\{(x_1, x_2)\}$ in the direction of the first coordinate to cover $P_1(H)$. In the second step, we construct an $\mathcal{H}$-augmenting sequence $\mathfrak{X}_L$ such that $H \subset \big((x_1, x_2) * \mathfrak{X}_U\big) * \mathfrak{X}_L$, i.e., $\mathfrak{X}_L$ stretches $(x_1, x_2) * \mathfrak{X}_U$ in the direction of the second coordinate to cover H.

Step 1: Let $H_1 = P_1(H) \in \mathcal{U}_1(\mathcal{H})$. Since $*_1$ is strongly ergodic, there exists a $\mathcal{U}_1(\mathcal{H})$-augmenting sequence $\mathfrak{X}_1$ such that $x_1 *_1 \mathfrak{X}_1 = H_1$. Let $\mathfrak{X}'_1 = (X'_{1,i})_{0 \le i < k'} = (\mathfrak{X}_1)^{\mathrm{per}(\mathcal{H})}$. For every $0 \le i < k' = |\mathfrak{X}'_1|$, we have $X'_{1,i} \in \mathcal{U}_1(\mathcal{H})^{i*_1} = \mathcal{U}_1(\mathcal{H}^{i*})$, and so from Definition 2.25 there exists $X'_i \in \mathcal{H}^{i*}$ such that $P_1(X'_i) = X'_{1,i}$. Define the $\mathcal{H}$-sequence $\mathfrak{X}'_U = (X'_i)_{0 \le i < k'}$. The sequence $\mathfrak{X}'_U$ is $\mathcal{H}$-repeatable since $\mathrm{per}(\mathcal{H})$ divides $|\mathfrak{X}'_U| = k' = |\mathfrak{X}_1| \cdot \mathrm{per}(\mathcal{H})$. By Lemma 2.7, there exists $l > 0$ such that $\mathfrak{X}_U := (\mathfrak{X}'_U)^l$ is $\mathcal{H}$-augmenting. We have:

$$\begin{aligned} H_1 &\overset{(a)}{\subset} H_1 *_1 (\mathfrak{X}_1)^{\mathrm{per}(\mathcal{H})l-1} = (x_1 *_1 \mathfrak{X}_1) *_1 (\mathfrak{X}_1)^{\mathrm{per}(\mathcal{H})l-1} \\ &= x_1 *_1 (\mathfrak{X}_1)^{\mathrm{per}(\mathcal{H})l} = x_1 *_1 (\mathfrak{X}'_1)^l = x_1 *_1 \big((X'_{1,i})_{0 \le i < k'}\big)^l \\ &= P_1\big((x_1, x_2)\big) *_1 \Big(\big(P_1(X'_i)\big)_{0 \le i < k'}\Big)^l = P_1\Big((x_1, x_2) * \big((X'_i)_{0 \le i < k'}\big)^l\Big) \\ &= P_1\big((x_1, x_2) * (\mathfrak{X}'_U)^l\big) = P_1\big((x_1, x_2) * \mathfrak{X}_U\big), \end{aligned} \tag{2.16}$$

where (a) follows from the fact that $\mathfrak{X}_1$ is $\mathcal{U}_1(\mathcal{H})$-augmenting.

Step 2: Define $X_U = (x_1, x_2) * \mathfrak{X}_U$. Since $\mathfrak{X}_U$ is $\mathcal{H}$-augmenting, we must have $X_U \subset K$, where $K \in \mathcal{K}_\mathcal{H}$ is such that $(x_1, x_2) \in K$ (see Theorem 2.1). Now since $\mathcal{K}_\mathcal{H}$ is a sub-stable partition of $\mathcal{H}$ (by Theorem 2.1) and since $(x_1, x_2) \in K \cap H$, we must have $K \subset H$. Therefore, $X_U \subset H$. On the other hand, from (2.16) we have $H_1 \subset P_1(X_U)$. We conclude that for every $a \in H_1$, we have $a \in P_1(X_U)$ and so there exists $b_a \in \mathcal{X}_2$ such that $(a, b_a) \in X_U \subset H$.

According to Theorem 2.4, there exist n disjoint sets $H_{1,1},\ldots,H_{1,n}\in\mathcal{L}_1(\mathcal{H})$ and n disjoint sets $H_{2,1},\ldots,H_{2,n}\in\mathcal{L}_2(\mathcal{H})$ such that $H=(H_{1,1}\times H_{2,1})\cup\ldots\cup(H_{1,n}\times H_{2,n})$. For every $a\in H_1=H_{1,1}\cup\ldots\cup H_{1,n}$, there exists a unique $1\leq i_a\leq n$ such that $a\in H_{1,i_a}$. We have:

$$\begin{aligned}H&=\bigcup_{1\leq i\leq n}(H_{1,i}\times H_{2,i})=\bigcup_{1\leq i\leq n}\bigcup_{a\in H_{1,i}}(\{a\}\times H_{2,i})\\&=\bigcup_{1\leq i\leq n}\bigcup_{a\in H_{1,i}}(\{a\}\times H_{2,i_a})=\bigcup_{a\in H_1}(\{a\}\times H_{2,i_a}).\end{aligned}\tag{2.17}$$

Fix $a\in H_1$. Since $(a,b_a)\in H=\bigcup_{a'\in H_1}(\{a'\}\times H_{2,i_{a'}})$, we must have $b_a\in H_{2,i_a}\in\mathcal{L}_2(\mathcal{H})$. Now since $*_2$ is strongly ergodic, there exists an $\mathcal{L}_2(\mathcal{H})$-augmenting sequence $\mathfrak{X}_{2,a}$ such that $b_a*_2\mathfrak{X}_{2,a}=H_{2,i_a}$. Let $\mathfrak{X}'_{2,a}=(X'_{2,a,i})_{0\leq i<k'_a}=(\mathfrak{X}_{2,a})^{\mathrm{per}(\mathcal{H})}$. For every $0\leq i<k'_a$, we have $X'_{2,a,i}\in\mathcal{L}_2(\mathcal{H})^{i*_2}=\mathcal{L}_2(\mathcal{H}^{i*})$, and so from Definition 2.26 there exist $x'_{1,a,i}\in\mathcal{X}_1$ and $X'_{a,i}\in\mathcal{H}^{i*}$ such that $X'_{2,a,i}=P_{2|x'_{1,a,i}}(X'_{a,i})$. Define the $\mathcal{H}$-sequence $\mathfrak{X}'_a=(X'_{a,i})_{0\leq i<k'_a}$. The sequence $\mathfrak{X}'_a$ is $\mathcal{H}$-repeatable since $\mathrm{per}(\mathcal{H})$ divides $|\mathfrak{X}'_a|=k'_a=|\mathcal{X}_{2,a}|\cdot\mathrm{per}(\mathcal{H})$.

Define the mapping $\pi_a:\mathcal{X}_1\to\mathcal{X}_1$ as $\pi_a(x)=(((x*_1x'_{1,a,0})*_1x'_{1,a,1})\ldots*_1x'_{1,a,k'_a-1})$ for every $x\in\mathcal{X}_1$. Since π_a is a permutation, there exists $p_a>0$ such that $\pi_a^{p_a}(x)=x$ for every $x\in\mathcal{X}_1$. $(\mathfrak{X}'_a)^{p_a}$ is $\mathcal{H}$-repeatable since $\mathfrak{X}'_a$ is $\mathcal{H}$-repeatable. Now by Lemma 2.7 there exists $l_a>0$ such that $\mathfrak{X}_a:=(\mathfrak{X}'_a)^{p_al_a}$ is $\mathcal{H}$-augmenting. We have:

$$\begin{aligned}\{a\}\times H_{2,i_a}&\overset{(a)}{\subset}\{a\}\times\left(H_{2,i_a}*_2(\mathfrak{X}_{2,a})^{\mathrm{per}(\mathcal{H})p_al_a-1}\right)\\&=\{a\}\times\left((b_a*_2\mathfrak{X}_{2,a})*_2(\mathfrak{X}_{2,a})^{\mathrm{per}(\mathcal{H})p_al_a-1}\right)\\&\overset{(b)}{=}\{\pi_a^{p_al_a}(a)\}\times\left(b_a*_2(\mathfrak{X}_{2,a})^{\mathrm{per}(\mathcal{H})p_al_a}\right)\\&=\{\pi_a^{p_al_a}(a)\}\times\left(b_a*_2(\mathfrak{X}'_{2,a})^{p_al_a}\right)\\&\overset{(c)}{=}(a,b_a)*\left(\left(\{x'_{1,a,i}\}\times X'_{2,a,i}\right)_{0\leq i<k_a}\right)^{p_al_a}\\&\overset{(d)}{\subset}(a,b_a)*\left((X'_{a,i})_{0\leq i<k_a}\right)^{p_al_a}\\&=(a,b_a)*(\mathfrak{X}'_a)^{p_al_a}=(a,b_a)*\mathfrak{X}_a.\end{aligned}$$

(a) follows from the fact that $\mathfrak{X}_{2,a}$ is $\mathcal{L}_2(\mathcal{H})$-augmenting, hence $(\mathfrak{X}_{2,a})^{\mathrm{per}(\mathcal{H})p_al_a-1}$ is $\mathcal{L}_2(\mathcal{H})$-augmenting (by Remark 2.7), and so $H_{2,i_a}\subset H_{2,i_a}*_2(\mathfrak{X}_{2,a})^{\mathrm{per}(\mathcal{H})p_al_a-1}$. (b) follows from the fact that $\pi_a^{p_a}(x)=x$ for every $x\in\mathcal{X}_1$, which implies that $\pi_a^{p_al_a}(a)=a$. (c) follows from the definition of π_a and from the fact that $\mathfrak{X}'_{2,a}=(X'_{2,a,i})_{0\leq i<k_a}$. (d) follows from the fact that $P_{2|x'_{1,a,i}}(X'_{a,i})=X'_{2,a,i}$, which implies that $\{x'_{1,a,i}\}\times X'_{2,a,i}\subset X'_{a,i}$ for every $0\leq i<k_a$.

Now let $\mathfrak{X}_L=(\mathfrak{X}_a)_{a\in H_1}$ be the $\mathcal{H}$-augmenting sequence obtained by concatenating the $\mathcal{H}$-augmenting sequences $\mathfrak{X}_a$ for all $a\in H_1$ (the order of the concatenation is not important). Since $\{a\}\times H_{2,i_a}\subset(a,b_a)*\mathfrak{X}_a$ for every $a\in H_1$, we must have

$$\{a\}\times H_{2,i_a}\subset(a,b_a)*\mathfrak{X}_L\text{ for every }a\in H_1.\tag{2.18}$$

Define $\mathfrak{X}=(\mathfrak{X}_U,\mathfrak{X}_L)$. We have $(x_1,x_2)*\mathfrak{X}=\left((x_1,x_2)*\mathfrak{X}_U\right)*\mathfrak{X}_L=X_U*\mathfrak{X}_L$. For every $a\in H_1$, we have already shown that $(a,b_a)\in X_U$ and so it follows from

(2.18) that:

$$\{a\} \times H_{2,i_a} \subset (a, b_a) * \mathfrak{X}_L \subset X_U * \mathfrak{X}_L = (x_1, x_2) * \mathfrak{X}.$$

Since this is true for every $a \in H_1$, we have:

$$H \overset{(a)}{=} \bigcup_{a \in H_1} \{a\} \times H_{2,i_a} \subset (x_1, x_2) * \mathfrak{X},$$

where (a) follows from (2.17).

Now since $\mathfrak{X}$ is $\mathcal{H}$-augmenting, Theorem 2.1 implies that $(x_1, x_2) * \mathfrak{X} \subset K$, where $K \in \mathcal{K}_{\mathcal{H}}$ is such that $(x_1, x_2) \in K$. Therefore, $\|\mathcal{H}\| = |H| \leq |(x_1, x_2) * \mathfrak{X}| \leq |K| = \|\mathcal{K}_{\mathcal{H}}\|$. Now since $\mathcal{K}_{\mathcal{H}}$ is a sub-stable partition of $\mathcal{H}$, we conclude that $\mathcal{K}_{\mathcal{H}} = \mathcal{H}$. But this is true for every stable partition $\mathcal{H}$ of $\mathcal{X}$, hence $*$ is strongly ergodic. □

Now we are ready to prove Theorem 2.5:

Proof of Theorem 2.5. Lemmas 2.37 and 2.38 show that Theorem 2.5 is true for $m = 2$. Now let $m > 2$ and suppose that the theorem is true for $m - 1$.

Let $*_1, \ldots, *_m$ be m binary operations such that $*_1 \otimes \ldots \otimes *_m$ is strongly ergodic. It is easy to see that $*_1 \otimes \ldots \otimes *_m$ can be identified to $(*_1 \otimes \ldots \otimes *_{m-1}) \otimes *_m$ (see Notation 2.5). Therefore, $(*_1 \otimes \ldots \otimes *_{m-1}) \otimes *_m$ is strongly ergodic. Lemma 2.37 implies that $*_1 \otimes \ldots \otimes *_{m-1}$ and $*_m$ are strongly ergodic. It then follows from the induction hypothesis that $*_1, \ldots, *_{m-1}$ are strongly ergodic. Therefore, $*_1, \ldots, *_m$ are strongly ergodic.

Conversely, let $*_1, \ldots, *_m$ be m strongly ergodic operations. From the induction hypothesis, we get that $*_1 \otimes \ldots \otimes *_{m-1}$ is strongly ergodic. Lemma 2.38 implies that $(*_1 \otimes \ldots \otimes *_{m-1}) \otimes *_m$ is strongly ergodic. But since $(*_1 \otimes \ldots \otimes *_{m-1}) \otimes *_m$ can be identified to $*_1 \otimes \ldots \otimes *_m$, we conclude that $*_1 \otimes \ldots \otimes *_m$ is strongly ergodic.

Therefore, Theorem 2.5 is true for all $m \geq 2$. □

3 Polarizing Binary Operations

In this chapter[1], we provide a necessary and sufficient condition for a binary operation to be polarizing (in the general multilevel sense). In Section 3.1, we formally define the concept of polarizing binary operations. In Section 3.2, we prove that a binary operation is polarizing if and only if it is uniformity-preserving and its right-inverse is strongly ergodic. In Section 3.3, we explain how we can use a polarizing operation to construct polar codes.

3.1 Formal Definition of Polarizing Binary Operations

Unless we state otherwise, every set that is considered in this chapter is finite.

3.1.1 Easy Channels

Notation 3.1. *A channel W with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ is denoted by $W : \mathcal{X} \longrightarrow \mathcal{Y}$. The transition probabilities of W are denoted by $W(y|x)$, where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Note that we use the long arrow ($\longrightarrow$) in the notation $W : \mathcal{X} \longrightarrow \mathcal{Y}$ and not the short arrow ($\rightarrow$) that we only use to describe mappings. For example, $W : \mathcal{X} \longrightarrow \mathcal{Y}$ denotes a channel, and $V : \mathcal{X} \rightarrow \mathcal{Y}$ denotes a mapping from $\mathcal{X}$ to $\mathcal{Y}$.*

The probability of error of the maximum-likelihood (ML) decoder[2] of W for uniformly distributed input is denoted as $P_e(W)$. The symmetric capacity of W, denoted $I(W)$, is the mutual information $I(X;Y)$, where X and Y are jointly distributed as $P_{X,Y}(x,y) = \frac{1}{|\mathcal{X}|}W(y|x)$ (i.e., X is uniform in $\mathcal{X}$ and it is used as input to the channel W while Y is the output).

Definition 3.1. *A channel $W : \mathcal{X} \longrightarrow \mathcal{Y}$ is said to be δ-*easy *if there exist an integer $L \leq |\mathcal{X}|$ and a random code $\mathcal{B}$ of block length 1 and rate $\log_2 L$ (i.e., $\mathcal{B} \in \mathcal{S} := \{C \subset \mathcal{X} : |C| = L\}$), which satisfy the following:*

- $|I(W) - \log_2 L| < \delta$.

[1] The material of this chapter is based on [15, 18].

[2] The ML decoder is the decoder that minimizes the probability of error.

- *For every $x \in \mathcal{X}$, we have $\sum_{C \in \mathcal{S}} \frac{1}{L} P_{\mathcal{B}}(C) \mathbb{1}_{x \in C} = \frac{1}{|\mathcal{X}|}$. In other words, if $C \in \mathcal{S}$ is chosen according to the distribution of $\mathcal{B}$ and X is chosen uniformly in C, then the marginal distribution of X as a random variable in $\mathcal{X}$ is uniform.*
- *If for each $C \in \mathcal{S}$ we fix a bijection $f_C : \{1, \dots, L\} \to C$, then $I(W_{\mathcal{B}}) > \log_2 L - \delta$, where $W_{\mathcal{B}} : \{1, \dots, L\} \longrightarrow \mathcal{Y} \times \mathcal{S}$ is the channel defined by:*

$$W_{\mathcal{B}}(y, C|a) = W(y|f_C(a)).P_{\mathcal{B}}(C).$$

Note that the value of $I(W_{\mathcal{B}})$ does not depend on the choice of the bijections $(f_C)_{C \in \mathcal{S}}$.

*If we also have $P_e(W_{\mathcal{B}}) < \epsilon$, we say that W is (δ, ϵ)-*easy*.*

If W is δ-easy for a small δ, then we can reliably transmit information near the symmetric capacity of W using a code of blocklength 1 (hence the easiness; there is no need to use codes of large blocklengths): We choose a random code according to $\mathcal{B}$, we reveal this code to the receiver, and then we transmit information using this code. The rate of this code is equal to $\log_2 L$ which is close to the symmetric capacity $I(W)$. On the other hand, the fact that $I(W_{\mathcal{B}}) > \log_2 L - \delta$ means that $W_{\mathcal{B}}$ is almost perfect, which ensures that our simple coding scheme has a low probability of error.

Note that we added (2) to our definition in order to induce a uniform distribution on the input. This is important for the polarization process (see the definition of W^- and W^+ in Definition 3.2: The distribution of U_1 and U_2 are assumed to be uniform in $\mathcal{X}$).

3.1.2 Polarization Process

In this section, we consider an ordinary (single user) channel W and a binary operation $*$ on its input alphabet.

Definition 3.2. *Let $\mathcal{X}$ be an arbitrary set and $*$ be a binary operation on $\mathcal{X}$. Let $W : \mathcal{X} \longrightarrow \mathcal{Y}$ be a channel. We define the two channels $W^- : \mathcal{X} \longrightarrow \mathcal{Y} \times \mathcal{Y}$ and $W^+ : \mathcal{X} \longrightarrow \mathcal{Y} \times \mathcal{Y} \times \mathcal{X}$ as follows:*

$$W^-(y_1, y_2|u_1) = \frac{1}{|\mathcal{X}|} \sum_{u_2 \in \mathcal{X}} W(y_1|u_1 * u_2) W(y_2|u_2),$$

$$W^+(y_1, y_2, u_1|u_2) = \frac{1}{|\mathcal{X}|} W(y_1|u_1 * u_2) W(y_2|u_2).$$

For every $s = (s_1, \dots, s_n) \in \{-, +\}^n$, we define W^s recursively as:

$$W^s := ((W^{s_1})^{s_2} \dots)^{s_n}.$$

Definition 3.3. *Let $(B_n)_{n \geq 1}$ be i.i.d. uniform random variables in $\{-, +\}$. For each channel W with input alphabet $\mathcal{X}$, we define the channel-valued process $(W_n)_{n \geq 0}$ recursively as follows:*

$$W_0 := W,$$
$$W_n := W_{n-1}^{B_n}, \ \forall n \geq 1.$$

Definition 3.4. *A binary operation $*$ is said to be* polarizing *if we have the following two properties:*

- *Conservation property: For every channel W with input alphabet $\mathcal{X}$, we have $I(W^-) + I(W^+) = 2I(W)$.*
- *Polarization property: For every channel W with input alphabet $\mathcal{X}$ and every $\delta > 0$, W_n almost surely becomes δ-easy, i.e.,*

$$\lim_{n\to\infty} \mathbb{P}\big[W_n \text{ is } \delta\text{-easy}\big] = 1.$$

Notation 3.2. *Throughout this chapter, we write $(U_1, U_2) \xrightarrow{f_*} (X_1, X_2) \xrightarrow{W} (Y_1, Y_2)$ to denote the following:*

- *U_1 and U_2 are two independent random variables uniformly distributed in $\mathcal{X}$.*
- *$X_1 = U_1 * U_2$ and $X_2 = U_2$.*
- *The conditional distribution $(Y_1, Y_2)|(X_1, X_2)$ is given by:*

$$P_{Y_1,Y_2|X_1,X_2}(y_1, y_2|x_1, x_2) = W(y_1|x_1)W(y_2|x_2).$$

  I.e., Y_1 and Y_2 are the outputs of two independent copies of the channel W with inputs X_1 and X_2 respectively.

- *$(U_1, U_2) - (X_1, X_2) - (Y_1, Y_2)$ is a Markov chain.*

Note that since $X_1 = U_1 * U_2$ and $X_2 = U_2$, the chain $(X_1, X_2) - (U_1, U_2) - (Y_1, Y_2)$ is also a Markov chain.

Remark 3.1. *Let $(U_1, U_2) \xrightarrow{f_*} (X_1, X_2) \xrightarrow{W} (Y_1, Y_2)$. From the definition of W^- and W^+, it is easy to see that we have $I(W^-) = I(U_1; Y_1, Y_2)$ and $I(W^+) = I(U_2; Y_1, Y_2, U_1)$. Therefore,*

$$\begin{aligned} I(W^-) + I(W^+) &= I(U_1; Y_1, Y_2) + I(U_2; Y_1, Y_2, U_1) \\ &= I(U_1, U_2; Y_1, Y_2) \overset{(a)}{=} I(X_1, X_2; Y_1, Y_2), \end{aligned}$$

where (a) follows from the fact that both $(U_1, U_2) - (X_1, X_2) - (Y_1, Y_2)$ and $(X_1, X_2) - (U_1, U_2) - (Y_1, Y_2)$ are Markov chains. We have the following:

- *If $*$ is not uniformity-preserving, then (X_1, X_2) is not uniform in $\mathcal{X}^2$. If W is a perfect channel, i.e., $I(W) = \log_2 |\mathcal{X}|$, we have*

$$I(W^-) + I(W^+) = I(X_1, X_2; Y_1, Y_2) \le H(X_1, X_2) \overset{(a)}{<} 2\log_2 |\mathcal{X}| = 2I(W), \tag{3.1}$$

  where (a) follows from the fact that (X_1, X_2) is not uniform in $\mathcal{X}^2$. (3.1) means that $*$ does not satisfy the conservation property of Definition 3.4. Therefore, every polarizing operation must be uniformity-preserving.

- *If $*$ is uniformity-preserving, then (X_1, X_2) is uniform in $\mathcal{X}^2$, i.e., X_1 and X_2 are independent and uniform in $\mathcal{X}$. Thus,*

$$I(W^-) + I(W^+) = I(X_1, X_2; Y_1, Y_2) = I(X_1; Y_1) + I(X_2; Y_2) = 2I(W).$$

Therefore, uniformity-preserving operations satisfy the conservation property.

We conclude that a binary operation $*$ satisfies the conservation property if and only if it is uniformity-preserving.

Definition 3.5. *Let $*$ be a polarizing operation on a set $\mathcal{X}$. We say that $\beta \geq 0$ is a* $*$-achievable exponent *if for every $\delta > 0$ and every channel W with input alphabet $\mathcal{X}$, W_n almost surely becomes $(\delta, 2^{-2^{\beta n}})$-easy, i.e.,*

$$\lim_{n\to\infty} \mathbb{P}\big[W_n \text{ is } (\delta, 2^{-2^{\beta n}})\text{-easy}\big] = 1.$$

We define the exponent *of $*$ as:*

$$E_* := \sup\{\beta \geq 0 : \ \beta \text{ is a } *\text{-achievable exponent}\}.$$

Note that E_* depends only on $*$ and it does not depend on any particular channel W. The definition of a $*$-achievable exponent ensures that it is achievable for every channel W with input alphabet $\mathcal{X}$.

Example 3.1. *If $\mathcal{X} = \mathbb{F}_2 = \{0, 1\}$ and $*$ is the addition modulo 2, then $E_* = \frac{1}{2}$ (see [19]).*

3.2 A Characterization of Polarizing Binary Operations

3.2.1 Necessary Condition

In this subsection, we show that if $*$ is polarizing, then $*$ is uniformity-preserving and $/^*$ (the right-inverse of $*$) is strongly ergodic. In order to prove this, we need the following two lemmas:

Lemma 3.1. *Let $*$ be an ergodic operation on a set $\mathcal{X}$. Let $\mathcal{H}$ be a stable partition of $\mathcal{X}$ such that $\mathcal{K}_\mathcal{H} \neq \mathcal{H}$, where $\mathcal{K}_\mathcal{H}$ is the first residue of $\mathcal{H}$ with respect to $*$. Define $\mathcal{A} = \mathcal{H} \cup \mathcal{K}_\mathcal{H}$. We have:*

1. *For every $A_1, A_2 \in \mathcal{A}$, we have:*
   - *$(A_1 \in \mathcal{K}_\mathcal{H}$ and $A_2 \in \mathcal{K}_\mathcal{H})$ if and only if $(A_1 * A_2 \in \mathcal{K}_\mathcal{H}{}^*$ and $A_2 \in \mathcal{K}_\mathcal{H})$.*
   - *$(A_1 \in \mathcal{K}_\mathcal{H}$ and $A_2 \in \mathcal{H})$ if and only if $(A_1 * A_2 \in \mathcal{K}_\mathcal{H}{}^*$ and $A_2 \in \mathcal{H})$.*
   - *$(A_1 \in \mathcal{H}$ and $A_2 \in \mathcal{K}_\mathcal{H})$ if and only if $(A_1 * A_2 \in \mathcal{H}^*$ and $A_2 \in \mathcal{K}_\mathcal{H})$.*
   - *$(A_1 \in \mathcal{H}$ and $A_2 \in \mathcal{H})$ if and only if $(A_1 * A_2 \in \mathcal{H}^*$ and $A_2 \in \mathcal{H})$.*
2. *For every $u_1, u_2 \in \mathcal{X}$ and every $A_1, A_2 \in \mathcal{A}$, we have*

$$(u_1 \in A_1 * A_2 \text{ and } u_2 \in A_2) \text{ if and only if } (u_1 /^* u_2 \in A_1 \text{ and } u_2 \in A_2).$$

Proof. 1) We have $\mathcal{A} = \mathcal{H} \cup \mathcal{K}_{\mathcal{H}}$. Therefore, for every $A_1, A_2 \in \mathcal{A}$, one of the following four conditions holds true:

(i) $A_1 \in \mathcal{K}_{\mathcal{H}}$ and $A_2 \in \mathcal{K}_{\mathcal{H}}$.

(ii) $A_1 \in \mathcal{K}_{\mathcal{H}}$ and $A_2 \in \mathcal{H}$.

(iii) $A_1 \in \mathcal{H}$ and $A_2 \in \mathcal{K}_{\mathcal{H}}$.

(iv) $A_1 \in \mathcal{H}$ and $A_2 \in \mathcal{H}$.

Now since $\mathcal{K}_{\mathcal{H}} \neq \mathcal{H}$ and $\mathcal{K}_{\mathcal{H}} \preceq \mathcal{H}$, we have $||\mathcal{K}_{\mathcal{H}}|| < ||\mathcal{H}||$. Therefore, for every $K \in \mathcal{K}_{\mathcal{H}}$ and every $H \in \mathcal{H}$, we have $|K| = ||\mathcal{K}_{\mathcal{H}}|| < ||\mathcal{H}|| = |H|$. This implies that $K \neq H$ for every $K \in \mathcal{K}_{\mathcal{H}}$ and every $H \in \mathcal{H}$, hence $\mathcal{K}_{\mathcal{H}} \cap \mathcal{H} = \emptyset$. Similarly, ${\mathcal{K}_{\mathcal{H}}}^* \cap \mathcal{H}^* = \emptyset$. We conclude that for every $A_1, A_2 \in \mathcal{A}$, the following four conditions are mutually exclusive:

(a) $A_1 * A_2 \in {\mathcal{K}_{\mathcal{H}}}^*$ and $A_2 \in \mathcal{K}_{\mathcal{H}}$.

(b) $A_1 * A_2 \in {\mathcal{K}_{\mathcal{H}}}^*$ and $A_2 \in \mathcal{H}$.

(c) $A_1 * A_2 \in \mathcal{H}^*$ and $A_2 \in \mathcal{K}_{\mathcal{H}}$.

(d) $A_1 * A_2 \in \mathcal{H}^*$ and $A_2 \in \mathcal{H}$.

We have:

- If $A_1 \in \mathcal{K}_{\mathcal{H}}$ and $A_2 \in \mathcal{K}_{\mathcal{H}}$, then $A_1 * A_2 \in {\mathcal{K}_{\mathcal{H}}}^*$. Therefore, (i) implies (a).
- If $A_1 \in \mathcal{K}_{\mathcal{H}}$ and $A_2 \in \mathcal{H}$, then $A_1 * A_2 \in {\mathcal{K}_{\mathcal{H}}}^*$ (see Theorem 2.1). Therefore, (ii) implies (b).
- If $A_1 \in \mathcal{H}$ and $A_2 \in \mathcal{K}_{\mathcal{H}}$, let $H \in \mathcal{H}$ be such that $A_2 \subset H$. (Note that there is no contradiction here between $A_2 \subset H \in \mathcal{H}$, $A_2 \in \mathcal{K}_{\mathcal{H}}$ and $\mathcal{H} \cap \mathcal{K}_{\mathcal{H}} = \emptyset$.) We have $A_1 * A_2 \subset A_1 * H$ and $|A_1 * A_2| \geq |A_1| = ||\mathcal{H}|| = ||\mathcal{H}^*|| = |A_1 * H|$. Therefore, $A_1 * A_2 = A_1 * H \in \mathcal{H}^*$. Hence (iii) implies (c).
- If $A_1 \in \mathcal{H}$ and $A_2 \in \mathcal{H}$, then $A_1 * A_2 \in \mathcal{H}^*$. Therefore, (iv) implies (d).

Now let $A_1, A_2 \in \mathcal{A}$ and suppose that (a) holds true (i.e., $A_1 * A_2 \in {\mathcal{K}_{\mathcal{H}}}^*$ and $A_2 \in \mathcal{K}_{\mathcal{H}}$). Since $A_1 \in \mathcal{A}$ then either $A_1 \in \mathcal{K}_{\mathcal{H}}$ or $A_1 \in \mathcal{H}$. But $A_2 \in \mathcal{K}_{\mathcal{H}}$, so either (i) or (iii) holds true. On the other hand, we have shown that (iii) implies (c), and (c) contradicts (a), so (iii) cannot be true. Therefore, (i) must be true. We conclude that (a) implies (i). Similarly, we can show that (b) implies (ii), (c) implies (iii), and (d) implies (iv).

2) Fix $A_1, A_2 \in \mathcal{A}$. We have:

- If $A_1 \in \mathcal{K}_{\mathcal{H}}$ and $A_2 \in \mathcal{K}_{\mathcal{H}}$, then $|A_1 * A_2| = ||{\mathcal{K}_{\mathcal{H}}}^*|| = ||\mathcal{K}_{\mathcal{H}}|| = |A_1|$.
- If $A_1 \in \mathcal{K}_{\mathcal{H}}$ and $A_2 \in \mathcal{H}$, then from 1) we have $A_1 * A_2 \in {\mathcal{K}_{\mathcal{H}}}^*$. Therefore, $|A_1 * A_2| = ||{\mathcal{K}_{\mathcal{H}}}^*|| = ||\mathcal{K}_{\mathcal{H}}|| = |A_1|$.
- If $A_1 \in \mathcal{H}$ and $A_2 \in \mathcal{K}_{\mathcal{H}}$ then from 1) we have $A_1 * A_2 \in \mathcal{H}^*$. Therefore, $|A_1 * A_2| = ||\mathcal{H}^*|| = ||\mathcal{H}|| = |A_1|$.

- If $A_1 \in \mathcal{H}$ and $A_2 \in \mathcal{H}$, then $|A_1 * A_2| = ||\mathcal{H}^*|| = ||\mathcal{H}|| = |A_1|$.

We conclude that in all cases, we have $|A_1 * A_2| = |A_1|$.

For every $u_1, u_2 \in \mathcal{X}$, we have:

- If $u_1 /^* u_2 \in A_1$ and $u_2 \in A_2$, then $u_1 = (u_1 /^* u_2) * u_2 \in A_1 * A_2$.

- If $u_1 \in A_1 * A_2$ and $u_2 \in A_2$, we have $A_1 * u_2 \subset A_1 * A_2$. On the other hand, we have $|A_1 * A_2| = |A_1| = |A_1 * u_2|$ (where the last equality holds true because $*$ is uniformity-preserving). We conclude that $A_1 * A_2 = A_1 * u_2$. Therefore, $(A_1 * A_2)/^* u_2 = A_1$ which implies that $u_1 /^* u_2 \in A_1$.

□

Definition 3.6. *A channel $W : \mathcal{X} \longrightarrow \mathcal{Y}$ is said to be equivalent to another channel $W' : \mathcal{X} \longrightarrow \mathcal{Z}$ if both channels are degraded from each other.*

Lemma 3.2. *Let $*$ be a uniformity-preserving operation on a set $\mathcal{X}$, and let $W : \mathcal{X} \longrightarrow \mathcal{Y}$. If $I(W^-) = I(W)$ then W^+ is equivalent to W.*

Proof. Since $I(W^+)+I(W^-) = 2I(W)$ and since $I(W^-) = I(W)$, we have $I(W^+) = I(W)$. Let $(U_1, U_2) \xrightarrow{f_*} (X_1, X_2) \xrightarrow{W} (Y_1, Y_2)$ (See Notation 3.2). We have:

$$\begin{aligned} I(W) = I(W^+) &= I(U_2; Y_1, Y_2, U_1) \\ &= I(U_2; Y_2) + I(U_2; Y_1, U_1 | Y_2) = I(W) + I(U_2; Y_1, U_1 | Y_2). \end{aligned}$$

This shows that $I(U_2; Y_1, U_1|Y_2) = 0$. This means that Y_2 is a sufficient statistic for the channel $U_2 \longrightarrow (Y_1, Y_2, U_1)$ (which is equivalent to W^+). We conclude that W^+ is equivalent to the channel $U_2 \longrightarrow Y_2$, which is equivalent to W. □

Proposition 3.1. *Let $*$ be a binary operation on a set $\mathcal{X}$. If $*$ is polarizing then $*$ is uniformity-preserving and $/^*$ is strongly ergodic.*

Proof. If $*$ is polarizing then $*$ must be uniformity-preserving (see Remark 3.1).

We first prove that $*$ is irreducible. Suppose to the contrary that $*$ is not irreducible. Proposition 2.1 shows that there exist two disjoint non-empty subsets A_1 and A_2 of $\mathcal{X}$ such that $A_1 \cup A_2 = \mathcal{X}$, $A_1 * \mathcal{X} = A_1$ and $A_2 * \mathcal{X} = A_2$. This means that for every $u_1, u_2 \in \mathcal{X}$ and every $y \in \{1, 2\}$, we have $u_1 \in A_y$ if and only if $u_1 * u_2 \in A_y$.

For each $\epsilon > 0$ define the channel $W_\epsilon : \mathcal{X} \longrightarrow \{1, 2, e\}$ as follows:

$$W_\epsilon(y|x) = \begin{cases} 1 - \epsilon & \text{if } y \in \{1, 2\} \text{ and } x \in A_y, \\ 0 & \text{if } y \in \{1, 2\} \text{ and } x \notin A_y, \\ \epsilon & \text{if } y = e. \end{cases}$$

$I(W_\epsilon) = (1 - \epsilon) h_2\big(\frac{|A_1|}{|\mathcal{X}|}\big)$, so there exists $\epsilon' > 0$ such that $I(W_{\epsilon'})$ is not the logarithm of any integer. For such ϵ', there exists $\delta > 0$ such that $W_{\epsilon'}$ is not δ-easy.

Let $(U_1, U_2) \xrightarrow{f_*} (X_1, X_2) \xrightarrow{W_{\epsilon'}} (Y_1, Y_2)$ (See Notation 3.2). Consider the channel $U_1 \longrightarrow (Y_1, Y_2)$ which is equivalent to $W^-_{\epsilon'}$. We have:

$$\begin{aligned} P_{Y_1,Y_2|U_1}(y_1, y_2|u_1) &= \frac{1}{|\mathcal{X}|} \sum_{u_2 \in \mathcal{X}} W_{\epsilon'}(y_1|u_1 * u_2) W_{\epsilon'}(y_2|u_2) \\ &\overset{(a)}{=} \frac{1}{|\mathcal{X}|} \sum_{u_2 \in \mathcal{X}} W_{\epsilon'}(y_1|u_1) W_{\epsilon'}(y_2|u_2) \\ &\overset{(b)}{=} \sum_{u_2 \in \mathcal{X}} W_{\epsilon'}(y_1|u_1) P_{Y_2|U_2}(y_2|u_2) P_{U_2}(u_2) \\ &= W_{\epsilon'}(y_1|u_1) P_{Y_2}(y_2), \end{aligned} \tag{3.2}$$

where (a) follows from the fact that if $y_1 = e$ then $W_{\epsilon'}(y_1|u_1 * u_2) = W_{\epsilon'}(y_1|u_1) = \epsilon'$ and if $y_1 \in \{1, 2\}$ then $u_1 \in A_{y_1}$ if and only if $u_1 * u_2 \in A_{y_1}$, which implies that $W_{\epsilon'}(y_1|u_1 * u_2) = W_{\epsilon'}(y_1|u_1)$. (b) follows from the fact that the channel $U_2 \longrightarrow Y_2$ is equivalent to $W_{\epsilon'}$ and the fact that U_2 is uniform in $\mathcal{X}$.

(3.2) implies that Y_1 is a sufficient statistic for the channel $U_1 \longrightarrow (Y_1, Y_2)$ (which is equivalent to $W^-_{\epsilon'}$). Moreover, since $P_{Y_1,Y_2|U_1}(y_1, y_2|u_1) = W_{\epsilon'}(y_1|u_1) P_{Y_2}(y_2)$, we conclude that the channel $W^-_{\epsilon'}$ is equivalent to $W_{\epsilon'}$. This implies that $I(W^-_{\epsilon'}) = I(W_{\epsilon'})$. Now Lemma 3.2 implies that $W^+_{\epsilon'}$ is equivalent to $W_{\epsilon'}$. Therefore, for every $l > 0$ and every $s \in \{-, +\}^l$, $W^s_{\epsilon'}$ is equivalent to $W_{\epsilon'}$ which is not δ-easy. This contradicts the fact that $*$ is polarizing. We conclude that $*$ must be irreducible.

Suppose that $*$ is not ergodic. Proposition 2.1 shows that there exists a partition $\{H_0, \ldots, H_{n-1}\}$ of $\mathcal{X}$ such that $H_i * \mathcal{X} = H_{i+1 \bmod n}$ for all $0 \leq i < n$ and $|H_0| = \ldots = |H_{n-1}|$. This means that for every $u_1, u_2 \in \mathcal{X}$ and every $y \in \{0, \ldots, n-1\}$, we have $u_1 * u_2 \in H_y$ if and only if $u_1 \in H_{y-1 \bmod n}$.

For each $0 \leq i < n$ and each $0 < \epsilon < 1$, define the channel $W_{i,\epsilon} : \mathcal{X} \longrightarrow \{0, \ldots, n-1, e\}$ as follows:

$$W_{i,\epsilon}(y|x) = \begin{cases} 1 - \epsilon & \text{if } y \in \{0, \ldots, n-1\} \text{ and } x \in H_{y+i \bmod n}, \\ 0 & \text{if } y \in \{0, \ldots, n-1\} \text{ and } x \notin H_{y+i \bmod n}, \\ \epsilon & \text{if } y = e. \end{cases}$$

$I(W_{i,\epsilon}) = (1 - \epsilon) \log_2 n$ so there exists $\epsilon' > 0$ such that $I(W_{i,\epsilon'})$ is not the logarithm of any integer. For such ϵ', there exists $\delta > 0$ such that $W_{i,\epsilon'}$ is not δ-easy for any $0 \leq i < n$.

Let $(U_1, U_2) \xrightarrow{f_*} (X_1, X_2) \xrightarrow{W_{i,\epsilon'}} (Y_1, Y_2)$. Consider the channel $U_1 \longrightarrow (Y_1, Y_2)$ which is equivalent to $W^-_{i,\epsilon'}$. We have:

$$\begin{aligned} P_{Y_1,Y_2|U_1}(y_1, y_2|u_1) &= \frac{1}{|\mathcal{X}|} \sum_{u_2 \in \mathcal{X}} W_{i,\epsilon'}(y_1|u_1 * u_2) W_{i,\epsilon'}(y_2|u_2) \\ &\overset{(a)}{=} \frac{1}{|\mathcal{X}|} \sum_{u_2 \in \mathcal{X}} W_{i-1 \bmod n,\epsilon'}(y_1|u_1) W_{i,\epsilon'}(y_2|u_2) \\ &\overset{(b)}{=} \sum_{u_2 \in \mathcal{X}} W_{i-1 \bmod n,\epsilon'}(y_1|u_1) P_{Y_2|U_2}(y_2|u_2) P_{U_2}(u_2) \\ &= W_{i-1 \bmod n,\epsilon'}(y_1|u_1) P_{Y_2}(y_2), \end{aligned} \tag{3.3}$$

where (a) follows from the fact that if $y_1 = e$ then

$$W_{i,\epsilon'}(y_1|u_1 * u_2) = W_{i-1 \bmod n,\epsilon'}(y_1|u_1) = \epsilon'$$

and if $y_1 \in \{0, \ldots, n-1\}$ then $u_1 * u_2 \in H_{y_1+i \bmod n}$ if and only if $u_1 \in H_{y_1+i-1 \bmod n}$ (which implies that $W_{i,\epsilon'}(y_1|u_1 * u_2) = W_{i-1 \bmod n,\epsilon'}(y_1|u_1)$). (b) follows from the fact that the channel $U_2 \longrightarrow Y_2$ is equivalent to $W_{i,\epsilon'}$ and the fact that U_2 is uniform in $\mathcal{X}$.

(3.3) implies that Y_1 is a sufficient statistic for the channel $U_1 \longrightarrow (Y_1, Y_2)$ (which is equivalent to $W^-_{i,\epsilon'}$). Moreover, since

$$P_{Y_1,Y_2|U_1}(y_1, y_2|u_1) = W_{i-1 \bmod n,\epsilon'}(y_1|u_1)P_{Y_2}(y_2),$$

we conclude that the channel $W^-_{i,\epsilon'}$ is equivalent to $W_{i-1 \bmod n,\epsilon'}$. This implies that $I(W^-_{i,\epsilon'}) = I(W_{i-1 \bmod n,\epsilon'}) = (1-\epsilon')\log_2 n = I(W_{i,\epsilon'})$. Now Lemma 3.2 implies that $W^+_{i,\epsilon'}$ is equivalent to $W_{i,\epsilon'}$. Therefore, for every $l > 0$ and every $s \in \{-,+\}^l$, $W^s_{i,\epsilon'}$ is equivalent to $W_{i-|s|^- \bmod n,\epsilon'}$ (where $|s|^-$ is the number of appearances of the $-$ sign in the sequence s) which is not δ-easy. This contradicts the fact that $*$ is polarizing. We conclude that $*$ must be ergodic.

Since $*$ is ergodic, $/^*$ is ergodic as well. Suppose that $/^*$ is not strongly ergodic. Theorem 2.2 implies the existence of a stable partition $\mathcal{H}$ of $(\mathcal{X}, /^*)$ such that $\mathcal{K}_\mathcal{H} \neq \mathcal{H}$ (where $\mathcal{K}_\mathcal{H}$ here denotes the first residue of $\mathcal{H}$ with respect to the right-inverse operation $/^*$). For each $i \geq 0$ and each $\epsilon > 0$ define the channel $W_{i,\epsilon} : \mathcal{X} \longrightarrow {\mathcal{K}_\mathcal{H}}^{i/^*} \cup {\mathcal{H}}^{i/^*}$ as follows:

$$W_{i,\epsilon}(y|x) = \begin{cases} 1-\epsilon & \text{if } x \in y \text{ and } y \in {\mathcal{K}_\mathcal{H}}^{i/^*}, \\ \epsilon & \text{if } x \in y \text{ and } y \in {\mathcal{H}}^{i/^*}, \\ 0 & \text{if } x \notin y. \end{cases}$$

We emphasize that y here is a subset of $\mathcal{X}$ and it is not an element of it. We have

$$I(W_{i,\epsilon}) = (1-\epsilon)\log_2 |{\mathcal{K}_\mathcal{H}}^{i/^*}| + \epsilon \log_2 |{\mathcal{H}}^{i/^*}| = (1-\epsilon)\log_2 |\mathcal{K}_\mathcal{H}| + \epsilon \log_2 |\mathcal{H}|.$$

Now since $\mathcal{K}_\mathcal{H} \neq \mathcal{H}$ and $\mathcal{K}_\mathcal{H} \preceq \mathcal{H}$, we have $|\mathcal{H}| \neq |\mathcal{K}_\mathcal{H}|$. Therefore, there exists $\epsilon' > 0$ such that $I(W_{i,\epsilon'})$ is not the logarithm of any integer. For such $\epsilon' > 0$, there exists $\delta > 0$ such that $I(W_{i,\epsilon'})$ is not δ-easy for any $i \geq 0$.

Let $(U_1, U_2) \xrightarrow{f_*} (X_1, X_2) \xrightarrow{W_{i,\epsilon'}} (Y_1, Y_2)$. Consider the channel $U_1 \longrightarrow (Y_1, Y_2)$,

which is equivalent to $W^{-}_{i,\epsilon'}$. We have:

$$
\begin{aligned}
&P_{Y_1,Y_2|U_1}(y_1,y_2|u_1)\\
&= \frac{1}{|\mathcal{X}|}\sum_{u_2\in\mathcal{X}} W_{i,\epsilon'}(y_1|u_1 * u_2)W_{i,\epsilon'}(y_2|u_2)\\
&= \frac{1}{|\mathcal{X}|}\sum_{u_2\in\mathcal{X}} \left[\mathbb{1}_{u_1*u_2\in y_1}\cdot\left((1-\epsilon')\mathbb{1}_{y_1\in\mathcal{K}_{\mathcal{H}}{}^{i/*}}+\epsilon'\mathbb{1}_{y_1\in\mathcal{H}^{i/*}}\right)\right]\\
&\qquad\qquad\times\left[\mathbb{1}_{u_2\in y_2}\cdot\left((1-\epsilon')\mathbb{1}_{y_2\in\mathcal{K}_{\mathcal{H}}{}^{i/*}}+\epsilon'\mathbb{1}_{y_2\in\mathcal{H}^{i/*}}\right)\right]\\
&= \frac{1}{|\mathcal{X}|}\sum_{u_2\in\mathcal{X}} \mathbb{1}_{u_1*u_2\in y_1,\, u_2\in y_2}\cdot\left((1-\epsilon')\mathbb{1}_{y_1\in\mathcal{K}_{\mathcal{H}}{}^{i/*}}+\epsilon'\mathbb{1}_{y_1\in\mathcal{H}^{i/*}}\right)\\
&\qquad\qquad\times\left((1-\epsilon')\mathbb{1}_{y_2\in\mathcal{K}_{\mathcal{H}}{}^{i/*}}+\epsilon'\mathbb{1}_{y_2\in\mathcal{H}^{i/*}}\right)\\
&\overset{(a)}{=} \frac{1}{|\mathcal{X}|}\sum_{u_2\in\mathcal{X}} \mathbb{1}_{u_1\in y_1/^*y_2,\, u_2\in y_2}\cdot\left((1-\epsilon')\mathbb{1}_{y_1\in\mathcal{K}_{\mathcal{H}}{}^{i/*}}+\epsilon'\mathbb{1}_{y_1\in\mathcal{H}^{i/*}}\right)\\
&\qquad\qquad\times\left((1-\epsilon')\mathbb{1}_{y_2\in\mathcal{K}_{\mathcal{H}}{}^{i/*}}+\epsilon'\mathbb{1}_{y_2\in\mathcal{H}^{i/*}}\right)\\
&= \frac{1}{|\mathcal{X}|}\sum_{u_2\in\mathcal{X}} \mathbb{1}_{u_1\in y_1/^*y_2,\, u_2\in y_2}\cdot\Big((1-\epsilon')^2\mathbb{1}_{y_1\in\mathcal{K}_{\mathcal{H}}{}^{i/*},\, y_2\in\mathcal{K}_{\mathcal{H}}{}^{i/*}}\\
&\qquad\qquad+(1-\epsilon')\epsilon'\mathbb{1}_{y_1\in\mathcal{K}_{\mathcal{H}}{}^{i/*},\, y_2\in\mathcal{H}^{i/*}}+\epsilon'(1-\epsilon')\mathbb{1}_{y_1\in\mathcal{H}^{i/*},\, y_2\in\mathcal{K}_{\mathcal{H}}{}^{i/*}}\\
&\qquad\qquad+\epsilon'^2\mathbb{1}_{y_1\in\mathcal{H}^{i/*},\, y_2\in\mathcal{H}^{i/*}}\Big)\\
&\overset{(b)}{=} \frac{1}{|\mathcal{X}|}\sum_{u_2\in\mathcal{X}} \mathbb{1}_{u_1\in y_1/^*y_2,\, u_2\in y_2}\cdot\Big((1-\epsilon')^2\mathbb{1}_{y_1/^*y_2\in\mathcal{K}_{\mathcal{H}}{}^{(i+1)/*},\, y_2\in\mathcal{K}_{\mathcal{H}}{}^{i/*}}\\
&\qquad\qquad+(1-\epsilon')\epsilon'\mathbb{1}_{y_1/^*y_2\in\mathcal{K}_{\mathcal{H}}{}^{(i+1)/*},\, y_2\in\mathcal{H}^{i/*}}+\epsilon'(1-\epsilon')\mathbb{1}_{y_1/^*y_2\in\mathcal{H}^{(i+1)/*},\, y_2\in\mathcal{K}_{\mathcal{H}}{}^{i/*}}\\
&\qquad\qquad+\epsilon'^2\mathbb{1}_{y_1/^*y_2\in\mathcal{H}^{(i+1)/*},\, y_2\in\mathcal{H}^{i/*}}\Big)\\
&= \frac{1}{|\mathcal{X}|}\sum_{u_2\in\mathcal{X}} \left[\mathbb{1}_{u_1\in y_1/^*y_2}\cdot\left((1-\epsilon')\mathbb{1}_{y_1/^*y_2\in\mathcal{K}_{\mathcal{H}}{}^{(i+1)/*}}+\epsilon'\mathbb{1}_{y_1/^*y_2\in\mathcal{H}^{(i+1)/*}}\right)\right]\\
&\qquad\qquad\times\left[\mathbb{1}_{u_2\in y_2}\cdot\left((1-\epsilon')\mathbb{1}_{y_2\in\mathcal{K}_{\mathcal{H}}{}^{i/*}}+\epsilon'\mathbb{1}_{y_2\in\mathcal{H}^{i/*}}\right)\right]\\
&= \frac{1}{|\mathcal{X}|}\sum_{u_2\in\mathcal{X}} W_{i+1,\epsilon'}(y_1/^*y_2|u_1)W_{i,\epsilon'}(y_2|u_2)\\
&\overset{(c)}{=} \sum_{u_2\in\mathcal{X}} W_{i+1,\epsilon'}(y_1/^*y_2|u_1)P_{Y_2|U_2}(y_2|u_2)P_{U_2}(u_2)\\
&= W_{i+1,\epsilon'}(y_1/^*y_2|u_1)P_{Y_2}(y_2),
\end{aligned}
\tag{3.4}
$$

where (a) follows from applying the second point of Lemma 3.1 on the ergodic operation $/^*$ and the stable partition $\mathcal{H}^{i/*}$. (b) follows from applying the first point of Lemma 3.1 on the ergodic operation $/^*$ and the stable partition $\mathcal{H}^{i/*}$. (c) follows from the fact that $W_{i,\epsilon'}$ is equivalent to the channel $U_2 \longrightarrow Y_2$ and from the fact that U_2 is uniform in $\mathcal{X}$.

(3.4) implies that $Y_1/^*Y_2$ is a sufficient statistic for the channel $U_1 \longrightarrow (Y_1, Y_2)$

(which is equivalent to $W^-_{i,\epsilon'}$). Moreover, since

$$P_{Y_1,Y_2|U_1}(y_1,y_2|u_1) = W_{i+1,\epsilon'}(y_1/^*y_2|u_1)P_{Y_2}(y_2),$$

we conclude that the channel $W^-_{i,\epsilon'}$ is equivalent to $W_{i+1,\epsilon'}$. This implies that $I(W^-_{i,\epsilon'}) = I(W_{i+1,\epsilon'}) = (1-\epsilon')\log_2|\mathcal{K}_\mathcal{H}| + \epsilon'\log_2|\mathcal{H}| = I(W_{i,\epsilon'})$. Now Lemma 3.2 implies that $W^+_{i,\epsilon'}$ is equivalent to $W_{i,\epsilon'}$. Therefore, for every $l > 0$ and every $s \in \{-,+\}^l$, $W^s_{i,\epsilon'}$ is equivalent to $W_{i+|s|^-,\epsilon'}$ (where $|s|^-$ is the number of appearances of the $-$ sign in the sequence s) which is not δ-easy. This again contradicts the fact that $*$ is polarizing. We conclude that $/^*$ must be strongly ergodic. □

3.2.2 Sufficient Condition

In this subsection, we prove a converse for Proposition 3.1. We will show that for any uniformity-preserving operation $*$, the strong ergodicity of $/^*$ implies that $*$ is polarizing. We will prove this in three steps.

Step 1: Polarized Channels are Projection Channels onto Stable Partitions

Notation 3.3. *For every sequence $\mathbf{x} = (x_i)_{0\le i<N}$ of N elements of $\mathcal{X}$, and for every $0 \le j \le k < N$, we define the subsequence $\mathbf{x}^k_j$ as the sequence $(x'_i)_{0\le i\le k-j}$, where $x'_i = x_{i+j}$ for every $0 \le i \le k-j$.*

Notation 3.4. *For every $k \ge 0$ and every sequence $\mathbf{x} = (x_i)_{0\le i<2^k}$ of $|\mathbf{x}| = 2^k$ elements of $\mathcal{X}$, we define $g_*(\mathbf{x}) \in \mathcal{X}$ recursively on k as follows:*

- *If $k = 0$ (i.e., $\mathbf{x} = (x_0)$), $g_*(\mathbf{x}) = x_0$.*
- *If $k > 0$, $g_*(\mathbf{x}) = g_*(\mathbf{x}_0^{|\mathbf{x}|/2-1}) * g_*(\mathbf{x}_{|\mathbf{x}|/2}^{|\mathbf{x}|-1}) = g_*(\mathbf{x}_0^{2^{k-1}-1}) * g_*(\mathbf{x}_{2^{k-1}}^{2^k-1})$.*

For example, we have:

- $g_*(\mathbf{x}_0^1) = x_0 * x_1$.
- $g_*(\mathbf{x}_0^3) = (x_0 * x_1) * (x_2 * x_3)$.
- $g_*(\mathbf{x}_0^7) = \big((x_0 * x_1) * (x_2 * x_3)\big) * \big((x_4 * x_5) * (x_6 * x_7)\big)$.

Definition 3.7. *Let A be a subset of $\mathcal{X}$. We define the probability distribution $\mathbb{I}_A$ on $\mathcal{X}$ as $\mathbb{I}_A(x) = \frac{1}{|A|}$ if $x \in A$ and $\mathbb{I}_A(x) = 0$ otherwise.*

Definition 3.8. *Let $\mathcal{Y}$ be an arbitrary set, $\mathcal{H}$ be a balanced partition of $\mathcal{X}$ and (X,Y) be a random pair in $\mathcal{X}\times\mathcal{Y}$. For every $\gamma > 0$, we define:*

$$\mathcal{Y}_{\mathcal{H},\gamma}(X,Y) = \Big\{y \in \mathcal{Y} :\ \exists H_y \in \mathcal{H},\ \|P_{X|Y=y} - \mathbb{I}_{H_y}\|_\infty < \gamma\Big\},$$

and

$$\mathcal{P}_{\mathcal{H},\gamma}(X,Y) = P_Y\big(\mathcal{Y}_{\mathcal{H},\gamma}(X,Y)\big).$$

Note that if $\mathcal{P}_{\mathcal{H},\gamma}(X,Y) \approx 1$ for a small γ then Y is "almost equivalent" to the projection of X onto $\mathcal{H}$. This will be proved rigorously in step 2. The next proposition will be used later to show that a relation $\mathcal{P}_{\mathcal{H},\gamma}(X,Y) \approx 1$ is satisfied between the input and output of a polarized channel, where $\mathcal{H}$ is a stable partition. This is why we say that polarized channels are projection channels onto stable partitions.

Proposition 3.2. *Let $*$ be a strongly ergodic operation on a set $\mathcal{X}$. Define $k = 2^{2^{|\mathcal{X}|}} + \mathrm{scon}(*)$ and let $\mathcal{Y}$ be an arbitrary set. For every $\gamma > 0$, there exists $\epsilon(\gamma) > 0$ depending only on $\mathcal{X}$ such that if $(X_i, Y_i)_{0 \le i < 2^k}$ is a sequence of 2^k random pairs satisfying:*

1. *$(X_i, Y_i)_{0 \le i < 2^k}$ are independent and identically distributed in $\mathcal{X} \times \mathcal{Y}$,*
2. *X_i is uniform in $\mathcal{X}$ for all $0 \le i < 2^k$,*
3. *$H\big(g_*(X_0^{2^k-1}) | Y_0^{2^k-1}\big) < H(X_0|Y_0) + \epsilon(\gamma)$,*

then there exists a stable partition $\mathcal{H}$ of $(\mathcal{X}, *)$ such that $\mathcal{P}_{\mathcal{H},\gamma}(X_0, Y_0) > 1 - \gamma$.

Proof. See Appendix 3.4.1. □

Step 2: Structure of Projection Channels

Lemma 3.3. *Let $\mathcal{X}$ be an arbitrary set and let $*$ be an ergodic operation on $\mathcal{X}$. For every $\delta > 0$, there exists $\gamma := \gamma(\delta) > 0$ such that for any stable partition $\mathcal{H}$ of $(\mathcal{X}, *)$, if (X, Y) is a pair of random variables in $\mathcal{X} \times \mathcal{Y}$ satisfying*

1. *X is uniform in $\mathcal{X}$,*
2. *$\mathcal{P}_{\mathcal{H},\gamma}(X;Y) > 1 - \gamma$,*

then $\left| I\big(\mathrm{Proj}_{\mathcal{H}'}(X); Y\big) - \log_2 \frac{|\mathcal{H}| \cdot \|\mathcal{H} \wedge \mathcal{H}'\|}{\|\mathcal{H}'\|} \right| < \delta$ for every stable partition $\mathcal{H}'$ of $(\mathcal{X}, *)$.

Proof. Let $\mathcal{H}'$ be a stable partition of $\mathcal{X}$. Note that the entropy function is continuous and the space of probability distributions on $\mathcal{H}'$ is compact. Therefore, the entropy function is uniformly continuous, which means that for every $\delta > 0$ there exists $\gamma'_{\mathcal{H}'}(\delta) > 0$ such that if p_1 and p_2 are two probability distributions on $\mathcal{H}'$ satisfying $\|p_1 - p_2\|_\infty < \gamma'_{\mathcal{H}'}(\delta)$ then $|H(p_1) - H(p_2)| < \frac{\delta}{2}$. Let $\delta > 0$ and define $\gamma_{\mathcal{H}'}(\delta) = \min\left\{ \frac{\delta}{2\log_2(|\mathcal{H}'|+1)}, \frac{1}{\|\mathcal{H}'\|} \gamma'_{\mathcal{H}'}(\delta) \right\}$. Now define $\gamma(\delta) = \min\{\gamma_{\mathcal{H}'}(\delta) : \mathcal{H}' \text{ is a stable partition}\}$ which depends only on $(\mathcal{X}, *)$ and δ. Clearly, $\|\mathcal{H}'\|\gamma(\delta) \le \gamma'_{\mathcal{H}'}(\delta)$ for every stable partition $\mathcal{H}'$ of $\mathcal{X}$.

Let $\mathcal{H}$ be a stable partition of $\mathcal{X}$ and suppose that $\mathcal{P}_{\mathcal{H},\gamma(\delta)}(X;Y) > 1 - \gamma(\delta)$, where X is uniform in $\mathcal{X}$. Fix $y \in \mathcal{Y}_{\mathcal{H},\gamma(\delta)}(X;Y)$. By the definition of $\mathcal{Y}_{\mathcal{H},\gamma(\delta)}(X;Y)$, there exists $H_y \in \mathcal{H}$ such that $|P_{X|Y}(x|y) - \mathbb{I}_{H_y}(x)| < \gamma(\delta)$ for every $x \in \mathcal{X}$.

Let $\mathcal{H}'$ be a stable partition of $\mathcal{X}$. Corollary 2.1 shows that $\mathcal{H} \wedge \mathcal{H}'$ is also a stable partition of $\mathcal{X}$. From the definition of $\mathcal{H} \wedge \mathcal{H}'$, for every $H' \in \mathcal{H}'$ we have either $H_y \cap H' = \emptyset$ or $H_y \cap H' \in \mathcal{H} \wedge \mathcal{H}'$. Therefore, we have either $|H_y \cap H'| = 0$ or

$|H_y \cap H'| = \|\mathcal{H} \wedge \mathcal{H}'\|$. Let $\mathcal{H}'_y = \{H' \in \mathcal{H}' : H_y \cap H' \neq \emptyset\}$, so $|H_y \cap H'| = \|\mathcal{H} \wedge \mathcal{H}'\|$ for all $H' \in \mathcal{H}'_y$. Now since $H_y = \bigcup_{H' \in \mathcal{H}'} (H_y \cap H')$, we have $\|\mathcal{H}\| = |H_y| = \sum_{H' \in \mathcal{H}'} |H_y \cap H'| = |\mathcal{H}'_y| \cdot \|\mathcal{H} \wedge \mathcal{H}'\|$. Therefore,

$$\frac{\|\mathcal{H}\|}{\|\mathcal{H} \wedge \mathcal{H}'\|} = |\mathcal{H}'_y| \leq |\mathcal{H}'|. \tag{3.5}$$

We will now show that for every $y \in \mathcal{Y}_{\mathcal{H},\gamma(\delta)}$, we have $\|P_{\mathrm{Proj}_{\mathcal{H}'}(X)|Y=y} - \mathbb{I}_{\mathcal{H}'_y}\|_\infty < \gamma'_{\mathcal{H}'}(\delta)$, where $\mathbb{I}_{\mathcal{H}'_y}$ is the probability distribution on $\mathcal{H}'$ defined as $\mathbb{I}_{\mathcal{H}'_y}(H') = \frac{1}{|\mathcal{H}'_y|}$ if $H' \in \mathcal{H}'_y$ and $\mathbb{I}_{\mathcal{H}'_y}(H') = 0$ otherwise. This will be useful to show that

$$\left| H(\mathrm{Proj}_{\mathcal{H}'}(X)|Y = y) - \log_2 \frac{\|\mathcal{H}\|}{\|\mathcal{H} \wedge \mathcal{H}'\|} \right| < \frac{\delta}{2} \text{ for all } y \in \mathcal{Y}_{\mathcal{H},\gamma(\delta)}.$$

Let $y \in \mathcal{Y}_{\mathcal{H},\gamma(\delta)}$ and $H' \in \mathcal{H}'$. We have $P_{\mathrm{Proj}_{\mathcal{H}'}(X)|Y}(H'|y) = \sum_{x \in H'} P_{X|Y}(x|y)$. But since $|P_{X|Y}(x|y) - \frac{1}{|H_y|}| < \gamma(\delta)$ for every $x \in H_y$, and since $P_{X|Y}(x|y) < \gamma(\delta)$ if $x \in \mathcal{X} \backslash H_y$, we conclude that $\left|P_{\mathrm{Proj}_{\mathcal{H}'}(X)|Y}(H'|y) - \frac{|H' \cap H_y|}{|H_y|}\right| < |H'|\gamma(\delta) = \|\mathcal{H}'\|\gamma(\delta) \leq \gamma'_{\mathcal{H}'}(\delta)$. We conclude:

- If $H' \in \mathcal{H}'_y$, we have $|H' \cap H_y| = \|\mathcal{H} \wedge \mathcal{H}'\|$ which means that $\frac{|H' \cap H_y|}{|H_y|} = \frac{\|\mathcal{H} \wedge \mathcal{H}'\|}{\|\mathcal{H}\|} \overset{(a)}{=} \frac{1}{|\mathcal{H}'_y|}$, where (a) follows from (3.5). Thus $|P_{\mathrm{Proj}_{\mathcal{H}'}(X)|Y}(H'|y) - \frac{1}{|\mathcal{H}'_y|}| < \gamma'_{\mathcal{H}'}(\delta)$.

- If $H' \in \mathcal{H}' \setminus \mathcal{H}'_y$, $\frac{|H' \cap H_y|}{|H_y|} = 0$ and so $P_{\mathrm{Proj}_{\mathcal{H}'}(X)|Y}(H'|y) < \gamma'_{\mathcal{H}'}(\delta)$.

Therefore, $\|P_{\mathrm{Proj}_{\mathcal{H}'}(X)|Y=y} - \mathbb{I}_{\mathcal{H}'_y}\|_\infty < \gamma'_{\mathcal{H}'}(\delta)$. This means that $\left|H(\mathrm{Proj}_{\mathcal{H}'}(X)|Y = y) - H(\mathbb{I}_{\mathcal{H}'_y})\right| < \frac{\delta}{2}$. But $H(\mathbb{I}_{\mathcal{H}'_y}) = \log_2 |\mathcal{H}'_y| \overset{(a)}{=} \log_2 \frac{\|\mathcal{H}\|}{\|\mathcal{H} \wedge \mathcal{H}'\|}$, where (a) follows from (3.5). Therefore,

$$\forall y \in \mathcal{Y}_{\mathcal{H},\gamma(\delta)}, \left| H(\mathrm{Proj}_{\mathcal{H}'}(X)|Y = y) - \log_2 \frac{\|\mathcal{H}\|}{\|\mathcal{H} \wedge \mathcal{H}'\|} \right| < \frac{\delta}{2}. \tag{3.6}$$

On the other hand, for every $y \in \mathcal{Y}^c_{\mathcal{H},\gamma(\delta)}$, $P_{\mathrm{Proj}_{\mathcal{H}'}(X)|Y=y}$ is a probability distribution on $\mathcal{H}'$ which implies that $0 \leq H(\mathrm{Proj}_{\mathcal{H}'}(X)|Y = y) \leq \log_2 |\mathcal{H}'|$. Moreover, we have $0 \leq \log_2 \frac{\|\mathcal{H}\|}{\|\mathcal{H} \wedge \mathcal{H}'\|} \leq \log_2 |\mathcal{H}'|$ from (3.5). Therefore,

$$\forall y \in \mathcal{Y}^c_{\mathcal{H},\gamma(\delta)}, \left| H(\mathrm{Proj}_{\mathcal{H}'}(X)|Y = y) - \log_2 \frac{\|\mathcal{H}\|}{\|\mathcal{H} \wedge \mathcal{H}'\|} \right| \leq \log_2 |\mathcal{H}'|. \tag{3.7}$$

We conclude that:

$$\begin{aligned}
&\left|H(\mathrm{Proj}_{\mathcal{H}'}(X)|Y) - \log_2 \frac{\|\mathcal{H}\|}{\|\mathcal{H}\wedge\mathcal{H}'\|}\right| \\
&\qquad \le \sum_{y\in\mathcal{Y}} \left|H(\mathrm{Proj}_{\mathcal{H}'}(X)|Y=y) - \log_2 \frac{\|\mathcal{H}\|}{\|\mathcal{H}\wedge\mathcal{H}'\|}\right| \cdot P_Y(y) \\
&\qquad \overset{(a)}{\le} \sum_{y\in\mathcal{Y}_{\mathcal{H},\gamma(\delta)}} \frac{\delta}{2}\cdot P_Y(y) + \sum_{y\in\mathcal{Y}^c_{\mathcal{H},\gamma(\delta)}} (\log_2|\mathcal{H}'|)\cdot P_Y(y) \\
&\qquad = \frac{\delta}{2}\cdot P_Y(\mathcal{Y}_{\mathcal{H},\gamma(\delta)}) + (\log_2|\mathcal{H}'|)P_Y(\mathcal{Y}^c_{\mathcal{H},\gamma(\delta)}) \overset{(b)}{<} \frac{\delta}{2} + (\log_2|\mathcal{H}'|)\gamma(\delta) \\
&\qquad \le \frac{\delta}{2} + (\log_2|\mathcal{H}'|)\cdot\frac{\delta}{2\log_2(|\mathcal{H}'|+1)} < \delta,
\end{aligned}$$

where (a) follows from (3.6) and (3.7). (b) follows from the second condition of the lemma.

Now since $\mathrm{Proj}_{\mathcal{H}'}(X)$ is uniform in $\mathcal{H}'$, we have $H(\mathrm{Proj}_{\mathcal{H}'}(X)) = \log_2|\mathcal{H}'|$. We conclude that if $\mathcal{P}_{\mathcal{H},\gamma(\delta)}(X,Y) > 1-\gamma(\delta)$ then for every stable partition $\mathcal{H}'$ of $(\mathcal{X},*)$, we have

$$\left|I\big(\mathrm{Proj}_{\mathcal{H}'}(X);Y\big) - \log_2\frac{|\mathcal{H}'|\cdot\|\mathcal{H}\wedge\mathcal{H}'\|}{\|\mathcal{H}\|}\right| < \delta,$$

which implies that $\left|I\big(\mathrm{Proj}_{\mathcal{H}'}(X);Y\big) - \log_2\frac{|\mathcal{H}|\cdot\|\mathcal{H}\wedge\mathcal{H}'\|}{\|\mathcal{H}'\|}\right| < \delta$ since $|\mathcal{H}|\cdot\|\mathcal{H}\| = |\mathcal{H}'|\cdot\|\mathcal{H}'\| = |\mathcal{X}|$. □

Step 3: Projection Channels are Easy

Definition 3.9. *Let $\mathcal{H}$ be a balanced partition of $\mathcal{X}$ and let $W:\mathcal{X}\longrightarrow\mathcal{Y}$. We define the channel $W[\mathcal{H}]:\mathcal{H}\longrightarrow\mathcal{Y}$ by:*

$$W[\mathcal{H}](y|H) = \frac{1}{\|\mathcal{H}\|}\sum_{\substack{x\in\mathcal{X}:\\ \mathrm{Proj}_{\mathcal{H}}(x)=H}} W(y|x) = \frac{1}{|H|}\sum_{x\in H} W(y|x).$$

Remark 3.2. *If X is a random variable uniformly distributed in $\mathcal{X}$ and Y is the output of the channel W when X is the input, then it is easy to see that $I(W[\mathcal{H}]) = I(\mathrm{Proj}_{\mathcal{H}}(X);Y)$.*

Theorem 3.1. *Let $\mathcal{X}$ be an arbitrary set and let $*$ be a uniformity-preserving operation on $\mathcal{X}$ such that $/^*$ is strongly ergodic. Let $W:\mathcal{X}\longrightarrow\mathcal{Y}$ be an arbitrary channel. Then for every $\delta>0$, we have:*

$$\lim_{n\to\infty}\frac{1}{2^n}\left|\left\{s\in\{-,+\}^n :\ \exists\mathcal{H}_s \text{ a stable partition of } (\mathcal{X},/^*),\right.\right.$$
$$\left.\left.\left|I(W^s[\mathcal{H}']) - \log_2\frac{|\mathcal{H}_s|\cdot\|\mathcal{H}_s\wedge\mathcal{H}'\|}{\|\mathcal{H}'\|}\right| < \delta \text{ for all stable partitions } \mathcal{H}' \text{ of } (\mathcal{X},/^*)\right\}\right| = 1.$$

Proof. Let $(W_n)_n$ be as in Definition 3.3. Since $*$ is uniformity-preserving, it satisfies the conservation property of Definition 3.4 (see Remark 3.1). Therefore, we have:

$$\mathbb{E}\big[I(W_{n+1})|W_n\big] = \frac{1}{2}I(W_n^-) + \frac{1}{2}I(W_n^+) = I(W_n).$$

This implies that the process $(I(W_n))_n$ is a martingale, and so it converges almost surely. Therefore, the process $\big(I(W_{n+k}) - I(W_n)\big)_n$ converges almost surely to zero, where $k = 2^{2^{|\mathcal{X}|}} + \mathrm{scon}(/^*)$. In particular, $\big(I(W_{n+k}) - I(W_n)\big)_n$ converges in probability to zero, hence for every $\delta > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}\Big[|I(W_{n+k}) - I(W_n)| \geq \epsilon\big(\gamma(\delta)\big)\Big] = 0,$$

where $\epsilon(.)$ is given by Proposition 3.2 and $\gamma(.)$ is given by Lemma 3.3. We have:

$$\mathbb{P}\Big[|I(W_{n+k}) - I(W_n)| \geq \epsilon\big(\gamma(\delta)\big)\Big] = \frac{1}{2^{n+k}}|A_{n,k}|,$$

where $A_{n,k} = \Big\{(s,s') \in \{-,+\}^n \times \{-,+\}^k : \ |I(W^{(s,s')}) - I(W^s)| \geq \epsilon\big(\gamma(\delta)\big)\Big\}$. Define:

$$B_{n,k} = \Big\{s \in \{-,+\}^n : \ |I(W^{(s,[k]^-)}) - I(W^s)| \geq \epsilon\big(\gamma(\delta)\big)\Big\},$$

where $[k]^- \in \{-,+\}^k$ is the sequence consisting of k minus signs. Clearly, $B_{n,k} \times \{[k]^-\} \subset A_{n,k}$ and so $|B_{n,k}| \leq |A_{n,k}|$. Now since

$$\lim_{n\to\infty} \frac{1}{2^{n+k}}|A_{n,k}| = \lim_{n\to\infty} \mathbb{P}\Big[|I(W_{n+k}) - I(W_n)| \geq \epsilon\big(\gamma(\delta)\big)\Big] = 0,$$

we must have $\lim_{n\to\infty} \frac{1}{2^{n+k}}|B_{n,k}| = 0$. Therefore, $\lim_{n\to\infty} \frac{1}{2^n}|B_{n,k}| = 2^k \times 0 = 0$ and so $\lim_{n\to\infty} \frac{1}{2^n}|B_{n,k}^c| = 1$.

Now suppose that $s \in B_{n,k}^c$, i.e., $|I(W^{(s,[k]^-)}) - I(W^s)| < \epsilon\big(\gamma(\delta)\big)$. Let $U_0, \ldots, U_{2^k-1}$ be 2^k independent random variables uniformly distributed in $\mathcal{X}$. For every $0 \leq j \leq k$, define the sequence $U_{j,0}, \ldots, U_{j,2^k-1}$ recursively as follows:

- $U_{0,i} = U_i$ for every $0 \leq i < 2^k$.
- For every $0 \leq j < k$ and every $0 \leq i < 2^k$, define $U_{j+1,i}$ as follows:

$$U_{j+1,i} = \begin{cases} U_{j,i} * U_{j,i+2^{k-j-1}} & \text{if } 0 \leq i \bmod 2^{k-j} < 2^{k-j-1}, \\ U_{j,i} & \text{if } 2^{k-j-1} \leq i \bmod 2^{k-j} < 2^{k-j}. \end{cases}$$

Since $*$ is uniformity-preserving, it is easy to see that for every $0 \leq i \leq k$, the 2^k random variables $U_{j,0}, \ldots, U_{j,2^k-1}$ are independent and uniform in $\mathcal{X}$. In particular, if we define $X_i = U_{k,i}$ for $0 \leq i < 2^k$, then $X_0, \ldots, X_{2^k-1}$ are 2^k independent random variables uniformly distributed in $\mathcal{X}$. Suppose that $X_0, \ldots, X_{2^k-1}$ are sent through 2^k independent copies of the channel W^s and let $Y_0, \ldots, Y_{2^k-1}$ be the output of each copy of the channel respectively. Clearly, $(X_i, Y_i)_{0\leq i<2^k}$ are independent and uniformly distributed in $\mathcal{X} \times \mathcal{Y}$. Moreover, $I(W^s) = I(X_i; Y_i)$ for every $0 \leq i < 2^k$. In particular, $I(W^s) = I(X_0; Y_0) = H(X_0) - H(X_0|Y_0) = \log_2 |\mathcal{X}| - H(X_0|Y_0)$. We will show by backward induction on $0 \leq j \leq k$ that for every $0 \leq q < 2^j$ we have:

- $W^{(s,[k-j]^-)}$ is equivalent to the channel $U_{j,q\cdot 2^{k-j}} \longrightarrow Y_{q\cdot 2^{k-j}}^{(q+1)\cdot 2^{k-j}-1}$.
- $U_{j,q\cdot 2^{k-j}} = g_{/^*}\big(X_{q\cdot 2^{k-j}}^{(q+1)\cdot 2^{k-j}-1}\big)$.

The claim is trivial for $j = k$. Now let $0 \leq j < k$ and suppose that the claim is true for $j+1$. Let $0 \leq q < 2^j$. From the induction hypothesis we have:

- $W^{(s,[k-j-1]^-)}$ is equivalent to the channel $U_{j+1,q\cdot 2^{k-j}} \longrightarrow Y_{q\cdot 2^{k-j}}^{(2q+1)\cdot 2^{k-j-1}-1}$.
- $U_{j+1,q\cdot 2^{k-j}} = g_{/^*}\big(X_{q\cdot 2^{k-j}}^{(2q+1)\cdot 2^{k-j-1}-1}\big)$.
- $W^{(s,[k-j-1]^-)}$ is equivalent to the channel $U_{j+1,(2q+1)\cdot 2^{k-j-1}} \longrightarrow Y_{(2q+1)\cdot 2^{k-j-1}}^{(q+1)\cdot 2^{k-j}-1}$.
- $U_{j+1,(2q+1)\cdot 2^{k-j-1}} = g_{/^*}\big(X_{(2q+1)\cdot 2^{k-j-1}}^{(q+1)\cdot 2^{k-j}-1}\big)$.

Now since

$$U_{j+1,q\cdot 2^{k-j}} = U_{j,q\cdot 2^{k-j}} * U_{j,(2q+1)\cdot 2^{k-j-1}}$$

and

$$U_{j+1,(2q+1)\cdot 2^{k-j-1}} = U_{j,(2q+1)\cdot 2^{k-j-1}},$$

it follows that $W^{(s,[k-j]^-)} = (W^{(s,[k-j-1]^-)})^-$ is equivalent to the channel $U_{j,q\cdot 2^{k-j}} \longrightarrow Y_{q\cdot 2^{k-j}}^{(q+1)\cdot 2^{k-j}-1}$ (see Remark 3.1). Moreover, we have

$$\begin{aligned} U_{j,q\cdot 2^{k-j}} &= U_{j+1,q\cdot 2^{k-j}} /^* U_{j,(2q+1)\cdot 2^{k-j-1}} = U_{j+1,q\cdot 2^{k-j}} /^* U_{j+1,(2q+1)\cdot 2^{k-j-1}} \\ &= g_{/^*}\big(X_{q\cdot 2^{k-j}}^{(2q+1)\cdot 2^{k-j-1}-1}\big) /^* g_{/^*}\big(X_{(2q+1)\cdot 2^{k-j-1}}^{(q+1)\cdot 2^{k-j}-1}\big) = g_{/^*}\big(X_{q\cdot 2^{k-j}}^{(q+1)\cdot 2^{k-j}-1}\big). \end{aligned}$$

This terminates the induction argument and so the claim is true for all $0 \leq j \leq k$. In particular, for $j = 0$ and $q = 0$, we have $U_0 = U_{0,0} = g_{/^*}\big(X_0^{2^k-1}\big)$ and $W^{(s,[k]^-)}$ is equivalent to the channel $U_0 \longrightarrow Y_0^{2^k-1}$. Thus,

$$I(W^{(s,[k]^-)}) = I(U_0; Y_0^{2^k-1}) = H(U_0) - H(U_0|Y_0^{2^k-1}) = \log_2|\mathcal{X}| - H(U_0|Y_0^{2^k-1}).$$

Hence

$$\begin{aligned} I(W^{(s,[k]^-)}) - I(W^s) &= \log_2|\mathcal{X}| - H(U_0|Y_0^{2^k-1}) - \log_2|\mathcal{X}| + H(X_0|Y_0) \\ &\overset{(a)}{=} H(X_0|Y_0) - H\big(g_{/^*}(X_0^{2^k-1})|Y_0^{2^k-1}\big), \end{aligned}$$

where (a) follows from the fact that $U_0 = g_{/^*}\big(X_0^{2^k-1}\big)$. We conclude that

$$\big|H\big(g_{/^*}(X_0^{2^k-1})|Y_0^{2^k-1}\big) - H(X_0|Y_0)\big| = |I(W^{(s,[k]^-)}) - I(W^s)| < \epsilon\big(\gamma(\delta)\big).$$

Proposition 3.2, applied to $/^*$, implies the existence of a stable partition $\mathcal{H}_s$ of $(\mathcal{X}, /^*)$ such that $\mathcal{P}_{\mathcal{H}_s,\gamma(\delta)}(X_0, Y_0) > 1 - \gamma(\delta)$. Now Lemma 3.3, applied to $/^*$, implies that for every stable partition $\mathcal{H}'$ of $(\mathcal{X}, /^*)$, we have $\left|I(W^s[\mathcal{H}']) - \log_2 \frac{|\mathcal{H}_s|\cdot\|\mathcal{H}_s\wedge\mathcal{H}'\|}{\|\mathcal{H}'\|}\right| =$

$\left|I\big(\mathrm{Proj}_{\mathcal{H}'}(X_0);Y_0\big) - \log_2 \frac{|\mathcal{H}_s|\cdot\|\mathcal{H}_s\wedge\mathcal{H}'\|}{\|\mathcal{H}'\|}\right| < \delta$. But this is true for every $s \in B^c_{n,k}$. Therefore, $B^c_{n,k} \subset D_n$, where D_n is defined as:

$$D_n = \Bigg\{ s \in \{-,+\}^n : \ \exists \mathcal{H}_s \text{ a stable partition of } (\mathcal{X}, /^*),$$
$$\left|I(W^s[\mathcal{H}']) - \log_2 \frac{|\mathcal{H}_s|\cdot\|\mathcal{H}_s\wedge\mathcal{H}'\|}{\|\mathcal{H}'\|}\right| < \delta \text{ for all stable partitions } \mathcal{H}' \text{ of } (\mathcal{X}, /^*) \Bigg\}.$$

Now since $\lim_{n\to\infty} \frac{1}{2^n}|B^c_{n,k}| = 1$ and $B^c_{n,k} \subset D_n$, we must have $\lim_{n\to\infty} \frac{1}{2^n}|D_n| = 1$. □

Corollary 3.1. *Let $\mathcal{X}$ be an arbitrary set and let $*$ be a uniformity-preserving operation on $\mathcal{X}$ such that $/^*$ is strongly ergodic, and let $W : \mathcal{X} \longrightarrow \mathcal{Y}$ be an arbitrary channel. Then for every $\delta > 0$, we have:*

$$\lim_{n\to\infty} \frac{1}{2^n}\Bigg|\Big\{ s \in \{-,+\}^n : \exists \mathcal{H}_s \text{ a stable partition of } (\mathcal{X}, /^*),$$
$$\big|I(W^s) - \log_2|\mathcal{H}_s|\big| < \delta, \big|I(W^s[\mathcal{H}_s]) - \log_2|\mathcal{H}_s|\big| < \delta \Big\}\Bigg| = 1.$$

Proof. We apply Theorem 3.1 and we consider the two particular cases where $\mathcal{H}' = \big\{\{x\} : \ x \in \mathcal{X}\big\}$ and $\mathcal{H}' = \mathcal{H}_s$. □

Remark 3.3. *Corollary 3.1 can be interpreted as follows: In a polarized channel W^s, we have $I(W^s) \approx I(W^s[\mathcal{H}_s]) \approx \log_2|\mathcal{H}_s|$ for some stable partition $\mathcal{H}_s$ of $(\mathcal{X}, /^*)$. Let X_s and Y_s be the input and output of the channel W^s respectively. $I(W^s[\mathcal{H}_s]) \approx \log_2|\mathcal{H}_s|$ means that Y_s "almost" determines $\mathrm{Proj}_{\mathcal{H}_s}(X_s)$. On the other hand, $I(W^s) \approx I(W^s[\mathcal{H}_s])$ means that there is "almost" no other information about X_s which can be determined from Y_s. Therefore, W^s is "almost" equivalent to the channel $X_s \longrightarrow \mathrm{Proj}_{\mathcal{H}_s}(X_s)$.*

Lemma 3.4. *Let $W : \mathcal{X} \longrightarrow \mathcal{Y}$ be an arbitrary channel. If there exists a balanced partition $\mathcal{H}$ of $\mathcal{X}$ such that $\big|I(W) - \log_2|\mathcal{H}|\big| < \delta$ and $\big|I(W[\mathcal{H}]) - \log_2|\mathcal{H}|\big| < \delta$, then W is δ-easy. Moreover, if we also have $P_e(W[\mathcal{H}]) < \epsilon$, then W is (δ,ϵ)-easy.*

Proof. Let $L = |\mathcal{H}|$ and let $H_1, \ldots, H_L$ be the L members of $\mathcal{H}$. Let $\mathcal{S} = \big\{C \subset \mathcal{X} : \ |C| = L\big\}$ and $\mathcal{S}_{\mathcal{H}} = \big\{\{x_1, \ldots, x_L\} : \ x_1 \in H_1, \ldots, x_L \in H_L\big\} \subset \mathcal{S}$. For each $1 \le i \le L$, let X_i be a random variable uniformly distributed in H_i. Define $\mathcal{B} = \{X_1, \ldots, X_L\}$, which is a random set taking values in $\mathcal{S}_{\mathcal{H}}$. Note that we can see $\mathcal{B}$ as a random variable in $\mathcal{S}$ since $\mathcal{S}_{\mathcal{H}} \subset \mathcal{S}$. For every $x \in \mathcal{X}$, let H_i be the unique element of $\mathcal{H}$ such that $x \in H_i$. We have:

$$\frac{1}{L}\sum_{C\in\mathcal{H}} P_{\mathcal{B}}(C)\mathbb{1}_{x\in C} = \frac{1}{|\mathcal{H}|}\mathbb{P}[x \in \mathcal{B}] \overset{(a)}{=} \frac{1}{|\mathcal{H}|}\mathbb{P}[X_i = x] = \frac{1}{|\mathcal{H}|}\cdot\frac{1}{|H_i|} = \frac{1}{|\mathcal{H}|}\cdot\frac{1}{\|\mathcal{H}\|} = \frac{1}{|\mathcal{X}|}, \tag{3.8}$$

where (a) follows from the fact that $x \in \mathcal{B}$ if and only if $X_i = x$. Now for each $C \in \mathcal{S}_{\mathcal{H}}$, define the bijection $f_C : \{1, \ldots, L\} \to C$ as follows: For each $1 \le i \le L$, $f_C(i)$ is the unique element in $C \cap H_i$ (so $\mathrm{Proj}_{\mathcal{H}}(f_C(i)) = H_i$). Let U be a random variable

chosen uniformly in $\{1, \ldots, L\}$ and independently from $\mathcal{B}$, and let $X = f_{\mathcal{B}}(U)$ (so $\text{Proj}_{\mathcal{H}}(X) = H_U$). From (3.8) we get that X is uniform in $\mathcal{X}$.

Let Y be the output of the channel W when X is the input. From Definition 3.1, we have $I(W_{\mathcal{B}}) = I(U; Y, \mathcal{B})$. On the other hand, $I(W[\mathcal{H}]) = I(\text{Proj}_{\mathcal{H}}(X); Y) = I(H_U; Y)$. Therefore, $I(W_{\mathcal{B}}) = I(U; Y, \mathcal{B}) \geq I(U; Y) \overset{(a)}{=} I(H_U; Y) = I(W[\mathcal{H}]) \overset{(b)}{>} \log_2 L - \delta$, where (a) follows from the fact that the mapping $u \to H_u$ is a bijection from $\{1, \ldots, L\}$ to $\mathcal{H}$ and (b) follows from the fact that $\left|I(W[\mathcal{H}]) - \log_2 |\mathcal{H}|\right| < \delta$. We conclude that W is δ-easy since $I(W_{\mathcal{B}}) > \log_2 L - \delta$ and $|I(W) - \log_2 L| < \delta$.

Now suppose that we also have $P_e(W[\mathcal{H}]) < \epsilon$. For every $C \in \mathcal{S}_{\mathcal{H}}$, define the mapping $g_C : \{1, \ldots, L\} \to \mathcal{H}$ as $g_C(i) = \text{Proj}_{\mathcal{H}}(f_C(i))$ for every $1 \leq i \leq L$. It is easy to see that g_C is a bijection for every $C \in \mathcal{S}_{\mathcal{H}}$. Furthermore, we have $\text{Proj}_{\mathcal{H}}(X) = g_{\mathcal{B}}(U)$.

Consider the following decoder for the channel $W_{\mathcal{B}}$:

- Compute an estimate $\hat{H}$ of $\text{Proj}_{\mathcal{H}}(X)$ using the ML decoder of the channel $W[\mathcal{H}]$.
- Compute $\hat{U} = g_{\mathcal{B}}^{-1}(\hat{H})$.

The probability of error of this decoder is:

$$\mathbb{P}[\hat{U} \neq U] = \mathbb{P}[\hat{H} \neq g_{\mathcal{B}}(U)] = \mathbb{P}[\hat{H} \neq \text{Proj}_{\mathcal{H}}(X)] = P_e(W[\mathcal{H}]) < \epsilon.$$

Now since the ML decoder of $W_{\mathcal{B}}$ minimizes the probability of error, we conclude that $P_e(W_{\mathcal{B}}) < \epsilon$. Therefore, W is a (δ, ϵ)-easy channel. □

Proposition 3.3. *If $*$ is a uniformity-preserving operation on a set $\mathcal{X}$ and $/^*$ is strongly ergodic, then $*$ is polarizing.*

Proof. We have the following:

- We know from Remark 3.1 that since $*$ is uniformity-preserving, it satisfies the conservation property of Definition 3.4.
- The polarization property of Definition 3.4 follows immediately from Corollary 3.1 and Lemma 3.4.

Therefore, $*$ is polarizing. □

Theorem 3.2. *If $*$ is a binary operation on a set $\mathcal{X}$, then $*$ is polarizing if and only if $*$ is uniformity-preserving and $/^*$ is strongly ergodic.*

Proof. The theorem follows from Propositions 3.1 and 3.3. □

3.3 Polar Code Construction

Let $*$ be a polarizing binary operation of exponent[3] $E_* > 0$ on a finite set $\mathcal{X}$. Fix a channel W with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. Choose $0 < \delta < 1$ and

[3]As we will see in Chapter 5, not every polarizing binary operation has a strictly positive exponent. In this section, we assume that $*$ is a polarizing binary operation that satisfies $E_* > 0$.

$0 < \beta < \beta' < E_*$, and let $n_0 \geq 0$ be such that for every $n \geq n_0$, we have

$$2^n 2^{-2^{\beta' n}} < 2^{-2^{\beta n}} \text{ and } \frac{1}{2^n}|E_n| > 1 - \frac{\delta}{2\log_2 |\mathcal{X}|},$$

where

$$E_n = \left\{ s \in \{-,+\}^n : W^s \text{ is } (\tfrac{\delta}{2}, 2^{-2^{\beta' n}})\text{-}easy \right\}.$$

Such an integer exists because $*$ is polarizing and $\beta' < E_*$ (see Definition 3.5). For every $s \in E_n$, W^s is $(\frac{\delta}{2}, 2^{-2^{\beta' n}})$-easy, hence there exist an integer $L^s \leq |\mathcal{X}|$ and a random code $\mathcal{B}^s$ of block length 1 and rate $\log_2 L^s$ (i.e., $\mathcal{B}^s \in \mathcal{S}^s := \{C \subset \mathcal{X} : |C| = L^s\}$), which satisfy the following:

- $|I(W^s) - \log_2 L^s| < \frac{\delta}{2}$.
- For every $x \in \mathcal{X}$, we have

$$\sum_{C \in \mathcal{S}^s} \frac{1}{L^s} P_{\mathcal{B}^s}(C) \mathbb{1}_{x \in C} = \frac{1}{|\mathcal{X}|}. \tag{3.9}$$

- If for each $C \in \mathcal{S}^s$ we fix a bijection $f_C^s : \{1, \ldots, L^s\} \to C$, then $I(W^s{}_{\mathcal{B}^s}) > \log_2 L^s - \frac{\delta}{2}$ and $P_e(W^s{}_{\mathcal{B}^s}) < 2^{-2^{\beta' n}}$, where $W^s{}_{\mathcal{B}^s} : \{1, \ldots, L^s\} \longrightarrow \mathcal{Y}^s \times \mathcal{S}^s$ is the channel defined as:

$$W^s{}_{\mathcal{B}^s}(y, C|a) = W^s(y|f_C^s(a)) \cdot P_{\mathcal{B}^s}(C).$$

  Note that $\mathcal{Y}^s$ denotes the output alphabet of W^s. In the rest of this section, we assume that the bijections $(f_C^s)_{s \in E_n, C \in \mathcal{S}^s}$ are fixed and known to the transmitter and the receiver.

A polar code is constructed as follows:

- If $s \notin E_n$, let U^s be a frozen symbol in $\mathcal{X}$, i.e., we suppose that the receiver knows U^s.
- If $s \in E_n$, let C^s be a frozen code of blocklength 1 and rate $\log_2 L^s$ (i.e., the code C^s is chosen from $\mathcal{S}^s$ and it is known to the receiver). Let $\tilde{U}^s$ be a random variable that is uniformly distributed in $\{1, \ldots, L^s\}$ and let $U^s = f_{C^s}^s(\tilde{U}^s)$.
- After computing U^s for every $s \in \{-,+\}^n$, we apply n polarization steps on the sequence $(U^s)_{s \in \{-,+\}^n}$ to obtain another sequence of 2^n symbols $(U_s)_{s \in \{-,+\}^n}$, which will be transmitted through 2^n independent copies of the channel W (see Section 3.3.1).

Since we have a freedom in the choice of the frozen symbols $(U^s)_{s \notin E_n}$ and the frozen codes $(C^s)_{s \in E_n}$, we can assume that these symbols and codes are randomly generated as follows:

- If $s \notin E_n$, we assume that U^s is chosen uniformly from $\mathcal{X}$.

- If $s \in E_n$, we assume that C^s is a random code taking values in $\mathcal{S}^s$ according to the distribution of $\mathcal{B}^s$. Equation (3.9) implies that $U^s = f^s_{C^s}(\tilde{U}^s)$ is uniformly distributed in $\mathcal{X}$.

Furthermore, we assume that the random variables $(U^s)_{s \notin E_n}$, $(\tilde{U}^s)_{s \in E_n}$ and $(C^s)_{s \in E_n}$ are independent.

3.3.1 Encoding

We associate the set $S_n := \{-,+\}^n$ with the strict total order $<$ that we define as $(s_1, \ldots, s_n) < (s'_1, \ldots, s'_n)$ if and only if $s_i = -, s'_i = +$ for some $i \in \{1, \ldots, n\}$ and $s_h = s'_h$ for all $i < h \leq n$.

For every $u = (u^s)_{s \in S_n} \in \mathcal{X}^{S_n}$, every $0 \leq n' \leq n$ and every $(s', s'') \in S_{n'} \times S_{n-n'}$, define $\mathcal{E}^{s''}_{s'}(u) \in \mathcal{X}$ recursively on $0 \leq n' \leq n$ as follows:

- $\mathcal{E}^s_{\emptyset}(u) = u^s$ if $n' = 0$ and $s \in S_n$.
- $\mathcal{E}^{s''}_{(s',-)}(u) = \mathcal{E}^{(s'',-)}_{s'}(u) * \mathcal{E}^{(s'',+)}_{s'}(u)$ if $n' > 0$, $s' \in S_{n'-1}$ and $s'' \in S_{n-n'}$.
- $\mathcal{E}^{s''}_{(s',+)}(u) = \mathcal{E}^{(s'',+)}_{s'}(u)$ if $n' > 0$, $s' \in S_{n'-1}$ and $s'' \in S_{n-n'}$.

For every $s \in S_n$, we write $\mathcal{E}^s_{\emptyset}(u)$ as $\mathcal{E}^s(u)$ and $\mathcal{E}^{\emptyset}_s(u)$ as $\mathcal{E}_s(u)$.

Let $\{W_s\}_{s \in S_n}$ be a set of 2^n independent copies of the channel W. W_s should not be confused with W^s: W_s is a copy of the channel W whereas W^s is a synthetic channel obtained from W as before.

Let $(U^s)_{s \in S_n} = (f^s_{C^s}(\tilde{U}^s))_{s \in S_n}$ be the sequence of 2^n independent random variables that were defined above. For every $0 \leq n' \leq n$, $s' \in S_{n'}$ and $s'' \in S_{n-n'}$, define $U^{s''}_{s'} = \mathcal{E}^{s''}_{s'}\big((U^s)_{s \in S_n}\big)$. We have:

- $U^s_{\emptyset} = U^s$ if $n' = 0$ and $s \in \{-,+\}^n$.
- $U^{s''}_{(s',-)} = U^{(s'',+)}_{s'} * U^{(s'',-)}_{s'}$ if $n' > 0$, $s' \in \{-,+\}^{n'-1}$ and $s'' \in \{-,+\}^{n-n'}$.
- $U^{s''}_{(s',+)} = U^{(s'',+)}_{s'}$ if $n' > 0$, $s' \in \{-,+\}^{n'-1}$ and $s'' \in \{-,+\}^{n-n'}$.

For every $s \in S_n$, let $U_s = U^{\emptyset}_s$. Since $*$ is polarizing, it is uniformity-preserving. This implies that $(U_s)_{s \in S_n}$ are independent and uniformly distributed in $\mathcal{X}$.

It is easy to see that the complexity of the encoding algorithm is $O(N \log N)$, where $N = 2^n$ is the blocklength of the polar code.

For every $s \in S_n$, we send U_s through the channel W_s. Let Y_s be the output of the channel W_s, and let $Y = \{Y_s\}_{s \in S_n}$. We can prove by backward induction on n' that for every $s'' \in S_{n-n'}$, the channel $U^{s''}_{s'} \longrightarrow \big(\{Y_s\}_{s \text{ has } s' \text{ as a prefix}}, \{U^r_{s'}\}_{r < s''}\big)$ is equivalent to the channel $W^{s''}$ for every $0 \leq n' \leq n$, $s' \in S_{n'}$ and $s'' \in S_{n-n'}$. In particular, the channel $U^s \longrightarrow \big(Y, \{U^r\}_{r<s}\big)$ is equivalent to the channel W^s for every $s \in S_n$. This implies that the channel $\tilde{U}^s \longrightarrow \big(Y, \{U^r\}_{r<s}, C^s\big)$ is equivalent to $W^s{}_{\mathcal{B}^s}$ for every $s \in E_n$.

Figure 3.1 is an illustration of a polar code construction for $n = 2$ (i.e., the blocklength is $N = 2^2 = 4$).

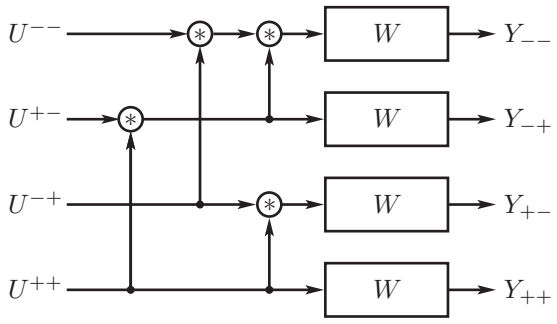

Figure 3.1 – Two polarization steps.

3.3.2 Decoding

If $s \notin E_n$, there is nothing to decode because the receiver knows U^s. Suppose now that $s \in E_n$. If we know $\{U^r\}_{r<s}$ then we can estimate $\tilde{U}^s$ from $\big(Y, \{U^r\}_{r<s}, C^s\big)$ using the maximum likelihood decoder of $W^s{}_{\mathcal{B}^s}$. After that, we can obtain an estimate of U^s by applying $f^s_{C^s}$ on the estimate of $\tilde{U}^s$.

This motivates us to consider the following successive cancellation decoder:

- $\hat{U}^s = U^s$ if $s \notin E_n$.
- $\hat{U}^s = f^s_{C^s}(\mathcal{D}^s(Y, \{\hat{U}^r\}_{r<s}, C^s))$ if $s \in E_n$, where $\mathcal{D}^s$ is the ML decoder of $W^s{}_{\mathcal{B}^s}$.

The symbols $(U^s)_{s\in S_n}$ are successively decoded according to the total order $<$ of S_n. By using essentially the same method that Arıkan used for binary-input channels [2], the successive cancellation decoder can be implemented with a complexity of $O(N \log N)$.

3.3.3 Performance of Polar Codes

We have

$$\begin{aligned}
\big\{\hat{U}^s = U^s,\ \forall s \in S_n\big\} &\Leftrightarrow \big\{\hat{U}^s = U^s,\ \forall s \in E_n\big\} \\
&\Leftrightarrow \Big\{f^s_{C^s}(\mathcal{D}^s(Y, \{\hat{U}^r\}_{r<s}, C^s)) = U^s,\ \forall s \in E_n\Big\} \\
&\Leftrightarrow \Big\{f^s_{C^s}(\mathcal{D}^s(Y, \{U^r\}_{r<s}, C^s)) = U^s,\ \forall s \in E_n\Big\} \\
&\Leftrightarrow \Big\{f^s_{C^s}(\mathcal{D}^s(Y, \{U^r\}_{r<s}, C^s)) = f^s_{C^s}(\tilde{U}^s),\ \forall s \in E_n\Big\} \\
&\Leftrightarrow \Big\{\mathcal{D}^s(Y, \{U^r\}_{r<s}, C^s) = \tilde{U}^s,\ \forall s \in E_n\Big\}.
\end{aligned}$$

Therefore, the probability of error of the above successive cancellation decoder is upper bounded by

$$\begin{aligned}
\sum_{s\in E_n} \mathbb{P}\big(\mathcal{D}^s(Y, \{U^r\}_{r<s}, C^s) \neq \tilde{U}^s\big) = \sum_{s\in E_n} P_e(W^s{}_{\mathcal{B}^s}) &\overset{(a)}{\leq} |E_n| 2^{-2^{\beta' n}} \\
&\leq 2^n 2^{-2^{\beta' n}} < 2^{-2^{\beta n}},
\end{aligned}$$

where (a) follows from the fact that $W^s{}_{\mathcal{B}^s}$ is $(\frac{\delta}{2}, 2^{-2^{\beta' n}})$-easy.

This upper bound was calculated on average over the random choice of the frozen symbols $(U^s)_{s\notin E_n}$ and codes $(C^s)_{s\in E_n}$. Therefore, there exists at least one choice of the frozen symbols and codes for which the upper bound of the probability of error still holds.

We should note here that unlike the case of binary-input symmetric memoryless channels where the frozen symbols can be chosen arbitrarily [2], the choice of the frozen symbols $(U^s)_{s\notin E_n}$ and codes $(C^s)_{s\in E_n}$ in our construction of polar codes cannot be arbitrary. The code designer should make sure that his choice of the frozen symbols and codes does indeed yield the desirable probability of error[4].

The last thing to discuss is the rate of polar codes. The rate at which we are communicating is $R = \frac{1}{2^n}\sum_{s\in E_n} \log_2 L^s$. On the other hand, we have $\big|I(W^s) - \log_2 L^s\big| < \frac{\delta}{2}$ for all $s \in E_n$. We conclude that:

$$\begin{aligned} I(W) &\overset{(a)}{=} \frac{1}{2^n}\sum_{s\in\{-,+\}^n} I(W^s) = \frac{1}{2^n}\sum_{s\in E_n} I(W^s) + \frac{1}{2^n}\sum_{s\in E_n^c} I(W^s) \\ &< \frac{1}{2^n}\sum_{s\in E_n}\Big(\log_2 L^s + \frac{\delta}{2}\Big) + \frac{1}{2^n}|E_n^c|\log_2|\mathcal{X}| \\ &< R + \frac{1}{2^n}|E_n|\frac{\delta}{2} + \frac{\delta}{2\log_2|\mathcal{X}|}\log_2|\mathcal{X}| \le R + \frac{\delta}{2} + \frac{\delta}{2} = R + \delta, \end{aligned}$$

where (a) follows from the conservation property of polarizing binary operations.

To this end we have shown the following proposition, which is the main result of this section:

Proposition 3.4. *If $*$ is a polarizing operation of exponent $E_* > 0$ on the set $\mathcal{X}$, then for every channel W with input alphabet $\mathcal{X}$, every $\beta < E_*$ and every $\delta > 0$, there exists $n_0 = n_0(W, \beta, \delta, *) > 0$ such that for every $n \ge n_0$, there exists a polar code of blocklength $N = 2^n$ and of rate at least $I(W) - \delta$ such that the probability of error of the successive cancellation decoder is at most 2^{-N^β}.*

3.4 Appendix

3.4.1 Proof of Proposition 3.2

Let $(X_i, Y_i)_{0\le i<2^k}$ be a sequence of 2^k random pairs that satisfy conditions 1) and 2) of Proposition 3.2.

Notation 3.5. *For every sequence $\mathbf{x} = (x_i)_{1\le i<2^k}$ of $2^k - 1$ elements of $\mathcal{X}$, define the mapping $\pi_{\mathbf{x}} : \mathcal{X} \to \mathcal{X}$ as $\pi_{\mathbf{x}}(x_0) = g_*\big((x_0, \mathbf{x})\big)$ for all $x_0 \in \mathcal{X}$, where $(x_0, \mathbf{x})$ is the sequence of 2^k elements obtained by concatenating x_0 and $\mathbf{x}$. Note that $\pi_{\mathbf{x}}$ is a bijection since $*$ is uniformity-preserving. Define:*

[4] In practice, the code designer can generate the frozen symbols $(U^s)_{s\notin E_n}$ and codes $(C^s)_{s\in E_n}$ randomly, and then runs a numerical simulation to assess the performance of the coding scheme. The code designer repeats this experiment until he finds a suitable choice for the frozen symbols $(U^s)_{s\notin E_n}$ and codes $(C^s)_{s\in E_n}$. With high probability, the code designer is expected to find good frozen symbols and codes after a few trials.

- $p_y(x) := P_{X_0|Y_0}(x|y)$ *for every* $x \in \mathcal{X}$ *and every* $y \in \mathcal{Y}$. *Note that* $p_y(x) = P_{X_i|Y_i}(x|y)$ *for every* $0 \leq i < 2^k$ *since* (X_i, Y_i) *and* (X_0, Y_0) *are identically distributed.*

- $p_{y_0,\mathbf{x}}(x) := p_{y_0}\big(\pi_{\mathbf{x}}^{-1}(x)\big)$ *for every* $x \in \mathcal{X}$, *every* $y_0 \in \mathcal{Y}$ *and every sequence* $\mathbf{x} = (x_i)_{1\leq i<2^k} \in \mathcal{X}^{2^k-1}$.

- *For every* $\mathbf{x} = (x_i)_{1\leq i<2^k} \in \mathcal{X}^{2^k-1}$, *and every* $y_1^{2^k-1} = (y_i)_{1\leq i<2^k} \in \mathcal{Y}^{2^k-1}$, *define*

$$p_{y_1^{2^k-1}}(\mathbf{x}) := \prod_{i=1}^{2^k-1} p_{y_i}(x_i) = P_{X_1^{2^k-1}|Y_1^{2^k-1}}(\mathbf{x}|y_1^{2^k-1}).$$

Fix $\gamma > 0$ and let $\gamma' = \min\left\{\dfrac{\gamma}{2^{|\mathcal{X}|}+1}, \dfrac{1}{(2^{|\mathcal{X}|}+2)|\mathcal{X}|}\right\}$.

Notation 3.6. *Define:*

$$\mathcal{C} = \Big\{y_0^{2^k-1} \in \mathcal{Y}^{2^k} : \forall \mathbf{x} \in \mathcal{X}^{2^k-1}, \forall \mathbf{x}' \in \mathcal{X}^{2^k-1},$$
$$\big(p_{y_1^{2^k-1}}(\mathbf{x}) \geq \gamma'^{2^k-1} \text{ and } p_{y_1^{2^k-1}}(\mathbf{x}') \geq \gamma'^{2^k-1}\big) \Rightarrow \|p_{y_0,\mathbf{x}} - p_{y_0,\mathbf{x}'}\|_\infty < \gamma'\Big\}.$$

Lemma 3.5. *There exists* $\epsilon(\gamma) > 0$ *such that if* $H\big(g_*(X_0^{2^k-1})\big|Y_0^{2^k-1}\big) < H(X_0|Y_0) + \epsilon(\gamma)$, *then*

$$P_{Y_0^{2^k-1}}(\mathcal{C}) > 1 - \gamma'^{2^k}.$$

Proof. For every $x \in \mathcal{X}$ and every $y_0^{2^k-1} \in \mathcal{Y}^{2^k}$, we have:

$$\begin{aligned}
P_{g_*(X_0^{2^k-1})|Y_0^{2^k-1}}(x|y_0^{2^k-1}) &= \sum_{\substack{x_0,\ldots,x_{2^k-1}\in\mathcal{X}:\\ g_*(x_0^{2^k-1})=x}} \left(\prod_{i=0}^{2^k-1} p_{y_i}(x_i)\right) \\
&= \sum_{\substack{\mathbf{x}\in\mathcal{X}^{2^k-1},\\ \mathbf{x}=(x_i)_{1\leq i<2^k}}} \sum_{\substack{x_0\in\mathcal{X}:\\ g_*((x_0,\mathbf{x}))=x}} \left(\prod_{i=1}^{2^k-1} p_{y_i}(x_i)\right) p_{y_0}(x_0) \\
&= \sum_{\mathbf{x}\in\mathcal{X}^{2^k-1}} p_{y_1^{2^k-1}}(\mathbf{x}) \sum_{\substack{x_0\in\mathcal{X}:\\ \pi_{\mathbf{x}}(x_0)=x}} p_{y_0}(x_0) \\
&= \sum_{\mathbf{x}\in\mathcal{X}^{2^k-1}} p_{y_1^{2^k-1}}(\mathbf{x}) p_{y_0}\big(\pi_{\mathbf{x}}^{-1}(x)\big) \\
&= \sum_{\mathbf{x}\in\mathcal{X}^{2^k-1}} p_{y_1^{2^k-1}}(\mathbf{x}) p_{y_0,\mathbf{x}}(x).
\end{aligned}$$

Therefore, for every $y_0^{2^k-1} \in \mathcal{Y}^{2^k}$ we have:

$$P_{g_*(X_0^{2^k-1})|Y_0^{2^k-1}}(x|y_0^{2^k-1}) = \sum_{\mathbf{x}\in\mathcal{X}^{2^k-1}} p_{y_1^{2^k-1}}(\mathbf{x}) p_{y_0,\mathbf{x}}(x). \tag{3.10}$$

Due to the concavity of the entropy function, it follows from (3.10) that for every sequence $y_0^{2^k-1} \in \mathcal{Y}^{2^k}$ we have:

$$\begin{aligned} H\big(g_*(X_0^{2^k-1})\big|Y_0^{2^k-1} = y_0^{2^k-1}\big) &\geq \sum_{\mathbf{x}\in\mathcal{X}^{2^k-1}} p_{y_1^{2^k-1}}(\mathbf{x}) H(p_{y_0,\mathbf{x}}) \\ &\overset{(a)}{=} \sum_{\mathbf{x}\in\mathcal{X}^{2^k-1}} p_{y_1^{2^k-1}}(\mathbf{x}) H(p_{y_0}) = H(p_{y_0}) \\ &= H(X_0|Y_0 = y_0), \end{aligned} \tag{3.11}$$

where (a) follows from the fact that the distribution $p_{y_0,\mathbf{x}}$ is a permuted version of the distribution p_{y_0}, which implies that $p_{y_0,\mathbf{x}}$ and p_{y_0} have the same entropy. Now if $y_0^{2^k-1} \in \mathcal{C}^c$, there exist $\mathbf{x} \in \mathcal{X}^{2^k-1}$ and $\mathbf{x}' \in \mathcal{X}^{2^k-1}$ such that $p_{y_1^{2^k-1}}(\mathbf{x}) \geq \gamma'^{2^k-1}$, $p_{y_1^{2^k-1}}(\mathbf{x}') \geq \gamma'^{2^k-1}$ and $\|p_{y_0,\mathbf{x}} - p_{y_0,\mathbf{x}'}\|_\infty \geq \gamma'$. Therefore, due to the strict concavity of the entropy function, it follows from (3.10) that there exists $\epsilon'(\gamma') > 0$ such that:

$$\begin{aligned} H\big(g_*(X_0^{2^k-1})\big|Y_0^{2^k-1} = y_0^{2^k-1}\big) &\geq \Big(\sum_{\mathbf{x}\in\mathcal{X}^{2^k-1}} p_{y_1^{2^k-1}}(\mathbf{x}) H(p_{y_0,\mathbf{x}})\Big) + \epsilon'(\gamma') \\ &= H(X_0|Y_0 = y_0) + \epsilon'(\gamma'). \end{aligned} \tag{3.12}$$

Moreover, since the space of probability distributions on $\mathcal{X}$ is compact, $\epsilon'(\gamma') > 0$ can be chosen so that it depends only on γ' and $|\mathcal{X}|$. We have:

$$\begin{aligned} &H\big(g_*(X_0^{2^k-1})\big|Y_0^{2^k-1}\big) \\ &\quad= \sum_{y_0^{2^k-1}\in\mathcal{C}} H\big(g_*(X_0^{2^k-1})\big|Y_0^{2^k-1} = y_0^{2^k-1}\big) P_{Y_0^{2^k-1}}(y_0^{2^k-1}) \\ &\qquad\qquad + \sum_{y_0^{2^k-1}\in\mathcal{C}^c} H\big(g_*(X_0^{2^k-1})\big|Y_0^{2^k-1} = y_0^{2^k-1}\big) P_{Y_0^{2^k-1}}(y_0^{2^k-1}) \\ &\quad\overset{(a)}{\geq} \sum_{y_0^{2^k-1}\in\mathcal{C}} H(X_0|Y_0 = y_0) P_{Y_0^{2^k-1}}(y_0^{2^k-1}) \\ &\qquad\qquad + \sum_{y_0^{2^k-1}\in\mathcal{C}^c} \big(H(X_0|Y_0 = y_0) + \epsilon'(\gamma')\big) P_{Y_0^{2^k-1}}(y_0^{2^k-1}) \\ &\quad= \Big(\sum_{y_0^{2^k-1}\in\mathcal{Y}^{2^k-1}} H(X_0|Y_0 = y_0) P_{Y_0^{2^k-1}}(y_0^{2^k-1})\Big) + \epsilon'(\gamma') P_{Y_0^{2^k-1}}(\mathcal{C}^c) \\ &\quad= H(X_0|Y_0) + \epsilon'(\gamma') P_{Y_0^{2^k-1}}(\mathcal{C}^c), \end{aligned}$$

where (a) follows from (3.11) and (3.12). Let $\epsilon(\gamma) = \epsilon'(\gamma')\gamma'^{2^k}$.

Clearly, if $H\big(g_*(X_0^{2^k-1})\big|Y_0^{2^k-1}\big) < H(X_0|Y_0) + \epsilon(\gamma)$, then we must have

$$P_{Y_0^{2^k-1}}(\mathcal{C}^c) < \gamma'^{2^k}.$$

□

In the next few definitions and lemmas, $(X_i, Y_i)_{0 \le i < 2^k}$ is a sequence of 2^k random pairs that satisfy conditions 1), 2) and 3) of Proposition 3.2 where $\epsilon(\gamma)$ is as in Lemma 3.5. In particular, we have $H\big(g_*(X_0^{2^k-1})\big|Y_0^{2^k-1}\big) < H(X_0|Y_0) + \epsilon(\gamma)$ and so by Lemma 3.5 we have $P_{Y_0^{2^k-1}}(\mathcal{C}) > 1 - \gamma'^{2^k}$, where

$$\gamma' = \min\left\{\frac{\gamma}{2^{|\mathcal{X}|}+1}, \frac{1}{(2^{|\mathcal{X}|}+2)|\mathcal{X}|}\right\}.$$

Notation 3.7. *Define the following:*

- *For each $y_0 \in \mathcal{Y}$, let $\mathcal{C}_{y_0} := \big\{y_1^{2^k-1} \in \mathcal{Y}^{2^k-1} :\ y_0^{2^k-1} \in \mathcal{C}\big\}$.*
- $\mathcal{C}_0 := \big\{y_0 \in \mathcal{Y} :\ P_{Y_1^{2^k-1}}(\mathcal{C}_{y_0}) > 1 - \gamma'^{2^k-1}\big\}$.
- *For each $y \in \mathcal{Y}$, let $A_y = \{x \in \mathcal{X} :\ p_y(x) \ge \gamma'\}$.*
- *For each $D \subset \mathcal{X}$, let $\mathcal{Y}_D = \{y \in \mathcal{Y} :\ A_y = D\}$.*
- $\mathcal{A} = \{D_0 \subset \mathcal{X} :\ P_{Y_0}(\mathcal{Y}_{D_0}) \ge \gamma'\}$.

We will show later that $\mathcal{A}$ is actually the stable partition $\mathcal{H}$ of $(\mathcal{X}, *)$ that is claimed in Proposition 3.2.

Lemma 3.6. *We have:*

- $P_{Y_0}(\mathcal{C}_0) > 1 - \gamma'$.
- *For every $D_0 \in \mathcal{A}$ there exists $y_0 \in \mathcal{C}_0$ such that $A_{y_0} = D_0$.*

Proof. We have

$$\begin{aligned} 1 - \gamma'^{2^k} < P_{Y_0^{2^k-1}}(\mathcal{C}) &= \sum_{y_0 \in \mathcal{C}_0} P_{Y_0}(y_0) P_{Y_1^{2^k-1}}(\mathcal{C}_{y_0}) + \sum_{y_0 \in \mathcal{C}_0^c} P_{Y_0}(y_0) P_{Y_1^{2^k-1}}(\mathcal{C}_{y_0}) \\ &\overset{(a)}{\le} P_{Y_0}(\mathcal{C}_0) + P_{Y_0}(\mathcal{C}_0^c)(1 - \gamma'^{2^k-1}) = 1 - \gamma'^{2^k-1} P_{Y_0}(\mathcal{C}_0^c), \end{aligned}$$

where (a) follows from the fact that $P_{Y_1^{2^k-1}}(\mathcal{C}_{y_0}) \le 1 - \gamma'^{2^k-1}$ for every $y_0 \in \mathcal{C}_0^c$. We conclude that $P_{Y_0}(\mathcal{C}_0^c) < \gamma'$, hence $P_{Y_0}(\mathcal{C}_0) > 1 - \gamma'$.

Now let $D_0 \in \mathcal{A}$. We have $P_{Y_0}(\mathcal{Y}_{D_0}) \ge \gamma'$ by definition. But we have just shown that $P_{Y_0}(\mathcal{C}_0) > 1 - \gamma'$, hence $1 \ge P_{Y_0}(\mathcal{Y}_{D_0} \cup \mathcal{C}_0) = P_{Y_0}(\mathcal{Y}_{D_0}) + P_{Y_0}(\mathcal{C}_0) - P_{Y_0}(\mathcal{Y}_{D_0} \cap \mathcal{C}_0) > \gamma' + 1 - \gamma' - P_{Y_0}(\mathcal{Y}_{D_0} \cap \mathcal{C}_0)$, thus $P_{Y_0}(\mathcal{Y}_{D_0} \cap \mathcal{C}_0) > 0$. This implies that $\mathcal{Y}_{D_0} \cap \mathcal{C}_0 \ne \emptyset$. Therefore, there exists $y_0 \in \mathcal{C}_0$ such that $A_{y_0} = D_0$. □

Lemma 3.7. *$\mathcal{A}$ is an $\mathcal{X}$-cover.*

Proof. For every $y_0 \in \mathcal{Y}$, let $a_{y_0} = \arg\max_x p_{y_0}(x)$. Clearly, $P_{Y_0}(a_{y_0}) \ge \frac{1}{|\mathcal{X}|} > \gamma'$. Therefore, $a_{y_0} \in A_{y_0}$ and so $A_{y_0} \ne \emptyset$ for every $y_0 \in \mathcal{Y}$. This means that $\mathcal{Y}_\emptyset = \emptyset$, hence $P_{Y_0}(\mathcal{Y}_\emptyset) = 0 < \gamma'$. We conclude that $\emptyset \notin \mathcal{A}$.

Suppose that $\mathcal{A}$ is not an $\mathcal{X}$-cover. This means that $\bigcup_{D_0\in\mathcal{A}} D_0 \neq \mathcal{X}$. Therefore, there exists $x_0 \in \mathcal{X}$ such that $x_0 \notin \bigcup_{D_0\in\mathcal{A}} D_0$ and so $x_0 \notin D_0$ for every $D_0 \in \mathcal{A}$. We have:

$$\begin{aligned}
\frac{1}{|\mathcal{X}|} = P_{X_0}(x_0) &= \sum_{y_0\in\mathcal{Y}} P_{Y_0}(y_0)p_{y_0}(x_0) \overset{(a)}{=} \sum_{D_0\subset\mathcal{X}} \sum_{y_0\in\mathcal{Y}_{D_0}} P_{Y_0}(y_0)p_{y_0}(x_0) \\
&= \sum_{D_0\in\mathcal{A}} \sum_{y_0\in\mathcal{Y}_{D_0}} P_{Y_0}(y_0)p_{y_0}(x_0) + \sum_{\substack{D_0\subset\mathcal{X}\\ D_0\notin\mathcal{A}}} \sum_{y_0\in\mathcal{Y}_{D_0}} P_{Y_0}(y_0)p_{y_0}(x_0) \\
&\overset{(b)}{\leq} \sum_{D_0\in\mathcal{A}} \sum_{y_0\in\mathcal{Y}_{D_0}} P_{Y_0}(y_0)\gamma' + \sum_{\substack{D_0\subset\mathcal{X}\\ D_0\notin\mathcal{A}}} \sum_{y_0\in\mathcal{Y}_{D_0}} P_{Y_0}(y_0) = \sum_{D_0\in\mathcal{A}} P_{Y_0}(\mathcal{Y}_{D_0})\gamma' + \sum_{\substack{D_0\subset\mathcal{X}\\ D_0\notin\mathcal{A}}} P_{Y_0}(\mathcal{Y}_{D_0}) \\
&\overset{(c)}{\leq} P_{Y_0}\Big(\bigcup_{D_0\in\mathcal{A}} \mathcal{Y}_{D_0}\Big)\gamma' + \sum_{\substack{D_0\subset\mathcal{X}\\ D_0\notin\mathcal{A}}} \gamma' \overset{(d)}{\leq} \gamma' + 2^{|\mathcal{X}|}\gamma' \leq (2^{|\mathcal{X}|}+1)\frac{1}{(2^{|\mathcal{X}|}+2)|\mathcal{X}|} < \frac{1}{|\mathcal{X}|},
\end{aligned}$$

where (a) follows from the fact that $\{\mathcal{Y}_{D_0} : D_0 \subset \mathcal{X}\}$ is a partition of $\mathcal{Y}$. (b) follows from the fact that if $D_0 \in \mathcal{A}$ and $y_0 \in \mathcal{Y}_{D_0}$, then $A_{y_0} = D_0 \in \mathcal{A}$ and so $x_0 \notin A_{y_0}$ (since $x_0 \notin D_0$ for every $D_0 \in \mathcal{A}$) which implies that $p_{y_0}(x_0) < \gamma'$. (c) follows from the fact that $\{\mathcal{Y}_{D_0} : D_0 \subset \mathcal{X}\}$ is a partition of $\mathcal{Y}$ and from the fact that $P_{Y_0}(\mathcal{Y}_{D_0}) < \gamma'$ for every $D_0 \notin \mathcal{A}$. (d) follows from the fact that there are at most $2^{|\mathcal{X}|}$ subsets of $\mathcal{X}$. We conclude that if $\mathcal{A}$ is not an $\mathcal{X}$-cover, then $\frac{1}{|\mathcal{X}|} < \frac{1}{|\mathcal{X}|}$ which is a contradiction. Therefore, $\mathcal{A}$ is an $\mathcal{X}$-cover. □

The next three lemmas will be used to show that $\mathcal{A}$ is a stable partition.

Lemma 3.8. *Let $k = 2^{2^{|\mathcal{X}|}} + \mathrm{scon}(*)$. For every $x \in \mathcal{X}$ there exists a sequence $\mathfrak{X} = (X_i)_{0\leq i<k}$ of length k such that $X_i \in \mathcal{A}^{i*}$ for every $0 \leq i < k$, and $x * \mathfrak{X} \in \langle\mathcal{A}\rangle^{k*}$.*

Proof. Since $\mathcal{A}$ is an $\mathcal{X}$-cover, we can apply Theorem 2.3. Therefore, there exists $0 \leq n < 2^{2^{|\mathcal{X}|}}$ such that $\mathrm{core}(\mathcal{A}^{n*}) = \langle\mathcal{A}\rangle$ and $\mathrm{per}(\langle\mathcal{A}\rangle)$ divides n. Fix $x \in \mathcal{X}$ and $X \in \langle\mathcal{A}\rangle^{k*} = \langle\mathcal{A}\rangle^{(k-n)*} = \mathrm{core}(\mathcal{A}^{n*})^{(k-n)*}$, and let $A \in \mathcal{A}$ be such that $x \in A$. Choose an arbitrary sequence $\mathfrak{X}_1 = (X_i)_{0\leq i<n}$ such that $X_i \in \mathcal{A}^{i*}$ for $0 \leq i < n$. Clearly, $A * \mathfrak{X}_1 \in \mathcal{A}^{n*}$. Since $\mathcal{A} \preceq \langle\mathcal{A}\rangle$, we have $\mathcal{A}^{n*} \preceq \langle\mathcal{A}\rangle^{n*} \overset{(a)}{=} \langle\mathcal{A}\rangle = \mathrm{core}(\mathcal{A}^{n*})$, where (a) follows from the fact that $\mathrm{per}(\langle\mathcal{A}\rangle)$ divides n. We conclude that there exists $B \in \mathrm{core}(\mathcal{A}^{n*})$ such that $A * \mathfrak{X}_1 \subset B$.

Since $k = 2^{2^{|\mathcal{X}|}} + \mathrm{scon}(*)$ and $0 \leq n < 2^{2^{|\mathcal{X}|}}$, we have $k - n > \mathrm{scon}(*)$. Let $x' \in x * \mathfrak{X}_1$. Since $k - n > \mathrm{scon}(*)$, we can apply Theorem 2.2 to get a sequence $\mathfrak{X}_2 = (X'_i)_{0\leq i<k-n}$ such that $X'_i \in \langle\mathcal{A}\rangle^{i*} = \mathrm{core}(\mathcal{A}^{n*})^{i*} \subset \mathcal{A}^{(n+i)*}$ for every $0 \leq i < k - n$, and $x' * \mathfrak{X}_2 = X$. Since $x' \in x * \mathfrak{X}_1 \subset A * \mathfrak{X}_1 \subset B$, we have $X = x' * \mathfrak{X}_2 \subset (x * \mathfrak{X}_1) * \mathfrak{X}_2 \subset B * \mathfrak{X}_2$. But both X and $B * \mathfrak{X}_2$ are elements of $\langle\mathcal{A}\rangle^{(k-n)*}$ which is a partition, so we must have $B * \mathfrak{X}_2 = X$. Now define $\mathfrak{X} = (\mathfrak{X}_1, \mathfrak{X}_2)$. We have $X = x' * \mathfrak{X}_2 \subset x * \mathfrak{X} \subset B * \mathfrak{X}_2 = X$. Therefore, $x * \mathfrak{X} = X \in \langle\mathcal{A}\rangle^{k*}$. □

Lemma 3.9. *For every $i \geq 0$ and every $X \in \mathcal{A}^{i*}$ there exist 2^i sets $B_0, \ldots, B_{2^i-1} \in \mathcal{A}$ such that*

$$X = \left\{ g_*(\mathbf{x}) : \mathbf{x} \in \prod_{j=0}^{2^i-1} B_j \right\} := \{ g_*(x_0, \ldots, x_{2^i-1}) : x_0 \in B_0, \ldots, x_{2^i-1} \in B_{2^i-1} \}.$$

Proof. We will show the lemma by induction on $i \geq 0$. The lemma is trivial for $i = 0$: Take $B_0 = X \in \mathcal{A}$, we get

$$X = \{x : x \in B_0\} = \left\{ g_*(\mathbf{x}) : \mathbf{x} \in \prod_{j=0}^{2^0-1} B_j \right\}.$$

Now let $i > 0$ and suppose that the lemma is true for $i - 1$. Let $X \in \mathcal{A}^{i*}$, and let $X', X'' \in \mathcal{A}^{(i-1)*}$ be such that $X = X' * X''$. It follows from the induction hypothesis that there exist 2^{i-1} sets $B'_0, \ldots, B'_{2^{i-1}-1} \in \mathcal{A}$ and 2^{i-1} sets $B''_0, \ldots, B''_{2^{i-1}-1} \in \mathcal{A}$ such that

$$X' = \left\{ g_*(\mathbf{x}') : \mathbf{x}' \in \prod_{j=0}^{2^{i-1}-1} B'_j \right\} \quad \text{and} \quad X'' = \left\{ g_*(\mathbf{x}'') : \mathbf{x}'' \in \prod_{j=0}^{2^{i-1}-1} B''_j \right\}.$$

We have:

$$\begin{aligned} X = X' * X'' &= \left\{ g_*(\mathbf{x}') : \mathbf{x}' \in \prod_{j=0}^{2^{i-1}-1} B'_j \right\} * \left\{ g_*(\mathbf{x}'') : \mathbf{x}'' \in \prod_{j=0}^{2^{i-1}-1} B''_j \right\} \\ &= \left\{ g_*(\mathbf{x}') * g_*(\mathbf{x}'') : \mathbf{x}' \in \prod_{j=0}^{2^{i-1}-1} B'_j,\ \mathbf{x}'' \in \prod_{j=0}^{2^{i-1}-1} B''_j \right\} \\ &= \left\{ g_*(\mathbf{x}) : \mathbf{x} \in \left(\prod_{j=0}^{2^{i-1}-1} B'_j \right) \times \left(\prod_{j=0}^{2^{i-1}-1} B''_j \right) \right\} = \left\{ g_*(\mathbf{x}) : \mathbf{x} \in \prod_{j=0}^{2^i-1} B_j \right\}, \end{aligned}$$

where

$$B_j = \begin{cases} B'_j & \text{if } 0 \leq j < 2^{i-1}, \\ B''_{j-2^{i-1}} & \text{if } 2^{i-1} \leq j < 2^i. \end{cases}$$

□

Lemma 3.10. *Let $\mathfrak{X} = (X_i)_{0 \leq i < l}$ be a sequence of length $l > 0$ such that $X_i \in \mathcal{A}^{i*}$ for every $0 \leq i < l$. There exist $2^l - 1$ sets $D_1, \ldots, D_{2^l-1} \in \mathcal{A}$ such that for every $x \in \mathcal{X}$, we have*

$$x * \mathfrak{X} = \left\{ g_*\big((x, \mathbf{x})\big) : \mathbf{x} \in \prod_{i=1}^{2^l-1} D_i \right\} = \left\{ \pi_{\mathbf{x}}(x) : \mathbf{x} \in \prod_{i=1}^{2^l-1} D_i \right\}.$$

Proof. We will show the lemma by induction on $l > 0$. If $l = 1$, the lemma is trivial: If we take $D_1 = X_0 \in \mathcal{A}$, then for every $x \in \mathcal{X}$ we have

$$x * \mathfrak{X} = \{x * x_0 : x_0 \in X_0\} = \big\{ g_*\big((x, x_0)\big) : x_0 \in D_1 \big\} = \left\{ g_*\big((x, \mathbf{x})\big) : \mathbf{x} \in \prod_{i=1}^{2^1-1} D_i \right\}.$$

Now let $l > 1$ and suppose that the lemma is true for $l-1$. Let $\mathfrak{X} = (X_i)_{0\le i<l}$ and define the sequence $\mathfrak{X}' = (X_i)_{0\le i<l-1}$. The induction hypothesis implies the existence of $2^{l-1}-1$ sets $D'_1, \ldots, D'_{2^{l-1}-1} \in \mathcal{A}$ such that for every $x \in \mathcal{X}$ we have

$$x * \mathfrak{X}' = \Big\{ g_*\big((x, \mathbf{x}')\big) : \mathbf{x}' \in \prod_{i=1}^{2^{l-1}-1} D'_i \Big\}.$$

On the other hand, since $X_{l-1} \in \mathcal{A}^{(l-1)*}$, Lemma 3.9 shows the existence of 2^{l-1} sets $D''_0, \ldots, D''_{2^{l-1}-1} \in \mathcal{A}$ such that

$$X_{l-1} = \Big\{ g_*(\mathbf{x}'') : \mathbf{x}'' \in \prod_{i=0}^{2^{l-1}-1} D''_i \Big\}.$$

Define the $2^l - 1$ sets $D_1, \ldots, D_{2^l-1} \in \mathcal{A}$ as follows:

$$D_i = \begin{cases} D'_i & \text{if } 1 \le i < 2^{l-1}, \\ D''_{i-2^{l-1}} & \text{if } 2^{l-1} \le i < 2^l. \end{cases}$$

For every $x \in \mathcal{X}$ we have:

$$\begin{aligned} x * \mathfrak{X} &= (x * \mathfrak{X}') * X_{l-1} \\ &= \Big\{ g_*\big((x, \mathbf{x}')\big) : \mathbf{x}' \in \prod_{i=1}^{2^{l-1}-1} D'_i \Big\} * \Big\{ g_*(\mathbf{x}'') : \mathbf{x}'' \in \prod_{i=0}^{2^{l-1}-1} D''_i \Big\} \\ &= \Big\{ g_*\big((x, \mathbf{x}')\big) * g_*(\mathbf{x}'') : \mathbf{x}' \in \prod_{i=1}^{2^{l-1}-1} D_i,\ \mathbf{x}'' \in \prod_{i=2^{l-1}}^{2^l-1} D_i \Big\} \\ &= \Big\{ g_*\big((x, \mathbf{x})\big) : \mathbf{x} \in \prod_{i=1}^{2^l-1} D_i \Big\}. \end{aligned}$$

□

Lemma 3.11. *We have the following:*

1. *$\mathcal{A}$ is a stable partition of $(\mathcal{X}, *)$.*
2. *If $y_0 \in \mathcal{C}_0$ and $A_{y_0} \in \mathcal{A}$ then $y_0 \in \mathcal{Y}_{\mathcal{A},\gamma}(X_0, Y_0)$.*

Proof. 1) Let $D_0 \in \mathcal{A}$. By Lemma 3.6, there exists $y_0 \in \mathcal{C}_0$ such that $D_0 = A_{y_0}$. Let $a_{y_0} = \arg\max_{x\in\mathcal{X}} p_{y_0}(x)$. Clearly, $p_{y_0}(a_{y_0}) \ge \frac{1}{|\mathcal{X}|} > \gamma'$ and so $a_{y_0} \in A_{y_0} = D_0$.

Since $\mathcal{A}$ is an $\mathcal{X}$-cover (Lemma 3.7), Theorem 2.3 implies the existence of an integer n satisfying $0 \le n < 2^{2^{|\mathcal{X}|}}$, $\mathrm{core}(\mathcal{A}^{n*}) = \langle\mathcal{A}\rangle$ and $\mathrm{per}(\langle\mathcal{A}\rangle)$ divides n. Moreover, Lemma 3.8 shows the existence of a sequence $\mathfrak{X} = (X_i)_{0\le i<k}$ such that $X_i \in \mathcal{A}^{i*}$ for all $0 \le i < k$ and $a_{y_0} * \mathfrak{X} \in \langle\mathcal{A}\rangle^{k*} = \langle\mathcal{A}\rangle^{(k-n)*}$. Let

$$B = a_{y_0} * \mathfrak{X} \in \langle\mathcal{A}\rangle^{k*} = \langle\mathcal{A}\rangle^{(k-n)*}. \tag{3.13}$$

Lemma 3.10 shows the existence of 2^k-1 sets $D_1,\dots,D_{2^k-1}\in\mathcal{A}$ such that

$$B=\Big\{g_*\big((a_{y_0},\mathbf{x})\big):\ \mathbf{x}\in\prod_{i=1}^{2^k-1}D_i\Big\}=\Big\{\pi_{\mathbf{x}}(a_{y_0}):\ \mathbf{x}\in\prod_{i=1}^{2^k-1}D_i\Big\}. \tag{3.14}$$

Define

$$\mathcal{C}'_{y_0}=\big\{y_1^{2^k-1}\in\mathcal{Y}^{2^k-1}:\ \forall 1\le i<2^k,\ A_{y_i}=D_i\big\}=\prod_{i=1}^{2^k-1}\mathcal{Y}_{D_i}.$$

Since $D_1,\dots,D_{2^k-1}\in\mathcal{A}$, we have

$$P_{Y_1^{2^k-1}}(\mathcal{C}'_{y_0})=\prod_{i=1}^{2^k-1}P_{Y_i}(\mathcal{Y}_{D_i})=\prod_{i=1}^{2^k-1}P_{Y_0}(\mathcal{Y}_{D_i})\ge\gamma'^{2^k-1}.$$

On the other hand, since $y_0\in\mathcal{C}_0$, we have $P_{Y_1^{2^k-1}}(\mathcal{C}_{y_0})>1-\gamma'^{2^k-1}$ from the definition of $\mathcal{C}_0$. Therefore, $P_{Y_1^{2^k-1}}(\mathcal{C}_{y_0}\cap\mathcal{C}'_{y_0})>0$ which implies that $\mathcal{C}_{y_0}\cap\mathcal{C}'_{y_0}\neq\emptyset$. Hence, there exists a sequence $(y_1,\dots,y_{2^k-1})\in\mathcal{C}_{y_0}$ such that $A_{y_i}=D_i$ for all $1\le i<2^k$.

Now fix a sequence

$$\mathbf{x}'=(x'_i)_{1\le i<2^k}\text{ such that }x'_i\in D_i\text{ for all }1\le i<2^k. \tag{3.15}$$

Let $x\in\pi_{\mathbf{x}'}^{-1}(B)$, then there exists $x'\in B$ such that $x'=\pi_{\mathbf{x}'}(x)$. Now from (3.14), since $x'\in B$, there exists a sequence $\mathbf{x}=(x_i)_{1\le i<2^k}$ such that $x_i\in D_i$ for all $1\le i<2^k$ and $x'=\pi_{\mathbf{x}}(a_{y_0})$. We have:

- $(y_i)_{0\le i<2^k}\in\mathcal{C}$ since $(y_1,\dots,y_{2^k-1})\in\mathcal{C}_{y_0}$.
- For every $1\le i<2^k$, we have $p_{y_i}(x_i)\ge\gamma'$ and $p_{y_i}(x'_i)\ge\gamma'$ since $x_i,x'_i\in D_i=A_{y_i}$. Therefore, $p_{y_1^{2^k-1}}(\mathbf{x})=\prod_{i=1}^{2^k-1}p_{y_i}(x_i)\ge\gamma'^{2^k-1}$ and similarly $p_{y_1^{2^k-1}}(\mathbf{x}')\ge\gamma'^{2^k-1}$.

From the definition of $\mathcal{C}$, we get $\|p_{y_0,\mathbf{x}}-p_{y_0,\mathbf{x}'}\|_\infty<\gamma'$ which implies that $|p_{y_0,\mathbf{x}}(x')-p_{y_0,\mathbf{x}'}(x')|<\gamma'$. Therefore,

$$|p_{y_0}(a_{y_0})-p_{y_0}(x)|=\big|p_{y_0}\big(\pi_{\mathbf{x}}^{-1}(x')\big)-p_{y_0}\big(\pi_{\mathbf{x}'}^{-1}(x')\big)\big|=|p_{y_0,\mathbf{x}}(x')-p_{y_0,\mathbf{x}'}(x')|<\gamma'.$$

We conclude that

$$\forall x\in\pi_{\mathbf{x}'}^{-1}(B),\ |p_{y_0}(a_{y_0})-p_{y_0}(x)|<\gamma', \tag{3.16}$$

and so $p_{y_0}(x)>p_{y_0}(a_{y_0})-\gamma'\ge\frac{1}{|\mathcal{X}|}-\gamma'\ge\frac{1}{|\mathcal{X}|}-\frac{1}{|\mathcal{X}|(2^{|\mathcal{X}|}+2)}>\frac{1}{|\mathcal{X}|}-\frac{1}{2|\mathcal{X}|}=\frac{1}{2|\mathcal{X}|}>\frac{1}{(2^{|\mathcal{X}|}+2)|\mathcal{X}|}\ge\gamma'$ which implies that $x\in A_{y_0}=D_0$. But this is true for every $x\in\pi_{\mathbf{x}'}^{-1}(B)$. We conclude that $\pi_{\mathbf{x}'}^{-1}(B)\subset D_0$. On the other hand, since $D_0\in\mathcal{A}\preceq\langle\mathcal{A}\rangle$, there exists $C\in\langle\mathcal{A}\rangle$ such that $D_0\subset C$. Therefore, $\pi_{\mathbf{x}'}^{-1}(B)\subset D_0\subset C$ and

$$\|\langle\mathcal{A}\rangle\|=\|\langle\mathcal{A}\rangle^{(k-n)*}\|=|B|\overset{(a)}{=}|\pi_{\mathbf{x}'}^{-1}(B)|\le|D_0|\le|C|=\|\langle\mathcal{A}\rangle\|,$$

where (a) follows from the fact that $\pi_{\mathbf{x}'}$ is a bijection. We conclude that $\|\langle\mathcal{A}\rangle\| = |\pi_{\mathbf{x}'}^{-1}(B)| = |D_0| = |C|$. But $\pi_{\mathbf{x}'}^{-1}(B) \subset D_0 \subset C$, so we must have

$$\pi_{\mathbf{x}'}^{-1}(B) = D_0 = C \in \langle\mathcal{A}\rangle. \tag{3.17}$$

Now since this is true for every $D_0 \in \mathcal{A}$, we conclude that $\mathcal{A} \subset \langle\mathcal{A}\rangle$. But $\mathcal{A}$ is an $\mathcal{X}$-cover and $\langle\mathcal{A}\rangle$ is a partition, so we must have $\mathcal{A} = \langle\mathcal{A}\rangle$. We conclude that $\mathcal{A}$ is a stable partition.

2) Let $y_0 \in \mathcal{C}_{y_0}$ and suppose that $D_0 = A_{y_0} \in \mathcal{A}$. Define $a_{y_0} = \arg\max_{x\in\mathcal{X}} p_{y_0}(x)$. Let $B \in \mathcal{A}^{k*}$ and $\mathbf{x}' \in \mathcal{X}^{2^k-1}$ be defined as in equations (3.13) and (3.15) respectively. Equation (3.17) shows that $D_0 = \pi_{\mathbf{x}'}^{-1}(B)$. By replacing $\pi_{\mathbf{x}'}^{-1}(B)$ by D_0 in equation (3.16), we conclude that for every $x \in D_0$ we have $|p_{y_0}(a_{y_0}) - p_{y_0}(x)| < \gamma'$, which means that

$$p_{y_0}(a_{y_0}) - \gamma' < p_{y_0}(x) < p_{y_0}(a_{y_0}) + \gamma'. \tag{3.18}$$

On the other hand, for every $x \in \mathcal{X} \setminus D_0 = \mathcal{X} \setminus A_{y_0}$, we have

$$0 \le p_{y_0}(x) < \gamma'. \tag{3.19}$$

By adding up the inequalities (3.18) for all $x \in D_0$ with the inequalities (3.19) for all $x \in \mathcal{X} \setminus D_0$, we get $|D_0| \cdot p_{y_0}(a_{y_0}) - |D_0| \cdot \gamma' < 1 < |D_0| \cdot p_{y_0}(a_{y_0}) + |\mathcal{X}| \cdot \gamma'$, from which we get $|p_{y_0}(a_{y_0}) - \frac{1}{|D_0|}| < \frac{|\mathcal{X}|}{|D_0|}\gamma' \le |\mathcal{X}|\gamma'$. We conclude that for every $x \in D_0$, we have

$$\left|p_{y_0}(x) - \frac{1}{|D_0|}\right| \le |p_{y_0}(x) - p_{y_0}(a_{y_0})| + \left|p_{y_0}(a_{y_0}) - \frac{1}{|D_0|}\right| < \gamma' + |\mathcal{X}|\gamma' < (2^{|\mathcal{X}|}+1)\gamma' \le \gamma,$$

and for every $x \in \mathcal{X} \setminus D_0 = \mathcal{X} \setminus A_{y_0}$, we have $p_{y_0}(x) < \gamma' < \gamma$. Therefore, $\|p_{y_0} - \mathbb{I}_{D_0}\|_\infty \le \gamma$ and so $y_0 \in \mathcal{Y}_{\mathcal{A},\gamma}(X_0, Y_0)$. □

Now we are ready to prove Proposition 3.2:

Proof of Proposition 3.2. According to Lemma 3.11, $\mathcal{A}$ is a stable partition. Moreover, for every $y_0 \in \mathcal{C}_0$ satisfying $A_{y_0} \in \mathcal{A}$, we have $y_0 \in \mathcal{Y}_{\mathcal{A},\gamma}(X_0, Y_0)$. Therefore, if we define

$$\mathcal{Y}'_{\mathcal{A}} = \big\{y \in \mathcal{Y} :\ A_y \in \mathcal{A}\big\},$$

then $\mathcal{Y}'_{\mathcal{A}} \cap \mathcal{C}_0 \subset \mathcal{Y}_{\mathcal{A},\gamma}(X_0, Y_0)$.

We have $\mathcal{Y}'^c_{\mathcal{A}} = \bigcup_{\substack{D\subset\mathcal{X}\\ D\notin\mathcal{A}}} \mathcal{Y}_D$. Now since $P_{Y_0}(\mathcal{Y}_D) < \gamma'$ for every $D \notin \mathcal{A}$, we have:

$$P_{Y_0}(\mathcal{Y}'^c_{\mathcal{A}}) \le \sum_{\substack{D\subset\mathcal{X}\\ D\notin\mathcal{A}}} P_{Y_0}(\mathcal{Y}_D) < 2^{|\mathcal{X}|}\gamma'.$$

But $P_{Y_0}(\mathcal{C}_0) > 1 - \gamma'$ by Lemma 3.6, so we have $P_{Y_0}(\mathcal{Y}'_{\mathcal{A}} \cap \mathcal{C}_0) > 1 - (2^{|\mathcal{X}|} + 1)\gamma' \ge 1 - \gamma$, which implies that $P_{Y_0}(\mathcal{Y}_{\mathcal{A},\gamma}(X_0, Y_0)) > 1 - \gamma$ since $\mathcal{Y}'_{\mathcal{A}} \cap \mathcal{C}_0 \subset \mathcal{Y}_{\mathcal{A},\gamma}(X_0, Y_0)$. By letting $\mathcal{H} = \mathcal{A}$, which is a stable partition, we get $\mathcal{P}_{\mathcal{H},\gamma}(X_0, Y_0) = P_{Y_0}(\mathcal{Y}_{\mathcal{H},\gamma}(X_0, Y_0)) > 1 - \gamma$. □

4 MAC Polarization Theory

In this chapter[1], we generalize the results of [8] and [9] to arbitrary multiple-access channels. In Section 4.1, we define the multiple-access channels and their capacity regions. In Section 4.2, we provide a formal definition of MAC-polarizing sequences of binary operations. In Section 4.3, we prove that a sequence of binary operations is MAC-polarizing if and only if every binary operation in the sequence is uniformity-preserving and its right-inverse is strongly ergodic. In Section 4.4, we explain how we can use a MAC-polarizing sequence of binary operations to construct MAC-polar codes for arbitrary multiple access channels. In Section 4.5, we show that if we use special binary operations (namely, the addition modulo the size of the input alphabets), the MAC-polar code construction becomes simpler.

4.1 Multiple-Access Channels

Definition 4.1. *A discrete m-user multiple-access channel (MAC) is an $(m+2)$-tuple $W = (\mathcal{X}_1, \ldots, \mathcal{X}_m, \mathcal{Y}, p_W)$, where $\mathcal{X}_1, \ldots, \mathcal{X}_m$ are finite sets that are called the* input alphabets *of W, $\mathcal{Y}$ is a finite set that is called the* output alphabet *of W, and $p_W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \times \mathcal{Y} \to [0,1]$ is a mapping that satisfies*

$$\forall (x_1, x_2, \ldots, x_m) \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_m, \ \sum_{y \in \mathcal{Y}} p_W(x_1, x_2, \ldots, x_m, y) = 1.$$

Notation 4.1. *We write $W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \longrightarrow \mathcal{Y}$ to denote that W has m users, $\mathcal{X}_1$, $\mathcal{X}_2$, $\ldots$, $\mathcal{X}_m$ as input alphabets, and $\mathcal{Y}$ as output alphabet. We denote $p_W(x_1, x_2, \ldots, x_m, y)$ as $W(y|x_1, x_2, \ldots, x_m)$ which is interpreted as the conditional probability of receiving y at the output, given that $(x_1, x_2, \ldots, x_m)$ is the input.*

Note that we use the long arrow ($\longrightarrow$) in the notation $W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \longrightarrow \mathcal{Y}$ and not the short arrow ($\to$) which we only use to describe mappings. For example, $W : \mathcal{X}_1 \times \mathcal{X}_2 \longrightarrow \mathcal{Y}$ denotes a 2-user MAC, whereas $V : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}$ denotes a mapping from $\mathcal{X}_1 \times \mathcal{X}_2$ to $\mathcal{Y}$.

[1]The material of this chapter is based on [17, 18, 20, 21].

Definition 4.2. *An m-user MAC-coding scheme $\mathcal{C}$ is a $(2m+2)$-tuple*

$$\mathcal{C} = (\mathcal{M}_1, \ldots, \mathcal{M}_m, N, f_1, \ldots, f_m, g).$$

$\mathcal{M}_k$ is the message set *of the k^{th} user, N is the blocklength, $f_k : \mathcal{M}_k \to \mathcal{X}_k^N$ is the encoder of the k^{th} user, and $g : \mathcal{Y}^N \to \mathcal{M}_1 \times \cdots \times \mathcal{M}_m$ is the decoder.*

The rate of transmission *for the k^{th} user is defined as $R_k = \frac{\log_2 |\mathcal{M}_k|}{N}$. The* rate vector *of the MAC-coding scheme is the m-tuple $(R_1, \ldots, R_m) \in \mathbb{R}^m$. The quantity $R_1 + \cdots + R_m$ is called* the sum-rate *of the MAC-coding scheme.*

The MAC-coding scheme $\mathcal{C} = (\mathcal{M}_1, \ldots, \mathcal{M}_m, N, f_1, \ldots, f_m, g)$ is implemented as follows:

- For every $1 \leq k \leq m$, a random message M_k is uniformly chosen from $\mathcal{M}_k$. M_k represents the message that the k^{th} user wishes to transmit to the receiver.
- For every $1 \leq k \leq m$, the k^{th} user computes $(X_{k,1}, \ldots, X_{k,N}) = f_k(M_k)$.
- For every $1 \leq k \leq m$, the k^{th} user transmits $X_{k,1}, \ldots, X_{k,N}$ to the receiver by using the MAC N times. More precisely, at the i^{th} use of the MAC, the k^{th} user transmits the symbol $X_{k,i}$.
- The receiver observes N output symbols $Y_1, \ldots, Y_N$.
- The receiver computes an estimate of the transmitted messages as

$$(\hat{M}_1, \ldots, \hat{M}_m) = g(Y_1, \ldots, Y_N).$$

The probability of error of the MAC-coding scheme $\mathcal{C}$ when it is used for the MAC W is given by

$$P_e(\mathcal{C}, W) = \mathbb{P}[(\hat{M}_1, \ldots, \hat{M}_m) \neq (M_1, \ldots, M_m)].$$

Definition 4.3. *A rate vector $R = (R_1, \ldots, R_m)$ is said to be achievable for the MAC $W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \longrightarrow \mathcal{Y}$ if for every $\delta, \epsilon > 0$, there exists a MAC-coding scheme of rate vector at least*[2] *$(R_1 - \delta, \ldots, R_m - \delta)$ and whose probability of error is at most ϵ. The* capacity region *of the MAC W is the set of all achievable rate vectors.*

Definition 4.4. *Given a MAC W and a collection of independent random variables $X_1, \ldots, X_m$ taking values in $\mathcal{X}_1, \ldots, \mathcal{X}_m$ respectively, we define the polymatroid region $\mathcal{J}_{X_1,\ldots,X_m}(W)$ in $\mathbb{R}^m$ as:*

$$\mathcal{J}_{X_1,\ldots,X_m}(W) := \Big\{ R = (R_1, \ldots, R_m) \in \mathbb{R}^m : \\ 0 \leq R(S) \leq I_{X_1,\ldots,X_m,S}(W) \ \text{ for all } \ S \subset \{1, \ldots, m\} \Big\},$$

where $R(S) := \sum_{k \in S} R_k$, $X(S) := (X_k)_{k \in S}$, and

$$I_{X_1,\ldots,X_m,S}(W) := I(X(S); Y X(S^c)).$$

[2] We consider that $(R_1, \ldots, R_m) \leq (R'_1, \ldots, R'_m)$ if $R_i \leq R'_i$ for every $1 \leq i \leq m$.

The mutual information is computed for the probability distribution on $\mathcal{X}_1 \times \cdots \times \mathcal{X}_m \times \mathcal{Y}$ *which is given by*

$$P_{X_1,\ldots,X_m,Y}(x_1,\ldots,x_m,y) = P_{X_1}(x_1)\cdots P_{X_m}(x_m)W(y|x_1,\ldots,x_m).$$

$\mathcal{J}_{X_1,\ldots,X_m}(W)$ *is called the* information-theoretic capacity region *of the MAC* W *for the input distributions* $X_1,\ldots,X_m$.

Theorem 4.1. *(Theorem 15.3.6 [3]) The capacity region of a MAC* W *is given by the closure of the convex hull of the union of all information-theoretic capacity regions of* W *for all the input distributions, i.e,*

$$\overline{\text{ConvexHull}}\left(\bigcup_{\substack{X_1,\ldots,X_m \\ \text{are independent} \\ \text{random variables in} \\ \mathcal{X}_1,\ldots,\mathcal{X}_m \text{ resp.}}} \mathcal{J}_{X_1,\ldots,X_m}(W)\right).$$

Definition 4.5. $I_{X_1,\ldots,X_m}(W) := I_{X_1,\ldots,X_m,\{1,\ldots,m\}}(W)$ *is called the* sum capacity *of* W *for the input distributions* $X_1,\ldots,X_m$. $I_{X_1,\ldots,X_m}(W)$ *is equal to the maximum value of* $R_1+\cdots+R_m$ *among all rate vectors* $(R_1,\ldots,R_m)$ *that belong to* $\mathcal{J}_{X_1,\ldots,X_m}(W)$. *The set of rate-vectors* $(R_1,\ldots,R_m)$ *in* $\mathcal{J}_{X_1,\ldots,X_m}(W)$ *which satisfy* $R_1+\cdots+R_m = I_{X_1,\ldots,X_m}(W)$ *is called* the dominant face *of* $\mathcal{J}_{X_1,\ldots,X_m}(W)$.

Notation 4.2. *For the sake of simplicity, if* $X_1,\ldots,X_m$ *are independent and uniform random variables in* $\mathcal{X}_1,\ldots,\mathcal{X}_m$ *respectively, we write* $\mathcal{J}(W)$, $I_S(W)$ *and* $I(W)$ *to denote* $\mathcal{J}_{X_1,\ldots,X_m}(W)$, $I_{X_1,\ldots,X_m,S}(W)$ *and* $I_{X_1,\ldots,X_m}(W)$, *respectively.*

$\mathcal{J}(W)$ *is called the* symmetric-capacity region *of* W, *and* $I(W)$ *is called the* symmetric sum-capacity *of* W.

4.2 MAC-Polarizing Sequences of Binary Operations

4.2.1 Easy MACs

Notation 4.3. *Let* W *be an* m*-user MAC. The probability of error of the maximum-likelihood (ML) decoder*[3] *of* W *for uniformly distributed input is denoted as* $P_e(W)$.

Definition 4.6. *An* m*-user MAC* $W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \longrightarrow \mathcal{Y}$ *is said to be* δ-easy *if there exist* m *integers* $L_1 \le |\mathcal{X}_1|,\ldots,L_m \le |\mathcal{X}_m|$, *and* m *independent random codes* $\mathcal{B}_1,\ldots,\mathcal{B}_m$ *taking values in the sets* $\mathcal{S}_1 = \{C_1 \subset \mathcal{X}_1 : |C_1| = L_1\}$, ..., $\mathcal{S}_m = \{C_m \subset \mathcal{X}_m : |C_m| = L_m\}$ *respectively, which satisfy the following:*

- $|I(W) - \log_2 L| < \delta$, *where* $L = L_1 \times \cdots \times L_m$.
- *For every* $1 \le i \le m$ *and every* $x_i \in \mathcal{X}_i$, *we have* $\displaystyle\sum_{C_i \in \mathcal{S}_i} \frac{1}{L_i} P_{\mathcal{B}_i}(C_i)\mathbb{1}_{x_i \in C_i} = \frac{1}{|\mathcal{X}_i|}$.

  In other words, if $C_i \in \mathcal{S}_i$ *is chosen according to the distribution of* $\mathcal{B}_i$ *and* X_i *is chosen uniformly in* C_i, *then the marginal distribution of* X_i *as a random variable in* $\mathcal{X}_i$ *is uniform.*

[3]The ML decoder is the decoder that minimizes the probability of error.

- *If for each* $1 \leq i \leq m$ *and each* $C_i \in \mathcal{S}_i$ *we fix a bijection* $f_{i,C_i} : \{1, ..., L_i\} \to C_i$, *then* $I(W_{\mathcal{B}_1,...,\mathcal{B}_m}) > \log_2 L - \delta$, *where*

$$W_{\mathcal{B}_1,...,\mathcal{B}_m} : \{1, ..., L_1\} \times \cdots \times \{1, ..., L_m\} \longrightarrow \mathcal{Y} \times \mathcal{S}_1 \times \cdots \times \mathcal{S}_m$$

is the MAC defined by:

$$W_{\mathcal{B}_1,...,\mathcal{B}_m}(y, C_1, \ldots, C_m | a_1, \ldots, a_m) \\ = W(y | f_{1,C_1}(a_1), \ldots, f_{m,C_m}(a_m)). \prod_{i=1}^{m} P_{\mathcal{B}_i}(C_i).$$

Note that the value of $I(W_{\mathcal{B}_1,...,\mathcal{B}_m})$ *does not depend on the choice of the bijections* $(f_{i,C_i})_{1 \leq i \leq m,\, C_i \in \mathcal{S}_i}$.

If we also have $P_e(W_{\mathcal{B}_1,...,\mathcal{B}_m}) < \epsilon$, *we say that* W *is* (δ, ϵ)-easy.

If W is a δ-easy MAC for a small δ, then we can reliably transmit information near the symmetric sum-capacity of W using a code of blocklength 1 (hence the easiness; there is no need to use codes of large blocklengths): We choose a random MAC-code according to $\mathcal{B}_1, \ldots, \mathcal{B}_m$, we reveal this code to the receiver, and then we transmit information using this code. The sum-rate of this code is equal to $\log_2 L_1 + \cdots + \log_2 L_m = \log_2 L$ which is close to the sum-capacity $I(W)$. On the other hand, the fact that $I(W_{\mathcal{B}_1,...,\mathcal{B}_m}) > \log_2 L - \delta$ means that $W_{\mathcal{B}_1,...,\mathcal{B}_m}$ is almost perfect, which ensures that our simple MAC-coding scheme has a low probability of error.

4.2.2 Polarization Process for MACs

Definition 4.7. *Let* $\mathcal{X}_1, \ldots, \mathcal{X}_m$ *be* m *arbitrary sets. Let* $*_1, \ldots, *_m$ *be* m *binary operations on* $\mathcal{X}_1, \ldots, \mathcal{X}_m$ *respectively, and let* $W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \longrightarrow \mathcal{Y}$ *be an* m*-user MAC. We define the two MACs* $W^- : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \longrightarrow \mathcal{Y} \times \mathcal{Y}$ *and* $W^+ : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \longrightarrow \mathcal{Y} \times \mathcal{Y} \times \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ *as follows:*

$$W^-(y_1, y_2 | u_{1,1}, \ldots, u_{m,1}) \\ = \frac{1}{|\mathcal{X}_1| \cdots |\mathcal{X}_m|} \sum_{\substack{u_{1,2} \in \mathcal{X}_1 \\ \vdots \\ u_{m,2} \in \mathcal{X}_m}} W(y_1 | u_{1,1} *_1 u_{1,2}, \ldots, u_{m,1} *_m u_{m,2}) W(y_2 | u_{1,2}, \ldots, u_{m,2}),$$

and

$$W^+(y_1, y_2, u_{1,1}, \ldots, u_{m,1} | u_{1,2}, \ldots, u_{m,2}) \\ = \frac{1}{|\mathcal{X}_1| \cdots |\mathcal{X}_m|} W(y_1 | u_{1,1} *_1 u_{1,2}, \ldots, u_{m,1} *_m u_{m,2}) W(y_2 | u_{1,2}, \ldots, u_{m,2}).$$

For every $s = (s_1, \ldots, s_n) \in \{-, +\}^n$, *we define* W^s *recursively as:*

$$W^s := ((W^{s_1})^{s_2} \ldots)^{s_n}.$$

Definition 4.8. *Let $(B_n)_{n\geq 1}$ be i.i.d. uniform random variables in $\{-,+\}$. For each MAC W with input alphabets $\mathcal{X}_1,\dots,\mathcal{X}_m$, we define the MAC-valued process $(W_n)_{n\geq 0}$ recursively as follows:*

$$W_0 := W,$$
$$W_n := W_{n-1}^{B_n},\ \forall n \geq 1.$$

Definition 4.9. *A sequence of m binary operations $(*_1,\dots,*_m)$ on the sets $\mathcal{X}_1,\dots,\mathcal{X}_m$ is said to be* MAC-polarizing *if we have the following two properties:*

- *Conservation property: For every MAC W with input alphabets $\mathcal{X}_1,\dots,\mathcal{X}_m$ we have*
$$I(W^-) + I(W^+) = 2I(W).$$
- *Polarization property: For every MAC W with input alphabets $\mathcal{X}_1,\dots,\mathcal{X}_m$ and every $\delta > 0$, W_n almost surely becomes δ-easy, i.e.,*
$$\lim_{n\to\infty} \mathbb{P}\big[W_n \text{ is } \delta\text{-easy}\big] = 1.$$

Notice that in the conservation property we only ask for the symmetric sum-capacity to be preserved and we do not ask for the whole symmetric-capacity region to be preserved. The reason for this is because MAC polarization sometimes induces a loss in the symmetric-capacity region (see [8] and [9]). There are, however, polar coding techniques that achieve the whole symmetric-capacity region (e.g., [22] and [23]) but those techniques are not based on MAC polarization; they are based on monotone chain rules and single-user channel polarization. In the above definition, we are only interested in the MAC polarization phenomenon itself. We note, however, that monotone chain rules can be used together with the general single-user polarization theory that was developed in Chapter 3 in order to construct MAC codes that achieve the whole symmetric-capacity region.

Remark 4.1. *As in Remark 3.1, a sequence of binary operations satisfies the conservation property if and only if every operation in the sequence is uniformity-preserving.*

Definition 4.10. *Let $(*_1,\dots,*_m)$ be a MAC-polarizing sequence on the sets $\mathcal{X}_1,\dots,\mathcal{X}_m$. We say that $\beta \geq 0$ is a* $(*_1,\dots,*_m)$-achievable exponent *if for every $\delta > 0$ and every MAC W with input alphabets $\mathcal{X}_1,\dots,\mathcal{X}_m$, W_n almost surely becomes $(\delta, 2^{-2^{\beta n}})$-easy, i.e.,*

$$\lim_{n\to\infty} \mathbb{P}\big[W_n \text{ is } (\delta, 2^{-2^{\beta n}})\text{-easy}\big] = 1.$$

We define the exponent *of $(*_1,\dots,*_m)$ as:*

$$E_{*_1,\dots,*_m} := \sup\{\beta \geq 0 :\ \beta \text{ is a } (*_1,\dots,*_m)\text{-achievable exponent}\}.$$

Remark 4.2. *For each $1 \leq i \leq m$ and each ordinary single-user channel $W_i : \mathcal{X}_i \longrightarrow \mathcal{Y}$ with input alphabet $\mathcal{X}_i$, consider the MAC $W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \longrightarrow \mathcal{Y}$ defined as $W(y|x_1,\dots,x_m) = W_i(y|x_i)$. Let $(W_{i,n})_{n\geq 0}$ be the single-user channel valued process obtained from W_i as in Definition 3.3, and let $(W_n)_{n\geq 0}$ be the MAC-valued process obtained from W as in Definition 4.8. It is easy to see that if W_n is*

δ-easy, then $W_{i,n}$ is δ-easy. This shows that if the sequence $(*_1, \ldots, *_m)$ is MAC-polarizing then $*_i$ is polarizing for each $1 \leq i \leq m$. Moreover, if W_n is (δ, ϵ)-easy, then $W_{i,n}$ is (δ, ϵ)-easy. This implies that $E_{*_1,\ldots,*_m} \leq E_{*_i}$ for each $1 \leq i \leq m$. Therefore, $E_{*_1,\ldots,*_m} \leq \min\{E_{*_1}, \ldots, E_{*_m}\}$.

4.3 Polarization Theory for MACs

Definition 4.11. *Let $W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \longrightarrow \mathcal{Y}$ be an m-user MAC. Let $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$. The single-user channel obtained from W is the channel $W' : \mathcal{X} \longrightarrow \mathcal{Y}$ defined by $W'\big(y\big|(x_1, \ldots, x_m)\big) = W(y|x_1, \ldots, x_m)$ for every $(x_1, \ldots, x_m) \in \mathcal{X}$.*

Notation 4.4. *Let $W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \longrightarrow \mathcal{Y}$ be an m-user MAC. Let $*_1, \ldots, *_m$ be m ergodic operations on $\mathcal{X}_1, \ldots, \mathcal{X}_m$ respectively, and let $* = *_1 \otimes \ldots \otimes *_m$, which is an ergodic operation on $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$. Let $\mathcal{H}$ be a stable partition of $(\mathcal{X}, *)$. $W[\mathcal{H}]$ denotes the single-user channel $W'[\mathcal{H}] : \mathcal{H} \longrightarrow \mathcal{Y}$ (see Definition 3.9), where W' is the single-user channel obtained from W.*

Lemma 4.1. *Let $W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \longrightarrow \mathcal{Y}$ be an m-user MAC. Let $*_1, \ldots, *_m$ be m ergodic operations on $\mathcal{X}_1, \ldots, \mathcal{X}_m$ respectively, and let $* = *_1 \otimes \ldots \otimes *_m$. If there exists $\delta > 0$ and a stable partition $\mathcal{H}$ of $(\mathcal{X}, *)$ such that $\big|I(W) - \log_2 |\mathcal{H}|\big| < \delta$ and $\big|I(W[\mathcal{H}]) - \log_2 |\mathcal{H}|\big| < \delta$, then W is a δ-easy MAC. Moreover, if we also have $P_e(W[\mathcal{H}]) < \epsilon$, then W is a (δ, ϵ)-easy MAC.*

Proof. Let $(\mathcal{H}_i)_{1 \leq i \leq m}$ be the canonical factorization of $\mathcal{H}$ (see Definition 2.20). Let $L = |\mathcal{H}|$. For each $1 \leq i \leq m$ let $L_i = |\mathcal{H}_i|$ and define $\mathcal{S}_i := \{C_i \subset \mathcal{X}_i : \ |C_i| = L_i\}$. We have $L = L_1 \times \cdots \times L_m$ (see Proposition 2.12). Moreover, we have

$$|I(W) - \log_2 L| = \big|I(W) - \log_2 |\mathcal{H}|\big| \leq \delta. \tag{4.1}$$

Now for each $1 \leq i \leq m$ let $H_{i,1}, \ldots, H_{i,L_i}$ be the elements of $\mathcal{H}_i$, and for each $1 \leq j \leq L_i$ let $X_{i,j}$ be a uniform random variable in $H_{i,j}$. We suppose that $X_{i,j}$ is independent from $X_{i',j'}$ for all $(i', j') \neq (i, j)$. Define $\mathcal{B}_i = \{X_{i,1}, \ldots, X_{i,L_i}\}$ which is a random subset of $\mathcal{X}_i$. Clearly, $|\mathcal{B}_i| = L_i$ since each $X_{i,j}$ is drawn from a different element of $\mathcal{H}_i$. Therefore, $\mathcal{B}_i$ takes values in $\mathcal{S}_i$ and $\mathcal{B}_1, \ldots, \mathcal{B}_m$ are independent.

For each $1 \leq i \leq m$ and each $x_i \in \mathcal{X}_i$, let j be the unique index $1 \leq j \leq L_i$ such that $x_i \in H_{i,j}$. Since we are sure that $x_i \notin H_{i,j'}$ for $j' \neq j$, then $x_i \in \mathcal{B}_i$ if and only if $X_{i,j} = x_i$. We have:

$$\begin{aligned} \sum_{C_i \in \mathcal{S}_i} \frac{1}{L_i} P_{\mathcal{B}_i}(C_i) \mathbb{1}_{x_i \in C_i} = \frac{1}{L_i} \mathbb{P}[x_i \in \mathcal{B}_i] &\overset{(a)}{=} \frac{1}{L_i} \mathbb{P}[X_{i,j} = x_i] \\ = \frac{1}{L_i} \cdot \frac{1}{|H_{i,j}|} = \frac{1}{|\mathcal{H}_i|} \cdot \frac{1}{\|\mathcal{H}_i\|} &= \frac{1}{|\mathcal{X}_i|}, \end{aligned} \tag{4.2}$$

where (a) follows from the fact that $x_i \in \mathcal{B}_i$ if and only if $X_{i,j} = x_i$.

Now for each $1 \leq i \leq m$ and each $C_i \subset \mathcal{S}_i$, let $f_{i,C_i} : \{1, \ldots, L_i\} \to C_i$ be a fixed bijection. Let $T_1, \ldots, T_m$ be m independent random variables that are uniform in $\{1, \ldots, L_1\}, \ldots, \{1, \ldots, L_m\}$ respectively, and which are independent of $\mathcal{B}_1, \ldots, \mathcal{B}_m$. For each $1 \leq i \leq m$, let $X_i = f_{i,\mathcal{B}_i}(T_i)$. Send $X_1, \ldots, X_m$ through the MAC W and let Y be the output. The MAC $T_1, \ldots, T_m \longrightarrow (Y, \mathcal{B}_1, \ldots, \mathcal{B}_m)$ is equivalent to the

MAC $W_{\mathcal{B}_1,\dots,\mathcal{B}_m}$ (see Definition 4.6). Our aim now is to show that $I(W_{\mathcal{B}_1,\dots,\mathcal{B}_m}) = I(T_1,\dots,T_m;Y,\mathcal{B}_1,\dots,\mathcal{B}_m) > \log_2 L - \delta$, which will imply that W is δ-easy (see Definition 4.6).

We have

$$I(T_1,\dots,T_m;Y,\mathcal{B}_1,\dots,\mathcal{B}_m) = H(T_1,\dots,T_m) - H(T_1,\dots,T_m|Y,\mathcal{B}_1,\dots,\mathcal{B}_m).$$

Now since $H(T_1,\dots,T_m) = H(T_1)+\cdots+H(T_m) = \log_2 L_1 + \dots + \log_2 L_m = \log_2 L$, it is sufficient to show that $H(T|Y,\mathcal{B}) < \delta$, where $T = (T_1,\dots,T_m)$ and $\mathcal{B} = (\mathcal{B}_1,\dots,\mathcal{B}_m) \in \mathcal{S}_1 \times \cdots \times \mathcal{S}_m$.

Now for each $1 \le i \le m$ and each $x_i \in \mathcal{X}_i$, we have:

$$\begin{aligned} P_{X_i}(x_i) = \mathbb{P}[f_{i,\mathcal{B}_i}(T_i) = x_i] &\overset{(a)}{=} \sum_{C_i \in \mathcal{S}_i:\, x_i \in C_i} \mathbb{P}[f_{i,C_i}(T_i) = x_i] P_{\mathcal{B}_i}(C_i) \\ &\overset{(b)}{=} \sum_{C_i \in \mathcal{S}_i:\, x_i \in C_i} \frac{1}{L_i} P_{\mathcal{B}_i}(C_i) = \sum_{C_i \in \mathcal{S}_i} \frac{1}{L_i} P_{\mathcal{B}_i}(C_i) \mathbb{1}_{x_i \in C_i} \overset{(c)}{=} \frac{1}{|\mathcal{X}_i|}, \end{aligned}$$

where (a) follows from the fact that $f_{i,C_i}(T_i) \in C_i$ and so if $x_i \notin C_i$ then there is a probability of zero to have $f_{i,C_i}(T_i) = x_i$. (b) follows from the fact that T_i is uniform in $\{1,\dots,L_i\}$ and f_{i,C_i} is a bijection from $\{1,\dots,L_i\}$ to C_i which imply that $f_{i,C_i}(T_i)$ is uniform in C_i and so $\mathbb{P}[f_{i,C_i}(T_i) = x_i] = \frac{1}{|C_i|} = \frac{1}{L_i}$. (c) follows from Equation (4.2). Therefore, $X := (X_1,\dots,X_m)$ is uniform in $\mathcal{X}$ since $X_1,\dots,X_m$ are independent and uniform in $\mathcal{X}_1,\dots,\mathcal{X}_m$ respectively. This means that

$$\begin{aligned} I(W[\mathcal{H}]) &= I(\mathrm{Proj}_{\mathcal{H}}(X);Y) = H(\mathrm{Proj}_{\mathcal{H}}(X)) - H(\mathrm{Proj}_{\mathcal{H}}(X)|Y) \\ &= \log_2 |\mathcal{H}| - H(\mathrm{Proj}_{\mathcal{H}}(X)|Y). \end{aligned}$$

Moreover, we have $\big|I(W[\mathcal{H}]) - \log_2|\mathcal{H}|\big| < \delta$ by hypothesis. We conclude that

$$H(\mathrm{Proj}_{\mathcal{H}}(X)|Y) < \delta. \tag{4.3}$$

For each $1 \le i \le m$, let $\mathcal{S}_{\mathcal{H}_i} = \big\{\{x_1,\dots,x_{L_i}\} :\ x_j \in H_{i,j},\ \forall 1 \le j \le L_i\big\}$ be the set of sections of $\mathcal{H}_i$ (see Definition 2.21). By construction, $\mathcal{B}_i$ takes values in $\mathcal{S}_{\mathcal{H}_i}$. Now define

$$\mathcal{S}_{\mathcal{H}} = \{C_1 \times \cdots \times C_m :\ C_1 \in \mathcal{S}_{\mathcal{H}_1},\dots,C_m \in \mathcal{S}_{\mathcal{H}_m}\}.$$

For each $C = C_1 \times \cdots \times C_m \in \mathcal{S}_{\mathcal{H}}$, define $f_C : \{1,\dots,L_1\} \times \cdots \times \{1,\dots,L_m\} \to \mathcal{H}$ as

$$f_C(t_1,\dots,t_m) = \mathrm{Proj}_{\mathcal{H}}\big(f_{1,C_1}(t_1),\dots,f_{m,C_m}(t_m)\big),$$

Since $C_1,\dots,C_m$ are sections of $\mathcal{H}_1,\dots,\mathcal{H}_m$ respectively, $C = C_1 \times \cdots \times C_m$ is a section of $\mathcal{H}$ (see Proposition 2.12). Therefore, for every $H \in \mathcal{H}$, there exists a unique $x = (x_1,\dots,x_m) \in C$ such that $H = \mathrm{Proj}_{\mathcal{H}}(x)$. This implies that there exist unique $t_1 \in \{1,\dots,L_1\}$, $\dots$,$t_m \in \{1,\dots,L_m\}$ such that $f_C(t_1,\dots,t_m) = H$. Therefore, f_C is a bijection from $\{1,\dots,L_1\} \times \cdots \times \{1,\dots,L_m\}$ to $\mathcal{H}$.

Now since f_C is a bijection for every $C \in \mathcal{S}_{\mathcal{H}}$ and since $\mathcal{B}_1 \times \cdots \times \mathcal{B}_m$ takes values in $\mathcal{S}_{\mathcal{H}}$, we have

$$\begin{aligned} H(T|Y,\mathcal{B}) &= H\big(f_{\mathcal{B}_1\times\cdots\times\mathcal{B}_m}(T)\big|Y,\mathcal{B}\big) \\ &= H\big(\mathrm{Proj}_{\mathcal{H}}\big(f_{1,\mathcal{B}_1}(T_1),\dots,f_{m,\mathcal{B}_m}(T_m)\big)\big|Y,\mathcal{B}\big) \\ &= H\big(\mathrm{Proj}_{\mathcal{H}}(X_1,\dots,X_m)\big|Y,\mathcal{B}\big) = H(\mathrm{Proj}_{\mathcal{H}}(X)|Y,\mathcal{B}) \\ &\le H(\mathrm{Proj}_{\mathcal{H}}(X)|Y) \overset{(a)}{<} \delta \end{aligned}$$

as required, where (a) follows from (4.3). We conclude that W is δ-easy.

Now suppose that we also have $P_e(W[\mathcal{H}]) < \epsilon$. Consider the following decoder for the MAC $W_{\mathcal{B}} = W_{\mathcal{B}_1,\ldots,\mathcal{B}_m}$:

- Compute an estimate $\hat{H}$ of $\mathrm{Proj}_{\mathcal{H}}(X)$ using the ML decoder of the channel $W[\mathcal{H}]$.
- Compute $\hat{T} = f^{-1}_{\mathcal{B}_1\times\cdots\times\mathcal{B}_m}(\hat{H})$.

The probability of error of this decoder is:

$$\begin{aligned}\mathbb{P}[\hat{T} \neq T] &= \mathbb{P}[\hat{H} \neq f_{\mathcal{B}_1\times\cdots\times\mathcal{B}_m}(T)] = \mathbb{P}\big[\hat{H} \neq \mathrm{Proj}_{\mathcal{H}}\big(f_{1,\mathcal{B}_1}(T_1),\ldots,f_{m,\mathcal{B}_m}(T_m)\big)\big]\\ &= \mathbb{P}[\hat{H} \neq \mathrm{Proj}_{\mathcal{H}}(X_1,\ldots,X_m)] = \mathbb{P}[\hat{H} \neq \mathrm{Proj}_{\mathcal{H}}(X)] = P_e(W[\mathcal{H}]) < \epsilon.\end{aligned}$$

Now since the ML decoder of $W_{\mathcal{B}}$ minimizes the probability of error, we conclude that $P_e(W_{\mathcal{B}}) < \epsilon$. Therefore, W is a (δ,ϵ)-easy MAC. □

Theorem 4.2. *Let $*_1,\ldots,*_m$ be m binary operations on $\mathcal{X}_1,\ldots,\mathcal{X}_m$ respectively. The sequence $(*_1,\ldots,*_m)$ is MAC-polarizing if and only if $*_1,\ldots,*_m$ are polarizing.*

Proof. Suppose that $(*_1,\ldots,*_m)$ is MAC-polarizing. By Remark 4.2, $*_1,\ldots,*_m$ are polarizing.

Conversely, suppose that $*_1,\ldots,*_m$ are polarizing. Theorem 3.2 implies that $*_1,\ldots,*_m$ are uniformity-preserving and $/^{*_1},\ldots,/^{*_m}$ are strongly ergodic. Now Theorem 2.5 implies that the binary operation $/^{*_1}\otimes\ldots\otimes/^{*_m}$ is strongly ergodic. By noticing that $/^{*_1\otimes\cdots\otimes *_m} = /^{*_1}\otimes\ldots\otimes/^{*_m}$, we conclude that $/^*$ is strongly ergodic, where $* = *_1\otimes\ldots\otimes *_m$.

Now let $W : \mathcal{X}_1\times\cdots\times\mathcal{X}_m \longrightarrow \mathcal{Y}$ be an m-user MAC. Let $\mathcal{X} = \mathcal{X}_1\times\cdots\times\mathcal{X}_m$ and let $W' : \mathcal{X} \longrightarrow \mathcal{Y}$ be the single-user channel obtained from W (see Definition 4.11).

For each $n > 0$ and each $s \in \{-,+\}^n$, let W'^s be obtained from W' using the operation $*$ (see Definition 3.2), and let W^s be obtained from W using the operations $*_1,\ldots,*_m$ (see Definition 4.7). Now since $/^*$ is strongly ergodic, then by Corollary 3.1, for every $\delta > 0$ we have:

$$\lim_{n\to\infty}\frac{1}{2^n}\Big|\Big\{s\in\{-,+\}^n :\ \exists\mathcal{H}_s \textit{ a stable partition of } (\mathcal{X},/^*),\\ \big|I(W'^s) - \log_2|\mathcal{H}_s|\big| < \delta, \big|I(W'^s[\mathcal{H}_s]) - \log_2|\mathcal{H}_s|\big| < \delta\Big\}\Big| = 1.$$

It is easy to see that W'^s is the single-user channel obtained from W^s. Therefore, $I(W^s) = I(W'^s)$ and $I(W^s[\mathcal{H}]) = I(W'^s[\mathcal{H}])$ (by definition). Therefore,

$$\lim_{n\to\infty}\frac{1}{2^n}\Big|\Big\{s\in\{-,+\}^n : \exists\mathcal{H}_s \textit{ a stable partition of } (\mathcal{X},/^*),\\ \big|I(W^s) - \log_2|\mathcal{H}_s|\big| < \delta, \big|I(W^s[\mathcal{H}_s]) - \log_2|\mathcal{H}_s|\big| < \delta\Big\}\Big| = 1.$$

Now Lemma 4.1, applied to $/^{*_1},\ldots,/^{*_m}$, implies that:

$$\lim_{n\to\infty}\frac{1}{2^n}\big|\big\{s\in\{-,+\}^n :\ W^s \text{ is } \delta\text{-easy}\big\}\big| = 1.$$

Therefore, $(*_1, \ldots, *_m)$ satisfies the polarization property of Definition 4.9. On the other hand, since $*_1, \ldots, *_m$ are uniformity-preserving, Remark 4.1 implies that $(*_1, \ldots, *_m)$ satisfies the conservation property of Definition 4.9. We conclude that $(*_1, \ldots, *_m)$ is MAC-polarizing. □

4.4 MAC-Polar Code Construction

Let $\mathcal{X}_1, \ldots, \mathcal{X}_m$ be m finite sets, and let $*_1, \ldots, *_m$ be m binary operations on $\mathcal{X}_1, \ldots, \mathcal{X}_m$, respectively. Assume that $(*_1, \ldots, *_m)$ is a MAC-polarizing sequence of binary operations of exponent[4] $E_{*_1,\ldots,*_m} > 0$. Fix an m-user MAC W with input alphabets $\mathcal{X}_1, \ldots, \mathcal{X}_m$ and output alphabet $\mathcal{Y}$. Choose $0 < \delta < 1$ and $0 < \beta < \beta' < E_{*_1,\ldots,*_m}$, and let $n_0 \geq 0$ be such that for every $n \geq n_0$, we have

$$2^n 2^{-2^{\beta' n}} < 2^{-2^{\beta n}} \text{ and } \frac{1}{2^n}|E_n| > 1 - \frac{\delta}{2\log_2(|\mathcal{X}_1| \times \cdots \times |\mathcal{X}_m|)},$$

where

$$E_n = \left\{ s \in \{-,+\}^n : W^s \text{ is } (\tfrac{\delta}{2}, 2^{-2^{\beta' n}})\text{-}easy \right\}.$$

Such an integer exists because $(*_1, \ldots, *_m)$ is MAC-polarizing and $\beta' < E_{*_1,\ldots,*_m}$ (see Definition 4.10). For every $s \in E_n$, W^s is $(\frac{\delta}{2}, 2^{-2^{\beta' n}})$-easy, hence there exist m integers $L_1^s \leq |\mathcal{X}_1|, \ldots, L_m^s \leq |\mathcal{X}_m|$, and m independent random codes $\mathcal{B}_1^s, \ldots, \mathcal{B}_m^s$ taking values in the sets $\mathcal{S}_1^s = \{C_1 \subset \mathcal{X}_1 : |C_1| = L_1^s\}$, $\ldots$, $\mathcal{S}_m^s = \{C_m \subset \mathcal{X}_m : |C_m| = L_m^s\}$ respectively, which satisfy the following:

- $|I(W^s) - \log_2 L^s| < \frac{\delta}{2}$, where $L^s = L_1^s \times \cdots \times L_m^s$.
- For every $1 \leq i \leq m$ and every $x_i \in \mathcal{X}_i$, we have

$$\sum_{C_i \in \mathcal{S}_i^s} \frac{1}{L_i^s} P_{\mathcal{B}_i^s}(C_i) \mathbb{1}_{x_i \in C_i} = \frac{1}{|\mathcal{X}_i|}. \tag{4.4}$$

- If for each $1 \leq i \leq m$ and each $C_i \in \mathcal{S}_i^s$ we fix a bijection $f_{i,C_i}^s : \{1, ..., L_i^s\} \to C_i$, then $I(W^s{}_{\mathcal{B}_1^s,\ldots,\mathcal{B}_m^s}) > \log_2 L^s - \frac{\delta}{2}$ and $P_e(W^s{}_{\mathcal{B}_1^s,\ldots,\mathcal{B}_m^s}) < 2^{-2^{\beta' n}}$, where

$$W^s{}_{\mathcal{B}_1^s,\ldots,\mathcal{B}_m^s} : \{1, ..., L_1^s\} \times \cdots \times \{1, ..., L_m^s\} \longrightarrow \mathcal{Y}^s \times \mathcal{S}_1^s \times \cdots \times \mathcal{S}_m^s$$

  is the MAC defined as:

$$W^s{}_{\mathcal{B}_1^s,\ldots,\mathcal{B}_m^s}(y, C_1, \ldots, C_m | a_1, \ldots, a_m) = W^s(y | f_{1,C_1}^s(a_1), \ldots, f_{m,C_m}^s(a_m)) . \prod_{i=1}^m P_{\mathcal{B}_i^s}(C_i).$$

  Note that $\mathcal{Y}^s$ denotes the output alphabet of W^s. In the rest of this section, we assume that the bijections $(f_{i,C_i}^s)_{1 \leq i \leq m, s \in E_n, C_i \in \mathcal{S}_i^s}$ are fixed and known to the m users and the receiver.

[4] As we will see in Chapter 5, not every MAC-polarizing sequence of binary operations has a strictly positive exponent. In this section, we assume that $(*_1, \ldots, *_m)$ is a MAC-polarizing sequence of binary operations which satisfies $E_{*_1,\ldots,*_m} > 0$.

A MAC-polar code is constructed as follows:

- If $s \notin E_n$ and $1 \leq i \leq m$, let U_i^s be a frozen symbol in $\mathcal{X}_i$, i.e., we suppose that the receiver knows U_i^s.
- If $s \in E_n$ and $1 \leq i \leq m$, let C_i^s be a frozen code of blocklength 1 and rate $\log_2 L_i^s$ (i.e., the code C_i^s is chosen from $\mathcal{S}_i^s$ and it is known to the receiver). Let $\tilde{U}_i^s$ be a random variable that is uniformly distributed in $\{1, \ldots, L_i^s\}$ and let $U_i^s = f_{i,C_i^s}^s(\tilde{U}_i^s)$.
- After computing U_i^s for every $s \in \{-,+\}^n$, the i^{th} user applies n polarization steps on the sequence $(U_i^s)_{s\in\{-,+\}^n}$ to obtain another sequence of 2^n symbols $(U_{i,s})_{s\in\{-,+\}^n}$, which will be transmitted through 2^n independent copies of the MAC W (see Section 4.4.1).

Since we have a freedom in the choice of the frozen symbols $(U_i^s)_{1\leq i\leq m, s\notin E_n}$ and the frozen codes $(C_i^s)_{1\leq i\leq m, s\in E_n}$, we can assume that these symbols and codes are randomly generated as follows:

- If $s \notin E_n$ and $1 \leq i \leq m$, we assume that U_i^s is chosen uniformly from $\mathcal{X}_i$.
- If $s \in E_n$ and $1 \leq i \leq m$, we assume that C_i^s is a random code taking values in $\mathcal{S}_i^s$ according to the distribution of $\mathcal{B}_i^s$. Equation (4.4) implies that $U_i^s = f_{i,C_i^s}^s(\tilde{U}_i^s)$ is uniformly distributed in $\mathcal{X}_i$.

Furthermore, we assume that the random variables $(U_i^s)_{1\leq i\leq m, s\notin E_n}$, $(\tilde{U}_i^s)_{1\leq i\leq m, s\in E_n}$ and $(C_i^s)_{1\leq i\leq m, s\in E_n}$ are independent.

4.4.1 Encoding

We associate the set $S_n := \{-,+\}^n$ with the strict total order $<$ that we define as $(s_1, \ldots, s_n) < (s'_1, \ldots, s'_n)$ if and only if $s_j = -, s'_j = +$ for some $j \in \{1, \ldots, n\}$ and $s_h = s'_h$ for all $j < h \leq n$.

Let $1 \leq i \leq m$. For every $u_i = (u_i^s)_{s\in S_n} \in \mathcal{X}^{S_n}$, every $0 \leq n' \leq n$ and every $(s', s'') \in S_{n'} \times S_{n-n'}$, define $\mathcal{E}_{i,s'}^{s''}(u_i) \in \mathcal{X}_i$ recursively on $0 \leq n' \leq n$ as follows:

- $\mathcal{E}_{i,\emptyset}^{s}(u_i) = u_i^s$ if $n' = 0$ and $s \in S_n$.
- $\mathcal{E}_{i,(s',-)}^{s''}(u_i) = \mathcal{E}_{i,s'}^{(s'',-)}(u_i) *_i \mathcal{E}_{i,s'}^{(s'',+)}(u_i)$ if $n' > 0$, $s' \in S_{n'-1}$ and $s'' \in S_{n-n'}$.
- $\mathcal{E}_{i,(s',+)}^{s''}(u_i) = \mathcal{E}_{i,s'}^{(s'',+)}(u_i)$ if $n' > 0$, $s' \in S_{n'-1}$ and $s'' \in S_{n-n'}$.

For every $s \in S_n$, we write $\mathcal{E}_{i,\emptyset}^{s}(u_i)$ as $\mathcal{E}_i^s(u_i)$ and $\mathcal{E}_{i,s}^{\emptyset}(u_i)$ as $\mathcal{E}_{i,s}(u_i)$.

Let $\{W_s\}_{s\in S_n}$ be a set of 2^n independent copies of the MAC W. W_s should not be confused with W^s: W_s is a copy of the MAC W whereas W^s is a synthetic MAC obtained from W as before.

Let $(U_i^s)_{s\in S_n} = (f_{i,C_i^s}^s(\tilde{U}_i^s))_{s\in S_n}$ be the sequence of 2^n independent random variables that were defined above. For every $0 \leq n' \leq n$, $s' \in S_{n'}$ and $s'' \in S_{n-n'}$, define $U_{i,s'}^{s''} = \mathcal{E}_{i,s'}^{s''}\big((U_i^s)_{s\in S_n}\big)$. We have:

- $U_{i,\emptyset}^{s} = U_i^s$ if $n' = 0$ and $s \in \{-,+\}^n$.

- $U^{s''}_{i,(s',-)} = U^{(s'',+)}_{i,s'} *_i U^{(s'',-)}_{i,s'}$ if $n' > 0$, $s' \in \{-,+\}^{n'-1}$ and $s'' \in \{-,+\}^{n-n'}$.
- $U^{s''}_{i,(s',+)} = U^{(s'',+)}_{i,s'}$ if $n' > 0$, $s' \in \{-,+\}^{n'-1}$ and $s'' \in \{-,+\}^{n-n'}$.

For every $s \in S_n$, let $U_{i,s} = U^{\emptyset}_{i,s}$. Since $(*_1, \ldots, *_m)$ is MAC-polarizing, $*_1, \ldots, *_m$ are uniformity-preserving. This implies that $(U_{i,s})_{s \in S_n}$ are independent and uniformly distributed in $\mathcal{X}_i$.

It is easy to see that the complexity of the encoding algorithm is $O(N \log N)$, where $N = 2^n$ is the blocklength of the MAC-polar code.

For every $s \in S_n$, the i^{th} user sends $U_{i,s}$ through the MAC W_s. Let Y_s be the output of the MAC W_s, and let $Y = \{Y_s\}_{s \in S_n}$. We can prove by backward induction on n' that for every $s'' \in S_{n-n'}$, the MAC $(U^{s''}_{1,s'}, \ldots, U^{s''}_{m,s'}) \longrightarrow \big(\{Y_s\}_{s \text{ has } s' \text{ as a prefix}}, \{U^r_{i,s'}\}_{1 \leq i \leq m, r < s''}\big)$ is equivalent to the MAC $W^{s''}$ for every $0 \leq n' \leq n$, $s' \in S_{n'}$ and $s'' \in S_{n-n'}$. In particular, the MAC $(U^s_1, \ldots, U^s_m) \longrightarrow \big(Y, \{U^r_i\}_{1 \leq i \leq m, r < s}\big)$ is equivalent to the MAC W^s for every $s \in S_n$. This implies that the MAC

$$(\tilde{U}^s_1, \ldots, \tilde{U}^s_m) \longrightarrow \big(Y, \{U^r_i\}_{1 \leq i \leq m, r < s}, \{C^s_i\}_{1 \leq i \leq m}\big)$$

is equivalent to $W^s{}_{\mathcal{B}^s_1, \ldots, \mathcal{B}^s_m}$ for every $s \in E_n$.

Figure 4.1 is an illustration of a MAC-polar code construction for $n = 1$ (i.e., the blocklength is $N = 2^1 = 2$).

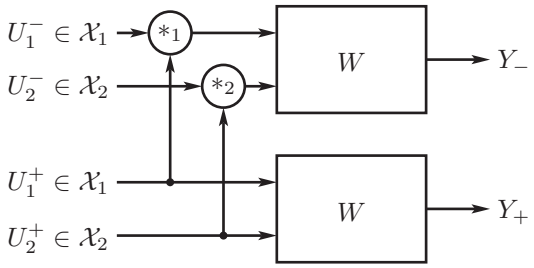

Figure 4.1 – One polarization step.

4.4.2 Decoding

If $s \notin E_n$, there is nothing to decode because the receiver knows $(U^s_i)_{1 \leq i \leq m}$. Suppose now that $s \in E_n$. If we know $\{U^r_i\}_{1 \leq i \leq m, r < s}$ then we can estimate $(\tilde{U}^s_i)_{1 \leq i \leq m}$ from $\big(Y, \{U^r_i\}_{1 \leq i \leq m, r < s}, \{C^s_i\}_{1 \leq i \leq m}\big)$ using the maximum-likelihood decoder of the MAC $W^s{}_{\mathcal{B}^s_1, \ldots, \mathcal{B}^s_m}$. After that, for every $1 \leq i \leq m$, we can obtain an estimate of U^s_i by applying f^s_{i,C^s_i} on the estimate of $\tilde{U}^s_i$.

This motivates us to consider the following successive cancellation decoder:

- If $s \notin E_n$, the receiver computes $\hat{U}^s_i = U^s_i$ for every $1 \leq i \leq m$.
- If $s \in E_n$, the receiver first computes

$$(\hat{\tilde{U}}^s_i)_{1 \leq i \leq m} = \mathcal{D}_s(Y, \{\hat{U}^r_i\}_{1 \leq i \leq m, r < s}, \{C^s_i\}_{1 \leq i \leq m})),$$

where $\mathcal{D}_s$ is the ML decoder of $W^s{}_{\mathcal{B}_1^s,\ldots,\mathcal{B}_m^s}$. The receiver then computes $\hat{U}_i^s = f_{i,C_i^s}^s(\hat{\tilde{U}}_i)$ for every $1 \leq i \leq m$.

The symbols $(U_i^s)_{1\leq i\leq m, s\in S_n}$ are successively decoded according to the total order $<$ of S_n. By using essentially the same method that Arıkan used for binary-input channels [2], the successive cancellation decoder can be implemented with a complexity of $O(N \log N)$.

4.4.3 Performance of MAC-Polar Codes

We have

$$\begin{aligned}
\big\{(\hat{U}_i^s)_{1\leq i\leq m} = (U_i^s)_{1\leq i\leq m},\ \forall s \in S_n\big\} \ &\Leftrightarrow\ \big\{(\hat{U}_i^s)_{1\leq i\leq m} = (U_i^s)_{1\leq i\leq m},\ \forall s \in E_n\big\} \\
&\Leftrightarrow\ \Big\{\big(f_{i,C_i^s}^s(\hat{\tilde{U}}_i^s)\big)_{1\leq i\leq m} = \big(f_{i,C_i^s}^s(\tilde{U}_i^s)\big)_{1\leq i\leq m},\ \forall s \in E_n\Big\} \\
&\Leftrightarrow\ \Big\{(\hat{\tilde{U}}_i^s)_{1\leq i\leq m} = (\tilde{U}_i^s)_{1\leq i\leq m},\ \forall s \in E_n\Big\} \\
&\Leftrightarrow\ \Big\{\mathcal{D}_s(Y, \{\hat{U}_i^r\}_{1\leq i\leq m, r<s}, \{C_i^s\}_{1\leq i\leq m})) = (\tilde{U}_i^s)_{1\leq i\leq m},\ \forall s \in E_n\Big\} \\
&\Leftrightarrow\ \Big\{\mathcal{D}_s(Y, \{U_i^r\}_{1\leq i\leq m, r<s}, \{C_i^s\}_{1\leq i\leq m})) = (\tilde{U}_i^s)_{1\leq i\leq m},\ \forall s \in E_n\Big\}.
\end{aligned}$$

Therefore, the probability of error of the above successive cancellation decoder is upper bounded as

$$\begin{aligned}
&\sum_{s\in E_n} \mathbb{P}\big(\mathcal{D}_s(Y, \{U_i^r\}_{1\leq i\leq m, r<s}, \{C_i^s\}_{1\leq i\leq m})) \neq (\tilde{U}_i^s)_{1\leq i\leq m}\big) \\
&\qquad = \sum_{s\in E_n} P_e(W^s{}_{\mathcal{B}_1^s,\ldots,\mathcal{B}_m^s}) \overset{(a)}{\leq} |E_n| 2^{-2^{\beta' n}} \leq 2^n 2^{-2^{\beta' n}} < 2^{-2^{\beta n}},
\end{aligned}$$

where (a) follows from the fact that $W^s{}_{\mathcal{B}_1^s,\ldots,\mathcal{B}_m^s}$ is $(\frac{\delta}{2}, 2^{-2^{\beta' n}})$-easy.

This upper bound was calculated on average over the random choice of the frozen symbols $(U_i^s)_{1\leq i\leq m, s\notin E_n}$ and codes $(C_i^s)_{1\leq i\leq m, s\in E_n}$. Therefore, there exists at least one choice of the frozen symbols and codes for which the upper bound of the probability of error still holds.

We should note here that unlike the case of binary-input symmetric memoryless channels where the frozen symbols can be chosen arbitrarily [2], the choice of the frozen symbols $(U_i^s)_{1\leq i\leq m, s\notin E_n}$ and codes $(C_i^s)_{1\leq i\leq m, s\in E_n}$ in our construction of MAC-polar codes cannot be arbitrary. The code designer should make sure that his choice of the frozen symbols and codes does indeed yield the desirable probability of error[5].

The last thing to discuss is the sum-rate of MAC-polar codes. The transmission rate at which the i^{th} user is communicating is $R_i = \dfrac{1}{2^n} \displaystyle\sum_{s\in E_n} \log_2 L_i^s$. Therefore, the

[5]In practice, the code designer can generate the frozen symbols $(U_i^s)_{1\leq i\leq m, s\notin E_n}$ and codes $(C_i^s)_{1\leq i\leq m, s\in E_n}$ randomly, and then runs a numerical simulation to assess the performance of the MAC-coding scheme. The code designer repeats this experiment until he finds a suitable choice for the frozen symbols $(U_i^s)_{1\leq i\leq m, s\notin E_n}$ and codes $(C_i^s)_{1\leq i\leq m, s\in E_n}$. With high probability, the code designer is expected to find good frozen symbols and codes after a few trials.

sum-rate is

$$R = \sum_{i=1}^{m} R_i = \frac{1}{2^n} \sum_{i=1}^{m} \sum_{s \in E_n} \log_2 L_i^s = \frac{1}{2^n} \sum_{s \in E_n} \log_2 L^s.$$

On the other hand, we have $\left| I(W^s) - \log_2 L^s \right| < \frac{\delta}{2}$ for all $s \in E_n$. We conclude that:

$$\begin{aligned} I(W) &\overset{(a)}{=} \frac{1}{2^n} \sum_{s \in \{-,+\}^n} I(W^s) = \frac{1}{2^n} \sum_{s \in E_n} I(W^s) + \frac{1}{2^n} \sum_{s \in E_n^c} I(W^s) \\ &< \frac{1}{2^n} \sum_{s \in E_n} \left(\log_2 L^s + \frac{\delta}{2} \right) + \frac{1}{2^n} |E_n^c| \log_2(|\mathcal{X}_1| \times \cdots \times |\mathcal{X}_m|) \\ &< R + \frac{1}{2^n} |E_n| \frac{\delta}{2} + \frac{\delta}{2 \log_2(|\mathcal{X}_1| \times \cdots \times |\mathcal{X}_m|)} \log_2(|\mathcal{X}_1| \times \cdots \times |\mathcal{X}_m|) \\ &\leq R + \frac{\delta}{2} + \frac{\delta}{2} = R + \delta, \end{aligned}$$

where (a) follows from the conservation property of MAC-polarizing sequences of binary operations.

To this end we have shown the following proposition, which is the main result of this section:

Proposition 4.1. *Let $(*_1, \ldots, *_m)$ be a MAC-polarizing sequence of binary operations on the sets $\mathcal{X}_1, \ldots, \mathcal{X}_m$. If $E_{*_1,\ldots,*_m} > 0$, then for every MAC W with input alphabets $\mathcal{X}_1, \ldots, \mathcal{X}_m$, every $\beta < E_{*_1,\ldots,*_m}$ and every $\delta > 0$, there exists $n_0 = n_0(W, \beta, \delta, *_1, \ldots, *_m) > 0$ such that for every $n \geq n_0$, there exists a MAC-polar code of blocklength $N = 2^n$ and of sum-rate at least $I(W) - \delta$ such that the probability of error of the successive cancellation decoder is at most 2^{-N^β}.*

4.5 A Special MAC-Polar Code Construction

If we have $|\mathcal{X}_k| = p_1^{r_1} p_2^{r_2} \ldots p_{n_k}^{r_{n_k}}$, where $p_1, \ldots, p_{n_k}$ are prime numbers, we can assume that $\mathcal{X}_k = \mathbb{F}_{p_1}^{r_1} \mathbb{F}_{p_2}^{r_2} \ldots \mathbb{F}_{p_{n_k}}^{r_{n_k}}$, where $\mathbb{F}_p$ denotes the Galois field of size p. This means that we can replace the k^{th} user by $r_1 + r_2 + \cdots + r_{n_k}$ virtual users such that r_1 virtual users have $\mathbb{F}_{p_1}$ as input alphabet, r_2 virtual users have $\mathbb{F}_{p_2}$ as input alphabet, and so on.

Therefore, we can assume without loss of generality that $\mathcal{X}_k = \mathbb{F}_{q_k}$ for every $1 \leq k \leq m$, where q_k is a prime number. In this section, we consider the polarization transformation of Definition 4.7, where for every $1 \leq k \leq m$, the binary operation that is used for the k^{th} user is the addition modulo q_k.

Let $p_1, p_2, \ldots, p_l$ be the distinct primes that appear in the sequence $q_1, \ldots, q_m$, and for each $1 \leq i \leq l$, let m_i be the number of times p_i appears in the sequence $q_1, \ldots, q_m$. We adopt two notations to indicate the users and their inputs:

- The first notation is the usual one: We have an index k taking value in $\{1, \ldots, m\}$, and the input of the k^{th} user is denoted as $X_k \in \mathbb{F}_{q_k}$.
- In the second notation, the m_i users having their inputs in $\mathbb{F}_{p_i}$ will be indexed by $(i,1), \ldots, (i,j), \ldots, (i,m_i)$, where $1 \leq i \leq l$ and $1 \leq j \leq m_i$. The input of the $(i,j)^{th}$ user is denoted as $X_{i,j} \in \mathbb{F}_{p_i}$. The vector $(X_{i,1}, \ldots, X_{i,m_i}) \in \mathbb{F}_{p_i}^{m_i}$ is denoted as $\vec{X}_i$.

Definition 4.12. *In order to simplify our notation, we will introduce the notion of generalized matrices:*

- *A generalized matrix* $A = (A_1, \ldots, A_l) \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i \times r_i}$ *is a collection of* l *matrices.* $\mathbb{F}_{p_i}^{m_i \times r_i}$ *denotes the set of* $m_i \times r_i$ *matrices with coefficients in* $\mathbb{F}_{p_i}$.
- *If* $r_i = 0$ *in* $A = (A_1, \ldots, A_l) \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i \times r_i}$, *we write* $A_i = \emptyset$. *In case* $A_i = \emptyset$ *for all* i, *we write* $A = \emptyset$.
- *A generalized vector* $\vec{x} = (\vec{x}_1, \ldots, \vec{x}_l) \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i}$ *is a collection of* l *vectors.*
- *Addition of generalized vectors is defined as component-wise addition.*
- *The transposition of a generalized matrix is obtained by transposing each matrix in it:* $A^T = (A_1^T, \ldots, A_l^T)$.
- *A generalized matrix acts on a generalized vector in a component-wise fashion: If* $A \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i \times r_i}$ *and* $\vec{x} \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i}$, *then* $\vec{y} = A^T \vec{x} \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{r_i}$ *is defined as*

$$\vec{y} = (A_1^T \vec{x}_1, \ldots, A_l^T \vec{x}_l).$$

  We adopt the convention that $\emptyset^T \vec{x}_i = \vec{0}$.
- *A generalized matrix* A *is said to be full rank if and only if each matrix component in it is full rank.*
- *The rank of a generalized matrix* $A \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i \times r_i}$ *is defined as:*

$$\operatorname{rank}(A) = \sum_{i=1}^{l} \operatorname{rank}(A_i).$$

- *The logarithmic rank of a generalized matrix is defined as:*

$$\operatorname{lrank}(A) = \sum_{i=1}^{l} \operatorname{rank}(A_i) \cdot \log_2 p_i.$$

- *If* A *is a generalized matrix satisfying* $A_i \neq \emptyset$ *and* $A_j = \emptyset$ *for all* $j \neq i$, *we say that* A *is an ordinary matrix and we identify* A *with* A_i.

Definition 4.13. *Let* $W : \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i} \longrightarrow \mathcal{Y}$ *be an* m*-user MAC, and let* $A \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i \times r_i}$ *be a full-rank generalized matrix. We define the* $\operatorname{rank}(A)$*-user MAC*

$$W[A] : \prod_{i=1}^{l} \mathbb{F}_{p_i}^{r_i} \longrightarrow \mathcal{Y}$$

as follows:

$$W[A](y|\vec{u}) = \frac{1}{\prod_{i=1}^{l} p_i^{m_i - r_i}} \sum_{\substack{\vec{x} \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i} \\ A^T \vec{x} = \vec{u}}} W(y|\vec{x}).$$

The main result of this section is that as the number of polarization steps becomes large, almost all the synthetic MACs W^s become MACs for which the output is "almost determined" by the action of a generalized matrix A_s on the input:

Theorem 4.3. *Let* $W : \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i} \longrightarrow \mathcal{Y}$ *be an* m*-user MAC. For every* $0 < \delta < 1$*, we have:*

$$\lim_{n\to\infty} \frac{1}{2^n} \Bigg| \Big\{ s \in \{-,+\}^n : \exists A_s \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i \times r_{i,s}},\ A_s \text{ is full rank}, \\ |I(W^s) - \operatorname{lrank}(A_s)| < \delta,\ |I(W^s[A_s]) - \operatorname{lrank}(A_s)| < \delta \Big\} \Bigg| = 1.$$

Proof. Since $G := \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i}$ is an Abelian group, we can view W as a channel from the Abelian group G to $\mathcal{Y}$. From Corollary 3.1, we have:

$$\lim_{n\to\infty} \frac{1}{2^n} \Bigg| \Big\{ s \in \{-,+\}^n : \exists \mathcal{H}_s \text{ a stable partition of } (G, /^+), \\ \big|I(W^s) - \log_2 |\mathcal{H}_s|\big| < \delta, \big|I(W^s[\mathcal{H}_s]) - \log_2 |\mathcal{H}_s|\big| < \delta \Big\} \Bigg| = 1.$$

On the other hand, from the proof of Proposition 2.4 we can see that every stable partition of $(G, /^+)$ is the quotient of G by a (normal[6]) subgroup of $(G, +)$. Therefore,

$$\lim_{n\to\infty} \frac{1}{2^n} \Bigg| \Big\{ s \in \{-,+\}^n :\ \exists H_s \text{ a subgroup of } (G, +), \\ \big|I(W^s) - \log_2 |G/H_s|\big| < \delta, \big|I(W^s[G/H_s]) - \log_2 |G/H_s|\big| < \delta \Big\} \Bigg| = 1.$$

Let $s \in \{-,+\}^n$ be such that there exists a subgroup H_s of G which satisfies:

- $\big|I(W^s) - \log_2 |G/H_s|\big| < \delta$.
- $\big|I(W^s[G/H_s]) - \log_2 |G/H_s|\big| < \delta$.

From the properties of Abelian groups, there exist l integers: $r_{1,s} \leq m_1, \ldots, r_{l,s} \leq m_l$ such that G/H_s is isomorphic to $\prod_{i=1}^{l} \mathbb{F}_{p_i}^{r_{i,s}}$ (Note that $r_{i,s}$ can be zero).

[6]Note that every subgroup of an Abelian group is normal.

Therefore, there exists a surjective homomorphism $f_s : \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i} \longrightarrow \prod_{i=1}^{l} \mathbb{F}_{p_i}^{r_{i,s}}$, such that for every $\vec{x} \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i}$, $f_s(\vec{x})$ can be determined from $\vec{x} \bmod H_s := \mathrm{Proj}_{G/H_s}(\vec{x})$ and vice versa.

For every $1 \leq i \leq l$ and $1 \leq j \leq m_i$, define the vector $\vec{e}^{\,i,j} \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i}$ in such a way that the $(i,j)^{th}$ component is equal to 1, and all the other components are equal to 0. Clearly, the order of $\vec{e}^{\,i,j}$ in the group G is equal to p_i. Define

$$\vec{y}^{\,i,j} = (\vec{y}_1^{\,i,j}, \vec{y}_2^{\,i,j}, \ldots, \vec{y}_l^{\,i,j}) = f_s(\vec{e}^{\,i,j}) \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{r_{i,s}}.$$

If $\vec{y}_{i'}^{\,i,j} \neq \vec{0}$ for some $i' \neq i$, then $p_{i'}$ divides the order of $\vec{y}^{\,i,j}$. But $\vec{y}^{\,i,j} = f_s(\vec{e}^{\,i,j})$, so the order of $\vec{y}^{\,i,j}$ divides the order of $\vec{e}^{\,i,j}$, which is equal to p_i. Therefore, if $\vec{y}_{i'}^{\,i,j} \neq \vec{0}$ for some $i' \neq i$, then $p_{i'}$ divides p_i, which is a contradiction. We conclude that $\vec{y}_{i'}^{\,i,j} = \vec{0}$ for every $i' \neq i$.

Now for every $\vec{x} \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i}$, we have $\vec{x} = \sum_{i=1}^{l} \sum_{j=1}^{m_i} x_{i,j} \vec{e}^{\,i,j}$. Therefore, $f_s(\vec{x}) = \sum_{i=1}^{l} \sum_{j=1}^{m_i} x_{i,j} \vec{y}^{\,i,j}$. Since $\vec{y}_{i'}^{\,i,j} = 0$ for all $i' \neq i$, then $f_s(\vec{x}) = A_s^T \vec{x}$, where $A_s = (A_{1,s}, \ldots, A_{l,s}) \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i \times r_{i,s}}$ is the generalized matrix whose components are given by $A_{i,s} = [\vec{y}_i^{\,i,1}\ \vec{y}_i^{\,i,2}\ \ldots\ \vec{y}_i^{\,i,m_i}]^T$. A_s is full rank since f_s is surjective. Furthermore, we have:

$$\mathrm{lrank}(A_s) = \sum_{i=1}^{l} r_{i,s} \cdot \log_2 p_i = \log_2 \Big(\prod_{i=1}^{l} p_i^{r_{i,s}} \Big) = \log_2 |G/H_s|.$$

Now recall that for every $\vec{x} \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i}$, $A_s^T \vec{x} = f_s(\vec{x})$ can be determined from $\vec{x} \bmod H_s$ and vice versa. We conclude that $W^s[G/H_s]$ is equivalent to $W^s[A_s]$. Therefore,

$$\lim_{n\to\infty} \frac{1}{2^n} \Bigg| \Big\{ s \in \{-,+\}^n : \exists A_s \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i \times r_{i,s}},\ A_s \text{ is full rank}, \\ |I(W^s) - \mathrm{lrank}(A_s)| < \delta,\ |I(W^s[A_s]) - \mathrm{lrank}(A_s)| < \delta \Big\} \Bigg| = 1.$$

□

4.5.1 MAC-Polar Code Construction

Choose $0 < \delta < 1$, and let n be an integer such that

$$\frac{1}{2^n}|E_n| > 1 - \frac{\delta}{2\sum_{i=1}^{l} m_i \log_2 p_i},$$

where

$$E_n = \Bigg\{ s \in \{-,+\}^n : \exists A_s \in \prod_{i=1}^{l} \mathbb{F}_{p_i}^{m_i \times r_{i,s}}, A_s \text{ is full rank}, \\ |I(W^s) - \operatorname{lrank}(A_s)| < \frac{\delta}{2},\ |I(W^s[A_s]) - \operatorname{lrank}(A_s)| < \frac{\delta}{2} \Bigg\}.$$

Such an integer exists due to Theorem 4.3.

For every $s \in E_n$, fix a generalized matrix $A_s = (A_{1,s}, \ldots, A_{l,s})$ that satisfies the conditions in E_n. Furthermore, for every $1 \le i \le l$, fix a set of $r_{i,s}$ indices

$$J_{i,s} = \{j_1, \ldots j_{r_{i,s}}\} \subset \{1, \ldots, m_i\}$$

such that the corresponding rows of $A_{i,s}$ are linearly independent.

Now for every $s \in \{-,+\}^n$, $1 \le i \le l$ and $1 \le j \le m_i$, define $F(s,i,j)$ as follows:

$$F(s,i,j) = \begin{cases} 0 & \text{if } s \in E_n \text{ and } j \in J_{i,s}, \\ 1 & \text{otherwise.} \end{cases}$$

$F(s,i,j) = 1$ indicates that the user (i,j) will be frozen in the channel W^s, i.e., no useful information will be sent.

A MAC-polar code is constructed as follows: The user (i,j) sends a symbol $U_{i,j}^s \in \mathbb{F}_{p_i}$ through a MAC that is equivalent to W^s. If $F(s,i,j) = 0$, $U_{i,j}^s$ is an information symbol, and if $F(s,i,j) = 1$, $U_{i,j}^s$ is a frozen symbol. Since we are free to choose any value for the frozen symbols, we will analyze the performance of the MAC-polar code averaged over all the possible choices of the frozen symbols, so we will consider that $U_{i,j}^s$ is a random variable that is uniformly distributed in $\mathbb{F}_{p_i}$ for every $s \in \{-,+\}^n$, $1 \le i \le l$ and $1 \le j \le m_i$. However, the value of $U_{i,j}^s$ will be revealed to the receiver if $F(s,i,j) = 1$, and if $F(s,i,j) = 0$ the receiver has to estimate $U_{i,j}^s$ from the output of the MAC.

Let $s \in \{-,+\}^n$. For every $1 \le i \le l$, we denote $(U_{i,1}^s, \ldots, U_{i,m_i}^s)$ as $\vec{U}_i^s$. Furthermore, we denote $(\vec{U}_1^s, \ldots, \vec{U}_l^s)$ as $\vec{U}^s$.

Encoding

We associate the set $S_n = \{-,+\}^n$ with the same strict total order $<$ that we defined in Section 4.4.1. Let $\{W_s\}_{s \in \{-,+\}^n}$ be a set of 2^n independent copies of the channel W. As in Section 4.4.1, W_s should not be confused with W^s: W_s is a copy of the MAC W, whereas W^s is a synthetic MAC obtained from W as before.

Let $1 \le i \le l$ and $1 \le j \le m_i$. For every $0 \le n' \le n$, $s' \in S_{n'}$ and $s'' \in S_{n-n'}$, define $U_{i,j,s'}^{s''}$ recursively on $0 \le n' \le n$ as follows:

- $U^s_{i,j,\emptyset} = U^s_{i,j}$ if $n' = 0$ and $s \in \{-,+\}^n$.
- $U^{s''}_{i,j,(s',-)} = U^{(s'',+)}_{i,j,s'} + U^{(s'',-)}_{i,j,s'}$ if $n' > 0$, $s' \in \{-,+\}^{n'-1}$ and $s'' \in \{-,+\}^{n-n'}$.
- $U^{s''}_{i,j,(s',+)} = U^{(s'',+)}_{i,j,s'}$ if $n' > 0$, $s' \in \{-,+\}^{n'-1}$ and $s'' \in \{-,+\}^{n-n'}$.

For every $s \in S_n$, let $U_{i,j,s} = U^{\emptyset}_{i,j,s}$. The user (i,j) sends $U_{i,j,s}$ through the MAC W_s for all $s \in \{-,+\}^n$. Let Y_s be the output of the MAC W_s, and let $Y = \{Y_s\}_{s\in\{-,+\}^n}$.

Let $0 \le n' \le n$, $s' \in S_{n'}$ and $s'' \in S_{n-n'}$. For every $1 \le i \le l$, we denote $(U^{s''}_{i,1,s'}, \ldots, U^{s''}_{i,m_i,s'})$ as $\vec{U}^{s''}_{i,s'}$. Furthermore, we denote $(\vec{U}^{s''}_{1,s'}, \ldots, \vec{U}^{s''}_{l,s'})$ as $\vec{U}^{s''}_{s'}$.

We can prove by backward induction on n' that for every $s'' \in S_{n-n'}$, the MAC $\vec{U}^{s''}_{s'} \longrightarrow \left(\{Y_s\}_{s \text{ has } s' \text{ as a prefix}}, \{\vec{U}^r_{s'}\}_{r<s''}\right)$ is equivalent to $W^{s''}$. In particular, the MAC $\vec{U}^s \longrightarrow \left(Y, \{\vec{U}^r\}_{r<s}\right)$ is equivalent to the MAC W^s for every $s \in S_n$.

Decoding

If $s \notin E_n$, there is nothing to decode because $F(s,i,j) = 1$ for all (i,j), i.e., the receiver knows $U^s_{i,j}$ for all (i,j).

Now suppose that $s \in E_n$. If we know $\{\vec{U}_r\}_{r<s}$ then we can estimate $\vec{U}_s$ as follows:

- If $F(s,i,j) = 1$ then we know $U^s_{i,j}$.
- We have $F(s,i,j) = 0$ for $r_{i,s}$ values of j corresponding to $r_{i,s}$ linearly independent rows of $A_{i,s}$. Therefore, if we know $A^T_{i,s}\vec{U}^s_i$, we can recover $U^s_{i,j}$ for all the indices j satisfying $F(s,i,j) = 0$.
- Since $A^T_s\vec{U}^s \longrightarrow \left(Y, \{\vec{U}^r\}_{r<s}\right)$ is equivalent to $W^s[A_s]$, we can estimate $A^T_s\vec{U}^s$ using the maximum likelihood decoder of the MAC $W^s[A_s]$.
- Let $\mathcal{D}_s(Y, \{\vec{U}^r\}_{r<s})$ be the estimate of $\vec{U}^s$ obtained from $(Y, \{\vec{U}^r\}_{r<s})$ by the above procedure.

This motivates the following successive cancellation decoder:

- $\hat{\vec{U}}^s = \vec{U}^s$ if $s \notin E_n$.
- $\hat{\vec{U}}^s = \mathcal{D}_s(Y, \{\hat{\vec{U}}^r\}_{r<s})$ if $s \in E_n$.

Performance of MAC-polar codes

As we will see in Chapter 5, the exponent of a sequence of quasigroup operations is equal to $\frac{1}{2}$. This means that the probability of error of the MAC-polar codes that we constructed in this section decays faster[7] than 2^{-N^β} for any $\beta < \frac{1}{2}$.

By changing our choice of the indices in $J_{i,s}$, we can achieve all the portion of the dominant face of the symmetric-capacity region that is achievable by MAC-polar codes. This portion of the dominant face that is achievable by MAC-polar codes might be strictly smaller than the dominant face. In such case, we say that we have a loss in the symmetric-capacity region. We study this loss in Chapter 6.

[7] In order to ensure that the probability of error of the MAC-polar code decays faster than 2^{-N^β}, we should add the condition $P_e(W^s[A_s]) < 2^{-2^{\beta' n}}$ to the definition of E_n, where $\beta < \beta' < \frac{1}{2}$.

5 Error Exponents

There are two common approaches to assess the performance of a family of codes:

- The error exponent approach: We fix a rate R and study the decay of the probability of error as the blocklength N increases. It is known that the decay of the probability of error of random codes is exponential in N (see e.g., [24]). Unlike random codes, the probability of error of polar codes does not decay exponentially in the blocklength. Arıkan and Telatar showed that the probability of error of polar codes for binary-input channels decays exponentially in the *square root* of the blocklength. This behavior was also shown for the polar codes of [4, 5, 6, 7], and the MAC-polar codes of [8, 9].

- The scaling exponent approach: We fix a probability of error and study the growth of the blocklength N as the gap to capacity $C(W)-R$ decreases towards zero. It was shown in [25, 26, 27, 28] that the blocklength of optimal codes grows as $O\left(\frac{1}{(C(W)-R)^2}\right)$. The scaling exponent of polar codes in the case of binary-input channels was studied in [29, 30, 31, 32].

In this chapter[1], we study the error exponents of polarizing binary operations and MAC-polarizing sequences of binary operations. In Section 5.1, we define the Bhattacharyya parameter of a channel, which is a very useful tool for the study of error exponents of polar codes. In Section 5.2, we show that the exponent of a polarizing binary operation cannot exceed $\frac{1}{2}$. We also provide a sufficient condition for a polarizing operation to have a zero exponent. In Section 5.3, we prove that the exponent of a quasigroup operation is exactly $\frac{1}{2}$. In Section 5.4, we show that the exponent of a MAC-polarizing sequence of binary operations is upper bounded by the exponent of the product of all the binary operations that are present in the sequence, which in turn is upper bounded by the exponent of every binary operation in the sequence. Furthermore, we prove that the exponent of a sequence of quasigroup operations is exactly $\frac{1}{2}$.

[1]The material of this chapter is based on [17, 18, 20].

5.1 The Bhattacharyya Parameter

Definition 5.1. *Let W be a channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. For every $x, x' \in \mathcal{X}$, we define the channel $W_{x,x'} : \{0,1\} \longrightarrow \mathcal{Y}$ as follows:*

$$W_{x,x'}(y|b) = \begin{cases} W(y|x) \text{ if } b = 0, \\ W(y|x') \text{ if } b = 1. \end{cases}$$

The Battacharyya parameter between x and x' of the channel W *is the Bhattacharyya parameter of the channel $W_{x,x'}$:*

$$Z(W_{x,x'}) := \sum_{y \in \mathcal{Y}} \sqrt{W_{x,x'}(y|0) W_{x,x'}(y|1)} = \sum_{y \in \mathcal{Y}} \sqrt{W(y|x) W(y|x')}.$$

It is easy to see that $0 \leq Z(W_{x,x'}) \leq 1$ for every $x, x' \in \mathcal{X}$. Moreover, if $x = x'$ we have $Z(W_{x,x'}) = Z(W_{x,x}) = 1$.

If $|\mathcal{X}| \geq 2$, the Battacharyya parameter of the channel W *is defined as:*

$$Z(W) := \frac{1}{|\mathcal{X}|(|\mathcal{X}| - 1)} \sum_{\substack{(x,x') \in \mathcal{X} \times \mathcal{X} \\ x \neq x'}} Z(W_{x,x'}).$$

We can easily see that $0 \leq Z(W) \leq 1$.

Proposition 5.1. *The Bhattacharyya parameter of a channel $W : \mathcal{X} \longrightarrow \mathcal{Y}$ has the following properties:*

1. $Z(W)^2 \leq 1 - \dfrac{I(W)}{\log_2 |\mathcal{X}|}$.
2. $I(W) \geq \log_2 \dfrac{|\mathcal{X}|}{1 + (|\mathcal{X}| - 1) Z(W)}$.
3. $\dfrac{1}{4} Z(W)^2 \leq P_e(W) \leq (|\mathcal{X}| - 1) Z(W)$, *where $P_e(W)$ is the probability of error of the maximum likelihood decoder of W for uniformly distributed input.*

Proof. Inequalities 1) and 2) are proved in Proposition 3.3 of [33], and the upper bound of 3) is shown in Proposition 3.2 of [33]. It remains to show the lower bound of 3).

Let $D_W^{\mathrm{ML}} : \mathcal{Y} \to \mathcal{X}$ be the ML decoder of the channel $W : \mathcal{X} \longrightarrow \mathcal{Y}$. I.e., for every $y \in \mathcal{Y}$, $D_W^{\mathrm{ML}}(y) = \arg\max_{x \in \mathcal{X}} W(y|x)$. For every $x \in \mathcal{X}$, let $P_{e,x}(W)$ be the probability of error of D_W^{ML} given that x was sent through W. Clearly, $P_e(W) = \dfrac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} P_{e,x}(W)$.

Now fix $x, x' \in \mathcal{X}$ such that $x \neq x'$ and define $P_{e,x,x'}(W) := \frac{1}{2} P_{e,x}(W) + \frac{1}{2} P_{e,x'}(W)$. Consider the channel $W_{x,x'} : \{0,1\} \longrightarrow \mathcal{Y}$. We can use D_W^{ML} to construct a decoder for $W_{x,x'}$ as follows:

- If $D_W^{\mathrm{ML}}(y) = x$, the decoder output is 0.

- If $D_W^{\mathrm{ML}}(y) = x'$, the decoder output is 1.
- If $D_W^{\mathrm{ML}}(y) \notin \{x, x'\}$ for $y \in \mathcal{Y}$, we consider that an error has occurred.

It is easy to see that the probability of error of the constructed decoder (assuming uniform binary input to the channel $W_{x,x'}$) is equal to $\frac{1}{2}P_{e,x}(W) + \frac{1}{2}P_{e,x'}(W) = P_{e,x,x'}(W)$. But since the ML decoder of $W_{x,x'}$ has the minimal probability of error among all decoders, we conclude that:

$$\begin{aligned} P_{e,x,x'}(W) \geq P_e(W_{x,x'}) &= \frac{1}{2}\sum_{y\in\mathcal{Y}} \min\big\{W_{x,x'}(y|0), W_{x,x'}(y|1)\big\} \\ &= \frac{1}{2}\sum_{y\in\mathcal{Y}} \min\big\{W(y|x), W(y|x')\big\}. \end{aligned} \tag{5.1}$$

On the other hand, we have:

$$\begin{aligned} Z(W_{x,x'}) &= \sum_{y\in\mathcal{Y}} \sqrt{W(y|x)W(y|x')} \\ &= \sum_{y\in\mathcal{Y}} \sqrt{\Big(\min\big\{W(y|x), W(y|x')\big\}\Big)\Big(\max\big\{W(y|x), W(y|x')\big\}\Big)} \\ &\overset{(a)}{\leq} \Big(\sum_{y\in\mathcal{Y}} \min\big\{W(y|x), W(y|x')\big\}\Big)^{1/2}\Big(\sum_{y\in\mathcal{Y}} \max\big\{W(y|x), W(y|x')\big\}\Big)^{1/2} \\ &\overset{(b)}{\leq} \sqrt{2P_{e,x,x'}(W)}\Big(\sum_{y\in\mathcal{Y}} W(y|x) + W(y|x')\Big)^{1/2} = \sqrt{2P_{e,x,x'}(W)}.\sqrt{2} \\ &= 2\sqrt{P_{e,x,x'}(W)}, \end{aligned}$$

where (a) follows from the Cauchy-Schwartz inequality. (b) follows from (5.1) and from the fact that $\max\big\{W(y|x), W(y|x')\big\} \leq W(y|x) + W(y|x')$. We conclude that:

$$P_{e,x,x'}(W) \geq \frac{1}{4}Z(W_{x,x'})^2. \tag{5.2}$$

Now since $P_e(W) = \dfrac{1}{|\mathcal{X}|}\displaystyle\sum_{x\in\mathcal{X}} P_{e,x}(W)$, we have:

$$\begin{aligned} \sum_{\substack{x,x'\in\mathcal{X}\\ x\neq x'}} P_{e,x,x'}(W) &= \frac{1}{2}\Big(\sum_{\substack{x,x'\in\mathcal{X}\\ x\neq x'}} P_{e,x}(W)\Big) + \frac{1}{2}\Big(\sum_{\substack{x,x'\in\mathcal{X}\\ x\neq x'}} P_{e,x'}(W)\Big) \\ &= \frac{1}{2}\Big(\sum_{x\in\mathcal{X}} (|\mathcal{X}|-1)P_{e,x}(W)\Big) + \frac{1}{2}\Big(\sum_{x'\in\mathcal{X}} (|\mathcal{X}|-1)P_{e,x'}(W)\Big) \\ &= \frac{1}{2}(|\mathcal{X}|-1)|\mathcal{X}|P_e(W) + \frac{1}{2}(|\mathcal{X}|-1)|\mathcal{X}|P_e(W) \\ &= (|\mathcal{X}|-1)|\mathcal{X}|P_e(W). \end{aligned}$$

Therefore,

$$P_e(W) = \frac{1}{(|\mathcal{X}|-1)|\mathcal{X}|} \sum_{\substack{x,x' \in \mathcal{X} \\ x \neq x'}} P_{e,x,x'}(W) \overset{(a)}{\geq} \frac{1}{(|\mathcal{X}|-1)|\mathcal{X}|} \sum_{\substack{x,x' \in \mathcal{X} \\ x \neq x'}} \frac{1}{4} Z(W_{x,x'})^2$$
$$\overset{(b)}{\geq} \frac{1}{4} \Big(\frac{1}{(|\mathcal{X}|-1)|\mathcal{X}|} \sum_{\substack{x,x' \in \mathcal{X} \\ x \neq x'}} Z(W_{x,x'}) \Big)^2 = \frac{1}{4} Z(W)^2,$$

where (a) follows from (5.2) and (b) follows from the convexity of the mapping $t \to t^2$. □

Remark 5.1. *Proposition 5.1 shows that $Z(W)$ measures the ability of the receiver to reliably decode the output and correctly estimate the input:*

- *If $Z(W)$ is low, the inequality $P_e(W) \leq (|\mathcal{X}|-1)Z(W)$ implies that $P_e(W)$ is also low and the receiver can determine the input from the output with high probability. This is also expressed by inequality 2) of Proposition 5.1: If $Z(W)$ is close to 0, $I(W)$ is close to $\log_2 |\mathcal{X}|$.*

- *If $Z(W)$ is close to 1, inequality 1) of Proposition 5.1 implies that $I(W)$ is close to 0, which means that the input and the output are "almost" independent and so it is not possible to recover the input reliably. This is also expressed by the inequality $P_e(W) \geq \frac{1}{4} Z(W)^2$: If $Z(W)$ is high, $P_e(W)$ cannot be too low.*

Since $W_{x,x'}$ is the binary input channel obtained by sending either x or x' through W, $Z(W_{x,x'})$ can be interpreted as a measure of the ability of the receiver to distinguish between x and x': If $Z(W_{x,x'}) \approx 0$, the receiver can reliably distinguish between x and x' and if $Z(W_{x,x'}) \approx 1$, the receiver cannot distinguish between x and x'.

5.2 Exponent of a Polarizing Operation

In this section, we study the exponent of polarizing operations.

Notation 5.1. *Let $x, x' \in \mathcal{X}$ and let $s \in \{-,+\}^n$. Throughout this section, $W^s_{x,x'}$ denotes $(W^s)_{x,x'}$. The channel $W^s_{x,x'}$ should not be confused with $(W_{x,x'})^s$ which is not defined unless a binary operation on $\{0,1\}$ is specified.*

Lemma 5.1. *For every $u_1, u_1', v \in \mathcal{X}$, we have $Z(W^-_{u_1,u_1'}) \geq \frac{1}{|\mathcal{X}|} Z(W_{u_1 * v, u_1' * v})$.*

Proof.

$$\begin{aligned}
Z(W^-_{u_1,u'_1}) &= \sum_{y_1,y_2\in\mathcal{Y}} \sqrt{W^-(y_1,y_2|u_1)W^-(y_1,y_2|u'_1)} \\
&= \sum_{y_1,y_2\in\mathcal{Y}} \sqrt{\sum_{u_2,u'_2\in\mathcal{X}} \frac{1}{|\mathcal{X}|^2} W(y_1|u_1*u_2)W(y_2|u_2)W(y_1|u'_1*u'_2)W(y_2|u'_2)} \\
&\geq \frac{1}{|\mathcal{X}|} \sum_{y_1,y_2\in\mathcal{Y}} \sqrt{W(y_1|u_1*v)W(y_2|v)W(y_1|u'_1*v)W(y_2|v)} \\
&= \frac{1}{|\mathcal{X}|} \sum_{y_1,y_2\in\mathcal{Y}} W(y_2|v)\sqrt{W(y_1|u_1*v)W(y_1|u'_1*v)} \\
&= \frac{1}{|\mathcal{X}|} \sum_{y_1\in\mathcal{Y}} \sqrt{W(y_1|u_1*v)W(y_1|u'_1*v)} = \frac{1}{|\mathcal{X}|} Z(W_{u_1*v,u'_1*v}).
\end{aligned}$$

□

Lemma 5.2. *For every $u_2, u'_2 \in \mathcal{X}$, we have*

$$Z(W^+_{u_2,u'_2}) = \frac{1}{|\mathcal{X}|} \sum_{u_1\in\mathcal{X}} Z(W_{u_1*u_2,u_1*u'_2}) Z(W_{u_2,u'_2}).$$

Proof.

$$\begin{aligned}
Z(W^+_{u_2,u'_2}) &= \sum_{y_1,y_2\in\mathcal{Y}} \sum_{u_1\in\mathcal{X}} \sqrt{W^+(y_1,y_2,u_1|u_2)W^+(y_1,y_2,u_1|u'_2)} \\
&= \sum_{y_1,y_2\in\mathcal{Y}} \sum_{u_1\in\mathcal{X}} \sqrt{\frac{1}{|\mathcal{X}|^2} W(y_1|u_1*u_2)W(y_2|u_2)W(y_1|u_1*u'_2)W(y_2|u'_2)} \\
&= \frac{1}{|\mathcal{X}|} \sum_{u_1\in\mathcal{X}} \sum_{y_1,y_2\in\mathcal{Y}} \sqrt{W(y_1|u_1*u_2)W(y_1|u_1*u'_2)}\sqrt{W(y_2|u_2)W(y_2|u'_2)} \\
&= \frac{1}{|\mathcal{X}|} \sum_{u_1\in\mathcal{X}} Z(W_{u_1*u_2,u_1*u'_2}) Z(W_{u_2,u'_2}).
\end{aligned}$$

□

Notation 5.2. *If W is a channel with input alphabet $\mathcal{X}$. We denote* $\max\limits_{\substack{x,x'\in\mathcal{X}\\ x\neq x'}} Z(W_{x,x'})$ *and* $\min\limits_{\substack{x,x'\in\mathcal{X}\\ x\neq x'}} Z(W_{x,x'})$ *by $Z_{\max}(W)$ and $Z_{\min}(W)$ respectively. Note that we can also express $Z_{\min}(W)$ as* $\min\limits_{x,x'\in\mathcal{X}} Z(W_{x,x'})$ *since $Z_{\min}(W) \leq 1$ and $Z_{x,x}(W) = 1$ for every $x \in \mathcal{X}$.*

Proposition 5.2. *Let $*$ be a polarizing operation on $\mathcal{X}$, where $|\mathcal{X}| \geq 2$. If for every $u_2, u'_2 \in \mathcal{X}$ there exists $u_1 \in \mathcal{X}$ such that $u_1 * u_2 = u_1 * u'_2$, then $E_* = 0$.*

Proof. Let $\beta > 0$ and $0 < \beta' < \beta$. Clearly, $\frac{1}{4}\left(2^{-2^{\beta' n}}\right)^2 > 2^{-2^{\beta n}}$ for n large enough. We have:

- For every $u_2, u_2' \in \mathcal{X}$ satisfying $u_2 \neq u_2'$, let $u_1 \in \mathcal{X}$ be such that $u_1 * u_2 = u_1 * u_2'$. Lemma 5.2 implies that $Z(W^+_{u_2,u_2'}) \geq \frac{1}{|\mathcal{X}|} Z(W_{u_1*u_2,u_1*u_2'}) Z(W_{u_2,u_2'}) = \frac{1}{|\mathcal{X}|} Z(W_{u_2,u_2'})$ since $Z(W_{u_1*u_2,u_1*u_2'}) = 1$. Therefore,

$$Z_{\max}(W^+) = \max_{\substack{x,x' \in \mathcal{X} \\ x \neq x'}} Z(W^+_{x,x'}) \geq \frac{1}{|\mathcal{X}|} \max_{\substack{x,x' \in \mathcal{X} \\ x \neq x'}} Z(W_{x,x'}) = \frac{1}{|\mathcal{X}|} Z_{\max}(W).$$

- By fixing $v \in \mathcal{X}$, Lemma 5.1 implies that

$$\begin{aligned} Z_{\max}(W^-) &= \max_{\substack{x,x' \in \mathcal{X} \\ x \neq x'}} Z(W^-_{x,x'}) \geq \frac{1}{|\mathcal{X}|} \max_{\substack{x,x' \in \mathcal{X} \\ x \neq x'}} Z(W_{x*v,x'*v}) \\ &\overset{(a)}{=} \frac{1}{|\mathcal{X}|} \max_{\substack{x,x' \in \mathcal{X} \\ x \neq x'}} Z(W_{x,x'}) = \frac{1}{|\mathcal{X}|} Z_{\max}(W), \end{aligned}$$

  where (a) follows from the fact that $*$ is uniformity-preserving, which implies that

$$\{(x * v, x' * v) : \ x, x' \in \mathcal{X}, \ x \neq x'\} = \{(x, x') : \ x, x' \in \mathcal{X}, \ x \neq x'\}.$$

By induction on $n > 0$, we conclude that for every $s \in \{-,+\}^n$ we have:

$$Z_{\max}(W^s) \geq \frac{1}{|\mathcal{X}|^n} Z_{\max}(W) = \frac{1}{2^{n \log_2 |\mathcal{X}|}} Z_{\max}(W).$$

If $Z(W) > 0$ we have $Z_{\max}(W) > 0$, and

$$Z(W^s) \geq \frac{1}{|\mathcal{X}|(|\mathcal{X}|-1)} Z_{\max}(W^s) \geq \frac{Z_{\max}(W)}{|\mathcal{X}|(|\mathcal{X}|-1) \cdot (2^n)^{\log_2 |\mathcal{X}|}},$$

which means that the decay of $Z(W^s)$ in terms of the blocklength 2^n can be at best polynomial. Therefore, for n large enough we have $Z(W^s) > 2^{-2^{\beta' n}}$ for every $s \in \{-,+\}^n$.

Now let $\delta = \frac{1}{3}\log_2 |\mathcal{X}| - \frac{1}{3}\log_2(|\mathcal{X}|-1) > 0$ and let W be any channel satisfying $\log_2 |\mathcal{X}| - \delta < I(W) < \log_2 |\mathcal{X}|$ (we can easily construct such a channel). Since $I(W) < \log_2 |\mathcal{X}|$, Proposition 5.1 implies that we have $Z(W) > 0$. Let W_n be the process introduced in Definition 3.3. Since $*$ is polarizing, we have $\mathbb{P}[W_n \text{ is } \delta\text{-easy}] > \frac{3}{4}$ (i.e., $\frac{1}{2^n}|\{s \in \{-,+\}^n : \ W^s \text{ is } \delta\text{-easy}\}| > \frac{3}{4}$) for n large enough. On the other hand, since $*$ satisfies the conservation property, we have $\mathbb{E}[I(W_n)] = \frac{1}{2^n} \sum_{s \in \{-,+\}^n} I(W^s) = I(W) > \log_2 |\mathcal{X}| - \delta$. Therefore, we must have $\mathbb{P}\big[I(W_n) > \log_2 |\mathcal{X}| - 2\delta\big] > \frac{1}{2}$ and so for n large enough, we have

$$\mathbb{P}\big[I(W_n) > \log_2 |\mathcal{X}| - 2\delta \text{ and } W_n \text{ is } \delta\text{-easy}\big] > \frac{1}{4}.$$

Now suppose $s \in \{-,+\}^n$ is such that W^s is δ-easy and $I(W^s) > \log_2 |\mathcal{X}| - 2\delta$, and let L and $\mathcal{B}$ be as in Definition 3.1. We have $I(W^s) - \log_2(|\mathcal{X}|-1) > 3\delta - 2\delta = \delta$

and so the only possible value for L is $|\mathcal{X}|$. But since the only subset of $\mathcal{X}$ of size $|\mathcal{X}|$ is $\mathcal{X}$, we have $\mathcal{B} = \mathcal{X}$ with probability 1. Therefore, $W_{\mathcal{B}}^s$ is equivalent to W^s which means that $Z(W_{\mathcal{B}}^s) = Z(W^s) > 2^{-2^{\beta' n}}$. Now Proposition 5.1 implies that $P_e(W_{\mathcal{B}}^s) > \frac{1}{4}\left(2^{-2^{\beta' n}}\right)^2 > 2^{-2^{\beta n}}$ and so W^s is not $(\delta, 2^{-\beta n})$-easy. Thus, $\mathbb{P}\big[W_n \text{ is } (\delta, 2^{-2^{\beta n}})\text{-easy}\big] < \frac{3}{4}$ for n large enough.

We conclude that no exponent $\beta > 0$ is $*$-achievable. Therefore, $E_* = 0$. □

Remark 5.2. *Consider the following uniformity-preserving operation:*

$*$	0	1	2	3
0	3	3	3	3
1	0	1	0	0
2	1	0	1	1
3	2	2	2	2

It is easy to see that $/^*$ is strongly ergodic and so $*$ is polarizing. Moreover, $*$ satisfies the property of Proposition 5.2, hence it has a zero exponent. This shows that the exponent of a polarizing operation can be as low as 0.

The following lemma will be used to show that $E_* \leq \frac{1}{2}$ for every polarizing operation $*$.

Lemma 5.3. *Let $*$ be a uniformity-preserving operation on $\mathcal{X}$ and let W be a channel with input alphabet $\mathcal{X}$. For every $n > 0$ and every $s \in \{-,+\}^n$, we have $Z_{\min}(W^s) \geq \left(\frac{Z_{\min}(W)}{|\mathcal{X}|}\right)^{(|s|^- + 1)2^{|s|^+}}$, where $|s|^-$ (resp. $|s|^+$) is the number of $-$ signs (resp. $+$ signs) in the sequence s.*

Proof. We will prove the lemma by induction on $n > 0$. If $n = 1$, then either $s = -$ or $s = +$. If $s = -$, let $v \in \mathcal{X}$. We have:

$$\begin{aligned} Z_{\min}(W^s) = Z_{\min}(W^-) &= \min_{u_1, u_1' \in \mathcal{X}} Z(W^-_{u_1, u_1'}) \\ &\overset{(a)}{\geq} \min_{u_1, u_1' \in \mathcal{X}} \frac{1}{|\mathcal{X}|} Z(W_{u_1 * v, u_1' * v}) \overset{(b)}{\geq} \left(\frac{Z_{\min}(W)}{|\mathcal{X}|}\right)^{(|s|^- + 1)2^{|s|^+}}, \end{aligned} \tag{5.3}$$

where (a) follows from Lemma 5.1 and (b) follows from the fact that $(|s|^- + 1)2^{|s|^+} = 2$ since $|s|^- = 1$ and $|s|^+ = 0$ when $s = -$.

If $s = +$, we have:

$$\begin{aligned} Z_{\min}(W^s) = Z_{\min}(W^+) &= \min_{u_2, u_2' \in \mathcal{X}} Z(W^+_{u_2, u_2'}) \\ &\overset{(a)}{\geq} \min_{u_2, u_2' \in \mathcal{X}} \frac{1}{|\mathcal{X}|} \sum_{u_1 \in \mathcal{X}} Z(W_{u_1 * u_2, u_1 * u_2'}) Z(W_{u_2, u_2'}) \\ &\geq Z_{\min}(W)^2 \overset{(b)}{\geq} \left(\frac{Z_{\min}(W)}{|\mathcal{X}|}\right)^{(|s|^- + 1)2^{|s|^+}}, \end{aligned} \tag{5.4}$$

where (a) follows from Lemma 5.2 and (b) follows from the fact that $(|s|^- + 1)2^{|s|^+} = 2$ since $|s|^- = 0$ and $|s|^+ = 1$ when $s = +$. Therefore, the lemma is true for $n = 1$. Now let $n > 1$ and suppose that it is true for $n - 1$. Let $s = (s', s_n) \in \{-, +\}^n$, where $s' \in \{-, +\}^{n-1}$ and $s_n \in \{-, +\}$. From the induction hypothesis, we have $Z_{\min}(W^{s'}) \geq \left(\frac{Z_{\min}(W)}{|\mathcal{X}|}\right)^{(|s'|^- + 1)2^{|s'|^+}}$.

If $s_n = -$, we can apply (5.3) on $W^{s'}$ to get:

$$\begin{aligned} Z_{\min}(W^s) &\geq \frac{1}{|\mathcal{X}|} Z_{\min}(W^{s'}) \geq \frac{1}{|\mathcal{X}|} \left(\frac{Z_{\min}(W)}{|\mathcal{X}|}\right)^{(|s'|^- + 1)2^{|s'|^+}} \\ &\geq \left(\frac{Z_{\min}(W)}{|\mathcal{X}|}\right)^{1 + (|s'|^- + 1)2^{|s'|^+}} \geq \left(\frac{Z_{\min}(W)}{|\mathcal{X}|}\right)^{(|s'|^- + 2)2^{|s'|^+}} \\ &= \left(\frac{Z_{\min}(W)}{|\mathcal{X}|}\right)^{(|s|^- + 1)2^{|s|^+}}. \end{aligned}$$

If $s_n = +$, we can apply (5.4) on $W^{s'}$ to get:

$$\begin{aligned} Z_{\min}(W^s) &\geq Z_{\min}(W^{s'})^2 \geq \left(\left(\frac{Z_{\min}(W)}{|\mathcal{X}|}\right)^{(|s'|^- + 1)2^{|s'|^+}}\right)^2 \\ &= \left(\frac{Z_{\min}(W)}{|\mathcal{X}|}\right)^{2(|s'|^- + 1)2^{|s'|^+}} = \left(\frac{Z_{\min}(W)}{|\mathcal{X}|}\right)^{(|s'|^- + 1)2^{|s'|^+ + 1}} \\ &= \left(\frac{Z_{\min}(W)}{|\mathcal{X}|}\right)^{(|s|^- + 1)2^{|s|^+}}. \end{aligned}$$

We conclude that the lemma is true for every $n > 0$. □

Proposition 5.3. *If $*$ is a polarizing operation on $\mathcal{X}$, where $|\mathcal{X}| \geq 2$, then $E_* \leq \frac{1}{2}$.*

Proof. Let $\beta > \frac{1}{2}$, and let $\frac{1}{2} < \beta' < \beta$. Let $\epsilon > 0$ be such that $(1 - \epsilon)\log_2 |\mathcal{X}| > \log_2 |\mathcal{X}| - \delta$, where $\delta = \frac{1}{3}|\mathcal{X}| - \frac{1}{3}(|\mathcal{X}| - 1)$. Let $e \notin \mathcal{X}$ and consider the channel $W : \mathcal{X} \longrightarrow \mathcal{X} \cup \{e\}$ defined as follows:

$$W(y|x) = \begin{cases} 1 - \epsilon & \text{if } y = x, \\ \epsilon & \text{if } y = e, \\ 0 & \text{otherwise.} \end{cases}$$

We have $I(W) = (1 - \epsilon)\log_2 |\mathcal{X}| > \log_2 |\mathcal{X}| - \delta$ and $Z(W_{x,x'}) = \epsilon$ for every $x, x' \in \mathcal{X}$ such that $x \neq x'$, and thus $Z_{\min}(W) = \epsilon$. We have the following:

- Since $\beta' > \frac{1}{2}$, the law of large numbers implies that

$$\frac{1}{2^n} \left|\left\{s \in \{-, +\}^n : |s|^+ \leq \beta' n\right\}\right|$$

converges to 1 as n goes to infinity. Therefore, for n large enough, we have $\frac{1}{2^n}|B_n| > \frac{7}{8}$, where

$$B_n = \left\{s \in \{-, +\}^n : |s|^+ \leq \beta' n\right\}.$$

- Since $\sum_{s\in\{-,+\}^n} I(W^s) = 2^n I(W) > 2^n(\log_2|\mathcal{X}| - \delta)$, we must have $\frac{1}{2^n}|C_n| > \frac{1}{2}$ where
$$C_n = \big\{s \in \{-,+\}^n : I(W^s) > \log_2|\mathcal{X}| - 2\delta\big\}.$$
- Since $*$ is polarizing, we have $\frac{1}{2^n}|D_n| > \frac{7}{8}$ for n large enough, where
$$D_n = \big\{s \in \{-,+\}^n :\ W^s \text{ is } \delta\text{-easy}\big\}.$$

We conclude that for n large enough, we have $\frac{1}{2^n}|A_n| > \frac{1}{4}$, where

$$\begin{aligned} A_n &= B_n \cap C_n \cap D_n \\ &= \big\{s \in \{-,+\}^n :\ |s|^+ \le \beta' n,\ W^s \text{ is } \delta\text{-easy and } I(W^s) > \log_2|\mathcal{X}| - 2\delta\big\}. \end{aligned}$$

Now let $s \in A_n$. Let L and $\mathcal{B}$ be as in Definition 3.1. We have $I(W^s) - \log_2(|\mathcal{X}|-1) > 3\delta - 2\delta = \delta$ and so the only possible value for L is $|\mathcal{X}|$, and since the only subset of $\mathcal{X}$ of size $|\mathcal{X}|$ is $\mathcal{X}$, we have $\mathcal{B} = \mathcal{X}$ with probability 1. Therefore, $W^s_{\mathcal{B}}$ is equivalent to W^s. Thus,

$$Z(W^s_{\mathcal{B}}) = Z(W^s) \ge Z_{\min}(W^s) \overset{(a)}{\ge} \left(\frac{Z_{\min}(W)}{|\mathcal{X}|}\right)^{(|s|^-+1)2^{|s|^+}} \overset{(b)}{\ge} \left(\frac{\epsilon}{|\mathcal{X}|}\right)^{(n+1)2^{\beta' n}},$$

where (a) follows from Lemma 5.3 and (b) follows from the fact that $|s|^- \le n$ and $|s|^+ \le \beta' n$ for $s \in A_n$, and from the fact that $Z_{\min}(W) = \epsilon$ which was proved earlier. Now Proposition 5.1 implies that $P_e(W^s_{\mathcal{B}}) \ge \frac{1}{4}\left(\frac{\epsilon}{|\mathcal{X}|}\right)^{2(n+1)2^{\beta' n}}$. On the other hand, since $\beta' < \beta$, we have $\frac{1}{4}\left(\frac{\epsilon}{|\mathcal{X}|}\right)^{2(n+1)2^{\beta' n}} > 2^{-2^{\beta n}}$ for n large enough. Therefore, W^s is not $(\delta, 2^{-\beta n})$-easy if $s \in A_n$ and n is large enough. Let W_n be the process introduced in Definition 3.3. For n large enough, we have

$$\mathbb{P}\big[W_n \text{ is } (\delta, 2^{-2^{\beta n}})\text{-easy}\big] \le 1 - \frac{1}{2^n}|A_n| < 1 - \frac{1}{4} = \frac{3}{4}.$$

We conclude that no exponent $\beta > \frac{1}{2}$ is $*$-achievable. Therefore, $E_* \le \frac{1}{2}$. □

5.3 Exponent of a Quasigroup Operation

Definition 5.2. *Let $(Q,*)$ be a quasigroup with $|Q| \ge 2$, and $\mathcal{Y}$ be an arbitrary set. Let $W : Q \longrightarrow \mathcal{Y}$ be an arbitrary channel, and $\mathcal{H}$ be a stable partition of $(Q, /^*)$. We define the channels $W[\mathcal{H}]^- : \mathcal{H}^{/^*} \longrightarrow \mathcal{Y}\times\mathcal{Y}$ and $W[\mathcal{H}]^+ : \mathcal{H} \longrightarrow \mathcal{Y}\times\mathcal{Y}\times\mathcal{H}^{/^*}$ as:*

$$W[\mathcal{H}]^+(y_1, y_2, H_1|H_2) = \frac{1}{|\mathcal{H}|} W[\mathcal{H}](y_1|H_1 * H_2) W[\mathcal{H}](y_2|H_2),$$

$$W[\mathcal{H}]^-(y_1, y_2|H_1) = \frac{1}{|\mathcal{H}|} \sum_{H_2\in\mathcal{H}} W[\mathcal{H}](y_1|H_1 * H_2) W[\mathcal{H}](y_2|H_2).$$

Lemma 5.4. *$W[\mathcal{H}]^+$ is degraded with respect to $W^+[\mathcal{H}]$, and $W[\mathcal{H}]^-$ is equivalent to $W^-[\mathcal{H}^{/^*}]$.*

Proof. Let $(H_1, H_2, y_1, y_2) \in \mathcal{H}^{/^*} \times \mathcal{H} \times \mathcal{Y} \times \mathcal{Y}$, we have:

$$\begin{aligned}
W[\mathcal{H}]^+(y_1, y_2, H_1|H_2) &= \frac{1}{|\mathcal{H}|} W[\mathcal{H}](y_1|H_1 * H_2) W[\mathcal{H}](y_2|H_2) \\
&= \frac{1}{|Q|.||\mathcal{H}||} \sum_{\substack{x_1 \in Q, \\ \mathrm{Proj}_{\mathcal{H}}(x_1) = H_1 * H_2}} W(y_1|x_1) \sum_{\substack{x_2 \in Q, \\ \mathrm{Proj}_{\mathcal{H}}(x_2) = H_2}} W(y_2|x_2) \\
&= \frac{1}{|Q|.||\mathcal{H}||} \sum_{\substack{x_1 \in Q, \\ \mathrm{Proj}_{\mathcal{H}^{/^*}}(x_1) = H_1}} \sum_{\substack{x_2 \in Q, \\ \mathrm{Proj}_{\mathcal{H}}(x_2) = H_2}} W(y_1|x_1 * x_2) W(y_2|x_2) \\
&= \frac{1}{||\mathcal{H}||} \sum_{\substack{x_1 \in Q, \\ \mathrm{Proj}_{\mathcal{H}^{/^*}}(x_1) = H_1}} \sum_{\substack{x_2 \in Q, \\ \mathrm{Proj}_{\mathcal{H}}(x_2) = H_2}} W^+(y_1, y_2, x_1|x_2) \\
&= \sum_{\substack{x_1 \in Q, \\ \mathrm{Proj}_{\mathcal{H}^{/^*}}(x_1) = H_1}} W^+[\mathcal{H}](y_1, y_2, x_1|H_2).
\end{aligned}$$

Therefore, $W[\mathcal{H}]^+$ is degraded with respect to $W^+[\mathcal{H}]$. Now let $(H_1, y_1, y_2) \in \mathcal{H}^{/^*} \times \mathcal{Y} \times \mathcal{Y}$, we have:

$$\begin{aligned}
W[\mathcal{H}]^-(y_1, y_2|H_1) &= \frac{1}{|\mathcal{H}|} \sum_{H_2 \in \mathcal{H}} W[\mathcal{H}](y_1|H_1 * H_2) W[\mathcal{H}](y_2|H_2) \\
&= \frac{1}{|Q|.||\mathcal{H}||} \sum_{H_2 \in \mathcal{H}} \sum_{\substack{x_1 \in Q, \\ \mathrm{Proj}_{\mathcal{H}}(x_1) = H_1 * H_2}} W(y_1|x_1) \sum_{\substack{x_2 \in Q, \\ \mathrm{Proj}_{\mathcal{H}}(x_2) = H_2}} W(y_2|x_2) \\
&= \frac{1}{|Q|.||\mathcal{H}||} \sum_{H_2 \in \mathcal{H}} \sum_{\substack{x_1 \in Q, \\ \mathrm{Proj}_{\mathcal{H}^{/^*}}(x_1) = H_1}} \sum_{\substack{x_2 \in Q, \\ \mathrm{Proj}_{\mathcal{H}}(x_2) = H_2}} W(y_1|x_1 * x_2) W(y_2|x_2) \\
&= \frac{1}{|Q|.||\mathcal{H}||} \sum_{\substack{x_1 \in Q, \\ \mathrm{Proj}_{\mathcal{H}^{/^*}}(x_1) = H_1}} \sum_{x_2 \in Q} W(y_1|x_1 * x_2) W(y_2|x_2) \\
&= \frac{1}{||\mathcal{H}||} \sum_{\substack{x_1 \in Q, \\ \mathrm{Proj}_{\mathcal{H}^{/^*}}(x_1) = H_1}} W^-(y_1, y_2|x_1) = W^-[\mathcal{H}^{/^*}](y_1, y_2|H_1).
\end{aligned}$$

Therefore, $W[\mathcal{H}]^-$ is equivalent to $W^-[\mathcal{H}^{/^*}]$. □

Definition 5.3. *Let $\mathcal{H}$ be a stable partition of $(Q, /^*)$, we define the stable partitions $\mathcal{H}^-$ and $\mathcal{H}^+$, by $\mathcal{H}^{/^*}$ and $\mathcal{H}$ respectively.*

Lemma 5.5. *Let $(B_n)_{n\geq 0}$ and $(W_n)_{n\geq 0}$ be defined as in definition 3.3. For each stable partition $\mathcal{H}$ of $(Q, /^*)$, we define the stable-partition-valued process $(\mathcal{H}_n)_{n\geq 0}$ by:*

$$\begin{aligned}
&\mathcal{H}_0 := \mathcal{H}, \\
&\mathcal{H}_n := \mathcal{H}_{n-1}^{B_n}, \ \forall n \geq 1.
\end{aligned}$$

Then $I(W_n[\mathcal{H}_n])$ converges almost surely to a number in

$$\mathcal{L}_{\mathcal{H}} := \big\{ \log_2 d : \ d \text{ divides } |\mathcal{H}| \big\}.$$

Proof. Since $W_n[\mathcal{H}_n]^-$ is equivalent to $W_n^-[\mathcal{H}_n^{/^*}]$ and $W_n[\mathcal{H}_n]^+$ is degraded with respect to $W_n^+[\mathcal{H}_n]$ (Lemma 5.4), we have:

$$\begin{aligned}\mathbb{E}\Big(I(W_{n+1}[\mathcal{H}_{n+1}])\Big|W_n\Big) &= \frac{1}{2}I(W_n^-[\mathcal{H}_n^{/^*}]) + \frac{1}{2}I(W_n^+[\mathcal{H}_n]) \\ &\geq \frac{1}{2}I(W_n[\mathcal{H}_n]^-) + \frac{1}{2}I(W_n[\mathcal{H}_n]^+) = I(W_n[\mathcal{H}_n]).\end{aligned}$$

This implies that the process $I(W_n[\mathcal{H}_n])$ is a sub-martingale and so it converges almost surely. Let $\delta > 0$, and define

$$\begin{aligned}D_{n,\delta} := \Big\{ & s \in \{-,+\}^n : \ \exists \mathcal{H}_s \text{ a stable partition of } (Q, /^*), \\ & \Big| I(W^s[\mathcal{H}']) - \log_2 \frac{|\mathcal{H}_s|.||\mathcal{H}_s \wedge \mathcal{H}'||}{||\mathcal{H}'||} \Big| < \delta \text{ for all stable partitions } \mathcal{H}' \text{ of } (Q, /^*) \Big\}.\end{aligned}$$

Theorem 3.1 implies that $\lim_{n\to\infty} \frac{1}{2^n}|D_{n,\delta}| = 1$. It is easy to see that almost surely, for every $\delta > 0$ and for every $n_0 > 0$, there exists $n > n_0$ such that $(B_1, \ldots, B_n) \in D_{l,\delta}$.

Let $(B_n)_{n\geq 0}$ be a realization that satisfies:

- The sequence $\big(I(W_n[\mathcal{H}_n])\big)_{n\geq 0}$ converges to a limit x.
- For every $\delta > 0$ and every $n_0 > 0$, there exists $n > n_0$ such that $(B_1, \ldots, B_n) \in D_{n,\delta}$.

Let $\delta > 0$ and let $n_0 > 0$ be chosen such that $|I(W_n[\mathcal{H}_n]) - x| < \delta$ for every $n > n_0$. Choose $n > n_0$ such that $s = (B_1, \ldots, B_n) \in D_{n,\delta}$. By taking $\mathcal{H}' = \mathcal{H}_n$ in (5.3), we obtain $\Big|I(W_n[\mathcal{H}_n]) - \log_2 \frac{|\mathcal{H}_s|.||\mathcal{H}_s \wedge \mathcal{H}_n||}{||\mathcal{H}_n||}\Big| < \delta$. Therefore,

$$\Big|x - \log_2 \frac{|\mathcal{H}_n|.||\mathcal{H}_s \wedge \mathcal{H}_n||}{||\mathcal{H}_s||}\Big| < 2\delta.$$

But $|Q| = |\mathcal{H}_s|.||\mathcal{H}_s|| = |\mathcal{H}_n|.||\mathcal{H}_n||$, hence $\Big|x - \log_2 \frac{|\mathcal{H}_s|.||\mathcal{H}_s \wedge \mathcal{H}_n||}{||\mathcal{H}_n||}\Big| < 2\delta$.

By noticing that $\frac{|\mathcal{H}_n|.||\mathcal{H}_s \wedge \mathcal{H}_n||}{||\mathcal{H}_s||}$ divides $|\mathcal{H}_n| = |\mathcal{H}|$, we conclude that

$$\min_{R\in\mathcal{L}_{\mathcal{H}}} |x - R| < 2\delta, \ \forall \delta > 0.$$

Therefore, $x \in \mathcal{L}_{\mathcal{H}}$. □

Lemma 5.6. *Let $(Q, *)$ be a quasigroup satisfying $|Q| \geq 2$, and let $W : Q \longrightarrow \mathcal{Y}$. Let $\mathcal{H}$ be a stable partition of $(Q, /^*)$. For every $0 < \delta < 1$ and every $0 < \beta < \frac{1}{2}$, we have:*

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{2^n}\Big|\Big\{ & s \in \{-,+\}^n : \ \exists \mathcal{H} \text{ a stable partition of } (Q, /^*), \\ & I(W^s[\mathcal{H}]) > \log_2 |\mathcal{H}| - \delta, \ Z(W^s[\mathcal{H}]) \geq 2^{-2^{n\beta}} \Big\}\Big| = 0.\end{aligned}$$

Proof. Let $0 < \delta < 1$ and $0 < \beta < \frac{1}{2}$, and let $\mathcal{H}$ be a stable partition of $(Q, /^*)$. $I(W_n[\mathcal{H}_n])$ converges almost surely to an element in $\mathcal{L}_{\mathcal{H}}$. Due to the relations between the quantities $I(W)$ and $Z(W)$ (see Proposition 5.1), we can see that $Z(W_n[\mathcal{H}_n])$ converges to 0 if and only if $I(W_n[\mathcal{H}_n])$ converges to $\log_2|\mathcal{H}|$, and there is a number $z_0 > 0$ such that $\liminf Z(W_n[H]) > z_0$ whenever $I(W_n[H])$ converges to a number in $\mathcal{L}_{\mathcal{H}}$ other than $\log_2|\mathcal{H}|$. Therefore, we can say that almost surely, we have:

$$\lim Z(W_n[\mathcal{H}_n]) = 0 \;\text{ or }\; \liminf Z(W_n[H]) > z_0$$

$Z(W_n^+[\mathcal{H}_n^+]) \leq Z(W_n[\mathcal{H}_n]^+)$ since $W_n[\mathcal{H}_n]^+$ is degraded with respect to $W_n^+[\mathcal{H}_n^+]$, and $Z(W_n^-[\mathcal{H}_n^-]) = Z(W_n[\mathcal{H}_n]^-)$ since $W_n[\mathcal{H}_n]^-$ and $W_n^-[\mathcal{H}_n^-]$ are equivalent (see Lemma 5.4). On the other hand, from [33, Lemma 3.5], we have:

- $Z(W_n[\mathcal{H}_n]^-) \leq \big(|\mathcal{H}|^2 - |\mathcal{H}| + 1\big) Z(W_n[\mathcal{H}_n])$.
- $Z(W_n[\mathcal{H}_n]^+) \leq \big(|\mathcal{H}| - 1\big) Z(W_n[\mathcal{H}_n])^2$.

Therefore, we have $Z(W_n^-[\mathcal{H}_n]) \leq K.Z(W_n[\mathcal{H}_n])$ and $Z(W_n^+[\mathcal{H}_n]) \leq K.Z(W_n[\mathcal{H}_n])^2$, where $K := \big(|\mathcal{H}|^2 - |\mathcal{H}| + 1\big)$. By applying exactly the same techniques that were used to prove [33, Theorem 3.5] we get:

$$\lim_{n\to\infty} \mathbb{P}\Big[I(W_n[\mathcal{H}_n]) > \log_2|\mathcal{H}| - \delta, Z(W_n[\mathcal{H}_n]) \geq 2^{-2^{n\beta}}\Big] = 0.$$

But this is true for every stable partition $\mathcal{H}$ of $(Q, /^*)$. Therefore,

$$\lim_{n\to\infty} \frac{1}{2^n}\Big|\Big\{s \in \{-,+\}^n :\ \exists \mathcal{H} \text{ a stable partition of } (Q, /^*),\\ I(W^s[\mathcal{H}^s]) > \log_2|\mathcal{H}| - \delta,\ Z(W^s[\mathcal{H}^s]) \geq 2^{-2^{n\beta}}\Big\}\Big| = 0.$$

By noticing that for every $s \in \{-,+\}^n$, there exists a stable partition $\mathcal{H}_s$ of $(Q, /^*)$ satisfying $\mathcal{H} = \mathcal{H}_s^s$, we conclude that:

$$\lim_{n\to\infty} \frac{1}{2^n}\Big|\Big\{s \in \{-,+\}^n :\ \exists \mathcal{H} \text{ a stable partition of } (Q, /^*),\\ I(W^s[\mathcal{H}]) > \log_2|\mathcal{H}| - \delta,\ Z(W^s[\mathcal{H}]) \geq 2^{-2^{n\beta}}\Big\}\Big| = 0.$$

□

Theorem 5.1. *The convergence of W_n to projection channels is almost surely fast:*

$$\lim_{n\to\infty} \frac{1}{2^n}\Big|\Big\{s \in \{-,+\}^n : \exists \mathcal{H}_s \text{ a stable partition of } (Q, /^*),\\ \big|I(W^s) - \log_2|\mathcal{H}_s|\big| < \delta, \big|I(W^s[\mathcal{H}_s]) - \log_2|\mathcal{H}_s|\big| < \delta,\ Z(W^s[\mathcal{H}_s]) < 2^{-2^{\beta n}}\Big\}\Big| = 1,$$

for every $0 < \delta < 1$, and every $0 < \beta < \frac{1}{2}$.

Proof. Let $0 < \delta < 1$, and $0 < \beta < \frac{1}{2}$. Define:

$$E_0 = \Big\{ s \in \{-,+\}^n : \exists \mathcal{H} \text{ a stable partition of } (Q, /^*), \\ I(W^s[\mathcal{H}]) > \log_2 |\mathcal{H}| - \delta,\ Z(W^s[\mathcal{H}]) \geq 2^{-2^{\beta n}} \Big\},$$

$$E_1 = \Big\{ s \in \{-,+\}^n : \exists \mathcal{H}_s \text{ a stable partition of } (Q, /^*), \\ \big|I(W^s) - \log_2 |\mathcal{H}_s|\big| < \delta, \big|I(W^s[\mathcal{H}_s]) - \log_2 |\mathcal{H}_s|\big| < \delta \Big\},$$

$$E_2 = \Big\{ s \in \{-,+\}^n : \exists \mathcal{H}_s \text{ a stable partition of } (Q, /^*), \\ \big|I(W^s) - \log_2 |\mathcal{H}_s|\big| < \delta, \big|I(W^s[\mathcal{H}_s]) - \log_2 |\mathcal{H}_s|\big| < \delta,\ Z(W^s[\mathcal{H}_s]) < 2^{-2^{\beta n}} \Big\}.$$

It is easy to see that $E_1 \setminus E_0 \subset E_2$ and $|E_2| \geq |E_1| - |E_0|$. From Corollary 3.1 and Lemma 5.6, we obtain

$$1 \geq \lim_{n\to\infty} \frac{1}{2^n}|E_2| \geq \lim_{n\to\infty} \frac{1}{2^n}\big(|E_1| - |E_0|\big) = 1 - 0 = 1.$$

□

Proposition 5.4. *If $(Q, *)$ is a quasigroup satisfying $|Q| \geq 2$, then $E_* = \frac{1}{2}$.*

Proof. Let $\beta < \beta' < \frac{1}{2}$. Let $W : Q \longrightarrow \mathcal{Y}$ be a channel. From Theorem 5.1, we have:

$$\lim_{n\to\infty} \frac{1}{2^n} \Big| \Big\{ s \in \{-,+\}^n :\ \exists \mathcal{H}_s \text{ a stable partition of } (\mathcal{X}, /^*), \\ \big|I(W^s) - \log_2 |\mathcal{H}_s|\big| < \delta,\ \big|I(W^s[\mathcal{H}_s]) - \log_2 |\mathcal{H}_s|\big| < \delta,\ Z(W^s[\mathcal{H}_s]) < 2^{-2^{\beta' n}} \Big\} \Big| = 1.$$

On the other hand, from Proposition 5.1, we have

$$P_e(W^s[\mathcal{H}_s]) \leq (|\mathcal{H}_s| - 1) Z(W^s[\mathcal{H}_s]) \leq (|\mathcal{X}| - 1) Z(W^s[\mathcal{H}_s]).$$

Therefore,

$$\lim_{n\to\infty} \frac{1}{2^n} \Big| \Big\{ s \in \{-,+\}^n :\ \exists \mathcal{H}_s \text{ a stable partition of } (\mathcal{X}, /^*), \\ \big|I(W^s) - \log_2 |\mathcal{H}_s|\big| < \delta,\ \big|I(W^s[\mathcal{H}_s]) - \log_2 |\mathcal{H}_s|\big| < \delta,\ P_e(W^s[\mathcal{H}_s]) < (|\mathcal{X}| - 1) 2^{-2^{\beta' n}} \Big\} \Big| = 1.$$

But $(|\mathcal{X}| - 1) 2^{-2^{\beta' n}} < 2^{-2^{\beta n}}$ for n large enough, hence

$$\lim_{n\to\infty} \frac{1}{2^n} \Big| \Big\{ s \in \{-,+\}^n :\ \exists \mathcal{H}_s \text{ a stable partition of } (\mathcal{X}, /^*), \\ \big|I(W^s) - \log_2 |\mathcal{H}_s|\big| < \delta, \big|I(W^s[\mathcal{H}_s]) - \log_2 |\mathcal{H}_s|\big| < \delta,\ P_e(W^s[\mathcal{H}_s]) < 2^{-2^{\beta n}} \Big\} \Big| = 1.$$

Lemma 3.4 now implies that:

$$\lim_{n\to\infty} \frac{1}{2^n}\big|\big\{s \in \{-,+\}^n :\ W^s \text{ is } (\delta, 2^{-2^{\beta n}})\text{-easy}\big\}\big| = 1.$$

We conclude that every $0 \leq \beta < \frac{1}{2}$ is a $*$-achievable exponent. Therefore, $E_* \geq \frac{1}{2}$. On the other hand, since $*$ is polarizing, Proposition 5.3 implies that $E_* \leq \frac{1}{2}$. Therefore, $E_* = \frac{1}{2}$. □

Corollary 5.1. *For every $\delta > 0$, every $\beta < \frac{1}{2}$, every channel $W : Q \longrightarrow \mathcal{Y}$, and every quasigroup operations $*$ on Q, there exists a polar code for the channel W constructed using $*$ such that its rate is at least $I(W) - \delta$ and its probability of error under successive cancellation decoding is less than 2^{-N^β}, where $N = 2^n$ is the blocklength.*

Proof. The corollary follows from Propositions 3.4 and 5.4. □

Conjecture 5.1. *If $*$ is a polarizing operation that is not a quasigroup operation, then $E_* < \frac{1}{2}$.*

Conjecture 5.1 implies that quasigroup operations are the best polarizing operations. Therefore, if the conjecture is true and we are looking for good polar codes with large blocklength, it is sufficient to consider quasigroup operations.

5.4 Exponent of a MAC-Polarizing Sequence of Binary Operations

Proposition 5.5. *Let $*_1, \ldots, *_m$ be m binary operations on $\mathcal{X}_1, \ldots, \mathcal{X}_m$ respectively. If $\max\limits_{1\leq i\leq m} |\mathcal{X}_i| \geq 2$ and $(*_1, \ldots, *_m)$ is MAC-polarizing, then*

$$E_{*_1,\ldots,*_m} \leq E_{*_1\otimes\ldots\otimes *_m} \leq \min\{E_{*_1}, \ldots, E_{*_m}\} \leq \frac{1}{2}.$$

Proof. Define $* = *_1\otimes\ldots\otimes *_m$. Let $W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \longrightarrow \mathcal{Y}$ be an m-user MAC and let $W' : \mathcal{X} \longrightarrow \mathcal{Y}$ be the single user channel obtained from W (see Definition 4.11). Note that every MAC polar code for the MAC W constructed using $(*_1, \ldots, *_m)$ can be seen as a polar code for the channel W' constructed using the operation $*$. Moreover, the probability of error of the ML decoder is the same. Therefore, every $(*_1, \ldots, *_m)$-achievable exponent is $*$-achievable. Hence, $E_{*_1,\ldots,*_m} \leq E_*$.

Now let $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$. For each $1 \leq i \leq m$ and each single user channel $W_i : \mathcal{X}_i \longrightarrow \mathcal{Y}$ with input alphabet $\mathcal{X}_i$, consider the single user channel $W : \mathcal{X} \longrightarrow \mathcal{Y}$ with input alphabet $\mathcal{X}$ defined as $W\big(y\big|(x_1, \ldots, x_m)\big) = W_i(y|x_i)$. Let $(W_{i,n})_{n\geq 0}$ be the single user channel valued process obtained from W_i using the operation $*_i$ as in Definition 3.3, and let $(W_n)_{n\geq 0}$ be the single user channel valued process obtained from W using the operation $*$ as in Definition 3.3. It is easy to see that for every $\delta > 0$ and every $\epsilon > 0$, $W_{i,n}$ is (δ, ϵ)-easy if and only if W_n is (δ, ϵ)-easy. This implies that each $*$-achievable exponent is $*_i$-achievable. Therefore, $E_* \leq E_{*_i}$ for every $1 \leq i \leq m$, hence $E_* \leq \min\{E_{*_1}, \ldots, E_{*_m}\}$. Now from Proposition 5.3, we have $\min\{E_{*_1}, \ldots, E_{*_m}\} \leq \frac{1}{2}$. □

Proposition 5.6. *Let $*_1, \ldots, *_m$ be m quasigroup operations on the sets $Q_1, \ldots, Q_m$, respectively. If $\max_{1 \leq i \leq m} |Q_i| \geq 2$, then $E_{*_1,\ldots,*_m} = \frac{1}{2}$.*

Proof. Let $* = *_1 \otimes \ldots \otimes *_m$, then $*$ is a quasigroup operation. Let $\beta < \beta' < \frac{1}{2}$. Let $W : Q_1 \times \cdots \times Q_m \longrightarrow \mathcal{Y}$ be an m-user MAC. Define $Q = Q_1 \times \cdots \times Q_m$ and let $W' : Q \longrightarrow \mathcal{Y}$ be the single user channel obtained from W (see Definition 4.11). For each $n > 0$ and each $s \in \{-,+\}^n$, let W'^s be obtained from W' using the operation $*$ (see Definition 3.2), and let W^s be obtained from W using the operations $*_1, \ldots, *_m$ (see Definition 4.7). From Theorem 5.1, we have:

$$\lim_{n\to\infty} \frac{1}{2^n} \Big| \Big\{ s \in \{-,+\}^n : \exists \mathcal{H}_s \text{ a stable partition of } (Q, /^*), \\ \big|I(W'^s) - \log_2 |\mathcal{H}_s|\big| < \delta,\ \big|I(W'^s[\mathcal{H}_s]) - \log_2 |\mathcal{H}_s|\big| < \delta,\ Z(W'^s[\mathcal{H}_s]) < 2^{-2^{\beta' n}} \Big\} \Big| = 1.$$

On the other hand, from Proposition 5.1, we have

$$P_e(W'^s[\mathcal{H}_s]) \leq (|\mathcal{H}_s| - 1) Z(W'^s[\mathcal{H}_s]) \leq (|Q| - 1) Z(W'^s[\mathcal{H}_s]).$$

Therefore,

$$\lim_{n\to\infty} \frac{1}{2^n} \Big| \Big\{ s \in \{-,+\}^n : \exists \mathcal{H}_s \text{ a stable partition of } (Q, /^*), \\ \big|I(W'^s) - \log_2 |\mathcal{H}_s|\big| < \delta,\ \big|I(W'^s[\mathcal{H}_s]) - \log_2 |\mathcal{H}_s|\big| < \delta,\ P_e(W'^s[\mathcal{H}_s]) < (|Q| - 1) 2^{-2^{\beta' n}} \Big\} \Big| = 1.$$

It is easy to see that W'^s is the single user channel obtained from W^s. Therefore, $I(W^s) = I(W'^s)$, $I(W^s[\mathcal{H}_s]) = I(W'^s[\mathcal{H}_s])$ (by definition) and $P_e(W^s[\mathcal{H}_s]) = P_e(W'^s[\mathcal{H}_s])$. On the other hand, we have $(|Q|-1)2^{-2^{\beta' n}} < 2^{-2^{\beta n}}$ for n large enough. We conclude that:

$$\lim_{n\to\infty} \frac{1}{2^n} \Big| \Big\{ s \in \{-,+\}^n : \exists \mathcal{H}_s \text{ a stable partition of } (Q, /^*), \\ \big|I(W^s) - \log_2 |\mathcal{H}_s|\big| < \delta, \big|I(W^s[\mathcal{H}_s]) - \log_2 |\mathcal{H}_s|\big| < \delta,\ P_e(W^s[\mathcal{H}_s]) < 2^{-2^{\beta n}} \Big\} \Big| = 1.$$

Now since $/^* = /^{*_1} \otimes \ldots \otimes /^{*_m}$ and since $/^{*_i}$ is ergodic (as it is a quasigroup operation) for every $1 \leq i \leq m$, Lemma 4.1 implies that:

$$\lim_{n\to\infty} \frac{1}{2^n} \big| \big\{ s \in \{-,+\}^n : W^s \text{ is } (\delta, 2^{-2^{\beta n}})\text{-easy} \big\} \big| = 1.$$

We conclude that every $0 \leq \beta < \frac{1}{2}$ is a $(*_1, \ldots, *_m)$-achievable exponent. Therefore, $E_{*_1,\ldots,*_m} \geq \frac{1}{2}$. On the other hand, we have $E_{*_1,\ldots,*_m} \leq \frac{1}{2}$ from Proposition 5.5. Hence $E_{*_1,\ldots,*_m} = \frac{1}{2}$. □

Corollary 5.2. *For every $\delta > 0$, every $\beta < \frac{1}{2}$, every MAC $W : Q_1 \times \cdots \times Q_m \longrightarrow \mathcal{Y}$, and every quasigroup operations $*_1, \ldots, *_m$ on $Q_1, \ldots, Q_m$ respectively, there exists a polar code for the MAC W constructed using $*_1, \ldots, *_m$ such that its sum-rate is at least $I(W) - \delta$ and its probability of error under successive cancellation decoding is less than 2^{-N^β}, where $N = 2^n$ is the blocklength.*

Proof. The corollary follows from Propositions 4.1 and 5.6. □

Fourier Analysis of MAC Polarization

6

We saw at the end of Chapter 4 that the multiple-access channel (MAC) polarization process might induce a loss in the symmetric-capacity region. This means that MAC-polar codes might not achieve the entire symmetric-capacity region.

In this chapter[1], we provide a single-letter necessary and sufficient condition that characterizes the set of MACs that do not lose any part of their symmetric-capacity region by polarization. The characterization that we provide works in the general setting where we have an arbitrary number of users and each user uses an arbitrary Abelian group operation on his input alphabet. We will show that the reason why a given MAC W loses parts of its symmetric-capacity region by polarization is because its transition probabilities are not "aligned", which makes W "incompatible" with polarization. The "alignment" condition will be expressed in terms of the Fourier transforms of the transition probabilities of W. The use of Fourier analysis in our study should not come as a surprise since the transition probabilities of W^- can be expressed as a convolution of the transition probabilities of W. This is what makes Fourier analysis useful for our study because it turns convolutions into multiplications, which are much easier to analyze.

Note that there are alternate polar coding solutions that can achieve the entire symmetric-capacity region without any loss. These techniques, which are not based on MAC polarization, are hybrid schemes combining single-user channel polarization with other techniques. In [8], Şaşoğlu et al. used the "rate splitting/onion peeling" scheme of [36] and [37] to transform any point on the dominant face of an m-user MAC into a corner point of a $(2m-1)$-user MAC and then applied single-user channel polarization to achieve this corner point. In [22], Arıkan used monotone chain rules to construct polar codes for the Slepian-Wolf problem, but the same technique can be used to achieve the entire symmetric-capacity region of a MAC.

Although the alternate solutions of [8] and [22] can achieve the entire symmetric-capacity region, they are more complicated than MAC-polar codes (i.e., those that are based on MAC polarization). The alternate solution in [8] requires more encoding and decoding complexity because it adds $m-1$ virtual users. Arıkan's solution [22]

[1]The material of this chapter is based on [34, 35].

does not add significant encoding and decoding complexity, but the code design is much more complicated than that of MAC-polar codes. So if we are given a MAC W whose symmetric-capacity region is preserved by polarization (i.e., MAC-polar codes can achieve the entire symmetric-capacity region of this MAC), then using MAC-polar codes for this MAC is preferable to the alternate solutions. One practical implication of this study is that it allows a code designer to determine whether he can use the preferable MAC-polar codes to achieve the symmetric-capacity region.

In Section 6.1, we introduce the preliminaries of this chapter: We describe the MAC polarization process and explain the discrete Fourier transforms on Abelian groups. In Section 6.2, we provide a sufficient condition for the preservation of the symmetric-capacity region. This sufficient condition, which is relatively easy to understand, provides an intuition that clarifies the necessary and sufficient condition that we prove later. In Section 6.3, we characterize the two-user MACs whose symmetric-capacity regions are preserved by polarization. Section 6.4 generalizes the results of Section 6.3 to MACs with arbitrary number of users.

6.1 Preliminaries

Throughout this chapter, $G_1, \ldots, G_m$ are finite Abelian groups. We will use the addition symbol $+$ to denote the group operations of $G_1, \ldots, G_m$. Since every finite Abelian group is isomorphic to the product of cyclic groups, we may assume without loss of generality that $G_1, \ldots, G_m$ are products of cyclic groups. In other words, for every $1 \leq i \leq m$, there exist k_i integers $N_{i,1}, \ldots, N_{i,k_i} > 0$ such that $G_i = \mathbb{Z}_{N_{i,1}} \times \cdots \times \mathbb{Z}_{N_{i,k_i}}$.

6.1.1 Polarization

Notation 6.1. *Let $W : G_1 \times \cdots \times G_m \longrightarrow \mathcal{Z}$ be an m-user MAC. We write $(X_1, \ldots, X_m) \xrightarrow{W} Z$ to denote the following:*

- *$X_1, \ldots, X_m$ are independent random variables uniformly distributed in $G_1, \ldots, G_m$ respectively.*
- *Z is the output of the MAC W when $X_1, \ldots, X_m$ are the inputs.*

Notation 6.2. *Fix $S \subset \{1, \ldots, m\}$ and let $S = \{i_1, \ldots, i_{|S|}\}$, where $i_1 < \ldots < i_{|S|}$. Define G_S as*

$$G_S := \prod_{i \in S} G_i = G_{i_1} \times \cdots \times G_{i_{|S|}}.$$

For every $(x_1, \ldots, x_m) \in G_1 \times \cdots \times G_m$, we write x_S to denote $(x_{i_1}, \ldots, x_{i_{|S|}})$.

Notation 6.3. *Let $W : G_1 \times \cdots \times G_m \longrightarrow \mathcal{Z}$ and $(X_1, \ldots, X_m) \xrightarrow{W} Z$. For every $S \subset \{1, \ldots, m\}$, we write $I_S(W)$ to denote $I(X_S; ZX_{S^c})$. If $S = \{i\}$, we denote $I_{\{i\}}(W)$ as $I_i(W)$.*

$I(W) := I_{\{1,\ldots,m\}}(W) = I(X_1, \ldots, X_m; Z)$ is called the symmetric sum-capacity *of W.*

The symmetric-capacity region *of an m-user MAC* $W : G_1 \times \cdots \times G_m \longrightarrow \mathcal{Z}$ *is defined as:*

$$\mathcal{J}(W) = \Big\{(R_1, \ldots, R_m) \in \mathbb{R}^m : \ \forall S \subset \{1, \ldots, m\},\ 0 \leq \sum_{i \in S} R_i \leq I_S(W)\Big\}.$$

Note that $I(W)$ is called the *symmetric* sum-capacity because it is computed using uniform input distributions. The same is true for $\mathcal{J}(W)$.

Notation 6.4. $\{-,+\}^* := \bigcup_{n \geq 0} \{-,+\}^n$, *where* $\{-,+\}^0 = \{\emptyset\}$.

Definition 6.1. *Let* $W : G_1 \times \cdots \times G_m \longrightarrow \mathcal{Z}$. *We define the m-user MACs* $W^- : G_1 \times \cdots \times G_m \longrightarrow \mathcal{Z}^2$ *and* $W^+ : G_1 \times \cdots \times G_m \longrightarrow \mathcal{Z}^2 \times G_1 \times \cdots \times G_m$ *as follows:*

$$\begin{aligned} &W^-(y_1, y_2 | u_{1,1}, \ldots, u_{m,1}) \\ &\quad = \frac{1}{|G_1| \cdots |G_m|} \sum_{\substack{u_{1,2} \in \mathcal{X}_1 \\ \vdots \\ u_{m,2} \in \mathcal{X}_m}} W(y_1 | u_{1,1} + u_{1,2}, \ldots, u_{m,1} + u_{m,2}) W(y_2 | u_{1,2}, \ldots, u_{m,2}), \end{aligned}$$

and

$$\begin{aligned} &W^+(y_1, y_2, u_{1,1}, \ldots, u_{m,1} | u_{1,2}, \ldots, u_{m,2}) \\ &\qquad = \frac{1}{|G_1| \cdots |G_m|} W(y_1 | u_{1,1} + u_{1,2}, \ldots, u_{m,1} + u_{m,2}) W(y_2 | u_{1,2}, \ldots, u_{m,2}). \end{aligned}$$

For every $s \in \{-,+\}^*$, *we define the MAC* W^s *as follows:*

$$W^s := \begin{cases} W & \text{if } s = \emptyset, \\ (\ldots((W^{s_1})^{s_2})\ldots)^{s_n} & \text{if } s = (s_1, \ldots, s_n). \end{cases}$$

The following remark explains why polarization might induce a loss in the symmetric-capacity region.

Remark 6.1. *Let* $U_1^m = (U_1, \ldots, U_m)$ *and* $\tilde{U}_1^m = (\tilde{U}_1, \ldots, \tilde{U}_m)$ *be two independent random variables uniformly distributed in* $G_1 \times \cdots \times G_m$. *Let* $X_1^m = U_1^m + \tilde{U}_1^m$ *and* $\tilde{X}_1^m = \tilde{U}_1^m$. *Let* $(X_1, \ldots, X_m) \xrightarrow{W} Z$ *and* $(\tilde{X}_1, \ldots, \tilde{X}_m) \xrightarrow{W} \tilde{Z}$. *We have:*

- $I(W) = I(X_1^m; Z) = I(\tilde{X}_1^m; \tilde{Z})$.
- $I(W^-) = I(U_1^m; Z\tilde{Z})$ *and* $I(W^+) = I(\tilde{U}_1^m; Z\tilde{Z}U_1^m)$.

Hence,

$$\begin{aligned} 2I(W) &= I(X_1^m; Z) + I(\tilde{X}_1^m; \tilde{Z}) = I(X_1^m \tilde{X}_1^m; Z\tilde{Z}) = I(U_1^m \tilde{U}_1^m; Z\tilde{Z}) \\ &= I(U_1^m; Z\tilde{Z}) + I(\tilde{U}_1^m; Z\tilde{Z} | U_1^m) \overset{(a)}{=} I(U_1^m; Z\tilde{Z}) + I(\tilde{U}_1^m; Z\tilde{Z}U_1^m) \\ &= I(W^-) + I(W^+), \end{aligned}$$

where (a) follows from the fact that U_1^m is independent of $\tilde{U}_1^m$.

Therefore, the symmetric sum-capacity is preserved by polarization. On the other hand, I_S might not be preserved if $S \subsetneq \{1,\ldots,m\}$.

For example, consider the two-user MAC case. Let $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$. Let (U_1, V_1) and (U_2, V_2) be two independent random pairs uniformly distributed in $G_1 \times G_2$. Let $X_1 = U_1 + U_2$, $X_2 = U_2$, $Y_1 = V_1 + V_2$ and $Y_2 = V_2$. Let $(X_1, Y_1) \xrightarrow{W} Z_1$ and $(X_2, Y_2) \xrightarrow{W} Z_2$. We have:

- *$I_1(W^-) = I(U_1; Z_1 Z_2 V_1)$ and $I_1(W^+) = I(U_2; Z_1 Z_2 U_1 V_1 V_2)$.*
- *$I_2(W^-) = I(V_1; Z_1 Z_2 U_1)$ and $I_2(W^+) = I(V_2; Z_1 Z_2 U_1 V_1 U_2)$.*

On the other hand, we have:

- *$I_1(W) = I(X_1; Z_1 Y_1) = I(X_2; Z_2 Y_2)$.*
- *$I_2(W) = I(Y_1; Z_1 X_1) = I(Y_2; Z_2 X_2)$.*

Therefore,

$$\begin{aligned} 2I_1(W) &= I(X_1; Z_1 Y_1) + I(X_2; Z_2 Y_2) = I(X_1 X_2; Z_1 Z_2 Y_1 Y_2) \\ &= I(U_1 U_2; Z_1 Z_2 V_1 V_2) = I(U_1; Z_1 Z_2 V_1 V_2) + I(U_2; Z_1 Z_2 V_1 V_2 U_1) \\ &\overset{(a)}{\geq} I(U_1; Z_1 Z_2 V_1) + I(U_2; Z_1 Z_2 V_1 V_2 U_1) = I_1(W^-) + I_1(W^+), \end{aligned} \tag{6.1}$$

where (a) follows from the fact that

$$I(U_1; Z_1 Z_2 V_1 V_2) = I(U_1; Z_1 Z_2 V_1) + I(U_1; V_2 | Z_1 Z_2 V_1) \geq I(U_1; Z_1 Z_2 V_1).$$

Similarly,

$$\begin{aligned} 2I_2(W) &= I(Y_1; Z_1 X_1) + I(Y_2; Z_2 X_2) = I(Y_1 Y_2; Z_1 Z_2 X_1 X_2) \\ &= I(V_1 V_2; Z_1 Z_2 U_1 U_2) = I(V_1; Z_1 Z_2 U_1 U_2) + I(V_2; Z_1 Z_2 U_1 U_2 V_1) \\ &\geq I(V_1; Z_1 Z_2 U_1) + I(V_2; Z_1 Z_2 U_1 U_2 V_1) = I_2(W^-) + I_2(W^+). \end{aligned}$$

Note that $\frac{1}{2^0} \sum_{s \in \{-,+\}^0} I_1(W^s) = \frac{1}{1} I_1(W^\emptyset) = I_1(W) \leq I_1(W)$. Now let $n \geq 0$ and assume that $\frac{1}{2^n} \sum_{s \in \{-,+\}^n} I_1(W^s) \leq I_1(W)$, then

$$\begin{aligned} \frac{1}{2^{n+1}} \sum_{s \in \{-,+\}^{n+1}} I_1(W^s) &= \frac{1}{2^{n+1}} \sum_{s \in \{-,+\}^n} \left(I_1\left(W^{(s,-)}\right) + I_1\left(W^{(s,+)}\right) \right) \\ &\overset{(a)}{\leq} \frac{1}{2^{n+1}} \sum_{s \in \{-,+\}^n} 2I_1(W) = \frac{1}{2^n} \sum_{s \in \{-,+\}^n} I_1(W^s) \leq I_1(W), \end{aligned}$$

where (a) follows from applying (6.1) to W^s. We conclude that for every $n \geq 0$ we have:

$$\frac{1}{2^n} \sum_{s \in \{-,+\}^n} I_1(W^s) \leq I_1(W). \tag{6.2}$$

Similarly,

$$\frac{1}{2^n}\sum_{s\in\{-,+\}^n} I_2(W^s) \le I_2(W). \tag{6.3}$$

By using a similar induction argument, but using the equality $I\left(W^{(s,-)}\right) + I\left(W^{(s,+)}\right) = 2I(W^s)$*, we can show that for every* $n \ge 0$*, we have:*

$$\frac{1}{2^n}\sum_{s\in\{-,+\}^n} I(W^s) = I(W). \tag{6.4}$$

While (6.4) *shows that polarization preserves the symmetric sum-capacity,* (6.2) *and* (6.3) *show that polarization might result into a loss in the symmetric-capacity region.*

Similarly, for the m*-user case, we have*

$$\frac{1}{2^n}\sum_{s\in\{-,+\}^n} I_S(W^s) \le I_S(W),\ \forall S \subsetneq \{1,\dots,m\}.$$

Definition 6.2. *Let* $S \subset \{1,\dots,m\}$*. We say that polarization* $*$-preserves I_S *for* W *if for all* $n \ge 0$ *we have:*

$$\frac{1}{2^n}\sum_{s\in\{-,+\}^n} I_S(W^s) = I_S(W).$$

If polarization $*$*-preserves* I_S *for every* $S \subset \{1,\dots,m\}$*, we say that polarization* $*$-preserves *the symmetric-capacity region for* W*.*

Remark 6.2. *If polarization* $*$*-preserves the symmetric-capacity region for* W*, then the entire symmetric-capacity region can be achieved by polar codes.*

Section 6.3 provides a characterization of two-user MACs whose I_1 is $*$-preserved by polarization. Section 6.4 generalizes the results of Section 6.3 and provides a characterization of m-user MACs whose I_S is $*$-preserved by polarization, where $S \subsetneq \{1,\dots,m\}$. This yields a complete characterization of the MACs with $*$-preserved symmetric-capacity regions.

6.1.2 Discrete Fourier Transform on Finite Abelian Groups

A tool that we are going to need for the analysis of the polarization process is the discrete Fourier transform (DFT) on finite Abelian groups. Since every finite Abelian group is isomorphic to the product of cyclic groups, the DFT on finite Abelian groups can be defined based on the usual multidimensional DFT.

Definition 6.3. *The* k*-dimensional discrete Fourier transform of a mapping* $f : \mathbb{Z}_{N_1} \times \cdots \times \mathbb{Z}_{N_k} \to \mathbb{C}$ *is the mapping* $\hat{f} : \mathbb{Z}_{N_1} \times \cdots \times \mathbb{Z}_{N_k} \to \mathbb{C}$ *defined as:*

$$\hat{f}(\hat{x}_1,\dots,\hat{x}_k) = \sum_{x_1\in\mathbb{Z}_{N_1},\dots,x_k\in\mathbb{Z}_{N_k}} f(x_1,\dots,x_k) e^{-j\frac{2\pi\hat{x}_1 x_1}{N_1}\cdots-j\frac{2\pi\hat{x}_k x_k}{N_k}}.$$

Notation 6.5. *Let $G = \mathbb{Z}_{N_1} \times \cdots \times \mathbb{Z}_{N_k}$ be a finite Abelian group. For every $x = (x_1, \ldots, x_k) \in G$ and every $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_k) \in G$, define $\langle \hat{x}, x \rangle \in \mathbb{R}$ as:*

$$\langle \hat{x}, x \rangle := \frac{\hat{x}_1 x_1}{N_1} + \cdots + \frac{\hat{x}_k x_k}{N_k} \in \mathbb{R}.$$

Using this notation, the DFT on G has a compact formula:

$$\hat{f}(\hat{x}) = \sum_{x \in G} f(x) e^{-j2\pi \langle \hat{x}, x \rangle}.$$

In the rest of this section, we recall well known properties of DFT.

Proposition 6.1. *The inverse DFT is given by the following formula:*

$$f(x) = \frac{1}{|G|} \sum_{\hat{x} \in G} \hat{f}(\hat{x}) e^{j2\pi \langle \hat{x}, x \rangle}.$$

Definition 6.4. *The convolution of two mappings $f : G \to \mathbb{C}$ and $g : G \to \mathbb{C}$ is the mapping $f * g : G \to \mathbb{C}$ defined as:*

$$(f * g)(x) = \sum_{x' \in G} f(x') g(x - x').$$

We will sometimes write $f(x) * g(x)$ to denote $(f * g)(x)$.

Proposition 6.2. *Let $f : G \to \mathbb{C}$ and $g : G \to \mathbb{C}$ be two mappings. We have:*

- $\widehat{(f * g)}(\hat{x}) = \hat{f}(\hat{x}) \hat{g}(\hat{x})$.
- $\widehat{(f \cdot g)}(\hat{x}) = \frac{1}{|G|} (\hat{f} * \hat{g})(\hat{x})$.
- *If $f_a : G \to \mathbb{C}$ is defined as $f_a(x) = f(x - a)$, then $\hat{f}_a(\hat{x}) = \hat{f}(\hat{x}) e^{-j2\pi \langle \hat{x}, a \rangle}$.*
- *If $\tilde{f} : G \to \mathbb{C}$ is defined as $\tilde{f}(x) = f(-x)$, then $\hat{\tilde{f}}(\hat{x}) = \hat{f}(\hat{x})^*$, where $\hat{f}(\hat{x})^*$ is the complex conjugate of $\hat{f}(\hat{x})$.*

6.1.3 Useful Notation

This subsection introduces useful notation that will be used throughout this chapter. The usefulness of this notation will be clear later. We added this subsection so that the reader may refer to it anytime.

Let $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$ be a two-user MAC and let $(X, Y) \xrightarrow{W} Z$. Define the following:

- $\mathrm{YZ}(W) := \{(y, z) \in G_2 \times \mathcal{Z} : \ P_{Y,Z}(y, z) > 0\}$. This is just the support of $P_{Y,Z}$.
- For every $(y, z) \in \mathrm{YZ}(W)$, define $p_{y,z,W} : G_1 \to [0, 1]$ as

$$p_{y,z,W}(x) = P_{X|Y,Z}(x|y, z), \ \ \forall x \in G_1.$$

For every $z \in \mathcal{Z}$, define:

- $\mathrm{Y}^z(W) := \{y \in G_2 :\ P_{Y,Z}(y,z) > 0\}$.
- $\Delta\mathrm{Y}^z(W) := \big\{y_1 - y_2 :\ y_1, y_2 \in \mathrm{Y}^z(W)\big\}$.
- $\hat{\mathrm{X}}^z(W) := \big\{\hat{x} \in G_1 : \exists y \in \mathrm{Y}^z(W), \hat{p}_{y,z,W}(\hat{x}) \neq 0\big\}$.
- $\mathrm{D}^z(W) := \hat{\mathrm{X}}^z(W) \times \Delta\mathrm{Y}^z(W) = \big\{(\hat{x}, y) :\ \hat{x} \in \hat{\mathrm{X}}^z(W),\ y \in \Delta\mathrm{Y}^z(W)\big\}$.

Now define:

- $\hat{\mathrm{X}}\mathrm{Z}(W) := \big\{(\hat{x}, z) :\ z \in \mathcal{Z},\ \hat{x} \in \hat{\mathrm{X}}^z(W)\big\}$.
- $\mathrm{D}(W) := \bigcup_{z \in \mathcal{Z}} \mathrm{D}^z(W)$.

6.1.4 Pseudo-Quadratic Functions

Definition 6.5. *Let $D \subset G_1 \times G_2$. Define the following sets:*

- $H_1(D) := \{x \in G_1 :\ \exists y \in G_2,\ (x,y) \in D\}$.
- *For every $x \in H_1(D)$, let* $H_2^x(D) := \{y \in G_2 :\ (x,y) \in D\}$.
- $H_2(D) := \{y \in G_2 :\ \exists x \in G_1,\ (x,y) \in D\}$.
- *For every $y \in H_2(D)$, let* $H_1^y(D) := \{x \in G_1 :\ (x,y) \in D\}$.

We say that D is a pseudo-quadratic domain *if:*

- *$H_1^y(D)$ is a subgroup of G_1 for every $y \in H_2(D)$.*
- *$H_2^x(D)$ is a subgroup of G_2 for every $x \in H_1(D)$.*

Definition 6.6. *Let $D \subset G_1 \times G_2$ and let $F : D \to \mathbb{T}$ be a mapping from D to $\mathbb{T} = \{\omega \in \mathbb{C} :\ |\omega| = 1\}$. We say that F is a* pseudo-quadratic function *if:*

- *D is a pseudo-quadratic domain.*
- *For every $y \in H_2(D)$, the mapping $x \to F(x,y)$ is a group homomorphism from $\big(H_1^y(D), +\big)$ to $(\mathbb{T}, \cdot)$.*
- *For every $x \in H_1(D)$, the mapping $y \to F(x,y)$ is a group homomorphism from $\big(H_2^x(D), +\big)$ to $(\mathbb{T}, \cdot)$.*

Definition 6.7. *We say that $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$ is* polarization compatible with respect to the first user *if there exists a pseudo-quadratic function $F : D \to \mathbb{T}$ such that:*

- $\mathrm{D}(W) \subset D \subset G_1 \times G_2$.
- *For every $(\hat{x}, z) \in \hat{\mathrm{X}}\mathrm{Z}(W)$ and every $y_1, y_2 \in \mathrm{Y}^z(W)$, we have $\hat{p}_{y_1,z,W}(\hat{x}) = F(\hat{x}, y_1 - y_2) \cdot \hat{p}_{y_2,z,W}(\hat{x})$.*

6.1.5 Main Result

The following theorem is the main result of this chapter:

Theorem 6.1. *If W is a two-user MAC, then polarization $*$-preserves I_1 for W if and only if W is polarization compatible with respect to the first user.*

Theorem 6.1 has the following implications:

- (Proposition 6.9) If $G_1 = G_2 = \mathbb{F}_q$ for a prime q and $(X, Y) \xrightarrow{W} Z$, then polarization $*$-preserves I_1 for W if and only if there exists $a \in \mathbb{F}_q$ such that $I(X + aY; Y|Z) = 0$.
- (Corollary 6.3) Polarization $*$-preserves the symmetric-capacity region for the binary adder channel.
- (Proposition 6.10) If $|G_1|$ and $|G_2|$ are co-prime and $(X, Y) \xrightarrow{W} Z$, then polarization $*$-preserves I_1 for W if and only if $I(X; Y|Z) = 0$ (i.e., if and only if the dominant face of $\mathcal{J}(W)$ is a single point).

The reader may find the polarization compatibility condition (Definition 6.7) too abstract at this stage and it may not be clear why the $*$-preservation of I_1 has anything to do with pseudo-quadratic functions. In order to clarify the meaning of polarization compatibility and make it more intuitive, we provide in Section 6.2 a sufficient condition for the $*$-preservation of I_1 that is easy to understand. After expressing this condition in terms of $\{\hat{p}_{y,z,W} : (y, z) \in \mathrm{YZ}(W)\}$, the link between the $*$-preservation of I_1 and pseudo-quadratic functions should become clear.

6.2 A Sufficient Condition for the $*$-Preservation of I_1

In this section, we only consider two-user MACs $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$, where G_1 and G_2 are finite Abelian groups. We derive a sufficient condition which ensures that polarization $*$-preserves I_1.

Definition 6.8. *Let $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$ be a two-user MAC. We say that I_1 is* preserved *for W if and only if $I_1(W^-) + I_1(W^+) = 2I_1(W)$.*

Lemma 6.1. *Polarization $*$-preserves I_1 for W if and only if I_1 is preserved for W^s for every $s \in \{-, +\}^*$.*

Proof. Polarization $*$-preserves I_1 for W if and only if

$$\begin{aligned}
\forall n \geq 0,\ I_1(W) = \frac{1}{2^n} \sum_{s \in \{-,+\}^n} I_1(W^s) & \\
\Leftrightarrow \quad \forall n \geq 0,\ \frac{1}{2^n} \sum_{s \in \{-,+\}^n} I_1(W^s) &= \frac{1}{2^{n+1}} \sum_{s' \in \{-,+\}^{n+1}} I_1(W^{s'}) \\
\Leftrightarrow \quad \forall n \geq 0,\ \sum_{s \in \{-,+\}^n} 2I_1(W^s) &= \sum_{s \in \{-,+\}^n} (I_1(W^{(s,-)}) + I_1(W^{(s,+)})) \\
\Leftrightarrow \quad \forall n \geq 0,\ \sum_{s \in \{-,+\}^n} \big(2I_1(W^s) - I_1(W^{(s,-)}) - I_1(W^{(s,+)})\big) &= 0.
\end{aligned}$$

But since $2I_1(W^s)-I_1(W^{(s,-)})-I_1(W^{(s,+)}) \geq 0$ (apply (6.1) to W^s), we conclude that polarization $*$-preserves I_1 for W if and only if

$$\forall n \geq 0, \forall s \in \{-,+\}^n,\ I_1(W^{(s,-)}) + I_1(W^{(s,+)}) = 2I_1(W^s).$$

In other words, polarization $*$-preserves I_1 for W if and only if I_1 is preserved for W^s for every $s \in \{-,+\}^*$. □

Suppose we want to prove that a given condition on W is sufficient for the $*$-preservation of I_1. Lemma 6.1 suggests a method to do that:

1. Show that if W satisfies the condition, then I_1 is preserved for W.
2. Show that if W satisfies the condition, then W^- and W^+ satisfy the condition as well.

By doing that, we would have shown that if W satisfies the condition, then W^s satisfies the same condition for all $s \in \{-,+\}^*$, which in turn implies that I_1 is preserved for W^s for all $s \in \{-,+\}^*$, hence polarization $*$-preserves I_1 for W due to Lemma 6.1.

Definition 6.9. *Let $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$ be a two-user MAC and let $(X,Y) \xrightarrow{W} Z$. We say that W is homomorphic-independent with respect to the first user if and only if there exists a subgroup H_2 of G_2, a group homomorphism $f : H_2 \to G_1$ and a mapping $g : \mathcal{Z} \to G_2$ such that:*

- $\mathbb{P}[Y - g(Z) \in H_2] = 1$.
- $I\big(X + f(Y - g(Z)); Y\big|Z\big) = 0$.

The condition $\mathbb{P}[Y - g(Z) \in H_2] = 1$ means that Y and $g(Z)$ belong to the same coset of H_2. In other words, given $Z = z$, Y belongs to a single coset of H_2, and this coset is determined by $g(z)$. On the other hand, the condition $I\big(X + f(Y - g(Z)); Y\big|Z\big) = 0$ is equivalent to say that given Z, a shifted version of X is conditionally independent of Y, and the amount by which X should be shifted is $f(Y - g(Z))$. One might be tempted to simplify the expression $I\big(X + f(Y - g(Z)); Y\big|Z\big)$ as follows:

$$I\big(X + f(Y - g(Z)); Y\big|Z\big) = I\big(X + f(Y) - f(g(Z)); Y\big|Z\big) = I\big(X + f(Y); Y\big|Z\big).$$

This would be correct if f were defined on the whole group G_2. However, f is only defined on a subgroup H_2 of G_2. This is why f can be applied on $Y - g(Z)$ which belongs to H_2, but cannot be applied to Y and $g(Z)$ individually because they can lie outside H_2.

In the rest of this section, we show that if W is homomorphic-independent with respect to the first user, then polarization $*$-preserves I_1 for W. For the sake of brevity, we will write homomorphic-independent to denote "homomorphic-independent with respect to the first user".

Lemma 6.2. *If $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$ is homomorphic-independent, then I_1 is preserved for W.*

Proof. Let $U_1, U_2, V_1, V_2, X_1, X_2, Y_1, Y_2, Z_1, Z_2$ be as in Remark 6.1. We can see from (6.1) that I_1 is preserved for W if and only if $I(U_1; Z_1Z_2V_1) = I(U_1; Z_1Z_2V_1V_2)$. Therefore, it is sufficient to show that $I(U_1; V_2|Z_1Z_2V_1) = 0$.

Let H_2, f and g be as in Definition 6.9. We have:

$$V_1 - g(Z_1) + g(Z_2) = Y_1 - Y_2 - g(Z_1) + g(Z_2) = \big(Y_1 - g(Z_1)\big) - \big(Y_2 - g(Z_2)\big) \overset{(a)}{\in} H_2,$$

where (a) is true because $\mathbb{P}[Y_1 - g(Z_1) \in H_2] = \mathbb{P}[Y_2 - g(Z_2) \in H_2] = 1$.

Let $\tilde{X}_1 = X_1 + f(Y_1 - g(Z_1))$ and $\tilde{X}_2 = X_2 + f(Y_2 - g(Z_2))$. We have:

$$\begin{aligned} U_1 + f\big(V_1 - g(Z_1) + g(Z_2)\big) &= X_1 - X_2 + f(Y_1 - Y_2 - g(Z_1) + g(Z_2)) \\ &= X_1 + f(Y_1 - g(Z_1)) - X_2 - f(Y_2 - g(Z_2)) \\ &= \tilde{X}_1 - \tilde{X}_2. \end{aligned} \tag{6.5}$$

Therefore,

$$\begin{aligned} I(U_1; V_2|Z_1Z_2V_1) &= I\big(U_1 - f\big(V_1 - g(Z_1) + g(Z_2)\big); V_2\big|Z_1Z_2V_1\big) \\ &= I(\tilde{X}_1 - \tilde{X}_2; V_2|Z_1Z_2V_1) \le I(\tilde{X}_1\tilde{X}_2; V_2|Z_1Z_2V_1) \\ &\le I(\tilde{X}_1\tilde{X}_2; V_1V_2|Z_1Z_2) = I(\tilde{X}_1\tilde{X}_2; Y_1Y_2|Z_1Z_2) \\ &= I(\tilde{X}_1; Y_1|Z_1) + I(\tilde{X}_2; Y_2|Z_2) \overset{(b)}{=} 0, \end{aligned}$$

where (b) follows from the fact that W is homomorphic-independent. We conclude that $I(U_1; V_2|Z_1Z_2V_1) = 0$ and so I_1 is preserved for W. □

Lemma 6.3. *If $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$ is homomorphic-independent, then W^- and W^+ are homomorphic-independent as well.*

Proof. Let $U_1, U_2, V_1, V_2, X_1, X_2, Y_1, Y_2, Z_1, Z_2$ be as in Remark 6.1. Let H_2, f and g be as in Definition 6.9. Define the mappings $g^- : \mathcal{Z}^2 \to G_2$ and $g^+ : \mathcal{Z}^2 \times G_1 \times G_2 \to G_2$ as follows:

$$g^-(z_1, z_2) = g(z_1) - g(z_2) \;\text{ and }\; g^+(z_1, z_2, u_1, v_1) = g(z_2).$$

Since W is homomorphic-independent, we have $\mathbb{P}[Y_1 - g(Z_1) \in H_2] = 1$ and $\mathbb{P}[Y_2 - g(Z_2) \in H_2] = 1$. Therefore, $\mathbb{P}[Y_1 - Y_2 - g(Z_1) + g(Z_2) \in H_2] = 1$ which implies that $\mathbb{P}\big[V_1 - g^-(Z_1, Z_2) \in H_2\big] = 1$. Similarly, $\mathbb{P}\big[V_2 - g^+(Z_1, Z_2, U_1, V_1) \in H_2\big] = \mathbb{P}[Y_2 - g(Z_2) \in H_2] = 1$.

Define $\tilde{X}_1 = X_1 + f(Y_1 - g(Z_1))$ and $\tilde{X}_2 = X_2 + f(Y_2 - g(Z_2))$ as in the proof of Lemma 6.2. From (6.5) we have $U_1 + f(V_1 - g^-(Z_1, Z_2)) = \tilde{X}_1 - \tilde{X}_2$. Therefore,

$$\begin{aligned} &I\big(U_1 + f(V_1 - g^-(Z_1, Z_2)); V_1\big|Z_1Z_2\big) \\ &\qquad = I(\tilde{X}_1 - \tilde{X}_2; V_1|Z_1Z_2) \le I(\tilde{X}_1\tilde{X}_2; V_1V_2|Z_1Z_2) \\ &\qquad = I(\tilde{X}_1\tilde{X}_2; Y_1Y_2|Z_1Z_2) = I(\tilde{X}_1; Y_1|Z_1) + I(\tilde{X}_2; Y_2|Z_2) \overset{(a)}{=} 0, \end{aligned}$$

where (a) follows from the fact that W is homomorphic-independent.

On the other hand, we have

$$\begin{aligned} I\big(U_2 + f(V_2 - g^+(Z_1,Z_2,U_1,V_1)); V_2\big|Z_1Z_2U_1V_1\big) \\ &= I\big(X_2 + f(Y_2 - g(Z_2)); V_2\big|Z_1Z_2U_1V_1\big) \\ &= I(\tilde{X}_2; V_2|Z_1Z_2, U_1 + f(V_1 - g^-(Z_1,Z_2)), V_1) \\ &= I(\tilde{X}_2; V_2|Z_1Z_2, \tilde{X}_1 - \tilde{X}_2, V_1) \le I(\tilde{X}_2, \tilde{X}_1 - \tilde{X}_2; V_2V_1|Z_1Z_2) \\ &= I(\tilde{X}_1\tilde{X}_2; Y_1Y_2|Z_1Z_2) = I(\tilde{X}_1; Y_1|Z_1) + I(\tilde{X}_2; Y_2|Z_2) = 0. \end{aligned}$$

We conclude that W^- and W^+ are homomorphic-independent. □

Proposition 6.3. *If $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$ is homomorphic-independent, then polarization $*$-preserves I_1 for W.*

Proof. We first show by induction on $n \ge 0$ that for every $s \in \{-,+\}^n$, W^s is homomorphic-independent. If $n = 0$, there is nothing to prove. Now let $n > 0$ and suppose that the claim is true for $n - 1$.

Let $s \in \{-,+\}^n$, then there exists $s' \in \{-,+\}^{n-1}$ such that $s = (s',-)$ or $s = (s',+)$. We know from the induction hypothesis that $W^{s'}$ is homomorphic-independent, and by applying Lemma 6.3 to $W^{s'}$ we deduce that both $W^{(s',-)}$ and $W^{(s',+)}$ are homomorphic-independent. Therefore, W^s is homomorphic-independent.

We conclude that for every $n \ge 0$ and every $s \in \{-,+\}^n$, W^s is homomorphic-independent. Lemma 6.2 implies that I_1 is preserved for W^s for every $s \in \{-,+\}^s$, and Lemma 6.1 shows that polarization $*$-preserves I_1 for W. □

One might try to simplify the sufficient condition that we have just shown in the following way. If H_2, f and g are as in Definition 6.9, let $\tilde{f} : G_2 \to G_1$ be an extension of f which is a homomorphism from $(G_2,+)$ to $(G_1,+)$. We have:

$$\begin{aligned} I\big(X + \hat{f}(Y); Y\big|Z\big) &= I\big(X + \hat{f}(Y) - \hat{f}(g(Z)); Y\big|Z\big) \\ &= I\big(X + \hat{f}(Y - g(Z)); Y\big|Z\big) = I\big(X + f(Y - g(Z)); Y\big|Z\big) = 0. \end{aligned}$$

This would suggest that homomorphic-independence is equivalent to the existence of a homomorphism $\hat{f} : G_2 \to G_1$ satisfying $I(X + \hat{f}(Y); Y|Z) = 0$, which is of course simpler than the way homomorphic-independence was defined in Definition 6.9. This argument breaks down when we realize that not every homomorphism $f : H_2 \to G_1$ can be extended to a homomorphism from $(G_2,+)$ to $(G_1,+)$. For example, if $G_1 = \mathbb{F}_2$, $G_2 = \mathbb{Z}_4$, $H_2 = \{0,2\} \subset G_2$ and $f : H_2 \to G_1$ is defined as $f(0) = 0$ and $f(2) = 1$, then f is clearly a homomorphism from H_2 to G_1. However, f is not extendable to a homomorphism $\hat{f} : G_2 \to G_1$ defined on the whole group G_2. If f were extendable, we would have $1 = \hat{f}(2) = \hat{f}(1) + \hat{f}(1) = 2\hat{f}(1) = 0$, which is a contradiction.

The existence of a homomorphism $f : G_2 \to G_1$ satisfying $I(X + f(Y); Y|Z)$ is of course a sufficient condition for the $*$-preservation of I_1 because it is a particular case of homomorphic-independence. However, homomorphic-independence is a strictly more general condition as we have shown in the previous paragraph.

Note that there is a large freedom on the choice of the mapping $g : \mathcal{Z} \to G_2$ in Definition 6.9. The main role of the mapping g is to find the coset of H_2 to which Y belongs, and any other mapping g' playing this role will satisfy the conditions of

Definition 6.9: Let H_2, f and g be as in Definition 6.9 and assume that $g' : \mathcal{Z} \to G_2$ satisfies $g'(z) - g(z) \in H_2$ for all $z \in \mathcal{Z}$. We have $Y - g'(Z) = Y - g(Z) + g(Z) - g'(Z) \in H_2$ with probability 1. On the other hand,

$$\begin{aligned} I\big(X + f(Y - g'(Z)); Y\big|Z\big) &= I\big(X + f\big(Y - g(Z) + g(Z) - g'(Z)\big); Y\big|Z\big) \\ &= I\big(X + f(Y - g(Z)) + f\big(g(Z) - g'(Z)\big); Y\big|Z\big) \\ &= I\big(X + f(Y - g(Z)); Y\big|Z\big) = 0. \end{aligned}$$

Therefore, H_2, f and g' also satisfy the conditions of Definition 6.9.

Let us now see how homomorphic-independence can be expressed in terms of $\{\hat{p}_{y,z,W} : (y,z) \in \mathrm{YZ}(W)\}$. For the sake of brevity, we will write $p_{y,z}$ to denote $p_{y,z,W}$.

The condition $I\big(X + f(Y - g(Z)); Y\big|Z\big) = 0$ is equivalent to the conditional independence of $X + f(Y - g(Z))$ and Y given Z. This is equivalent to say that for every $x \in G_1$, every $y_1, y_2 \in G$ and every $z \in \mathcal{Z}$ satisfying $P_{Y,Z}(y_1, z) > 0$ and $P_{Y,Z}(y_2, z) > 0$, we have

$$P_{X+f(Y-g(Z))|Y,Z}\big(x\big|y_1, z\big) = P_{X+f(Y-g(Z))|Y,Z}\big(x\big|y_2, z\big).$$

On the other hand, we have

$$\begin{aligned} P_{X+f(Y-g(Z))|Y,Z}\big(x\big|y_1, z\big) &= P_{X|Y,Z}\big(x - f(y_1 - g(z))\big|y_1, z\big) \\ &= p_{y_1,z}(x - f(y_1 - g(z))), \end{aligned}$$

and

$$\begin{aligned} P_{X+f(Y-g(Z))|Y,Z}\big(x\big|y_2, z\big) &= P_{X|Y,Z}\big(x - f(y_2 - g(z))\big|y_2, z\big) \\ &= p_{y_2,z}(x - f(y_2 - g(z))). \end{aligned}$$

Therefore, the condition $I\big(X + f(Y - g(Z)); Y\big|Z\big) = 0$ is equivalent to say that for every $z \in \mathcal{Z}$ and every $y_1, y_2 \in \mathrm{Y}^z(W)$, we have

$$\begin{aligned} & p_{y_1,z}(x - f(y_1 - g(z))) = p_{y_2,z}(x - f(y_2 - g(z))), \ \ \forall x \in G_1 \\ \Leftrightarrow\ & p_{y_1,z}(x) = p_{y_2,z}\big(x + f(y_1 - g(z)) - f(y_2 - g(z))\big), \ \ \forall x \in G_1 \\ \Leftrightarrow\ & p_{y_1,z}(x) = p_{y_2,z}\big(x + f(y_1 - g(z) - y_2 + g(z))\big), \ \ \forall x \in G_1 \\ \Leftrightarrow\ & p_{y_1,z}(x) = p_{y_2,z}\big(x + f(y_1 - y_2)\big), \ \ \forall x \in G_1. \end{aligned}$$

This shows that if we want to get rid of the mapping g in the second condition of Definition 6.9, we have to express the homomorphic-independence condition in terms of the conditional probability distributions $\{p_{y,z} : (y,z) \in \mathrm{YZ}(W)\}$. On the other hand, due to the freedom on the choice of the mapping g that we have shown above, we can see that g in the condition $\mathbb{P}[Y - g(Z) \in H_2]$ just serves the purpose of saying that given $Z = z$, Y belongs to a single coset of H_2. In other words, $y_1 - y_2 \in H_2$ for all $z \in \mathcal{Z}$ and every $y_1, y_2 \in \mathrm{Y}^z(W)$, which is equivalent to say that $\Delta\mathrm{Y}^z(W) \subset H_2$ for all $z \in \mathcal{Z}$. Therefore, the first condition of Definition 6.9 can be replaced by $\mathrm{D}(W) \subset G_1 \times H_2$. This shows the following lemma:

Lemma 6.4. *W is homomorphic-independent if and only if there exists a subgroup H_2 of G_2 and a homomorphism $f : H_2 \to G_1$ satisfying:*

- $\mathrm{D}(W) \subset G_1 \times H_2$.
- *For every* $z \in \mathcal{Z}$ *and every* $y_1, y_2 \in \mathrm{Y}^z(W)$*, we have:*

$$\begin{aligned} & p_{y_1,z}(x) = p_{y_2,z}(x + f(y_1 - y_2)), \ \ \forall x \in G_1 \\ \Leftrightarrow \ & \hat{p}_{y_1,z}(\hat{x}) = \hat{p}_{y_2,z}(\hat{x}) e^{j2\pi\langle \hat{x}, f(y_1-y_2)\rangle}, \ \ \forall \hat{x} \in G_1. \end{aligned}$$

Lemma 6.4 suggests that if we are given a two-user MAC W and we want to check whether it is homomorphic-independent, then one way to do that is to compute $Q(\hat{x}, y_1, y_2, z) = \frac{\hat{p}_{y_1,z}(\hat{x})}{\hat{p}_{y_2,z}(\hat{x})}$ for every $z \in \mathcal{Z}$, every $y_1, y_2 \in \mathrm{Y}^z(W)$ and every $\hat{x}$ satisfying $\hat{p}_{y_2,z}(\hat{x}) \neq 0$, and then make sure that $Q(\hat{x}, y_1, y_2, z)$ can be expressed as $e^{j2\pi\langle \hat{x}, f(y_1-y_2)\rangle}$ for some homomorphism $f : H_2 \to G_1$, where H_2 is a subgroup of G_2 that satisfies $\mathrm{D}(W) \subset G_1 \times H_2$.

We can now make the following remarks:

- $e^{j2\pi\langle \hat{x}, f(y_1-y_2)\rangle} \in \mathbb{T} := \{\omega \in \mathbb{C} : \ |\omega| = 1\}$.
- $e^{j2\pi\langle \hat{x}, f(y_1-y_2)\rangle}$ depends only on $\hat{x}$ and $y_1 - y_2$.
- For every $y \in H_2$, the mapping $\hat{x} \to e^{j2\pi\langle \hat{x}, f(y)\rangle}$ is a group homomorphism from $(G_1, +)$ to $(\mathbb{T}, \cdot)$.
- For every $\hat{x} \in G_1$, the mapping $y \to e^{j2\pi\langle \hat{x}, f(y)\rangle}$ is a group homomorphism from $(H_2, +)$ to $(\mathbb{T}, \cdot)$.

Therefore, the mapping $(\hat{x}, y) \to e^{j2\pi\langle \hat{x}, f(y)\rangle}$ is a pseudo-quadratic function from $G_1 \times H_2$ to $\mathbb{T}$.

We can now show the following characterization of homomorphic-independent MACs:

Proposition 6.4. *Let* $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$ *be a two-user MAC.* W *is homomorphic-independent if and only if there exists a subgroup* H_2 *of* G_2 *and a pseudo-quadratic function* $F : G_1 \times H_2 \to \mathbb{T}$ *satisfying:*

- $\mathrm{D}(W) \subset G_1 \times H_2$.
- *For every* $(\hat{x}, z) \in \hat{\mathrm{X}}\mathrm{Z}(W)$ *and every* $y_1, y_2 \in \mathrm{Y}^z(W)$*, we have* $\hat{p}_{y_1,z}(\hat{x}) = F(\hat{x}, y_1 - y_2)\hat{p}_{y_2,z}(\hat{x})$.

Proof. The above discussion shows that the existence of such H_2 and F is a necessary condition for the homomorphic-independence of W. For the proof that it is also sufficient, see Appendix 6.5.1. □

Note that the only difference between polarization compatibility (Definition 6.7) and the characterization of homomorphic-independence of Proposition 6.4 is that the domain D of the pseudo-quadratic function F in Definition 6.7 can be an arbitrary pseudo-quadratic domain, whereas the domain of F in Proposition 6.4 needs to be of the form $G_1 \times H_2$ for some subgroup H_2 of G_2. In the next section, we show that polarization compatibility is a necessary and sufficient condition for the $*$-preservation of I_1.

6.3 Two-user MACs with $*$-Preserved I_1

Throughout this section, we fix a two-user MAC $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$, where G_1 and G_2 are finite Abelian groups. This section is dedicated to proving Theorem 6.1.

6.3.1 Polarization Compatibility is Necessary

For the sake of simplicity, we write $p_{y,z}(x)$ to denote $p_{y,z,W}(x)$.

According to (6.1), I_1 is preserved for W if and only if $I(U_1; V_2|Z_1Z_2V_1) = 0$, which means that for every $z_1, z_2 \in \mathcal{Z}$ and every $v_1, v_2 \in G_2$, if

$$P_{V_2,Z_1,Z_2,V_1}(v_2, z_1, z_2, v_1) > 0,$$

then $P_{U_1|V_2,Z_1,Z_2,V_1}(u_1|v_2, z_1, z_2, v_1)$ does not depend on v_2.

In order to study this condition, we should keep track of the values of $z_1, z_2 \in \mathcal{Z}$ and $v_1, v_2 \in G_2$ for which $P_{V_2,Z_1,Z_2,V_1}(v_2, z_1, z_2, v_1) > 0$. But

$$P_{V_2,Z_1,Z_2,V_1}(v_2, z_1, z_2, v_1) = P_{Y_1,Z_1}(v_1 + v_2, z_1)P_{Y_2,Z_2}(v_2, z_2),$$

so it is sufficient to keep track of the pairs $(y, z) \in G_2 \times \mathcal{Z}$ satisfying $P_{Y,Z}(y, z) > 0$. This is where $\mathrm{YZ}(W)$ and $\{\mathrm{Y}^z(W) : \ z \in \mathcal{Z}\}$ become useful.

The following lemma gives a characterization of two-user MACs with preserved I_1 in terms of the Fourier transform of the distributions $p_{y,z}$.

Lemma 6.5. *I_1 is preserved for W if and only if for every $y_1, y_2, y_1', y_2' \in G_2$ and every $z_1, z_2 \in \mathcal{Z}$ satisfying*

- $y_1 - y_2 = y_1' - y_2'$,
- $y_1, y_1' \in \mathrm{Y}^{z_1}(W)$ *and* $y_2, y_2' \in \mathrm{Y}^{z_2}(W)$,

we have

$$\hat{p}_{y_1,z_1}(\hat{x}) \cdot \hat{p}_{y_2,z_2}(\hat{x})^* = \hat{p}_{y_1',z_1}(\hat{x}) \cdot \hat{p}_{y_2',z_2}(\hat{x})^*, \ \ \forall \hat{x} \in G_1.$$

Proof. Let $U_1, U_2, V_1, V_2, X_1, X_2, Y_1, Y_2, Z_1, Z_2$ be as in Remark 6.1. We know that I_1 is preserved for W if and only if $I(U_1; V_2|Z_1Z_2V_1) = 0$, which is equivalent to say that U_1 is conditionally independent of V_2 given (Z_1, Z_2, V_1).

In other words, for any fixed $(z_1, z_2, v_1) \in \mathcal{Z} \times \mathcal{Z} \times G_2$ satisfying

$$P_{Z_1,Z_2,V_1}(z_1, z_2, v_1) > 0,$$

if $v_2, v_2' \in G_2$ satisfy $P_{V_2|Z_1,Z_2,V_1}(v_2|z_1, z_2, v_1) > 0$ and $P_{V_2|Z_1,Z_2,V_1}(v_2'|z_1, z_2, v_1) > 0$, then we have

$$\forall u_1 \in G_1, \ P_{U_1|V_2,Z_1,Z_2,V_1}(u_1|v_2, z_1, z_2, v_1) = P_{U_1|V_2,Z_1,Z_2,V_1}(u_1|v_2', z_1, z_2, v_1).$$

This condition is equivalent to saying that, for every $z_1, z_2 \in \mathcal{Z}$ and every $v_1, v_2, v_2' \in G_2$ satisfying $P_{Z_1,Z_2,Y_1,Y_2}(z_1, z_2, v_1+v_2, v_2) > 0$ and $P_{Z_1,Z_2,Y_1,Y_2}(z_1, z_2, v_1+v_2', v_2') > 0$, we have

$$\forall u_1 \in G_1, \ P_{X_1-X_2|Z_1,Z_2,Y_1,Y_2}(u_1|z_1, z_2, v_1 + v_2, v_2) = P_{X_1-X_2|Z_1,Z_2,Y_1,Y_2}(u_1|z_1, z_2, v_1 + v_2', v_2').$$

By denoting v_1+v_2, v_2, v_1+v_2' and v_2' as y_1, y_2, y_1' and y_2' respectively (so that $y_1-y_2=y_1'-y_2'=v_1$), we can deduce that I_1 is preserved for W if and only if for every $y_1, y_2, y_1', y_2' \in G_2$ and every $z_1, z_2 \in \mathcal{Z}$ satisfying $y_1-y_2=y_1'-y_2'$, $P_{Z_1,Z_2,Y_1,Y_2}(z_1,z_2,y_1,y_2)>0$ and $P_{Z_1,Z_2,Y_1,Y_2}(z_1,z_2,y_1',y_2')>0$ (i.e., $y_1, y_1' \in \mathrm{Y}^{z_1}(W)$ and $y_2, y_2' \in \mathrm{Y}^{z_2}(W)$), we have

$$\forall u_1 \in G_1,\ P_{X_1-X_2|Z_1,Z_2,Y_1,Y_2}(u_1|z_1,z_2,y_1,y_2)=P_{X_1-X_2|Z_1,Z_2,Y_1,Y_2}(u_1|z_1,z_2,y_1',y_2').$$

On the other hand, we have:

$$\begin{aligned}
&P_{X_1-X_2|Z_1,Z_2,Y_1,Y_2}(u_1|z_1,z_2,y_1,y_2)\\
&\qquad=\sum_{u_2\in G_1} P_{X_1|Z_1,Y_1}(u_1+u_2|z_1,y_1)P_{X_2|Z_2,Y_2}(u_2|z_2,y_2)\\
&\qquad=\sum_{u_2\in G_1} p_{y_1,z_1}(u_1+u_2)p_{y_2,z_2}(u_2)=(p_{y_1,z_1}*\tilde{p}_{y_2,z_2})(u_1),
\end{aligned}$$

where we define $\tilde{p}_{y_2,z_2}(x):=p_{y_2,z_2}(-x)$. Similarly,

$$P_{X_1-X_2|Z_1,Z_2,Y_1,Y_2}(u_1|z_1,z_2,y_1',y_2')=(p_{y_1',z_1}*\tilde{p}_{y_2',z_2})(u_1).$$

Therefore, for every $u_1 \in G_1$, we have

$$(p_{y_1,z_1}*\tilde{p}_{y_2,z_2})(u_1)=(p_{y_1',z_1}*\tilde{p}_{y_2',z_2})(u_1),$$

which is equivalent to $\hat{p}_{y_1,z_1}(\hat{u}_1)\cdot\hat{p}_{y_2,z_2}(\hat{u}_1)^*=\hat{p}_{y_1',z_1}(\hat{u}_1)\cdot\hat{p}_{y_2',z_2}(\hat{u}_1)^*$ for every $\hat{u}_1 \in G_1$. □

Definition 6.10. *Let $W: G_1\times G_2 \longrightarrow \mathcal{Z}$ be a two-user MAC. We say that I_1 is $*^-$* preserved *for W if and only if I_1 is preserved for $W^{[n]^-}$ for every $n\geq 0$, where $[n]^- \in \{-,+\}^n$ is the sequence containing n minus signs (e.g., $[0]^-=\emptyset$, $[2]^-=(-,-)$).*

The following three lemmas study the MACs W for which I_1 is $*^-$ preserved.

Lemma 6.6. *If I_1 is $*^-$ preserved for W, then for every $n>0$, every $y_1,\ldots,y_{2^n}$, $y_1',\ldots,y_{2^n}' \in G_2$ and every $z_1,\ldots,z_{2^n} \in \mathcal{Z}$ satisfying*

- $\sum_{i=1}^{2^n} y_i = \sum_{i=1}^{2^n} y_i'$,
- $y_1 \in \mathrm{Y}^{z_1}(W),\ldots,y_{2^n}\in\mathrm{Y}^{z_{2^n}}(W)$, *and*
- $y_1' \in \mathrm{Y}^{z_1}(W),\ldots,y_{2^n}'\in\mathrm{Y}^{z_{2^n}}(W)$,

we have

$$\prod_{i=1}^{2^n}\hat{p}_{y_i,z_i}(\hat{x})=\prod_{i=1}^{2^n}\hat{p}_{y_i',z_i}(\hat{x}),\ \ \forall \hat{x}\in G_1.$$

Proof. See Appendix 6.5.2. □

Lemma 6.7. *If I_1 is $*^-$ preserved for W then for every $(\hat{x},z)\in\hat{\mathrm{X}}\mathrm{Z}(W)$, we have:*

- $\hat{p}_{y,z}(\hat{x}) \neq 0$ *for all* $y \in \mathrm{Y}^z(W)$.
- $\dfrac{\hat{p}_{y,z}(\hat{x})}{\hat{p}_{y',z}(\hat{x})} \in \mathbb{T}$ *for every* $y, y' \in \mathrm{Y}^z(W)$, *where* $\mathbb{T} := \{\omega \in \mathbb{C} : \ |\omega| = 1\}$.

Proof. If $\hat{x} \in \hat{\mathrm{X}}^z(W)$, there exists $y' \in \mathrm{Y}^z(W)$ satisfying $\hat{p}_{y',z}(\hat{x}) \neq 0$. Fix $y \in \mathrm{Y}^z(W)$ and let $a > 0$ be the order of $y - y'$ in G_2 (i.e., $a(y - y') = 0$). Let $n > 0$ be such that $a < 2^n$ and define the two sequences $(y_i)_{1 \leq i \leq 2^n}$ and $(y'_i)_{1 \leq i \leq 2^n}$ as follows:

- If $1 \leq i \leq a$, $y_i = y$ and $y'_i = y'$.
- If $a < i \leq 2^n$, $y_i = y'_i = y'$.

Since $a(y - y') = 0$, we have $ay = ay'$ and so

$$\sum_{i=1}^{2^n} y_i = ay + (2^n - a)y' = ay' + (2^n - a)y' = \sum_{i=1}^{2^n} y'_i.$$

By applying Lemma 6.6, we get

$$\begin{aligned}\big(\hat{p}_{y,z}(\hat{x})\big)^a \big(\hat{p}_{y',z}(\hat{x})\big)^{2^n - a} &= \prod_{i=1}^{2^n} \hat{p}_{y_i,z}(\hat{x}) = \prod_{i=1}^{2^n} \hat{p}_{y'_i,z}(\hat{x}) \\ &= \big(\hat{p}_{y',z}(\hat{x})\big)^{2^n} \neq 0.\end{aligned}$$

Therefore, $\hat{p}_{y,z}(\hat{x}) \neq 0$. Moreover,

$$\Big(\frac{\hat{p}_{y,z}(\hat{x})}{\hat{p}_{y',z}(\hat{x})}\Big)^a = 1,$$

which means that $\dfrac{\hat{p}_{y,z}(\hat{x})}{\hat{p}_{y',z}(\hat{x})}$ is a root of unity. Hence $\dfrac{\hat{p}_{y,z}(\hat{x})}{\hat{p}_{y',z}(\hat{x})} \in \mathbb{T}$. □

Lemma 6.8. *If* I_1 *is* $*^-$ *preserved for* W, *there exists a unique mapping* $\hat{f}_W : \mathrm{D}(W) \to \mathbb{T}$ *such that for every* $(\hat{x}, z) \in \hat{\mathrm{X}}\mathrm{Z}(W)$ *and every* $y_1, y_2 \in \mathrm{Y}^z(W)$, *we have*

$$\hat{p}_{y_1,z}(\hat{x}) = \hat{f}_W(\hat{x}, y_1 - y_2) \cdot \hat{p}_{y_2,z}(\hat{x}).$$

Proof. Let $(\hat{x}, y) \in \mathrm{D}(W)$. Let z be such that $(\hat{x}, y) \in \mathrm{D}^z(W) = \hat{\mathrm{X}}^z(W) \times \Delta\mathrm{Y}^z(W)$, and let $y_1, y_2 \in \mathrm{Y}^z(W)$ be such that $y_1 - y_2 = y$. We want to show that $\dfrac{\hat{p}_{y_1,z}(\hat{x})}{\hat{p}_{y_2,z}(\hat{x})}$ depends only on $(\hat{x}, y) = (\hat{x}, y_1 - y_2)$ and that $\dfrac{\hat{p}_{y_1,z}(\hat{x})}{\hat{p}_{y_2,z}(\hat{x})} \in \mathbb{T}$.

Suppose there exist $z' \in \mathcal{Z}$ and $y'_1, y'_2 \in \mathrm{Y}^{z'}(W)$ which satisfy $\hat{x} \in \hat{\mathrm{X}}^{z'}(W)$ and $y'_1 - y'_2 = y = y_1 - y_2$. We need to show that $\dfrac{\hat{p}_{y_1,z}(\hat{x})}{\hat{p}_{y_2,z}(\hat{x})} = \dfrac{\hat{p}_{y'_1,z'}(\hat{x})}{\hat{p}_{y'_2,z'}(\hat{x})} \in \mathbb{T}$.

From Lemma 6.7 we have $p_{y_1,z}(\hat{x}) \neq 0$, $p_{y_2,z}(\hat{x}) \neq 0$, $p_{y'_1,z'}(\hat{x}) \neq 0$ and $p_{y'_2,z'}(\hat{x}) \neq 0$. On the other hand, since $y_1 + y'_2 = y_2 + y'_1$, Lemma 6.6 shows that $p_{y_1,z}(\hat{x}) \cdot p_{y'_2,z'}(\hat{x}) = p_{y_2,z}(\hat{x}) \cdot p_{y'_1,z'}(\hat{x})$. Therefore,

$$\frac{p_{y_1,z}(\hat{x})}{p_{y_2,z}(\hat{x})} = \frac{p_{y'_1,z'}(\hat{x})}{p_{y'_2,z'}(\hat{x})} \overset{(a)}{\in} \mathbb{T},$$

where (a) follows from Lemma 6.7. This shows that the value of $\frac{p_{y_1,z}(\hat{x})}{p_{y_2,z}(\hat{x})} \in \mathbb{T}$ depends only on $(\hat{x}, y)$ and does not depend on the choice of z, y_1, y_2. We conclude that there exists a unique $\hat{f}_W(\hat{x}, y) \in \mathbb{T}$ such that for every $z \in \mathcal{Z}$ and every $y_1, y_2 \in \mathrm{Y}^z(W)$ satisfying $\hat{x} \in \hat{\mathrm{X}}^z(W)$ and $y_1 - y_2 = y$, we have $\hat{p}_{y_1,z}(\hat{x}) = \hat{f}_W(\hat{x}, y) \cdot \hat{p}_{y_2,z}(\hat{x})$. □

Notice that the only difference between the mapping $\hat{f}_W$ in Lemma 6.8 and the function F in Definition 6.7 is that F is a pseudo-quadratic function defined on a pseudo-quadratic domain D containing $\mathrm{D}(W)$, whereas $\hat{f}_W$ is only defined on $\mathrm{D}(W)$. Therefore, if we want to prove that W is polarization compatible, we have to show that $\hat{f}_W$ can be extended to a pseudo-quadratic function.

Another important remark is that if polarization $*$-preserves I_1 for W, then from Lemma 6.1 we can see that I_1 is preserved for $W^{(s,[n]^-)}$ for every $s \in \{-,+\}^*$ and every $n \geq 0$. Therefore, I_1 is $*^-$ preserved for W^s for every $s \in \{-,+\}^*$. Lemma 6.8 now implies that for every $s \in \{-,+\}^*$, there exists a function $\hat{f}_{W^s} : \mathrm{D}(W^s) \to \mathbb{T}$ such that for every $(\hat{x}, z^s) \in \hat{\mathrm{X}}\mathrm{Z}(W^s)$ and every $y_1, y_2 \in \mathrm{Y}^{z^s}(W^s)$, we have

$$\hat{p}_{y_1,z^s,W^s}(\hat{x}) = \hat{f}_{W^s}(\hat{x}, y_1 - y_2) \cdot \hat{p}_{y_2,z^s,W^s}(\hat{x}).$$

By studying the relations between $\mathrm{D}(W)$ and $\hat{f}_W$ on one hand and $\mathrm{D}(W^s)$ and $\hat{f}_{W^s}$ on the other hand, we can deduce restrictions on $\hat{f}_W$ which will allow us to extend it to a pseudo-quadratic function.

The following proposition shows how $\mathrm{D}(W^-)$ and $\hat{f}_{W^-}$ are related to $\mathrm{D}(W)$ and $\hat{f}_W$ in the case where I_1 is $*^-$ preserved for W.

Proposition 6.5. *If I_1 is $*^-$ preserved for W, we have:*

1. $\mathrm{D}(W^-) = \{(\hat{x}, y_1 + y_2) : (\hat{x}, y_1), (\hat{x}, y_2) \in \mathrm{D}(W)\}$.
2. *For every $\hat{x} \in G_1$ and every $y_1, y_2 \in G_2$ satisfying $(\hat{x}, y_1), (\hat{x}, y_2) \in \mathrm{D}(W)$, we have*
$$\hat{f}_{W^-}(\hat{x}, y_1 + y_2) = \hat{f}_W(\hat{x}, y_1) \cdot \hat{f}_W(\hat{x}, y_2).$$

Proof. See Appendix 6.5.3. □

Corollary 6.1. *If I_1 is $*^-$ preserved for W, then $\mathrm{D}(W) \subset \mathrm{D}(W^-)$ and $\hat{f}_{W^-}(\hat{x}, y) = \hat{f}_W(\hat{x}, y)$ for every $(\hat{x}, y) \in \mathrm{D}(W)$, i.e., $\hat{f}_{W^-}$ is an extension of $\hat{f}_W$.*

Proof. Let $(\hat{x}, y) \in \mathrm{D}(W)$. There exists $z \in \mathcal{Z}$ and $y_1, y_2 \in \mathrm{Y}^z(W)$ such that $y = y_1 - y_2$, $\hat{p}_{y_1,z}(\hat{x}) \neq 0$ and $\hat{p}_{y_2,z}(\hat{x}) \neq 0$. Since $y_1 \in \mathrm{Y}^z(W)$, we have $0 = y_1 - y_1 \in \Delta\mathrm{Y}^z(W)$. Therefore, we have $(\hat{x}, 0) \in \mathrm{D}(W)$ and $\hat{f}_W(\hat{x}, 0) = \frac{\hat{p}_{y_1,z}(\hat{x})}{\hat{p}_{y_1,z}(\hat{x})} = 1$.

Since $(\hat{x}, y) \in \mathrm{D}(W)$ and $(\hat{x}, 0) \in \mathrm{D}(W)$, Proposition 6.5 implies that $(\hat{x}, y) = (\hat{x}, y + 0) \in \mathrm{D}(W^-)$ and $\hat{f}_{W^-}(\hat{x}, y) = \hat{f}_W(\hat{x}, y) \cdot \hat{f}_W(\hat{x}, 0) = \hat{f}_W(\hat{x}, y)$. □

The following proposition shows how $\mathrm{D}(W^+)$ and $\hat{f}_{W^+}$ are related to $\mathrm{D}(W)$ and $\hat{f}_W$ in the case where polarization $*$-preserves I_1 for W.

Proposition 6.6. *If polarization $*$-preserves I_1 for W, we have:*

1. $\big\{(\hat{x}_1+\hat{x}_2,y):\ (\hat{x}_1,y),(\hat{x}_2,y)\in \mathrm{D}(W)\big\}\subset \mathrm{D}(W^+)$.
2. *For every* $\hat{x}_1,\hat{x}_2\in G_1$ *and every* $y\in G_2$ *satisfying* $(\hat{x}_1,y),(\hat{x}_2,y)\in \mathrm{D}(W)$, *we have*
$$\hat{f}_{W^+}(\hat{x}_1+\hat{x}_2,y)=\hat{f}_W(\hat{x}_1,y)\cdot\hat{f}_W(\hat{x}_2,y).$$

Proof. See Appendix 6.5.4. □

Corollary 6.2. *If polarization* $*$*-preserves* I_1 *for* W, *then* $\mathrm{D}(W)\subset\mathrm{D}(W^+)$ *and* $\hat{f}_{W^+}(\hat{x},y)=\hat{f}_W(\hat{x},y)$ *for every* $(\hat{x},y)\in\mathrm{D}(W)$, *i.e.,* $\hat{f}_{W^+}$ *is an extension of* $\hat{f}_W$.

Proof. For every $(\hat{x},y)\in\mathrm{D}(W)$, there exists $z\in\mathcal{Z}$ and $y_1,y_2\in\mathrm{Y}^z(W)$ such that $\hat{x}\in\hat{\mathrm{X}}^z(W)$ and $y=y_1-y_2$. Lemma 6.7 implies that $\hat{p}_{y_1,z}(\hat{x})\neq 0$ and $\hat{p}_{y_2,z}(\hat{x})\neq 0$. We have:
$$\hat{p}_{y_1,z}(0)=\sum_{x\in G_1}p_{y_1,z}(x)e^{-j2\pi\langle 0,x\rangle}=\sum_{x\in G_1}p_{y_1,z}(x)=1\neq 0.$$
Similarly, $\hat{p}_{y_2,z}(0)=1\neq 0$. Therefore, we have $0\in\hat{\mathrm{X}}^z(W)$ and $y\in\Delta\mathrm{Y}^z(W)$. Hence, $(0,y)\in\mathrm{D}(W)$ and $\hat{f}_W(0,y)=\dfrac{\hat{p}_{y_1,z}(0)}{\hat{p}_{y_2,z}(0)}=1$.

Since $(\hat{x},y)\in\mathrm{D}(W)$ and $(0,y)\in\mathrm{D}(W)$, Proposition 6.6 implies that $(\hat{x},y)=(\hat{x}+0,y)\in\mathrm{D}(W^+)$ and $\hat{f}_{W^+}(\hat{x},y)=\hat{f}_W(\hat{x},y)\hat{f}_W(0,y)=\hat{f}_W(\hat{x},y)$. □

The next proposition gives a necessary condition for the $*$-preservation of I_1:

Proposition 6.7. *If polarization* $*$*-preserves* I_1 *for* W, *then* $\hat{f}_W$ *can be extended to a pseudo-quadratic function.*

Proof. Define the sequence $(W_n)_{n\geq 0}$ of MACs recursively as follows:

- $W_0=W$.
- $W_n=W_{n-1}^-$ if $n>0$ is odd.
- $W_n=W_{n-1}^+$ if $n>0$ is even.

For example, we have $W_1=W^-$, $W_2=W^{(-,+)}$, $W_3=W^{(-,+,-)}$, $W_4=W^{(-,+,-,+)}$ ...

It follows from Corollaries 6.1 and 6.2 that:

- The sequence of sets $\big(\mathrm{D}(W_n)\big)_{n\geq 0}$ is increasing.
- $\hat{f}_{W_n}$ is an extension of $\hat{f}_W$ for every $n>0$.

Since $\big(\mathrm{D}(W_n)\big)_{n\geq 0}$ is increasing and since $G_1\times G_2$ is finite, there exists $n_0>0$ such that for every $n\geq n_0$ we have $\mathrm{D}(W_n)=\mathrm{D}(W_{n_0})$ for all $n\geq n_0$. We may assume without loss of generality that n_0 is even. Define the following sets:

- $\hat{H}_1=\{\hat{x}\in G_1:\ \exists y\in G_2,\ (\hat{x},y)\in\mathrm{D}(W_{n_0})\}$.
- For every $\hat{x}\in\hat{H}_1$, let $H_2^{\hat{x}}=\{y\in G_2:\ (\hat{x},y)\in\mathrm{D}(W_{n_0})\}$.

- $H_2 = \{y \in G_2 : \exists \hat{x} \in G_1,\ (\hat{x}, y) \in \mathrm{D}(W_{n_0})\}$.
- For every $y \in H_2$, let $\hat{H}_1^y = \{\hat{x} \in G_1 : (\hat{x}, y) \in \mathrm{D}(W_{n_0})\}$.

We have the following:

- For every fixed $y \in H_2$, let $\hat{x}_1, \hat{x}_2 \in \hat{H}_1^y$ so that $(\hat{x}_1, y), (\hat{x}_2, y) \in \mathrm{D}(W_{n_0}) \subset \mathrm{D}(W_{n_0+1})$. It follows from Proposition 6.6 that $(\hat{x}_1 + \hat{x}_2, y) \in \mathrm{D}(W_{n_0+1}^+) = \mathrm{D}(W_{n_0+2}) = \mathrm{D}(W_{n_0})$ which implies that $\hat{x}_1 + \hat{x}_2 \in \hat{H}_1^y$. Hence $\hat{H}_1^y$ is a subgroup of $(G_1, +)$. Moreover, we have:

$$\begin{aligned}\hat{f}_{W_{n_0}}(\hat{x}_1 + \hat{x}_2, y) &\overset{(a)}{=} \hat{f}_{W_{n_0+2}}(\hat{x}_1 + \hat{x}_2, y) = \hat{f}_{W_{n_0+1}^+}(\hat{x}_1 + \hat{x}_2, y) \\ &\overset{(b)}{=} \hat{f}_{W_{n_0+1}}(\hat{x}_1, y) \cdot \hat{f}_{W_{n_0+1}}(\hat{x}_2, y) \overset{(c)}{=} \hat{f}_{W_{n_0}}(\hat{x}_1, y) \cdot \hat{f}_{W_{n_0}}(\hat{x}_2, y),\end{aligned}$$

  where (a) and (c) follow from corollaries 6.1 and 6.2 and (b) follows from Proposition 6.6. Therefore the mapping $\hat{x} \to \hat{f}_{W_{n_0}}(\hat{x}, y)$ is a group homomorphism from $(\hat{H}_1^y, +)$ to $(\mathbb{T}, \cdot)$.

- For every fixed $\hat{x} \in \hat{H}_1$, let $y_1, y_2 \in H_2^{\hat{x}}$ so that $(\hat{x}, y_1), (\hat{x}, y_2) \in \mathrm{D}(W_{n_0})$. It follows from Proposition 6.5 that $(\hat{x}, y_1 + y_2) \in \mathrm{D}(W_{n_0}^-) = \mathrm{D}(W_{n_0+1}) = \mathrm{D}(W_{n_0})$ which implies that $y_1 + y_2 \in H_2^{\hat{x}}$. Hence $H_2^{\hat{x}}$ is a subgroup of $(G_2, +)$. Moreover, we have

$$\begin{aligned}\hat{f}_{W_{n_0}}(\hat{x}, y_1 + y_2) &\overset{(a)}{=} \hat{f}_{W_{n_0+1}}(\hat{x}, y_1 + y_2) = \hat{f}_{W_{n_0}^-}(\hat{x}, y_1 + y_2) \\ &\overset{(b)}{=} \hat{f}_{W_{n_0}}(\hat{x}, y_1) \cdot \hat{f}_{W_{n_0}}(\hat{x}, y_2),\end{aligned}$$

  where (a) follows from corollary 6.1 and (b) follows from Proposition 6.5. Therefore the mapping $y \to \hat{f}_{W_{n_0}}(\hat{x}, y)$ is a group homomorphism from $(H_2^{\hat{x}}, +)$ to $(\mathbb{T}, \cdot)$.

We conclude that $\hat{f}_{W_{n_0}}$ (which is an extension of $\hat{f}_W$) is pseudo-quadratic. □

Proposition 6.7 shows that if polarization $*$-preserves I_1 for W then W must be polarization compatible with respect to the first user.

6.3.2 Polarization Compatibility is Sufficient

For the sake of brevity, we will write "polarization compatible" to denote "polarization compatible with respect to the first user". In this subsection, we show that polarization compatibility is a sufficient condition for the $*$-preservation of I_1.

Lemma 6.9. *If $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$ is polarization compatible then I_1 is preserved for W.*

Proof. Let $F : D \to \mathbb{T}$ be the pseudo-quadratic function of Definition 6.7. Suppose that $y_1, y_2, y_1', y_2' \in G_2$ and $z_1, z_2 \in \mathcal{Z}$ satisfy:

- $y_1 - y_2 = y_1' - y_2'$.

- $y_1, y_1' \in \mathrm{Y}^{z_1}(W)$ and $y_2, y_2' \in \mathrm{Y}^{z_2}(W)$.

For every $\hat{x} \in G_1$, we have:

- If $(\hat{x}, z_1) \notin \hat{\mathrm{XZ}}(W)$ then $\hat{p}_{y_1,z_1}(\hat{x}) = 0$ and $\hat{p}_{y_1',z_1}(\hat{x}) = 0$, so

$$\hat{p}_{y_1,z_1}(\hat{x})\hat{p}_{y_2,z_2}(\hat{x})^* = \hat{p}_{y_1',z_1}(\hat{x})\hat{p}_{y_2',z_2}(\hat{x})^* = 0.$$

- If $(\hat{x}, z_2) \notin \hat{\mathrm{XZ}}(W)$ then $\hat{p}_{y_2,z_2}(\hat{x}) = 0$ and $\hat{p}_{y_2',z_2}(\hat{x}) = 0$, so

$$\hat{p}_{y_1,z_1}(\hat{x})\hat{p}_{y_2,z_2}(\hat{x})^* = \hat{p}_{y_1',z_1}(\hat{x})\hat{p}_{y_2',z_2}(\hat{x})^* = 0.$$

- If $(\hat{x}, z_1) \in \hat{\mathrm{XZ}}(W)$ and $(\hat{x}, z_2) \in \hat{\mathrm{XZ}}(W)$, then

$$\begin{aligned}\hat{p}_{y_1,z_1}(\hat{x})\hat{p}_{y_2,z_2}(\hat{x})^* &= \hat{p}_{y_1',z_1}(\hat{x})F(\hat{x}, y_1 - y_1')\hat{p}_{y_2',z_2}(\hat{x})^*F(\hat{x}, y_2 - y_2')^* \\ &\overset{(a)}{=} \hat{p}_{y_1',z_1}(\hat{x})\hat{p}_{y_2',z_2}(\hat{x})^*,\end{aligned}$$

  where (a) follows from the fact that $y_1 - y_1' = y_2 - y_2'$ and so

$$F(\hat{x}, y_1 - y_1')F(\hat{x}, y_2 - y_2')^* = |F(\hat{x}, y_1 - y_1')|^2 = 1.$$

Therefore, we have $\hat{p}_{y_1,z_1}(\hat{x})\hat{p}_{y_2,z_2}(\hat{x})^* = \hat{p}_{y_1',z_1}(\hat{x})\hat{p}_{y_2',z_2}(\hat{x})^*$ for all $\hat{x} \in G_1$. Lemma 6.5 now implies that I_1 is preserved for W. □

Lemma 6.10. *If $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$ is polarization compatible then W^- and W^+ are polarization compatible as well.*

Proof. See Appendix 6.5.5. □

Proposition 6.8. *If W is polarization compatible then polarization $*$-preserves I_1 for W.*

Proof. Suppose that W is polarization compatible. Using Lemma 6.10, we can show by induction that W^s is polarization compatible for every $s \in \{-,+\}^*$. Lemma 6.9 now implies that I_1 is preserved for W^s for every $s \in \{-,+\}^*$. By applying Lemma 6.1, we deduce that polarization $*$-preserves I_1 for W. □

Propositions 6.7 and 6.8 show that polarization $*$-preserves I_1 for W if and only if W is polarization compatible. This completes the proof of Theorem 6.1.

6.3.3 Special Cases

The characterization found in Theorem 6.1 (i.e., polarization compatibility) takes a simple form in the special case where $G_1 = G_2 = \mathbb{F}_q$ for a prime q:

Proposition 6.9. *Let $W : \mathbb{F}_q \times \mathbb{F}_q \longrightarrow \mathcal{Z}$ be a two-user MAC and let $(X, Y) \xrightarrow{W} Z$. Polarization $*$-preserves I_1 for W if and only if there exists $a \in \mathbb{F}_q$ such that $I(X + aY; Y|Z) = 0$.*

Proof. If polarization $*$-preserves I_1 for W then W is polarization compatible. Let $F : D \to \mathbb{T}$ be the pseudo-quadratic function of Definition 6.7. We have the following:

- If there exists $(\hat{x}, y) \in D$ such that $\hat{x} \neq 0$ and $y \neq 0$ then $D = \mathbb{F}_q \times \mathbb{F}_q$ since D is a pseudo-quadratic domain and since q is prime.
- If for all $(\hat{x}, y) \in D$ we have either $\hat{x} = 0$ or $y = 0$, then $F(\hat{x}, y) = 1$ for every $(\hat{x}, y) \in D$. Hence the mapping $F' : \mathbb{F}_q \times \mathbb{F}_q \to \mathbb{T}$ defined as $F'(\hat{x}, y) = 1$ is an extension of F which preserves the pseudo-quadratic property.

Therefore, we can assume without loss of generality that $D = \mathbb{F}_q \times \mathbb{F}_q$. Now since $F(1,1)^q = F(1, q \cdot 1) = F(1, 0) = 1$, $F(1,1)$ is a q^{th} root of unity. Therefore, there exists $a \in \mathbb{F}_q$ such that $F(1,1) = e^{j2\pi \frac{a}{q}}$.

Fix $z \in \mathcal{Z}$ and $y_1, y_2 \in \mathrm{Y}^z(W)$. For every $\hat{x} \in \mathbb{F}_q$ we have

$$\hat{p}_{y_1,z}(\hat{x}) = \hat{p}_{y_2,z}(\hat{x}) \cdot F(\hat{x}, y_1 - y_2) = \hat{p}_{y_2,z}(\hat{x}) \cdot e^{j2\pi a \frac{(y_1 - y_2)\hat{x}}{q}},$$

which is equivalent to say that for every $x' \in \mathbb{F}_q$, we have

$$p_{y_1,z}(x') = p_{y_2,z}(x' + a(y_1 - y_2)),$$

i.e.,

$$P_{X|Y,Z}(x'|y_1, z) = P_{X|Y,Z}(x' + a(y_1 - y_2)|y_2, z). \tag{6.6}$$

By applying the change of variable $x' = x - ay_1$, we can see that (6.6) is equivalent to

$$\begin{aligned} P_{X+aY|Y,Z}(x|y_1, z) &= P_{X|Y,Z}(x - ay_1|y_1, z) = P_{X|Y,Z}(x'|y_1, z) \\ &= P_{X|Y,Z}(x' + a(y_1 - y_2)|y_2, z) \\ &= P_{X|Y,Z}(x - ay_1 + a(y_1 - y_2)|y_2, z) \\ &= P_{X|Y,Z}(x - ay_2|y_2, z) = P_{X+aY|Y,Z}(x|y_2, z). \end{aligned}$$

This shows that $X + aY$ is conditionally independent of Y given Z, i.e., $I(X + aY; Y|Z) = 0$.

On the other hand, let $W : \mathbb{F}_q \times \mathbb{F}_q \longrightarrow \mathcal{Z}$ be a two-user MAC and let $(X, Y) \xrightarrow{W} Z$. If there exists $a \in \mathbb{F}_q$ such that $I(X + aY; Y|Z) = 0$, then Proposition 6.3 implies that polarization $*$-preserves I_1 for W. □

Corollary 6.3. *Polarization $*$-preserves the symmetric-capacity region for the binary adder channel.*

Proof. Let X and Y be two independent uniform random variables in $\{0, 1\}$, and let $Z = X + Y \in \{0, 1, 2\}$ (where $+$ here denotes addition in $\mathbb{R}$). It is easy to check that $I(X \oplus Y; Y|Z) = I(X \oplus Y; X|Z) = 0$. Therefore, polarization $*$-preserves I_1 and I_2 for W. We conclude that polarization $*$-preserves the symmetric-capacity region for W. □

Remark 6.3. *It may seem promising to try to generalize Proposition 6.9 to the case where $G_1 = \mathbb{F}_q^k$ and $G_2 = \mathbb{F}_q^l$ by considering the condition $I(X + AY; Y|Z) = 0$ for some matrix $A \in \mathbb{F}_q^{k \times l}$. Although this condition is sufficient for the $*$-preservation of I_1 (Proposition 6.3), it turns out that it is not necessary.*

Proposition 6.10. *If $|G_1|$ and $|G_2|$ are co-prime and $(X,Y) \xrightarrow{W} Z$, then polarization $*$-preserves I_1 for W if and only if $I(X;Y|Z) = 0$ (i.e., if and only if the dominant face of $\mathcal{J}(W)$ is a single point).*

Proof. Let $F : D \to \mathbb{T}$ be a pseudo-quadratic function. For every $(\hat{x}, y) \in D$, we have:

- $F(\hat{x}, y)^{|G_1|} = F(|G_1| \cdot \hat{x}, y) = F(0, y) = 1.$
- $F(\hat{x}, y)^{|G_2|} = F(\hat{x}, |G_2| \cdot y) = F(\hat{x}, 0) = 1.$

Therefore, $F(\hat{x}, y)$ is both a $|G_1|^{th}$ root of unity and a $|G_2|^{th}$ root of unity. This shows that $F(\hat{x}, y)$ must be equal to 1 because $|G_1|$ and $|G_2|$ are co-prime. We conclude that every pseudo-quadratic function $F : D \to \mathbb{T}$ must be equal to 1. Therefore, polarization $*$-preserves I_1 for W if and only if $\hat{p}_{y_1,z}(\hat{x}) = \hat{p}_{y_2,z}(\hat{x})$ for every $(\hat{x}, z) \in \hat{\mathrm{X}}\mathrm{Z}(W)$ and every $y_1, y_2 \in \mathrm{Y}^z(W)$.

Now since $\hat{p}_{y_1,z}(\hat{x}) = \hat{p}_{y_2,z}(\hat{x}) = 0$ for every $\hat{x} \notin \hat{\mathrm{X}}^z(W)$ and every $y_1, y_2 \in \mathrm{Y}^z(W)$, we conclude that polarization $*$-preserves I_1 for W if and only if $\hat{p}_{y_1,z}(\hat{x}) = \hat{p}_{y_2,z}(\hat{x})$ for every $(\hat{x}, z) \in G_1 \times \mathcal{Z}$ and every $y_1, y_2 \in \mathrm{Y}^z(W)$. This is equivalent to say that $p_{y_1,z}(x) = p_{y_2,z}(x)$ for every $(x, z) \in G_1 \times \mathcal{Z}$ and every $y_1, y_2 \in \mathrm{Y}^z(W)$. This just means that X and Y are conditionally independent given Z. □

6.4 Generalization to Multiple User MACs

Definition 6.11. *Let $W : G_1 \times \cdots \times G_m \longrightarrow \mathcal{Z}$ be an m-user MAC. For every $S \subset \{1, \ldots, m\}$, we define the two-user MAC $W_S : G_S \times G_{S^c} \longrightarrow \mathcal{Z}$ as $W_S(y|x_S, x_{S^c}) = W(y|x_1, \ldots, x_m)$.*

Remark 6.4. *It is easy to see that for every $s \in \{-,+\}^*$ and every $S \subset \{1, \ldots, m\}$, we have $(W^s)_S = (W_S)^s$. Therefore, polarization $*$-preserves I_S for W if and only if polarization $*$-preserves I_1 for W_S.*

Theorem 6.2. *Let $W : G_1 \times \cdots \times G_m \longrightarrow \mathcal{Z}$ be an m-user MAC. Polarization $*$-preserves I_S for W if and only if W_S is polarization compatible.*

Proof. Direct corollary of Theorem 6.1 and Remark 6.4. □

6.5 Appendix

6.5.1 Proof of Proposition 6.4

We need the following lemma:

Lemma 6.11. *Let $(G,+)$ be an Abelian group and let $\hat{f} : G \to \mathbb{T}$ be a group homomorphism from $(G,+)$ to $(\mathbb{T},\cdot)$. There exists $x_f \in G$ satisfying:*

- *$\hat{f}(\hat{x}) = e^{j2\pi\langle \hat{x}, x_f \rangle}$ for every $\hat{x} \in G$.*
- *If $n > 0$ is such that $\hat{f}(\hat{x})^n = 1$ for every $\hat{x} \in G$, then $nx_f = 0$.*

Proof. Let $N_1, \ldots, N_k > 0$ be k integers such that $G = \mathbb{Z}_{N_1} \times \cdots \times \mathbb{Z}_{N_k}$. For every $1 \leq i \leq k$, let $e_i = (0, \ldots, 0, 1, 0, \ldots, 0) \in G$ be the element of G whose j^{th} coordinate is 1 if $j = i$ and 0 otherwise.

Since $N_i e_i = 0$, we have $\hat{f}(e_i)^{N_i} = \hat{f}(N_i e_i) = \hat{f}(0) = 1$. Therefore, $\hat{f}(e_i)$ is an N_i^{th} root of unity, so there exists $0 \leq x_i < N_i$ such that $\hat{f}(e_i) = e^{\frac{j2\pi x_i}{N_i}}$.

Let $x_f := (x_1, \ldots, x_k) \in G$. For every $\hat{x} = (\hat{x}_1, \ldots \hat{x}_k) \in G$ we have:

$$\hat{f}(\hat{x}) = \hat{f}\left(\sum_{i=1}^{k} \hat{x}_i e_i\right) = \prod_{i=1}^{k} \hat{f}(e_i)^{\hat{x}_i} = \prod_{i=1}^{k} \left(e^{\frac{j2\pi x_i}{N_i}}\right)^{\hat{x}_i} = e^{\sum_{i=1}^{k} \frac{j2\pi \hat{x}_i x_i}{N_i}} = e^{j2\pi\langle \hat{x}, x_f \rangle}.$$

If $n > 0$ is such that $\hat{f}(\hat{x})^n = 1$ for every $\hat{x} \in G$, then $e^{\frac{j2\pi n x_i}{N_i}} = \hat{f}(e_i)^n = 1$ for every $1 \leq i \leq k$. This means that N_i divides nx_i for every $1 \leq i \leq k$. Therefore,

$$nx_f = (nx_1 \bmod N_1, \ldots, nx_k \bmod N_k) = 0.$$

□

Now we are ready to prove Proposition 6.4.

Since we have shown the necessary condition in the discussion before the statement of Proposition 6.4, we only need to show the sufficient condition.

Let $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$ be a two-user MAC, and assume that there exists a subgroup H_2 of G_2 and a pseudo-quadratic function $F : G_1 \times H_2 \to \mathbb{T}$ satisfying:

- $\mathrm{D}(W) \subset G_1 \times H_2$.
- For every $(\hat{x}, z) \in \hat{\mathrm{X}}\mathrm{Z}(W)$ and every $y_1, y_2 \in \mathrm{Y}^z(W)$, we have $\hat{p}_{y_1,z}(\hat{x}) = F(\hat{x}, y_1 - y_2)\hat{p}_{y_2,z}(\hat{x})$.

Since $(H_2, +)$ is an Abelian group, it is isomorphic to the product of cyclic groups. Let $N'_1, \ldots, N'_{k'} > 0$ be k' integers such that H_2 is isomorphic to $\mathbb{Z}_{N'_1} \times \cdots \times \mathbb{Z}_{N'_{k'}}$. Because of this isomorphism, we can find k' elements $e'_1, \ldots, e'_{k'} \in H_2$ such that:

- e'_i is of order N'_i for every $1 \leq i \leq k'$.
- For every $y \in H_2$, there exist unique integers $0 \leq y_1 < N'_1, \ldots, 0 \leq y_{k'} < N'_{k'}$ such that $y = \sum_{i=1}^{k'} y_i e'_i$.

For every $1 \leq i \leq k'$, the mapping $\hat{x} \to F(\hat{x}, e'_i)$ is a group homomorphism from $(G_1, +)$ to $(\mathbb{T}, \cdot)$. Lemma 6.11 shows that there exists $f_i \in G_1$ such that $F(\hat{x}, e'_i) = e^{j2\pi\langle \hat{x}, f_i \rangle}$ for every $\hat{x} \in G_1$. Moreover, for every $1 \leq i \leq k'$, we have $F(\hat{x}, e'_i)^{N'_i} = F(\hat{x}, N'_i e'_i) = F(\hat{x}, 0) = 1$ for every $\hat{x} \in G_1$, hence $N'_i f_i = 0$.

For every $y \in H_2$, define $f(y) = \sum_{i=1}^{k'} y_i f_i$, where $0 \leq y_1 < N'_1, \ldots, 0 \leq y_{k'} < N'_{k'}$ satisfy $y = \sum_{i=1}^{k'} y_i e'_i$. We can show that f is a group homomorphism from $(H_2, +)$ to

$(G_1,+)$: Let $y,y'\in H_2$ and let $0\le y_1,y_1',y_1''\le N_1'$, …, $0\le y_{k'},y_{k'}',y_{k'}''\le N_{k'}'$ be such that $y=\sum_{i=1}^{k'} y_i e_i'$, $y'=\sum_{i=1}^{k'} y_i' e_i'$ and $y+y'=\sum_{i=1}^{k'} y_i'' e_i'$. We have:

$$0=y+y'-y-y'=\sum_{i=1}^{k'}(y_i''-y_i-y_i')e_i' \overset{(a)}{=} \sum_{i=1}^{k'}(y_i''-y_i-y_i' \bmod N_i')e_i',$$

where (a) follows from the fact that e_i' is of order N_i'. Thus, $y_i''=y_i+y_i' \bmod N_i'$ for every $1\le i\le k'$. Therefore,

$$f(y+y')=\sum_{i=1}^{k'} y_i'' f_i \overset{(b)}{=} \sum_{i=1}^{k'}(y_i+y_i')f_i=\left(\sum_{i=1}^{k'} y_i f_i\right)+\left(\sum_{i=1}^{k'} y_i' f_i\right)=f(y)+f(y'),$$

where (b) follows from the fact that $N_i' f_i=0$ and $y_i''=y_i+y_i' \bmod N_i'$ for every $1\le i\le k'$. We conclude that f is a group homomorphism from $(H_2,+)$ to $(G_1,+)$.

On the other hand, for every $\hat{x}\in G_1$, we have:

$$\begin{aligned}F(\hat{x},y)&=F\left(\hat{x},\sum_{i=1}^{k'} y_i e_i'\right)=\prod_{i=1}^{k'} F\left(\hat{x},e_i'\right)^{y_i}=\prod_{i=1}^{k'}\left(e^{j2\pi\langle \hat{x},f_i\rangle}\right)^{y_i}\\&=e^{\sum_{i=1}^{k'} j2\pi y_i\langle \hat{x},f_i\rangle}=e^{j2\pi\left\langle \hat{x},\sum_{i=1}^{k'} y_i f_i\right\rangle}=e^{j2\pi\langle \hat{x},f(y)\rangle}.\end{aligned}$$

Let $z\in\mathcal{Z}$ and $y_1,y_2\in \mathrm{Y}^z(W)$. For every $\hat{x}\in G_1$, we have:

- If $\hat{x}\notin \hat{\mathrm{X}}^z(W)$, we have $\hat{p}_{y_1,z}(\hat{x})=\hat{p}_{y_2,z}(\hat{x})=0$, hence
$$\hat{p}_{y_1,z}(\hat{x})=\hat{p}_{y_2,z}(\hat{x})e^{j2\pi\langle \hat{x},f(y_1-y_2)\rangle}.$$
- If $\hat{x}\in \hat{\mathrm{X}}^z(W)$, we have
$$\hat{p}_{y_1,z}(\hat{x})=\hat{p}_{y_2,z}(\hat{x})F(\hat{x},y_1-y_2)=\hat{p}_{y_2,z}(\hat{x})e^{j2\pi\langle \hat{x},f(y_1-y_2)\rangle}.$$

We conclude that

$$\begin{aligned}&\hat{p}_{y_1,z}(\hat{x})=\hat{p}_{y_2,z}(\hat{x})e^{j2\pi\langle \hat{x},f(y_1-y_2)\rangle}\ \ \forall \hat{x}\in G_1\\\Leftrightarrow\ &p_{y_1,z}(x)=p_{y_2,z}(x+f(y_1-y_2))\ \ \forall x\in G_1.\end{aligned}$$

Lemma 6.4 now shows that W is homomorphic-independent.

6.5.2 Proof of Lemma 6.6

We need the following two lemmas:

Lemma 6.12. *Suppose that I_1 is $*^-$ preserved for W. Fix $n>0$ and let $(U_i,V_i)_{0\le i<2^n}$ be a sequence of random pairs which are independent and uniformly distributed in $G_1\times G_2$. Let*

$$F=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}.$$

Define $X_0^{2^n-1}=F^{\otimes n}\cdot U_0^{2^n-1}$ and $Y_0^{2^n-1}=F^{\otimes n}\cdot V_0^{2^n-1}$, and for each $0\le i<2^n$ let $(X_i,Y_i)\xrightarrow{W} Z_i$. We have the following:

- *The MAC* $(U_0, V_0) \longrightarrow Z_0^{2^n-1}$ *is equivalent to* $W^{[n]^-}$.
- $I(U_0; V_1^{2^n-1} | Z_0^{2^n-1} V_0) = 0$.

Proof. We will show the lemma by induction on $n > 0$. For $n = 1$, the claim follows from Remark 6.1 and from the fact that I_1 is preserved for W if and only if $I(U_0; V_1 | Z_0 Z_1 V_0) = 0$ (see (6.1)).

Now let $n > 1$ and suppose that the claim is true for $n - 1$. Let $N = 2^{n-1}$. We have $X_0^{2^n-1} = F^{\otimes n} \cdot U_0^{2^n-1}$ and $Y_0^{2^n-1} = F^{\otimes n} \cdot V_0^{2^n-1}$, i.e., $X_0^{2N-1} = F^{\otimes n} \cdot U_0^{2N-1}$ and $Y_0^{2N-1} = F^{\otimes n} \cdot V_0^{2N-1}$. Therefore, we have:

$$X_0^{N-1} = F^{\otimes(n-1)} \cdot (U_0^{N-1} + U_N^{2N-1}),$$

$$X_N^{2N-1} = F^{\otimes(n-1)} \cdot U_N^{2N-1},$$

$$Y_0^{N-1} = F^{\otimes(n-1)} \cdot (V_0^{N-1} + V_N^{2N-1}),$$

and

$$Y_N^{2N-1} = F^{\otimes(n-1)} \cdot V_N^{2N-1}.$$

This means that

$$(U_0^{N-1} + U_N^{2N-1}, V_0^{N-1} + V_N^{2N-1}, Z_0^{N-1})$$

and

$$(U_N^{2N-1}, V_N^{2N-1}, Z_N^{2N-1})$$

satisfy the conditions of the induction hypothesis. Therefore,

- $I(U_0 + U_N; V_1^{N-1} + V_{N+1}^{2N-1} | Z_0^{N-1}, V_0 + V_N) = 0$.
- $I(U_N; V_{N+1}^{2N-1} | Z_N^{2N-1}, V_N) = 0$.

Moreover, since

$$(U_0^{N-1} + U_N^{2N-1}, V_0^{N-1} + V_N^{2N-1}, Z_0^{N-1})$$

is independent of

$$(U_N^{2N-1}, V_N^{2N-1}, Z_N^{2N-1}),$$

we can combine the above two equations to get:

$$I(U_0 + U_N, U_N; V_1^{N-1} + V_{N+1}^{2N-1}, V_{N+1}^{2N-1} | Z_0^{2N-1}, V_0 + V_N, V_N) = 0,$$

which can be rewritten as

$$I(U_0 U_N; V_1^{N-1} V_{N+1}^{2N-1} | Z_0^{2N-1} V_0 V_N) = 0. \tag{6.7}$$

On the other hand, it also follows from the induction hypothesis that:

- The MAC $(U_0 + U_N, V_0 + V_N) \longrightarrow Z_0^{N-1}$ is equivalent to $W^{[n-1]^-}$.
- The MAC $(U_N, V_N) \longrightarrow Z_N^{2N-1}$ is equivalent to $W^{[n-1]^-}$.

This implies that the MAC $(U_0, V_0) \longrightarrow Z_0^{2N-1}$ is equivalent to $W^{[n]^-}$. Now since I_1 is $*^-$ preserved for W, I_1 must be preserved for $W^{[n-1]^-}$. Therefore,

$$I(U_0; V_N | Z_0^{2N-1} V_0) = I(U_0; V_N | Z_0^{N-1} Z_N^{2N-1} V_0) \stackrel{(a)}{=} 0, \tag{6.8}$$

where (a) follows from (6.1). We conclude that:

$$\begin{aligned} I(U_0; V_1^{2N-1} | Z_0^{2N-1} V_0) &= I(U_0; V_N | Z_0^{2N-1} V_0) + I(U_0; V_1^{N-1} V_{N+1}^{2N-1} | Z_0^{2N-1} V_0 V_N) \\ &\leq I(U_0; V_N | Z_0^{2N-1} V_0) + I(U_0 U_N; V_1^{N-1} V_{N+1}^{2N-1} | Z_0^{2N-1} V_0 V_N) \\ &\stackrel{(b)}{=} 0, \end{aligned}$$

where (b) follows from (6.7) and (6.8). □

Lemma 6.13. *For every $n > 0$, if $X_0^{2^n-1} = F^{\otimes n} U_0^{2^n-1}$, then $U_0 = \sum_{i=0}^{2^n-1} (-1)^{|i|_b} X_i$, where $|i|_b$ is the number of ones in the binary expansion of i.*

Proof. We will show the lemma by induction on $n > 0$. For $n = 1$, the fact that $X_0^1 = F^{\otimes 1} \cdot U_0^1 = F \cdot U_0^1$ implies that $X_0 = U_0 + U_1$ and $X_1 = U_1$. Therefore $U_0 = X_0 - X_1 = \sum_{i=0}^{1} (-1)^{|i|_b} X_i$.

Now let $n > 1$ and suppose that the claim is true for $n - 1$. Let $N = 2^{n-1}$. The fact that $X_0^{2N-1} = F^{\otimes n} \cdot U_0^{2N-1}$ implies that:

- $X_0^{N-1} = F^{\otimes(n-1)} \cdot (U_0^{N-1} + U_N^{2N-1})$.
- $X_N^{2N-1} = F^{\otimes(n-1)} \cdot U_N^{2N-1}$.

We can apply the induction hypothesis to get:

- $U_0 + U_N = \sum_{i=0}^{N-1} (-1)^{|i|_b} X_i$.
- $U_N = \sum_{i=0}^{N-1} (-1)^{|i|_b} X_{i+N}$.

Therefore,

$$\begin{aligned} U_0 &= \sum_{i=0}^{N-1} (-1)^{|i|_b} X_i - \sum_{i=0}^{N-1} (-1)^{|i|_b} X_{i+N} = \sum_{i=0}^{N-1} (-1)^{|i|_b} X_i + \sum_{i=0}^{N-1} (-1)^{1+|i|_b} X_{i+N} \\ &= \sum_{i=0}^{N-1} (-1)^{|i|_b} X_i + \sum_{i=N}^{2N-1} (-1)^{1+|i-N|_b} X_i \stackrel{(a)}{=} \sum_{i=0}^{N-1} (-1)^{|i|_b} X_i + \sum_{i=N}^{2N-1} (-1)^{|i|_b} X_i \\ &= \sum_{i=0}^{2N-1} (-1)^{|i|_b} X_i, \end{aligned}$$

where (a) follows from the fact that for $2^n = N \leq i < 2N = 2^{n+1}$, we have $|i - N|_b = |i - 2^n|_b = |i|_b - 1$. □

We are now ready to prove Lemma 6.6:

Let W be a two-user MAC such that I_1 is $*^-$ preserved for W. Let $n > 0$, $y_1, \ldots, y_{2^n}, y'_1, \ldots, y'_{2^n} \in G_2$ and $z_1, \ldots, z_{2^n} \in \mathcal{Z}$ be such that

- $\sum_{i=1}^{2^n} y_i = \sum_{i=1}^{2^n} y'_i$,
- $y_1 \in \mathrm{Y}^{z_1}(W), \ldots, y_{2^n} \in \mathrm{Y}^{z_{2^n}}(W)$, and
- $y'_1 \in \mathrm{Y}^{z_1}(W), \ldots, y'_{2^n} \in \mathrm{Y}^{z_{2^n}}(W)$.

Now fix $\hat{x} \in G_1$. If $\hat{p}_{y,z}(\hat{x}) = 0$ for every $(y, z) \in \mathrm{YZ}(W)$, then we clearly have

$$\prod_{i=1}^{2^n} \hat{p}_{y_i,z_i}(\hat{x}) = \prod_{i=1}^{2^n} \hat{p}_{y'_i,z_i}(\hat{x}).$$

Therefore, we can assume without loss of generality that there exists $(y, z) \in \mathrm{YZ}(W)$ which satisfies $\hat{p}_{y,z}(\hat{x}) \neq 0$.

Let $U_0^{2^{n+1}-1}$, $V_0^{2^{n+1}-1}$, $X_0^{2^{n+1}-1}$, $Y_0^{2^{n+1}-1}$ and $Z_0^{2^{n+1}-1}$ be as in Lemma 6.12 and let $N = 2^{n+1}$ so that we have

$$I(U_0; V_1^{N-1} | Z_0^{N-1} V_0) = 0. \tag{6.9}$$

Since $X_0^{N-1} = F^{\otimes(n+1)} \cdot U_0^{N-1}$ and $Y_0^{N-1} = F^{\otimes(n+1)} \cdot V_0^{N-1}$, Lemma 6.13 implies that

$$U_0 = \sum_{i=0}^{N-1} (-1)^{|i|_b} X_i \text{ and } V_0 = \sum_{i=0}^{N-1} (-1)^{|i|_b} Y_i. \tag{6.10}$$

Notice that $\left|\left\{0 \leq i < N = 2^{n+1} : |i|_b \equiv 0 \bmod 2\right\}\right| = \left|\left\{0 \leq i < N = 2^{n+1} : |i|_b \equiv 1 \bmod 2\right\}\right| = 2^n$. Let $k_1, \ldots, k_{2^n}$ be the elements of $\left\{0 \leq i < N : |i|_b \equiv 0 \bmod 2\right\}$ and let $l_1, \ldots, l_{2^n}$ be the elements of $\left\{0 \leq i < N : |i|_b \equiv 1 \bmod 2\right\}$.

Define $(\tilde{y}_i, \tilde{y}'_i, \tilde{z}_i)_{0 \leq i < N}$ as follows:

- For every $1 \leq i \leq 2^n$, let $\tilde{y}_{k_i} = y_i$, $\tilde{y}'_{k_i} = y'_i$ and $\tilde{z}_{k_i} = z_i$.
- For every $1 \leq i \leq 2^n$, let $\tilde{y}_{l_i} = \tilde{y}'_{l_i} = y$ and $\tilde{z}_{l_i} = z$ (where (y, z) is any fixed pair in $\mathrm{YZ}(W)$ satisfying $\hat{p}_{y,z}(\hat{x}) \neq 0$).

Now let $\tilde{v}_0^{N-1} = (F^{\otimes(n+1)})^{-1} \cdot \tilde{y}_0^{N-1}$ and $\tilde{v}_0'^{N-1} = (F^{\otimes(n+1)})^{-1} \cdot \tilde{y}_0'^{N-1}$. We have

$$\begin{aligned}
\tilde{v}_0 &\overset{(a)}{=} \sum_{i=0}^{N-1} (-1)^{|i|_b} \tilde{y}_i = \sum_{i=1}^{2^n} (\tilde{y}_{k_i} - \tilde{y}_{l_i}) = \left(\sum_{i=1}^{2^n} y_i\right) - 2^n y \\
&\overset{(b)}{=} \left(\sum_{i=1}^{2^n} y'_i\right) - 2^n y = \sum_{i=1}^{2^n} (\tilde{y}'_{k_i} - \tilde{y}'_{l_i}) = \sum_{i=0}^{N-1} (-1)^{|i|_b} \tilde{y}'_i \overset{(c)}{=} \tilde{v}'_0,
\end{aligned}$$

where (a) and (c) follow from Lemma 6.13. (b) follows from the fact that $\sum_{i=1}^{2^n} y_i = \sum_{i=1}^{2^n} y'_i$. Therefore,

$$(\tilde{v}_0, \tilde{z}_0^{N-1}) = (\tilde{v}'_0, \tilde{z}_0^{N-1}). \tag{6.11}$$

On the other hand, since $\tilde{y}_i \in \mathrm{Y}^{\tilde{z}_i}(W)$ for every $0 \le i < N$, we have

$$\begin{aligned} P_{V_0,V_1^{N-1},Z_0^{N-1}}(\tilde{v}_0, \tilde{v}_1^{N-1}, \tilde{z}_0^{N-1}) &= P_{V_0^{N-1},Z_0^{N-1}}(\tilde{v}_0^{N-1}, \tilde{z}_0^{N-1}) \\ &= P_{Y_0^{N-1},Z_0^{N-1}}(\tilde{y}_0^{N-1}, \tilde{z}_0^{N-1}) > 0. \end{aligned} \tag{6.12}$$

Similarly, since $\tilde{y}'_i \in \mathrm{Y}^{\tilde{z}_i}(W)$ for every $0 \le i < N$, we have

$$\begin{aligned} P_{V_0,V_1^{N-1},Z_0^{N-1}}(\tilde{v}'_0, \tilde{v}_1'^{N-1}, \tilde{z}_0^{N-1}) &= P_{V_0^{N-1},Z_0^{N-1}}\big(\tilde{v}_0'^{N-1}, \tilde{z}_0^{N-1}\big) \\ &= P_{Y_0^{N-1},Z_0^{N-1}}(\tilde{y}_0'^{N-1}, \tilde{z}_0^{N-1}) > 0. \end{aligned} \tag{6.13}$$

Equation (6.9) implies that given (V_0, Z_0^{N-1}), U_0 is conditionally independent of V_1^{N-1}. Equations (6.11), (6.12) and (6.13) now imply that for every $u_0 \in G_1$, we have:

$$\begin{aligned} &P_{U_0|V_1^{N-1},V_0,Z_0^{N-1}}(u_0|\tilde{v}_1^{N-1}, \tilde{v}_0, \tilde{z}_0^{N-1}) = P_{U_0|V_1^{N-1},V_0,Z_0^{N-1}}(u_0|\tilde{v}_1'^{N-1}, \tilde{v}'_0, \tilde{z}_0^{N-1}) \\ &\quad\Leftrightarrow\ P_{U_0|V_0^{N-1},Z_0^{N-1}}(u_0|\tilde{v}_0^{N-1}, \tilde{z}_0^{N-1}) = P_{U_0|V_0^{N-1},Z_0^{N-1}}(u_0|\tilde{v}_0'^{N-1}, \tilde{z}_0^{N-1}) \\ &\quad\Leftrightarrow\ P_{U_0|Y_0^{N-1},Z_0^{N-1}}(u_0|\tilde{y}_0^{N-1}, \tilde{z}_0^{N-1}) = P_{U_0|Y_0^{N-1},Z_0^{N-1}}(u_0|\tilde{y}_0'^{N-1}, \tilde{z}_0^{N-1}) \\ &\quad\overset{(a)}{\Leftrightarrow} \sum_{\substack{\tilde{x}_0^{N-1}\in G_1^N:\\ \sum_{i=0}^{N-1}(-1)^{|i|_b}\tilde{x}_i = u_0}} \prod_{i=0}^{N-1} P_{X_i|Y_i,Z_1}(\tilde{x}_i|\tilde{y}_i, \tilde{z}_i) = \sum_{\substack{\tilde{x}_0^{N-1}\in G_1^N:\\ \sum_{i=0}^{N-1}(-1)^{|i|_b}\tilde{x}_i = u_0}} \prod_{i=0}^{N-1} P_{X_i|Y_i,Z_1}(\tilde{x}_i|\tilde{y}'_i, \tilde{z}_i) \\ &\quad\overset{(b)}{\Leftrightarrow} \sum_{\substack{x_1^N\in G_1^N:\\ \sum_{i=1}^{2^n} x_i - \sum_{i=2^n+1}^{N} x_i = u_0}} \prod_{i=1}^{2^n} p_{y_i,z_i}(x_i) \prod_{i=2^n+1}^{N} p_{y,z}(x_i) \\ &\qquad\qquad = \sum_{\substack{x_1^N\in G_1^N:\\ \sum_{i=1}^{2^n} x_i - \sum_{i=2^n+1}^{N} x_i = u_0}} \prod_{i=1}^{2^n} p_{y'_i,z_i}(x_i) \prod_{i=2^n+1}^{N} p_{y,z}(x_i), \end{aligned} \tag{6.14}$$

where (a) follows from (6.10) and (b) follows from the following change of variables:

$$x_i = \begin{cases} \tilde{x}_{k_i} & \text{if } 1 \le i \le 2^n, \\ \tilde{x}_{l_{i-2^n}} & \text{if } 2^n \le i \le 2^{n+1} = N. \end{cases}$$

Now notice that the left hand side of (6.14) is the convolution of $(p_{y_i,z_i})_{1\le i\le 2^n}$ with 2^n copies of $\tilde{p}_{y,z}$ (where we define $\tilde{p}_{y,z}(x) = p_{y,z}(-x)$). Likewise, the right hand

side of (6.14) is the convolution of $(p_{y'_i,z_i})_{1\le i\le 2^n}$ with 2^n copies of $\tilde{p}_{y,z}$. By applying the DFT on (6.14), we get:

$$\prod_{i=1}^{2^n}\hat{p}_{y_i,z_i}(\hat{u}_0)\prod_{i=2^n+1}^{N}\hat{p}_{y,z}(\hat{u}_0)^*=\prod_{i=1}^{2^n}\hat{p}_{y'_i,z_i}(\hat{u}_0)\prod_{i=2^n+1}^{N}\hat{p}_{y,z}(\hat{u}_0)^*,\ \ \forall\hat{u}_0\in G_1.$$

In particular,

$$\prod_{i=1}^{2^n}\hat{p}_{y_i,z_i}(\hat{x})\prod_{i=2^n+1}^{N}\hat{p}_{y,z}(\hat{x})^*=\prod_{i=1}^{2^n}\hat{p}_{y'_i,z_i}(\hat{x})\prod_{i=2^n+1}^{N}\hat{p}_{y,z}(\hat{x})^*.$$

Now since $\hat{p}_{y,z}(\hat{x})\neq 0$, we conclude that

$$\prod_{i=1}^{2^n}\hat{p}_{y_i,z_i}(\hat{x})=\prod_{i=1}^{2^n}\hat{p}_{y'_i,z_i}(\hat{x}).$$

6.5.3 Proof of Proposition 6.5

We need the following lemmas.

Lemma 6.14. *For every two-user MAC $W:G_1\times G_2\longrightarrow\mathcal{Z}$ and every $z_1,z_2\in\mathcal{Z}$, we have:*

$$\mathrm{Y}^{(z_1,z_2)}(W^-)=\mathrm{Y}^{z_1}(W)-\mathrm{Y}^{z_2}(W)=\big\{y_1-y_2:\ y_1\in\mathrm{Y}^{z_1}(W),y_2\in\mathrm{Y}^{z_2}(W)\big\}.$$

Proof. Let $U_1,U_2,V_1,V_2,X_1,X_2,Y_1,Y_2,Z_1,Z_2$ be as in Remark 6.1. For every $v_1\in G_2$ and every $z_1,z_2\in\mathcal{Z}$, we have:

$$\begin{aligned}P_{V_1,Z_1,Z_2}(v_1,z_1,z_2)&=\sum_{\substack{y_1,y_2\in G_2:\\ v_1=y_1-y_2}}P_{Y_1,Y_2,Z_1,Z_2}(y_1,y_2,z_1,z_2)\\&=\sum_{\substack{y_1,y_2\in G_2:\\ v_1=y_1-y_2}}P_{Y_1,Z_1}(y_1,z_1)P_{Y_2,Z_2}(y_2,z_2).\end{aligned}$$

Therefore, $v_1\in\mathrm{Y}^{(z_1,z_2)}(W^-)$ if and only if there exist $y_1,y_2\in G_2$ such that $y_1\in\mathrm{Y}^{z_1}(W)$, $y_2\in\mathrm{Y}^{z_2}(W)$ and $v_1=y_1-y_2$. Hence,

$$\mathrm{Y}^{(z_1,z_2)}(W^-)=\big\{y_1-y_2:\ y_1\in\mathrm{Y}^{z_1}(W),y_2\in\mathrm{Y}^{z_2}(W)\big\}.$$

□

Lemma 6.15. *Let $U_1,U_2,V_1,V_2,X_1,X_2,Y_1,Y_2,Z_1,Z_2$ be as in Remark 6.1. For every $z_1,z_2\in\mathcal{Z}$, every $v_1\in\mathrm{Y}^{(z_1,z_2)}(W^-)$ and every $\hat{u}_1\in G_1$, we have:*

$$\hat{p}_{v_1,(z_1,z_2),W^-}(\hat{u}_1)=\sum_{\substack{v_2\in\mathrm{Y}^{z_2}(W):\\ v_1+v_2\in\mathrm{Y}^{z_1}(W)}}\frac{P_{Y_1|Z_1}(v_1+v_2|z_1)P_{Y_2|Z_2}(v_2|z_2)}{P_{V_1|Z_1,Z_2}(v_1|z_1,z_2)}\hat{p}_{v_1+v_2,z_1}(\hat{u}_1)\cdot\hat{p}_{v_2,z_2}(\hat{u}_1)^*.\quad(6.15)$$

Proof. Fix $z_1, z_2 \in \mathcal{Z}$ and $v_1 \in \mathrm{Y}^{(z_1,z_2)}(W^-)$, and let $\beta = P_{V_1|Z_1,Z_2}(v_1|z_1,z_2) > 0$. For every $u_1 \in G_1$, we have:

$$\begin{aligned}
&p_{v_1,(z_1,z_2),W^-}(u_1) \\
&= P_{U_1|V_1,Z_1,Z_2}(u_1|v_1,z_1,z_2) = \frac{1}{\beta} P_{U_1,V_1|Z_1,Z_2}(u_1,v_1|z_1,z_2) \\
&= \frac{1}{\beta} \sum_{\substack{u_2 \in G_1, \\ v_2 \in G_2}} P_{U_1,U_2,V_1,V_2|Z_1,Z_2}(u_1,u_2,v_1,v_2|z_1,z_2) \\
&= \frac{1}{\beta} \sum_{\substack{u_2 \in G_1, \\ v_2 \in G_2}} P_{X_1,X_2,Y_1,Y_2|Z_1,Z_2}(u_1+u_2,u_2,v_1+v_2,v_2|z_1,z_2) \\
&= \frac{1}{\beta} \sum_{v_2 \in G_2} \sum_{u_2 \in G_1} P_{X_1,Y_1|Z_1}(u_1+u_2,v_1+v_2|z_1) P_{X_2,Y_2|Z_2}(u_2,v_2|z_2) \\
&= \frac{1}{\beta} \sum_{\substack{v_2 \in \mathrm{Y}^{z_2}(W): \\ v_1+v_2 \in \mathrm{Y}^{z_1}(W)}} \sum_{u_2 \in G_1} P_{X_1,Y_1|Z_1}(u_1+u_2,v_1+v_2|z_1) P_{X_2,Y_2|Z_2}(u_2,v_2|z_2) \\
&= \frac{1}{\beta} \sum_{\substack{v_2 \in \mathrm{Y}^{z_2}(W): \\ v_1+v_2 \in \mathrm{Y}^{z_1}(W)}} P_{Y_1|Z_1}(v_1+v_2|z_1) P_{Y_2|Z_2}(v_2|z_2) \sum_{u_2 \in G_1} p_{v_1+v_2,z_1}(u_1+u_2) p_{v_2,z_2}(u_2) \\
&= \frac{1}{\beta} \sum_{\substack{v_2 \in \mathrm{Y}^{z_2}(W): \\ v_1+v_2 \in \mathrm{Y}^{z_1}(W)}} P_{Y_1|Z_1}(v_1+v_2|z_1) P_{Y_2|Z_2}(v_2|z_2) (p_{v_1+v_2,z_1} * \tilde{p}_{v_2,z_2})(u_1),
\end{aligned}$$

where we define $\tilde{p}_{v_2,z_2}(x) = p_{v_2,z_2}(-x)$ for every $x \in G_1$. Therefore, for every $\hat{u}_1 \in G_1$, we have:

$$\hat{p}_{v_1,(z_1,z_2),W^-}(\hat{u}_1) = \sum_{\substack{v_2 \in \mathrm{Y}^{z_2}(W): \\ v_1+v_2 \in \mathrm{Y}^{z_1}(W)}} \frac{P_{Y_1|Z_1}(v_1+v_2|z_1) P_{Y_2|Z_2}(v_2|z_2)}{P_{V_1|Z_1,Z_2}(v_1|z_1,z_2)} \hat{p}_{v_1+v_2,z_1}(\hat{u}_1) \cdot \hat{p}_{v_2,z_2}(\hat{u}_1)^*.$$

□

Lemma 6.16. *If I_1 is $*^-$ preserved for W, then*

$$\mathrm{D}(W^-) \subset \{(\hat{x}, y_1+y_2) : (\hat{x},y_1), (\hat{x},y_2) \in \mathrm{D}(W)\}.$$

Proof. Let $U_1, U_2, V_1, V_2, X_1, X_2, Y_1, Y_2, Z_1, Z_2$ be as in Remark 6.1. Let $(\hat{u}_1, v_1) \in \mathrm{D}(W^-)$. There exists $z^- = (z_1, z_2) \in \mathcal{Z}^2$ such that $(\hat{u}_1, v_1) \in \mathrm{D}^{z^-}(W^-)$, i.e., $\hat{u}_1 \in \hat{\mathrm{X}}^{z^-}(W^-)$ and $v_1 \in \Delta\mathrm{Y}^{z^-}(W^-)$. This implies the existence of $v_1', v_1'' \in \mathrm{Y}^{z^-}(W^-)$ such that $v_1 = v_1' - v_1''$. Since $\hat{u}_1 \in \hat{\mathrm{X}}^{z^-}(W^-)$, Lemma 6.7 shows that $\hat{p}_{v_1',z^-,W^-}(\hat{u}_1) \neq 0$ and $\hat{p}_{v_1'',z^-,W^-}(\hat{u}_1) \neq 0$. From (6.15), we have:

$$\hat{p}_{v_1',z^-,W^-}(\hat{u}_1) = \sum_{\substack{v_2' \in \mathrm{Y}^{z_2}(W): \\ v_1'+v_2' \in \mathrm{Y}^{z_1}(W)}} \frac{P_{Y_1|Z_1}(v_1'+v_2'|z_1) P_{Y_2|Z_2}(v_2'|z_2)}{P_{V_1|Z_1,Z_2}(v_1'|z_1,z_2)} \hat{p}_{v_1'+v_2',z_1}(\hat{u}_1) \cdot \hat{p}_{v_2',z_2}(\hat{u}_1)^*.$$

Since $\hat{p}_{v_1',z^-,W^-}(\hat{u}_1) \neq 0$, the terms in the above sum cannot all be zero. Therefore, there exists $v_2' \in \mathrm{Y}^{z_2}(W)$ such that $v_1' + v_2' \in \mathrm{Y}^{z_1}(W)$, $\hat{p}_{v_1'+v_2',z_1}(\hat{u}_1) \neq 0$ and $\hat{p}_{v_2',z_2}(\hat{u}_1) \neq 0$. Similarly, since $\hat{p}_{v_1'',z^-,W^-}(\hat{u}_1) \neq 0$, there exists $v_2'' \in \mathrm{Y}^{z_2}(W)$ such that $v_1'' + v_2'' \in \mathrm{Y}^{z_1}(W)$, $\hat{p}_{v_1''+v_2'',z_1}(\hat{u}_1) \neq 0$ and $\hat{p}_{v_2'',z_2}(\hat{u}_1) \neq 0$. Therefore, we have

- $\hat{u}_1 \in \hat{\mathrm{X}}^{z_1}(W)$ (because $\hat{p}_{v_1'+v_2',z_1}(\hat{u}_1) \neq 0$).
- $v_1 + v_2' - v_2'' = (v_1' + v_2') - (v_1'' + v_2'') \in \Delta\mathrm{Y}^{z_1}(W)$.
- $\hat{u}_1 \in \hat{\mathrm{X}}^{z_2}(W)$ (because $\hat{p}_{v_2',z_2}(\hat{u}_1) \neq 0$).
- $v_2'' - v_2' \in \Delta\mathrm{Y}^{z_2}(W)$.

We can now see that $(\hat{u}_1, v_1 + v_2' - v_2'') \in \mathrm{D}^{z_1}(W) \subset \mathrm{D}(W)$ and $(\hat{u}_1, v_2'' - v_2') \in \mathrm{D}^{z_2}(W) \subset \mathrm{D}(W)$. By noticing that $v_1 = (v_1 + v_2' - v_2'') + (v_2'' - v_2')$, we conclude that:

$$\mathrm{D}(W^-) \subset \big\{(\hat{x}, y_1 + y_2) : \ (\hat{x}, y_1), (\hat{x}, y_2) \in \mathrm{D}(W)\big\}.$$

□

Now we are ready to prove Proposition 6.5.

Let W be a two-user MAC such that I_1 is $*^-$ preserved for W. Let $U_1, U_2, V_1, V_2, X_1, X_2, Y_1, Y_2, Z_1, Z_2$ be as in Remark 6.1.

1. Let $\hat{x} \in G_1$ and $y_1, y_2 \in G_2$ be such that $(\hat{x}, y_1), (\hat{x}, y_2) \in \mathrm{D}(W)$. There exist $z_1, z_2 \in \mathcal{Z}$, $y_1', y_1'' \in \mathrm{Y}^{z_1}(W)$ and $y_2', y_2'' \in \mathrm{Y}^{z_2}(W)$ such that $\hat{x} \in \hat{\mathrm{X}}^{z_1}(W)$, $\hat{x} \in \hat{\mathrm{X}}^{z_2}(W)$, $y_1 = y_1' - y_1''$ and $y_2 = y_2' - y_2''$. Lemma 6.7 implies that $\hat{p}_{y_1',z_1}(\hat{x}) \neq 0$, $\hat{p}_{y_1'',z_1}(\hat{x}) \neq 0$, $\hat{p}_{y_2',z_2}(\hat{x}) \neq 0$ and $\hat{p}_{y_2'',z_2}(\hat{x}) \neq 0$. Now from Lemma 6.14 we get $y_1' - y_2'' \in \mathrm{Y}^{(z_1,z_2)}(W^-)$ and $y_1'' - y_2' \in \mathrm{Y}^{(z_1,z_2)}(W^-)$.

   For every $v_2 \in \mathrm{Y}^{z_2}(W)$ satisfying $y_1' - y_2'' + v_2 \in \mathrm{Y}^{z_1}(W)$, we have:

$$\begin{aligned} &\hat{p}_{y_1'-y_2''+v_2,z_1}(\hat{x}) \cdot \hat{p}_{v_2,z_2}(\hat{x})^* \\ &\qquad = \hat{p}_{y_1',z_1}(\hat{x}) \hat{f}_W(\hat{x}, v_2 - y_2'') \cdot \hat{p}_{y_2'',z_2}(\hat{x})^* \hat{f}_W(\hat{x}, v_2 - y_2'')^* \\ &\qquad \overset{(a)}{=} \hat{p}_{y_1',z_1}(\hat{x}) \hat{p}_{y_2'',z_2}(\hat{x})^*, \end{aligned} \tag{6.16}$$

   where (a) follows from the fact that $\hat{f}_W(\hat{x}, v_2 - y_2'') \in \mathbb{T}$, which means that

$$\hat{f}_W(\hat{x}, v_2 - y_2'') \hat{f}_W(\hat{x}, v_2 - y_2'')^* = |\hat{f}_W(\hat{x}, v_2 - y_2'')|^2 = 1.$$

Let $z^- = (z_1, z_2) \in \mathcal{Z}^2$. From (6.15), we have:

$$\begin{aligned}
&\hat{p}_{y_1'-y_2'',z^-,W^-}(\hat{x}) \\
&= \sum_{\substack{v_2 \in \mathrm{Y}^{z_2}(W):\\ y_1'-y_2''+v_2 \in \mathrm{Y}^{z_1}(W)}} \frac{P_{Y_1|Z_1}(y_1'-y_2''+v_2|z_1)P_{Y_2|Z_2}(v_2|z_2)}{P_{V_1|Z_1,Z_2}(y_1'-y_2''|z_1,z_2)} \hat{p}_{y_1'-y_2''+v_2,z_1}(\hat{x}) \cdot \hat{p}_{v_2,z_2}(\hat{x})^* \\
&\overset{(a)}{=} \sum_{\substack{v_2 \in \mathrm{Y}^{z_2}(W):\\ y_1'-y_2''+v_2 \in \mathrm{Y}^{z_1}(W)}} \frac{P_{Y_1|Z_1}(y_1'-y_2''+v_2|z_1)P_{Y_2|Z_2}(v_2|z_2)}{P_{V_1|Z_1,Z_2}(y_1'-y_2''|z_1,z_2)} \hat{p}_{y_1',z_1}(\hat{x})\hat{p}_{y_2'',z_2}(\hat{x})^* \\
&= \hat{p}_{y_1',z_1}(\hat{x})\hat{p}_{y_2'',z_2}(\hat{x})^* \sum_{\substack{v_2 \in \mathrm{Y}^{z_2}(W):\\ y_1'-y_2''+v_2 \in \mathrm{Y}^{z_1}(W)}} \frac{P_{Y_1|Z_1}(y_1'-y_2''+v_2|z_1)P_{Y_2|Z_2}(v_2|z_2)}{P_{V_1|Z_1,Z_2}(y_1'-y_2''|z_1,z_2)} \\
&= \hat{p}_{y_1',z_1}(\hat{x})\hat{p}_{y_2'',z_2}(\hat{x})^* \neq 0,
\end{aligned}$$

where (a) follows from (6.16). This shows that $\hat{x} \in \hat{\mathrm{X}}^{z^-}(W^-)$. Now since $y_1' - y_2'' \in \mathrm{Y}^{z^-}(W^-)$ and $y_1'' - y_2' \in \mathrm{Y}^{z^-}(W^-)$, we have $(y_1' - y_2'') - (y_1'' - y_2') \in \Delta\mathrm{Y}^{z^-}(W^-)$. Therefore,

$$(\hat{x}, y_1 + y_2) = (\hat{x}, y_1' - y_1'' + y_2' - y_2'') = \big(\hat{x}, (y_1' - y_2'') - (y_1'' - y_2')\big) \in \mathrm{D}(W^-).$$

Hence, $\big\{(\hat{x}, y_1 + y_2) :\ (\hat{x}, y_1), (\hat{x}, y_2) \in \mathrm{D}(W)\big\} \subset \mathrm{D}(W^-)$. We conclude that

$$\mathrm{D}(W^-) = \big\{(\hat{x}, y_1 + y_2) :\ (\hat{x}, y_1), (\hat{x}, y_2) \in \mathrm{D}(W)\big\}$$

since the other inclusion was shown in Lemma 6.16.

2. Let $\hat{x}, y_1, y_2$ be such that $(\hat{x}, y_1), (\hat{x}, y_2) \in \mathrm{D}(W)$. Define $y_1', y_1'', y_2', y_2'', z_1, z_2, z^-$ as in 1). We have shown that $\hat{p}_{y_1'-y_2'',z^-,W^-}(\hat{x}) = \hat{p}_{y_1',z_1}(\hat{x})\hat{p}_{y_2'',z_2}(\hat{x})^*$. Similarly, we can show that $\hat{p}_{y_1''-y_2',z^-,W^-}(\hat{x}) = \hat{p}_{y_1'',z_1}(\hat{x})\hat{p}_{y_2',z_2}(\hat{x})^*$. Therefore,

$$\begin{aligned}
\hat{f}_{W^-}(\hat{x}, y_1 + y_2) &= \hat{f}_{W^-}(\hat{x}, y_1' - y_1'' + y_2' - y_2'') = \hat{f}_{W^-}\big(\hat{x}, (y_1' - y_2'') - (y_1'' - y_2')\big) \\
&= \frac{\hat{p}_{y_1'-y_2'',z^-,W^-}(\hat{x})}{\hat{p}_{y_1''-y_2',z^-,W^-}(\hat{x})} = \frac{\hat{p}_{y_1',z_1}(\hat{x})\hat{p}_{y_2'',z_2}(\hat{x})^*}{\hat{p}_{y_1'',z_1}(\hat{x})\hat{p}_{y_2',z_2}(\hat{x})^*} = \frac{\hat{f}_W(\hat{x}, y_1)}{\hat{f}_W(\hat{x}, y_2)^*} \\
&\overset{(a)}{=} \hat{f}_W(\hat{x}, y_1) \cdot \hat{f}_W(\hat{x}, y_2),
\end{aligned}$$

where (a) follows from the fact that $\hat{f}_W(\hat{x}, y_2) \cdot \hat{f}_W(\hat{x}, y_2)^* = |\hat{f}_W(\hat{x}, y_2)|^2 = 1$.

6.5.4 Proof of Proposition 6.6

We need the following lemmas.

Lemma 6.17. *For every $y_1, y_2 \in G_2$ and every $z_1, z_2 \in \mathcal{Z}$, we have:*

- *If $(y_1, z_1) \notin \mathrm{YZ}(W)$ or $(y_2, z_2) \notin \mathrm{YZ}(W)$, then $\big(y_2, (z_1, z_2, u_1, y_1 - y_2)\big) \notin \mathrm{YZ}(W^+)$ for every $u_1 \in G_1$.*
- *If $(y_1, z_1) \in \mathrm{YZ}(W)$ and $(y_2, z_2) \in \mathrm{YZ}(W)$, there exists $u_1 \in G_1$ such that $\big(y_2, (z_1, z_2, u_1, y_1 - y_2)\big) \in \mathrm{YZ}(W^+)$.*

Proof. Let $U_1, U_2, V_1, V_2, X_1, X_2, Y_1, Y_2, Z_1, Z_2$ be as in Remark 6.1. For every $u_1 \in G_1$, every $y_1, y_2 \in G_2$ and every $z_1, z_2 \in \mathcal{Z}$, we have:

$$\begin{aligned} P_{V_2,Z_1,Z_2,U_1,V_1}(y_2, z_1, z_2, u_1, y_1 - y_2) \\ &= \sum_{u_2 \in G_1} P_{U_2,V_2,Z_1,Z_2,U_1,V_1}(u_2, y_2, z_1, z_2, u_1, y_1 - y_2) \\ &= \sum_{u_2 \in G_1} P_{X_1,X_2,Y_1,Y_2,Z_1,Z_2}(u_1 + u_2, u_2, y_1, y_2, z_1, z_2) \\ &= \sum_{u_2 \in G_1} P_{X_1,Y_1,Z_1}(u_1 + u_2, y_1, z_1) \cdot P_{X_2,Y_2,Z_2}(u_2, y_2, z_2). \end{aligned}$$

Therefore, we have:

- If $(y_1, z_1) \notin \mathrm{YZ}(W)$ or $(y_2, z_2) \notin \mathrm{YZ}(W)$, then for all $u_1, u_2 \in G_1$, we have $P_{X_1,Y_1,Z_1}(u_1 + u_2, y_1, z_1) \leq P_{Y_1,Z_1}(y_1, z_1) = 0$ or $P_{X_2,Y_2,Z_2}(u_2, y_2, z_2) \leq P_{Y_2,Z_2}(y_2, z_2) = 0$, which means that $P_{V_2,Z_1,Z_2,U_1,V_1}(y_2, z_1, z_2, u_1, y_1 - y_2) = 0$. Hence $\big(y_2, (z_1, z_2, u_1, y_1 - y_2)\big) \notin \mathrm{YZ}(W^+)$ for every $u_1 \in G_1$.
- If $(y_1, z_1) \in \mathrm{YZ}(W)$ and $(y_2, z_2) \in \mathrm{YZ}(W)$, then $P_{Y_1,Z_1}(y_1, z_1) > 0$ and $P_{Y_2,Z_2}(y_2, z_2) > 0$. This means that there exist $x_1, x_2 \in G_1$ such that
$$P_{X_1,Y_1,Z_1}(x_1, y_1, z_1) > 0$$
and
$$P_{X_2,Y_2,Z_2}(x_2, y_2, z_2) > 0.$$
Let $u_1 = x_1 - x_2$ and $u_2 = x_2$. We have
$$P_{X_1,Y_1,Z_1}(u_1 + u_2, y_1, z_1) \cdot P_{X_2,Y_2,Z_2}(u_2, y_2, z_2) > 0,$$
which implies that $P_{V_2,Z_1,Z_2,U_1,V_1}(y_2, z_1, z_2, u_1, y_1 - y_2) > 0$ hence
$$\big(y_2, (z_1, z_2, u_1, y_1 - y_2)\big) \in \mathrm{YZ}(W^+).$$

□

Lemma 6.18. *Let $U_1, U_2, V_1, V_2, X_1, X_2, Y_1, Y_2, Z_1, Z_2$ be as in Remark 6.1. For every $\big(v_2, (z_1, z_2, u_1, v_1)\big) \in \mathrm{YZ}(W^+)$, we have:*

$$\hat{p}_{v_2,(z_1,z_2,u_1,v_1),W^+}(\hat{u}_2) = \sum_{\hat{u}_2' \in G_1} \frac{\hat{p}_{v_1+v_2,z_1}(\hat{u}_2') \cdot \hat{p}_{v_2,z_2}(\hat{u}_2 - \hat{u}_2')}{|G_1|\alpha(u_1, z_1, z_2, v_1, v_2)} e^{j2\pi\langle \hat{u}_2', u_1 \rangle}, \tag{6.17}$$

where $\alpha(u_1, z_1, z_2, v_1, v_2) = P_{U_1|Z_1,Z_2,V_1,V_2}(u_1|z_1, z_2, v_1, v_2)$.

Proof. For every $\big(v_2, (z_1, z_2, u_1, v_1)\big) \in \mathrm{YZ}(W^+)$ and every $u_2 \in G_2$, we have:

$$\begin{aligned} p_{v_2,(z_1,z_2,u_1,v_1),W^+}(u_2) &= P_{U_2|V_2,Z_1,Z_2,U_1,V_1}(u_2|v_2, z_1, z_2, u_1, v_1) \\ &= \frac{P_{U_1,U_2|Z_1,Z_2,V_1,V_2}(u_1, u_2|z_1, z_2, v_1, v_2)}{P_{U_1|Z_1,Z_2,V_1,V_2}(u_1|z_1, z_2, v_1, v_2)} \\ &= \frac{P_{X_1,X_2|Z_1,Z_2,Y_1,Y_2}(u_1 + u_2, u_2|z_1, z_2, v_1 + v_2, v_2)}{\alpha(u_1, z_1, z_2, v_1, v_2)} \\ &= \frac{P_{X_1|Z_1,Y_1}(u_1 + u_2|z_1, v_1 + v_2) P_{X_2|Z_2,Y_2}(u_2|z_2, v_2)}{\alpha(u_1, z_1, z_2, v_1, v_2)} \\ &= \frac{p_{v_1+v_2,z_1}(u_1 + u_2) p_{v_2,z_2}(u_2)}{\alpha(u_1, z_1, z_2, v_1, v_2)}. \end{aligned}$$

Therefore, for every $\hat{u}_2 \in G_2$, we have:

$$\begin{aligned}\hat{p}_{v_2,(z_1,z_2,u_1,v_1),W^+}(\hat{u}_2) &= \frac{\frac{1}{|G_1|}\left(\hat{p}_{v_1+v_2,z_1}(\hat{u}_2)e^{j2\pi\langle\hat{u}_2,u_1\rangle}\right) * \hat{p}_{v_2,z_2}(\hat{u}_2)}{\alpha(u_1,z_1,z_2,v_1,v_2)} \\ &= \frac{\sum_{\hat{u}_2'\in G_1}\hat{p}_{v_1+v_2,z_1}(\hat{u}_2')e^{j2\pi\langle\hat{u}_2',u_1\rangle}\hat{p}_{v_2,z_2}(\hat{u}_2-\hat{u}_2')}{|G_1|\alpha(u_1,z_1,z_2,v_1,v_2)} \\ &= \sum_{\hat{u}_2'\in G_1}\frac{\hat{p}_{v_1+v_2,z_1}(\hat{u}_2')\cdot\hat{p}_{v_2,z_2}(\hat{u}_2-\hat{u}_2')}{|G_1|\alpha(u_1,z_1,z_2,v_1,v_2)}e^{j2\pi\langle\hat{u}_2',u_1\rangle}.\end{aligned}$$

□

Lemma 6.19. *Let $(y_1,z_1),(y_2,z_2) \in \mathrm{YZ}(W)$ and $\hat{x} \in G_1$. If there exists $u_1 \in G_1$ such that*

$$\sum_{\hat{u}\in G_1}\hat{p}_{y_1,z_1}(\hat{u})\cdot\hat{p}_{y_2,z_2}(\hat{x}-\hat{u})e^{j2\pi\langle\hat{u},u_1\rangle} \neq 0, \tag{6.18}$$

then we have:

- $(y_2, z^+) \in \mathrm{YZ}(W^+)$, *where* $z^+ = (z_1, z_2, u_1, y_1 - y_2)$.
- $\hat{x} \in \hat{\mathrm{X}}^{z^+}(W^+)$.

Proof. Let $U_1, U_2, V_1, V_2, X_1, X_2, Y_1, Y_2, Z_1, Z_2$ be as in Remark 6.1. Let $v_1 = y_1 - y_2$ and $v_2 = y_2$. Notice that the expression in (6.18) is the DFT of the mapping $K : G_1 \to \mathbb{C}$ defined as

$$K(x) = |G_1| \cdot p_{y_1,z_1}(u_1 + x) \cdot p_{y_2,z_2}(x).$$

Equation (6.18) shows that $\hat{K}$ is not zero everywhere which implies that K is not zero everywhere. Therefore, there exists $x \in G_1$ such that $K(x) \neq 0$. We have:

$$\begin{aligned}P_{V_2,Z_1,Z_2,U_1,V_1}(v_2,z_1,z_2,u_1,v_1) &\geq P_{U_1,U_2,V_1,V_2,Z_1,Z_2}(u_1,x,y_1-y_2,y_2,z_1,z_2) \\ &= P_{X_1,X_2,Y_1,Y_2,Z_1,Z_2}(u_1+x,x,y_1,y_2,z_1,z_2) \\ &= P_{X_1,Y_1,Z_1}(u_1+x,y_1,z_1)P_{X_2,Y_2,Z_2}(x,y_2,z_2) \\ &= P_{Y_1,Z_1}(y_1,z_1)p_{y_1,z_1}(u_1+x)\cdot P_{Y_2,Z_2}(y_2,z_2)p_{y_2,z_2}(x) \\ &= P_{Y_1,Z_1}(y_1,z_1)\cdot P_{Y_2,Z_2}(y_2,z_2)\cdot\frac{K(x)}{|G_1|} \overset{(a)}{>} 0,\end{aligned}$$

where (a) follows from the fact that $y_1 \in \mathrm{Y}^{z_1}(W)$, $y_2 \in \mathrm{Y}^{z_2}(W)$ and $K(x) > 0$. We conclude that $\big(v_2,(z_1,z_2,u_1,v_1)\big) \in \mathrm{YZ}(W^+)$ and so we can apply (6.17) to $(v_2, z_1, z_2, u_1, v_1)$:

$$\hat{p}_{v_2,(z_1,z_2,u_1,v_1),W^+}(\hat{x}) \overset{(a)}{=} \sum_{\hat{u}\in G_1}\frac{\hat{p}_{y_1,z_1}(\hat{u})\cdot\hat{p}_{y_2,z_2}(\hat{x}-\hat{u})}{|G_1|\alpha(u_1,z_1,z_2,v_1,v_2)}e^{j2\pi\langle\hat{u},u_1\rangle} \overset{(b)}{\neq} 0,$$

where (b) follows from (6.18). Therefore, $\hat{p}_{y_2,z^+,W^+}(\hat{x}) \neq 0$, where

$$z^+ = (z_1, z_2, u_1, y_1 - y_2).$$

Hence $\hat{x} \in \hat{\mathrm{X}}^{z^+}(W^+)$. □

Now we are ready to prove Proposition 6.6.

Let W be a two-user MAC and assume that polarization $*$-preserves I_1 for W. Let $U_1, U_2, V_1, V_2, X_1, X_2, Y_1, Y_2, Z_1, Z_2$ be as in Remark 6.1.

1. Suppose that $\hat{x}_1, \hat{x}_2 \in G_1$ and $y \in G_2$ satisfy $(\hat{x}_1, y), (\hat{x}_2, y) \in \mathrm{D}(W)$ and let $\hat{x} = \hat{x}_1 + \hat{x}_2$. There exist $z_1, z_2 \in \mathcal{Z}$, $y_1, y_1' \in \mathrm{Y}^{z_1}(W)$ and $y_2, y_2' \in \mathrm{Y}^{z_2}(W)$ such that

   - $\hat{x}_1 \in \hat{\mathrm{X}}^{z_1}(W)$ and $y = y_1 - y_1'$.
   - $\hat{x}_2 \in \hat{\mathrm{X}}^{z_2}(W)$ and $y = y_2 - y_2'$.

   Lemma 6.7 implies that $\hat{p}_{y_1,z_1}(\hat{x}_1) \neq 0$, $\hat{p}_{y_1',z_1}(\hat{x}_1) \neq 0$, $\hat{p}_{y_2,z_2}(\hat{x}_2) \neq 0$ and $\hat{p}_{y_2',z_2}(\hat{x}_2) \neq 0$.

   Let $v_1 = y_1 - y_2 = y_1' - y_2'$, $v_2 = y_2$ and $v_2' = y_2'$. Define the mapping $\hat{L} : G_1 \to \mathbb{C}$ as

   $$\hat{L}(\hat{u}) = \hat{p}_{y_1,z_1}(\hat{u}) \cdot \hat{p}_{y_2,z_2}(\hat{x} - \hat{u}).$$

   We have: $\hat{L}(\hat{x}_1) = \hat{p}_{y_1,z_1}(\hat{x}_1) \cdot \hat{p}_{y_2,z_2}(\hat{x}_2) \neq 0$. Therefore, the mapping $\hat{L}$ is not zero everywhere, which implies that its inverse DFT is not zero everywhere. Hence there exists $u_1 \in G_1$ such that:

   $$\sum_{\hat{u} \in G_1} \hat{p}_{y_1,z_1}(\hat{u}) \cdot \hat{p}_{y_2,z_2}(\hat{x} - \hat{u}) e^{j2\pi\langle \hat{u}, u_1 \rangle} \neq 0.$$

   It follows from Lemma 6.19 that $(v_2, z^+) \in \mathrm{YZ}(W^+)$ and $\hat{x} \in \hat{\mathrm{X}}^{z^+}(W^+)$, where $z^+ = (z_1, z_2, u_1, v_1)$. If we can also show that $(v_2', z^+) \in \mathrm{YZ}(W^+)$ we will be able to conclude that $(\hat{x}, y) \in \mathrm{D}(W^+)$ since $y = v_2 - v_2'$. We have the following:

   - Since $(v_2, z^+) \in \mathrm{YZ}(W^+)$, we have

     $$P_{U_1,Z_1,Z_2,V_1}(u_1, z_1, z_2, v_1) \geq P_{V_2,Z_1,Z_2,U_1,V_1}(v_2, z_1, z_2, u_1, v_1) > 0.$$

     Hence,

     $$P_{U_1|Z_1,Z_2,V_1}(u_1|z_1, z_2, v_1) > 0.$$

   - Since $y_1' \in \mathrm{Y}^{z_1}(W)$ and $y_2' \in \mathrm{Y}^{z_2}(W)$, we have

     $$P_{V_2,Z_1,Z_2,V_1}(v_2', z_1, z_2, v_1) = P_{Y_1,Z_1,Y_2,Z_2}(y_1', z_1, y_2', z_2) > 0.$$

     Thus,

     $$P_{V_2|Z_1,Z_2,V_1}(v_2'|z_1, z_2, v_1) > 0.$$

   But I_1 is preserved for W, so we must have $I(U_1; V_2|Z_1Z_2V_1) = 0$. Therefore,

   $$\begin{aligned} &P_{U_1,V_2|Z_1,Z_2,V_1}(u_1, v_2'|z_1, z_2, v_1) \\ &\quad = P_{U_1|Z_1,Z_2,V_1}(u_1|z_1, z_2, v_1) \cdot P_{V_2|Z_1,Z_2,V_1}(v_2'|z_1, z_2, v_1) > 0. \end{aligned} \tag{6.19}$$

   We conclude that $P_{V_2,Z_1,Z_2,U_1,V_1}(v_2', z_1, z_2, u_1, v_1) > 0$, i.e., $(v_2', z^+) \in \mathrm{YZ}(W^+)$. Hence, $(\hat{x}, y) \in \mathrm{D}(W^+)$. We conclude that $(\hat{x}_1 + \hat{x}_2, y) \in \mathrm{D}(W^+)$ for every $\hat{x}_1, \hat{x}_2 \in G_1$ and every $y \in G_2$ satisfying $(\hat{x}_1, y), (\hat{x}_2, y) \in \mathrm{D}(W)$. Therefore,

   $$\big\{(\hat{x}_1 + \hat{x}_2, y) : \ (\hat{x}_1, y), (\hat{x}_2, y) \in \mathrm{D}(W)\big\} \subset \mathrm{D}(W^+).$$

2. Suppose that $\hat{x}_1, \hat{x}_2 \in G_1$ and $y \in G_2$ satisfy $(\hat{x}_1, y), (\hat{x}_2, y) \in \mathrm{D}(W)$ and let $\hat{x} = \hat{x}_1 + \hat{x}_2$. Let $y_1, y_2, y_1', y_2', v_1, v_2, v_2', z_1, z_2, z^+$ be defined as in 1) so that $v_2, v_2' \in \mathrm{Y}^{z+}(W^+)$, $y = v_2 - v_2'$ and $\hat{x} \in \hat{\mathrm{X}}^{z^+}(W^+)$. Lemma 6.7 implies that $\hat{p}_{v_2,z^+,W^+}(\hat{x}) \neq 0$ and $\hat{p}_{v_2',z^+,W^+}(\hat{x}) \neq 0$. Now since $(\hat{x}, y) = (\hat{x}, v_2 - v_2') \in \mathrm{D}(W^+)$, we have:

$$\begin{aligned}\hat{p}_{v_2,(z_1,z_2,u_1,v_1),W^+}(\hat{x}) &= \hat{p}_{v_2,z^+,W^+}(\hat{x}) = \hat{f}_{W^+}(\hat{x}, y) \cdot \hat{p}_{v_2',z^+,W^+}(\hat{x}) \\ &= \hat{f}_{W^+}(\hat{x}, y) \cdot \hat{p}_{v_2',(z_1,z_2,u_1,v_1),W^+}(\hat{x}).\end{aligned}$$

Define $F : G_1 \to \mathbb{C}$ and $F' : G_1 \to \mathbb{C}$ as follows:

$$F(u_1') = \sum_{\hat{u} \in G_1} \hat{p}_{y_1,z_1}(\hat{u}) \cdot \hat{p}_{y_2,z_2}(\hat{x} - \hat{u}) e^{j2\pi\langle \hat{u}, u_1' \rangle}.$$

$$F'(u_1') = \sum_{\hat{u} \in G_1} \hat{p}_{y_1',z_1}(\hat{u}) \cdot \hat{p}_{y_2',z_2}(\hat{x} - \hat{u}) e^{j2\pi\langle \hat{u}, u_1' \rangle}.$$

For every $u_1' \in G_1$, we have:

- If $F(u_1') \neq 0$ then from Lemma 6.19, we have

$$\big(v_2, (z_1, z_2, u_1', v_1)\big) \in \mathrm{YZ}(W^+) \ \text{ and } \ \hat{x} \in \hat{\mathrm{X}}^{(z_1,z_2,u_1',v_1)}(W^+).$$

  By replacing u_1 by u_1' in (6.19), we get $\big(v_2', (z_1, z_2, u_1', v_1)\big) \in \mathrm{YZ}(W^+)$. Therefore,

$$\hat{p}_{v_2,(z_1,z_2,u_1',v_1),W^+}(\hat{x}) = \hat{f}_{W^+}(\hat{x}, y) \cdot \hat{p}_{v_2',(z_1,z_2,u_1',v_1),W^+}(\hat{x}). \tag{6.20}$$

  We have:

$$\begin{aligned}&F(u_1') \\ &= \sum_{\hat{u} \in G_1} \hat{p}_{y_1,z_1}(\hat{u}) \cdot \hat{p}_{y_2,z_2}(\hat{x} - \hat{u}) e^{j2\pi\langle \hat{u}, u_1' \rangle} \\ &\overset{(a)}{=} |G_1| \cdot \alpha(u_1', z_1, z_2, v_1, v_2) \hat{p}_{v_2,(z_1,z_2,u_1',v_1),W^+}(\hat{x}) \\ &\overset{(b)}{=} \frac{\alpha(u_1', z_1, z_2, v_1, v_2)}{\alpha(u_1', z_1, z_2, v_1, v_2')} |G_1| \alpha(u_1', z_1, z_2, v_1, v_2') \hat{p}_{v_2',(z_1,z_2,u_1',v_1),W^+}(\hat{x}) \hat{f}_{W^+}(\hat{x}, y) \\ &\overset{(c)}{=} \frac{\alpha(u_1', z_1, z_2, v_1, v_2)}{\alpha(u_1', z_1, z_2, v_1, v_2')} \sum_{\hat{u} \in G_1} \hat{p}_{y_1',z_1}(\hat{u}) \cdot \hat{p}_{y_2',z_2}(\hat{x} - \hat{u}) e^{j2\pi\langle \hat{u}, u_1' \rangle} \hat{f}_{W^+}(\hat{x}, y) \\ &= \frac{P_{U_1|Z_1,Z_2,V_1,V_2}(u_1'|z_1, z_2, v_1, v_2)}{P_{U_1|Z_1,Z_2,V_1,V_2}(u_1'|z_1, z_2, v_1, v_2')} \hat{f}_{W^+}(\hat{x}, y) F'(u_1') \\ &\overset{(d)}{=} \hat{f}_{W^+}(\hat{x}, y) F'(u_1'),\end{aligned}$$

  where (a) and (c) follow from (6.17), (b) follows from (6.20) and (d) follows from the fact that $I(U_1; V_2 | Z_1 Z_2 V_1) = 0$. Therefore, $F'(u_1') \neq 0$ and $F(u_1') = \hat{f}_{W^+}(\hat{x}, y) F'(u_1')$.

- If $F(u_1')=0$ then we must have $F'(u_1')=0$ (because $F'(u_1')\neq 0$ would yield $F(u_1')\neq 0$ in a similar way as above, which is a contradiction). Therefore, we have $F(u_1')=0=\hat{f}_{W^+}(\hat{x},y)F'(u_1')$.

We conclude that for every $u_1'\in G_1$, we have

$$F(u_1')=\hat{f}_{W^+}(\hat{x},y)F'(u_1')=\sum_{\hat{u}\in G_1}\hat{f}_{W^+}(\hat{x},y)\cdot\hat{p}_{y_1',z_1}(\hat{u})\cdot\hat{p}_{y_2',z_2}(\hat{x}-\hat{u})e^{j2\pi\langle\hat{u},u_1'\rangle}. \tag{6.21}$$

Now define $g:G_1\times G_2\to\mathbb{C}$ as follows:

$$g(\hat{x}',y')=\begin{cases}\hat{f}_W(\hat{x}',y') & \text{if } (\hat{x}',y')\in \mathrm{D}(W),\\ 0 & \text{otherwise.}\end{cases} \tag{6.22}$$

For every $\hat{x}'\in G_1$, we have:

- If $\hat{p}_{y_1,z_1}(\hat{x}')\neq 0$ then $\hat{p}_{y_1',z_1}(\hat{x}')\neq 0$ (by Lemma 6.7) and $\hat{p}_{y_1,z_1}(\hat{x}')=\hat{f}_W(\hat{x}',y_1-y_1')\hat{p}_{y_1',z_1}(\hat{x}')=g(\hat{x}',y)\hat{p}_{y_1',z_1}(\hat{x}')$.
- If $\hat{p}_{y_1,z_1}(\hat{x}')=0$ then $\hat{p}_{y_1',z_1}(\hat{x}')=0$ (by Lemma 6.7) and so $\hat{p}_{y_1,z_1}(\hat{x}')=0=g(\hat{x}',y)\hat{p}_{y_1',z_1}(\hat{x}')$.

Therefore, for every $\hat{x}'\in G_1$ we have $\hat{p}_{y_1,z_1}(\hat{x}')=g(\hat{x}',y)\hat{p}_{y_1',z_1}(\hat{x}')$. Similarly, $\hat{p}_{y_2,z_2}(\hat{x}')=g(\hat{x}',y)\hat{p}_{y_2',z_2}(\hat{x}')$ for all $\hat{x}'\in G_1$. Hence,

$$\begin{aligned}F(u_1')&=\sum_{\hat{u}\in G_1}\hat{p}_{y_1,z_1}(\hat{u})\cdot\hat{p}_{y_2,z_2}(\hat{x}-\hat{u})e^{j2\pi\langle\hat{u},u_1'\rangle}\\&=\sum_{\hat{u}\in G_1}g(\hat{u},y)\hat{p}_{y_1',z_1}(\hat{u})\cdot g(\hat{x}-\hat{u},y)\hat{p}_{y_2',z_2}(\hat{x}-\hat{u})e^{j2\pi\langle\hat{u},u_1'\rangle}.\end{aligned} \tag{6.23}$$

We conclude that for every $u_1'\in G_1$, we have:

$$\begin{aligned}\sum_{\hat{u}\in G_1}\Big[\hat{f}_{W^+}(\hat{x},y)-g(\hat{u},y)g(\hat{x}-\hat{u},y)\Big]\hat{p}_{y_1',z_1}(\hat{u})\cdot\hat{p}_{y_2',z_2}(\hat{x}-\hat{u})e^{j2\pi\langle\hat{u},u_1'\rangle}&\\ \overset{(a)}{=}F(u_1')-F(u_1')=0,&\end{aligned} \tag{6.24}$$

where (a) follows from (6.21) and (6.23). Notice that the sum in (6.24) is the inverse DFT of the function $\hat{K}:G_1\to\mathbb{C}$ defined as:

$$\hat{K}(\hat{u})=|G_1|\cdot\Big[\hat{f}_{W^+}(\hat{x},y)-g(\hat{u},y)g(\hat{x}-\hat{u},y)\Big]\hat{p}_{y_1',z_1}(\hat{u})\cdot\hat{p}_{y_2',z_2}(\hat{x}-\hat{u}).$$

Now (6.24) implies that the inverse DFT of $\hat{K}$ is zero everywhere. Therefore, $\hat{K}$ is also zero everywhere. In particular,

$$\hat{K}(\hat{x_1})=|G_1|\cdot\Big[\hat{f}_{W^+}(\hat{x},y)-g(\hat{x}_1,y)g(\hat{x}_2,y)\Big]\hat{p}_{y_1',z_1}(\hat{x}_1)\cdot\hat{p}_{y_2',z_2}(\hat{x}_2)=0.$$

But $\hat{p}_{y_1',z_1}(\hat{x}_1)\neq 0$ and $\hat{p}_{y_2',z_2}(\hat{x}_2)\neq 0$, so we must have

$$\hat{f}_{W^+}(\hat{x},y)-g(\hat{x}_1,y)g(\hat{x}_2,y)=0.$$

Therefore,

$$\hat{f}_{W^+}(\hat{x},y)=g(\hat{x}_1,y)g(\hat{x}_2,y)=\hat{f}_W(\hat{x}_1,y)\cdot\hat{f}_W(\hat{x}_2,y).$$

6.5.5 Proof of Lemma 6.10

Lemma 6.20. *If $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$ is polarization compatible then W^- is also polarization compatible.*

Proof. Let $U_1, U_2, V_1, V_2, X_1, X_2, Y_1, Y_2, Z_1, Z_2$ be as in Remark 6.1. Let $F : D \to \mathbb{T}$ be the pseudo-quadratic function of Definition 6.7.

Let $(\hat{u}_1, v) \in \mathrm{D}(W^-)$. There exists $z^- = (z_1, z_2) \in \mathcal{Z}^2$ such that $\hat{u}_1 \in \hat{\mathrm{X}}^{z^-}(W^-)$ and $v \in \Delta\mathrm{Y}^{z^-}(W^-)$. We have:

- Since $\hat{u}_1 \in \hat{\mathrm{X}}^{z^-}(W^-)$, there exists $v_1 \in \mathrm{Y}^{z^-}(W^-)$ such that $\hat{p}_{v_1,z^-,W^-}(\hat{u}_1) \neq 0$. From (6.15), we have:

$$\hat{p}_{v_1,z^-,W^-}(\hat{u}_1) = \sum_{\substack{v_2 \in \mathrm{Y}^{z_2}(W): \\ v_1+v_2 \in \mathrm{Y}^{z_1}(W)}} \frac{P_{Y_1|Z_1}(v_1+v_2|z_1)P_{Y_2|Z_2}(v_2|z_2)}{P_{V_1|Z_1,Z_2}(v_1|z_1,z_2)} \hat{p}_{v_1+v_2,z_1}(\hat{u}_1) \cdot \hat{p}_{v_2,z_2}(\hat{u}_1)^*.$$

  Since $\hat{p}_{v_1,z^-,W^-}(\hat{u}_1) \neq 0$, the terms in the above sum cannot all be zero. Therefore, there exists $v_2 \in \mathrm{Y}^{z_2}(W)$ such that $v_1 + v_2 \in \mathrm{Y}^{z_1}(W)$, $\hat{p}_{v_1+v_2,z_1}(\hat{u}_1) \neq 0$ and $\hat{p}_{v_2,z_2}(\hat{u}_1) \neq 0$. Hence, $\hat{u}_1 \in \hat{\mathrm{X}}^{z_1}(W)$ and $\hat{u}_1 \in \hat{\mathrm{X}}^{z_2}(W)$.

- From Lemma 6.14 we have $\mathrm{Y}^{z^-}(W^-) = \mathrm{Y}^{z_1}(W) - \mathrm{Y}^{z_2}(W)$ which implies that $\Delta\mathrm{Y}^{z^-}(W^-) = \Delta\mathrm{Y}^{z_1}(W) - \Delta\mathrm{Y}^{z_2}(W)$. Now since $v \in \Delta\mathrm{Y}^{z^-}(W^-)$, there exists $y_1 \in \Delta\mathrm{Y}^{z_1}(W)$ and $y_2 \in \Delta\mathrm{Y}^{z_2}(W)$ such that $v = y_1 - y_2$.

We conclude that

$$(\hat{u}_1, y_1) \in \hat{\mathrm{X}}^{z_1}(W) \times \Delta\mathrm{Y}^{z_1}(W) = \mathrm{D}^{z_1}(W) \subset \mathrm{D}(W) \subset D,$$

and

$$(\hat{u}_1, y_2) \in \hat{\mathrm{X}}^{z_2}(W) \times \Delta\mathrm{Y}^{z_2}(W) = \mathrm{D}^{z_2}(W) \subset \mathrm{D}(W) \subset D.$$

Therefore, $(\hat{u}_1, v) = (\hat{u}_1, y_1 - y_2) \in D$ since D is a pseudo-quadratic domain. Since this is true for every $(\hat{u}_1, v) \in \mathrm{D}(W^-)$, we conclude that $\mathrm{D}(W^-) \subset D$.

Now let $(\hat{u}_1, z^-) \in \hat{\mathrm{X}}\mathrm{Z}(W^-)$ (where $z^- = (z_1, z_2) \in \mathcal{Z}^2$). We have shown that $\hat{u}_1 \in \hat{\mathrm{X}}^{z_1}(W)$ and $\hat{u}_1 \in \hat{\mathrm{X}}^{z_2}(W)$ and so $(\hat{u}_1, z_1) \in \hat{\mathrm{X}}\mathrm{Z}(W)$ and $(\hat{u}_1, z_2) \in \hat{\mathrm{X}}\mathrm{Z}(W)$. Fix $y_1 \in \mathrm{Y}^{z_1}(W)$ and $y_2 \in \mathrm{Y}^{z_2}(W)$. For every $v_1' \in \mathrm{Y}^{z^-}(W^-)$, we have:

$$\begin{aligned}
&\hat{p}_{v_1',z^-,W^-}(\hat{u}_1) \\
&= \sum_{\substack{v_2' \in \mathrm{Y}^{z_2}(W): \\ v_1'+v_2' \in \mathrm{Y}^{z_1}(W)}} \frac{P_{Y_1|Z_1}(v_1'+v_2'|z_1)P_{Y_2|Z_2}(v_2'|z_2)}{P_{V_1|Z_1,Z_2}(v_1'|z_1,z_2)} \hat{p}_{v_1'+v_2',z_1}(\hat{u}_1) \cdot \hat{p}_{v_2',z_2}(\hat{u}_1)^* \\
&\overset{(a)}{=} \sum_{\substack{v_2' \in \mathrm{Y}^{z_2}(W): \\ v_1'+v_2' \in \mathrm{Y}^{z_1}(W)}} \frac{P_{Y_1|Z_1}(v_1'+v_2'|z_1)P_{Y_2|Z_2}(v_2'|z_2)}{P_{V_1|Z_1,Z_2}(v_1'|z_1,z_2)} \hat{p}_{y_1,z_1}(\hat{u}_1) \cdot F(\hat{u}_1, v_1'+v_2'-y_1) \cdot \frac{\hat{p}_{y_2,z_2}(\hat{u}_1)^*}{F(\hat{u}_1, v_2'-y_2)} \\
&\overset{(b)}{=} \hat{p}_{y_1,z_1}(\hat{u}_1) \cdot \hat{p}_{y_2,z_2}(\hat{u}_1)^* \sum_{\substack{v_2' \in \mathrm{Y}^{z_2}(W): \\ v_1'+v_2' \in \mathrm{Y}^{z_1}(W)}} \frac{P_{Y_1|Z_1}(v_1'+v_2'|z_1)P_{Y_2|Z_2}(v_2'|z_2)}{P_{V_1|Z_1,Z_2}(v_1'|z_1,z_2)} F(\hat{u}_1, v_1'+v_2'-y_1-v_2'+y_2) \\
&= \hat{p}_{y_1,z_1}(\hat{u}_1) \cdot \hat{p}_{y_2,z_2}(\hat{u}_1)^* \cdot F(\hat{u}_1, v_1'-y_1+y_2) \sum_{\substack{v_2' \in \mathrm{Y}^{z_2}(W): \\ v_1'+v_2' \in \mathrm{Y}^{z_1}(W)}} \frac{P_{Y_1|Z_1}(v_1'+v_2'|z_1)P_{Y_2|Z_2}(v_2'|z_2)}{P_{V_1|Z_1,Z_2}(v_1'|z_1,z_2)} \\
&= \hat{p}_{y_1,z_1}(\hat{u}_1) \cdot \hat{p}_{y_2,z_2}(\hat{u}_1)^* \cdot F(\hat{u}_1, v_1'-y_1+y_2),
\end{aligned}$$

where (a) follows from the polarization compatibility of W and from the fact that $F(\hat{u}_1, v_2' - y_2) \in \mathbb{T}$ which implies that $F(\hat{u}_1, v_2' - y_2)^* = \dfrac{1}{F(\hat{u}_1, v_2' - y_2)}$. (b) follows from the fact that the mapping $y \to F(\hat{u}_1, y)$ is a group homomorphism from $(H_2^{\hat{u}_1}(D), +)$ to $(\mathbb{T}, \cdot)$. Therefore, for every $v_1', v_1'' \in \mathrm{Y}^{z^-}(W^-)$, we have:

$$\begin{aligned}
\hat{p}_{v_1',z^-,W^-}(\hat{u}_1) &= \hat{p}_{y_1,z_1}(\hat{u}_1) \cdot \hat{p}_{y_2,z_2}(\hat{u}_1)^* \cdot F(\hat{u}_1, v_1' - y_1 + y_2) \\
&= \hat{p}_{y_1,z_1}(\hat{u}_1) \cdot \hat{p}_{y_2,z_2}(\hat{u}_1)^* \cdot F(\hat{u}_1, v_1'' - y_1 + y_2 + v_1' - v_1'') \\
&= \hat{p}_{y_1,z_1}(\hat{u}_1) \cdot \hat{p}_{y_2,z_2}(\hat{u}_1)^* \cdot F(\hat{u}_1, v_1'' - y_1 + y_2) \cdot F(\hat{u}_1, v_1' - v_1'') \\
&= \hat{p}_{v_1'',z^-,W^-}(\hat{u}_1) \cdot F(\hat{u}_1, v_1' - v_1'').
\end{aligned}$$

Hence, $\hat{p}_{v_1',z^-,W^-}(\hat{u}_1) = F(\hat{u}_1, v_1' - v_1'') \cdot \hat{p}_{v_1'',z^-,W^-}(\hat{u}_1)$. We conclude that W^- is polarization compatible. □

Lemma 6.21. *If $W : G_1 \times G_2 \longrightarrow \mathcal{Z}$ is polarization compatible then W^+ is also polarization compatible.*

Proof. Let $U_1, U_2, V_1, V_2, X_1, X_2, Y_1, Y_2, Z_1, Z_2$ be as in Remark 6.1. Let $F : D \to \mathbb{T}$ be the pseudo-quadratic function of Definition 6.7.

Let $(\hat{u}_2, v) \in \mathrm{D}(W^+)$. There exists $z^+ = (z_1, z_2, u_1, v_1) \in \mathcal{Z}^+$ such that $\hat{u}_2 \in \hat{\mathrm{X}}^{z^+}(W^+)$ and $v \in \Delta\mathrm{Y}^{z^+}(W^+)$. We have:

- Since $\hat{u}_2 \in \hat{\mathrm{X}}^{z^+}(W^+)$, there exists $v_2 \in \mathrm{Y}^{z^+}(W^+)$ such that $\hat{p}_{v_2,z^+}(\hat{u}_2) \neq 0$. From (6.17) we have

$$\hat{p}_{v_2,z^+,W^+}(\hat{u}_2) = \sum_{\hat{u}_2' \in G_1} \frac{\hat{p}_{v_1+v_2,z_1}(\hat{u}_2') \cdot \hat{p}_{v_2,z_2}(\hat{u}_2 - \hat{u}_2')}{|G_1|\alpha(u_1, z_1, z_2, v_1, v_2)} e^{j2\pi\langle \hat{u}_2', u_1\rangle}.$$

  Since $\hat{p}_{v_2,z^+,W^+}(\hat{u}_2) \neq 0$, there must exist $\hat{u}_2' \in G_1$ such that $\hat{p}_{v_1+v_2,z_1}(\hat{u}_2') \neq 0$ and $\hat{p}_{v_2,z_2}(\hat{u}_2 - \hat{u}_2') \neq 0$. Therefore, $\hat{u}_2' \in \hat{\mathrm{X}}^{z_1}(W)$ and $(\hat{u}_2 - \hat{u}_2') \in \hat{\mathrm{X}}^{z_2}(W)$.

- Since $v \in \Delta\mathrm{Y}^{z^+}(W^+)$, there exist $v_2', v_2'' \in \mathrm{Y}^{z^+}(W^+)$ such that $v = v_2' - v_2''$. Now Lemma 6.17 implies that $v_1 + v_2' \in \mathrm{Y}^{z_1}(W)$, $v_2' \in \mathrm{Y}^{z_2}(W)$, $v_1 + v_2'' \in \mathrm{Y}^{z_1}(W)$ and $v_2'' \in \mathrm{Y}^{z_2}(W)$. Therefore, $v = (v_1 + v_2') - (v_1 + v_2'') \in \Delta\mathrm{Y}^{z_1}(W)$ and $v = v_2' - v_2'' \in \Delta\mathrm{Y}^{z_2}(W)$.

We conclude that

$$(\hat{u}_2', v) \in \hat{\mathrm{X}}^{z_1}(W) \times \Delta\mathrm{Y}^{z_1}(W) = \mathrm{D}^{z_1}(W) \subset \mathrm{D}(W) \subset D$$

and

$$(\hat{u}_2 - \hat{u}_2', v) \in \hat{\mathrm{X}}^{z_2}(W) \times \Delta\mathrm{Y}^{z_2}(W) = \mathrm{D}^{z_2}(W) \subset \mathrm{D}(W) \subset D.$$

Now since D is a pseudo-quadratic domain, we have $(\hat{u}_2, v) = \big(\hat{u}_2' + (\hat{u}_2 - \hat{u}_2'), v\big) \in D$. We conclude that $\mathrm{D}(W^+) \subset D$.

Now let $(\hat{u}_2, z^+) \in \hat{\mathrm{X}}\mathrm{Z}(W^+)$, where $z^+ = (z_1, z_2, u_1, v_1) \in \mathcal{Z}^+$. For every $v_2', v_2'' \in \mathrm{Y}^{z^+}(W^+)$, we have $v_1 + v_2' \in \mathrm{Y}^{z_1}(W)$, $v_2' \in \mathrm{Y}^{z_2}(W)$, $v_1 + v_2'' \in \mathrm{Y}^{z_1}(W)$ and $v_2'' \in$

$\mathrm{Y}^{z_2}(W)$ from Lemma 6.17. Therefore,

$$
\begin{aligned}
&\hat{p}_{v_2',z^+,W^+}(\hat{u}_2) \\
&= \sum_{\hat{u}_2' \in G_1} \frac{\hat{p}_{v_1+v_2',z_1}(\hat{u}_2') \cdot \hat{p}_{v_2',z_2}(\hat{u}_2 - \hat{u}_2')}{|G_1| \alpha(u_1, z_1, z_2, v_1, v_2')} e^{j2\pi\langle \hat{u}_2', u_1 \rangle} \\
&\overset{(a)}{=} \sum_{\substack{\hat{u}_2' \in G_1: \\ \hat{u}_2' \in \hat{\mathrm{X}}^{z_1}(W), \\ \hat{u}_2 - \hat{u}_2' \in \hat{\mathrm{X}}^{z_2}(W)}} \frac{\hat{p}_{v_1+v_2',z_1}(\hat{u}_2') \cdot \hat{p}_{v_2',z_2}(\hat{u}_2 - \hat{u}_2')}{|G_1| \alpha(u_1, z_1, z_2, v_1, v_2')} e^{j2\pi\langle \hat{u}_2', u_1 \rangle} \\
&\overset{(b)}{=} \sum_{\substack{\hat{u}_2' \in G_1: \\ \hat{u}_2' \in \hat{\mathrm{X}}^{z_1}(W), \\ \hat{u}_2 - \hat{u}_2' \in \hat{\mathrm{X}}^{z_2}(W)}} \frac{\hat{p}_{v_1+v_2'',z_1}(\hat{u}_2') F(\hat{u}_2', v_2' - v_2'') \cdot \hat{p}_{v_2'',z_2}(\hat{u}_2 - \hat{u}_2') F(\hat{u}_2 - \hat{u}_2', v_2' - v_2'')}{|G_1| \cdot P_{U_1|Z_1,Z_2,V_1,V_2}(u_1|z_1, z_2, v_1, v_2')} e^{j2\pi\langle \hat{u}_2', u_1 \rangle} \\
&\overset{(c)}{=} \sum_{\hat{u}_2' \in G_1} \frac{\hat{p}_{v_1+v_2'',z_1}(\hat{u}_2') \cdot \hat{p}_{v_2'',z_2}(\hat{u}_2 - \hat{u}_2')}{|G_1| \cdot P_{U_1|Z_1,Z_2,V_1,V_2}(u_1|z_1, z_2, v_1, v_2'')} F(\hat{u}_2' + \hat{u}_2 - \hat{u}_2', v_2' - v_2'') \cdot e^{j2\pi\langle \hat{u}_2', u_1 \rangle} \\
&= F(\hat{u}_2, v_2' - v_2'') \sum_{\hat{u}_2' \in G_1} \frac{\hat{p}_{v_1+v_2'',z_1}(\hat{u}_2') \cdot \hat{p}_{v_2'',z_2}(\hat{u}_2 - \hat{u}_2')}{|G_1| \cdot P_{U_1|Z_1,Z_2,V_1,V_2}(u_1|z_1, z_2, v_1, v_2'')} e^{j2\pi\langle \hat{u}_2', u_1 \rangle} \\
&= F(\hat{u}_2, v_2' - v_2'') \hat{p}_{v_2'',z^+,W^+}(\hat{u}_2),
\end{aligned}
$$

where (a) follows from the fact that $\hat{p}_{v_1+v_2',z_1}(\hat{u}_2') = 0$ if $\hat{u}_2' \notin \hat{\mathrm{X}}^{z_1}(W)$, and $\hat{p}_{v_2',z_2}(\hat{u}_2 - \hat{u}_2') = 0$ if $(\hat{u}_2 - \hat{u}_2') \notin \hat{\mathrm{X}}^{z_2}(W)$. (b) follows from the fact that W is polarization compatible. (c) follows from the fact that F is pseudo-quadratic and the fact that U_1 is conditionally independent of V_2 given (Z_1, Z_2, V_1) (since the polarization compatibility of W implies that I_1 is preserved for W by Lemma 6.9, which implies that $I(U_1; V_2|Z_1 Z_2 V_1) = 0$). Therefore, for every $v_2', v_2'' \in \mathrm{Y}^{z^+}(W^+)$, we have

$$\hat{p}_{v_2',z^+,W^+}(\hat{u}_2) = F(\hat{u}_2, v_2' - v_2'') \cdot \hat{p}_{v_2'',z^+,W^+}(\hat{u}_2).$$

We conclude that W^+ is polarization compatible. □

Lemma 6.10 follows from Lemmas 6.20 and 6.21.

Erasure Schemes Using Generalized Polar Codes

7

The probability of error of polar codes for binary-input channels under successive cancellation decoding was shown to be equal to $o(2^{-N^{1/2-\epsilon}})$ [19], where N is the blocklength. A more refined estimation of the probability of error, which explicitly depends on the transmission rate R, was obtained by Hassani et al. [38]. They showed that the probability of error under successive cancellation decoding of the polar code is equal to $2^{-2^{\frac{n}{2}+\frac{\sqrt{n}}{2}Q^{-1}\left(\frac{R}{I(W)}\right)+o(\sqrt{n})}}$, where $N = 2^n$ is the blocklength, R is the transmission rate, $I(W)$ is the capacity of the binary-input memoryless symmetric (BMS) channel W, and Q is the well known Q-function[1]. They also showed that the probability of error under MAP decoding has the same asymptotic behavior. This does not show a good performance of polar codes in terms of the probability of error because the decay is too slow in the blocklength. One attempt to enhance the performance of polar codes was to apply list decoding with CRC error detection [39].

Another possible way to enhance the performance of polar codes is through decoding with erasure; it is sometimes desirable to allow the receiver not to decide which message was transmitted, especially when there is a feedback from the receiver to the transmitter: If a confusing string of symbols was received (in the sense that there is a high probability of a decoding error to occur, no matter which message the receiver chooses as the decoded message), the receiver can ask the transmitter to resend the message, in the hope that the received string will not be confusing in the next transmission.

There are two types of error when we allow decoding with erasure:

- If the receiver decides on the transmitted message and makes an error, we say that an undetected error occurs.
- If the receiver does not decide, we say that an erasure occurs.

[1] $Q(x) = \mathbb{P}[X \geq x]$, where X is a Gaussian random variable of mean 0 and variance 1.

In general, there is a trade-off between the probability of undetected error p_{ue} and the erasure probability p_{er}: p_{ue} can be made smaller at the expense of a higher p_{er}. The trade-off between these parameters was first studied by Forney [40].

In this chapter[2], we study the tradeoff between these parameters for generalized polar[3] (GP) codes, which are a family of codes that contains, among others, the standard polar codes of Arıkan [2] and Reed-Muller codes[4]. In Section 7.1, we provide the preliminaries of this chapter: We provide a formal definition of erasure schemes, GP codes, and successive cancellation decoders with erasure. In Section 7.2, we study the erasure schemes that are based on GP Codes: We compute the zero-undetected-error capacity of GP codes under the low-complexity successive cancellation decoder with erasure, and we derive an estimate of the erasure probability of GP codes for rates that are below the zero-undetected-error capacity.

7.1 Preliminaries

7.1.1 Useful Notations

For every $0 \leq \epsilon, \epsilon' \leq 1$, define the following:

- $\overline{\epsilon} = 1 - \epsilon$.
- $\epsilon * \epsilon' = \epsilon\overline{\epsilon'} + \overline{\epsilon}\epsilon'$.
- $m(\epsilon) = \min\{\epsilon, \overline{\epsilon}\}$.

For every $x \in \mathbb{F}_2^N$ and every $\mathcal{I} \subset [N] = \{1, \ldots, N\}$, we write $x_{\mathcal{I}} \in \mathbb{F}_2^{\mathcal{I}}$ to denote the subvector containing the components of x whose indices appear in $\mathcal{I}$.

7.1.2 Erasure Schemes

Let $W : \mathbb{F}_2 \longrightarrow \mathcal{Y}$ be a binary-input channel. A coding scheme with erasure is a 4-tuple $\mathcal{C} = (\mathcal{M}, N, f, g)$ where $\mathcal{M}$ is the set of messages, N is the blocklength of the code, $f : \mathcal{M} \to \mathbb{F}_2^N$ is the encoder mapping, and $g : \mathcal{Y}^N \to \mathcal{M} \cup \{\mathbf{e}\}$ is the decoder mapping, where $\mathbf{e} \notin \mathcal{M}$ represents erasure.

The scheme is used as follows:

- The transmitter chooses a message M uniformly in $\mathcal{M}$ and computes $X^N = (X_1, \ldots, X_N) = f(M)$.
- The transmitter sends $X_1, \ldots, X_N$ through N independent copies of the channel W, i.e., he uses the channel N times. The rate R of the coding scheme is the amount of information that is sent per channel use: $R = \dfrac{\log_2 |\mathcal{M}|}{N}$.
- The receiver obtains $Y_1, \ldots, Y_N$ and computes $\hat{M} = g(Y^N) = g(Y_1, \ldots, Y_N)$.

[2] The material of this chapter is based on [41].

[3] See Section 7.1.5 for the definition of generalized polar codes.

[4] The invention of polar codes brought back attention to Reed-Muller codes because of their similarity. It was recently shown that Reed-Muller codes achieve the capacity of binary erasure channels under MAP decoding [42].

- If $\hat{M} = \mathbf{e}$, we say that an erasure has occurred. Thus, the erasure probability of the scheme is $p_{er}(W, \mathcal{C}) = \mathbb{P}(\{\hat{M} = \mathbf{e}\})$.
- If $\hat{M} \neq \mathbf{e}$ and $\hat{M} \neq M$, we say that an undetected error has occured. Therefore, the undetected error probability of the scheme is $p_{ue}(W, \mathcal{C}) = \mathbb{P}\big(\{\hat{M} \notin \{\mathbf{e}, M\}\}\big)$.

In practice, it is desirable to maximize the rate R while minimizing the erasure probability $p_{er}(W, \mathcal{C})$, the undetected-error probability $p_{ue}(W, \mathcal{C})$, the blocklength N, as well as the computational complexity of both the encoder and the decoder. The trade-off between all these performance parameters is one of the important problems in information theory. In this chapter we are interested in studying the trade-off between these parameters asymptotically in N under the following assumptions:

(i) A BMS channel W is used.

(ii) Only GP codes are considered.

(iii) Only successive cancellation decoders with erasure[5] are considered.

7.1.3 Binary-Input Memoryless Symmetric Channels

We encountered binary-input memoryless symmetric (BMS) channels in Definition 1.2. In this chapter, we will adopt a more general definition.

BMS channels generalize binary symmetric channels (BSC). One can think of a BMS channel as "a combination of BSCs": Let $\mathrm{BSC}(\epsilon_1), \ldots, \mathrm{BSC}(\epsilon_l)$ be a collection of l binary symmetric channels of crossover probabilities $\epsilon_1, \ldots, \epsilon_l$ respectively. Let $p_1, \ldots, p_l$ be a probability distribution over $[l] := \{1, \ldots, l\}$ and consider the binary-input channel W which operates as follows: During each use of the channel W, one of the channels $\mathrm{BSC}(\epsilon_1), \ldots, \mathrm{BSC}(\epsilon_l)$ is chosen with probability $p_1, \ldots, p_l$ respectively. The bit at the input of W is transmitted to the receiver through the chosen BSC. Moreover, we assume that the receiver knows which BSC was used in each channel use of W. Formally, the channel $W : \mathbb{F}_2 \to [l] \times \mathbb{F}_2$ can be defined as follows:

$$W(i, y|x) = \begin{cases} p_i \cdot (1 - \epsilon_i) & \text{if } x = y, \\ p_i \cdot \epsilon_i & \text{if } x \neq y. \end{cases} \tag{7.1}$$

We denote this channel W as

$$W = \sum_{i=1}^{l} p_i \cdot \mathrm{BSC}(\epsilon_i).$$

Definition 7.1. *A channel W is said to be binary-input memoryless symmetric (BMS) if there exist $0 \leq \epsilon_1, \ldots, \epsilon_l \leq 1$ and a probability distribution $\{p_1, \ldots, p_l\}$ over $[l] = \{1, \ldots, l\}$ such that W is equivalent (in the sense of Definition 3.6) to the channel $\sum_{i=1}^{l} p_i \cdot \mathrm{BSC}(\epsilon_i)$. In this case, we write*

$$W \equiv \sum_{i=1}^{l} p_i \cdot \mathrm{BSC}(\epsilon_i), \tag{7.2}$$

[5] See Section 7.1.6 for the definition of successive cancellation decoders with erasure.

and we say that this is a BSC*-decomposition of* W*.*

Note that one can define general BMS channels by considering infinite collections of BSCs. The binary-input additive white Gaussian noise channels are examples of general BMS channels with continuous output alphabet. For the sake of simplicity, we will only consider in this chapter BMS channels with finite output alphabets. However, all the main results of this chapter are also valid for general BMS channels.

Another remark that is worth mentioning is that there are infinitely many BSC-decompositions of a given BMS channel W. The reason for this is twofold:

(i) We can decompose or unite BSC-components having the same crossover probability by decomposing or adding their fractions (i.e., the p_i parameters) respectively.

(ii) For every $\epsilon > 0$, we have $\mathrm{BSC}(\epsilon) \equiv \mathrm{BSC}(\overline{\epsilon})$. Therefore, we can change the crossover probability of any BSC component to its complement.

This motivates the following definition:

Definition 7.2. *If* $\epsilon_i \leq \frac{1}{2}$ *for all* $1 \leq i \leq l$*, we say that* $W \equiv \sum_{i=1}^{l} p_i \cdot \mathrm{BSC}(\epsilon_i)$ *is a natural* BSC*-decomposition of* W*. Note that any* BSC*-decomposition can be naturalized as follows:*

$$W \equiv \sum_{i=1}^{l} p_i \cdot \mathrm{BSC}(\epsilon_i) \equiv \sum_{i=1}^{l} p_i \cdot \mathrm{BSC}\left(m(\epsilon_i)\right).$$

If $0 \leq \epsilon_1 < \ldots < \epsilon_l \leq \frac{1}{2}$ *and* $p_i > 0$ *for all* $1 \leq i \leq l$*, we say that* $W \equiv \sum_{i=1}^{l} p_i \cdot \mathrm{BSC}(\epsilon_i)$ *is the canonical* BSC*-decomposition of* W*. It can be shown that the canonical* BSC*-decomposition of* W *is unique.*

Example 7.1. *For every* $0 \leq \epsilon \leq 1$*, the binary erasure channel* $\mathrm{BEC}(\epsilon)$ *is BMS. Moreover, for* $0 < \epsilon < 1$*, its canonical* BSC*-decomposition is*

$$\mathrm{BEC}(\epsilon) \equiv (1-\epsilon) \cdot \mathrm{BSC}(0) + \epsilon \cdot \mathrm{BSC}\left(\frac{1}{2}\right).$$

Definition 7.3. *Let* $W \equiv \sum_{i=1}^{l} p_i \cdot \mathrm{BSC}(\epsilon_i)$*. For every* $0 \leq \epsilon \leq \frac{1}{2}$*, define the fraction* $p_W(\epsilon)$ *of* $\mathrm{BSC}(\epsilon)$ *in* W *as follows:*

$$p_W(\epsilon) = \sum_{i=1}^{l} p_i \cdot \mathbb{1}_{\{m(\epsilon_i)=\epsilon\}}.$$

$p_W(\epsilon)$ *is well defined because it does not depend on the* BSC*-decomposition of* W*. I.e., if* $\sum_{i=1}^{l} p_i \cdot \mathrm{BSC}(\epsilon_i) \equiv \sum_{j=1}^{l'} p'_j \cdot \mathrm{BSC}(\epsilon'_j)$ *then* $\sum_{i=1}^{l} p_i \cdot \mathbb{1}_{\{m(\epsilon_i)=\epsilon\}} = \sum_{j=1}^{l'} p'_j \cdot \mathbb{1}_{\{m(\epsilon'_j)=\epsilon\}}$.

As we will see later, the parameter $p_W(0)$ will play an important role in our analysis. We introduce another parameter which is also of interest for our study:

Definition 7.4. *Let W be a BMS channel. We define the* best imperfect component *of W, denoted $\epsilon_{\text{bic}}(W)$, as follows:*

$$\epsilon_{\text{bic}}(W) = \begin{cases} 0 & \text{if } I(W) = 1, \\ \min\limits_{\substack{\epsilon \in]0, \frac{1}{2}]: \\ p_W(\epsilon) > 0}} \epsilon & \text{if } I(W) < 1. \end{cases}$$

$$= \begin{cases} 0 & \text{if } I(W) = 1, \\ \min\limits_{\substack{1 \le i \le l, \\ p_i > 0,\ 0 < \epsilon_i < 1}} m(\epsilon_i) & \text{if } I(W) < 1. \end{cases}$$

7.1.4 $\mathcal{D}_t$ Decoders for BMS Channels

Definition 7.5. *Let $W = \sum_{i=1}^{l} p_i \cdot \text{BSC}(\epsilon_i)$ and let $0 \le t \le \frac{1}{2}$. Define the decoder $\mathcal{D}_t : [l] \times \mathbb{F}_2 \to \{0, 1, \mathbf{e}\}$ of W as follows:*

$$\mathcal{D}_t(i, x) = \begin{cases} x & \text{if } \epsilon_i \le t, \\ 1 \oplus x & \text{if } \epsilon_i \ge 1 - t, \\ \mathbf{e} & \text{otherwise.} \end{cases}$$

Remark 7.1. *$\mathcal{D}_t$ decoders are desirable because no other decoder with erasure can provide a strictly better trade-off between p_{ue} and p_{er} for the code of blocklength 1 and rate 1. Moreover, $\mathcal{D}_t$ decoders are very easy to implement: We compute the log-likelihood ratio $\text{LLR}(y) = \log \frac{P_{X|Y}(1|y)}{P_{X|Y}(0|y)}$ (where X and Y are the input and output of W respectively) and then compare with $T = \log \frac{1-t}{t}$:*

$$\mathcal{D}_t(y) = \begin{cases} 0 & \text{if } \text{LLR}(y) \le -T, \\ 1 & \text{if } \text{LLR}(y) \ge T, \\ \mathbf{e} & \text{otherwise.} \end{cases}$$

7.1.5 Generalized Polar Codes

Definition 7.6. *A code $f : \mathcal{M} \to \mathbb{F}_2^N$ is said to be a* generalized polar (GP) code *of parameters $(n, r, \mathcal{I}, b)$ if it satisfies the following:*

- *$N = 2^n$, $\mathcal{M} = \mathbb{F}_2^{\mathcal{I}}$ and $b \in \mathbb{F}_2^{N-r}$.*
- *$\mathcal{I} \subset [N] = \{1, \ldots, N\}$ and $|\mathcal{I}| = r$.*
- *$f(u) = F^{\otimes n} \cdot \tilde{u}$, where*

$$F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

  and $\tilde{u} \in \mathbb{F}_2^N$ is such that $\tilde{u}_{\mathcal{I}} = u$ and $\tilde{u}_{\mathcal{I}^c} = b$.

n is called the number of polarization steps *of the GP code. We denote the code f as $\text{GP}(n, r, \mathcal{I}, b)$. Moreover, if $b = 0 \in \mathbb{F}_2^{N-r}$, we write $\text{GP}(n, r, \mathcal{I})$.*

Example 7.2. *Here are two examples of GP codes:*

- *Standard polar codes of Arıkan: Take $\mathcal{I}$ to be the set of indices of the r synthetic channels having the lowest Bhattacharyya parameters, and take b to be the vector of frozen bits.*
- *Reed-Muller codes: Take $\mathcal{I}$ to be the set of indices of the r columns of $F^{\otimes n}$ having the largest number of ones, and take $b = 0 \in \mathbb{F}_2^{N-r}$.*

7.1.6 Successive Cancellation Decoder with Erasure of GP codes

Because of the recursive construction of $F^{\otimes n}$, one can implement the encoder of any GP code in $O(N \log N)$ time exactly like polar codes.

On the other hand, for any given $\mathrm{GP}(n, r, \mathcal{I}, b)$ code, there are various decoders that can be considered. One attractive choice is what we call *successive cancellation decoder with erasure (SCE)* which operates similarly like the successive cancellation decoder of polar codes, but instead of applying the ML decoder for each bit u_i, we apply a $\mathcal{D}_{t_i}$ decoder for some $0 \leq t_i \leq \frac{1}{2}$. The reason why SCE decoders are desirable is because they have low computational complexity.

Definition 7.7. *For every $i \in \mathcal{I}$ let $0 \leq t_i \leq \frac{1}{2}$ and let $t = (t_i)_{i \in \mathcal{I}} \in [0, \frac{1}{2}]^{\mathcal{I}}$. The $\mathcal{D}_t$ successive cancellation decoder with erasure (denoted* SCE-$\mathcal{D}_t$ *or simply $\mathcal{D}_t$) for a $\mathrm{GP}(n, r, \mathcal{I}, b)$ code operates as follows:*

- *For each $i \in \mathcal{I}$, compute $\hat{u}_i$ by applying the $\mathcal{D}_{t_i}$ decoder. The bits are successively decoded exactly in the same order as in the successive cancellation decoder of polar codes.*
- *If $\hat{u}_i = \mathbf{e}$ for any $i \in \mathcal{I}$, stop decoding immediately and declare erasure.*
- *If $\hat{u}_i \neq \mathbf{e}$ for every $i \in \mathcal{I}$, the output is $\hat{u} = (\hat{u}_i)_{i \in \mathcal{I}}$.*

Two remarks are worth mentioning here:

- The computational complexity of any SCE decoder is $O(N \log N)$.
- If $t_i = 0$ for every $i \in \mathcal{I}$, we get a zero-undetected-error scheme.

7.2 Erasure Schemes Using GP Codes

Definition 7.8. *Let $W : \mathbb{F}_2 \longrightarrow \mathcal{Y}$ be a BMS channel and define*

$$I_0^{\mathrm{GP}}(W) := \sum_{\substack{y \in \mathcal{Y}: \\ W(y|1)=0}} W(y|0) = \sum_{\substack{y \in \mathcal{Y}: \\ W(y|0)=0}} W(y|1). \tag{7.3}$$

It can be easily shown that $I_0^{\mathrm{GP}}(W) = p_W(0)$.

The following theorem, which is the main result of this chapter, shows that $I_0^{\mathrm{GP}}(W)$ is the *zero-undetected-error capacity of GP codes for W under* SCE *decoders.*

Theorem 7.1. *Let W be a fixed BMS channel. We have the following:*

- *For every $R < I_0^{\mathrm{GP}}(W)$, every $\beta < \frac{1}{2}$ and every n large enough, there exists a GP code of blocklength $N = 2^n$ and of rate at least R for which the low-complexity $\mathcal{D}_0$-SCE decoder (which induces a zero-undetected-error scheme) has an erasure probability of order $2^{-2^{\beta \cdot n}}$.*

- *For every $\alpha > 0$, every $\beta > \frac{1}{2}$, every n large enough, and every GP code of rate $I_0^{\mathrm{GP}}(W) < R < I(W)$ and blocklength $N = 2^n$, if $p_{er} < 1 - \alpha$ then $p_{ue} > 2^{-2^{\beta \cdot n}}$. In other words, the undetected error probability cannot be made better than $2^{-N^{\frac{1}{2}+o(1)}}$ unless the erasure probability is of order $1 - o(1)$.*

In order to prove Theorem 7.1, we need a few lemmas and propositions. The next proposition shows the first point of the theorem. In fact, it provides a better estimate for the erasure probability:

Proposition 7.1. *Let $W : \mathbb{F}_2 \longrightarrow \mathcal{Y}$ be a BMS channel. For every $R < I_0^{\mathrm{GP}}(W)$, there exists a GP code of blocklength $N = 2^n$ and of rate at least R for which the low-complexity $\mathcal{D}_0$-SCE decoder (which induces a zero-undetected-error scheme) has an erasure probability of order $2^{-2^{\frac{n}{2}+Q^{-1}\left(\frac{R}{I_0^{\mathrm{GP}}(W)}\right)\frac{\sqrt{n}}{2}+o(\sqrt{n})}}$, where $Q(x) = \mathbb{P}(\{\mathcal{N}(0,1) \geq x\})$ is the standard Q-function.*

Proof. Define $W' : \mathbb{F}_2 \longrightarrow \mathbb{F}_2 \cup \{\mathbf{e}\}$ as follows:

$$W'(y'|x) = \begin{cases} \sum\limits_{\substack{y \in \mathcal{Y}: \\ W(y|x\oplus 1)=0}} W(y|x) & \text{if } y' = x, \\ \sum\limits_{\substack{y \in \mathcal{Y}: \\ W(y|x\oplus 1)>0}} W(y|x) & \text{if } y' = \mathbf{e}, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, for each $x \in \mathbb{F}_2$ we contract all the output symbols of W for which we can decide without error that the input was x to one output symbol of W' that we also denote by x. Moreover, we contract all the remaining uncontracted symbols to the erasure symbol $\mathbf{e}$.

Let $\epsilon = 1 - I_0^{\mathrm{GP}}(W)$. One can easily check that $W' = \mathrm{BEC}(\epsilon) \preceq W$. Now for every $R < I_0^{\mathrm{GP}}(W) = 1 - \epsilon = I(W')$, there exists a polar code for W' of rate at least R and whose probability of error under successive cancellation decoder is equal to $2^{-2^{\frac{n}{2}+Q^{-1}\left(\frac{R}{I(W')}\right)\frac{\sqrt{n}}{2}+o(\sqrt{n})}}$ (see [38]). One can use the same code for W and apply the $\mathcal{D}_0$-SCE decoder. This induces a zero-undetected-error scheme.

It can be easily seen that the erasure probability for the $\mathcal{D}_0$-SCE decoder of the GP code for W is of the same order as the error probability of the successive cancellation decoder of the polar code for W'. □

In order to prove the second point of Theorem 7.1, we will need the analysis tools of polarization theory. Let us first recall the basic notations and definitions.

Let $W : \mathbb{F}_2 \longrightarrow \mathcal{Y}$ be a binary-input channel. We define the two channels $W^- : \mathbb{F}_2 \longrightarrow \mathcal{Y} \times \mathcal{Y}$ and $W^+ : \mathbb{F}_2 \longrightarrow \mathcal{Y} \times \mathcal{Y} \times \mathbb{F}_2$ as follows:

$$W^-(y_1, y_2|u_1) = \frac{1}{2} \sum_{u_2 \in \mathbb{F}_2} W(y_1|u_1 \oplus u_2)W(y_2|u_2), \tag{7.4}$$

$$W^+(y_1, y_2, u_1|u_2) = \frac{1}{2} W(y_1|u_1 \oplus u_2)W(y_2|u_2). \tag{7.5}$$

For every $s = (s_1, \ldots, s_n) \in \{-, +\}^n$, we define W^s recursively as

$$W^s := ((W^{s_1})^{s_2} \ldots)^{s_n}.$$

Proposition 7.2. *If W is BMS, then W^- and W^+ are BMS as well. More precisely, if $W \equiv \sum_{i=1}^{l} p_i \cdot \mathrm{BSC}(\epsilon_i)$, then*

$$W^- \equiv \sum_{i=1}^{l} \sum_{j=1}^{l} p_i p_j \cdot \mathrm{BSC}(\epsilon_i * \epsilon_j), \tag{7.6}$$

and

$$W^+ \equiv \sum_{i=1}^{l} \sum_{j=1}^{l} p_i p_j \cdot \left((\epsilon_i * \epsilon_j) \cdot \mathrm{BSC}\left(\frac{\epsilon_i \overline{\epsilon}_j}{\epsilon_i * \epsilon_j} \right) + (\epsilon_i * \overline{\epsilon}_j) \cdot \mathrm{BSC}\left(\frac{\epsilon_i \epsilon_j}{\epsilon_i * \overline{\epsilon}_j} \right) \right). \tag{7.7}$$

Proof. We use Equations (7.1), (7.4) and (7.5) and we apply the fact that $\mathrm{BSC}(\epsilon) \equiv \mathrm{BSC}(\overline{\epsilon})$ for every $\epsilon \in [0, 1]$. □

Proposition 7.2 can be used to derive the effect of polarization on $I_0^{\mathrm{GP}}(W)$ and $\epsilon_{\mathrm{bic}}(W)$:

Corollary 7.1. *$I_0^{\mathrm{GP}}(W^-) = I_0^{\mathrm{GP}}(W)^2$ and $I_0^{\mathrm{GP}}(W^+) = 2I_0^{\mathrm{GP}}(W) - I_0^{\mathrm{GP}}(W)^2$.*

Proof. Let $W \equiv \sum_{i=1}^{l} p_i \cdot \mathrm{BSC}(\epsilon_i)$ be a BSC-decomposition of W. Using the equations of Proposition 7.2, one can see that:

- $I_0^{\mathrm{GP}}(W^-) = p_{W^-}(0) \overset{(a)}{=} p_W(0)^2 = I_0^{\mathrm{GP}}(W)^2$, where (a) follows from the fact that $m(\epsilon_i * \epsilon_j) = 0$ if and only if $m(\epsilon_i) = m(\epsilon_j) = 0$.
- $I_0^{\mathrm{GP}}(W^+) = p_{W^+}(0) \overset{(b)}{=} 2p_W(0) - p_W(0)^2 = 2I_0^{\mathrm{GP}}(W) - I_0^{\mathrm{GP}}(W)^2$, where (b) follows from the fact that

  $$m\left(\frac{\epsilon_i \overline{\epsilon}_j}{\epsilon_i * \epsilon_j} \right) = 0 \quad \Leftrightarrow \quad m(\epsilon_i) = 0 \text{ or } m(\epsilon_j) = 0,$$

  and

  $$m\left(\frac{\epsilon_i \epsilon_j}{\epsilon_i * \overline{\epsilon}_j} \right) = 0 \quad \Leftrightarrow \quad m(\epsilon_i) = 0 \text{ or } m(\epsilon_j) = 0.$$

□

Corollary 7.2. *We have:*

$$\epsilon_{\mathrm{bic}}(W^-) = \begin{cases} 2\epsilon_{\mathrm{bic}}(W) \cdot \overline{\epsilon_{\mathrm{bic}}(W)} & \text{if } p_W(0) = 0, \\ \epsilon_{\mathrm{bic}}(W) & \text{otherwise.} \end{cases}$$

$$\epsilon_{\mathrm{bic}}(W^+) = \frac{\epsilon_{\mathrm{bic}}(W)^2}{\epsilon_{\mathrm{bic}}(W)^2 + (1 - \epsilon_{\mathrm{bic}}(W))^2}.$$

Proof. If $I(W) = 1$ (i.e., $\epsilon_{\mathrm{bic}}(W) = 0$), then $I(W^-) = I(W^+) = 1$ which implies that $\epsilon_{\mathrm{bic}}(W^-) = \epsilon_{\mathrm{bic}}(W^+) = 0$. This shows the corollary for $I(W) = 1$.

Assume now that $I(W) < 1$ so that $\epsilon_{\mathrm{bic}}(W) > 0$. Let $W \equiv \sum_{i=1}^{l} p_i \cdot \mathrm{BSC}(\epsilon_i)$ be the canonical BSC-decomposition of W.

Since $0 \leq \epsilon_i, \epsilon_j \leq \frac{1}{2}$ for every $1 \leq i, j \leq l$, it is easy to see that:

- $0 \leq \epsilon_i * \epsilon_j \leq \frac{1}{2}$. This means that the crossover probabilities appearing in (7.6) do not need to be complemented.
- $\epsilon_i * \epsilon_j = 0$ if and only if $\epsilon_i = \epsilon_j = 0$.

Now since the function $\epsilon * \epsilon'$ is increasing in both ϵ and ϵ' (assuming $0 \leq \epsilon, \epsilon' \leq \frac{1}{2}$), we conclude that

$$\begin{aligned} \epsilon_{\mathrm{bic}}(W^-) &= \min_{\substack{1 \leq i,j \leq l, \\ m(\epsilon_i * \epsilon_j) > 0}} m(\epsilon_i * \epsilon_j) \\ &= \begin{cases} 2\epsilon_{\mathrm{bic}}(W) \cdot (1 - \epsilon_{\mathrm{bic}}(W)) & \text{if } p_W(0) = 0, \\ \epsilon_{\mathrm{bic}}(W) & \text{otherwise.} \end{cases} \end{aligned}$$

We apply a similar reasoning on $m\left(\frac{\epsilon_i \bar{\epsilon}_j}{\epsilon_i * \epsilon_j}\right)$ and $m\left(\frac{\epsilon_i \epsilon_j}{\epsilon_i * \bar{\epsilon}_j}\right)$. We obtain:

$$\begin{aligned} \epsilon_{\mathrm{bic}}(W^+) &= \min\left\{ \frac{\epsilon_i \epsilon_j}{\epsilon_i * \bar{\epsilon}_j}, \frac{\epsilon_i \bar{\epsilon}_j}{\epsilon_i * \epsilon_j}, 1 - \frac{\epsilon_i \bar{\epsilon}_j}{\epsilon_i * \epsilon_j} : \ 1 \leq i, j \leq l, \ \epsilon_i > 0, \ \epsilon_j > 0 \right\} \\ &= \frac{\epsilon_{\mathrm{bic}}(W)^2}{\epsilon_{\mathrm{bic}}(W)^2 + (1 - \epsilon_{\mathrm{bic}}(W))^2}. \end{aligned}$$

□

Proposition 7.3. *Let $W : \mathbb{F}_2 \longrightarrow \mathcal{Y}$ be a BMS channel and let $GP(n, r, \mathcal{I}, b)$ be a generalized polar code of rate $R = \frac{r}{2^n}$ and blocklength $N = 2^n$. If $I_0^{\mathrm{GP}}(W) < R < I(W)$ then for every $\beta > \frac{1}{2}$, every $\alpha > 0$ and every n large enough, there is no* SCE *decoder which can make the undetected error probability lower than 2^{-N^β} unless it makes the erasure probability at least $1 - \alpha$.*

Proof. Let $(B_n)_{n \geq 1}$ be i.i.d. uniform random variables in $\{-, +\}$. Define the channel-valued process $(W_n)_{n \geq 0}$ as follows:

$$\begin{aligned} W_0 &:= W, \\ W_n &:= W_{n-1}^{B_n} \ \forall n \geq 1. \end{aligned}$$

Let $\frac{1}{2} < \beta' < \beta$ and let n be large enough so that we have $\frac{\alpha}{2} \cdot 2^{-N^{\beta'}} \geq 2^{-N^{\beta}}$, where $N = 2^n$.

Corollary 7.1 shows that the process $I_0^{\mathrm{GP}}(W_n)$ is a martingale process. Therefore, $I_0^{\mathrm{GP}}(W_n)$ converges almost surely. Moreover, one can show by standard polarization theory techniques that $I_0^{\mathrm{GP}}(W_n) = p_{W_n}(0)$ converges almost surely to 0 or 1. Furthermore, for every $\epsilon > 0$ we have:

$$\lim_{n\to\infty} \mathbb{P}(\{p_{W_n}(0) < \epsilon\}) = 1 - p_W(0).$$

Therefore, as n becomes large, the fraction of indices $s \in \{-,+\}^n$ such that $p_{W^s}(0) \geq \epsilon$ is roughly at most $I_0^{\mathrm{GP}}(W) = p_W(0)$.

On the other hand, from Corollary 7.2, we can easily see that $\epsilon_{\mathrm{bic}}(W^-) \geq \epsilon_{\mathrm{bic}}(W)$ and $\epsilon_{\mathrm{bic}}(W^+) \geq \epsilon_{\mathrm{bic}}(W)^2$. By applying the same analysis of [19], but to ϵ_{bic} instead of the Bhattacharyya parameter, one can show that if $I(W) < 1$, then the fraction of indices $s \in \{-,+\}^n$ such that $\epsilon_{\mathrm{bic}}(W^s) \geq 2^{-2^{\beta' n}}$ goes to 1. Therefore, for n large enough, if $R > I_0^{\mathrm{GP}}(W) = p_W(0)$, there exists at least one index $s \in \{-,+\}^n$ whose corresponding index in $F^{\otimes n}$ appears in the generator matrix of the GP code and which satisfies $\epsilon_{\mathrm{bic}}(W^s) > 2^{-2^{\beta' n}}$ and $p_{W^s}(0) < \frac{\alpha}{2}$. Let $i \in [2^n]$ be the index of the column of $F^{\otimes n}$ corresponding to s and let $0 \leq t_i \leq \frac{1}{2}$ be the threshold used for W^s in an SCE $-\mathcal{D}_t$ decoder. Let $p_{ue}^{(i)}$ and $p_{er}^{(i)}$ be the erasure probability and undetected error probability of the $\mathcal{D}_{t_i}$ decoder applied to W^s respectively. We have:

$$p_{er}^{(i)} = \sum_{\epsilon > t_i} p_W(\epsilon),$$

and

$$\begin{aligned} p_{ue}^{(i)} &= \sum_{\epsilon \leq t_i} \epsilon \cdot p_{W^s}(\epsilon) = \sum_{\epsilon_{\mathrm{bic}}(W^s) \leq \epsilon \leq t_i} \epsilon \cdot p_{W^s}(\epsilon) \\ &\geq \sum_{\epsilon_{\mathrm{bic}}(W^s) \leq \epsilon \leq t_i} \epsilon_{\mathrm{bic}}(W^s) \cdot p_{W^s}(\epsilon) \\ &= \epsilon_{\mathrm{bic}}(W^s) \cdot (1 - p_{W^s}(0) - p_{er}^{(i)}) \\ &\geq 2^{-N^{\beta'}} \cdot \left(1 - \frac{\alpha}{2} - p_{er}^{(i)}\right). \end{aligned} \tag{7.8}$$

Therefore, if $p_{er}^{(i)} \leq 1 - \alpha$ then $p_{ue}^{(i)} \geq \frac{\alpha}{2} \cdot 2^{-N^{\beta'}} \geq 2^{-N^{\beta}}$. Hence $p_{ue}^{(i)}$ cannot be made less than $2^{-N^{\beta}}$ unless $p_{er}^{(i)}$ is at least $1 - \alpha$. The proposition now follows from the fact that the erasure probability and the undetected error probability of the whole scheme are lower bounded by $p_{er}^{(i)}$ and $p_{ue}^{(i)}$ respectively. □

The proof of Theorem 7.1 now follows from Propositions 7.1 and 7.3.

Polar Codes for Arbitrary Classical-Quantum Channels

8

The polarization phenomenon can be generalized to the setting where the input of the channel is classical and the output is a quantum state. Wilde and Guha constructed polar codes for binary-input classical-quantum channels[1] (cq-channel) in [43]. They showed that using the same polarization transformation of Arıkan yields polarization of the synthetic cq-channels to almost useless and almost perfect channels. Wilde and Guha proposed a quantum successive cancellation decoder and showed that its probability of error decays faster than $2^{-N^{\beta}}$ for any $\beta < \frac{1}{2}$. In [44], Hirche et. al. constructed codes for binary-input classical-quantum multiple-access channels[2] (cq-MAC) by combining the polarization results of [43] with the monotone chain rule method of Arıkan [22].

In this chapter[3], we construct polar codes for arbitrary cq-channels and arbitrary cq-MACs by using arbitrary Abelian group operations on the input alphabets. The polarization transformation that we use is similar to the one in [6]. Since we are proving a quantum version of the results in [4] and [6], many ideas of these two papers were adopted and adapted to the quantum setting. However, some inequalities that were used in [4] and [6] do not have quantum analogues. Therefore, other inequalities that serve the same purpose needed to be shown for cq-channels.

In Section 8.1, we provide a very brief introduction to quantum mechanics. For a more detailed discussion of quantum mechanics, see [47, Chapter 2]. The main purpose of Section 8.1 is to make this chapter accessible for readers who are not familiar with quantum mechanics. Readers already familiar with quantum mechanics may skip ahead to Section 8.1.4 where we describe the non-commutative union bound. In Section 8.2, we define classical-quantum channels and explain some basic results that we will use later. In Section 8.3, we describe the polarization process. In Section 8.4, we show that we have a two-level polarization if the cq-channel has $\mathbb{F}_q$ as its input alphabet, where q is a prime number. In Section 8.5, we prove multilevel polarization for arbitrary cq-channels using an arbitrary Abelian group operation on the input alphabet. We show that the synthetic cq-channels converge to deter-

[1] The definition of classical-quantum channels can be found in Section 8.2.

[2] The definition of classical-quantum multiple-access channels can be found in Section 8.7.

[3] The material of this chapter is based on [45, 46].

ministic homomorphism channels that project their input onto a quotient group of the input alphabet. We discuss the rate of polarization (i.e., how fast the synthetic cq-channels polarize) in Section 8.6. We discuss the construction of polar codes in Section 8.7. As in all polar coding schemes, the encoder can be implemented in $O(N \log N)$ operations, where N is the blocklength of the polar code. We prove that the probability of error of the quantum successive cancellation decoder decays faster than 2^{-N^β} for any $\beta < \frac{1}{2}$, but we do not have an efficient implementation of the decoder. We discuss the polarization of arbitrary cq-MACs in Section 8.8. We show that while cq-MAC-polar codes might not achieve the entire symmetric-capacity region, they always achieve points on the dominant face. We show that the entire symmetric-capacity region can be achieved by combining the cq-channel polarization result either with the rate-splitting method of [8], or with the monotone chain rule method of [22].

8.1 Introduction to Quantum Mechanics

From a pedagogical point of view, the conventional wisdom in writing an introduction to any field is to start by an informal discussion (in order to build an intuition about the topic), and then provide a formal description of the subject. However, we do not believe that this is the best approach to follow in the case of quantum mechanics: The purpose of informal discussions is to explain the ideas of the subject in terms of concepts that the reader is already familiar with, whereas quantum mechanics is fundamentally different than everything that we are used to in our everyday life.

Any informal description of quantum mechanics is bound to use philosophical statements and interpretations that are inaccurate (or at best misleading). In our introduction to quantum mechanics, we will avoid using such interpretations and try to be as philosophically neutral as possible.

8.1.1 Closed quantum systems

We start by providing the mathematical formalism describing closed quantum systems[4].

The state space

We first describe the simplest quantum system: the quantum bit (qubit). Unlike the classical bit, which can only be in one of two states (either 0 or 1), the qubit[5] can be in an "arbitrary superposition of the states 0 and 1". By superposition, we mean a "linear combination of the states 0 and 1". We represent the state of a qubit as a unit complex vector $|\psi\rangle$ of dimension 2. The states 0 and 1 are represented by the vectors

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \in \mathbb{C}^2 \quad \text{and} \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

[4] A closed quantum system is a physical system that does not interact with its environment.

[5] The polarization of one photon can be represented as a one-qubit system: The states 0 and 1 correspond to horizontal and vertical polarization, respectively.

respectively. A general state of a one-qubit system can be represented by the complex vector

$$|\psi\rangle = \begin{pmatrix}\alpha_0\\ \alpha_1\end{pmatrix} = \alpha_0|0\rangle + \alpha_1|1\rangle \in \mathbb{C}^2,$$

where $\alpha_0, \alpha_1 \in \mathbb{C}$ and $|\alpha_0|^2 + |\alpha_1|^2 = 1$. We can see that there are infinitely many possible states for a qubit.

In a system of two qubits, the state is described by a unit complex vector of dimension 4. The states 00, 01, 10 and 11 are represented by the vectors

$$|00\rangle = \begin{pmatrix}1\\0\\0\\0\end{pmatrix} \in \mathbb{C}^4, \quad |01\rangle = \begin{pmatrix}0\\1\\0\\0\end{pmatrix} \in \mathbb{C}^4, \quad |10\rangle = \begin{pmatrix}0\\0\\1\\0\end{pmatrix} \in \mathbb{C}^4, \quad \text{and} \quad |11\rangle = \begin{pmatrix}0\\0\\0\\1\end{pmatrix} \in \mathbb{C}^4,$$

respectively. A general state of a two-qubits system can be represented by the complex vector

$$|\psi\rangle = \begin{pmatrix}\alpha_{00}\\ \alpha_{01}\\ \alpha_{10}\\ \alpha_{11}\end{pmatrix} = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle,$$

where $\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11} \in \mathbb{C}$ and $|\alpha_{00}|^2 + |\alpha_{01}|^2 + |\alpha_{10}|^2 + |\alpha_{11}|^2 = 1$.

In a system of n qubits, the state is described by a unit complex vector of dimension 2^n, which is a superposition of the states $\big\{|b_1 \dots b_n\rangle : \ b_1, \dots, b_n \in \{0,1\}\big\}$:

$$|\psi\rangle = \sum_{(b_1,\dots,b_n)\in\{0,1\}^n} \alpha_{b_1\dots b_n}|b_1 \dots b_n\rangle \in \mathbb{C}^{2^n},$$

where $\alpha_{b_1\dots b_n} \in \mathbb{C}$ for every $(b_1, \dots, b_n) \in \{0,1\}^n$ and $\sum_{(b_1,\dots,b_n)\in\{0,1\}^n} |\alpha_{b_1\dots b_n}|^n = 1$.

The state of a general closed quantum system A is determined by a unit vector in a complex Hilbert space $\mathcal{H}_A$ that is called the *state space* of the system A. For example, the state space of a quantum system of n qubits is $\mathbb{C}^{2^n}$. In this thesis, we only consider quantum systems whose state spaces are finite dimensional. Therefore, we can assume without loss of generality that the state space is $\mathbb{C}^d$, where d is the dimension of the state space.

Remark 8.1. *The unit vector that can represent the physical state is not unique: If $|\psi\rangle, |\phi\rangle \in \mathcal{H}_A$ are two unit vectors satisfying $|\psi\rangle = e^{j\theta}|\phi\rangle$ for some $\theta \in \mathbb{R}$, then $|\psi\rangle$ and $|\phi\rangle$ represent the same physical state. In other words, the "global" phase of the unit vector is physically irrelevant.*

If we want a one-to-one representation of the physical states, we have to consider the projective Hilbert space[6] *corresponding to the state space: The set of physical states is in one-to-one correspondence with the rays of the projective Hilbert space.*

[6]The projective Hilbert space corresponding to a Hilbert space $\mathcal{H}$ is the quotient of $\mathcal{H} \setminus \{0\}$ by the equivalence relation $\equiv$ defined as $v \equiv w \ \Leftrightarrow \ \exists \lambda \in \mathbb{C}, \ v = \lambda w$. The equivalence classes are called rays.

Notation 8.1. *The inner product between the states $|\psi\rangle$ and $|\phi\rangle$ is denoted as*

$$\langle\psi|\phi\rangle.$$

This is called the bra-ket notation, which is widely used in quantum mechanics. The first part (namely, $\langle\psi|$) is called the bra part of the braket. The second part (namely, $|\phi\rangle$) is called the ket part. This is why the state vectors are also called ket-vectors. $\langle\psi|$ is called the bra-vector that is associated to the ket-vector $|\psi\rangle$.

It is useful to think of a ket-vector $|\psi\rangle$ as a column matrix, and to interpret its associated bra-vector $\langle\psi|$ as the complex conjugate of the column matrix $|\psi\rangle$. In this case, the bra-ket $\langle\psi|\phi\rangle$, which is the inner product between $|\psi\rangle$ and $|\phi\rangle$, is exactly the result of the matrix multiplication of the bra-vector $\langle\psi|$ with the ket-vector $|\phi\rangle$, i.e., $\langle\psi|\phi\rangle = (\langle\psi|) \cdot (|\phi\rangle)$.

For example, if $|\psi\rangle, |\phi\rangle \in \mathbb{C}^2$ are two ket-vectors in the state space of a one-qubit system

$$|\psi\rangle = \begin{pmatrix}\alpha_0\\ \alpha_1\end{pmatrix} = \alpha_0|0\rangle + \alpha_1|1\rangle \in \mathbb{C}^2 \quad \text{and} \quad |\phi\rangle = \begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix} = \beta_0|0\rangle + \beta_1|1\rangle \in \mathbb{C}^2,$$

then

$$\langle\psi|\phi\rangle = \begin{pmatrix}\alpha_0^* & \alpha_1^*\end{pmatrix}\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix} = \alpha_0^*\beta_0 + \alpha_1^*\beta_1.$$

Evolution in time

If a quantum system A is closed, then for every $t_1, t_2 \in \mathbb{R}$, there exists a unitary operator $U_{t_1,t_2} : \mathcal{H}_A \to \mathcal{H}_A$ such that if $|\psi_{t_1}\rangle$ and $|\psi_{t_2}\rangle$ are the states of the system at time t_1 and t_2, respectively, then

$$|\psi_{t_2}\rangle = U_{t_1,t_2}|\psi_{t_1}\rangle.$$

In other words, the state of a closed quantum system evolves unitarily[7].

Measurements

In contrast with the classical world, the state of a quantum system cannot be perfectly determined by observation and experiment. For example, let $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ be the state of a one-qubit system, and assume that $|\psi\rangle$ is unknown. There is no experiment that enables us to know exactly the state of the system. We emphasize that this impossibility is not due to the ambiguity in the global phase[8].

There is a set of measurements that are physically possible, but none of them enables us to determine the state of the quantum system perfectly. In the following, we describe the set of measurements that we can "perform on a closed quantum system"[9].

[7] The time evolution operator U_{t_1,t_2} can be determined by the general Schrödinger equation.

[8] There is no experiment that enables us to find two complex numbers α_0', α_1' such that $\alpha_0'|0\rangle + \alpha_1'|1\rangle = e^{j\theta}|\psi\rangle$ for some $\theta \in \mathbb{R}$.

[9] By definition, a measurement is an interaction between the quantum system with a measuring device M. This means that it is impossible to measure a closed quantum system without making it open. By "performing a measurement on a closed quantum system A", we mean that the measurement is performed while the composite system AM is closed. We emphasize that the measurement (at least in the presented formalism) is *not* a unitary evolution of the composite system AM.

Let A be a closed quantum system whose state space is $\mathcal{H}_A$. A measurement of the system A is a physical process that is applied on the system at the end of which we obtain one of several possible outcomes[10].

A measurement is characterized by a collection of n orthogonal projections[11] $P_1, \ldots, P_n$ of $\mathcal{H}_A$, which satisfy:

- $P_1, \ldots, P_n$ are mutually orthogonal: $P_i P_j = 0$ for every $i, j \neq 0$.
- $P_1, \ldots, P_n$ add up to the identity: $\sum_{i=1}^{n} P_i = I$, where I is the identity mapping on $\mathcal{H}_A$.

$m \in \{1, \ldots, n\}$ represents the different possible outcomes of the measurement.

If $|\psi\rangle$ is the state of the quantum system before applying the measurement $\{P_1, \ldots, P_n\}$, then the measurement outcome will be equal to $m \in \{1, \ldots, n\}$ with probability[12]

$$\mathbb{P}(\{\text{outcome} = m\}) = \langle\psi|P_m|\psi\rangle.$$

Furthermore, if the measurement outcome m occurs, then the post-measurement state is equal to

$$\frac{1}{\sqrt{\langle\psi|P_m|\psi\rangle}} P_m|\psi\rangle.$$

For example, consider a one-qubit system, and consider the measurement $\{P_0, P_1\}$, where

$$P_0 = |0\rangle\langle 0| = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad P_1 = |1\rangle\langle 1| = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Let $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ be the state of the qubit. A simple calculation shows that if we measure the state $|\psi\rangle$ with the measurement $\{P_0, P_1\}$, then the outcome 0 (resp. 1) occurs with probability $|\alpha_0|^2$ (resp. $|\alpha_1|^2$). Moreover, if the outcome 0 (resp. 1) occurs, then the post-measurement state is $|0\rangle$ (resp. $|1\rangle$). This is why we say that "$\{P_0, P_1\}$ measures the bit-value of the qubit".

Composite quantum systems

If two quantum systems A and B have state spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, then the state space of the composite system AB is equal to the tensor product of the individual state spaces:

$$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B.$$

[10] For example, if we are measuring the bit-value of a qubit, we get one of two possible outcomes: 0 or 1.

[11] An orthogonal projection on a Hilbert space $\mathcal{H}$ is a linear mapping P from $\mathcal{H}$ to itself which satisfies the following:

- P is a projection: $P^2|\psi\rangle = P|\psi\rangle$ for every $|\psi\rangle \in \mathcal{H}$.
- P is self-adjoint: $P^\dagger = P$.

[12] The philosophical interpretation of the probabilitic nature of measurement is left to the reader. We stick to the frequentist interpretation because it is exactly what is tested in practice: If there is a large number of copies of the system, all of which are in the state $|\psi\rangle$, and if we perform the same measurement $\{P_1, \ldots, P_n\}$ on all of the copies, then the fraction of times we get the outcome m will be very close to $\langle\psi|P_m|\psi\rangle$.

For example, if A represents a system of n_A qubits and B represents a system of n_B qubits, then

$$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B = \mathbb{C}^{2^{n_A}} \otimes \mathbb{C}^{2^{n_B}} \equiv \mathbb{C}^{2^{n_A} \times 2^{n_B}} = \mathbb{C}^{2^{n_A + n_B}},$$

which means that the composite system AB represents a system of $n_A + n_B$ qubits, as expected.

8.1.2 Open quantum systems

In Section 8.1.1, we described the quantum mechanics of closed systems, which are too idealistic. Perhaps the only truly closed system is the whole universe. Assume that we are only interested in a system A that is open. Let E be the system describing the rest of the universe, i.e., E represents the environment of A.

The formalism of Section 8.1.1 does not help us in describing the quantum mechanics of the system A because it is open. Nevertheless, we can still use this formalism to describe the quantum mechanics of the composite system AE. Let $|\psi\rangle \in \mathcal{H}_{AE} = \mathcal{H}_A \otimes \mathcal{H}_E$ be the ket-vector that describes the state of AE.

Assume that we perform a measurement that acts only on the system A. This measurement must be of the form $\{P_1 \otimes I_E, \ldots, P_n \otimes I_E\}$, where $P_1, \ldots, P_n$ are orthogonal projections acting on $\mathcal{H}_A$, and I_E is the identity operator on E. The probability of getting the outcome $m \in \{1, \ldots, n\}$ is equal to

$$\begin{aligned}\mathbb{P}(\{\text{outcome} = m\}) &= \langle\psi|P_m \otimes I_E|\psi\rangle = \operatorname{Tr}\left((P_m \otimes I_E)|\psi\rangle\langle\psi|\right) \\ &= \operatorname{Tr}_A \operatorname{Tr}_E\left((P_m \otimes I_E)|\psi\rangle\langle\psi|\right) = \operatorname{Tr}_A\left(P_m \operatorname{Tr}_E(|\psi\rangle\langle\psi|)\right),\end{aligned}$$

where Tr_A (resp. Tr_E) is the partial trace with respect to the system A (resp. E). If we define

$$\rho = \operatorname{Tr}_E(|\psi\rangle\langle\psi|),$$

we get

$$\mathbb{P}(\{\text{outcome} = m\}) = \operatorname{Tr}_A(P_m\rho) = \operatorname{Tr}(P_m\rho).$$

This means that if we know ρ, then we can compute the probability distribution of the outcome of any measurement that acts on A. ρ is called the *density-matrix* that represents the state of the open system A. It is easy to see that if the outcome of the measurement is m, then the post-measurement density matrix is equal to

$$\frac{P_m \rho P_m}{\operatorname{Tr}(P_m\rho)}.$$

We can also show that if $|\psi\rangle$ is the state of the composite system AE and if $|\psi\rangle$ was subjected to a unitary operator $U \otimes I_E$ (i.e., it acts only on A), then the resulting density matrix after the unitary evolution is $U\rho U^\dagger$. We conclude that if we are only interested in the system A and if all the quantum operations that are applied act only on A, then we do not need to know the state $|\psi\rangle$ of the whole system: We just need to know the density matrix of the system A. This motivates us to represent the state of an open system by its density matrix.

Density matrices have the following properties:

- ρ is a positive semi-definite operator acting on $\mathcal{H}_A$.
- The trace of ρ is equal to 1.

8.1.3 POVM measurements

In practice, it is not easy to perfectly devise a measurement that *only* acts on a desired system A. A general measurement cannot be described by a *projective measurement*[13].

If we are not interested in specifying the post-measurement state, then one possible way to describe a general measurement is through the POVM[14] formalism.

A POVM measurement is described by a collection of n operators $\{E_1, \dots, E_n\}$ that satisfy:

- $E_1, \dots, E_n$ are positive semi-definite operators acting on $\mathcal{H}_A$.
- $\sum_{i=1}^{n} E_i = I$, where I is the identity operator on $\mathcal{H}_A$.

If a POVM measurement $\{E_1, \dots, E_n\}$ is applied on an open system of state ρ, then the probability that the outcome $m \in \{1, \dots, n\}$ will occur is:

$$\mathbb{P}(\{\text{outcome} = m\}) = \text{Tr}(E_m \rho).$$

We emphasize that the POVM formalism does not enable us to specify the post-measurement state. This is because POVM measurements do not have unique physical implementations. Nevertheless, for every POVM measurement $\{E_1, \dots, E_n\}$, there exists one implementation of it such that the post-measurement state corresponding to the outcome m is $\frac{\sqrt{E_m}\rho\sqrt{E_m}}{\text{Tr}(E_m\rho)}$.

8.1.4 Non-Commutative Union Bound

Sen proved in [48] the following "non-commutative union bound":

$$1 - \text{Tr}(\Pi_r \dots \Pi_1 \rho \Pi_1 \dots \Pi_r) \leq 2\sqrt{\sum_{i=1}^{r}(1 - \text{Tr}(\Pi_i \rho))}, \tag{8.1}$$

where $\Pi_1, \dots, \Pi_r$ are projection operators. This inequality was used in [43] to upper bound the probability of error of the quantum successive cancellation decoder of the polar code constructed for a binary-input cq-channel. This was possible because the measurements used in [43] are projective. In this chapter, the quantum successive cancellation decoder that we propose uses general POVM measurement. Therefore, we cannot use the inequality (8.1).

We provide a "non-commutative union bound" that is looser than (8.1) by a multiplicative factor of $\sqrt{r}$, but it is more general so that it can be applied to general POVMs.

Lemma 8.1. *Let $\Pi_1, \dots, \Pi_r$ be r semi-definite positive operators satisfying $\Pi_1 \leq I, \dots, \Pi_r \leq I$. We have:*

$$1 - \text{Tr}\left(\sqrt{\Pi_r} \dots \sqrt{\Pi_1}\rho\sqrt{\Pi_1} \dots \sqrt{\Pi_r}\right) \leq 2\sqrt{r}\sqrt{\sum_{i=1}^{r}(1 - \text{Tr}(\Pi_i \rho))}.$$

[13] The measurement procedure that was described in Section 8.1.1 is called a *projective measurement*.

[14] POVM stands for *Positive Operator-Valued Measure*.

Proof. See Appendix 8.9.1. □

8.2 Classical-Quantum Channels

A classical-quantum (cq) channel $W : x \in G \longrightarrow \rho_x \in \mathcal{DM}(k)$ takes a classical input $x \in G$ and has a quantum output $\rho_x \in \mathcal{DM}(k)$, where $\mathcal{DM}(k)$ is the space of density matrices of dimension $k < \infty$. We assume that the input alphabet G is finite but its size $q = |G|$ can be arbitrary.

8.2.1 Coding for a Classical-Quantum Channel

A coding scheme for a cq-channel $W : x \in G \longrightarrow \rho_x \in \mathcal{DM}(k)$ is a 4-tuple $(\mathcal{M}, N, f, \mathcal{D})$. $\mathcal{M}$ is the *message set*, N is the *blocklength*, $f : \mathcal{M} \to \mathcal{X}^N$ is the *encoder* and $\mathcal{D}$ is the *(quantum) decoder*. $\mathcal{D} = \{D_m\}_{m\in\mathcal{M}}$ is a POVM measurement that is indexed by $\mathcal{M}$. Every operator D_m acts on $(\mathbb{C}^k)^{\otimes N}$, which is the state-space of the system describing N cq-channel outputs.

The coding scheme is implemented as follows:

- A random message M is uniformly chosen from $\mathcal{M}$.
- The transmitter computes $(X_1, \ldots, X_N) = f(M)$.
- The transmitter sends $X_1, \ldots, X_N$ to the receiver by using the cq-channel N times.
- The receiver observes the output system, which is in the state $\rho_{X_1} \otimes \cdots \otimes \rho_{X_N}$.
- The receiver applies the POVM measurement $\mathcal{D} = \{D_m\}_{m\in\mathcal{M}}$ on the output system. Let $\hat{M} \in \mathcal{M}$ be the measurement outcome.

The rate of the coding scheme is $\frac{\log_2|\mathcal{M}|}{N}$. The probability of error is given by

$$\mathbb{P}[\{\hat{M} \neq M\}] = 1 - \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \operatorname{Tr}\left(D_m \cdot (\rho_{f_1(m)} \otimes \cdots \otimes \rho_{f_N(m)})\right),$$

where $f(m) = (f_1(m), \ldots, f_N(m))$.

8.2.2 Quantum-Information Theoretic Quantities

If the input to the cq-channel $W : x \in G \longrightarrow \rho_x \in \mathcal{DM}(k)$ is uniformly distributed, we can describe the state of the joint input-output system as the state $\rho^{XB} \in \mathcal{DM}(q \cdot k)$ defined as:

$$\rho^{XB} := \frac{1}{q}\sum_{x\in G} |x\rangle\langle x| \otimes \rho_x.$$

A very important quantity associated with W is the symmetric Holevo information $I(W)$ defined as:

$$\begin{aligned} I(W) := I(X;B)_\rho &:= H(X)_\rho + H(B)_\rho - H(XB)_\rho \\ &:= H(\rho^X) + H(\rho^B) - H(\rho^{XB}) \\ &:= H\big(\operatorname{Tr}_B(\rho^{XB})\big) + H\big(\operatorname{Tr}_X(\rho^{XB})\big) - H(\rho^{XB}), \end{aligned}$$

where $H(\sigma)$ is the von Neumann entropy of the density matrix σ:

$$H(\sigma) = -\operatorname{Tr}(\sigma \log_2 \sigma).$$

It is easy to show that

$$I(W) = H\left(\frac{1}{q}\sum_{x\in G}\rho_x\right) - \frac{1}{q}\sum_{x\in G} H(\rho_x).$$

The quantity $I(W)$ is the capacity for transmitting classical information over the cq-channel W when the prior input distribution is restricted to be uniform in G. We have $0 \le I(W) \le \log_2 q$.

Besides $I(W)$, we will need another parameter that measures the reliability of the cq-channel W. For the binary-input case, the fidelity between the two output states was used as a measure of reliability in [43]. In our case, we have q output states, so we will consider the average pairwise fidelity between them (similarly to the average Bhattacharyya distance defined in [4]):

$$F(W) := \frac{1}{q(q-1)} \sum_{\substack{x,x'\in G,\\ x\neq x'}} F(\rho_x, \rho_{x'}),$$

where $F(\rho,\sigma) = \operatorname{Tr}\sqrt{\rho^{\frac{1}{2}}\sigma\rho^{\frac{1}{2}}} = \left\|\sqrt{\sigma}\sqrt{\rho}\right\|_1$, and $\|A\|_1$ is the nuclear norm of the matrix A:

$$\|A\|_1 = \operatorname{Tr}\sqrt{A^\dagger A}.$$

Clearly, $0 \le F(W) \le 1$. We adopt the convention $F(W) := 0$ if $|G| = 1$.

It was shown in [49] that $P_e(W) \le (q-1)F(W)$, where $P_e(W)$ is the probability of error of the optimal decoder of W. This shows that if $F(W)$ is small then $P_e(W)$ is also small and so W is reliable. Intuitively, this is true because a small $F(W)$ means that all the pairwise fidelities are small, which implies that all the output states are easily distinguishable from each other, which in turn should allow a reliable decoding.

The following proposition provides three inequalities that relate $I(W)$ and $F(W)$.

Proposition 8.1. *We have:*

(i) $I(W) \ge \log_2 \dfrac{q}{1+(q-1)F(W)}$.

(ii) $I(W) \le \log_2(q/2) + \sqrt{1-F(W)^2}$.

(iii) $I(W) \le \log_2\left(1 + \sqrt{q^2 - (1+(q-1)F(W))^2}\right)$.

Proof. See Appendix 8.9.2. □

In the above proposition, the first inequality implies that if $I(W)$ is close to 0 then $F(W)$ is close to 1. The same inequality also implies that if $F(W)$ is close to 0 then $I(W)$ is close to $\log_2 q$. The second inequality implies that if $I(W)$ is close to $\log_2 q$ then $F(W)$ is close to 0. The third inequality implies that if $F(W)$ is close to 1 then $I(W)$ is close to 0.

8.3 Polarization Process for Classical-Quantum Channels

Since any set can be endowed with an Abelian group operation, we may assume that one such operation on G is fixed. We will denote this Abelian group operation additively.

Let $W : x \in G \longrightarrow \rho_x \in \mathcal{DM}(k)$ be a cq-channel. Define the cq-channels $W^- : u_1 \in G \longrightarrow \rho^-_{u_1} \in \mathcal{DM}(k^2)$ and $W^+ : u_2 \in G \longrightarrow \rho^+_{u_2} \in \mathcal{DM}(k^2 \cdot q)$ as:

$$\rho^-_{u_1} = \frac{1}{q} \sum_{u_2 \in G} \rho_{u_1+u_2} \otimes \rho_{u_2},$$

and

$$\rho^+_{u_2} = \frac{1}{q} \sum_{u_1 \in G} \rho_{u_1+u_2} \otimes \rho_{u_2} \otimes |u_1\rangle\langle u_1|.$$

Moreover for every $n > 0$ and every $s = (s_1, \ldots, s_n) \in \{-,+\}^n$, define $W^s = (\ldots((W^{s_1})^{s_2})\ldots)^{s_n}$.

Remark 8.2. *W^- and W^+ can be constructed as follows:*

- *Two independent and uniform random variables U_1, U_2 are generated in G.*
- *$X_1 = U_1 + U_2$ and $X_2 = U_2$ are computed.*
- *X_1 is sent through one copy of the cq-channel W. Let B_1 be the quantum system describing the output.*
- *X_2 is sent through another copy of the cq-channel W (independent from the one that was used for X_1). Let B_2 be the quantum system describing the output.*

It can be easily seen that the cq-channels $U_1 \longrightarrow B_1B_2$ and $U_2 \longrightarrow B_1B_2U_1$ simulate W^- and W^+ respectively.

We have:

$$\begin{aligned} I(W^-) + I(W^+) &= I(U_1; B_1B_2) + I(U_2; B_1B_2U_1) = I(U_1; B_1B_2) + I(U_2; B_1B_2|U_1) \\ &= I(U_1U_2; B_1B_2) = I(X_1X_2; B_1B_2) = I(X_1; B_1) + I(X_2; B_2) \\ &= 2I(W). \end{aligned}$$

This shows that the total symmetric Holevo information is conserved. Moreover,

$$I(W^+) = I(U_2; B_1B_2U_1) \geq I(U_2; B_2) = I(X_2; B_2) = I(W)$$

and

$$I(W^-) = 2I(W) - I(W^+) \leq I(W).$$

Let us now study the reliability of the cq-channel and how it is affected after one step of polarization. But first let us define the quantity $F_d(W)$ for every $d \in G$:

$$F_d(W) = \frac{1}{q} \sum_{x \in G} F(\rho_x, \rho_{x+d}).$$

Clearly, $0 \leq F_d(W) \leq 1$ and $F_0(W) = 1$. Note that

$$F(W) = \frac{1}{q-1} \sum_{\substack{d \in G, \\ d \neq 0}} F_d(W).$$

Define $F_{\max}(W) = \max_{\substack{d \in G, \\ d \neq 0}} F_d(W)$. Clearly, $F(W) \leq F_{\max}(W) \leq (q-1)F(W)$.

Proposition 8.2. *For every $d \in G$, we have:*

- $F_d(W^+) = F_d(W)^2$.
- $F_d(W) \leq F_d(W^-) \leq 2F_d(W) + \sum_{\substack{\Delta \in G, \\ \Delta \neq 0, \\ \Delta \neq -d}} F_\Delta(W) F_{d+\Delta}(W)$.

Proof. See Appendix 8.9.3. □

Corollary 8.1. *We have:*

- $F_{\max}(W^+) = F_{\max}(W)^2$.
- $F_{\max}(W) \leq F_{\max}(W^-) \leq qF_{\max}(W)$.
- $F(W^+) \leq \min\left\{F(W),\ (q-1)^2 F(W)^2\right\}$.
- $F(W) \leq F(W^-) \leq q(q-1)F(W)$

Proof. First equation:

$$F_{\max}(W^+) = \max_{\substack{d \in G, \\ d \neq 0}} F_d(W^+) = \max_{\substack{d \in G, \\ d \neq 0}} F_d(W)^2 = \left(\max_{\substack{d \in G, \\ d \neq 0}} F_d(W) \right)^2 = F_{\max}(W)^2.$$

Second equation:

$$\begin{aligned} F_{\max}(W) &= \max_{\substack{d \in G, \\ d \neq 0}} F_d(W) \leq \max_{\substack{d \in G, \\ d \neq 0}} F_d(W^-) = F_{\max}(W^-) \\ &\leq \max_{\substack{d \in G, \\ d \neq 0}} \left(2F_d(W) + \sum_{\substack{\Delta \in G, \\ \Delta \neq 0, \\ \Delta \neq -d}} F_\Delta(W) F_{d+\Delta}(W) \right) \\ &\leq 2F_{\max}(W) + (q-2)F_{\max}(W)^2 \leq qF_{\max}(W). \end{aligned}$$

First part of third equation:

$$F(W^+) = \frac{1}{q-1} \sum_{\substack{d \in G, \\ d \neq 0}} F_d(W^+) = \frac{1}{q-1} \sum_{\substack{d \in G, \\ d \neq 0}} F_d(W)^2 \leq \frac{1}{q-1} \sum_{\substack{d \in G, \\ d \neq 0}} F_d(W) = F(W).$$

Second part of third equation:

$$F(W^+) \leq F_{\max}(W^+) = F_{\max}(W)^2 \leq (q-1)^2 F(W)^2.$$

First inequality of the fourth equation:

$$F(W^-) = \frac{1}{q-1} \sum_{\substack{d \in G, \\ d \neq 0}} F_d(W^-) \geq \frac{1}{q-1} \sum_{\substack{d \in G, \\ d \neq 0}} F_d(W) = F(W).$$

Second inequality of the fourth equation:

$$F(W^-) \leq F_{\max}(W^-) \leq q F_{\max}(W) \leq q(q-1) F(W).$$

□

The following lemma is very useful to prove polarization results.

Lemma 8.2. *[6] Let $\{B_n\}_{n \geq 0}$ be a sequence of independent and uniformly distributed $\{-,+\}$-valued random variables. Suppose $\{I_n\}_{n \geq 0}$ and $\{T_n\}_{n \geq 0}$ are two processes adapted to the process $\{B_n\}_{n \geq 0}$ satisfying:*

(1) $0 \leq I_n \leq \log_2 q$.

(2) $\{I_n\}_{n \geq 0}$ *converges almost surely to a random variable* I_∞.

(3) $0 \leq T_n \leq 1$.

(4) $T_{n+1} = T_n^2$ *when* $B_{n+1} = +$.

(5) There exists a function $f(\epsilon)$ (depending only on q) satisfying $\lim_{\epsilon \to 0} f(\epsilon) = 0$ such that for all n, if $T_n < \epsilon$ then $I_n > \log_2 q - f(\epsilon)$.

(6) There exists a function $g(\epsilon)$ (depending only on q) satisfying $\lim_{\epsilon \to 0} g(\epsilon) = 0$ such that for all n, if $T_n > 1 - \epsilon$ then $I_n < g(\epsilon)$.

Then $T_\infty = \lim_{n \to \infty} T_n$ exists almost surely. Moreover, we have $I_\infty \in \{0, \log_2 q\}$ and $T_\infty \in \{0, 1\}$ with probability 1.

8.4 Polarization for $G = \mathbb{F}_q$

In this section, we focus on the particular case where $G = \mathbb{F}_q$ where q is prime. The main result of this section is the following theorem.

Theorem 8.1. *Let $W : x \in \mathbb{F}_q \longrightarrow \rho_x \in \mathcal{DM}(k)$ be a cq-channel with input in $\mathbb{F}_q$. For every $\delta > 0$, we have:*

$$\lim_{n \to \infty} \frac{1}{2^n} \left| \left\{ s \in \{-,+\}^n : \ \delta \leq I(W^s) \leq \log_2 q - \delta \right\} \right| = 0. \tag{8.2}$$

Moreover, for every $\beta < \frac{1}{2}$, we have:

$$\lim_{n\to\infty} \frac{1}{2^n} \left|\left\{ s \in \{-,+\}^n : \ I(W^s) \geq \log_2 q - \delta, \ F(W^s) < 2^{-2^{\beta n}} \right\}\right| = \frac{1}{\log_2 q} I(W). \tag{8.3}$$

Proof. Let $\{B_n\}_{n\geq 0}$ be a sequence of independent and uniformly distributed $\{-,+\}$-valued random variables. Define the cq-channel-valued process $\{W_n\}_{n\geq 0}$ as follows:

- $W_0 = W$.
- $W_n = W_{n-1}^{B_n}$ for every $n \geq 1$.

Let $I_n = I(W_n)$ and $T_n = F_{\max}(W_n)$. Let us check the conditions of Lemma 8.2. Conditions (1) and (3) follow from the properties of $I(W)$ and $F_{\max}(W)$. Condition (4) is satisfied because of Corollary 8.1.

We have $\mathbb{E}(I_{n+1}|W_n) = \frac{1}{2} I(W_n^-) + \frac{1}{2} I(W_n^+) = I(W_n)$. This shows that $\{I_n\}_{n\geq 0}$ is a bounded martingale and so it converges almost surely. This shows that condition (2) is satisfied.

Condition (5) follows from the following inequality:

$$I(W) \overset{(a)}{\geq} \log_2 \frac{q}{1 + (q-1)F(W)} \geq \log_2 \frac{q}{1 + (q-1)F_{\max}(W)},$$

where (a) is from Proposition 8.1. By choosing $f(\epsilon) = \log_2(1 + (q-1)\epsilon)$, we can see that condition (5) is satisfied.

In order to show condition (6), we need to prove that if $F_{\max}(W)$ is close to 1 then $I(W)$ is close to 0. Let d be such that $F_d(W) = F_{\max}(W)$. We have:

$$1 - F_d(W) = \frac{1}{q} \sum_{x\in G} \Big(1 - F(\rho_x, \rho_{x+d})\Big).$$

Therefore, for every $x \in G$ we have $1 - F(\rho_x, \rho_{x+d}) \leq q(1 - F_d(W))$ and so

$$F(\rho_x, \rho_{x+d}) \geq 1 - q(1 - F_d(W)).$$

Assume that $F_d(W)$ is high enough so that

$$1 - q(1 - F_d(W)) \geq \cos \frac{\pi}{2(q-1)}. \tag{8.4}$$

Now let $x, x' \in G$ be such that $x \neq x'$. Define $A(\rho_x, \rho_{x'}) = \arccos F(\rho_x, \rho_{x'})$ and let $l = \dfrac{x' - x}{d} \bmod q$. We have:

$$\begin{aligned}
F(\rho_x, \rho_{x'}) = \cos\big(A(\rho_x, \rho_{x+ld})\big) &\overset{(a)}{\geq} \cos\left(\sum_{i=0}^{l-1} A(\rho_{x+id}, \rho_{x+(i+1)d})\right) \\
&= \cos\left(\sum_{i=0}^{l-1} \arccos F(\rho_{x+id}, \rho_{x+(i+1)d})\right) \\
&\overset{(b)}{\geq} \cos\Big(l \cdot \arccos\Big(1 - q\big(1 - F_d(W)\big)\Big)\Big) \\
&\overset{(c)}{\geq} \cos\Big((q-1) \cdot \arccos\Big(1 - q\big(1 - F_d(W)\big)\Big)\Big),
\end{aligned}$$

where (a) follows from the fact that $A(\rho_x, \rho_{x'})$ is a metric distance [47]. (a), (b) and (c) are true because cos is a decreasing function on $\left[0, \frac{\pi}{2}\right]$ and we assumed Equation (8.4). We deduce that

$$F(W) = \frac{1}{q(q-1)} \sum_{\substack{x,x' \in G, \\ x \neq x'}} F(\rho_x, \rho_{x'}) \geq \cos\Big((q-1)\cdot \arccos\Big(1-q\big(1-F_d(W)\big)\Big)\Big). \quad (8.5)$$

By combining Equation (8.5) and inequality (iii) of Proposition 8.1, we get condition (6) of Lemma 8.2. Therefore, all the conditions of Lemma 8.2 are satisfied. We conclude that $\{I(W_n)\}_{n\geq 0}$ converges almost surely to a random variable $I_\infty \in \{0, \log_2 q\}$. This proves Equation (8.2).

From Corollary 8.1 we can deduce that $F(W^-) \leq q^2 F(W)$ and $F(W^+) \leq q^2 F(W)^2$. Therefore, we can apply the same techniques that were used to prove [33, Theorem 3.5] in order to get Equation (8.3). □

Theorem 8.1 can be used to construct polar codes for any cq-channel whose input alphabet size is prime. The polar code construction, encoder and decoder are similar to the one described in [43]. The main idea is to send information only through synthetic cq-channels for which the symmetric Holevo information is close to $\log_2 q$ and for which the average pairwise fidelity is less than 2^{-N^β}, where $N = 2^n$ is the blocklength of the polar code and $\beta < \frac{1}{2}$. We send frozen symbols that are known to the receiver through the remaining synthetic cq-channels. A quantum successive cancellation decoder that is similar to the one in [43] is applied. The probability of error can be shown to decay faster than 2^{-N^β} for any $\beta < \frac{1}{2}$. We postpone the accurate description and the study of the polar code till Section 8.7 where we construct polar codes in the more general case where $(G, +)$ is an arbitrary Abelian group.

8.5 Polarization for Arbitrary $(G, +)$

In this section, $(G, +)$ is an arbitrary Abelian group. For every cq-channel $W : x \in G \longrightarrow \rho_x \in \mathcal{DM}(k)$ and for every subgroup H of G, define the cq-channel $W[H] : D \in G/H \longrightarrow \rho_D \in \mathcal{DM}(k)$ as follows:

$$\rho_D = \frac{1}{|D|} \sum_{x \in D} \rho_x.$$

$W[H]$ can be simulated as follows: If a coset $D \in G/H$ is chosen as input, a random variable X is chosen uniformly from D and then sent through the cq-channel W.

It is easy to see that if $\rho^{XB} = \frac{1}{q} \sum_{x \in G} |x\rangle\langle x|^X \otimes \rho_x^B$, then $I(W[H]) = I(X \bmod H; B)_\rho$.

The main result of this section is the following theorem.

Theorem 8.2. *Let $W : x \in G \longrightarrow \rho_x \in \mathcal{DM}(k)$ be a cq-channel. For every $\delta > 0$, we have:*

$$\lim_{n\to\infty} \frac{1}{2^n}\Big|\Big\{s \in \{-,+\}^n : \exists H_s \text{ a subgroup of } G, \\ \big|I(W^s) - \log_2|G/H_s|\big| < \delta,\ \big|I(W^s[H_s]) - \log_2|G/H_s|\big| < \delta\Big\}\Big| = 1.$$

Theorem 8.2 can be interpreted as follows: As the number of polarization steps becomes large, the synthetic cq-channels polarize to homomorphism cq-channels projecting their input onto a quotient group of G. The inequality

$$\big|I(W^s[H_s]) - \log_2|G/H_s|\big| < \delta$$

means that from the output of W^s, one can determine with high probability the coset of H_s to which the input belongs. The inequality

$$\big|I(W^s) - \log_2|G/H_s|\big| < \delta$$

means that there is almost no other information about the input that can be determined from the output of W^s.

In order to prove Theorem 8.2 we need several definitions and lemmas. Let $\{B_n\}_{n\geq 0}$ be a sequence of independent and uniformly distributed $\{-,+\}$-valued random variables. Define the cq-channel-valued process $\{W_n\}_{n\geq 0}$ as follows:

- $W_0 = W$.
- $W_n = W_{n-1}^{B_n}$ for every $n \geq 1$.

Lemma 8.3. *For every subgroup H of G, the process $\{I(W_n[H])\}_{n\geq 0}$ is a submartingale.*

Proof. It is sufficient to show that $I(W^-[H]) + I(W^+[H]) \geq 2I(W[H])$. Let U_1, U_2, X_1, X_2, B_1 and B_2 be as in Remark 8.2. We have:

$$\begin{aligned} I(W^-[H]) + I(W^+[H]) &= I(U_1 \bmod H; B_1B_2) + I(U_2 \bmod H; B_1B_2U_1) \\ &\geq I(U_1 \bmod H; B_1B_2) + I(U_2 \bmod H; B_1B_2, U_1 \bmod H) \\ &= I(U_1 \bmod H, U_2 \bmod H; B_1B_2) \\ &= I(X_1 \bmod H, X_2 \bmod H; B_1B_2) \\ &= I(X_1 \bmod H, B_1) + I(X_2 \bmod H; B_2) = 2I(W[H]). \end{aligned}$$

□

Let $M \subset H$ be two subgroups of G. For every coset D of H, let $D/M = \{C \in G/M : C \subset D\}$ be the set of cosets of M which are subsets of D. Define the cq-channel $W[M|D] : C \in D/M \longrightarrow \rho_C \in \mathcal{DM}(k)$ as follows:

$$\rho_C = \frac{1}{|C|}\sum_{x\in C} \rho_x.$$

$W[M|D]$ can be simulated as follows: If a coset $C \in D/M$ is chosen as input, a random variable X is chosen uniformly from C and then sent through the cq-channel W.

Define the following:

- $I_{M|H}(W) = I(W[M]) - I(W[H])$.
- $F_{\max}^{M|H}(W) = \max_{\substack{d\in H,\\ d\notin M}} F_d(W)$.

The following lemma relates $I_{M|H}(W)$ to $\{I(W[M|D]) :\ D \in G/H\}$.

Lemma 8.4. $I_{M|H}(W) = \dfrac{1}{|G/H|}\displaystyle\sum_{D\in G/H} I(W[M|D])$.

Proof. Let $\rho^{XB} = \dfrac{1}{q}\displaystyle\sum_{x\in G} |x\rangle\langle x|^X \otimes \rho_x^B$. We have $I(W[M]) = I(X \bmod M; B)_\rho$ and $I(W[M]) = I(X \bmod H; B)_\rho$. Therefore,

$$\begin{aligned} I_{M|H}(W) &= I(W[M]) - I(W[H]) = I(X \bmod M; B)_\rho - I(X \bmod M; B)_\rho \\ &= I(X \bmod M, X \bmod H; B)_\rho - I(X \bmod H; B)_\rho = I(X \bmod M; B|X \bmod H)_\rho \\ &= \sum_{D\in G/H} \frac{1}{|G/H|} I(X \bmod M; B|X \bmod H = D)_\rho \overset{(a)}{=} \sum_{D\in G/H} \frac{1}{|G/H|} I(W[M|D]), \end{aligned}$$

where (a) follows from the fact that conditioning on $X \bmod H = D$, the state of the input-output system becomes $\dfrac{1}{|D|}\displaystyle\sum_{x\in D} |x\rangle\langle x|^X \otimes \rho_x^B$ and so the mutual information between $X \bmod M$ and B becomes exactly $I(W[M|D])$. □

The following lemma relates $F(W[M|D])$ to $F_{\max}^{M|H}(W)$.

Lemma 8.5. *For every $D \in G/H$, we have:*

(1) $F(W[M|D]) \le \dfrac{q\cdot|M|}{|H|} F_{\max}^{M|H}(W)$.

(2) There exists $\epsilon_q > 0$ depending only on q such that if M is maximal in H (i.e., $|H/M|$ is prime) and if $F_{\max}^{M|H}(W) \ge 1 - \epsilon_q$, then

$$F(W[M|D]) \ge \cos\left(\frac{|H|-|M|}{|M|}\arccos\left(1 - \sqrt{1 - \left(1 - q\big(1 - F_{\max}^{M|H}(W)\big)\right)^2}\right)\right).$$

Proof. See Appendix 8.9.4. □

Lemma 8.6. *For every two subgroups $M \subset H$ of G where M is maximal in H (i.e., $|H/M|$ is prime), the process $\{I_{M|H}(W_n)\}_{n\ge 0}$ converges almost surely to a random variable $I_{M|H}^{(\infty)} \in \{0, \log_2|H/M|\}$ and the process $\{F_{\max}^{M|H}(W_n)\}_{n\ge 0}$ converges almost surely to a random variable $F_{M|H}^{(\infty)} \in \{0,1\}$.*

Proof. Let $I_n = I_{M|H}(W_n)$ and $T_n = F_{\max}^{M|H}(W_n)$. We will show that I_n and T_n satisfy the conditions of Lemma 8.2, where q is replaced with $q' = |H/M|$. Conditions (1) and (3) are obviously satisfied. Condition (4) is also satisfied because of Proposition 8.2.

Since $I_{M|H}(W_n) = I(W_n[M]) - I(W_n[H])$ and since the processes $\{I(W_n[M])\}_{n\geq 0}$ and $\{I(W_n[H])\}_{n\geq 0}$ are sub-martingales by Lemma 8.3, we conclude that $\{I_n\}_{n\geq 0}$ converges almost surely. Therefore, condition (2) is satisfied.

To see that condition (5) is satisfied, assume that $F_{\max}^{M|H}(W)$ is close to zero, then the first inequality of Lemma 8.5 implies that $F(W[M|D])$ is close to zero for every $D \in G/H$. The first inequality of Proposition 8.1 then shows that $I(W[M|D])$ is close to $\log_2 q'$, for every $D \in G/H$. Lemma 8.4 now implies that $I_{M|H}(W)$ is close to $\log_2 q'$.

To see that condition (6) is satisfied, assume that $F_{\max}^{M|H}(W)$ is close to 1, then the second inequality of Lemma 8.5 implies that $F(W[M|D])$ is close to 1 for every $D \in G/H$. The third inequality of Proposition 8.1 then shows that $I(W[M|D])$ is close to zero, for every $D \in G/H$. Lemma 8.4 now implies that $I_{M|H}(W)$ is close to zero.

We conclude that $\{I_{M|H}(W_n)\}_{n\geq 0}$ converges almost surely to a random variable taking values in $\{0, \log_2 q'\} = \{0, \log_2 |H/M|\}$ and $\{F_{\max}^{M|H}(W_n)\}_{n\geq 0}$ converges almost surely to a random variable taking values in $\{0,1\}$. □

Lemma 8.7. *Let $d_1, \ldots, d_r \in G$. If $F_{d_i}(W) \geq 1 - \frac{1}{q}\left(1 - \cos\frac{\pi}{2r}\right)$ for all $1 \leq i \leq r$, then*

$$F_{d_1+\cdots+d_r}(W) \geq \cos\left(\sum_{i=1}^{r} \arccos\Big(1 - q\big(1 - F_{d_i}(W)\big)\Big)\right).$$

Proof. We may assume without loss of generality that $d_1 \neq 0, \ldots, d_r \neq 0$ and $d := d_1 + \cdots + d_r \neq 0$. Define $d'_1 = 0$, and for every $2 \leq i \leq r$, let $d'_i = \sum_{j=1}^{i-1} d_j$.

For every $1 \leq i \leq r$, we have $1 - F_{d_i}(W) = \frac{1}{q}\sum_{x\in G}\big(1 - F(\rho_x, \rho_{x+d_i})\big)$. Therefore, for every $x \in G$, we have $1 - F(\rho_x, \rho_{x+d_i}) \leq q\big(1 - F_{d_i}(W)\big)$ and so $F(\rho_x, \rho_{x+d_i}) \geq 1 - q\big(1 - F_{d_i}(W)\big)$. Therefore,

$$\begin{aligned}
&F(\rho_x, \rho_{x+d}) \\
&= F(\rho_{x+d'_1}, \rho_{x+d'_r+d_r}) = \cos A(\rho_{x+d'_1}, \rho_{x+d'_r+d_r}) \overset{(a)}{\geq} \cos\left(\sum_{i=1}^{r} A(\rho_{x+d'_i}, \rho_{x+d'_i+d_i})\right) \\
&= \cos\left(\sum_{i=1}^{r} \arccos F(\rho_{x+d'_i}, \rho_{x+d'_i+d_i})\right) \overset{(b)}{\geq} \cos\left(\sum_{i=1}^{r} \arccos\Big(1 - q\big(1 - F_{d_i}(W)\big)\Big)\right),
\end{aligned}$$

where (a) follows from the fact that $A(\rho', \rho'') = \arccos F(\rho', \rho'')$ is a metric distance [47]. (a) and (b) are true because cos is a decreasing function on $\left[0, \frac{\pi}{2}\right]$ and we assumed that $F_{d_i}(W) \geq 1 - \frac{1}{q}\left(1 - \cos\frac{\pi}{2r}\right)$ for every $1 \leq i \leq r$. We conclude that

$$F_d(W) = \frac{1}{q}\sum_{x\in G} F(\rho_x, \rho_{x+d}) \geq \cos\left(\sum_{i=1}^{r} \arccos\Big(1 - q\big(1 - F_{d_i}(W)\big)\Big)\right).$$

□

Lemma 8.8. *Let $d \in G$ be such that $d \neq 0$ and let $H = \langle d \rangle$ be the subgroup generated by d. We have:*

- *If $F_d(W) \leq F_{\max}^{M|H}(W)$ for every maximal subgroup M of H.*
- *If $F_{\max}^{M|H}(W) \geq 1 - \frac{1}{q}\left(1 - \cos\frac{\pi}{2q}\right)$ for every maximal subgroup M of H, then*

$$F_d(W) \geq \cos\left(q \cdot \arccos\left(1 - q\left(1 - \min_{\substack{M \text{ is a maximal} \\ \text{subgroup of } H}} F_{\max}^{M|H}(W)\right)\right)\right).$$

Proof. Let M be a maximal subgroup of H. Since $H = \langle d \rangle$, then we must have $d \in H$ and $d \notin M$. Therefore,

$$F_d(W) \leq \max_{\substack{d' \in H, \\ d' \notin M}} F_{d'}(W) = F_{\max}^{M|H}(W).$$

Now let $M_1, \ldots, M_r$ be the maximal subgroups of $H = \langle d \rangle$. For every $1 \leq i \leq r$, let $d_i \in H$ be such that $d_i \notin M_i$ and $F_{d_i}(W) = F_{\max}^{M_i|H}(W)$. It was shown in [6] that $d \in \langle d_1, \ldots, d_r \rangle$, which means that there are $l_1, \ldots, l_r \in \mathbb{N}$ such that $d = \sum_{i=1}^{r} l_i d_i$. Moreover, $l_1, \ldots, l_r \in \mathbb{N}$ can be chosen so that $l_1 + \cdots + l_r \leq q$.

Since $F_{d_i}(W) \geq 1 - \frac{1}{q}\left(1 - \cos\frac{\pi}{2q}\right) \geq 1 - \frac{1}{q}\left(1 - \cos\frac{\pi}{2(l_1 + \cdots + l_r)}\right)$ for all $1 \leq i \leq r$, Lemma 8.7 implies that

$$\begin{aligned}
F_d(W) &= F_{l_1 d_1 + \cdots + l_r d_r}(W) \\
&\geq \cos\left(\sum_{i=1}^{r} l_i \arccos\left(1 - q\big(1 - F_{d_i}(W)\big)\right)\right) \\
&\overset{(a)}{\geq} \cos\left((l_1 + \cdots + l_r)\arccos\left(1 - q\big(1 - \min_{1 \leq i \leq r} F_{d_i}(W)\big)\right)\right) \\
&\overset{(b)}{\geq} \cos\left(q \cdot \arccos\left(1 - q\big(1 - \min_{1 \leq i \leq r} F_{d_i}(W)\big)\right)\right),
\end{aligned}$$

where (a) and (b) are true because cos is decreasing on $\left[0, \frac{\pi}{2}\right]$ and because we assumed that $F_{d_i}(W) \geq 1 - \frac{1}{q}\left(1 - \cos\frac{\pi}{2q}\right)$ for all $1 \leq i \leq r$. □

Proposition 8.3. *For every $d \in G$, the process $\{F_d(W_n)\}_{n \geq 0}$ converges almost surely to a random variable $F_d^{(\infty)} \in \{0, 1\}$. Moreover, the random set $\{d \in G : F_d^{(\infty)} = 1\}$ is almost surely a subgroup of G.*

Proof. Let $d \in G$ be such that $d \neq 0$. Let $H = \langle d \rangle$ be the subgroup generated by d. Lemma 8.6 shows that for every maximal subgroup M of H, the process $\left\{F_{\max}^{M|H}(W_n)\right\}_{n \geq 0}$ converges almost surely to a random variable taking values in $\{0, 1\}$.

Take a sample of the process $\{W_n\}_{n \geq 0}$ for which $\left\{F_{\max}^{M|H}(W_n)\right\}_{n \geq 0}$ converges to either 0 or 1 for every maximal subgroup M of H. We have:

- If there exists a maximal subgroup M of H for which $\left\{F_{\max}^{M|H}(W_n)\right\}_{n\geq 0}$ converges to 0, then the first point of Lemma 8.8 implies that $\{F_d(W_n)\}_{n\geq 0}$ converges to 0 as well.
- If $\left\{F_{\max}^{M|H}(W_n)\right\}_{n\geq 0}$ converges to 1 for all maximal subgroups M of H, then the second point of Lemma 8.8 implies that $\{F_d(W_n)\}_{n\geq 0}$ converges to 1 as well.

We conclude that for every $d \in G$, the process $\{F_d(W_n)\}_{n\geq 0}$ converges almost surely to a random variable $F_d^{(\infty)} \in \{0,1\}$. (Note that for $d=0$, we have $F_0(W_n)=1$ for all n.)

Now take a sample of the process $\{W_n\}_{n\geq 0}$ for which $\{F_d(W_n)\}_{n\geq 0}$ converges to either 0 or 1 for every $d \in G$. If $d_1, d_2 \in G$ are such that $\{F_{d_1}(W_n)\}_{n\geq 0}$ and $\{F_{d_2}(W_n)\}_{n\geq 0}$ converge to 1, then Lemma 8.7 implies that $\{F_{d_1+d_2}(W_n)\}_{n\geq 0}$ converges to 1 as well. We conclude that the set $\left\{d \in G : \{F_d(W_n)\}_{n\geq 0} \text{ converges to } 1\right\}$ is a subgroup of G. □

Corollary 8.2. *For every $\epsilon > 0$, we have*

$$\lim_{n\to\infty} \frac{1}{2^n}\Big|\Big\{s \in \{-,+\}^n : \exists H_s \text{ a subgroup of } G, \\ F_d(W) > 1-\epsilon \text{ for every } d \in H_s, \text{ and } F_d(W) < \epsilon \text{ for every } d \notin H_s\Big\}\Big| = 1.$$

Lemma 8.9. *For every $\delta > 0$, there exists $\epsilon > 0$ depending only on δ and q such that for every cq-channel W, if there exists a subgroup H of G satisfying $F_d(W) > 1-\epsilon$ for all $d \in H$ and $F_d(W) < \epsilon$ for all $d \notin H$, then $\left|I(W) - \log_2|G/H|\right| < \delta$ and $\left|I(W[H]) - \log_2|G/H|\right| < \delta$.*

Proof. If $H = G$, then $I(W[G]) = 0 = \log_2|G/G|$ and so $\left|I(W[G]) - \log_2|G/G|\right| = 0 < \delta$. On the other hand, since $H = G$, we have $F_d(W) > 1-\epsilon$ for every $d \in G$. Therefore, $F(W) = \dfrac{1}{q-1}\displaystyle\sum_{\substack{d\in G,\\ d\neq 0}} F_d(W) > 1-\epsilon$. The third inequality of Proposition 8.1 now implies $I(W) < \delta_q^{(1)}$ for some function $\epsilon \to \delta_q^{(1)}(\epsilon)$ (depending only on ϵ and q) which satisfies $\lim\limits_{\epsilon\to 0}\delta_q^{(1)}(\epsilon) = 0$.

Now assume that $H \neq G$. We have

$$F(W[H]) = F(W[H|G]) \overset{(a)}{\leq} \frac{q\cdot|H|}{q} F_{\max}^{H|G}(W) \leq q \max_{\substack{d\in G,\\ d\notin H}} F_d(W) \leq q\epsilon,$$

where (a) follows from the first inequality of Lemma 8.5. The first inequality of Proposition 8.1 implies that $I(W[H]) > \log_2|G/H| - \delta_q^{(2)}(\epsilon)$ for some function $\epsilon \to \delta_q^{(2)}(\epsilon)$ (depending only on ϵ and q) which satisfies $\lim\limits_{\epsilon\to 0}\delta_q^{(2)}(\epsilon) = 0$.

On the other hand, we have $F_{\max}^{\{0\}|H}(W) = \max\limits_{\substack{d\in H,\\ d\neq 0}} F_d(W) \geq 1-\epsilon$. Assume that $\epsilon < \epsilon_q$, where ϵ_q is given by Lemma 8.5. For every $D \in G/H$, we have

$$F(W[\{0\}|D]) \geq \cos\left((|H|-1)\cdot\arccos\left(1-\sqrt{1-\left(1-q\big(1-F_{\max}^{\{0\}|H}(W)\big)\right)^2}\right)\right).$$

This means that $F(W[\{0\}|D])$ is close to 1 as well. The third inequality of Proposition 8.1 now implies that $I(W[\{0\}|D]) < \delta_q^{(3)}(\epsilon)$ for some function $\epsilon \to \delta_q^{(3)}(\epsilon)$ (depending only on ϵ and q) which satisfies $\lim_{\epsilon\to 0} \delta_q^{(3)}(\epsilon) = 0$. We conclude that

$$\begin{aligned} I(W) - I(W[H]) &= I(W[\{0\}]) - I(W[H]) = I_{\{0\}|H}(W) \\ &\overset{(a)}{=} \frac{1}{|G/H|} \sum_{D\in G/H} I(W[\{0\}|D]) < \delta_q^(3), \end{aligned}$$

where (a) follows from Lemma 8.4. We conclude that

$$\left|I(W)-\log_2|G/H|\right| \leq |I(W)-I(W[H])|+\left|I(W[H])-\log_2|G/H|\right| < \delta_q^{(2)}(\delta)+\delta_q^{(3)}(\delta).$$

If we define $\delta_q(\epsilon) = \max\left\{\delta_q^{(1)}(\epsilon), \delta_q^{(2)}(\epsilon) + \delta_q^{(3)}(\epsilon)\right\}$, we get $\left|I(W)-\log_2|G/H|\right| < \delta_q(\epsilon)$ and $\left|I(W[H]) - \log_2|G/H|\right| < \delta_q(\epsilon)$ in all cases. Moreover, $\lim_{\epsilon\to 0} \delta_q(\epsilon) = 0$.

This concludes the proof of the lemma. □

The proof of Theorem 8.2 now follows immediately from Corollary 8.2 and Lemma 8.9.

8.6 Rate of Polarization

In order to derive the rate of polarization (i.e., how fast the synthetic cq-channels polarize), we need the following two lemmas.

Lemma 8.10. *For every subgroup H of G, we have:*

- $F(W^-[H]) \leq |H|q(q-|H|)F(W[H])$.
- $F(W^+[H]) \leq |H|(q-|H|)^2F(W[H])^2$.

Proof. See Appendix 8.9.5 □

Lemma 8.11. *For any $0 < \delta < 1$ and any $0 < \beta < \frac{1}{2}$, we have*

$$\lim_{n\to\infty} \frac{1}{2^n}\left|\left\{s \in \{-,+\}^n : \ I(W^s[H]) > \log_2|G/H| - \delta, \ F(W^s[H]) \geq 2^{-2^{\beta n}}\right\}\right| = 0.$$

Proof. The lemma is trivial if $H = G$, so let us assume that $H \neq G$. Let $H_1, \ldots, H_r$ be a sequence of subgroups of G satisfying:

- $H = H_1 \subset \ldots \subset H_r = G$.
- H_i is maximal in H_{i+1} for every $1 \leq i < r$.

Let $\{W_n\}_{n\geq 0}$ be the process defined in the previous section. Lemma 8.6 implies that $\{I_{H_i|H_{i+1}}(W_n)\}_{n\geq 0}$ converges almost surely to a random variable $I^{(\infty)}_{H_i|H_{i+1}} \in \{0, \log_2|H_{i+1}/H_i|\}$. On the other hand, we have

$$\begin{aligned} I(W_n[H]) = I(W_n[H]) - I(W_n[G]) &= \sum_{i=1}^{r-1} \big(I(W_n[H_i]) - I(W_n[H_{i+1}])\big) \\ &= \sum_{i=1}^{r-1} I_{H_i|H_{i+1}}(W_n). \end{aligned}$$

This shows that the process $\{I(W_n[H])\}_{n\geq 0}$ converges almost surely to a random variable $I^{(\infty)}_H$ satisfying

$$I^{(\infty)}_H \in \{\log_2 m : \ m \text{ divides } |G/H|\}.$$

Due to the relations between the quantities $I(W)$ and $F(W)$ in Proposition 8.1, we can see that $\{F(W_n[H])\}_{n\geq 0}$ converges to 0 whenever $\{I(W_n[H])\}_{n\geq 0}$ converges to $\log_2|G/H|$, and there is a number $f_0 > 0$ such that $\liminf_{n\to\infty} F(W_n[H]) > f_0$ whenever $\{I(W_n[H])\}_{n\geq 0}$ converges to a number in $\{\log_2 m : \ m \text{ divides } |G/H|\}$ other than $\log_2|G/H|$. Therefore, we can say that almost surely, we have:

$$\lim_{n\to\infty} F(W_n[H]) = 0 \ \text{ or } \ \liminf_{n\to\infty} F(W_n[H]) > f_0.$$

Now from Lemma 8.10, we have $F(W_n^-[H]) \leq q^3 F(W_n[H])$ and $F(W_n^+[H]) \leq q^3 F(W_n[H])^2$. By applying exactly the same techniques that were used to prove [33, Theorem 3.5] we get:

$$\lim_{n\to\infty} \mathbb{P}\Big(\Big\{I(W_n[H]) > \log_2|G/H| - \delta, F(W_n[H]) \geq 2^{-2^{n\beta}}\Big\}\Big) = 0.$$

By examining the explicit expression of this probability we get the lemma. □

Theorem 8.3. *The polarization of W_n is almost surely fast:*

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{2^n}\Big|\Big\{s\in\{-,+\}^n : \exists H_s \textit{ subgroup of } G, \\ &\big|I(W^s) - \log_2|G/H_s|\big| < \delta, \big|I(W^s[H_s]) - \log_2|G/H_s|\big| < \delta, F(W^s[H_s]) < 2^{-2^{\beta n}}\Big\}\Big| = 1, \end{aligned}$$

for any $0<\delta<1$ and any $0<\beta<\frac{1}{2}$.

Proof. For every subgroup H of G, define:

$$E_H = \Big\{s\in\{-,+\}^n : \ I(W^s[H]) > \log_2|G/H| - \delta, F(W^s[H]) \geq 2^{-2^{\beta n}}\Big\},$$

$$\begin{aligned} E_1 = \Big\{s\in\{-,+\}^n : &\exists H_s \text{ subgroup of } G, \\ &\big|I(W^s) - \log_2|G/H_s|\big| < \delta, \big|I(W^s[H_s]) - \log_2|G/H_s|\big| < \delta\Big\}, \end{aligned}$$

and

$$E_2 = \Big\{ s \in \{-,+\}^n : \exists H_s \text{ subgroup of } G, \\ \big|I(W^s) - \log_2 |G/H_s|\big| < \delta, \big|I(W^s[H_s]) - \log_2 |G/H_s|\big| < \delta, F(W^s[H_s]) < 2^{-2^{\beta n}} \Big\}.$$

If $s \in E_1 / \Big(\bigcup_{H \text{ subgroup of } G} E_H \Big)$ then $s \in E_2$. Therefore,

$$E_1 / \Big(\bigcup_{H \text{ subgroup of } G} E_H \Big) \subset E_2,$$

and $|E_2| \geq |E_1| - \sum_{H \text{ subgroup of } G} |E_H|$. By Theorem 8.2 and Lemma 8.11 we have:

$$1 \geq \lim_{n\to\infty} \frac{1}{2^n}|E_2| \geq \lim_{n\to\infty} \frac{1}{2^n}\Big(|E_1| - \sum_{H \text{ subgroup of } G} |E_H|\Big) = 1 - 0 = 1.$$

□

8.7 Polar Code Construction

Let $W : x \in G \longrightarrow \rho_x \in \mathcal{DM}(k)$ be an arbitrary cq-channel.

Choose $0 < \delta < 1$ and $0 < \beta < \beta' < \frac{1}{2}$, and let n be an integer such that

$$2\sqrt{2^n}\sqrt{(q-1)2^n 2^{-2^{\beta' n}}} \leq 2^{-2^{\beta n}} \text{ and } \frac{1}{2^n}|E_n| > 1 - \frac{\delta}{2\log_2 q},$$

where

$$E_n = \Big\{ s \in \{-,+\}^n : \exists H_s \text{ subgroup of } G, \\ \big|I(W^s) - \log_2 |G/H_s|\big| < \frac{\delta}{2}, \big|I(W^s[H_s]) - \log_2 |G/H_s|\big| < \frac{\delta}{2}, F(W^s[H_s]) < 2^{-2^{\beta' n}} \Big\}.$$

Such an integer exists due to Theorem 8.3. For every $s \in \{-,+\}^n$ choose a subgroup H_s of G as follows:

- If $s \notin E_n$, define $H_s = G$. We clearly have $F(W^s[H_s]) = 0 < 2^{-2^{\beta' n}}$.
- If $s \in E_n$, choose a subgroup H_s of G such that $F(W^s[H_s]) < 2^{-2^{\beta' n}}$, $\big|I(W^s) - \log_2 |G/H_s|\big| < \frac{\delta}{2}$ and $\big|I(W^s[H_s]) - \log_2 |G/H_s|\big| < \frac{\delta}{2}$.

Now for every $s \in \{-,+\}^n$, let $f_s : G/H_s \longrightarrow G$ be a frozen mapping (in the sense that the receiver knows f_s) such that $f_s(a) \bmod H_s = a$ for all $a \in G/H_s$. We call such mapping *a section mapping* of G/H_s. Let $\tilde{U}^s$ be a random coset chosen uniformly in G/H_s and let $U^s = f_s(\tilde{U}^s)$. Note that if the receiver can determine $U^s \bmod H_s = \tilde{U}^s$ accurately, then he can also determine U^s since he knows f_s.

If $H_s \neq \{0\}$, we have some freedom on the choice of the section mapping f_s. We will analyze the performance of polar codes averaged over all possible section

mappings. I.e., we assume that f_s is chosen uniformly from the set of all possible section mappings of G/H_s. We can easily see that the induced distributions of $\{U^s : s \in \{-,+\}^n\}$ are independent and uniform in G. Note that for every $s \in \{-,+\}^n$, the receiver has to determine $\tilde{U}^s = U^s \bmod H_s$ in order to successfully determine U^s.

8.7.1 Encoder

We associate the set $S_n := \{-,+\}^n$ with the strict total order $<$ defined as $(s_1,...,s_n) < (s'_1,...,s'_n)$ if and only if $s_i = -, s'_i = +$ for some $i \in \{1,...,n\}$ and $s_h = s'_h$ for all $i < h \leq n$.

For every $u = (u^s)_{s\in S_n} \in G^{S_n}$, every $0 \leq n' \leq n$ and every $(s', s'') \in S_{n'} \times S_{n-n'}$, define $\mathcal{E}^{s''}_{s'}(u) \in G$ recursively on $0 \leq n' \leq n$ as follows:

- $\mathcal{E}^{s}_{\emptyset}(u) = u^s$ if $n' = 0$ and $s \in S_n$.
- $\mathcal{E}^{s''}_{(s',-)}(u) = \mathcal{E}^{(s'',-)}_{s'}(u) + \mathcal{E}^{(s'',+)}_{s'}(u)$ if $n' > 0$, $s' \in S_{n'-1}$ and $s'' \in S_{n-n'}$.
- $\mathcal{E}^{s''}_{(s',+)}(u) = \mathcal{E}^{(s'',+)}_{s'}(u)$ if $n' > 0$, $s' \in S_{n'-1}$ and $s'' \in S_{n-n'}$.

For every $s \in S_n$, we write $\mathcal{E}^{s}_{\emptyset}(u)$ as $\mathcal{E}^{s}(u)$ and $\mathcal{E}^{\emptyset}_{s}(u)$ as $\mathcal{E}_s(u)$.

Let $\{W_s\}_{s\in S_n}$ be a set of 2^n independent copies of the cq-channel W. W_s should not be confused with W^s: W_s is a copy of the cq-channel W and W^s is a synthetic cq-channel obtained from W as before.

Let $(U^s)_{s\in S_n} = (f_s(\tilde{U}^s))_{s\in S_n}$ be the sequence of 2^n independent random variables that were defined before. For every $0 \leq n' \leq n$, $s' \in S_{n'}$ and $s'' \in S_{n-n'}$, define $U^{s''}_{s'} = \mathcal{E}^{s''}_{s'}\big((U^s)_{s\in S_n}\big)$. We have:

- $U^{s}_{\emptyset} = U^s$ if $n' = 0$ and $s \in \{-,+\}^n$.
- $U^{s''}_{(s',-)} = U^{(s'',+)}_{s'} + U^{(s'',-)}_{s'}$ if $n' > 0$, $s' \in \{-,+\}^{n'-1}$ and $s'' \in \{-,+\}^{n-n'}$.
- $U^{s''}_{(s',+)} = U^{(s'',+)}_{s'}$ if $n' > 0$, $s' \in \{-,+\}^{n'-1}$ and $s'' \in \{-,+\}^{n-n'}$.

For every $s \in S_n$, let $U_s = U^{\emptyset}_s$. It is easy to see that $(U_s)_{s\in S_n}$ are independent and uniformly distributed in G.

For every $s \in S_n$, we send U_s through the cq-channel W_s. Let B_s be the system describing the output of the cq-channel W_s, and let $B = \{B_s\}_{s\in S_n}$. We can prove by backward induction on n' that for every $s'' \in S_{n-n'}$, the cq-channel $U^{s''}_{s'} \to \big(\{B_s\}_{s \text{ has } s' \text{ as a prefix}}, \{U^r_{s'}\}_{r<s''}\big)$ is equivalent to the cq-channel $W^{s''}$ for every $0 \leq n' \leq n$, $s' \in S_{n'}$ and $s'' \in S_{n-n'}$. In particular, the cq-channel $U^s \to \big(B, \{U^r\}_{r<s}\big)$ is equivalent to the cq-channel W^s for every $s \in S_n$.

Note that the encoding algorithm described above has a complexity of $O(N \log N)$, where $N = 2^n$ is the blocklength of the polar code.

8.7.2 Quantum Successive Cancellation decoder

Before describing the decoder, let us fix a few useful notations.

For every $s \in S_n$, define $\mathcal{L}_s = \{r \in S_n : r < s\}$ and $\mathcal{U}_s = \{r \in S_n : r > s\}$. For every $u = (u^s)_{s\in S_n} \in G^{S_n}$, define the following:

- For every $S \subset S_n$, let $u^S := (u^s)_{s\in S}$.
- For every $s \in S_n$, let $u_s := \mathcal{E}_s(u)$.
- Define $\rho_u^B := \bigotimes_{s\in S_n} \rho_{u_s}^{B_s}$. This means that if $U^s = u^s$ for every $s \in S_n$, then the receiver sees the state ρ_u^B at the output.

It is easy to see that for every $s \in S_n$, we have $W^s : u^s \in G \longrightarrow \rho_{s,u^s}^{B,U^{\mathcal{L}_s}} \in \mathcal{DM}\left(k^{2^n} \cdot q^{|\mathcal{L}_s|}\right)$, where

$$\rho_{s,u^s}^{B,U^{\mathcal{L}_s}} = \frac{1}{q^{|\mathcal{L}_s|}} \sum_{u^{\mathcal{L}_s}\in G^{\mathcal{L}_s}} \overline{\rho}_{u^s,u^{\mathcal{L}_s}}^B \otimes \left|u^{\mathcal{L}_s}\right\rangle \left\langle u^{\mathcal{L}_s}\right|^{U^{\mathcal{L}_s}},$$

and

$$\overline{\rho}_{u^s,u^{\mathcal{L}_s}}^B = \frac{1}{q^{|\mathcal{U}_s|}} \sum_{u^{\mathcal{U}_s}\in G^{\mathcal{U}_s}} \rho_u^B.$$

Moreover, we have $W^s[H_s] : \tilde{u}^s \in G/H_s \longrightarrow \rho_{s,\tilde{u}^s}^{B,U^{\mathcal{L}_s}} \in \mathcal{DM}\left(k^{2^n} \cdot q^{|\mathcal{L}_s|}\right)$, where

$$\rho_{s,\tilde{u}^s}^{B,U^{\mathcal{L}_s}} = \frac{1}{|H_s|} \sum_{u_s\in\tilde{u}_s} \rho_{s,u^s}^{B,U^{\mathcal{L}_s}} = \frac{1}{q^{|\mathcal{L}_s|}} \sum_{u^{\mathcal{L}_s}\in G^{\mathcal{L}_s}} \overline{\rho}_{\tilde{u}^s,u^{\mathcal{L}_s}}^B \otimes \left|u^{\mathcal{L}_s}\right\rangle \left\langle u^{\mathcal{L}_s}\right|^{U^{\mathcal{L}_s}},$$

and

$$\overline{\rho}_{\tilde{u}^s,u^{\mathcal{L}_s}}^B = \frac{1}{|H_s| \cdot q^{|\mathcal{U}_s|}} \sum_{u^s\in\tilde{u}^s} \sum_{u^{\mathcal{U}_s}\in G^{\mathcal{U}_s}} \rho_u^B.$$

Lemma 8.12. *For every $u^{\mathcal{L}_s} \in G^{\mathcal{L}_s}$, there exists a POVM $\left\{\Pi_{(s),u^{\mathcal{L}_s},\tilde{u}^s}^B : \tilde{u}^s \in G/H_s\right\}$ such that the POVM $\left\{\Pi_{(s),\tilde{u}^s}^{B,U^{\mathcal{L}_s}} : \tilde{u}^s \in G/H_s\right\}$ defined as*

$$\Pi_{(s),\tilde{u}^s}^{B,U^{\mathcal{L}_s}} = \sum_{u^{\mathcal{L}_s}\in G^{\mathcal{L}_s}} \Pi_{(s),u^{\mathcal{L}_s},\tilde{u}^s}^B \otimes \left|u^{\mathcal{L}_s}\right\rangle \left\langle u^{\mathcal{L}_s}\right|^{U^{\mathcal{L}_s}},$$

satisfies

$$1 - \frac{1}{|G/H_s|} \sum_{\tilde{u}^s\in G/H_s} \mathrm{Tr}\left(\Pi_{(s),\tilde{u}^s}^{B,U^{\mathcal{L}_s}} \rho_{s,\tilde{u}^s}^{B,U^{\mathcal{L}_s}}\right) < (|G/H_s| - 1)F(W[H_s]).$$

Proof. See Appendix 8.9.6. □

For every $s \in S_n$ and every $u^{\mathcal{L}_s} \in G^{\mathcal{L}_s}$, define the POVM $\left\{\Pi_{(s),u^{\mathcal{L}_s},u^s}^B : u^s \in G\right\}$ as:

$$\Pi_{(s),u^{\mathcal{L}_s},u^s}^B = \begin{cases} \Pi_{(s),u^{\mathcal{L}_s},u^s \bmod H_s}^B & \text{if } u^s = f_s(u^s \bmod H_s), \\ 0 & \text{otherwise.} \end{cases}$$

Now we are ready to describe the quantum successive cancellation decoder. We will decode $\{U^s\}_{s\in S_n}$ successively by respecting the order $<$ on S_n. At the stage $s \in S_n$, we would have decoded $U^{\mathcal{L}_s} = (U^r)_{r<s}$ and obtained an estimate $\hat{u}^{\mathcal{L}_s} = (\hat{u}^r)_{r<s}$ of it, so we apply the POVM $\left\{\Pi_{(s),\hat{u}^{\mathcal{L}_s},u^s}^B : u^s \in G\right\}$ on the output system

$B = (B_s)_{s\in S_n}$ and we let $\hat{u}^s$ be the measurement result. We assume that the POVM measurement is designed so that if σ^B was the state of the B system before the measurement, and if the output $\hat{u}^s$ occurs, then the post-measurement state is $\frac{\sqrt{\Pi^B_{(s),\hat{u}^{\mathcal{L}_s},\hat{u}^s}}\sigma^B\sqrt{\Pi^B_{(s),\hat{u}^{\mathcal{L}_s},\hat{u}^s}}}{\mathrm{Tr}\left(\Pi^B_{(s),\hat{u}^{\mathcal{L}_s},\hat{u}^s}\sigma^B\right)}$.

The whole procedure is equivalent to applying the POVM

$$\left\{\Lambda^B_u :\ u = (u^s)_{s\in S_n} \in G^{S_n}\right\}$$

defined as:

$$\Lambda^B_u = \sqrt{\Pi^B_{(s_1),u^{s_1}}}\cdots\sqrt{\Pi^B_{(s_i),u^{\mathcal{L}_{s_i}},u^{s_i}}}\cdots\sqrt{\Pi^B_{(s_N),u^{\mathcal{L}_{s_N}},u^{s_N}}}\sqrt{\Pi^B_{(s_N),u^{\mathcal{L}_{s_N}},u^{s_N}}}\cdots \sqrt{\Pi^B_{(s_i),u^{\mathcal{L}_{s_i}},u^{s_i}}}\cdots\sqrt{\Pi^B_{(s_1),u^{s_1}}},$$

where $s_1 < s_2 < \ldots < s_N$ are the $N = 2^n$ elements of S_n ordered according to the order relation $<$.

It is easy to see that $\Lambda_u \geq 0$ for every $u \in G^{S_n}$, and $\sum_{u\in G^{S_n}} \Lambda_u = I$.

8.7.3 Performance of Polar Codes

For every $s \in S_n$, let $\mathcal{F}_s$ be the set of section mappings of G/H_s. We have:

$$\mathcal{F}_s = \left\{f_s \in G^{G/H_s} :\ f_s(\tilde{u}^s) \in \tilde{u}^s \text{ for all } \tilde{u}^s \in G/H_s\right\}.$$

It is easy to see that $|\mathcal{F}_s| = |H_s|^{|G/H_s|}$. Define

$$\mathcal{F} := \prod_{s\in S_n} \mathcal{F}_s.$$

For every $f = (f_s)_{s\in S_n} \in \mathcal{F}$ and every $\tilde{u} = (\tilde{u}^s)_{s\in S_n} \in \prod_{s\in S_n}(G/H_s)$, define $f(\tilde{u}) = \left(f_s(\tilde{u}^s)\right)_{s\in S_n} \in G^{S_n}$.

The probability of error of the quantum successive cancellation decoder for a particular choice of $f = (f_s)_{s\in S_n} \in \mathcal{F} = \prod_{s\in S_n} \mathcal{F}_s$ is given by:

$$\begin{aligned} P_e(f) &= \frac{1}{\prod_{s\in S_n}|G/H_s|}\sum_{\tilde{u}\in\prod_{s\in S_n}(G/H_s)}\left(1 - \mathrm{Tr}\left(\Lambda^B_{f(\tilde{u})}\rho^B_{f(\tilde{u})}\right)\right) \\ &= \mathbb{E}_{\tilde{U}}\left(1 - \mathrm{Tr}\left(\Lambda^B_{f(\tilde{U})}\rho^B_{f(\tilde{U})}\right)\right), \end{aligned}$$

where $\tilde{U} = (\tilde{U}^s)_{s\in S_n}$ is uniformly distributed in $\prod_{s\in S_n}(G/H_s)$.

The probability of error averaged over all the choices of $f = (f_s)_{s\in S_n} \in \mathcal{F} = \prod_{s\in S_n} \mathcal{F}_s$ is:

$$\begin{aligned}\overline{P}_e &= \frac{1}{|\mathcal{F}|}\sum_{f\in\mathcal{F}} P_e(f) = \frac{1}{|\mathcal{F}|}\sum_{f\in\mathcal{F}} \mathbb{E}_{\tilde{U}}\left(1-\mathrm{Tr}\left(\Lambda_{f(\tilde{U})}\rho^B_{f(\tilde{U})}\right)\right)\\ &= \mathbb{E}_{F,\tilde{U}}\left(1-\mathrm{Tr}\left(\Lambda^B_{F(\tilde{U})}\rho^B_{F(\tilde{U})}\right)\right) = \mathbb{E}_{F,\tilde{U}}\left(1-\mathrm{Tr}\left(\Lambda^B_U\rho^B_U\right)\right),\end{aligned}$$

where $F=(F_s)_{s\in S_n}$ is uniformly distributed in $\mathcal{F}=\prod_{s\in S_n}\mathcal{F}_s$, and $U=(U^s)_{s\in S_n} = F(U)=\left(F_s(\tilde{U}^s)\right)_{s\in S_n}$. It is easy to see that $\{U^s:\ s\in S_n\}$ are independent and uniformly distributed in G. We have:

$$\begin{aligned}\overline{P}_e &= \mathbb{E}_{F,\tilde{U}}\left(1-\mathrm{Tr}\left(\Lambda^B_U\rho^B_U\right)\right)\\ &= \mathbb{E}_{F,\tilde{U}}\left(1-\mathrm{Tr}\left(\sqrt{\Pi^B_{(s_N),U^{\mathcal{L}_{s_N}},U^{s_N}}}\cdots\sqrt{\Pi^B_{(s_1),U^{s_1}}}\rho^B_U\sqrt{\Pi^B_{(s_1),U^{s_1}}}\cdots\sqrt{\Pi^B_{(s_N),U^{\mathcal{L}_{s_N}},U^{s_N}}}\right)\right)\\ &\overset{(a)}{\le} \mathbb{E}_{F,\tilde{U}}\left(2\sqrt{N}\sqrt{\sum_{i=1}^N\left(1-\mathrm{Tr}\left(\Pi^B_{(s_i),U^{\mathcal{L}_{s_i}},U^{s_i}}\rho^B_U\right)\right)}\right)\\ &\overset{(b)}{\le} 2\sqrt{N}\sqrt{\mathbb{E}_{F,\tilde{U}}\left(\sum_{i=1}^N\left(1-\mathrm{Tr}\left(\Pi^B_{(s_i),U^{\mathcal{L}_{s_i}},U^{s_i}}\rho^B_U\right)\right)\right)}\\ &= 2\sqrt{N}\sqrt{\sum_{s\in S_n}\mathbb{E}_{F,\tilde{U}}\left(1-\mathrm{Tr}\left(\Pi^B_{(s),U^{\mathcal{L}_s},U^s}\rho^B_U\right)\right)}\\ &\overset{(c)}{=} 2\sqrt{N}\sqrt{\sum_{s\in S_n}\mathbb{E}_{F,\tilde{U}}\left(1-\mathrm{Tr}\left(\Pi^B_{(s),U^{\mathcal{L}_s},\tilde{U}^s}\rho^B_U\right)\right)}\\ &\overset{(d)}{=} 2\sqrt{N}\sqrt{\sum_{s\in S_n}\mathbb{E}_{U,\tilde{U}^s}\left(1-\mathrm{Tr}\left(\Pi^B_{(s),U^{\mathcal{L}_s},\tilde{U}^s}\rho^B_U\right)\right)}\\ &= 2\sqrt{N}\sqrt{\sum_{s\in S_n}\left(1-\mathbb{E}_{\tilde{U}^s,U^{\mathcal{L}_s}}\mathrm{Tr}\left(\Pi^B_{(s),U^{\mathcal{L}_s},\tilde{U}^s}\mathbb{E}_{U^s,U^{\mathcal{U}_s}|\tilde{U}^s,U^{\mathcal{L}_s}}\left(\rho^B_U\right)\right)\right)}\\ &\overset{(e)}{=} 2\sqrt{N}\sqrt{\sum_{s\in S_n}\left(1-\mathbb{E}_{\tilde{U}^s,U^{\mathcal{L}_s}}\left(\mathrm{Tr}\left(\Pi^B_{(s),U^{\mathcal{L}_s},\tilde{U}^s}\overline{\rho}^B_{\tilde{U}^s,U^{\mathcal{L}_s}}\right)\right)\right)},\end{aligned}$$

where (a) follows from the "non-commutative union bound" of Lemma 8.1. (b) follows from the concavity of the square root. (c) follows from the fact that $U^s = f_s(\tilde{U}^s)$, which implies that $U^s \bmod H_s = \tilde{U}^s$ and $U^s = f_s(U^s \bmod H_s)$, which in turn implies that $\Pi^B_{(s),U^{\mathcal{L}_s},U^s} = \Pi^B_{(s),U^{\mathcal{L}_s},U^s \bmod H_s}$. (d) follows from the fact that $\mathrm{Tr}\left(\Pi^B_{(s),U^{\mathcal{L}_s},\tilde{U}^s}\rho^B_U\right)$ depends only on $\tilde{U}^s$ and U. (e) follows from the fact that for every $\tilde{u}^s\in G/H_s$ and every $u^{\mathcal{L}_s}\in G^{\mathcal{L}_s}$, we have:

$$\mathbb{E}_{U^s,U^{\mathcal{U}_s}|\tilde{U}^s=\tilde{u}^s,U^{\mathcal{L}_s}=u^{\mathcal{L}_s}}\left(\rho^B_U\right) = \frac{1}{|H_s|\cdot q^{|\mathcal{U}_s|}}\sum_{u^s\in\tilde{u}^s}\sum_{u^{\mathcal{U}_s}\in G^{\mathcal{U}_s}}\rho^B_u = \overline{\rho}^B_{\tilde{u}^s,u^{\mathcal{L}_s}}.$$

On the other hand, we have:

$$\begin{aligned}
&\mathbb{E}_{\tilde{U}^s,U^{\mathcal{L}_s}}\left(\mathrm{Tr}\left(\Pi^B_{(s),U^{\mathcal{L}_s},\tilde{U}^s}\overline{\rho}^B_{\tilde{U}^s,U^{\mathcal{L}_s}}\right)\right)\\
&\qquad=\frac{1}{|G/H_s|}\sum_{\tilde{u}^s\in G/H_s}\frac{1}{q^{|\mathcal{L}_s|}}\sum_{u^{\mathcal{L}_s}\in G^{\mathcal{L}_s}}\mathrm{Tr}\left(\Pi^B_{(s),u^{\mathcal{L}_s},\tilde{u}^s}\overline{\rho}^B_{\tilde{u}^s,u^{\mathcal{L}_s}}\right)\\
&\qquad=\frac{1}{|G/H_s|}\sum_{\tilde{u}^s\in G/H_s}\mathrm{Tr}\left(\Pi^{B,U^{\mathcal{L}_s}}_{(s),\tilde{u}^s}\rho^{B,U^{\mathcal{L}_s}}_{s,\tilde{u}^s}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\overline{P}_e &\le 2\sqrt{N}\sqrt{\sum_{s\in S_n}\left(1-\mathbb{E}_{\tilde{U}^s,U^{\mathcal{L}_s}}\left(\mathrm{Tr}\left(\Pi^B_{(s),U^{\mathcal{L}_s},\tilde{U}^s}\overline{\rho}^B_{\tilde{U}^s,U^{\mathcal{L}_s}}\right)\right)\right)}\\
&= 2\sqrt{N}\sqrt{\sum_{s\in S_n}\left(1-\frac{1}{|G/H_s|}\sum_{\tilde{u}^s\in G/H_s}\mathrm{Tr}\left(\Pi^{B,U^{\mathcal{L}_s}}_{(s),\tilde{u}^s}\rho^{B,U^{\mathcal{L}_s}}_{s,\tilde{u}^s}\right)\right)}\\
&\overset{(a)}{\le} 2\sqrt{N}\sqrt{\sum_{s\in S_n}(|G/H_s|-1)F(W[H_s])} \le 2\sqrt{N}\sqrt{\sum_{s\in S_n}(q-1)2^{-2^{\beta' n}}}\\
&\le 2\sqrt{2^n}\sqrt{(q-1)2^n 2^{-2^{\beta' n}}} \le 2^{-2^{\beta n}},
\end{aligned}$$

where (a) follows from Lemma 8.12.

The above upper bound was calculated on average over a random choice of the frozen section mappings. Therefore, there is at least one choice of the frozen section mappings for which the upper bound of the probability of error still holds.

It remains to study the rate of the constructed polar code. The rate at which we are communicating is $R=\frac{1}{2^n}\sum_{s\in\{-,+\}^n}\log_2|G/H_s|=\frac{1}{2^n}\sum_{s\in E_n}\log_2|G/H_s|$. On the other hand, we have $\left|I(W^s)-\log_2|G/H_s|\right|<\frac{\delta}{2}$ for all $s\in E_n$. Now since we have $\sum_{s\in\{-,+\}^n}I(W^s)=2^nI(W)$, we conclude that:

$$\begin{aligned}
I(W)&=\frac{1}{2^n}\sum_{s\in\{-,+\}^n}I(W^s)=\frac{1}{2^n}\sum_{s\in E_n}I(W^s)+\frac{1}{2^n}\sum_{s\in E_n^c}I(W^s)\\
&<\frac{1}{2^n}\sum_{s\in E_n}\left(\log_2|G/H_s|+\frac{\delta}{2}\right)+\frac{1}{2^n}|E_n^c|\log_2 q\\
&<R+\frac{1}{2^n}|E_n|\frac{\delta}{2}+\frac{\delta}{2\log_2 q}\log_2 q\\
&\le R+\frac{\delta}{2}+\frac{\delta}{2}=R+\delta,
\end{aligned}$$

where $E_n^c=\{-,+\}^n\setminus E_n$.

To this end we have proven the following theorem which is the main result of this chapter:

Theorem 8.4. *Let $W : x \in G \longrightarrow \rho_x \in \mathcal{DM}(k)$ be an arbitrary cq-channel, where the input alphabet is endowed with an Abelian group operation. For every $\delta > 0$ and every $0 < \beta < \frac{1}{2}$, there exists a polar code of blocklength $N = 2^n$ based on the group operation which has a rate $R > I(W) - \delta$ and an encoder algorithm of complexity $O(N \log N)$. Moreover, the probability of error of the quantum successive cancellation decoder is less than 2^{-N^β}.*

8.8 Polar Codes for Arbitrary Classical-Quantum MACs

An m-user classical-quantum multiples access channel (cq-MAC)

$$W : (x_1, \ldots, x_m) \in G_1 \times \cdots \times G_m \longrightarrow \rho_{x_1,\ldots,x_m} \in \mathcal{DM}(k)$$

takes classical inputs $\{x_i \in G_i : \ 1 \leq i \leq m\}$ from the m users and produces a quantum output $\rho_{x_1,\ldots,x_m} \in \mathcal{DM}(k)$. We assume that the input alphabets G_i are finite but their sizes $q_i = |G_i|$ can be arbitrary.

The achievable rate-region is described by a collection of inequalities [50]:

$$\forall S \subset \{1, \ldots, m\},\ 0 \leq R_S \leq I(X_S; B|X_{S^c})_\rho = I(X_S; BX_{S^c})_\rho,$$

where $R_S = \sum_{i \in S} R_i$, $X_S = (X_i)_{i \in S}$, $S^c = \{1, \ldots, m\} \setminus S$, and the mutual information $I(X_S; Y|X_{S^c})_\rho$ is computed according to the following state:

$$\rho^{X_1,\ldots,X_m,B} = \sum_{\substack{x_1 \in G_1, \\ \vdots \\ x_m \in G_m}} \left(\prod_{i=1}^m P_{X_i}(x_i) \right) \left(\bigotimes_{1 \leq i \leq m} |x_i\rangle\langle x_i|^{X_i} \right) \otimes \rho^B_{x_1,\ldots,x_m},$$

for some independent probability distributions $\{P_{X_i}(x_i) : \ x_i \in G_i\}$ on G_i for $1 \leq i \leq m$.

We are interested in the case where the probability distributions of $X_1, \ldots, X_m$ are uniform in $G_1, \ldots, G_m$ respectively. We define the symmetric-capacity region $\mathcal{J}(W)$ of W as

$$\mathcal{J}(W) = \Big\{ (R_1, \ldots, R_m) \in \mathbb{R}^m : \ 0 \leq R_S \leq I_S(W),\ \forall S \subset \{1, \ldots, m\} \Big\},$$

where $I_S(W) := I(X_S; BX_{S^c})_\rho$ is computed according to

$$\rho^{X_1,\ldots,X_m,B} = \frac{1}{q_1 \cdots q_m} \sum_{\substack{x_1 \in G_1, \\ \vdots \\ x_m \in G_m}} \left(\bigotimes_{1 \leq i \leq m} |x_i\rangle\langle x_i|^{X_i} \right) \otimes \rho^B_{x_1,\ldots,x_m}.$$

The set $\big\{ (R_1, \ldots, R_m) \in \mathcal{J}(W) : \ R_1 + \cdots + R_m = I(W) \big\}$ is called the dominant face of $\mathcal{J}(W)$, where $I(W) := I[\{1, \ldots, m\}](W) = I(X_1 \ldots X_m; B)_\rho$ is the symmetric sum-capacity of W.

For every $1 \leq i \leq m$, we fix an Abelian group operation on G_i and we denote it additively. It is possible to construct cq-MAC codes which achieve the rates in the region $\mathcal{J}(W)$ using one of the following two methods:

- By using the monotone chain rule method of Arıkan [22] and applying a polarization transformation using the Abelian group operation for each user.
- By using the rate-splitting method described in [8] and applying a polarization transformation using the Abelian group operation for each user.

By using the cq-channel polarization results of this chapter and a similar analysis as in [22], [8] and [44], we can show that both methods yield cq-MAC codes that achieve the whole region $\mathcal{J}(W)$ for which the probability of error of the quantum successive cancellation decoder decays faster than 2^{-N^β} for any $\beta < \frac{1}{2}$, where N is the blocklength of the code.

However, one may hesitate to call the codes obtained using these methods as cq-MAC-polar codes because they are not based on the polarization of cq-MACs. These methods are hybrid schemes which combine cq-channel polarization (not cq-MAC polarization) with other techniques. Moreover, the code construction for these methods is more complicated than cq-MAC-polar codes. In the rest of this section, we describe how cq-MAC-polar codes are constructed.

We define the cq-MACs W^- and W^+ as follows:

$$W^- : (u_{1,1}, \dots, u_{m,1}) \in G_1 \times \cdots \times G_m \longrightarrow \rho^-_{u_{1,1},\dots,u_{m,1}} \in \mathcal{DM}(k^2),$$

$$W^+ : (u_{1,2}, \dots, u_{m,2}) \in G_1 \times \cdots \times G_m \longrightarrow \rho^+_{u_{1,2},\dots,u_{m,2}} \in \mathcal{DM}(k^2 q_1 \cdots q_m),$$

where

$$\rho^-_{u_{1,1},\dots,u_{m,1}} = \frac{1}{q_1 \cdots q_m} \sum_{\substack{u_{1,2} \in G_1, \\ \vdots \\ u_{m,2} \in G_m}} \rho_{u_{1,1}+u_{1,2},\dots,u_{m,1}+u_{m,2}} \otimes \rho_{u_{1,2},\dots,u_{m,2}},$$

and

$$\rho^+_{u_{1,2},\dots,u_{m,2}} = \frac{1}{q_1 \cdots q_m} \sum_{\substack{u_{1,1} \in G_1, \\ \vdots \\ u_{m,1} \in G_m}} \rho_{u_{1,1}+u_{1,1},\dots,u_{m,1}+u_{m,2}} \otimes \rho_{u_{1,1},\dots,u_{m,2}} \otimes \left(\bigotimes_{1 \le i \le m} |u_{i,1}\rangle\langle u_{i,1}| \right).$$

Note that the cq-MAC W can be seen as a cq-channel with input in $G := G_1 \times \cdots \times G_m$. Moreover, W^- and W^+ when seen as cq-channels can be obtained from the cq-channel W by applying the polarization transformation which uses the Abelian group operation of the product group G. Therefore, the cq-channel polarization results of the previous sections can be applied to W. In particular, we have:

- $I(W^-) + I(W^+) = 2I(W)$. This shows that the symmetric sum-capacity is conserved by the polarization transformation and that for every $n > 0$, the region $\dfrac{1}{2^n} \displaystyle\sum_{s \in \{-,+\}^n} \mathcal{J}(W^s)$ contains points on the dominant face of $\mathcal{J}(W)$.
- For every subgroup H of G, we have $I(W^-[H]) + I(W^-[H]) \ge 2I(W[H])$ by Lemma 8.3. Therefore, for every $S \subset \{1, \dots, m\}$, we have

$$\begin{aligned} I_S(W^-) + I_S(W^+) &= \big(I(W^-) - I(W^-[G_S])\big) + \big(I(W^+) - I(W^+[G_S])\big) \\ &\le 2I(W) - 2I(W[G_S]) = 2I_S(W), \end{aligned} \tag{8.6}$$

where,

$$G_S = \left(\prod_{i\in S} G_i\right) \times \left(\prod_{j\notin S} \{0\}\right).$$

Equation (8.6) shows that although the symmetric-sum capacity is conserved by polarization, the highest achievable individual rates can decrease. In other words, polarization can induce a loss in the symmetric-capacity region.

- Theorem 8.3 implies that

$$\lim_{n\to\infty} \frac{1}{2^n} \Big| \big\{ s \in \{-,+\}^n : \exists H_s \text{ subgroup of } G, \\ \big| I(W^s) - \log_2 |G/H_s| \big| < \delta, \big| I(W^s[H_s]) - \log_2 |G/H_s| \big| < \delta, F(W^s[H_s]) < 2^{-2^{\beta n}} \big\} \Big| = 1.$$

In other words, as the number of polarization steps becomes large, the synthetic cq-MACs become close to deterministic homomorphism cq-channels which project the input $(U_1^s, \ldots, U_m^s)$ onto some quotient group G/H_s of the product group G.

One can employ the properties of subgroups of product groups to show that the polarized cq-MAC W^s is an "easy" cq-MAC in a sense similar to the way easy MACs were defined in Definition 4.6. This allows the construction of cq-MAC-polar codes for which the probability of error of the quantum successive cancellation decoder decays faster than 2^{-N^β} for any $0 < \beta < \frac{1}{2}$, where $N = 2^n$ is the blocklength of the code. The region of rates that are achievable by cq-MAC-polar codes is given by:

$$\begin{aligned} \mathcal{J}^{\text{pol}}(W) &= \bigcap_{n\geq 0} \left(\frac{1}{2^n} \sum_{s\in\{-,+\}^n} \mathcal{J}(W^s) \right) \\ &= \Big\{ (R_1, \ldots, R_m) \in \mathbb{R}^m : \; R_S \leq I_S^{\text{pol}}(W), \; \forall S \subset \{1, \ldots, m\} \Big\}, \end{aligned}$$

where

$$I_S^{\text{pol}}(W) = \lim_{n\to\infty} \frac{1}{2^n} \sum_{s\in\{-,+\}^n} I_S(W^s).$$

The cq-MAC-polar codes can be compared to the two cq-MAC coding methods that were described at the beginning of this section:

- The cq-MAC-polar codes have the advantage that the code construction is simpler.

- The other two coding methods have the advantage that they always achieve the whole symmetric-capacity region $\mathcal{J}(W)$, which may not be the case for cq-MAC-polar codes in general.

8.9 Appendix

8.9.1 Proof of Lemma 8.1

Let $\Pi_{r+1} = I$. We have:

$$
\begin{aligned}
&1 - \mathrm{Tr}\left(\sqrt{\Pi_r}\ldots\sqrt{\Pi_1}\rho\sqrt{\Pi_1}\ldots\sqrt{\Pi_r}\right) \\
&\quad = \mathrm{Tr}\left(\sqrt{\Pi_{r+1}}\rho\sqrt{\Pi_{r+1}}\right) - \mathrm{Tr}\left(\sqrt{\Pi_{r+1}}\ldots\sqrt{\Pi_1}\rho\sqrt{\Pi_1}\ldots\sqrt{\Pi_{r+1}}\right) \\
&\quad = \sum_{i=1}^{r} \mathrm{Tr}\left(\sqrt{\Pi_{r+1}}\ldots\sqrt{\Pi_{i+1}}\rho\sqrt{\Pi_{i+1}}\ldots\sqrt{\Pi_{r+1}}\right) - \mathrm{Tr}\left(\sqrt{\Pi_{r+1}}\ldots\sqrt{\Pi_i}\rho\sqrt{\Pi_i}\ldots\sqrt{\Pi_{r+1}}\right) \\
&\quad = \sum_{i=1}^{r} \mathrm{Tr}\left(\sqrt{\Pi_{r+1}}\ldots\sqrt{\Pi_{i+1}}\left(\rho - \sqrt{\Pi_i}\rho\sqrt{\Pi_i}\right)\sqrt{\Pi_{i+1}}\ldots\sqrt{\Pi_{r+1}}\right) \\
&\quad \overset{(a)}{\leq} \sum_{i=1}^{r} \mathrm{Tr}\left(\sqrt{\Pi_{r+1}}\ldots\sqrt{\Pi_{i+1}}\cdot\left|\rho - \sqrt{\Pi_i}\rho\sqrt{\Pi_i}\right|\cdot\sqrt{\Pi_{i+1}}\ldots\sqrt{\Pi_{r+1}}\right) \\
&\quad \overset{(b)}{\leq} \sum_{i=1}^{r} \mathrm{Tr}\left|\rho - \sqrt{\Pi_i}\rho\sqrt{\Pi_i}\right| = \sum_{i=1}^{r}\left\|\rho - \sqrt{\Pi_i}\rho\sqrt{\Pi_i}\right\|_1 \overset{(c)}{\leq} 2\sum_{i=1}^{r}\sqrt{\mathrm{Tr}\left(\rho - \sqrt{\Pi_i}\rho\sqrt{\Pi_i}\right)} \\
&\quad = 2r\frac{1}{r}\sum_{i=1}^{r}\sqrt{1-\mathrm{Tr}(\Pi_i\rho)} \overset{(d)}{\leq} 2r\sqrt{\frac{1}{r}\sum_{i=1}^{r}(1-\mathrm{Tr}(\Pi_i\rho))} = 2\sqrt{r}\sqrt{\sum_{i=1}^{r}(1-\mathrm{Tr}(\Pi_i\rho))},
\end{aligned}
$$

where (a) follows from the fact that $\sqrt{\Pi_j} \geq 0$ for every $i+1 \leq j \leq r+1$, $\rho - \sqrt{\Pi_i}\rho\sqrt{\Pi_i} \leq \left|\rho - \sqrt{\Pi_i}\rho\sqrt{\Pi_i}\right|$ and the fact that if $A \leq B$ and $C \geq 0$, then $\mathrm{Tr}(AC) \leq \mathrm{Tr}(BC)$. (b) follows from the fact that $0 \leq \sqrt{\Pi_j} \leq I$ for every $i+1 \leq j \leq r+1$, $\left|\rho - \sqrt{\Pi_i}\rho\sqrt{\Pi_i}\right| \geq 0$, and the fact that if A, B are two positive operators with $B \leq I$, then $\mathrm{Tr}(AB) \leq \mathrm{Tr}(AB) + \mathrm{Tr}(A(I-B)) = \mathrm{Tr}(A)$. (c) follows from the fact that $\left\|\rho - \sqrt{X}\rho\sqrt{X}\right\|_1 \leq 2\sqrt{\mathrm{Tr}\left(\rho - \sqrt{X}\rho\sqrt{X}\right)}$ for every positive operator $X \leq I$ (see [51]). (d) follows from the concavity of the square root.

8.9.2 Proof of Proposition 8.1

In [52, Prop. 1], it was shown that for every $0 \leq s \leq 1$, we have:

$$
I(W) \geq -\frac{1}{s}\log_2 \mathrm{Tr}\left[\left(\sum_{x\in G} P_X(x)\cdot\rho_x^{\frac{1}{1+s}}\right)^{1+s}\right].
$$

By taking $s = 1$, we obtain:

$$
\begin{aligned}
I(W) &\geq -\log_2 \operatorname{Tr}\left[\left(\sum_{x\in G}\frac{1}{q}\cdot\sqrt{\rho_x}\right)^2\right] = -\log_2 \operatorname{Tr}\left(\frac{1}{q^2}\sum_{x,x'\in G}\sqrt{\rho_x}\sqrt{\rho_{x'}}\right)\\
&= -\log_2 \operatorname{Tr}\left(\frac{1}{q^2}\sum_{x\in G}\rho_x + \frac{1}{q^2}\sum_{\substack{x,x'\in G,\\ x\neq x'}}\sqrt{\rho_x}\sqrt{\rho_{x'}}\right)\\
&= -\log_2\left(\frac{1}{q} + \frac{1}{q^2}\sum_{\substack{x,x'\in G,\\ x\neq x'}}\operatorname{Tr}(\sqrt{\rho_x}\sqrt{\rho_{x'}})\right)\\
&\overset{(a)}{\geq} -\log_2\left(\frac{1}{q} + \frac{1}{q^2}\sum_{\substack{x,x'\in G,\\ x\neq x'}}F(\rho_x,\rho_{x'})\right) = \log_2\frac{q}{1+(q-1)F(W)},
\end{aligned}
$$

where (a) follows from the fact that $\operatorname{Tr}(\sqrt{\rho_x}\sqrt{\rho_{x'}}) \leq \operatorname{Tr}(|\sqrt{\rho_x}\sqrt{\rho_{x'}}|) = \|\sqrt{\rho_x}\sqrt{\rho_{x'}}\|_1 = F(\rho_x,\rho_{x'})$.

In order to prove the second inequality, define the cq-channel $\tilde{W} : x \in G \longrightarrow \tilde{\rho}_x \in \mathcal{DM}(k\cdot q^2)$ as follows:

$$
\tilde{\rho}_x^{BS_1S_2} = \rho_x^B \otimes \left(\frac{1}{2(q-1)}\sum_{\substack{x'\in G,\\ x'\neq x}}\left(|x\rangle\langle x|^{S_1}\otimes|x'\rangle\langle x'|^{S_2} + |x'\rangle\langle x'|^{S_1}\otimes|x\rangle\langle x|^{S_2}\right)\right).
$$

The two additional systems S_1 and S_2 can be interpreted as additional side information about the input which is provided to the receiver. Note that if S_1S_2 are traced out, we recover the cq-channel W.

Let $\tilde{\rho}^{XBS_1S_2} = \frac{1}{q}\sum_{x\in G}|x\rangle\langle x|^X \otimes \tilde{\rho}_x^{BS_1S_2}$. We have:

$$
\begin{aligned}
I(W) &= I(X;B)_{\tilde{\rho}} \leq I(X;BS_1S_2)_{\tilde{\rho}} = I(X;S_1S_2)_{\tilde{\rho}} + I(X;B|S_1S_2)_{\tilde{\rho}}\\
&= H(X) - H(X|S_1S_2) + I(X;B|S_1S_2)_{\tilde{\rho}}\\
&\overset{(a)}{=} \log_2(q) - 1 + \sum_{s_1,s_2\in G} I(X;B|S_1=s_1,S_2=s_2)P_{S_1,S_2}(s_1,s_2)\\
&\overset{(b)}{=} \log_2(q/2) + \frac{1}{q(q-1)}\sum_{\substack{s_1,s_2\in G,\\ s_1\neq s_2}} I(X;B|S_1=s_1,S_2=s_2)\\
&\overset{(c)}{=} \log_2(q/2) + \frac{1}{q(q-1)}\sum_{\substack{s_1,s_2\in G,\\ s_1\neq s_2}} I(W_{s_1,s_2}),
\end{aligned}
$$

where (a) follows from the fact that given $\{S_1 = s_1, S_2 = s_2\}$, the conditional probability distribution of X is uniform in $\{s_1, s_2\}$. (b) follows from the fact that

the distribution of (S_1, S_2) is uniform in the set

$$\{(s_1, s_2) \in G \times G : \ s_1 \neq s_2\}.$$

(c) is true because conditioning $\tilde{\rho}^{XBS_1S_2}$ on $\{S_1 = s_2, S_2 = s_2\}$ and then tracing out S_1S_2 gives the state $\frac{1}{2}|s_1\rangle\langle s_1|^X \otimes \rho_{s_1}^B + \frac{1}{2}|s_2\rangle\langle s_2|^X \otimes \rho_{s_2}^B$ which just represents W_{s_1,s_2} with uniform input, where $W_{s_1,s_2} : x \in \{0,1\} \longrightarrow \rho_{x,s_1,s_2} \in \mathcal{DM}(k)$ is the binary-input cq-channel defined as $\rho_{0,s_1,s_2} = \rho_{s_1}$ and $\rho_{1,s_1,s_2} = \rho_{s_2}$. In other words, the cq-channel W_{s_1,s_2} is obtained from W by restricting the input to $\{s_1, s_2\}$.

Now since W_{s_1,s_2} is a binary-input cq-channel, we have from [43, Prop. 1] that

$$I(W_{s_1,s_2}) \leq \sqrt{1 - F(W_{s_1,s_2})^2} = \sqrt{1 - F(\rho_{s_1}, \rho_{s_2})^2}.$$

Therefore,

$$\begin{aligned} I(W) &\leq \log_2(q/2) + \frac{1}{q(q-1)} \sum_{\substack{s_1,s_2 \in G, \\ s_1 \neq s_2}} \sqrt{1 - F(\rho_{s_1}, \rho_{s_2})^2} \\ &\leq \log_2(q/2) + \sqrt{1 - F(W)^2}, \end{aligned}$$

where the last inequality follows from the concavity of the function $t \to \sqrt{1-t^2}$.

It remains to show the last inequality of Proposition 8.1. Define the following:

- $\rho^{XB} = \frac{1}{q} \sum_{x \in G} |x\rangle\langle x|^X \otimes \rho_x^B$.
- $\Lambda^{XB} = \sum_{x \in G} |x\rangle\langle x|^X \otimes E_x^B$, where $\{E_x^B : \ x \in G\}$ is an optimal POVM that decodes W with the lowest probability of error.

We have:

- $\rho^X = \mathrm{Tr}_B(\rho^{XB}) = \frac{1}{q} \sum_{x \in G} |x\rangle\langle x|^X$.
- $\rho^B = \mathrm{Tr}_X(\rho^{XB}) = \frac{1}{q} \sum_{x \in G} \rho_x^B$.

From [47, Sec 9.2.3], we have

$$D\left(\rho^{XB}, \rho^X \otimes \rho^B\right)^2 + F\left(\rho^{XB}, \rho^X \otimes \rho^B\right)^2 \leq 1, \tag{8.7}$$

where $D(\rho', \rho'') = \frac{1}{2}\|\rho' - \rho''\|_1$ is the trace distance between ρ' and ρ''. We have:

$$
\begin{aligned}
F\left(\rho^{XB}, \rho^X \otimes \rho^B\right) &= \left\| \sqrt{\rho^{XB}} \sqrt{\rho^X \otimes \rho^B} \right\|_1 \\
&= \left\| \frac{1}{q} \left(\sum_{x \in G} |x\rangle\langle x|^X \otimes \sqrt{\rho_x^B} \right) \cdot \left(\sum_{x \in G} |x\rangle\langle x|^X \otimes \sqrt{\rho^B} \right) \right\|_1 \\
&= \frac{1}{q} \left\| \sum_{x \in G} |x\rangle\langle x|^X \otimes \sqrt{\rho_x^B} \sqrt{\rho^B} \right\|_1 = \frac{1}{q} \sum_{x \in G} \left\| \sqrt{\rho_x^B} \sqrt{\rho^B} \right\|_1 \\
&= \frac{1}{q} \sum_{x \in G} F\left(\rho_x^B, \rho^B\right) = \frac{1}{q} \sum_{x \in G} F\left(\rho_x^B, \frac{1}{q} \sum_{x' \in G} \rho_{x'}^B\right) \\
&\overset{(a)}{\geq} \frac{1}{q^2} \sum_{x, x' \in G} F\left(\rho_x^B, \rho_{x'}^B\right) = \frac{1}{q^2} \left(q + \sum_{\substack{x, x' \in G, \\ x \neq x'}} F\left(\rho_x^B, \rho_{x'}^B\right) \right) \\
&= \frac{1}{q} \left(1 + (q-1) F(W)\right),
\end{aligned}
\tag{8.8}
$$

where (a) follows from the concavity of the fidelity.

Now let $P_c(W) = 1 - P_e(W)$ be the probability of correct guess of the optimal decoder $\{E_x^B : x \in G\}$. We have:

$$
P_c(W) = \frac{1}{q} \sum_{x \in G} \operatorname{Tr}\left(E_x^B \rho_x^B\right) = \frac{1}{q} \sum_{x \in G} \operatorname{Tr}\left(|x\rangle\langle x|^X \otimes E_x^B \rho_x^B\right) = \operatorname{Tr}\left(\Lambda^{XB} \rho^{XB}\right).
$$

Therefore,

$$
\begin{aligned}
\operatorname{Tr}\left(\Lambda^{XB}\left(\rho^{XB} - \rho^X \otimes \rho^B\right)\right) &= P_c(W) - \operatorname{Tr}\left(\frac{1}{q} \sum_{x \in G} |x\rangle\langle x|^X \otimes E_x^B \rho^B\right) \\
&= P_c(W) - \frac{1}{q} \sum_{x \in G} \operatorname{Tr}(E_x^B \rho^B) = P_c(W) - \frac{1}{q} \overset{(a)}{\geq} 0,
\end{aligned}
$$

where (a) follows from the fact that a random guess gives a probability of correct guess $\frac{1}{q}$.

On the other hand, we know that $D(\rho^{XB}, \rho^X \otimes \rho^B) = \max_{0 \leq \Gamma \leq I} \operatorname{Tr}(\Gamma(\rho^{XB} - \rho^X \otimes \rho^B))$. Therefore,

$$
\begin{aligned}
0 \leq P_c(W) - \frac{1}{q} = \operatorname{Tr}\left(\Lambda^{XB}\left(\rho^{XB} - \rho^X \otimes \rho^B\right)\right) &\overset{(b)}{\leq} \max_{0 \leq \Gamma \leq I} \operatorname{Tr}(\Gamma(\rho^{XB} - \rho^X \otimes \rho^B)) \\
&= D\left(\rho^{XB}, \rho^X \otimes \rho^B\right),
\end{aligned}
\tag{8.9}
$$

where (b) follows from the fact that $0 \leq \Lambda^{XB} \leq I$.

By combining (8.7), (8.8) and (8.9), we get:

$$
\left(P_c(W) - \frac{1}{q}\right)^2 + \frac{1}{q^2} \left(1 + (q-1) F(W)\right)^2 \leq 1.
$$

Thus,

$$P_c(W) \leq \frac{1}{q} + \sqrt{1 - \frac{1}{q^2}\left(1 + (q-1)F(W)\right)^2} = \frac{1 + \sqrt{q^2 - \left(1 + (q-1)F(W)\right)^2}}{q},$$

which implies that

$$H(X|B) \overset{(a)}{\geq} -\log_2 P_c(W) \geq \log_2 q - \log_2 \left(1 + \sqrt{q^2 - \left(1 + (q-1)F(W)\right)^2}\right),$$

where (a) follows from [53, Prop 4.3] and the operational interpretation of the conditional min-entropy of a cq-state in terms of the guessing probability [54]. Therefore,

$$\begin{aligned} I(W) = I(X;B) &= H(X) - H(X|B) \\ &= \log_2 q - H(X|B) \leq \log_2 \left(1 + \sqrt{q^2 - \left(1 + (q-1)F(W)\right)^2}\right). \end{aligned}$$

8.9.3 Proof of Proposition 8.2

Lemma 8.13. *Let A and B be two positive semi-definite $k \times k$ matrices. We have*[15]*:*

$$\operatorname{Tr}\sqrt{A+B} \leq \operatorname{Tr}\sqrt{A} + \operatorname{Tr}\sqrt{B}.$$

Proof. Let us first assume that A and B are invertible. Since the mapping $C \to C^{-1}$ is monotonically decreasing [56], we have $(A+B)^{-1} \leq A^{-1}$. Moreover, since the square root is operator monotone [56], we have $(A+B)^{-\frac{1}{2}} \leq A^{-\frac{1}{2}}$. Similarly, $(A+B)^{-\frac{1}{2}} \leq B^{-\frac{1}{2}}$. Therefore,

$$\begin{aligned} \operatorname{Tr}\sqrt{A+B} &= \operatorname{Tr}\left((A+B)\cdot(A+B)^{-\frac{1}{2}}\right) = \operatorname{Tr}\left(A\cdot(A+B)^{-\frac{1}{2}}\right) + \operatorname{Tr}\left(B\cdot(A+B)^{-\frac{1}{2}}\right) \\ &\overset{(a)}{\leq} \operatorname{Tr}\left(A\cdot A^{-\frac{1}{2}}\right) + \operatorname{Tr}\left(B\cdot B^{-\frac{1}{2}}\right) = \operatorname{Tr}\sqrt{A} + \operatorname{Tr}\sqrt{B}, \end{aligned}$$

where (a) follows from the fact that if $C \leq D$ and $A \geq 0$, then $\operatorname{Tr}(AC) \leq \operatorname{Tr}(AD)$.

Now let A and B be two arbitrary positive semi-definite $k \times k$ matrices. We have:

$$\operatorname{Tr}\sqrt{A+B} = \lim_{\epsilon\to 0} \operatorname{Tr}\sqrt{A+B+2\epsilon I} \leq \lim_{\epsilon\to 0} \operatorname{Tr}\sqrt{A+\epsilon I} + \operatorname{Tr}\sqrt{B+\epsilon I} = \operatorname{Tr}\sqrt{A} + \operatorname{Tr}\sqrt{B}.$$

□

Lemma 8.14. *Let $\rho_1, \ldots, \rho_n$ and $\sigma_1, \ldots, \sigma_m$ be $n+m$ density matrices of the same dimension. Let $\{p_1, \ldots, p_n\}$ and $\{q_1, \ldots, q_m\}$ be probability distributions on $\{1, \ldots, n\}$ and $\{1, \ldots, m\}$ respectively. We have:*

$$F\left(\sum_{i=1}^n p_i\rho_i, \sum_{j=1}^m q_j\sigma_j\right) \leq \sum_{i=1}^n \sum_{j=1}^m \sqrt{p_i q_j} F(\rho_i, \sigma_j).$$

[15] The proof of Lemma 8.13 is due to Martin Argerami who thankfully answered my question on Math Stack Exchange. In an earlier version of this work, we used a weaker inequality $\operatorname{Tr}\sqrt{\sum_{i=1}^n A_i} \leq n\sum_{i=1}^n \operatorname{Tr}\sqrt{A_i}$ which we proved using Weyl's inequality [55] that relates the eigenvalues of $A+B$ with those of A and B.

Proof. It is sufficient to show the lemma for the case where $n = 1$:

$$F\left(\rho, \sum_{j=1}^{m} q_j \sigma_j\right) = \operatorname{Tr} \sqrt{\rho^{\frac{1}{2}} \left(\sum_{j=1}^{m} q_j \sigma_j\right) \rho^{\frac{1}{2}}} \overset{(a)}{\leq} \sum_{j=1}^{m} \sqrt{q_j} \operatorname{Tr} \sqrt{\rho^{\frac{1}{2}} \sigma_j \rho^{\frac{1}{2}}} = \sum_{j=1}^{m} \sqrt{q_j} F(\rho, \sigma_j),$$

where (a) follows from Lemma 8.13. □

Now we are ready to prove Proposition 8.2:

$$\begin{aligned}
F_d(W^+) &= \frac{1}{q} \sum_{x \in G} F(\rho_x^+, \rho_{x+d}^+) \\
&= \frac{1}{q} \sum_{x \in G} F\left(\frac{1}{q} \sum_{u_1 \in G} \rho_{u_1+x} \otimes \rho_x \otimes |u_1\rangle\langle u_1|, \frac{1}{q} \sum_{u_1 \in G} \rho_{u_1+x+d} \otimes \rho_{x+d} \otimes |u_1\rangle\langle u_1|\right) \\
&= \frac{1}{q} \sum_{x \in G} F\left(\left(\frac{1}{q} \sum_{u_1 \in G} |u_1\rangle\langle u_1| \otimes \rho_{u_1+x}\right) \otimes \rho_x, \left(\frac{1}{q} \sum_{u_1 \in G} |u_1\rangle\langle u_1| \otimes \rho_{u_1+x+d}\right) \otimes \rho_{x+d}\right) \\
&= \frac{1}{q} \sum_{x \in G} F\left(\frac{1}{q} \sum_{u_1 \in G} |u_1\rangle\langle u_1| \otimes \rho_{u_1+x}, \frac{1}{q} \sum_{u_1 \in G} |u_1\rangle\langle u_1| \otimes \rho_{u_1+x+d}\right) \cdot F\left(\rho_x, \rho_{x+d}\right) \\
&= \frac{1}{q} \sum_{x \in G} \left(\frac{1}{q} \sum_{u_1 \in G} F\left(\rho_{u_1+x}, \rho_{u_1+x+d}\right)\right) \cdot F\left(\rho_x, \rho_{x+d}\right) \\
&= \frac{1}{q} \sum_{x \in G} F_d(W) \cdot F\left(\rho_x, \rho_{x+d}\right) = F_d(W)^2.
\end{aligned}$$

$$\begin{aligned}
F_d(W^-) &= \frac{1}{q} \sum_{x \in G} F(\rho_x^-, \rho_{x+d}^-) \\
&= \frac{1}{q} \sum_{x \in G} F\left(\frac{1}{q} \sum_{u_2 \in G} \rho_{x+u_2} \otimes \rho_{u_2}, \frac{1}{q} \sum_{u_2 \in G} \rho_{x+d+u_2} \otimes \rho_{u_2}\right) \\
&\overset{(a)}{\geq} \frac{1}{q^2} \sum_{x, u_2 \in G} F(\rho_{x+u_2} \otimes \rho_{u_2}, \rho_{x+d+u_2} \otimes \rho_{u_2}) \\
&= \frac{1}{q^2} \sum_{x, u_2 \in G} F(\rho_{x+u_2}, \rho_{x+d+u_2}) = F_d(W),
\end{aligned}$$

where (a) follows from the joint concavity of the fidelity.

$$
\begin{aligned}
F_d(W^-) &= \frac{1}{q}\sum_{x\in G} F\left(\frac{1}{q}\sum_{u_2\in G}\rho_{x+u_2}\otimes\rho_{u_2}, \frac{1}{q}\sum_{u_2'\in G}\rho_{x+d+u_2'}\otimes\rho_{u_2'}\right)\\
&\overset{(a)}{\le} \frac{1}{q}\sum_{x\in G}\sum_{u_2,u_2'\in G}\frac{1}{\sqrt{q^2}}F\left(\rho_{x+u_2}\otimes\rho_{u_2}, \rho_{x+d+u_2'}\otimes\rho_{u_2'}\right)\\
&= \frac{1}{q^2}\sum_{x,u_2,u_2'\in G} F\left(\rho_{x+u_2}, \rho_{x+d+u_2'}\right)\cdot F\left(\rho_{u_2}, \rho_{u_2'}\right)\\
&= \frac{1}{q^2}\sum_{x,u_2\in G} F\left(\rho_{x+u_2}, \rho_{x+d+u_2}\right) + \frac{1}{q^2}\sum_{x,u_2\in G} F\left(\rho_{u_2}, \rho_{u_2-d}\right)\\
&\qquad\qquad + \frac{1}{q^2}\sum_{\substack{x,u_2,u_2'\in G,\\ u_2'\neq u_2,\\ u_2'\neq u_2-d}} F\left(\rho_{x+u_2}, \rho_{x+d+u_2'}\right)\cdot F\left(\rho_{u_2}, \rho_{u_2'}\right)\\
&= 2F_d(W) + \frac{1}{q^2}\sum_{\substack{\Delta\in G,\\ \Delta\neq 0,\\ \Delta\neq -d}}\sum_{x',u_2\in G} F\left(\rho_{x'}, \rho_{x'+d+\Delta}\right)F\left(\rho_{u_2}, \rho_{u_2+\Delta}\right)\\
&= 2F_d(W) + \sum_{\substack{\Delta\in G,\\ \Delta\neq 0,\\ \Delta\neq -d}} F_\Delta(W)F_{d+\Delta}(W),
\end{aligned}
$$

where (a) follows from Lemma 8.14.

8.9.4 Proof of Lemma 8.5

$$
\begin{aligned}
&F(W[M|D])\\
&\quad= \frac{1}{|D/M|(|D/M|-1)}\sum_{\substack{C,C'\in D/M,\\ C\neq C'}} F(\rho_C, \rho_{C'})\\
&\quad= \frac{|M|^2}{|H|(|H|-|M|)}\sum_{\substack{C,C'\in D/M,\\ C\neq C'}} F\left(\frac{1}{|C|}\sum_{x\in C}\rho_x, \frac{1}{|C'|}\sum_{x'\in C'}\rho_{x'}\right)\\
&\quad\overset{(a)}{\le} \frac{|M|^2}{|H|(|H|-|M|)\sqrt{|C|\cdot|C'|}}\sum_{\substack{C,C'\in D/M,\\ C\neq C'}}\sum_{\substack{x\in C,\\ x'\in C'}} F(\rho_x, \rho_{x'})\\
&\quad\overset{(b)}{\le} \frac{|M|}{|H|(|H|-|M|)}\sum_{\substack{x\in D,\\ d\in H,\\ d\notin M}} F(\rho_x, \rho_{x+d}) \le \frac{|M|}{|H|(|H|-|M|)}\sum_{\substack{d\in H,\\ d\notin M}}\frac{q}{q}\sum_{x\in G} F(\rho_x, \rho_{x+d})\\
&\quad= \frac{q\cdot|M|}{|H|(|H|-|M|)}\sum_{\substack{d\in H,\\ d\notin M}} F_d(W) \le \frac{q\cdot|M|}{|H|(|H|-|M|)}(|H|-|M|)F_{\max}^{M|H}(W),
\end{aligned}
$$

where (a) follows from Lemma 8.14, and (b) follows from the fact that $|C| = |C'| = |M|$ and the fact that $\Big\{\exists C, C' \in D/M\colon\ x \in C,\ x' \in C' \text{ and } C \neq C'\Big\}$ if and only if $\Big\{x \in D,\ x' - x \in H \text{ and } x' - x \notin M\Big\}$.

Now let us show the second inequality of Lemma 8.5. Assume that M is maximal in H and let $d \in H$ be such that $d \notin M$ and $F_{\max}^{M|H}(W) = F_d(W)$. Since $1 - F_d(W) = \frac{1}{q}\sum_{x\in G}\big(1 - F(\rho_x, \rho_{x+d})\big)$, we have $F(\rho_x, \rho_{x+d}) \geq 1 - q(1 - F_d(W)) = 1 - q\big(1 - F_{\max}^{M|H}(W)\big)$ for every $x \in G$.

For every $C \in D/M$, we have:

$$\begin{aligned}
F(\rho_C, \rho_{d+C}) &\overset{(a)}{\geq} 1 - D(\rho_C, \rho_{d+C}) = 1 - D\left(\frac{1}{|C|}\sum_{x\in C}\rho_x, \frac{1}{|C|}\sum_{x\in C}\rho_{x+d}\right) \\
&= 1 - \frac{1}{2}\left\|\frac{1}{|C|}\sum_{x\in C}(\rho_x - \rho_{x+d})\right\|_1 \geq 1 - \frac{1}{|C|}\sum_{x\in C}\frac{1}{2}\left\|\rho_x - \rho_{x+d}\right\|_1 \\
&= 1 - \frac{1}{|C|}\sum_{x\in C} D(\rho_x, \rho_{x+d}) \overset{(b)}{\geq} 1 - \frac{1}{|C|}\sum_{x\in C}\sqrt{1 - F(\rho_x, \rho_{x+d})^2} \\
&\geq 1 - \sqrt{1 - \Big(1 - q\big(1 - F_{\max}^{M|H}(W)\big)\Big)^2},
\end{aligned}$$

where (a) follows from the fact that $D(\rho', \rho'') + F(\rho', \rho'') \geq 1$ (see [47]). (here $D(\rho', \rho'') = \frac{1}{2}\|\rho' - \rho''\|_1$ is the trace distance between ρ' and ρ''.) (b) follows from the fact that $D(\rho', \rho'')^2 + F(\rho', \rho'')^2 \leq 1$ (see [47]).

Now let $C, C' \in D/M$ be such that $C \neq C'$. Since $|H/M|$ is prime, we can write $C' = ld + C$ for some $0 \leq l < |H/M|$. We have:

$$\begin{aligned}
F(\rho_C, \rho_{C'}) = F(\rho_C, \rho_{ld+C}) = \cos A(\rho_C, \rho_{ld+C}) &\overset{(a)}{\geq} \cos\left(\sum_{i=0}^{l-1} A\big(\rho_{id+C}, \rho_{(i+1)d+C}\big)\right) \\
&= \cos\left(\sum_{i=0}^{l-1}\arccos F\big(\rho_{id+C}, \rho_{(i+1)d+C}\big)\right) \\
&\overset{(b)}{\geq} \cos\left(l \cdot \arccos\left(1 - \sqrt{1 - \Big(1 - q\big(1 - F_{\max}^{M|H}(W)\big)\Big)^2}\right)\right) \\
&\overset{(c)}{\geq} \cos\left(\frac{|H| - |M|}{|M|}\arccos\left(1 - \sqrt{1 - \Big(1 - q\big(1 - F_{\max}^{M|H}(W)\big)\Big)^2}\right)\right),
\end{aligned}$$

where (a) follows from the fact that $A(\rho', \rho'') = \arccos F(\rho', \rho'')$ is a metric [47]. Note that since cos is a decreasing function on $\left[0, \frac{\pi}{2}\right]$, (a), (b) and (c) become true if we assume that $1 - \sqrt{1 - \Big(1 - q\big(1 - F_{\max}^{M|H}(W)\big)\Big)^2} \geq \cos\left(\frac{\pi}{2(q-1)}\right)$. In other words, we can take

$$\delta_q = \frac{1}{q}\left(1 - \sqrt{1 - \left(1 - \cos\left(\frac{\pi}{2(q-1)}\right)\right)^2}\right).$$

We conclude that

$$F(W[M|D]) = \frac{1}{|D/M|(|D/M|-1)} \sum_{\substack{C,C' \in D/M,\\ C \neq C'}} F(\rho_C, \rho_{C'})$$
$$\geq \cos\left(\frac{|H|-|M|}{|M|} \arccos\left(1 - \sqrt{1 - \left(1 - q\big(1 - F_{\max}^{M|H}(W)\big)\right)^2}\right)\right).$$

8.9.5 Proof of Lemma 8.10

Lemma 8.15. *For every subgroup H of G, we have:*

$$F_{\max}^{H|G}(W) \leq (q - |H|) F(W[H])$$

Proof.

$$F(W[H]) = \frac{1}{|G/H|(|G/H|-1)} \sum_{\substack{C,C' \in G/H,\\ C \neq C'}} F(\rho_C, \rho_{C'})$$
$$= \frac{1}{|G/H|(|G/H|-1)} \sum_{\substack{C,C' \in G/H,\\ C \neq C'}} F\left(\frac{1}{|C|}\sum_{x \in C} \rho_x, \frac{1}{|C'|}\sum_{x' \in C'} \rho_{x'}\right)$$
$$\overset{(a)}{\geq} \frac{1}{|G/H|(|G/H|-1)} \cdot \frac{1}{|H|^2} \sum_{\substack{C,C' \in G/H,\\ C \neq C'}} \sum_{\substack{x \in C,\\ x' \in C'}} F(\rho_x, \rho_{x'})$$
$$= \frac{1}{q(q-|H|)} \sum_{\substack{x,d \in G,\\ d \notin H}} F(\rho_x, \rho_{x+d}) = \frac{1}{q-|H|} \sum_{\substack{d \in G,\\ d \notin H}} F_d(W)$$
$$\geq \frac{1}{q-|H|} F_{\max}^{H|G}(W),$$

where (a) follows from the concavity of the fidelity and from the fact that $|C| = |C'| = |H|$. □

Now we are ready to prove Lemma 8.10. The lemma is trivial for $H = G$. Assume that $H \neq G$. We have:

$$F(W^-[H]) = F(W^-[H|G]) \overset{(a)}{\leq} \frac{q \cdot |H|}{q} F_{\max}^{H|G}(W^-) = |H| \max_{\substack{d \in G,\\ d \notin H}} F_d(W^-)$$
$$\overset{(b)}{\leq} |H| \max_{\substack{d \in G,\\ d \notin H}} \left\{ 2F_d(W) + \sum_{\substack{\Delta \in G,\\ \Delta \neq 0,\\ \Delta \neq -d}} F_\Delta(W) F_{d+\Delta}(W) \right\}$$
$$\overset{(c)}{\leq} |H| \Big(2F_{\max}^{H|G}(W) + (q-2) F_{\max}^{H|G}(W) \Big) = |H| q F_{\max}^{H|G}(W)$$
$$\overset{(d)}{\leq} |H| q (q - |H|) F(W[H]),$$

where (a) follows from Lemma 8.5. (b) follows from Proposition 8.2. (c) follows from the fact that for every $d, \Delta \in G$, if $d \notin H$ then either $\Delta \notin H$ or $d + \Delta \notin H$, and so $F_\Delta(W)F_{d+\Delta}(W) \leq F_{\max}^{H|G}(W)$. (d) follows from Lemma 8.15.

On the other hand,

$$F(W^+[H]) = F(W^+[H|G]) \overset{(a)}{\leq} \frac{q \cdot |H|}{q} F_{\max}^{H|G}(W^+) = |H| \max_{\substack{d \in G, \\ d \notin H}} F_d(W^+)$$

$$\overset{(b)}{=} |H| \max_{\substack{d \in G, \\ d \notin H}} F_d(W)^2 = |H| F_{\max}^{H|G}(W)^2 \overset{(c)}{\leq} |H|(q - |H|)^2 F(W[H])^2,$$

where (a) follows from Lemma 8.5, (b) follows from Proposition 8.2 and (c) follows from Lemma 8.15.

8.9.6 Proof of Lemma 8.12

It is sufficient to show the following simpler version:

Lemma 8.16. *If $W : x \in G \longrightarrow \rho_x \in \mathcal{DM}(k \cdot r)$ is a cq-channel such that*

$$\rho_x^{BU} = \frac{1}{r} \sum_{u=1}^{r} \rho_{x,u}^B \otimes |u\rangle\langle u|^U,$$

where $\rho_{x,u}^B \in \mathcal{DM}(k)$ and $\{|u\rangle^U : 1 \leq u \leq r\}$ is an orthonormal basis of the Hilbert space of dimension r, then for every $1 \leq u \leq r$, there exists a POVM $\left\{\Pi_{u,x}^B : x \in G\right\}$ such that the POVM $\left\{\Pi_x^{BU} : x \in G\right\}$ defined as

$$\Pi_x^{BU} = \sum_{u=1}^{r} \Pi_{u,x}^B \otimes |u\rangle\langle u|^U,$$

satisfies

$$1 - \frac{1}{q} \sum_{x \in G} \mathrm{Tr}\left(\Pi_x^{BU} \rho_x^{BU}\right) < (q-1)F(W).$$

Proof. For every $1 \leq u \leq r$, define the cq-channel $W_u : x \in G \longrightarrow \rho_{x,u} \in \mathcal{DM}(k)$. The optimal decoder for W_u satisfies $P_e(W_u) \leq (q-1)F(W_u)$ [49]. Therefore, there exists a POVM $\left\{\Pi_{u,x}^B : x \in G\right\}$ satisfying,

$$1 - \frac{1}{q} \sum_{x \in G} \mathrm{Tr}\left(\Pi_{u,x}^B \rho_{u,x}^B\right) < (q-1)F(W_u).$$

For every $x \in G$, define

$$\Pi_x^{BU} = \sum_{u=1}^{r} \Pi_{u,x}^B \otimes |u\rangle\langle u|^U.$$

It is easy to see that $\left\{\Pi_x^{BU} : x \in G\right\}$ is a valid POVM. We have:

$$
\begin{aligned}
1 - \frac{1}{q}\sum_{x\in G} \mathrm{Tr}\left(\Pi_x^{BU}\rho_x^{BU}\right) & \\
&= 1 - \frac{1}{qr}\sum_{x\in G}\sum_{u=1}^{r} \mathrm{Tr}\left(\Pi_{u,x}^{B}\rho_{u,x}^{B}\right) = \frac{1}{r}\sum_{u=1}^{r}\left(1 - \frac{1}{q}\sum_{x\in G}\mathrm{Tr}\left(\Pi_{u,x}^{B}\rho_{u,x}^{B}\right)\right) \\
&\leq \frac{1}{r}\sum_{u=1}^{r}(q-1)F(W_u) = \frac{q-1}{r}\sum_{u=1}^{r}\sum_{\substack{x,x'\in G,\\ x\neq x'}} F(\rho_{u,x}^{B}, \rho_{u,x'}^{B}) \\
&= (q-1)\sum_{\substack{x,x'\in G,\\ x\neq x'}} F\left(\frac{1}{r}\sum_{u=1}^{r}\rho_{x,u}^{B}\otimes|u\rangle\langle u|^{U}, \frac{1}{r}\sum_{u=1}^{r}\rho_{x',u}^{B}\otimes|u\rangle\langle u|^{U}\right) \\
&= (q-1)F(W).
\end{aligned}
$$

□

9 Conclusion of Part I

In this chapter, we summarize the main contributions of the first part of this thesis. Furthermore, we briefly discuss some open problems and possible future directions in polarization theory.

9.1 Ergodic Theory of Binary Operations

In Chapter 2, we developed an ergodic theory for binary operations. This theory was applied in Chapter 3 to characterize the polarizing binary operations. The potential applications of the ergodic theory of binary operations might extend beyond polarization theory. The mathematical framework that is developed in chapter 2 is fairly general and might be useful to areas outside polarization and information theory.

As we saw in Chapter 2, a uniformity-preserving operation is ergodic (resp. irreducible, quasigroup operation) if and only if its right-inverse is ergodic (resp. irreducible, quasigroup operation). A natural question to ask is whether the strong ergodicity of a binary operation implies the strong ergodicity of its right-inverse. This question remains an open problem.

9.2 Polarizing Binary Operations

In Chapter 3, we provided a complete characterization of polarizing binary operations. We showed that a binary operation is polarizing if and only if it is uniformity-preserving and its right-inverse is strongly ergodic.

9.2.1 Structure of polarized channels

Let $*$ be a polarizing binary operation on a finite set $\mathcal{X}$, and let W be a channel with input alphabet $\mathcal{X}$. We showed in Chapter 3 that as the number n of polarization steps becomes large, the synthetic channels $(W^s)_{s\in\{-,+\}^n}$ polarize to channels that project their input onto a stable partition of $(\mathcal{X}, /^*)$. For every stable partition $\mathcal{H}$

of $(\mathcal{X}, /^*)$, define

$$\#_{\mathcal{H}}(W) = \lim_{\delta \to 0} \lim_{n \to \infty} \frac{1}{2^n} \Big| \Big\{ s \in \{-,+\}^n : \\ \big| I(W^s) - \log_2 |\mathcal{H}| \big| < \delta, \ \big| I(W^s[\mathcal{H}]) - \log_2 |\mathcal{H}| \big| < \delta \Big\} \Big|.$$

It is not difficult to show that the limit in the above equation exists. The quantity $\#_{\mathcal{H}}(W)$ represents the asymptotic fraction of polarized synthetic channels that project their input onto $\mathcal{H}$. Clearly,

$$\sum_{\substack{\mathcal{H} \text{ is a stable} \\ \text{partition of } (\mathcal{X},/^*)}} \#_{\mathcal{H}}(W) = 1.$$

One problem that remains open is to find a method[1] to compute $\#_{\mathcal{H}}(W)$ for an arbitrary channel W and an arbitrary stable partition $\mathcal{H}$ of $(\mathcal{X}, /^*)$.

9.2.2 General Arıkan-Style constructions

The Arıkan-style constructions that we considered in chapter 3 combine exactly two channels in one polarization step. In the following, we explain more general Arıkan-style constructions that can combine more than two channels in one polarization step.

An l-ary kernel on the set $\mathcal{X}$ is a mapping $f : \mathcal{X}^l \to \mathcal{X}^l$. For every $1 \leq i \leq l$, we denote the i^{th} component of $f(u_1, \ldots, u_l)$ as $f_i(u_1, \ldots, u_l)$, i.e.,

$$f(u_1, \ldots, u_l) = \big(f_1(u_1, \ldots, u_l), \ldots, f_l(u_1, \ldots, u_l) \big).$$

For every l-ary kernel f on $\mathcal{X}$, every $1 \leq i \leq l$ and every channel $W : \mathcal{X} \longrightarrow \mathcal{Y}$, define the channel $W^{(i)} : \mathcal{X} \longrightarrow \mathcal{Y}^l \times \mathcal{X}^{i-1}$ as follows:

$$W^{(i)}(y_1, \ldots, y_l, u_1, \ldots, u_{i-1} | u_i) \\ = \frac{1}{|\mathcal{X}|^{l-1}} \sum_{u_{i+1}, \ldots, u_l \in \mathcal{X}} W(y_1 | f_1(u_1, \ldots, u_l)) \times \cdots \times W(y_l | f_l(u_1, \ldots, u_l)).$$

For every $n \geq 1$ and every $s = (s_1, \ldots, s_n) \in \{1, \ldots, l\}^n$, define

$$W^s = (\ldots (W^{(s_1)})^{(s_2)} \ldots)^{(s_n)}.$$

An l-ary kernel f on $\mathcal{X}$ is said to be polarizing if it satisfies the following two properties:

- Conservation property: For every channel W with input alphabet $\mathcal{X}$, we have

$$\sum_{i=1}^{l} I(W^{(i)}) = l \cdot I(W).$$

[1] We seek a closed-form formula, or a low-complexity algorithm that can approximate $\#_{\mathcal{H}}(W)$ with arbitrary precision.

- Polarization property: For every channel W with input alphabet $\mathcal{X}$ and every $\delta > 0$, we have

$$\lim_{n\to\infty} \frac{1}{l^n} \big|\{s \in \{1,\dots,l\}^n :\ W^s \text{ is } \delta\text{-easy}\}\big| = 1.$$

It is easy to see that an l-ary kernel f satisfies the conservation property if and only if f is a bijection.

The results of Chapter 3 can be seen as a characterization of the polarizing 2-ary kernels of the form $f_*(u_1, u_2) = (u_1 * u_2, u_2)$ for some binary operation $*$ on $\mathcal{X}$. A characterization of polarizing *linear* kernels over finite fields is given in [57, 58]. Sufficient conditions for a non-linear kernel to be polarizing can be found in [59, 60].

One problem that remains open is to find a necessary and sufficient condition that characterizes all the polarizing kernels. A generalization of the ergodic theory of binary operations that we developed in Chapter 2 is likely to provide such a characterization.

9.3 MAC Polarization Theory

In Chapter 4, we showed that a sequence of binary operations is MAC-polarizing if and only if every operation in the sequence is uniformity-preserving and its right-inverse is strongly ergodic.

9.3.1 Region of Achievable Rate-Vectors

Let $W : \mathcal{X}_1 \times \ldots \times \mathcal{X}_m \longrightarrow \mathcal{Y}$ be an m-user MAC, and let $(*_1, \dots, *_m)$ be a MAC-polarizing sequence of binary operations. It is easy to see that the region of rate-vectors that are achievable by MAC-polar codes is given by

$$\mathcal{J}^{\text{pol}}(W) := \Big\{ R = (R_1, \dots, R_m) \in \mathbb{R}^m : \\ 0 \le R(S) \le I_S^{\text{pol}}(W) \ \text{ for all } \ S \subset \{1,\dots,m\} \Big\},$$

where $R(S) := \sum_{k\in S} R_k$, and $I_S^{\text{pol}}(W) = \lim_{n\to\infty} \frac{1}{2^n} \sum_{s\in\{-,+\}^n} I_S(W^s)$ (See Section 4.1 for the definition of $I_S(W^s)$).

An open problem in MAC polarization theory is to find a method to compute $\mathcal{J}^{\text{pol}}(W)$ for an arbitrary MAC W. In other words, we seek a method to compute $I_S^{\text{pol}}(W)$ for every $S \subset \{1,\dots,m\}$. It is not difficult to show that

$$I_S^{\text{pol}}(W) = \sum_{\substack{\mathcal{H} \text{ is a} \\ \text{stable partition of} \\ (\mathcal{X}, /^{*_1\otimes\cdots\otimes *_m})}} \#_{\mathcal{H}}(W) \cdot \log_2 |\mathcal{L}_S(\mathcal{H})|.$$

(See Notation 2.5 for the definition of $\mathcal{L}_S(\mathcal{H})$.)

We conclude that in order to compute $\mathcal{J}^{\text{pol}}(W)$, it is sufficient to solve the problem described in Section 9.2.1.

9.3.2 General Arıkan-Style constructions

We can generalize the Arıkan-style construction of Section 9.2.2 to multiple-access channels: Let $f_1, \ldots, f_m$ be l-ary kernels on $\mathcal{X}_1, \ldots, \mathcal{X}_m$, respectively. For every $1 \le k \le m$ and every $1 \le i \le l$, let $f_{k,i}$ be the i^{th} component of f_k, i.e.,

$$f_k(u_{k,1}, \ldots, u_{k,l}) = \big(f_{k,1}(u_{k,1}, \ldots, u_{k,l}), \ldots, f_{k,l}(u_{k,1}, \ldots, u_{k,l})\big).$$

For every $1 \le i \le l$ and every MAC $W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \longrightarrow \mathcal{Y}$, define the MAC $W^{(i)} : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \longrightarrow \mathcal{Y}^l \times (\mathcal{X}_1 \times \cdots \times \mathcal{X}_m)^{i-1}$ as follows:

$$\begin{aligned} W^{(i)}\big(y_1, \ldots, y_l, (u_{k,1})_{1\le k\le m}, \ldots, (u_{k,i-1})_{1\le k\le m} \big| u_{1,i}, \ldots, u_{m,i}\big) \\ = \frac{1}{|\mathcal{X}_1|^{l-1} \cdots |\mathcal{X}_m|^{l-1}} \sum_{\substack{u_{1,i+1},\ldots,u_{1,l} \in \mathcal{X}_1 \\ \vdots \\ u_{m,i+1},\ldots,u_{m,l} \in \mathcal{X}_m}} W\big(y_1 \big| f_{1,1}(u_{1,1}, \ldots, u_{1,l}), \ldots, f_{m,1}(u_{m,1}, \ldots, u_{m,l})\big) \\ \times \cdots \times W\big(y_l \big| f_{1,l}(u_{1,1}, \ldots, u_{1,l}), \ldots, f_{m,l}(u_{m,1}, \ldots, u_{m,l})\big). \end{aligned}$$

For every $n \ge 1$ and every $s = (s_1, \ldots, s_n) \in \{1, \ldots, l\}^n$, define

$$W^s = (\ldots(W^{(s_1)})^{(s_2)} \ldots)^{(s_n)}.$$

The sequence $(f_1, \ldots, f_l)$ is said to be MAC-polarizing if it satisfies the following two properties:

- Conservation property: For every m-user MAC W with input alphabets $\mathcal{X}_1, \ldots, \mathcal{X}_m$, we have
$$\sum_{i=1}^{l} I(W^{(i)}) = l \cdot I(W).$$
- Polarization property: For every m-user MAC W with input alphabets $\mathcal{X}_1, \ldots, \mathcal{X}_m$, and every $\delta > 0$, we have
$$\lim_{n\to\infty} \frac{1}{l^n} \big|\big\{s \in \{1, \ldots, l\}^n : \ W^s \text{ is } \delta\text{-easy}\big\}\big| = 1.$$

It is easy to see that the sequence $(f_1, \ldots, f_m)$ satisfies the conservation property if and only if $f_1, \ldots, f_m$ are bijections.

An open problem in MAC polarization theory is to find a necessary and sufficient condition for a sequence of l-ary kernels to be MAC-polarizing.

9.4 Error Exponents

In Chapter 5, we showed that the exponent E_* of a polarizing binary operation $*$ cannot exceed $\frac{1}{2}$. We proved that if $*$ is a quasigroup operation, then $E_* = \frac{1}{2}$. We conjectured that $E_* < \frac{1}{2}$ if $*$ is not a quasigroup operation. Finding a closed-form formula for E_* is an open problem.

If we wish to construct polar codes with an exponent that is strictly better than $\frac{1}{2}$, we have to use Arıkan-style constructions that are not based on binary operations. Korada et. al. showed that it is possible to achieve exponents that exceed $\frac{1}{2}$ by using linear l-ary kernels [57].

Presman et. al. showed that nonlinear kernels can achieve strictly better exponents compared to all linear kernels [60]. A non-linear kernel is said to be *excellent* if it outperforms all the linear kernels of the same size. One drawback of the excellent kernels of [60] is that they have a large size (i.e., a large arity). Is it possible to find excellent kernels of small size? Finding a necessary and sufficient condition for a non-linear kernel to be excellent is an open problem.

9.5 Fourier Analysis of MAC Polarization

In Chapter 6, we characterized all the MACs W that do not lose any part of their symmetric-capacity region by polarization (i.e., $\mathcal{J}^{\text{pol}}(W) = \mathcal{J}(W)$). The necessary and sufficient condition that we provided is a single-letter characterization: The mapping $\hat{f}_W$ can be directly computed using the transition probabilities of W. Moreover, since the number of pseudo-quadratic functions is finite, checking whether $\hat{f}_W$ is extendable to a pseudo-quadratic function can be accomplished in a finite number of computations.

The characterization that we provided works in the setting where we use an Abelian group operation on the input alphabet of each user. Generalizing the results of Chapter 6 to arbitrary MAC-polarizing sequences of binary operations remains an open problem. Following a similar approach to that of Chapter 6 might solve the problem in the case of non-Abelian groups because there is a notion of discrete Fourier transforms on these groups. A completely different approach might be needed to solve the problem in the general case of an arbitrary MAC-polarizing sequence of binary operations.

9.6 Erasure Schemes Using Generalized Polar Codes

In Chapter 7, we studied the erasure schemes that use generalized polar (GP) codes. We provided a closed-form formula for the zero-undetected-error capacity $I_0^{\text{GP}}(W)$ of GP codes for a given binary-input memoryless symmetric channel W under the low-complexity successive cancellation decoder with erasure. We showed that for every $R < I_0^{\text{GP}}(W)$, there exists a GP code of blocklength N and of rate at least R where the undetected-error probability is zero and the erasure probability is less than $2^{-N^{\frac{1}{2}-\epsilon}}$. Conversely, we showed that for any GP code of rate $I_0^{\text{GP}}(W) < R < I(W)$ and blocklength N, the undetected-error probability cannot be made less than $2^{-N^{\frac{1}{2}+\epsilon}}$ unless the erasure probability is close to 1.

The tradeoff that we obtained between the undetected-error probability and the erasure probability for rates $R > I_0^{\text{GP}}(W)$ is very sharp and does not depend on the rate R. A more refined estimation of the tradeoff between p_{ue} and p_{er}, which explicitly depends on R, remains an open problem.

Another problem that remains open is to generalize the results of Chapter 7 to channels with arbitrary input alphabet, and Arıkan-style constructions that are based on arbitrary polarizing operations.

9.7 Polar Codes for Arbitrary Classical-Quantum Channels

In Chapter 8, we showed that using an Arıkan-style construction that is based on an Abelian group operation yields multilevel polarization for arbitrary classical-quantum channels (in a similar way as in the case of classical channels). This result made it possible to construct polar codes for arbitrary cq-channels and arbitrary cq-MACs.

One weakness of the results presented in Chapter 8 is that the proposed quantum successive cancellation decoder does not seem to have an efficient implementation. This was also the case for the polar codes that were constructed for binary-input cq-channels [43]. Finding an efficient decoder for the polar codes remains an open problem.

If we define cq-polarizing binary operations as those that can polarize an arbitrary cq-channel to "easy" cq-channels (in a sense similar to Definitions 3.1 and 3.4), then Chapter 8 shows that Abelian group operations are cq-polarizing. Therefore, being an Abelian group operation is a sufficient condition to be cq-polarizing. On the other hand, from the results of Chapter 3 we can deduce that being uniformity-preserving and having a right-inverse that is strongly ergodic are necessary conditions because classical channels are particular cq-channels. Finding a necessary and sufficient condition for a binary operation to be cq-polarizing remains an open problem. Trying to prove a quantum version of the results in Chapter 3 by using a similar approach might not be successful because the proof of the sufficient condition relies heavily on the entropy of the input conditioned on a particular output symbol, and this does not have an analog in the case of cq-channels.

We also showed that cq-MAC polarization can induce a loss in the symmetric capacity region. A necessary and sufficient condition for $\mathcal{J}^{\mathrm{pol}}(W) = \mathcal{J}(W)$ in the case of classical MACs was given in Chapter 6. Generalizing the results of Chapter 6 to cq-MACs is an open problem. Recall that the condition in Chapter 6 was given in terms of the Fourier transform of the probability distribution of one input conditioned on the output and on the other input. Since this conditional probability does not have an analog in the case of cq-MACs, generalizing the results of Chapter 6 to cq-MACs might be challenging, and a completely different approach might be needed.

Part II

Channel Ordering

Characterizations of Various Channel Orderings

10

In this chapter[1], we provide several characterizations for various channel orderings. In Section 10.1, we provide the preliminaries of this chapter. In Section 10.2, we recall known properties of the output-degradedness ordering. In Section 10.3, we introduce the input-degradedness ordering of communication channels. We show that if W is input-degraded from another channel W', then any decoder that is good for W is also good for W'. We provide two characterizations for input-degradedness, one of which is similar to the Blackwell-Sherman-Stein (BSS) theorem. In Section 10.4, we study the Shannon ordering of communication channels. We show that W' contains W if and only if W is the skew-composition of W' with a convex-product channel. We use this fact to derive a characterization of the Shannon ordering that is similar to the BSS theorem.

10.1 Preliminaries

10.1.1 Set-Theoretic Notations

For every integer $n > 0$, we denote the set $\{1, \ldots, n\}$ as $[n]$.

The set of mappings from a set A to a set B is denoted as B^A.

Let A be a subset of B. The *indicator mapping* $\mathbb{1}_{A,B} : B \to \{0, 1\}$ of A in B is defined as:

$$\mathbb{1}_{A,B}(x) = \mathbb{1}_{x \in A} = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

If the superset B is clear from the context, we write $\mathbb{1}_A$ to denote the indicator mapping of A in B.

The *power set* of B is the set of subsets of B. Since every subset of B can be identified with its indicator mapping, we denote the power set of B as $2^B := \{0, 1\}^B$.

[1]The material of this chapter is based on [61, 62, 63, 64, 65, 66].

10.1.2 Probability Measures

If $\mathcal{A} \subset 2^M$ is a collection of subsets of M, we denote the σ-algebra that is *generated* by $\mathcal{A}$ as $\sigma(\mathcal{A})$.

The set of probability measures on (M, Σ) is denoted as $\mathcal{P}(M, \Sigma)$. If the σ-algebra Σ is known from the context, we write $\mathcal{P}(M)$ to denote the set of probability measures.

If $P \in \mathcal{P}(M, \Sigma)$ and $\{x\}$ is a measurable singleton, we write $P(x)$ to denote $P(\{x\})$.

For every $P_1, P_2 \in \mathcal{P}(M, \Sigma)$, the *total variation distance* between P_1 and P_2 is defined as:

$$\|P_1 - P_2\|_{TV} = \sup_{A \in \Sigma} |P_1(A) - P_2(A)|.$$

The space $\mathcal{P}(M, \Sigma)$ is a complete metric space under the total variation distance.

10.1.3 Probabilities on Finite Sets

We always endow finite sets with their finest σ-algebra, i.e., the power set. In this case, every probability measure is completely determined by its value on singletons, i.e., if P is a measure on a finite set $\mathcal{X}$, then for every $A \subset \mathcal{X}$, we have

$$P(A) = \sum_{x \in A} P(x).$$

If $\mathcal{X}$ is a finite set, we denote the set of probability distributions on $\mathcal{X}$ as $\Delta_{\mathcal{X}}$. Note that $\Delta_{\mathcal{X}}$ is an $(|\mathcal{X}| - 1)$-dimensional simplex in $\mathbb{R}^{\mathcal{X}}$.

10.1.4 Meta-Probability Measures

Let $\mathcal{X}$ be a finite set and let $\Delta_{\mathcal{X}}$ be the set of probability measures on $\mathcal{X}$. A *meta-probability measure* on $\mathcal{X}$ is a probability measure on the Borel sets of $\Delta_{\mathcal{X}}$. It is called a meta-probability measure because it is a probability measure on the space of probability distributions on $\mathcal{X}$.

We denote the set of meta-probability measures on $\mathcal{X}$ as $\mathcal{MP}(\mathcal{X})$. Clearly, $\mathcal{MP}(\mathcal{X}) = \mathcal{P}(\Delta_{\mathcal{X}})$.

A meta-probability measure MP on $\mathcal{X}$ is said to be *balanced* if it satisfies

$$\int_{\Delta_{\mathcal{X}}} p \cdot d\mathrm{MP}(p) = \pi_{\mathcal{X}},$$

where $\pi_{\mathcal{X}}$ is the uniform probability distributions on $\mathcal{X}$.

A meta-probability measure MP on $\mathcal{X}$ is said to be *finitely supported* if there exists a finite subset A of $\Delta_{\mathcal{X}}$ such that $\mathrm{MP}(A) = 1$. In this case, the support of MP is defined as:

$$\mathrm{supp}(\mathrm{MP}) = \{p \in \Delta_{\mathcal{X}} : \mathrm{MP}(p) > 0\}.$$

We denote the set of all balanced meta-probability measures on $\mathcal{X}$ as $\mathcal{MP}_b(\mathcal{X})$. The set of all balanced and finitely supported meta-probability measures on $\mathcal{X}$ is denoted as $\mathcal{MP}_{bf}(\mathcal{X})$.

10.1.5 Convex-Extreme Points

Let $\mathcal{X}$ be a finite set. For every $A \subset \Delta_{\mathcal{X}}$, let $\mathrm{co}(A)$ be the convex hull of A. We say that $p \in A$ is *convex-extreme* if it is an extreme point of $\mathrm{co}(A)$, i.e., for every $p_1, \ldots, p_n \in \mathrm{co}(A)$ and every $\lambda_1, \ldots, \lambda_n > 0$ satisfying $\sum_{i=1}^{n} \lambda_i = 1$ and $\sum_{i=1}^{n} \lambda_i p_i = p$, we have $p_1 = \ldots = p_n = p$. It is easy to see that if A is finite, then the convex-extreme points of A coincide with the extreme points of $\mathrm{co}(A)$. We denote the set of convex-extreme points of A as $\mathrm{CE}(A)$.

10.1.6 The Space of Channels from $\mathcal{X}$ to $\mathcal{Y}$

Let $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ be the set of all channels having $\mathcal{X}$ as input alphabet and $\mathcal{Y}$ as output alphabet.

If $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ and $V \in \mathrm{DMC}_{\mathcal{Y},\mathcal{Z}}$, we define the composition $V \circ W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Z}}$ of W and V as follows:

$$(V \circ W)(z|x) = \sum_{y \in \mathcal{Y}} V(z|y) W(y|x), \quad \forall x \in \mathcal{X},\ \forall z \in \mathcal{Z}.$$

It is easy to see that the mapping $(W, V) \to V \circ W$ from $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \times \mathrm{DMC}_{\mathcal{Y},\mathcal{Z}}$ to $\mathrm{DMC}_{\mathcal{X},\mathcal{Z}}$ is continuous.

For every mapping $f : \mathcal{X} \to \mathcal{Y}$, define the *deterministic channel* $D_f \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ as follows:

$$D_f(y|x) = \begin{cases} 1 & \text{if } y = f(x), \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that if $f : \mathcal{X} \to \mathcal{Y}$ and $g : \mathcal{Y} \to \mathcal{Z}$, then $D_g \circ D_f = D_{g \circ f}$.

For every two channels $W_1 \in \mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1}$ and $W_2 \in \mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}_2}$, define the *channel product* $W_1 \otimes W_2 \in \mathrm{DMC}_{\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{Y}_1 \times \mathcal{Y}_2}$ of W_1 and W_2 as:

$$(W_1 \otimes W_2)(y_1, y_2|x_1, x_2) = W_1(y_1|x_1) W_2(y_2|x_2).$$

$W_1 \otimes W_2$ arises when the transmitter has two channels W_1 and W_2 at his disposal and he uses both of them at each channel use. Channel products were first introduced by Shannon in [67].

10.2 Output-Degradedness and Output-Equivalence

Let $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ and $W' \in \mathrm{DMC}_{\mathcal{X},\mathcal{Z}}$ be two channels having the same input alphabet. We say that W' is *output-degraded* from W if there exists a channel $V \in \mathrm{DMC}_{\mathcal{Y},\mathcal{Z}}$ such that $W' = V \circ W$. W and W' are said to be *output-equivalent* if each one is output-degraded from the other. In the rest of this section, we describe one way to check whether two given channels are output-equivalent.

Let $\Delta_{\mathcal{X}}$ and $\Delta_{\mathcal{Y}}$ be the space of probability distributions on $\mathcal{X}$ and $\mathcal{Y}$ respectively. Define $P_W^o \in \Delta_{\mathcal{Y}}$ as

$$P_W^o(y) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} W(y|x), \quad \forall y \in \mathcal{Y}.$$

This can be interpreted as the probability distribution of the output when the input is uniformly distributed in $\mathcal{X}$. The *image* of W is the set of output-symbols $y \in \mathcal{Y}$ having strictly positive probabilities:

$$\mathrm{Im}(W) = \{y \in \mathcal{Y} : \ P_W^o(y) > 0\}.$$

For every $y \in \mathrm{Im}(W)$, define $W_y^{-1} \in \Delta_{\mathcal{X}}$ as follows:

$$W_y^{-1}(x) = \frac{W(y|x)}{|\mathcal{X}| P_W^o(y)}, \ \ \forall x \in \mathcal{X}.$$

$W_y^{-1}(x)$ can be interpreted as the posterior probability of x, given that the output is y, and assuming a uniform prior distribution on the input. In other words, if X is a random variable uniformly distributed in $\mathcal{X}$ and Y is the output of the channel W when X is the input, then:

- $P_W^o(y) = P_Y(y)$ for every $y \in \mathcal{Y}$.
- $W_y^{-1}(x) = P_{X|Y}(x|y)$ for every $(x, y) \in \mathcal{X} \times \mathrm{Im}(W)$.

Let $(x, y) \in \mathcal{X} \times \mathcal{Y}$. If $P_W^o(y) = P_Y(y) > 0$, we have

$$W(y|x) = P_{Y|X}(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)} = |\mathcal{X}| P_Y(y) P_{X|Y}(x|y) = |\mathcal{X}| P_W^o(y) W_y^{-1}(x).$$

On the other hand, if $P_W^o(y) = 0$, then we must have $W(y|x) = 0$. We conclude that P_W^o and the collection $\{W_y^{-1}\}_{y \in \mathrm{Im}(W)}$ uniquely determine W.

The *Blackwell measure*[2] (denoted MP_W) of W is a probability distribution on $\Delta_{\mathcal{X}}$ having masses $P_W^o(y)$ on W_y^{-1} for each $y \in \mathrm{Im}(W)$:

$$\mathrm{MP}_W(B) = \sum_{\substack{y \in \mathrm{Im}(W), \\ W_y^{-1} \in B}} P_W^o(y), \ \ \forall B \in \mathcal{B}(\Delta_{\mathcal{X}}).$$

Another way to express MP_W is as follows:

$$\mathrm{MP}_W = \sum_{y \in \mathrm{Im}(W)} P_W^o(y) \cdot \delta_{W_y^{-1}},$$

where $\delta_{W_y^{-1}}$ is a Dirac measure centered at $W_y^{-1} \in \Delta_{\mathcal{X}}$.

MP_W can be interpreted as follows: After the receiver obtains the output of the channel, he can compute the posterior probabilities of the input as the conditional probability distribution of the input given the output symbol that he received. But before receiving the output symbol, the receiver does not know what he we will receive. He just has different probabilities for different possible output symbols. Therefore, the posterior probability distribution that will be computed by the receiver is itself random, and so we need a meta-probability measure to describe it. MP_W is exactly this meta-probability measure.

[2] In an earlier version of this work, I called MP_W the *posterior meta-probability distribution* of W. Maxim Raginsky thankfully brought to my attention the fact that MP_W is called *Blackwell measure*.

Since $\mathrm{Im}(W)$ is finite, the support of MP_W is finite and it consists of all points in $\Delta_{\mathcal{X}}$ having strictly positive mass:

$$\mathrm{supp}(\mathrm{MP}_W) = \{p \in \Delta_{\mathcal{X}} : \ \mathrm{MP}_W(p) > 0\}.$$

The *rank* of W is the size of the support of its Blackwell measure:

$$\mathrm{rank}(W) = |\,\mathrm{supp}(\mathrm{MP}_W)|.$$

Notice that for every $x \in \mathcal{X}$, we have

$$\begin{aligned}\int_{\Delta_{\mathcal{X}}} p(x) \cdot d\mathrm{MP}_W(p) &= \sum_{p \in \mathrm{supp}(\mathrm{MP}_W)} \mathrm{MP}_W(p) \cdot p(x) = \sum_{y \in \mathrm{Im}(W)} P_W^o(y) W_y^{-1}(x) \\ &= \sum_{y \in \mathrm{Im}(W)} \frac{1}{|\mathcal{X}|} W(y|x) \overset{(a)}{=} \sum_{y \in \mathcal{Y}} \frac{1}{|\mathcal{X}|} W(y|x) = \frac{1}{|\mathcal{X}|},\end{aligned}$$

where (a) follows from the fact that $W(y|x) = 0$ for every $y \notin \mathrm{Im}(W)$. Therefore, we can write

$$\int_{\Delta_{\mathcal{X}}} p \cdot d\mathrm{MP}_W(p) = \pi_{\mathcal{X}}, \tag{10.1}$$

where $\pi_{\mathcal{X}}$ is the uniform probability distribution on $\mathcal{X}$. This shows that MP_W is a balanced meta-probability measure.

The following proposition characterizes the Blackwell measures of DMCs with input alphabet $\mathcal{X}$:

Proposition 10.1. *[68] A meta-probability measure* MP *on* $\mathcal{X}$ *is the Blackwell measure of some DMC with input alphabet* $\mathcal{X}$ *if and only if* MP *is balanced and finitely supported.*

Proof. This proposition is known [68], but we provide a proof for the sake of completeness.

The above discussion shows that if MP is the Blackwell measure of some channel with input alphabet $\mathcal{X}$, then it is balanced and finitely supported.

Now assume that MP is balanced and finitely supported, and let $\mathcal{Y} = \mathrm{supp}(\mathrm{MP})$. Define the channel $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ as $W(p|x) = |\mathcal{X}|\mathrm{MP}(p)p(x)$ for every $x \in \mathcal{X}$ and every $p \in \mathcal{Y} = \mathrm{supp}(\mathrm{MP})$. For every $x \in \mathcal{X}$, we have:

$$\sum_{p \in \mathcal{Y}} W(p|x) = \sum_{p \in \mathrm{supp}(\mathrm{MP})} |\mathcal{X}|p(x)\mathrm{MP}(p) = |\mathcal{X}| \int_{\Delta_{\mathcal{X}}} p(x) \cdot d\mathrm{MP}(p) = |\mathcal{X}|\pi_{\mathcal{X}}(x) = 1.$$

Therefore, W is a valid channel. For every $p \in \mathcal{Y}$, we have

$$\begin{aligned}P_W^o(p) &= \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} W(p|x) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} |\mathcal{X}|p(x)\mathrm{MP}(p) \\ &= \sum_{x \in \mathcal{X}} p(x)\mathrm{MP}(p) = \mathrm{MP}(p) > 0,\end{aligned}$$

which implies that $\mathrm{Im}(W) = \mathcal{Y}$. For every $(x, p) \in \mathcal{X} \times \mathcal{Y}$ we have:

$$W_p^{-1}(x) = \frac{W(p|x)}{|\mathcal{X}|P_W^o(p)} = \frac{|\mathcal{X}|\mathrm{MP}(p)p(x)}{|\mathcal{X}|\,\mathrm{MP}(p)} = p(x).$$

Therefore, $W_p^{-1} = p$ for every $p \in \mathcal{Y}$. For every Borel subset B of $\Delta_{\mathcal{X}}$, we have:

$$\mathrm{MP}_W(B) = \sum_{\substack{p \in \mathrm{Im}(W),\\ W_p^{-1} \in B}} P_W^o(p) = \sum_{\substack{p \in \mathrm{supp}(\mathrm{MP}),\\ p \in B}} \mathrm{MP}(p) = \mathrm{MP}(B).$$

We conclude that $\mathrm{MP}_W = \mathrm{MP}$. □

In [69], equivalent representations for binary memoryless symmetric (BMS) channels (namely L, D and G densities) were provided. A necessary and sufficient condition for the output-degradation of a BMS channel W' with respect to another BMS channel W was given in [69] in terms of the $|D|$-densities of W and W'. It immediately follows from this condition that two BMS channels are output-equivalent if and only if they have the same $|D|$-densities. One can deduce from this that two BMS channels (with finite output alphabets) are output-equivalent if and only if they have the same Blackwell measure. The following proposition shows that this is also true for channels with arbitrary (but finite) input and output alphabets:

Proposition 10.2. *[68] Let $\mathcal{X}, \mathcal{Y}$ and $\mathcal{Z}$ be three finite sets. Two channels $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ and $W' \in \mathrm{DMC}_{\mathcal{X},\mathcal{Z}}$ are output-equivalent if and only if $\mathrm{MP}_W = \mathrm{MP}_{W'}$.*

Proof. This proposition is known [68], but we provide a proof in Appendix 10.5.1 for the sake of completeness. □

Corollary 10.1. *If $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ and $\mathrm{rank}(W) > |\mathcal{Z}|$, then W is not output-equivalent to any channel in $\mathrm{DMC}_{\mathcal{X},\mathcal{Z}}$.*

Proof. Since $\mathrm{rank}(W') = |\mathrm{supp}(\mathrm{MP}_{W'})| \leq |\mathcal{Z}|$ for every $W' \in \mathrm{DMC}_{\mathcal{X},\mathcal{Z}}$, it is impossible for W to be output-equivalent to any channel W' in $\mathrm{DMC}_{\mathcal{X},\mathcal{Z}}$. □

Corollary 10.2. *If $|\mathcal{X}| = 1$, all channels with input alphabet $\mathcal{X}$ are output-equivalent.*

10.3 Input-Degradedness and Input-Equivalence

Let $\mathcal{X}, \mathcal{X}'$ and $\mathcal{Y}$ be three finite sets. Let $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ and $W' \in \mathrm{DMC}_{\mathcal{X}',\mathcal{Y}}$. We say that W is *input-degraded* from W' if there exists a channel $V' \in \mathrm{DMC}_{\mathcal{X},\mathcal{X}'}$ such that $W = W' \circ V'$. The channels W and W' are said to be *input-equivalent* if each one is input-degraded from the other.

Let $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ be a fixed channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. For every $x \in \mathcal{X}$, define $W_x \in \Delta_{\mathcal{Y}}$ as:

$$W_x(y) = W(y|x), \quad \forall y \in \mathcal{Y}.$$

Proposition 10.3. *Let $\mathcal{X}', \mathcal{X}$ and $\mathcal{Y}$ be three finite sets. $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ is input-degraded from $W' \in \mathrm{DMC}_{\mathcal{X}',\mathcal{Y}}$ if and only if $\mathrm{co}(\{W_x : x \in \mathcal{X}\}) \subset \mathrm{co}(\{W'_{x'} : x' \in \mathcal{X}'\})$.*

Proof. Assume that W is input-degraded from W'. There exists $V' \in \mathrm{DMC}_{\mathcal{X},\mathcal{X}'}$ such that $W = W' \circ V'$. For every $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, we have:

$$W_x(y) = W(y|x) = \sum_{x' \in \mathcal{X}'} W'(y|x')V'(x'|x) = \sum_{x' \in \mathcal{X}'} V'(x'|x)W'_{x'}(y).$$

Therefore, $W_x = \sum_{x' \in \mathcal{X}'} V'(x'|x) W'_{x'}$ which means that $W_x \in \mathrm{co}(\{W'_{x'} : x' \in \mathcal{X}'\})$ for every $x \in \mathcal{X}$, hence $\mathrm{co}(\{W_x : x \in \mathcal{X}\}) \subset \mathrm{co}(\{W'_{x'} : x' \in \mathcal{X}'\})$.

Conversely, assume that $\mathrm{co}(\{W_x : x \in \mathcal{X}\}) \subset \mathrm{co}(\{W'_{x'} : x' \in \mathcal{X}'\})$ and let $x \in \mathcal{X}$. Since $W_x \in \mathrm{co}(\{W'_{x'} : x' \in \mathcal{X}'\})$, there exists a set of numbers $\alpha_{x,x'} \geq 0$ satisfying $\sum_{x' \in \mathcal{X}'} \alpha_{x,x'} = 1$ such that $W_x = \sum_{x' \in \mathcal{X}'} \alpha_{x,x'} W_{x'}$. Define $V' \in \mathrm{DMC}_{\mathcal{X},\mathcal{X}'}$ as $V(x'|x) = \alpha_{x,x'}$ for every $x \in \mathcal{X}$ and every $x' \in \mathcal{X}'$. We have $W = W' \circ V'$ and so W is input-degraded from W'. □

For every channel $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, we define the *input-equivalence characteristic of* W, or simply the *characteristic of* W, as $\mathrm{CE}(W) := \mathrm{CE}(\{W_x : x \in \mathcal{X}\})$. The *input-rank* of $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ is the size of its characteristic: $\mathrm{irank}(W) = |\mathrm{CE}(W)|$.

Proposition 10.4. *Let $\mathcal{X}', \mathcal{X}$ and $\mathcal{Y}$ be three finite sets. $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ is input-equivalent to $W' \in \mathrm{DMC}_{\mathcal{X}',\mathcal{Y}}$ if and only if $\mathrm{CE}(W) = \mathrm{CE}(W')$.*

Proof. It follows from Proposition 10.3 that W is input-equivalent to W' if and only if $\mathrm{co}(\{W_x : x \in \mathcal{X}\}) = \mathrm{co}(\{W'_{x'} : x' \in \mathcal{X}'\})$, which happens if and only if $\mathrm{CE}(W) = \mathrm{CE}(\mathrm{co}(\{W_x : x \in \mathcal{X}\})) = \mathrm{CE}(\mathrm{co}(\{W'_{x'} : x' \in \mathcal{X}'\})) = \mathrm{CE}(W')$. □

10.3.1 Operational Implication in Terms of Decoders

Let $\mathcal{Y}$ be a finite set. An (n, M)-decoder on $\mathcal{Y}$ is a mapping $\mathcal{D} : \mathcal{Y}^n \to \mathcal{M}$, where $|\mathcal{M}| = M$. The set $\mathcal{M}$ is the *message set* of $\mathcal{D}$, n is the *blocklength* of $\mathcal{D}$, M is the *size* of $\mathcal{D}$ and $\frac{1}{n} \log_2 |\mathcal{M}|$ is the *rate* of $\mathcal{D}$.

Let $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ be a channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, and let $\mathcal{D} : \mathcal{Y}^n \to \mathcal{M}$ be a decoder on $\mathcal{Y}$. A *maximum-likelihood (ML) encoder* for $\mathcal{D}$ when it is used for W is any encoder $\mathcal{E} : \mathcal{M} \to \mathcal{X}^n$ satisfying

$$\sum_{\substack{y_1^n \in \mathcal{Y}^n: \\ \mathcal{D}(y_1^n)=m}} \prod_{i=1}^{n} W(y_i|\mathcal{E}_i(m)) \geq \sum_{\substack{y_1^n \in \mathcal{Y}^n: \\ \mathcal{D}(y_1^n)=m}} \prod_{i=1}^{n} W(y_i|x_i), \quad \forall m \in \mathcal{M},\ \forall x_1^n \in \mathcal{X}^n,$$

where $(\mathcal{E}_1(m), \ldots, \mathcal{E}_n(m)) = \mathcal{E}(m) \in \mathcal{X}^n$.

It is easy to see that a maximum-likelihood encoder has the best probability of error among all encoders (assuming that the decoder $\mathcal{D}$ is used). The *probability of error of $\mathcal{D}$ under ML-encoding for W* is given by:

$$P_{e,\mathcal{D}}(W) = 1 - \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \max_{x_1^n \in \mathcal{X}^n} \left\{ \sum_{\substack{y_1^n \in \mathcal{Y}^n: \\ \mathcal{D}(y_1^n)=m}} \prod_{i=1}^{n} W(y_i|x_i) \right\}.$$

Proposition 10.5. *Let $\mathcal{X}', \mathcal{X}$ and $\mathcal{Y}$ be three finite sets. If $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ is input-degraded from $W' \in \mathrm{DMC}_{\mathcal{X}',\mathcal{Y}}$, then $P_{e,\mathcal{D}}(W') \leq P_{e,\mathcal{D}}(W)$ for every decoder $\mathcal{D}$ on $\mathcal{Y}$. Moreover, if W and W' are input-equivalent, then $P_{e,\mathcal{D}}(W) = P_{e,\mathcal{D}}(W')$ for every decoder $\mathcal{D}$ on $\mathcal{Y}$.*

Proof. Assume that $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ is input-degraded from $W' \in \mathrm{DMC}_{\mathcal{X}',\mathcal{Y}}$. Let $V' \in \mathrm{DMC}_{\mathcal{X},\mathcal{X}'}$ be such that $W = W' \circ V'$.

Fix an (n, M) decoder $\mathcal{D}$ on $\mathcal{Y}$ and let $\mathcal{M}$ be its message set. We have:

$$\begin{aligned}
1 - P_{e,\mathcal{D}}(W) &= \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \max_{x_1^n \in \mathcal{X}^n} \left\{ \sum_{\substack{y_1^n\in\mathcal{Y}^n:\\ \mathcal{D}(y_1^n)=m}} \prod_{i=1}^n W(y_i|x_i) \right\} \\
&= \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \max_{x_1^n \in \mathcal{X}^n} \left\{ \sum_{\substack{y_1^n\in\mathcal{Y}^n:\\ \mathcal{D}(y_1^n)=m}} \prod_{i=1}^n \left(\sum_{x_i'\in\mathcal{X}'} W'(y_i|x_i')V'(x_i'|x_i) \right) \right\} \\
&= \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \max_{x_1^n \in \mathcal{X}^n} \left\{ \sum_{\substack{y_1^n\in\mathcal{Y}^n:\\ \mathcal{D}(y_1^n)=m}} \sum_{x_1'^n\in\mathcal{X}'^n} \prod_{i=1}^n \left(W'(y_i|x_i')V'(x_i'|x_i) \right) \right\} \\
&= \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \max_{x_1^n \in \mathcal{X}^n} \left\{ \sum_{x_1'^n\in\mathcal{X}'^n} \left(\prod_{i=1}^n V'(x_i'|x_i) \right) \sum_{\substack{y_1^n\in\mathcal{Y}^n:\\ \mathcal{D}(y_1^n)=m}} \prod_{i=1}^n W'(y_i|x_i') \right\} \\
&\le \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \max_{x_1'^n \in \mathcal{X}'^n} \left\{ \sum_{\substack{y_1^n\in\mathcal{Y}^n:\\ \mathcal{D}(y_1^n)=m}} \prod_{i=1}^n W'(y_i|x_i') \right\} = 1 - P_{e,\mathcal{D}}(W').
\end{aligned}$$

Therefore, $P_{e,\mathcal{D}}(W') \le P_{e,\mathcal{D}}(W)$.

If W and W' are input-degraded from each other, then $P_{e,\mathcal{D}}(W') \le P_{e,\mathcal{D}}(W)$ and $P_{e,\mathcal{D}}(W) \le P_{e,\mathcal{D}}(W')$, hence $P_{e,\mathcal{D}}(W') = P_{e,\mathcal{D}}(W)$. □

10.3.2 A Characterization of Input-Degradedness

Let $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ and let $\mathcal{U}$ be a finite set. For every $p \in \Delta_{\mathcal{U}}$ and every $D \in \mathrm{DMC}_{\mathcal{Y},\mathcal{U}}$, define

$$P_c(p, W, D) = \sup_{E\in\mathrm{DMC}_{\mathcal{U},\mathcal{X}}} \sum_{\substack{u\in\mathcal{U},\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} p(u)E(x|u)W(y|x)D(u|y).$$

$P_c(p, W, D)$ can be interpreted as follows: Let U be a random variable in $\mathcal{U}$ distributed as p. Assume that U was encoded using the random encoder $E \in \mathrm{DMC}_{\mathcal{U},\mathcal{X}}$ to get $X \in \mathcal{X}$. Send X through the channel W and let $Y \in \mathcal{Y}$ be the output. Apply the random decoder $D \in \mathrm{DMC}_{\mathcal{Y},\mathcal{U}}$ on $\mathcal{Y}$ to get an estimate $\hat{U}$ of U. We have:

$$\mathbb{P}[\{\hat{U} = U\}] = \sum_{\substack{u\in\mathcal{U},\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} p(u)E(x|u)W(y|x)D(u|y).$$

Therefore, $P_c(p, W, D)$ is the optimal probability of successfully estimating U by the fixed decoder D among all random encoders $E \in \mathrm{DMC}_{\mathcal{U},\mathcal{X}}$. Note that the optimal encoder can always be chosen to be deterministic.

The following theorem provides a characterization of input-degradedness that is somewhat similar to the characterization of output-degradedness given in [70].

Theorem 10.1. *A channel $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ is input-degraded from another channel $W' \in \mathrm{DMC}_{\mathcal{X}',\mathcal{Y}}$ if and only if $P_c(p,W,D) \leq P_c(p,W',D)$ for every $p \in \Delta_{\mathcal{U}}$, every $D \in \mathrm{DMC}_{\mathcal{Y},\mathcal{U}}$ and every finite set $\mathcal{U}$.*

Proof. Assume that W is input-degraded from W'. There exists $V' \in \mathrm{DMC}_{\mathcal{X},\mathcal{X}'}$ such that $W = W' \circ V'$. For every finite set $\mathcal{U}$, every $p \in \Delta_{\mathcal{U}}$ and every $D \in \mathrm{DMC}_{\mathcal{Y},\mathcal{U}}$, we have:

$$\begin{aligned}
P_c(p,W,D) &= \sup_{E \in \mathrm{DMC}_{\mathcal{U},\mathcal{X}}} \sum_{\substack{u\in\mathcal{U},\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} p(u)E(x|u)W(y|x)D(u|y) \\
&= \sup_{E \in \mathrm{DMC}_{\mathcal{U},\mathcal{X}}} \sum_{\substack{u\in\mathcal{U},\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} p(u)E(x|u)\left(\sum_{x'\in\mathcal{X}'} W'(y|x')V'(x'|x)\right)D(u|y) \\
&= \sup_{E \in \mathrm{DMC}_{\mathcal{U},\mathcal{X}}} \sum_{\substack{u\in\mathcal{U},\\ x'\in\mathcal{X}',\\ y\in\mathcal{Y}}} p(u)\left(\sum_{x\in\mathcal{X}} V'(x'|x)E(x|u)\right)W'(y|x')D(u|y) \\
&= \sup_{E \in \mathrm{DMC}_{\mathcal{U},\mathcal{X}}} \sum_{\substack{u\in\mathcal{U},\\ x'\in\mathcal{X}',\\ y\in\mathcal{Y}}} p(u)(V'\circ E)(x'|u)W'(y|x')D(u|y) \\
&\leq \sup_{E' \in \mathrm{DMC}_{\mathcal{U},\mathcal{X}'}} \sum_{\substack{u\in\mathcal{U},\\ x'\in\mathcal{X}',\\ y\in\mathcal{Y}}} p(u)E'(x'|u)W'(y|x')D(u|y) = P_c(p,W',D).
\end{aligned}$$

Conversely, assume that $P_c(p,W,D) \leq P_c(p,W',D)$ for every $p \in \Delta_{\mathcal{U}}$, every $D \in \mathrm{DMC}_{\mathcal{Y},\mathcal{U}}$ and every finite set $\mathcal{U}$.

Let x_0 be any symbol that does belong to $\mathcal{X}$ and let $\mathcal{U} = \mathcal{X} \cup \{x_0\}$. For every $n \geq 1$, define $p_n \in \Delta_{\mathcal{U}}$ as follows:

$$p_n(u) = \begin{cases} \dfrac{1}{|\mathcal{X}|}\left(1 - \dfrac{1}{n+1}\right) & \text{if } u \in \mathcal{X}, \\ \dfrac{1}{n+1} & \text{if } u = x_0. \end{cases}$$

p_n was chosen in such a way that $\frac{p_n(x_0)}{p_n(x)} = \frac{|\mathcal{X}|}{n}$ for every $x \in \mathcal{X}$. This is going to be useful later. Define the channel $W_0 \in \mathrm{DMC}_{\mathcal{U},\mathcal{Y}}$ as follows:

$$W_0(y|u) = \begin{cases} W(y|u) & \text{if } u \in \mathcal{X}, \\ \dfrac{1}{|\mathcal{X}|}\displaystyle\sum_{x\in\mathcal{X}} W(y|x) & \text{if } u = x_0. \end{cases}$$

Fix the encoder $E \in \mathrm{DMC}_{\mathcal{U},\mathcal{X}}$ as follows:

$$E(x|u) = \begin{cases} 1 & \text{if } u = x, \\ \dfrac{1}{|\mathcal{X}|} & \text{if } u = x_0, \\ 0 & \text{otherwise.} \end{cases}$$

For every $D \in \mathrm{DMC}_{\mathcal{Y},\mathcal{U}}$, we have:

$$\begin{aligned}
&\sum_{\substack{u\in\mathcal{U},\\ y\in\mathcal{Y}}} p_n(u)W_0(y|u)D(u|y) \\
&\quad= \left(\sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} p_n(x)W_0(y|x)D(x|y)\right) + \sum_{y\in\mathcal{Y}} p_n(x_0)W_0(y|x_0)D(x_0|y) \\
&\quad= \left(\sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} p_n(x)W(y|x)D(x|y)\right) + \sum_{y\in\mathcal{Y}} p_n(x_0)\frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} W(y|x)D(x_0|y) \\
&\quad= \left(\sum_{\substack{u\in\mathcal{X},\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} p_n(u)E(x|u)W(y|x)D(u|y)\right) + \sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} p_n(x_0)E(x|x_0)W(y|x)D(x_0|y) \\
&\quad= \sum_{\substack{u\in\mathcal{U},\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} p_n(u)E(x|u)W(y|x)D(u|y) \le P_c(p_n,W,D) \le P_c(p_n,W',D) \\
&\quad= \sup_{E'\in\mathrm{DMC}_{\mathcal{U},\mathcal{X}'}} \sum_{\substack{u\in\mathcal{U},\\ x'\in\mathcal{X}',\\ y\in\mathcal{Y}}} p_n(u)E'(x'|u)W'(y|x')D(u|y).
\end{aligned}$$

Therefore,

$$\min_{E'\in\mathrm{DMC}_{\mathcal{U},\mathcal{X}'}} \sum_{\substack{u\in\mathcal{U},\\ y\in\mathcal{Y}}} p_n(u)\left(W_0(y|x) - \sum_{x'\in\mathcal{X}'} E'(x'|u)W'(y|x')\right)D(u|y) \le 0,$$

hence

$$\max_{D\in\mathrm{DMC}_{\mathcal{Y},\mathcal{U}}} \min_{E'\in\mathrm{DMC}_{\mathcal{U},\mathcal{X}'}} \sum_{\substack{u\in\mathcal{U},\\ y\in\mathcal{Y}}} p_n(u)\left(W_0(y|u) - \sum_{x'\in\mathcal{X}'} E'(x'|u)W'(y|x')\right)D(u|y) \le 0,$$

or equivalently

$$\max_{D\in\mathrm{DMC}_{\mathcal{Y},\mathcal{U}}} \min_{E'\in\mathrm{DMC}_{\mathcal{U},\mathcal{X}'}} \sum_{\substack{u\in\mathcal{U},\\ y\in\mathcal{Y}}} p_n(u)\Big(W_0(y|u) - (W'\circ E')(y|u)\Big)D(u|y) \le 0. \tag{10.2}$$

Note that the sets $\mathrm{DMC}_{\mathcal{Y},\mathcal{U}}$ and $\mathrm{DMC}_{\mathcal{U},\mathcal{X}'}$ are compact and convex. On the other hand, since the function $\displaystyle\sum_{\substack{u\in\mathcal{U},\\ y\in\mathcal{Y}}} p_n(u)\Big(W_0(y|u) - (W'\circ E')(y|u)\Big)D(u|y)$ is affine in both $D \in \mathrm{DMC}_{\mathcal{Y},\mathcal{U}}$ and $E' \in \mathrm{DMC}_{\mathcal{U},\mathcal{X}'}$, it is continuous, concave in D and convex in E'. Therefore, we can apply the minimax theorem [71] to exchange the max and the min in Equation (10.2). We obtain:

$$\min_{E'\in\mathrm{DMC}_{\mathcal{U},\mathcal{X}'}} \max_{D\in\mathrm{DMC}_{\mathcal{Y},\mathcal{U}}} \sum_{\substack{u\in\mathcal{U},\\ y\in\mathcal{Y}}} p_n(u)\Big(W_0(y|u) - (W'\circ E')(y|u)\Big)D(u|y) \le 0.$$

Therefore, there exists $E'_n \in \mathrm{DMC}_{\mathcal{U},\mathcal{X}'}$ such that

$$\begin{aligned}
0 &\geq \max_{D\in\mathrm{DMC}_{\mathcal{Y},\mathcal{U}}} \sum_{\substack{u\in\mathcal{U},\\ y\in\mathcal{Y}}} p_n(u)\Big(W_0(y|u) - (W'\circ E'_n)(y|u)\Big)D(u|y) \\
&\overset{(a)}{=} \sum_{y\in\mathcal{Y}} \max_{u\in\mathcal{U}} p_n(u)\Big(W_0(y|u) - (W'\circ E'_n)(y|u)\Big) \\
&\geq \sum_{y\in\mathcal{Y}} \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}} p_n(u)\Big(W_0(y|u) - (W'\circ E'_n)(y|u)\Big) \\
&= \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}} p_n(u)\sum_{y\in\mathcal{Y}}\Big(W_0(y|u) - (W'\circ E'_n)(y|u)\Big) = 0,
\end{aligned}$$

where (a) follows from the fact that $\sum_{\substack{u\in\mathcal{U},\\ y\in\mathcal{Y}}} p_n(u)\Big(W_0(y|u) - (W'\circ E'_n)(y|u)\Big)D(u|y)$ is maximized when D is chosen to be deterministic in such a way that for every $y\in\mathcal{Y}$, $D(u_y|y) = 1$ for some $u_y \in \mathcal{U}$ satisfying $p_n(u_y)(W_0(y|u_y) - (W'\circ E'_n)(y|u_y)) = \max_{u\in\mathcal{U}}\Big\{p_n(u)\big(W_0(y|u) - (W'\circ E'_n)(y|u)\big)\Big\}$. We conclude that

$$\sum_{y\in\mathcal{Y}} \max_{u\in\mathcal{U}} p_n(u)\Big(W_0(y|u) - (W'\circ E'_n)(y|u)\Big) = 0.$$

Assume there exists $y\in\mathcal{Y}$ and $\tilde{u}\in\mathcal{U}$ such that

$$p_n(\tilde{u})\big(W_0(y|\tilde{u}) - (W'\circ E'_n)(y|\tilde{u})\big) < \max_{u\in\mathcal{U}} p_n(u)\Big(W_0(y|u) - (W'\circ E'_n)(y|u)\Big).$$

In this case, we have

$$\begin{aligned}
0 &= \sum_{u\in\mathcal{U}} p_n(u)\sum_{y\in\mathcal{Y}}\Big(W_0(y|u) - (W'\circ E'_n)(y|u)\Big) \\
&= \sum_{y\in\mathcal{Y}}\sum_{u\in\mathcal{U}} p_n(u)\Big(W_0(y|u) - (W'\circ E'_n)(y|u)\Big) \\
&< \sum_{y\in\mathcal{Y}} |\mathcal{U}|\cdot\max_{u\in\mathcal{U}} p_n(u)\Big(W_0(y|u) - (W'\circ E'_n)(y|u)\Big) = 0,
\end{aligned}$$

which is a contradiction. Therefore, for every $y\in\mathcal{Y}$ and every $x\in\mathcal{X}$, we have

$$\begin{aligned}
p_n(x)\big(W(y|x) - (W'\circ E'_n)(y|x)\big) &= \max_{u\in\mathcal{U}} p_n(u)\Big(W_0(y|u) - (W'\circ E'_n)(y|u)\Big) \\
&= p_n(x_0)\big(W_0(y|x_0) - (W'\circ E'_n)(y|x_0)\big),
\end{aligned}$$

which implies that

$$\begin{aligned}
\big|W(y|x) - (W'\circ E'_n)(y|x)\big| &= \frac{p_n(x_0)}{p_n(x)}\big|W_0(y|x_0) - (W'\circ E'_n)(y|x_0)\big| \\
&\leq \frac{p_n(x_0)}{p_n(x)} = \frac{|\mathcal{X}|}{n}.
\end{aligned}$$

Since the space $\mathrm{DMC}_{\mathcal{U},\mathcal{X}'}$ is compact, there exists a converging subsequence $(E'_{n_k})_{k\geq 0}$ of $(E'_n)_{n\geq 1}$. Let E' be the limit of $(E'_{n_k})_{k\geq 0}$. For every $x \in \mathcal{X}$ and every $y \in \mathcal{Y}$, we have:

$$\big|W(y|x) - (W' \circ E')(y|x)\big| = \lim_{k\to\infty} \big|W(y|x) - (W' \circ E'_{n_k})(y|x)\big| \leq \lim_{k\to\infty} \frac{|\mathcal{X}|}{n_k} = 0,$$

which means that $W(y|x) = (W' \circ E')(y|x)$. Define $V' \in \mathrm{DMC}_{\mathcal{X},\mathcal{X}'}$ as $V'(x'|x) = E'(x'|x)$ for every $x \in \mathcal{X}$ and every $x' \in \mathcal{X}'$. For every $x \in \mathcal{X}$ and every $y \in \mathcal{Y}$, we have:

$$\begin{aligned}(W' \circ V')(y|x) &= \sum_{x'\in\mathcal{X}'} W'(y|x')V'(x'|x) \\ &= \sum_{x'\in\mathcal{X}'} W'(y|x')E'(x'|x) = (W' \circ E')(y|x) = W(y|x).\end{aligned}$$

Therefore, $W = W' \circ V'$. We conclude that W is input-degraded from W'. □

10.3.3 A Characterization in Terms of Randomized Games

A *randomized game* is a 5-tuple $\mathcal{G} = (\mathcal{Z}, \mathcal{X}, \mathcal{Y}, l, W)$ such that $\mathcal{X}, \mathcal{Y}$ and $\mathcal{Z}$ are finite sets, l is a mapping from $\mathcal{Z} \times \mathcal{Y}$ to $\mathbb{R}$, and $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$. The mapping l is called the *payoff function* of the game $\mathcal{G}$, and the channel W is called the *randomizer* of $\mathcal{G}$. During the game, a player sees a symbol $z \in \mathcal{Z}$ and decides on a symbol $x \in \mathcal{X}$. A random symbol $y \in \mathcal{Y}$ is then randomly generated according to the conditional probability distribution $W(y|x)$ and the player gets the payoff $l(z, y)$.

A *strategy* for the game $\mathcal{G}$ is a channel $S \in \mathrm{DMC}_{\mathcal{Z},\mathcal{X}}$. For every $z \in \mathcal{Z}$, the *payoff gained by the strategy S for z in the game $\mathcal{G}$* is given by:

$$\$(z, S, \mathcal{G}) = \sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} S(x|z)W(y|x)l(z, y).$$

The *payoff vector gained by the strategy S in the game $\mathcal{G}$* is given by:

$$\vec{\$}(S, \mathcal{G}) = \big(\$(z, S, \mathcal{G})\big)_{z\in\mathcal{Z}} \in \mathbb{R}^{\mathcal{Z}}.$$

It is easy to see that for every $\alpha \in [0, 1]$ and every $S_1, S_2 \in \mathrm{DMC}_{\mathcal{Z},\mathcal{X}}$, we have

$$\vec{\$}(\alpha S_1 + (1-\alpha)S_2, \mathcal{G}) = \alpha\vec{\$}(S_1, \mathcal{G}) + (1-\alpha)\vec{\$}(S_2, \mathcal{G}).$$

The *achievable payoff region for the game $\mathcal{G}$* is given by:

$$\$_{\mathrm{ach}}(\mathcal{G}) = \left\{\vec{\$}(S, \mathcal{G}) :\ S \in \mathrm{DMC}_{\mathcal{Z},\mathcal{X}}\right\} \subset \mathbb{R}^{\mathcal{Z}}.$$

Clearly, $\$_{\mathrm{ach}}(\mathcal{G})$ is a convex subset of $\mathbb{R}^{\mathcal{Z}}$. Moreover, since $\mathrm{DMC}_{\mathcal{Z},\mathcal{X}}$ is compact and since the mapping $S \to \vec{\$}(S, \mathcal{G})$ is a continuous mapping from $\mathrm{DMC}_{\mathcal{Z},\mathcal{X}}$ to $\mathbb{R}^{\mathcal{Z}}$, the region $\$_{\mathrm{ach}}(\mathcal{G})$ is a compact subset of $\mathbb{R}^{\mathcal{Z}}$.

The *average payoff for the strategy $S \in \mathrm{DMC}_{\mathcal{Z},\mathcal{X}}$ for the game $\mathcal{G}$* is given by:

$$\hat{\$}(S, \mathcal{G}) = \frac{1}{|\mathcal{Z}|}\sum_{z\in\mathcal{Z}} \$(z, S, \mathcal{G}) = \sum_{\substack{z\in\mathcal{Z},\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} \frac{1}{|\mathcal{Z}|} S(x|z)W(y|x)l(z, y).$$

The *optimal average payoff for the game* $\mathcal{G}$ is given by

$$\$_{\text{opt}}(\mathcal{G}) = \sup_{S \in \text{DMC}_{\mathcal{Z},\mathcal{X}}} \hat{\$}(S, \mathcal{G}).$$

Note that we can always find an optimal strategy that is deterministic.

The following theorem provides a characterization of input-degradedness that is similar to the famous Blackwell-Sherman-Stein theorem [11, 12, 13].

Theorem 10.2. *Let* $\mathcal{X}, \mathcal{X}'$ *and* $\mathcal{Y}$ *be three finite sets. Let* $W \in \text{DMC}_{\mathcal{X},\mathcal{Y}}$ *and* $W' \in \text{DMC}_{\mathcal{X}',\mathcal{Y}}$*. The following conditions are equivalent:*

(a) W *is input-degraded from* W'*.*

(b) For every finite set $\mathcal{Z}$ *and every payoff function* $l : \mathcal{Z} \times \mathcal{Y} \to \mathbb{R}$*, we have*

$$\$_{\text{ach}}(\mathcal{Z}, \mathcal{X}, \mathcal{Y}, l, W) \subset \$_{\text{ach}}(\mathcal{Z}, \mathcal{X}', \mathcal{Y}, l, W').$$

(c) For every finite set $\mathcal{Z}$ *and every payoff function* $l : \mathcal{Z} \times \mathcal{Y} \to \mathbb{R}$*, we have*

$$\$_{\text{opt}}(\mathcal{Z}, \mathcal{X}, \mathcal{Y}, l, W) \leq \$_{\text{opt}}(\mathcal{Z}, \mathcal{X}', \mathcal{Y}, l, W').$$

Proof. Assume that (a) is true. There exists $V' \in \text{DMC}_{\mathcal{X}',\mathcal{X}}$ such that $W = W' \circ V'$. Fix a finite set $\mathcal{Z}$ and a payoff function $l : \mathcal{Z} \times \mathcal{Y} \to \mathbb{R}$. Define $\mathcal{G} = (\mathcal{Z}, \mathcal{X}, \mathcal{Y}, l, W)$ and $\mathcal{G}' = (\mathcal{Z}, \mathcal{X}', \mathcal{Y}, l, W')$.

Fix $\vec{v} = (v_z)_{z \in \mathcal{Z}} \in \$_{\text{ach}}(\mathcal{G})$. There exists $S \in \text{DMC}_{\mathcal{Z},\mathcal{X}}$ such that $(v_z)_{z \in \mathcal{Z}} = \vec{v} = \big(\$(z, S, \mathcal{G})\big)_{z \in \mathcal{Z}}$. Let $S' = V' \circ S$. For every $z \in \mathcal{Z}$, we have:

$$\begin{aligned}
\$(z, S', \mathcal{G}') &= \sum_{\substack{x' \in \mathcal{X}', \\ y \in \mathcal{Y}}} S'(x'|z) W'(y|x') l(z, y) \\
&= \sum_{\substack{x' \in \mathcal{X}', \\ y \in \mathcal{Y}}} \left(\sum_{x \in \mathcal{X}} V'(x'|x) S(x|z) \right) W'(y|x') l(z, y) \\
&= \sum_{\substack{x \in \mathcal{X}, \\ y \in \mathcal{Y}}} S(x|z) \Big(\sum_{x' \in \mathcal{X}'} W'(y|x') V'(x'|x) \Big) l(z, y) \\
&= \sum_{\substack{x \in \mathcal{X}, \\ y \in \mathcal{Y}}} S(x|z) W(y|x) l(z, y) = \$(z, S, \mathcal{G}).
\end{aligned}$$

Therefore, $\vec{v} = \vec{\$}(S', \mathcal{G}') \in \$_{\text{ach}}(\mathcal{G}')$. Since this is true for every $\vec{v} \in \$_{\text{ach}}(\mathcal{G})$, we have $\$_{\text{ach}}(\mathcal{G}) \subset \$_{\text{ach}}(\mathcal{G}')$. We conclude that (a) implies (b).

Now assume that (b) is true. Fix a finite set $\mathcal{Z}$ and a payoff function $l : \mathcal{Z} \times \mathcal{Y} \to \mathbb{R}$. Define $\mathcal{G} = (\mathcal{Z}, \mathcal{X}, \mathcal{Y}, l, W)$ and $\mathcal{G}' = (\mathcal{Z}, \mathcal{X}', \mathcal{Y}, l, W')$. We have $\$_{\text{ach}}(\mathcal{G}) \subset \$_{\text{ach}}(\mathcal{G}')$. Therefore,

$$\$_{\text{opt}}(\mathcal{G}) = \sup_{(v_z)_{z \in \mathcal{Z}} \in \$_{\text{ach}}(\mathcal{G})} \frac{1}{|\mathcal{Z}|} \sum_{z \in \mathcal{Z}} v_z \overset{(*)}{\leq} \sup_{(v'_z)_{z \in \mathcal{Z}} \in \$_{\text{ach}}(\mathcal{G}')} \frac{1}{|\mathcal{Z}|} \sum_{z \in \mathcal{Z}} v'_z = \$_{\text{opt}}(\mathcal{G}'),$$

where $(*)$ follows from the fact that $\$_{\text{ach}}(\mathcal{G}) \subset \$_{\text{ach}}(\mathcal{G}')$. This shows that (b) implies (c).

Now assume that (c) is true. Fix a finite set $\mathcal{U}$, $p \in \Delta_{\mathcal{U}}$ and $D \in \text{DMC}_{\mathcal{Y},\mathcal{U}}$. Define the payoff function $l : \mathcal{U} \times \mathcal{Y} \to \mathbb{R}$ as $l(u,y) = |\mathcal{U}|p(u)D(u|y)$. Define the randomized games $\mathcal{G} = (\mathcal{U}, \mathcal{X}, \mathcal{Y}, W, l)$ and $\mathcal{G}' = (\mathcal{U}, \mathcal{X}', \mathcal{Y}, W', l)$. We have:

$$\begin{aligned} P_c(p,W,D) &= \sup_{E \in \text{DMC}_{\mathcal{U},\mathcal{X}}} \sum_{\substack{u\in\mathcal{U},\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} p(u)E(x|u)W(y|x)D(u|y) \\ &= \sup_{E \in \text{DMC}_{\mathcal{U},\mathcal{X}}} \sum_{\substack{u\in\mathcal{U},\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} \frac{1}{|\mathcal{U}|}E(x|u)W(y|x)l(u,y) \\ &= \sup_{E \in \text{DMC}_{\mathcal{U},\mathcal{X}}} \hat{\$}(E,\mathcal{G}) = \$_{\text{opt}}(\mathcal{G}). \end{aligned}$$

Similarly, we can show that $P_c(p,W',D) = \$_{\text{opt}}(\mathcal{G}')$. Since we assumed that (c) is true, we have $\$_{\text{opt}}(\mathcal{G}) \leq \$_{\text{opt}}(\mathcal{G}')$. Therefore, for every finite set $\mathcal{U}$, every $p \in \Delta_{\mathcal{U}}$ and every $D \in \text{DMC}_{\mathcal{Y},\mathcal{U}}$, we have $P_c(p,W,D) \leq P_c(p,W',D)$. Theorem 10.1 now implies that W is input-degraded from W', hence (c) implies (a). We conclude that (a), (b) and (c) are equivalent. □

10.4 Shannon Ordering and Shannon Equivalence

Let $\mathcal{X}, \mathcal{X}', \mathcal{Y}$ and $\mathcal{Y}'$ be four finite sets. Let $W \in \text{DMC}_{\mathcal{X},\mathcal{Y}}$ and $W' \in \text{DMC}_{\mathcal{X}',\mathcal{Y}'}$. We say that W' contains W if there exist n pairs of channels $(R_i, T_i)_{1\leq i\leq n}$ and a probability distribution $\alpha \in \Delta_{[n]}$ such that $R_i \in \text{DMC}_{\mathcal{X},\mathcal{X}'}$ and $T_i \in \text{DMC}_{\mathcal{Y}',\mathcal{Y}}$ for every $1 \leq i \leq n$, and $W = \sum_{i=1}^{n} \alpha(i) T_i \circ W' \circ R_i$, i.e.,

$$W(y|x) = \sum_{i=1}^{n} \alpha(i) \sum_{\substack{x'\in\mathcal{X}',\\ y'\in\mathcal{Y}'}} T_i(y|y')W'(y'|x')R_i(x'|x).$$

The channels W and W' are said to be *Shannon-equivalent* if each one contains the other.

A channel $V \in \text{DMC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ is said to be a *convex-product channel* if it is the convex combination of the products of channels in $\text{DMC}_{\mathcal{X},\mathcal{X}'}$ with channels in $\text{DMC}_{\mathcal{Y}',\mathcal{Y}}$. More precisely, $V \in \text{DMC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ is a convex-product channel if there exist n pairs of channels $(R_i, T_i)_{1\leq i\leq n}$ and a probability distribution $\alpha \in \Delta_{[n]}$ such that $R_i \in \text{DMC}_{\mathcal{X},\mathcal{X}'}$ and $T_i \in \text{DMC}_{\mathcal{Y}',\mathcal{Y}}$ for every $1 \leq i \leq n$, and

$$V(x',y|x,y') = \sum_{i=1}^{n} \alpha(i) R_i(x'|x) T_i(y|y').$$

We denote the set of convex-product channels from $\mathcal{X}\times\mathcal{Y}'$ to $\mathcal{X}'\times\mathcal{Y}$ as $\text{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$.

Proposition 10.6. *The space* $\text{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ *is a compact and convex subset of* $\text{DMC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$.

Proof. Define the set of product channels

$$\mathrm{PC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}} = \{R \otimes T :\ R \in \mathrm{DMC}_{\mathcal{X},\mathcal{X}'},\ T \in \mathrm{DMC}_{\mathcal{Y}',\mathcal{Y}}\}.$$

Clearly, $\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ is the convex hull of $\mathrm{PC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ and so $\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ is convex. Now since $\mathrm{PC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ can be seen as a subset of $\mathbb{R}^{\mathcal{X}\times\mathcal{Y}'\times\mathcal{X}'\times\mathcal{Y}}$, it follows from the Carathéodory theorem that every channel V in $\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ can be written as a convex combination of at most

$$n = |\mathcal{X}\times\mathcal{Y}'\times\mathcal{X}'\times\mathcal{Y}| + 1$$

product channels in $\mathrm{PC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$. Define the mapping

$$f : \Delta_{[n]} \times (\mathrm{DMC}_{\mathcal{X},\mathcal{X}'} \times \mathrm{DMC}_{\mathcal{Y}',\mathcal{Y}})^n \to \mathrm{DMC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$$

as

$$f\big(\alpha, (R_i, T_i)_{1\le i\le n}\big) = \sum_{i=1}^{n} \alpha(i) R_i \otimes T_i.$$

Since $\Delta_{[n]}$, $\mathrm{DMC}_{\mathcal{X},\mathcal{X}'}$ and $\mathrm{DMC}_{\mathcal{Y}',\mathcal{Y}}$ are compact, the space $\Delta_{[n]} \times (\mathrm{DMC}_{\mathcal{X},\mathcal{X}'} \times \mathrm{DMC}_{\mathcal{Y}',\mathcal{Y}})^n$ is compact. Moreover, since f is continuous, it follows that

$$\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}} = f\big(\Delta_{[n]} \times (\mathrm{DMC}_{\mathcal{X},\mathcal{X}'} \times \mathrm{DMC}_{\mathcal{Y}',\mathcal{Y}})^n\big)$$

is compact. □

Let $\mathcal{X}, \mathcal{X}', \mathcal{X}'', \mathcal{Y}, \mathcal{Y}'$ and $\mathcal{Y}''$ be finite sets. For every $V \in \mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ and every $V' \in \mathrm{DMC}_{\mathcal{X}'\times\mathcal{Y}'',\mathcal{X}''\times\mathcal{Y}'}$, define the *skew-composition* $V \circ_s V' \in \mathrm{DMC}_{\mathcal{X}\times\mathcal{Y}'',\mathcal{X}''\times\mathcal{Y}}$ of V' with V as follows:

$$(V \circ_s V')(x'', y|x, y'') = \sum_{\substack{x'\in\mathcal{X}',\\ y'\in\mathcal{Y}'}} V(x', y|x, y') V'(x'', y'|x', y''), \tag{10.3}$$

for every $x'' \in \mathcal{X}''$, $y \in \mathcal{Y}$, $x \in \mathcal{X}$ and $y'' \in \mathcal{Y}''$. It may not be immediately clear from (10.3) that $V \circ_s V'$ is a valid channel in $\mathrm{DMC}_{\mathcal{X}\times\mathcal{Y}'',\mathcal{X}''\times\mathcal{Y}}$. In the following, we show that $V \circ_s V' \in \mathrm{DMC}_{\mathcal{X}\times\mathcal{Y}'',\mathcal{X}''\times\mathcal{Y}}$.

Let $n \ge 1$, $\alpha \in \Delta_{[n]}$, $(R_i, T_i)_{1\le i\le n}$ be such that $R_i \in \mathrm{DMC}_{\mathcal{X},\mathcal{X}'}$ and $T_i \in \mathrm{DMC}_{\mathcal{Y}',\mathcal{Y}}$ for every $1 \le i \le n$, and

$$V = \sum_{i=1}^{n} \alpha(i) R_i \otimes T_i.$$

For every $(x, y'') \in \mathcal{X}\times\mathcal{Y}''$, we have

$$\begin{aligned}
\sum_{\substack{x''\in\mathcal{X}'',\\ y\in\mathcal{Y}}} (V \circ_s V')(x'', y|x, y'') &= \sum_{\substack{x''\in\mathcal{X}'',\\ y\in\mathcal{Y}}} \sum_{\substack{x'\in\mathcal{X}',\\ y'\in\mathcal{Y}'}} V(x', y|x, y') V'(x'', y'|x', y'') \\
&= \sum_{\substack{x''\in\mathcal{X}'',\\ y\in\mathcal{Y}}} \sum_{\substack{x'\in\mathcal{X}',\\ y'\in\mathcal{Y}'}} \sum_{i=1}^{n} \alpha(i) R_i(x'|x) T_i(y|y') V'(x'', y'|x', y'') \\
&= \sum_{i=1}^{n} \alpha(i) \sum_{\substack{x''\in\mathcal{X}'',\\ y\in\mathcal{Y}}} \sum_{\substack{x'\in\mathcal{X}',\\ y'\in\mathcal{Y}'}} R_i(x'|x) T_i(y|y') V'(x'', y'|x', y'').
\end{aligned}$$

Hence,

$$\sum_{\substack{x''\in\mathcal{X}'',\\ y\in\mathcal{Y}}} (V\circ_s V')(x'',y|x,y'') = \sum_{i=1}^{n}\alpha(i)\sum_{x''\in\mathcal{X}''}\sum_{\substack{x'\in\mathcal{X}',\\ y'\in\mathcal{Y}'}} R_i(x'|x)V'(x'',y'|x',y'')$$
$$= \sum_{i=1}^{n}\alpha(i)\sum_{x'\in\mathcal{X}'} R_i(x'|x) = \sum_{i=1}^{n}\alpha(i) = 1.$$

Therefore, $V\circ_s V'\in \mathrm{DMC}_{\mathcal{X}\times\mathcal{Y}'',\mathcal{X}''\times\mathcal{Y}}$. Note that if $V\in\mathrm{DMC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ and $V\notin \mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$, then the skew-composition of V' with V as defined in Equation (10.3) does not always yield a valid channel in $\mathrm{DMC}_{\mathcal{X}\times\mathcal{Y}'',\mathcal{X}''\times\mathcal{Y}}$.

Lemma 10.1. *If $V\in\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ and $V'\in\mathrm{CPC}_{\mathcal{X}'\times\mathcal{Y}'',\mathcal{X}''\times\mathcal{Y}'}$, then $V\circ_s V'\in\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}'',\mathcal{X}''\times\mathcal{Y}}$.*

Proof. Let $n\geq 1$, $\alpha\in\Delta_{[n]}$, $(R_i,T_i)_{1\leq i\leq n}$ be such that $R_i\in\mathrm{DMC}_{\mathcal{X},\mathcal{X}'}$ and $T_i\in\mathrm{DMC}_{\mathcal{Y}',\mathcal{Y}}$ for every $1\leq i\leq n$, and

$$V=\sum_{i=1}^{n}\alpha(i)R_i\otimes T_i.$$

Let $n'\geq 1$, $\alpha'\in\Delta_{[n']}$, $(R'_j,T'_j)_{1\leq j\leq n'}$ be such that $R'_j\in\mathrm{DMC}_{\mathcal{X}',\mathcal{X}''}$ and $T'_j\in\mathrm{DMC}_{\mathcal{Y}'',\mathcal{Y}'}$ for every $1\leq j\leq n'$, and

$$V'=\sum_{j=1}^{n'}\alpha'(j)R'_j\otimes T'_j.$$

We have

$$\begin{aligned}(V\circ_s V')(x'',y|x,y'') &= \sum_{\substack{x'\in\mathcal{X}',\\ y'\in\mathcal{Y}'}} V(x',y|x,y')V'(x'',y'|x',y'')\\ &= \sum_{\substack{x'\in\mathcal{X}',\\ y'\in\mathcal{Y}'}}\sum_{i=1}^{n}\alpha(i)R_i(x'|x)T_i(y|y')\sum_{j=1}^{n'}\alpha'(j)R'_j(x''|x')T'_j(y'|y'')\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n'}\alpha(i)\alpha'(j)\sum_{\substack{x'\in\mathcal{X}',\\ y'\in\mathcal{Y}'}} R_i(x'|x)T_i(y|y')R'_j(x''|x')T'_j(y'|y'')\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n'}\alpha(i)\alpha'(j)(R'_j\circ R_i)(x''|x)(T_i\circ T'_j)(y|y'').\end{aligned}$$

Therefore, $V\circ_s V'\in\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}'',\mathcal{X}''\times\mathcal{Y}}$. □

For every $W'\in\mathrm{DMC}_{\mathcal{X}',\mathcal{Y}'}$ and every $V\in\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$, we define the *skew-composition* $V\circ_s W'\in\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ of W' with V as follows:

$$(V\circ_s W')(y|x)=\sum_{\substack{x'\in\mathcal{X}',\\ y'\in\mathcal{Y}'}} V(x',y|x,y')W'(y'|x'). \tag{10.4}$$

Note that Equation (10.4) can be seen as a particular case of Equation (10.3) if we let $\mathcal{X}'' = \mathcal{Y}'' = \{0\}$ (i.e., a singleton) and we identify $\mathrm{DMC}_{\mathcal{X}',\mathcal{Y}'}$ with $\mathrm{DMC}_{\mathcal{X}'\times\mathcal{Y}'',\mathcal{X}''\times\mathcal{Y}'}$.

The following lemma is trivial:

Lemma 10.2. *Let $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ and $W' \in \mathrm{DMC}_{\mathcal{X}',\mathcal{Y}'}$. W' contains W if and only if there exists $V \in \mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ such that $W = V \circ_s W'$.*

10.4.1 A Characterization of the Shannon Ordering

A *blind randomized in the middle (BRM) game* is a 6-tuple $\mathcal{G} = (\mathcal{U}, \mathcal{X}, \mathcal{Y}, \mathcal{V}, l, W)$ such that $\mathcal{U}, \mathcal{X}, \mathcal{Y}$ and $\mathcal{V}$ are finite sets, l is a mapping from $\mathcal{U} \times \mathcal{V}$ to $\mathbb{R}$, and $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$. The mapping l is called the *payoff function* of the BRM game $\mathcal{G}$, and the channel W is called the *randomizer* of $\mathcal{G}$. The BRM game consists of two players that we call Alice and Bob. The BRM game takes place in two stages:

- Alice chooses a symbol $u \in \mathcal{U}$ and writes her choice on a piece of paper. Bob chooses two functions $f : \mathcal{U} \to \mathcal{X}$ and $g : \mathcal{Y} \to \mathcal{V}$, and writes a description of f and g on a piece of paper. At this stage, no player has knowledge of the choice of the other player.

- Alice and Bob simultaneously reveal their papers. They compute $x = f(u) \in \mathcal{X}$ and then randomly generate a symbol $y \in \mathcal{Y}$ according to the conditional probability distribution $W(y|x)$. Finally, $v = g(y)$ is computed and then Alice pays[3] Bob an amount of money that is equal to $l(u, v)$.

A *strategy* (for Bob) in the BRM game $\mathcal{G}$ is a 4-tuple $S = (n, \alpha, \mathbf{f}, \mathbf{g})$ satisfying:

- $n \geq 1$ is a strictly positive integer.

- $\alpha \in \Delta_{[n]}$.

- $\mathbf{f} = (f_i)_{1\leq i\leq n} \in (\mathcal{X}^{\mathcal{U}})^n$, where $\mathcal{X}^{\mathcal{U}}$ is the set of functions from $\mathcal{U}$ to $\mathcal{X}$.

- $\mathbf{g} = (g_i)_{1\leq i\leq n} \in (\mathcal{V}^{\mathcal{Y}})^n$.

We denote n and α as n_S and α_S respectively. For every $1 \leq i \leq n = n_S$, we denote f_i and g_i as $f_{i,S}$ and $g_{i,S}$ respectively. The set of strategies is denoted as $\mathcal{S}_{\mathcal{U},\mathcal{X},\mathcal{Y},\mathcal{V}}$.

Bob implements the strategy S as follows: He randomly picks an index $i \in \{1, \ldots, n_S\}$ according to the distribution $\alpha_S \in \Delta_{[n_S]}$, and then commits to the choice $(f_{i,S}, g_{i,S})$.

For every $u \in \mathcal{U}$, the *payoff gained by the strategy S for u in the BRM game $\mathcal{G}$* is given by:

$$\$(u, S, \mathcal{G}) = \sum_{i=1}^{n_S} \alpha_S(i) \sum_{y\in\mathcal{Y}} W(y|f_{i,S}(u)) l(u, g_{i,S}(y)).$$

The *payoff vector gained by the strategy S in the game $\mathcal{G}$* is given by:

$$\vec{\$}(S, \mathcal{G}) = \big(\$(u, S, \mathcal{G})\big)_{u\in\mathcal{U}} \in \mathbb{R}^{\mathcal{U}}.$$

[3] If $l(u, v) < 0$, then Bob pays Alice an amount of money that is equal to $-l(u, v)$.

The *achievable payoff region for the game* $\mathcal{G}$ is given by:

$$\$_{\text{ach}}(\mathcal{G}) = \left\{\vec{\$}(S,\mathcal{G}) :\ S \in \mathcal{S}_{\mathcal{U},\mathcal{X},\mathcal{Y},\mathcal{V}}\right\} \subset \mathbb{R}^{\mathcal{U}}.$$

The *average payoff for the strategy* $S \in \mathcal{S}_{\mathcal{U},\mathcal{X},\mathcal{Y},\mathcal{V}}$ *in the game* $\mathcal{G}$ is given by:

$$\hat{\$}(S,\mathcal{G}) = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}} \$(u,S,\mathcal{G}).$$

$\hat{\$}(S,\mathcal{G})$ is the expected gain of Bob assuming that Alice chooses $u \in \mathcal{U}$ uniformly at random.

The *optimal average payoff for the game* $\mathcal{G}$ is given by

$$\$_{\text{opt}}(\mathcal{G}) = \sup_{S\in\mathcal{S}_{\mathcal{U},\mathcal{X},\mathcal{Y},\mathcal{V}}} \hat{\$}(S,\mathcal{G}).$$

For every strategy $S \in \mathcal{S}_{\mathcal{U},\mathcal{X},\mathcal{Y},\mathcal{V}}$, we associate the convex-product channel $V_S \in \text{CPC}_{\mathcal{U}\times\mathcal{Y},\mathcal{X}\times\mathcal{V}}$ defined as

$$V_S = \sum_{i=1}^{n_S} \alpha_S(i) D_{f_{i,S}} \otimes D_{g_{i,S}}.$$

For every $u \in \mathcal{U}$, we have

$$\begin{aligned}
\$(u,S,\mathcal{G}) &= \sum_{i=1}^{n_S} \alpha_S(i) \sum_{y\in\mathcal{Y}} W(y|f_{i,S}(u)) l(u,g_{i,S}(y)) \\
&= \sum_{i=1}^{n_S} \alpha_S(i) \sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y},\\ v\in\mathcal{V}}} D_{f_{i,S}}(x|u) W(y|x) D_{g_{i,S}}(v|y) l(u,v) \\
&= \sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y},\\ v\in\mathcal{V}}} \left(\sum_{i=1}^{n_S} \alpha_S(i) D_{f_{i,S}}(x|u) D_{g_{i,S}}(v|y)\right) W(y|x) l(u,v) \\
&= \sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y},\\ v\in\mathcal{V}}} V_S(x,v|u,y) W(y|x) l(u,v).
\end{aligned} \tag{10.5}$$

Lemma 10.3. *For every* $V \in \text{CPC}_{\mathcal{U}\times\mathcal{Y},\mathcal{X}\times\mathcal{V}}$*, there exists* $S \in \mathcal{S}_{\mathcal{U},\mathcal{X},\mathcal{Y},\mathcal{V}}$ *such that* $V = V_S$.

Proof. Let $n \geq 1$, $\alpha \in \Delta_{[n]}$, $(R_i,T_i)_{1\leq i\leq n}$ be such that $R_i \in \text{DMC}_{\mathcal{U},\mathcal{X}}$ and $T_i \in \text{DMC}_{\mathcal{Y},\mathcal{V}}$ for every $1 \leq i \leq n$, and

$$V = \sum_{i=1}^{n} \alpha(i) R_i \otimes T_i. \tag{10.6}$$

Since every channel can be written as a convex combination of deterministic channels [10], we can rewrite (10.6) as a convex combination of products of deterministic channels. Therefore, there exists $S \in \mathcal{S}_{\mathcal{U},\mathcal{X},\mathcal{Y},\mathcal{V}}$ such that $V = V_S$. □

Equation (10.5) and Lemma 10.3 imply that $\$_{\text{ach}}(\mathcal{G})$ is the image of $\text{CPC}_{\mathcal{U}\times\mathcal{Y},\mathcal{X}\times\mathcal{V}}$ by a linear function. Since $\text{CPC}_{\mathcal{U}\times\mathcal{Y},\mathcal{X}\times\mathcal{V}}$ is convex and compact (Proposition 10.6), $\$_{\text{ach}}(\mathcal{G})$ is convex and compact as well.

Let $\mathcal{U}$ and $\mathcal{V}$ be two finite sets and let $l : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ be a payoff function. We say that l is *normalized and positive* if $l(u, v) \geq 0$ for every $u \in \mathcal{U}$ and every $v \in \mathcal{V}$, and

$$\sum_{\substack{u\in\mathcal{U},\\ v\in\mathcal{V}}} l(u, v) = 1.$$

In other words, l is normalized and positive if $l \in \Delta_{\mathcal{U}\times\mathcal{V}}$.

The following theorem provides a characterization of the Shannon ordering of communication channels that is similar to the BSS theorem.

Theorem 10.3. *Let $\mathcal{X}, \mathcal{X}', \mathcal{Y}$ and $\mathcal{Y}'$ be four finite sets. Let $W \in \text{DMC}_{\mathcal{X},\mathcal{Y}}$ and $W' \in \text{DMC}_{\mathcal{X}',\mathcal{Y}'}$. The following conditions are equivalent:*

(a) W' contains W.

(b) For every two finite sets $\mathcal{U}$ and $\mathcal{V}$, and every payoff function $l : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$, we have

$$\$_{\text{ach}}(\mathcal{U}, \mathcal{X}, \mathcal{Y}, \mathcal{V}, l, W) \subset \$_{\text{ach}}(\mathcal{U}, \mathcal{X}', \mathcal{Y}', \mathcal{V}, l, W').$$

(c) For every two finite sets $\mathcal{U}$ and $\mathcal{V}$, and every payoff function $l : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$, we have

$$\$_{\text{opt}}(\mathcal{U}, \mathcal{X}, \mathcal{Y}, \mathcal{V}, l, W) \leq \$_{\text{opt}}(\mathcal{U}, \mathcal{X}', \mathcal{Y}', \mathcal{V}, l, W').$$

(d) For every two finite sets $\mathcal{U}$ and $\mathcal{V}$, and every normalized and positive payoff function $l \in \Delta_{\mathcal{U}\times\mathcal{V}}$, we have

$$\$_{\text{ach}}(\mathcal{U}, \mathcal{X}, \mathcal{Y}, \mathcal{V}, l, W) \subset \$_{\text{ach}}(\mathcal{U}, \mathcal{X}', \mathcal{Y}', \mathcal{V}, l, W').$$

(e) For every two finite sets $\mathcal{U}$ and $\mathcal{V}$, and every normalized and positive payoff function $l \in \Delta_{\mathcal{U}\times\mathcal{V}}$, we have

$$\$_{\text{opt}}(\mathcal{U}, \mathcal{X}, \mathcal{Y}, \mathcal{V}, l, W) \leq \$_{\text{opt}}(\mathcal{U}, \mathcal{X}', \mathcal{Y}', \mathcal{V}, l, W').$$

Proof. Assume that (a) is true. Lemma 10.2 implies that there exists a convex-product channel $V \in \text{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ such that $W = V \circ_s W'$. Let $\mathcal{U}$ and $\mathcal{V}$ be two finite sets, and let $l : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ be a payoff function. Define $\mathcal{G} = (\mathcal{U}, \mathcal{X}, \mathcal{Y}, \mathcal{V}, l, W)$ and $\mathcal{G}' = (\mathcal{U}, \mathcal{X}', \mathcal{Y}', \mathcal{V}, l, W')$.

Fix $\vec{v} \in \$_{\text{ach}}(\mathcal{G})$. There exists $S \in \mathcal{S}_{\mathcal{U},\mathcal{X},\mathcal{Y},\mathcal{V}}$ such that

$$\vec{v} = \vec{\$}(S, \mathcal{G}) = \big(\$(u, S, \mathcal{G})\big)_{u\in\mathcal{U}}.$$

From equation (10.5) we have:

$$\begin{aligned}
\$(u,S,\mathcal{G}) &= \sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y},\\ v\in\mathcal{V}}} V_S(x,v|u,y)W(y|x)l(u,v) \\
&= \sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y},\\ v\in\mathcal{V}}} V_S(x,v|u,y)\left(\sum_{\substack{x'\in\mathcal{X}',\\ y'\in\mathcal{Y}'}} V(x',y|x,y')W'(y'|x')\right)l(u,v) \\
&= \sum_{\substack{x'\in\mathcal{X}',\\ y'\in\mathcal{Y}',\\ v\in\mathcal{V}}} \left(\sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} V_S(x,v|u,y)V(x',y|x,y')\right)W'(y'|x')l(u,v) \\
&= \sum_{\substack{x'\in\mathcal{X}',\\ y'\in\mathcal{Y}',\\ v\in\mathcal{V}}} (V_S \circ_s V)(x',v|u,y')W'(y'|x')l(u,v).
\end{aligned}$$

Lemma 10.1 implies that $V_S \circ_s V \in \mathrm{CPC}_{\mathcal{U}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{V}}$ and Lemma 10.3 implies that there exists $S' \in \mathcal{S}_{\mathcal{U},\mathcal{X}',\mathcal{Y}',\mathcal{V}}$ such that $V_{S'} = V_S \circ_s V$. Therefore,

$$\$(u,S,\mathcal{G}) = \sum_{\substack{x'\in\mathcal{X}',\\ y'\in\mathcal{Y}',\\ v\in\mathcal{V}}} V_{S'}(x',v|u,y')W'(y'|x')l(u,v) \overset{(*)}{=} \$(u,S',\mathcal{G}'),$$

where $(*)$ follows from Equation (10.5). This shows that $\vec{v} = \big(\$(u,S',\mathcal{G}')\big)_{u\in\mathcal{U}}$, hence $\$_{\mathrm{ach}}(\mathcal{G}) \subset \$_{\mathrm{ach}}(\mathcal{G}')$. Therefore, (a) implies (b).

Now assume that (b) is true. Let $\mathcal{U}$ and $\mathcal{V}$ be two finite sets, and let $l : \mathcal{U}\times\mathcal{V} \to \mathbb{R}$ be a payoff function. Define $\mathcal{G} = (\mathcal{U},\mathcal{X},\mathcal{Y},\mathcal{V},l,W)$ and $\mathcal{G}' = (\mathcal{U},\mathcal{X}',\mathcal{Y}',\mathcal{V},l,W')$. We have $\$_{\mathrm{ach}}(\mathcal{G}) \subset \$_{\mathrm{ach}}(\mathcal{G}')$. Therefore,

$$\$_{\mathrm{opt}}(\mathcal{G}) = \sup_{(v_u)_{u\in\mathcal{U}}\in\$_{\mathrm{ach}}(\mathcal{G})} \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}} v_u \overset{(**)}{\leq} \sup_{(v'_u)_{u\in\mathcal{U}}\in\$_{\mathrm{ach}}(\mathcal{G}')} \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}} v'_u = \$_{\mathrm{opt}}(\mathcal{G}'),$$

where $(**)$ follows from the fact that $\$_{\mathrm{ach}}(\mathcal{G}) \subset \$_{\mathrm{ach}}(\mathcal{G}')$. This shows that (b) implies (c). We can show similarly that (d) implies (e).

Trivially, (b) implies (d), and (c) implies (e).

Now assume that (e) is true. Fix a normalized and positive payoff function $l \in \Delta_{\mathcal{X}\times\mathcal{Y}}$, and define the BRM games

$$\mathcal{G} = (\mathcal{X},\mathcal{X},\mathcal{Y},\mathcal{Y},l,W) \text{ and } \mathcal{G}' = (\mathcal{X},\mathcal{X}',\mathcal{Y}',\mathcal{Y},l,W').$$

We have $\$_{\mathrm{opt}}(\mathcal{G}) \leq \$_{\mathrm{opt}}(\mathcal{G}')$.

Fix a strategy $S \in \mathcal{S}_{\mathcal{X},\mathcal{X},\mathcal{Y},\mathcal{Y}}$ satisfying $n_S = 1$, $f_{1,S}(x) = x$ for all $x \in \mathcal{X}$ and $g_{1,S}(y) = y$ for all $y \in \mathcal{Y}$. Clearly $\alpha_S(1) = 1$, hence

$$\begin{aligned}
\hat{\$}(S,\mathcal{G}) &= \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} \$(x,S,\mathcal{G}) \\
&= \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} W(y|f_{1,S}(x))l\big(x,g_{1,S}(y)\big) = \frac{1}{|\mathcal{X}|}\sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} W(y|x)l(x,y).
\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{1}{|\mathcal{X}|}\sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} W(y|x)l(x,y) = \hat{\$}(S,\mathcal{G}) \le \$_{\mathrm{opt}}(\mathcal{G}) \le \$_{\mathrm{opt}}(\mathcal{G}') &= \sup_{S'\in\mathcal{S}_{\mathcal{X},\mathcal{X}',\mathcal{Y}',\mathcal{Y}}} \hat{\$}(S',\mathcal{G}') \\ &= \sup_{S'\in\mathcal{S}_{\mathcal{X},\mathcal{X}',\mathcal{Y}',\mathcal{Y}}} \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} \$(x,S',\mathcal{G}') \\ &= \sup_{S'\in\mathcal{S}_{\mathcal{X},\mathcal{X}',\mathcal{Y}',\mathcal{Y}}} \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}}\sum_{\substack{x'\in\mathcal{X}',\\ y'\in\mathcal{Y}',\\ y\in\mathcal{Y}}} V_{S'}(x',y|x,y')W'(y'|x')l(x,y) \\ &= \sup_{S'\in\mathcal{S}_{\mathcal{X},\mathcal{X}',\mathcal{Y}',\mathcal{Y}}} \frac{1}{|\mathcal{X}|}\sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} (V_{S'}\circ_s W')(y|x)l(x,y) \\ &\overset{(\dagger)}{=} \sup_{V\in\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}} \frac{1}{|\mathcal{X}|}\sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} (V\circ_s W')(y|x)l(x,y),\end{aligned}$$

where (†) follows from Lemma 10.3. Therefore,

$$\inf_{V\in\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}} \frac{1}{|\mathcal{X}|}\sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} \big(W(y|x) - (V\circ_s W')(y|x)\big)l(x,y) \le 0.$$

Since this is true for every $l\in\Delta_{\mathcal{X}\times\mathcal{Y}}$, we have:

$$\sup_{l\in\Delta_{\mathcal{X}\times\mathcal{Y}}}\inf_{V\in\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}} \sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} \big(W(y|x) - (V\circ_s W')(y|x)\big)l(x,y) \le 0.$$

Moreover, since $\Delta_{\mathcal{X}\times\mathcal{Y}}$ and $\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ are compact (see Proposition 10.6), the sup and the inf are attainable. Therefore, we can write:

$$\max_{l\in\Delta_{\mathcal{X}\times\mathcal{Y}}}\min_{V\in\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}} \sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} \big(W(y|x) - (V\circ_s W')(y|x)\big)l(x,y) \le 0. \tag{10.7}$$

Since the function $\sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} \big(W(y|x) - (V\circ_s W')(y|x)\big)l(x,y)$ is affine in both $l\in\Delta_{\mathcal{X}\times\mathcal{Y}}$ and $V\in\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$, it is continuous, concave in l and convex in V. On the other hand, the sets $\Delta_{\mathcal{X}\times\mathcal{Y}}$ and $\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ are compact and convex (see Proposition 10.6). Therefore, we can apply the minimax theorem [71] to exchange the max and the min in Equation (10.7). We obtain:

$$\min_{V\in\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}}\max_{l\in\Delta_{\mathcal{X}\times\mathcal{Y}}} \sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} \big(W(y|x) - (V\circ_s W')(y|x)\big)l(x,y) \le 0.$$

Therefore, there exists $V\in\mathrm{CPC}_{\mathcal{X}\times\mathcal{Y}',\mathcal{X}'\times\mathcal{Y}}$ such that

$$\begin{aligned}0 &\ge \max_{l\in\Delta_{\mathcal{X}\times\mathcal{Y}}} \sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} \big(W(y|x) - (V\circ_s W')(y|x)\big)l(x,y) \\ &\overset{(\dagger\dagger)}{=} \max_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} \big(W(y|x) - (V\circ_s W')(y|x)\big),\end{aligned}$$

where (††) follows from the fact that $\sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} \big(W(y|x)-(V\circ_s W')(y|x)\big)l(x,y)$ is maximized when we choose $l\in\Delta_{\mathcal{X},\mathcal{Y}}$ in such a way that $l(x_0,y_0)=1$ for any $(x_0,y_0)\in\mathcal{X}\times\mathcal{Y}$ satisfying

$$\big(W(y_0|x_0)-(V\circ_s W')(y_0|x_0)\big)=\max_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}}\big(W(y|x)-(V\circ_s W')(y|x)\big).$$

We conclude that for every $(x,y)\in\mathcal{X}\times\mathcal{Y}$, we have

$$W(y|x)\le(V\circ_s W')(y|x).$$

Now since $\sum_{y\in\mathcal{Y}}W(y|x)=\sum_{y\in\mathcal{Y}}(V\circ_s W')(y|x)$ for every $x\in\mathcal{X}$, we must have $W(y|x)=(V\circ_s W')(y|x)$ for every $(x,y)\in\mathcal{X}\times\mathcal{Y}$. Therefore, $W=V\circ_s W'$. Lemma 10.2 now implies that W' contains W, hence (e) implies (a). We conclude that the conditions (a), (b), (c), (d) and (e) are equivalent. □

10.5 Appendix

10.5.1 Proof of Proposition 10.2

For every $A\subset\Delta_{\mathcal{X}}$, let $\mathrm{co}(A)$ be the convex hull of A. We say that $p\in A$ is *convex-extreme* if it is an extreme point of $\mathrm{co}(A)$, i.e., for every $p_1,\ldots,p_n\in\mathrm{co}(A)$ and every $\lambda_1,\ldots,\lambda_n>0$ satisfying $\sum_{i=1}^n\lambda_i=1$ and $\sum_{i=1}^n\lambda_i p_i=p$, we have $p_1=\ldots=p_n=p$. It is easy to see that if A is finite, then the convex-extreme points of A coincide with the extreme points of $\mathrm{co}(A)$. We denote the set of convex-extreme points of A as $\mathrm{CE}(A)$.

Let $W\in\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ and $W'\in\mathrm{DMC}_{\mathcal{X},\mathcal{Z}}$ be such that W' is output-degraded from W. There exists $V\in\mathrm{DMC}_{\mathcal{Y},\mathcal{Z}}$ such that $W'=V\circ W$. Let X be a random variable uniformly distributed in $\mathcal{X}$, let Y be the output of W when X is the input, and let Z be the output of V when Y is the input in such a way that $X-Y-Z$ is a Markov chain. Clearly, $P_{Z|X}(z|x)=W'(z|x)$ for every $(x,z)\in\mathcal{X}\times\mathcal{Z}$.

For every $z\in\mathcal{Z}$, we have:

$$P^o_{W'}(z)=P_Z(z)=\sum_{y\in\mathcal{Y}}P_Y(y)P_{Z|Y}(z|y)=\sum_{y\in\mathrm{Im}(W)}V(z|y)P^o_W(y). \tag{10.8}$$

Define $V^{-1}\in\mathrm{DMC}_{\mathrm{Im}(W'),\mathrm{Im}(W)}$ as

$$V^{-1}(y|z)=P_{Y|Z}(y|z)=\frac{P_Y(y)P_{Z|Y}(z|y)}{P_Z(z)}=\frac{V(z|y)P^o_W(y)}{\sum_{y'\in\mathrm{Im}(W)}V(z|y')P^o_W(y')}.$$

Note that for every $(y,z)\in\mathrm{Im}(W)\times\mathrm{Im}(W')$, we have $V^{-1}(y|z)=0$ if and only if $V(z|y)=0$.

For every $(x,z) \in \mathcal{X} \times \mathrm{Im}(W')$, we have:

$$\begin{aligned} W_z'^{-1}(x) = P_{X|Z}(x|z) &= \sum_{y\in\mathcal{Y}} P_{X,Y|Z}(x,y|z) = \sum_{\substack{y\in\mathcal{Y},\\ P_Y(y)>0}} P_{X,Y|Z}(x,y|z) \\ &= \sum_{y\in\mathrm{Im}(W)} P_{Y|Z}(y|z)P_{X|Y,Z}(x|y,z) \overset{(a)}{=} \sum_{y\in\mathrm{Im}(W)} V^{-1}(y|z)P_{X|Y}(x|y) \\ &= \sum_{y\in\mathrm{Im}(W)} V^{-1}(y|z)W_y^{-1}(x), \end{aligned} \tag{10.9}$$

where (a) follows from the fact that $X - Y - Z$ is a Markov chain.

Equation (10.9) shows that for every $z \in \mathrm{Im}(W')$, we have

$$W_z'^{-1} \in \mathrm{co}(\{W_y^{-1} :\ y \in \mathrm{Im}(W)\}) = \mathrm{co}(\mathrm{supp}(\mathrm{MP}_W)).$$

Therefore,

$$\mathrm{co}(\mathrm{supp}(\mathrm{MP}_{W'})) = \mathrm{co}(\{W_z'^{-1} :\ z \in \mathrm{Im}(W')\}) \subset \mathrm{co}(\mathrm{supp}(\mathrm{MP}_W)). \tag{10.10}$$

Now for every $p \in \Delta_{\mathcal{X}}$, define

$$\mathcal{Y}_p := \{y \in \mathrm{Im}(W) :\ W_y^{-1} = p\}.$$

Similarly,

$$\mathcal{Z}_p := \{z \in \mathrm{Im}(W') :\ W_z'^{-1} = p\}.$$

Let $p_{ext} \in \mathrm{CE}(\mathrm{supp}(\mathrm{MP}_W))$ and let $z \in \mathrm{Im}(W')$. Equation (10.9) shows that if $z \in \mathcal{Z}_{p_{ext}}$, then $V^{-1}(y|z) = 0$ for every $y \in \mathrm{Im}(W) \setminus \mathcal{Y}_{p_{ext}}$. Now since $V^{-1}(y|z) = 0 \Leftrightarrow V(z|y) = 0$ for every $(y,z) \in \mathrm{Im}(W) \times \mathrm{Im}(W')$, we deduce that if $z \in \mathcal{Z}_{p_{ext}}$ then $V(z|y) = 0$ for every $y \in \mathrm{Im}(W) \setminus \mathcal{Y}_{p_{ext}}$. Therefore,

$$\begin{aligned} \mathrm{MP}_{W'}(p_{ext}) = \sum_{z\in\mathcal{Z}_{p_{ext}}} P_{W'}^o(z) &\overset{(a)}{=} \sum_{z\in\mathcal{Z}_{p_{ext}}} \sum_{y\in\mathrm{Im}(W)} V(z|y)P_W^o(y) \\ &\overset{(b)}{=} \sum_{z\in\mathcal{Z}_{p_{ext}}} \sum_{y\in\mathcal{Y}_{p_{ext}}} V(z|y)P_W^o(y) \le \sum_{z\in\mathrm{Im}(W')} \sum_{y\in\mathcal{Y}_{p_{ext}}} V(z|y)P_W^o(y) \\ &= \sum_{y\in\mathcal{Y}_{p_{ext}}} P_W^o(y) = \mathrm{MP}_W(p_{ext}), \end{aligned} \tag{10.11}$$

where (a) follows from Equation (10.8), and (b) follows from the fact that for every $y \in \mathrm{Im}(W) \setminus \mathcal{Y}_{p_{ext}}$, we have $V(z|y) = 0$.

Now assume that W and W' are output-equivalent. Equation (10.10) (applied twice) implies that we must have $\mathrm{co}(\mathrm{supp}(\mathrm{MP}_{W'})) = \mathrm{co}(\mathrm{supp}(\mathrm{MP}_W))$ which implies that $\mathrm{supp}(\mathrm{MP}_{W'})$ and $\mathrm{supp}(\mathrm{MP}_W)$ have the same convex-extreme points. Now fix a convex-extreme point $p_{ext} \in \mathrm{CE}(\mathrm{supp}(\mathrm{MP}_{W'})) = \mathrm{CE}(\mathrm{supp}(\mathrm{MP}_W))$. Equation (10.11) (applied twice) implies that $\mathrm{MP}_W(p_{ext}) = \mathrm{MP}_{W'}(p_{ext})$. By using Equation (10.11) again we obtain:

$$\sum_{z\in\mathcal{Z}_{p_{ext}}} \sum_{y\in\mathcal{Y}_{p_{ext}}} V(z|y)P_W^o(y) = \sum_{z\in\mathrm{Im}(W')} \sum_{y\in\mathcal{Y}_{p_{ext}}} V(z|y)P_W^o(y),$$

hence

$$\sum_{z\in\mathrm{Im}(W')\setminus\mathcal{Z}_{p_{ext}}}\sum_{y\in\mathcal{Y}_{p_{ext}}}V(z|y)P^o_W(y)=0.$$

But $P^o_W(y)>0$ for every $y\in\mathcal{Y}_{p_{ext}}$. Therefore, for every $z\in\mathrm{Im}(W')\setminus\mathcal{Z}_{p_{ext}}$ and every $y\in\mathcal{Y}_{p_{ext}}$, we must have $V(z|y)=0$ (which implies that $V^{-1}(y|z)=0$). We conclude that for every $z\in\mathrm{Im}(W')\setminus\mathcal{Z}_{p_{ext}}$, we can rewrite Equations (10.8) and (10.9) as:

$$P^o_{W'}(z)=\sum_{y\in\mathrm{Im}(W)\setminus\mathcal{Y}_{p_{ext}}}V(z|y)P^o_W(y),$$

and

$$W'^{-1}_z=\sum_{y\in\mathrm{Im}(W)\setminus\mathcal{Y}_{p_{ext}}}V^{-1}(y|z)W^{-1}_y.$$

We can now repeat the above argument but on $\mathrm{supp}(\mathrm{MP}_W)\setminus\{p_{ext}\}$ and $\mathrm{supp}(\mathrm{MP}_{W'})\setminus\{p_{ext}\}$ instead of $\mathrm{supp}(\mathrm{MP}_W)$ and $\mathrm{supp}(\mathrm{MP}_{W'})$. We deduce that $\mathrm{co}(\mathrm{supp}(\mathrm{MP}_W)\setminus\{p_{ext}\})=\mathrm{co}(\mathrm{supp}(\mathrm{MP}_{W'})\setminus\{p_{ext}\})$ so $\mathrm{supp}(\mathrm{MP}_W)\setminus\{p_{ext}\}$ and $\mathrm{supp}(\mathrm{MP}_{W'})\setminus\{p_{ext}\}$ have the same convex-extreme points. We can also prove that $\mathrm{MP}_W(p'_{ext})=\mathrm{MP}_{W'}(p'_{ext})$ for every $p'_{ext}\in\mathrm{CE}(\mathrm{supp}(\mathrm{MP}_{W'})\setminus\{p_{ext}\})=\mathrm{CE}(\mathrm{supp}(\mathrm{MP}_W)\setminus\{p_{ext}\})$.

Notice that any point of $\mathrm{supp}(\mathrm{MP}_W)$ (respectively $\mathrm{supp}(\mathrm{MP}_{W'})$) becomes convex-extreme after removing a finite number of elements from $\mathrm{supp}(\mathrm{MP}_W)$ (respectively $\mathrm{supp}(\mathrm{MP}_{W'})$). Therefore, after inductively applying the above argument a finite number of times, we can deduce that $\mathrm{supp}(\mathrm{MP}_W)=\mathrm{supp}(\mathrm{MP}_{W'})$ and $\mathrm{MP}_W(p)=\mathrm{MP}_{W'}(p)$ for every $p\in\mathrm{supp}(\mathrm{MP}_W)=\mathrm{supp}(\mathrm{MP}_{W'})$, hence $\mathrm{MP}_W=\mathrm{MP}_{W'}$.

Now let $W\in\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ and $W'\in\mathrm{DMC}_{\mathcal{X},\mathcal{Z}}$ be any two channels satisfying $\mathrm{MP}_W=\mathrm{MP}_{W'}$. We have $\mathrm{supp}(\mathrm{MP}_W)=\mathrm{supp}(\mathrm{MP}_{W'})$. Furthermore, for every $p\in\mathrm{supp}(\mathrm{MP}_W)=\mathrm{supp}(\mathrm{MP}_{W'})$, we have

$$\sum_{y\in\mathcal{Y}_p}P^o_W(y)=\mathrm{MP}_W(p)=\mathrm{MP}_{W'}(p)=\sum_{z\in\mathcal{Z}_p}P^o_{W'}(z).$$

Define the channel $V\in\mathrm{DMC}_{\mathcal{Y},\mathcal{Z}}$ as

$$V(z|y)=\begin{cases}\dfrac{1}{|\mathcal{Z}|} & \text{if } y\notin\mathrm{Im}(W),\\ \dfrac{P^o_{W'}(z)}{\mathrm{MP}_{W'}(W^{-1}_y)} & \text{if } y\in\mathrm{Im}(W) \text{ and } z\in\mathcal{Z}_{W^{-1}_y},\\ 0 & \text{otherwise.}\end{cases}$$

A simple calculation shows that $\sum_{z\in\mathcal{Z}}V(z|y)=1$ for every $y\in\mathcal{Y}$, so V is a valid channel.

Notice that for every $(y,z)\in\mathrm{Im}(W)\times\mathrm{Im}(W')$, we have:

$$z\in\mathcal{Z}_{W^{-1}_y}\ \Leftrightarrow\ W'^{-1}_z=W^{-1}_y\ \Leftrightarrow\ y\in\mathcal{Y}_{W'^{-1}_z}.$$

Moreover, if $z \in \mathrm{Im}(W')$ and $y \in \mathcal{Y}_{W'^{-1}_z}$, we have $\mathrm{MP}_{W'}(W_y^{-1}) = \mathrm{MP}_W(W'^{-1}_z)$. Therefore, we can rewrite V as:

$$V(z|y) = \begin{cases} \dfrac{P^o_{W'}(z)}{\mathrm{MP}_W(W'^{-1}_z)} & \text{if } z \in \mathrm{Im}(W') \text{ and } y \in \mathcal{Y}_{W'^{-1}_z}, \\ \dfrac{1}{|\mathcal{Z}|} & \text{if } y \notin \mathrm{Im}(W), \\ 0 & \text{otherwise.} \end{cases}$$

Let $W'' = V \circ W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Z}}$. For every $z \in \mathcal{Z} \setminus \mathrm{Im}(W')$, Equation (10.8) implies that:

$$P^o_{W''}(z) = \sum_{y \in \mathrm{Im}(W)} V(z|y) P^o_W(y) \overset{(a)}{=} 0 = P^o_{W'}(z),$$

where (a) follows from the fact that $V(z|y) = 0$ if $y \in \mathrm{Im}(W)$ and $z \notin \mathrm{Im}(W')$.

On the other hand, for every $z \in \mathrm{Im}(W')$, Equation (10.8) implies that:

$$\begin{aligned} P^o_{W''}(z) &= \sum_{y \in \mathrm{Im}(W)} V(z|y) P^o_W(y) = \sum_{y \in \mathcal{Y}_{W'^{-1}_z}} \frac{P^o_{W'}(z)}{\mathrm{MP}_W(W'^{-1}_z)} P^o_W(y) \\ &= \frac{P^o_{W'}(z)}{\mathrm{MP}_W(W'^{-1}_z)} \sum_{y \in \mathcal{Y}_{W'^{-1}_z}} P^o_W(y) = \frac{P^o_{W'}(z)}{\mathrm{MP}_W(W'^{-1}_z)} \mathrm{MP}_W(W'^{-1}_z) = P^o_{W'}(z). \end{aligned}$$

Therefore, $P^o_{W''}(z) = P^o_{W'}(z)$ for every $z \in \mathcal{Z}$, which implies that $\mathrm{Im}(W'') = \mathrm{Im}(W')$.

Now define $V^{-1} \in \mathrm{DMC}_{\mathrm{Im}(W''),\mathrm{Im}(W)}$ as

$$V^{-1}(y|z) = \frac{V(z|y) P^o_W(y)}{\sum_{y' \in \mathrm{Im}(W)} V(z|y') P^o_W(y')}.$$

Equation (10.9) implies that for every $z \in \mathrm{Im}(W'') = \mathrm{Im}(W')$, we have:

$$\begin{aligned} W''^{-1}_z &= \sum_{y \in \mathrm{Im}(W)} V^{-1}(y|z) W_y^{-1} \overset{(a)}{=} \sum_{y \in \mathcal{Y}_{W'^{-1}_z}} V^{-1}(y|z) W_y^{-1} \\ &= \sum_{y \in \mathcal{Y}_{W'^{-1}_z}} V^{-1}(y|z) W'^{-1}_z \overset{(b)}{=} \sum_{y \in \mathrm{Im}(W)} V^{-1}(y|z) W'^{-1}_z = W'^{-1}_z, \end{aligned}$$

where (a) and (b) follow from the fact that for every $(y, z) \in \mathrm{Im}(W) \times \mathrm{Im}(W'')$, we have $V^{-1}(y|z) = 0$ if and only if $V(z|y) = 0$.

We conclude that $P^o_{W''} = P^o_{W'}$, and for every $z \in \mathrm{Im}(W'') = \mathrm{Im}(W')$, we have $W''^{-1}_z = W'^{-1}_z$. Therefore, $W' = W'' = V \circ W$ and so W' is output-degraded from W. By exchanging the roles of W and W' we get that W is also output-degraded from W', hence W and W' are output-equivalent.

Topological Structures on DMC Spaces

11

Let $\mathcal{X}$ and $\mathcal{Y}$ be two fixed finite sets. Every discrete memoryless channel (DMC) with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ can be determined by its transition probabilities. Since there are $|\mathcal{X}| \times |\mathcal{Y}|$ such probabilities, the space of all channels from $\mathcal{X}$ to $\mathcal{Y}$ can be seen as a subset of $\mathbb{R}^{|\mathcal{X}|\times|\mathcal{Y}|}$. Therefore, this space can be naturally endowed with the Euclidean metric, or any other equivalent metric. A generalization of this topology to infinite input and output alphabets was considered in [72].

There are a few drawbacks for this approach. For example, consider the case where $\mathcal{X} = \mathcal{Y} = \mathbb{F}_2 := \{0, 1\}$. The binary symmetric channels BSC(ϵ) and BSC($1-\epsilon$) have non-zero Euclidean distance if $\epsilon \neq \frac{1}{2}$. On the other hand, BSC(ϵ) and BSC($1-\epsilon$) are completely equivalent from an operational point of view: Both channels have exactly the same probability of error under optimal decoding for any fixed code. Moreover, any sub-optimal decoder for one channel can be transformed to a sub-optimal decoder for the other channel without changing the probability of error nor the computational complexity. This is why it makes sense, from an information-theoretic point of view, to identify output-equivalent channels and consider them as one point in the space of "output-equivalent channels".

The limitation of the Euclidean metric is clearer when we consider channels with different output alphabets. For example, BSC $\left(\frac{1}{2}\right)$ and BEC (1) are completely equivalent but they do not have the same output alphabet, and so there is no way to compare them with the Euclidean metric because they do not belong to the same space.

The standard approach to solve this problem is to find a "canonical sufficient statistic" and find a representation of each channel in terms of this sufficient statistic. This makes it possible to compare channels with different output-alphabets. One standard sufficient statistic that has been widely used for binary-input channels is the log-likelihood ratio. Each binary-input channel can be represented as a density of log-likelihood ratios (called L-density in [69]). This representation makes it possible to "topologize" the space of "output-equivalent binary-input channels" by considering the topology of convergence in distribution [69]. A similar approach can be adopted for non-binary-input channels (see [73] and [74]). Another (equiv-

alent) way to "topologize" the space of output-equivalent channels is by using the Le Cam deficiency distance [75].

One drawback[1] of the current formulation of this topology is that it does not allow us to see it as a "natural topology". Consider a fixed output alphabet $\mathcal{Y}$ and let us focus on the space of "equivalent channels" from $\mathcal{X}$ to $\mathcal{Y}$. Since this space is the quotient of the space of channels from $\mathcal{X}$ to $\mathcal{Y}$, which is naturally topologized by the Euclidean metric, it seems that the most natural topology on this space is the quotient of the Euclidean topology by the output-equivalence relation. This motivates us to consider a topology on the space of "output-equivalent channels" with input alphabet $\mathcal{X}$ and arbitrary but finite output alphabet as *natural* if and only if it induces the quotient topology on the subspaces of "output-equivalent channels" from $\mathcal{X}$ to $\mathcal{Y}$ for any finite output alphabet $\mathcal{Y}$. A legitimate question to ask now is whether the L-density topology is natural in this sense or not.

In this chapter[2], we construct and study several topologies on the quotients of the spaces of discrete memoryless channels by the output-equivalence, the input-equivalence and the Shannon-equivalence relations.

In Section 11.1, we provide a brief summary of the basic concepts and theorems in general topology. In Section 11.2, we introduce the measure-theoretic notations that we use in this chapter. In Section 11.3, We define and study the space of channels from $\mathcal{X}$ to $\mathcal{Y}$.

In Section 11.4, we define and study the space of output-equivalent channels with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. In Section 11.5, we introduce the space of output-equivalent channels with fixed input alphabet $\mathcal{X}$ and arbitrary but finite output alphabet. We investigate the properties of general natural topologies, and we study the finest natural topology. We introduce the *noisiness metric* on the space of output-equivalent channels, and we show that its induced topology, which we call the *noisiness topology*, is natural. We also study the topologies that are inherited from the space of meta-probability measures by identifying output-equivalent channels with their Blackwell measures. We show that the weak-$*$ topology (which is the standard generalization of the L-density topology to non-binary-input channels) is exactly the same as the noisiness topology. Furthermore, we show that the Borel σ-algebra is the same for all Hausdorff natural topologies.

In Section 11.6, we define and study the space of input-equivalent channels with fixed input and output alphabets. In Section 11.7, we introduce the space of input-equivalent channels with fixed output alphabet $\mathcal{Y}$ and arbitrary but finite input alphabet. A topology on this space is said to be natural if it induces the quotient topology on the subspaces of input-equivalent channels with fixed input alphabet. We investigate the properties of general natural topologies, and we study the finest natural topology. We introduce the similarity metric on the space of input-equivalent channels, and we show that the topology induced by this metric is natural.

In Section 11.8, we define and study the space of Shannon-equivalent channels with fixed input and output alphabets. In Section 11.9, we introduce the space of Shannon-equivalent channels with arbitrary but finite input and output alphabets. A topology on this space is said to be natural if it induces the quotient topology on the subspaces of Shannon-equivalent channels with fixed input and output alphabets.

[1] The mentioned drawback is secondary, and it is relevant only for conceptual purposes.

[2] The material of this chapter is based on [61, 62, 63, 64, 65, 66].

We investigate the properties of general natural topologies, and we study the finest natural topology. We introduce the BRM metric on the space of Shannon-equivalent channels, and we show that the topology induced by this metric is natural.

11.1 Introduction to General Topology

In this section, we recall basic definitions and well known theorems in general topology. The reader who is already familiar with the basic concepts of topology may skip this section and refer to it later if necessary. Proofs of all non-referenced facts can be found in any standard textbook on general topology (e.g., [76]). Definitions and theorems that may not be widely known can be found in Sections 11.1.10, 11.1.14 and 11.1.15.

11.1.1 Set-Theoretic Notations

A collection $\mathcal{A} \subset 2^B$ of subsets of B is said to be *finer* than another collection $\mathcal{A}' \subset 2^B$ if $\mathcal{A}' \subset \mathcal{A}$. If this is the case, we also say that $\mathcal{A}'$ is *coarser* than $\mathcal{A}$.

Let $(A_i)_{i\in I}$ be a collection of arbitrary sets indexed by I. The *disjoint union* of $(A_i)_{i\in I}$ is defined as $\coprod_{i\in I} A_i = \bigcup_{i\in I}(A_i \times \{i\})$. For every $i \in I$, the i^{th}*-canonical injection* is the mapping $\phi_i : A_i \to \coprod_{j\in I} A_j$ defined as $\phi_i(x_i) = (x_i, i)$. If no confusions can arise, we can identify A_i with $A_i \times \{i\}$ through the canonical injection. Therefore, we can see A_i as a subset of $\coprod_{j\in I} A_j$ for every $i \in I$.

A *relation* R on a set T is a subset of $T \times T$. For every $x, y \in T$, we write xRy to denote $(x, y) \in R$.

A relation is said to be *reflexive* if xRx for every $x \in T$. It is *symmetric* if xRy implies yRx for every $x, y \in T$. It is *anti-symmetric* if xRy and yRx imply $x = y$ for every $x, y \in T$. It is *transitive* if xRy and yRz imply xRz for every $x, y, z \in T$.

An *order relation* is a relation that is reflexive, anti-symmetric and transitive. An *equivalence relation* is a relation that is reflexive, symmetric and transitive.

Let R be an equivalence relation on T. For every $x \in T$, the set $\hat{x} = \{y \in T : xRy\}$ is the *R-equivalence class* of x. The collection of R-equivalence classes, which we denote as T/R, forms a partition of T, and it is called the *quotient space of* T *by* R. The mapping $\mathrm{Proj}_R : T \to T/R$ defined as $\mathrm{Proj}_R(x) = \hat{x}$ for every $x \in T$ is the *projection mapping onto* T/R.

11.1.2 Topological Spaces

A *topological space* is a pair $(T,\mathcal{U})$, where $\mathcal{U} \subset 2^T$ is a collection of subsets of T satisfying:

- $\emptyset \in \mathcal{U}$ and $T \in \mathcal{U}$.
- The intersection of a finite collection of members of $\mathcal{U}$ is also a member of $\mathcal{U}$.
- The union of an arbitrary collection of members of $\mathcal{U}$ is also a member of $\mathcal{U}$.

If $(T,\mathcal{U})$ is a topological space, we say that $\mathcal{U}$ is a *topology* on T.

The power set 2^T of T is clearly a topology. It is called the *discrete topology* on T.

If $\mathcal{A}$ is a an arbitrary collection of subsets of T, we can construct a topology on T starting from $\mathcal{A}$ as follows:

$$\bigcap_{\substack{\mathcal{A}\subset\mathcal{V}\subset 2^T,\\ \mathcal{V}\text{ is a topology on }T}} \mathcal{V}.$$

This is the coarsest topology on T that contains $\mathcal{A}$. It is called the *topology on T generated by $\mathcal{A}$*.

Let $(T,\mathcal{U})$ be a topological space. The subsets of T that are members of $\mathcal{U}$ are called the *open sets* of T. Complements of open sets are called *closed sets*. We can easily see that the closed sets satisfy the following:

- $\emptyset$ and T are closed.
- The union of a finite collection of closed sets is closed.
- The intersection of an arbitrary collection of closed sets is closed.

Let A be an arbitrary subset of T. The *closure* $\mathrm{cl}(A)$ of A is the smallest closed set containing A:

$$\mathrm{cl}(A) = \bigcap_{\substack{A\subset F\subset T,\\ F\text{ is closed}}} F.$$

The *interior* A° of A is the largest open subset of A:

$$A^\circ = \bigcup_{\substack{U\subset A,\\ U\text{ is open}}} U.$$

If $A \subset T$ and $\mathrm{cl}(A) = T$, we say that A is *dense* in T.

$(T,\mathcal{U})$ is said to be *separable* if there exists a countable subset of T that is dense in T.

A subset O of T is said to be a *neighborhood* of $x \in T$ if there exists an open set $U \in \mathcal{U}$ such that $x \in U \subset O$.

A *neighborhood basis* of $x \in T$ is a collection $\mathcal{O}$ of neighborhoods of x such that for every neighborhood O of x, there exists $O' \in \mathcal{O}$ such that $O' \subset O$.

We say that $(T,\mathcal{U})$ is *first-countable* if every point $x \in T$ has a countable neighborhood basis.

A collection of open sets $\mathcal{B} \subset \mathcal{U}$ is said to be a *base* for the topology $\mathcal{U}$ if every open set $U \in \mathcal{U}$ can be written as the union of elements of $\mathcal{B}$.

We say that $(T,\mathcal{U})$ is a *second-countable* space if the topology $\mathcal{U}$ has a countable base.

It is a well known fact that every second-countable space is first-countable and separable.

We say that a sequence $(x_n)_{n\geq 0}$ of elements of T *converges* to $x \in T$ if for every neighborhood O of x, there exists $n_0 \geq 0$ such that for every $n \geq n_0$, we have $x_n \in O$. We say that x is a *limit* of the sequence $(x_n)_{n\geq 0}$. Note that the limit does not need to be unique if there is no constraint on the topology.

11.1.3 Separation Axioms

$(T,\mathcal{U})$ is said to be a T_1-*space* if for every $x, y \in T$, there exists an open set $U \in \mathcal{U}$ such that $x \in U$ and $y \notin U$. It is easy to see that $(T,\mathcal{U})$ is T_1 if and only if all singletons are closed.

$(T,\mathcal{U})$ is said to be a *Hausdorff* space (or T_2-*space*) if for every $x, y \in T$, there exist two open sets $U, V \in \mathcal{U}$ such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$.

If $(T,\mathcal{U})$ is Hausdorff, then the limit of every converging sequence is unique.

$(T,\mathcal{U})$ is said to be *regular* if for every $x \in T$ and every closed set F not containing x, there exist two open sets $U, V \in \mathcal{U}$ such that $x \in U$, $F \subset V$ and $U \cap V = \emptyset$.

$(T,\mathcal{U})$ is said to be *normal* if for every two disjoint closed sets A and B, there exist two open sets $U, V \in \mathcal{U}$ such that $A \subset U$, $B \subset V$ and $U \cap V = \emptyset$.

If $(T,\mathcal{U})$ is normal, disjoint closed sets can be separated by disjoint closed neighborhoods. I.e., for every two disjoint closed sets A and B, there exist two open sets $U, U' \in \mathcal{U}$ and two closed sets K, K' such that $A \subset U \subset K$, $B \subset U' \subset K'$ and $K \cap K' = \emptyset$.

$(T,\mathcal{U})$ is said to be a T_3-*space* if it is both T_1 and regular.

$(T,\mathcal{U})$ is said to be a T_4-*space* if it is both T_1 and normal.

It is easy to see that $T_4 \Rightarrow T_3 \Rightarrow T_2 \Rightarrow T_1$.

11.1.4 Relativization

If $(T,\mathcal{U})$ is a topological space and A is an arbitrary subset of T, then A inherits a topology $\mathcal{U}_A$ from $(T,\mathcal{U})$ as follows:

$$\mathcal{U}_A = \{A \cap U : U \in \mathcal{U}\}.$$

It is easy to check that $\mathcal{U}_A$ is a topology on A.

If $(T,\mathcal{U})$ is first-countable (respectively second-countable, or Hausdorff), then $(A,\mathcal{U}_A)$ is first-countable (respectively second-countable, or Hausdorff).

If $(T,\mathcal{U})$ is normal and A is closed, then $(A,\mathcal{U}_A)$ is normal.

The union of a countable number of separable subspaces is separable.

11.1.5 Continuous Mappings

Let $(T,\mathcal{U})$ and $(S,\mathcal{V})$ be two topological spaces. A mapping $f : T \to S$ is said to be *continuous* if for every $V \in \mathcal{V}$, we have $f^{-1}(V) \in \mathcal{U}$.

$f : T \to S$ is an *open* mapping if $f(U) \in \mathcal{V}$ whenever $U \in \mathcal{U}$. $f : T \to S$ is a *closed* mapping if $f(F)$ is closed in S whenever F is closed in T.

A bijection $f : T \to S$ is a *homeomorphism* if both f and f^{-1} are continuous. In this case, for every $A \subset T$, $A \in \mathcal{U}$ if and only if $f(A) \in \mathcal{V}$. This means that $(T,\mathcal{U})$ and $(S,\mathcal{V})$ have the same topological structure and share the same topological properties.

11.1.6 Compact and Sequentially Compact Spaces

$(T,\mathcal{U})$ is a *compact* space if every open cover of T admits a finite sub-cover. I.e., if $(U_i)_{i\in I}$ is a collection of open sets such that $T = \bigcup_{i\in I} U_i$ then there exists $n > 0$ and $i_1, \ldots, i_n \in I$ such that $T = \bigcup_{j=1}^{n} U_{i_j}$.

If $(T,\mathcal{U})$ is compact, then every closed subset of T is compact (with respect to the inherited topology).

If $f : T \to S$ is a continuous mapping from a compact space $(T,\mathcal{U})$ to an arbitrary topological space $(S,\mathcal{V})$, then $f(T)$ is compact.

If A is a compact subset of a Hausdorff topological space, then A is closed.

$(T,\mathcal{U})$ is said to be *locally compact* if every point has at least one compact neighborhood. A compact space is automatically locally compact.

If $(T,\mathcal{U})$ is Hausdorff and locally compact, then for every point $x \in T$ and every neighborhood O of x, O contains a compact neighborhood of x.

A compact Hausdorff space is always normal.

$(T,\mathcal{U})$ is a *σ-compact* space if it is the union of a countable collection of compact subspaces.

$(T,\mathcal{U})$ is *countably compact* if every countable open cover of T admits a finite sub-cover. This is a weaker condition compared to compactness.

$(T,\mathcal{U})$ is said to be *sequentially compact* if every sequence in T has a converging subsequence. In general, compactness does not imply sequential compactness nor the other way around.

11.1.7 Connected Spaces

$(T,\mathcal{U})$ is a *connected* space if it satisfies one of the following equivalent conditions:

- T cannot be written as the union of two disjoint non-empty open sets.
- T cannot be written as the union of two disjoint non-empty closed sets.
- The only subsets of T that are both open and closed are $\emptyset$ and T.
- Every continuous mapping from T to $\{0,1\}$ is constant, where $\{0,1\}$ is endowed with the discrete topology.

$(T,\mathcal{U})$ is *path-connected* if every two points of T can be joined by a continuous path. I.e., for every $x, y \in T$, there exists a continuous mapping $f : [0,1] \to T$ such that $f(0) = x$ and $f(1) = y$, where $[0,1]$ is endowed with the well known Euclidean topology[3].

A path-connected space is connected but the converse is not true in general.

A subset A of T is said to be connected (respectively path-connected) if $(A,\mathcal{U}_A)$ is connected (respectively path-connected).

If $(A_i)_{i\in I}$ is a collection of connected (respectively path-connected) subsets of T such that $\bigcap_{i\in I} A_i \neq \emptyset$, then $\bigcup_{i\in I} A_i$ is connected (respectively path-connected).

11.1.8 Product of Topological Spaces

Let $\{(T_i,\mathcal{U}_i)\}_{i\in I}$ be a collection of topological spaces indexed by I. Let $T = \prod_{i\in I} T_i$ be the product of this collection. For every $j \in I$, the *j^{th}-canonical projection* is the mapping $\mathrm{Proj}_j : T \to T_j$ defined as $\mathrm{Proj}_j\left((x_i)_{i\in I}\right) = x_j$.

[3]See Section 11.1.11 for the definition of the Euclidean metric and its induced topology

The *product topology* $\mathcal{U} := \bigotimes_{i\in I} \mathcal{U}_i$ on T is the coarsest topology that makes all the canonical projections continuous. It can be shown that $\mathcal{U}$ is generated by the collection of sets of the form $\prod_{i\in I} U_i$, where $U_i \in \mathcal{U}_i$ for all $i \in I$, and $U_i \neq T_i$ for only finitely many $i \in I$.

The product of T_1 (respectively, Hausdorff, regular, T_3, compact, connected, or path-connected) spaces is T_1 (respectively, Hausdorff, regular, T_3, compact, connected, or path-connected).

11.1.9 Disjoint Union

Let $\{(T_i,\mathcal{U}_i)\}_{i\in I}$ be a collection of topological spaces indexed by I. Let $T = \coprod_{i\in I} T_i$ be the disjoint union of this collection. The *disjoint union topology* $\mathcal{U} := \bigoplus_{i\in I} \mathcal{U}_i$ on T is the finest topology which makes all the canonical injections continuous. It can be shown that $U \in \mathcal{U}$ if and only if $U \cap T_i \in \mathcal{U}_i$ for every $i \in I$.

A mapping $f : T \to S$ from $(T,\mathcal{U})$ to a topological space $(S,\mathcal{V})$ is continuous if and only if it is continuous on T_i for every $i \in I$.

The disjoint union of T_1 (respectively Hausdorff) spaces is T_1 (respectively Hausdorff). The disjoint union of two or more non-empty spaces is always disconnected.

Products are distributive with respect to the disjoint union, i.e., if $(S,\mathcal{V})$ is a topological space then $S\times\left(\coprod_{i\in I} T_i\right) = \coprod_{i\in I}(S\times T_i)$ and $\mathcal{V}\otimes\left(\bigoplus_{i\in I}\mathcal{U}_i\right) = \bigoplus_{i\in I}(\mathcal{V}\otimes\mathcal{U}_i)$.

11.1.10 Quotient Topology

Let $(T,\mathcal{U})$ be a topological space and let R be an equivalence relation on T. The *quotient topology* on T/R is the finest topology that makes the projection mapping Proj_R continuous. It is given by

$$\mathcal{U}/R = \left\{\hat{U} \subset T/R :\ \mathrm{Proj}_R^{-1}(\hat{U}) \in \mathcal{U}\right\}.$$

Lemma 11.1. *Let $f : T \to S$ be a continuous mapping from $(T,\mathcal{U})$ to $(S,\mathcal{V})$. If $f(x) = f(x')$ for every $x, x' \in T$ satisfying xRx', then we can define a* transcendent *mapping $f : T/R \to S$ such that $f(\hat{x}) = f(x')$ for any $x' \in \hat{x}$. f is well defined on T/R . Moreover, f is a continuous mapping from $(T/R,\mathcal{U}/R)$ to $(S,\mathcal{V})$.*

If $(T,\mathcal{U})$ is compact (respectively, connected, or path-connected), then $(T/R,\mathcal{U}/R)$ is compact (respectively, connected, or path-connected).

T/R is said to be *upper semi-continuous* if for every $\hat{x} \in T/R$ and every open set $U \in \mathcal{U}$ satisfying $\hat{x} \subset U$, there exists an open set $V \in \mathcal{U}$ such that $\hat{x} \subset V \subset U$, and V can be written as the union of members of T/R.

The following Lemma characterizes upper semi-continuous quotient spaces:

Lemma 11.2. *[76] T/R is upper semi-continuous if and only if* Proj_R *is a closed mapping.*

The following theorem is very useful to prove many topological properties for the quotient space:

Theorem 11.1. *[76] Let $(T,\mathcal{U})$ be a topological space, and let R be an equivalence relation on T such that T/R is upper semi-continuous and $\hat{x}$ is a compact subset of T for every $\hat{x} \in T/R$. If $(T,\mathcal{U})$ is Hausdorff (respectively, regular, locally compact, or second-countable) then $(T/R,\mathcal{U}/R)$ is Hausdorff (respectively, regular, locally compact, or second-countable).*

11.1.11 Metric Spaces

A *metric space* is a pair (M, d), where $d : M \times M \to \mathbb{R}^+$ satisfies:

- $d(x, y) = 0$ if and only if $x = y$ for every $x, y \in M$.
- Symmetry: $d(x, y) = d(y, x)$ for every $x, y \in M$.
- Triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$ for every $x, y, z \in M$.

If (M, d) is a metric space, we say that d is a *metric* (or *distance*) on M.

For every $x \in M$ and every $\epsilon > 0$, we define the *open ball* of center x and radius ϵ as:

$$B_\epsilon(x) = \{y \in M : \ d(x, y) < \epsilon\}.$$

The *metric topology* $\mathcal{U}_d$ on M *induced by* d is the coarsest topology on M which makes d a continuous mapping from $M \times M$ to $\mathbb{R}^+$. It is generated by all the open balls.

The metric topology is always T_4 and first-countable. Moreover, $(M,\mathcal{U}_d)$ is separable if and only if it is second-countable.

Since every metric space is Hausdorff, we can see that every subset of a compact metric space is closed if and only if it is compact.

Every σ-compact metric space is second-countable.

For metric spaces, compactness and sequential compactness are equivalent.

A function $f : M_1 \to M_2$ from a metric space (M_1, d_1) to a metric space (M_2, d_2) is said to be *uniformly continuous* if for every $\epsilon > 0$, there exists $\delta > 0$ such that for every $x, x' \in M_1$ satisfying $d_1(x, x') < \delta$ we have $d_2(f(x), f(x')) < \epsilon$.

If $f : M_1 \to M_2$ is a continuous mapping from a compact metric space (M_1, d_1) to an arbitrary metric space (M_2, d_2), then f is uniformly continuous.

A topological space $(T,\mathcal{U})$ is said to be *metrizable* if there exists a metric d on T such that $\mathcal{U}$ is the metric topology on T induced by d.

The disjoint union of metrizable spaces is always metrizable.

The following theorem shows that all separable metrizable spaces are characterized topologically:

Theorem 11.2. *[76] A topological space $(T,\mathcal{U})$ is metrizable and separable if and only if it is Hausdorff, regular and second countable.*

The *Euclidean metric* on $\mathbb{R}^n$ is defined as $d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$, where $x = (x_i)_{1\leq i\leq n}$ and $y = (y_i)_{1\leq i\leq n}$.

$\mathbb{R}^n$ is second countable. Moreover, a subset of $\mathbb{R}^n$ is compact if and only if it is bounded and closed.

11.1.12 Complete Metric Spaces

A sequence $(x_n)_{n\geq 0}$ is said to be a Cauchy sequence in (M,d) if for every $\epsilon > 0$, there exists $n_0 \geq 0$ such that for every $n_1, n_2 \geq n_0$ we have $d(x_{n_1}, x_{n_2}) < \epsilon$.

Every converging sequence is Cauchy, but the converse is not true in general.

A metric space is said to be *complete* if every Cauchy sequence converges in it.

A closed subset of a complete space is always complete.

A complete subspace of an arbitrary metric space is always closed.

Every compact metric space is complete, but the converse is not true in general.

For every metric space (M,d), there exists a superspace $(\overline{M},\overline{d})$ containing M such that:

- $(\overline{M},\overline{d})$ is complete.
- M is dense in $(\overline{M},\overline{d})$.
- $\overline{d}(x,y) = d(x,y)$ for every $x, y \in M$.

The space $(\overline{M},\overline{d})$ is said to be *a completion* of (M,d).

11.1.13 Polish and Baire Spaces

A topological space $(T,\mathcal{U})$ that is both separable and completely metrizable (i.e., has a metrization that is complete) is called a *Polish space*.

A topological space is said to be a *Baire space* if the intersection of countably many dense open subsets is dense. The following facts can be found in [77]:

- Every completely metrizable space is Baire.
- Every compact Hausdorff space is Baire.
- Every open subset of a Baire space is Baire.

11.1.14 Sequential Spaces

Sequential spaces were introduced by Franklin [78] to answer the following question: Assume we know all the converging sequences of a topological space. Is this enough to uniquely determine the topology of the space? *Sequential spaces* are the most general category of spaces for which converging sequences suffice to determine the topology.

Let $(T,\mathcal{U})$ be a topological space. A subset $U \subset T$ is said to be *sequentially open* if for every sequence $(x_n)_{n\geq 0}$ that converges to a point of U lies eventually in U, i.e., there exists $n_0 \geq 0$ such that $x_n \in U$ for every $n \geq n_0$. Clearly, every open subset of T is sequentially open, but the converse is not true in general.

A topological space $(T,\mathcal{U})$ is said to be *sequential* if every sequentially open subset of T is open.

A mapping $f : T \to S$ from a sequential topological space $(T,\mathcal{U})$ to an arbitrary topological space $(S,\mathcal{V})$ is continuous if and only if for every sequence $(x_n)_{n\geq 0}$ in T that converges to $x \in T$, the sequence $(f(x_n))_{n\geq 0}$ converges to $f(x)$ in $(S,\mathcal{V})$ [78].

The following facts were shown in [78]:

- Every first-countable space is sequential. Therefore, every metrizable space is sequential.
- The quotient of a sequential space is sequential.
- All closed and open subsets of a sequential space are sequential.
- Every countably compact sequential Hausdorff space is sequentially compact.
- A topological space is sequential if and only if it is the quotient of a metric space.

11.1.15 Compactly Generated Spaces

A topological space $(T,\mathcal{U})$ is *compactly generated* if it is Hausdorff and for every subset F of T, F is closed if and only if $F \cap K$ is closed for every compact subset K of T. Equivalently, $(T,\mathcal{U})$ is *compactly generated* if it is Hausdorff and for every subset U of T, U is open in T if and only if $U \cap K$ is open in K for every compact subset K of T.

The following facts can be found in [79]:

- All locally compact Hausdorff spaces are compactly generated.
- All first-countable Hausdorff spaces are compactly generated. Therefore, every metrizable space is compactly generated.
- A Hausdorff quotient of a compactly generated space is compactly generated.
- If $(T,\mathcal{U})$ is compactly generated and $(S,\mathcal{V})$ is Hausdorff locally compact, then $(T \times S, \mathcal{U} \otimes \mathcal{V})$ is compactly generated.

11.1.16 The Hausdorff Metric

Let (M,d) be a metric space. Let $\mathcal{K}(M)$ be the set of compact subsets of M. The Hausdorff metric on $\mathcal{K}(M)$ is defined as:

$$\begin{aligned} d_H(K_1,K_2) &= \max\left\{\sup_{x_1\in K_1} d(x_1,K_2), \sup_{x_2\in K_2} d(x_2,K_1)\right\} \\ &= \max\left\{\sup_{x_1\in K_1}\inf_{x_2\in K_2} d(x_1,x_2), \sup_{x_2\in K_2}\inf_{x_1\in K_1} d(x_2,x_1)\right\}. \end{aligned}$$

11.2 Measure-Theoretic Notations

In this section, we introduce the measure-theoretic notations that we are using. We assume that the reader is familiar with the basic definitions and theorems of measure theory.

11.2.1 Probabilities on Finite Sets

If $\mathcal{X}$ is a finite set, we denote the set of probability distributions on $\mathcal{X}$ as $\Delta_{\mathcal{X}}$. Note that $\Delta_{\mathcal{X}}$ is an $(|\mathcal{X}|-1)$-dimensional simplex in $\mathbb{R}^{\mathcal{X}}$. We always endow $\Delta_{\mathcal{X}}$ with the total-variation distance and its induced topology. For every $p_1, p_2 \in \Delta_{\mathcal{X}}$, we have:

$$\|p_1 - p_2\|_{TV} = \frac{1}{2}\sum_{x\in\mathcal{X}} |p_1(x) - p_2(x)| = \frac{1}{2}\|p_1 - p_2\|_1.$$

Note that the total-variation topology on $\Delta_{\mathcal{X}}$ is the same as the one inherited from the Euclidean topology of $\mathbb{R}^{\mathcal{X}}$ by relativisation. Since $\Delta_{\mathcal{X}}$ is a closed and bounded subset of $\mathbb{R}^{\mathcal{X}}$, it is compact.

11.2.2 Borel Sets and the Support of a Measure

Let $(T,\mathcal{U})$ be a Hausdorff topological space. The *Borel σ-algebra* of $(T,\mathcal{U})$ is the σ-algebra generated by $\mathcal{U}$. We denote the Borel σ-algebra of $(T,\mathcal{U})$ as $\mathcal{B}(T,\mathcal{U})$. If the topology $\mathcal{U}$ is known from the context, we write $\mathcal{B}(T)$ to denote the Borel σ-algebra. The sets in $\mathcal{B}(T)$ are called the *Borel sets* of T.

The *support* of a probability measure $P \in \mathcal{P}(T,\mathcal{B}(T))$ is the set of all points $x \in T$ for which every neighborhood has a strictly positive measure:

$$\operatorname{supp}(P) = \{x \in T :\ P(O) > 0 \text{ for every neighborhood } O \text{ of } x\}.$$

If P is a probability measure on a Polish space, then $P\big(T \setminus \operatorname{supp}(P)\big) = 0$.

11.2.3 Convergence of Probability Measures and the weak-$*$ Topology

We have many notions of convergence of probability measures. If the measurable space does not have a topological structure, we have two notions of convergence:

- The *total-variation convergence*: We say that a sequence $(P_n)_{n\geq 0}$ of probability measures in $\mathcal{P}(M,\Sigma)$ converges in total-variation to $P \in \mathcal{P}(M,\Sigma)$ if and only if $\lim_{n\to\infty} \|P_n - P\|_{TV} = 0$.
- The *strong convergence*: We say that a sequence $(P_n)_{n\geq 0}$ in $\mathcal{P}(M,\Sigma)$ strongly converges to $P \in \mathcal{P}(M,\Sigma)$ if and only if $\lim_{n\to\infty} P_n(A) = P(A)$ for every $A \in \Sigma$.

Clearly, total-variation convergence implies strong convergence. The converse is not true in general. However, if we are working in the Borel σ-algebra of a Polish space T and $(P_n)_{n\geq 0}$ strongly converges to a finitely supported probability measure P, then

$$\begin{aligned}
&\|P_n - P\|_{TV} \\
&\quad= \sup_{B\in\mathcal{B}(T)} |P_n(B) - P(B)| \\
&\quad\leq \sup_{B\in\mathcal{B}(T)} \left(\big|P_n\big(B \setminus \operatorname{supp}(P)\big) - P\big(B \setminus \operatorname{supp}(P)\big)\big| + \sum_{x\in\operatorname{supp}(P)} |P_n(x) - P(x)| \right) \\
&\quad= \sup_{B\in\mathcal{B}(T)} \left(\big|P_n\big(B \setminus \operatorname{supp}(P)\big)\big| + \sum_{x\in\operatorname{supp}(P)} |P_n(x) - P(x)| \right)
\end{aligned}$$

hence,

$$\begin{aligned}\|P_n - P\|_{TV} &\leq \left|P_n\big(T \setminus \mathrm{supp}(P)\big)\right| + \sum_{x \in \mathrm{supp}(P)} |P_n(x) - P(x)| \\ &= \left|P_n\big(T \setminus \mathrm{supp}(P)\big) - P\big(T \setminus \mathrm{supp}(P)\big)\right| + \sum_{x \in \mathrm{supp}(P)} |P_n(x) - P(x)| \overset{n\to\infty}{\longrightarrow} 0,\end{aligned}$$

which implies that $(P_n)_{n\geq 0}$ also converges to P in total-variation. Therefore, in a Polish space, total-variation convergence and strong convergence to finitely supported probability measures are equivalent.

Let $(T,\mathcal{U})$ be a Hausdorff topological space. We say that a sequence $(P_n)_{n\geq 0}$ of probability measures in $\mathcal{P}(T,\mathcal{B}(T))$ weakly-$*$ converges to $P \in \mathcal{P}(T,\mathcal{B}(T))$ if and only if for every bounded and continuous function f from T to $\mathbb{R}$, we have

$$\lim_{n\to\infty} \int_T f \cdot dP_n = \int_T f \cdot dP.$$

Note that many authors call this notion "weak convergence" rather than weak-$*$ convergence. We will refrain from using the term "weak convergence" in order to be consistent with the functional analysis notation.

The *weak-$*$ topology* on $\mathcal{P}(T,\mathcal{B}(T))$ is the coarsest topology which makes the mappings

$$P \to \int_{\Delta_{\mathcal{X}}} f \cdot dP$$

continuous over $\mathcal{P}(T,\mathcal{B}(T))$, for every bounded and continuous function f from T to $\mathbb{R}$.

11.2.4 Metrization of the Weak-$*$ Topology

If $(T,\mathcal{U})$ is a Polish space (i.e., separable and completely metrizable), then the weak-$*$ topology on $\mathcal{P}(T,\mathcal{B}(T))$ is also Polish [80]. There are many known metrizations for the weak-$*$ topology. One metrization that is particularly convenient for us is the Wasserstein metric.

The 1^{st}-Wasserstein distance on $\mathcal{P}(T,\mathcal{B}(T))$ is defined as

$$W_1(P,P') = \inf_{\gamma \in \Gamma(P,P')} \int_{T\times T} d(x,x') \cdot d\gamma(x,x'),$$

where $\Gamma(P,P')$ is the collection of all probability measures on $T \times T$ with marginals P and P' on the first and second factors respectively, and d is a metric on T that induces the topology $\mathcal{U}$. $\Gamma(P,P')$ is also called the set of *couplings* of P and P'.

If d is bounded and (T,d) is separable and complete, then W_1 metrizes the weak-$*$ topology [80]. If $(T,\mathcal{U})$ is compact, then $(\mathcal{P}(T),W_1)$ is also compact [80].

If $D = \sup_{x,x' \in T} d(x,x')$ is the diameter of (T,d), then $W_1(P,P') \leq D\|P - P'\|_{TV}$ [80]. In other words, the Wasserstein metric is controlled by total-variation.

11.3 The Space of Channels from $\mathcal{X}$ to $\mathcal{Y}$

Let $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ be the set of all channels having $\mathcal{X}$ as input alphabet and $\mathcal{Y}$ as output alphabet.

For every $W, W' \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, define the distance between W and W' as follows:

$$d_{\mathcal{X},\mathcal{Y}}(W, W') = \frac{1}{2} \max_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} |W'(y|x) - W(y|x)|.$$

It is easy to check the following properties of $d_{\mathcal{X},\mathcal{Y}}$:

- $0 \leq d_{\mathcal{X},\mathcal{Y}}(W, W') \leq 1$.
- $d_{\mathcal{X},\mathcal{Y}} : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \to \mathbb{R}^+$ is a metric distance on $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$.

Throughout this chapter, we always associate the space $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ with the metric distance $d_{\mathcal{X},\mathcal{Y}}$ and the metric topology $\mathcal{T}_{\mathcal{X},\mathcal{Y}}$ induced by it.

For every $x \in \mathcal{X}$, the mapping $y \to W(y|x)$ is a probability distributions on $\mathcal{Y}$. Therefore, every channel W can be seen as a collection of probability distributions on $\mathcal{Y}$, and the collection is indexed by $x \in \mathcal{X}$. This allows us to identify the space $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ with $(\Delta_{\mathcal{Y}})^{\mathcal{X}} = \prod_{x \in \mathcal{X}} \Delta_{\mathcal{Y}}$, where $\Delta_{\mathcal{Y}}$ is the set of probability distributions on $\mathcal{Y}$. It is easy to see that the topology given by the metric $d_{\mathcal{X},\mathcal{Y}}$ on $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ is the same as the product topology on $(\Delta_{\mathcal{Y}})^{\mathcal{X}}$, which is also the same as the topology inherited from the Euclidean topology of $\mathbb{R}^{\mathcal{X} \times \mathcal{Y}}$ by relativization.

It is known that $\Delta_{\mathcal{Y}}$ is a closed and bounded subset of $\mathbb{R}^{\mathcal{Y}}$. Therefore, $\Delta_{\mathcal{Y}}$ is compact, which implies that $(\Delta_{\mathcal{Y}})^{\mathcal{X}}$ is compact. We conclude that the metric space $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \equiv (\Delta_{\mathcal{Y}})^{\mathcal{X}}$ is compact. Moreover, since $\Delta_{\mathcal{Y}}$ a convex subset of $\mathbb{R}^{\mathcal{Y}}$, it is path-connected, hence $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \equiv (\Delta_{\mathcal{Y}})^{\mathcal{X}}$ is path-connected as well.

If $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ and $V \in \mathrm{DMC}_{\mathcal{Y},\mathcal{Z}}$, we define the composition $V \circ W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Z}}$ of W and V as follows:

$$(V \circ W)(z|x) = \sum_{y \in \mathcal{Y}} V(z|y) W(y|x), \quad \forall x \in \mathcal{X},\ \forall z \in \mathcal{Z}.$$

It is easy to see that the mapping $(W, V) \to V \circ W$ from $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \times \mathrm{DMC}_{\mathcal{Y},\mathcal{Z}}$ to $\mathrm{DMC}_{\mathcal{X},\mathcal{Z}}$ is continuous.

For every mapping $f : \mathcal{X} \to \mathcal{Y}$, define the *deterministic channel* $D_f \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ as follows:

$$D_f(y|x) = \begin{cases} 1 & \text{if } y = f(x), \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that if $f : \mathcal{X} \to \mathcal{Y}$ and $g : \mathcal{Y} \to \mathcal{Z}$, then $D_g \circ D_f = D_{g \circ f}$.

11.4 Space of Output-Equivalent Channels from $\mathcal{X}$ to $\mathcal{Y}$

11.4.1 The $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ Space

Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets. Define the relation $R^{(o)}_{\mathcal{X},\mathcal{Y}}$ on $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ as follows:

$$\forall W, W' \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}, \quad W R^{(o)}_{\mathcal{X},\mathcal{Y}} W' \ \Leftrightarrow \ W \text{ is output-equivalent to } W'.$$

It is easy to see that $R^{(o)}_{\mathcal{X},\mathcal{Y}}$ is an equivalence relation on $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$. $R^{(o)}_{\mathcal{X},\mathcal{Y}}$ is called the *output-equivalence relation* on $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$.

Definition 11.1. *The space of output-equivalent channels with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ is the quotient of the space of channels from $\mathcal{X}$ to $\mathcal{Y}$ by the output-equivalence relation:*

$$\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} = \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} / R^{(o)}_{\mathcal{X},\mathcal{Y}}.$$

We define the topology $\mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}}$ on $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ as the quotient topology $\mathcal{T}_{\mathcal{X},\mathcal{Y}}/R^{(o)}_{\mathcal{X},\mathcal{Y}}$.

Unless we explicitly state otherwise, we always associate $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ with the quotient topology $\mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}}$.

For every $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, let $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ be the $R^{(o)}_{\mathcal{X},\mathcal{Y}}$-equivalence class containing W.

Lemma 11.3. *The projection mapping* $\mathrm{Proj} : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \to \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ *defined as* $\mathrm{Proj}(W) = \hat{W}$ *is continuous and closed.*

Proof. See Appendix 11.10.1. □

Corollary 11.1. *For every $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, $\hat{W}$ is a compact subset of* $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$.

Proof. Since $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ is compact, then $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} = \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} / R^{(o)}_{\mathcal{X},\mathcal{Y}}$ is compact as well.

Let $\mathrm{Proj} : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \to \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ be as in Lemma 11.3. Since Proj is closed and since $\{W\}$ is closed in $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, $\{\hat{W}\} = \mathrm{Proj}(\{W\})$ is closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$. Therefore, $\hat{W} = \mathrm{Proj}^{-1}(\{\hat{W}\})$ is closed in $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ because Proj is continuous. Now since $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ is compact, $\hat{W}$ is compact as well. □

Theorem 11.3. $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ *is a compact, path-connected and metrizable space.*

Proof. Since $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ is compact and path-connected, $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} = \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} / R^{(o)}_{\mathcal{X},\mathcal{Y}}$ is compact and path-connected as well.

Since the projection map Proj of Lemma 11.3 is closed, Lemma 11.2 implies that the quotient space $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} = \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} / R^{(o)}_{\mathcal{X},\mathcal{Y}}$ is upper semi-continuous. On the other hand, Corollary 11.1 shows that all the members of $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ are compact in $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$. Therefore, the conditions of Theorem 11.1 are satisfied.

Since $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ is a metric space, it is Hausdorff and regular. Moreover, since it can be seen as a subspace of $\mathbb{R}^{|\mathcal{X}|\cdot|\mathcal{Y}|}$, it is also second-countable. By Theorem 11.1 we get that $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} = \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} / R^{(o)}_{\mathcal{X},\mathcal{Y}}$ is Hausdorff, regular and second-countable, and from Theorem 11.2 we conclude that $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ is separable and metrizable. □

11.4.2 Canonical Embedding and Canonical Identification

Let $\mathcal{X}, \mathcal{Y}_1$ and $\mathcal{Y}_2$ be three finite sets such that $|\mathcal{Y}_1| \leq |\mathcal{Y}_2|$. We will show that there is a canonical embedding from $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}$ to $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$. In other words, there exists an explicitly constructable compact subset A of $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$ such that A is homeomorphic to $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}$. A and the homeomorphism depend only on $\mathcal{X}, \mathcal{Y}_1$ and $\mathcal{Y}_2$ (this is why we say that they are canonical). Moreover, we can show that A depends only on $|\mathcal{Y}_1|$, $\mathcal{X}$ and $\mathcal{Y}_2$.

Lemma 11.4. *For every $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_1}$ and every injection f from $\mathcal{Y}_1$ to $\mathcal{Y}_2$, W is output-equivalent to $D_f \circ W$.*

Proof. Clearly $D_f \circ W$ is output-degraded from W. Now let f' be any mapping from $\mathcal{Y}_2$ to $\mathcal{Y}_1$ such that $f'(f(y_1)) = y_1$ for every $y_1 \in \mathcal{Y}_1$. We have $W = (D_{f'} \circ D_f) \circ W = D_{f'} \circ (D_f \circ W)$, and so W is also output-degraded from $D_f \circ W$. □

Corollary 11.2. *For every $W, W' \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_1}$ and every two injections f, g from $\mathcal{Y}_1$ to $\mathcal{Y}_2$, we have:*

$$W R^{(o)}_{\mathcal{X},\mathcal{Y}_1} W' \;\Leftrightarrow\; (D_f \circ W) R^{(o)}_{\mathcal{X},\mathcal{Y}_2} (D_g \circ W').$$

Proof. Since W is output-equivalent to $D_f \circ W$ and W' is output-equivalent to $D_g \circ W'$, then W is output-equivalent to W' if and only if $D_f \circ W$ is output-equivalent to $D_g \circ W'$. □

For every $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_1}$, we denote the $R^{(o)}_{\mathcal{X},\mathcal{Y}_1}$-equivalence class of W as $\hat{W}$, and for every $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2}$, we denote the $R^{(o)}_{\mathcal{X},\mathcal{Y}_2}$-equivalence class of W as $\tilde{W}$.

Proposition 11.1. *Let $f : \mathcal{Y}_1 \to \mathcal{Y}_2$ be any fixed injection from $\mathcal{Y}_1$ to $\mathcal{Y}_2$. Define the mapping $F : \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1} \to \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$ as $F(\hat{W}) = \widetilde{D_f \circ W'} = \mathrm{Proj}_2(D_f \circ W')$, where $W' \in \hat{W}$ and $\mathrm{Proj}_2 : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2} \to \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$ is the projection onto the $R^{(o)}_{\mathcal{X},\mathcal{Y}_2}$-equivalence classes. We have:*

- *F is well defined, i.e., $\mathrm{Proj}_2(D_f \circ W')$ does not depend on $W' \in \hat{W}$.*
- *F is a homeomorphism from $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}$ to $F\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}\big)$.*
- *F does not depend on f, i.e., F depends only on $\mathcal{X}, \mathcal{Y}_1$ and $\mathcal{Y}_2$, hence it is canonical.*
- *$F\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}\big)$ depends only on $|\mathcal{Y}_1|$, $\mathcal{X}$ and $\mathcal{Y}_2$.*
- *For every $W' \in \hat{W}$ and every $W'' \in F(\hat{W})$, W' is output-equivalent to W''.*

Proof. Corollary 11.2 implies that $\mathrm{Proj}_2(D_f \circ W) = \mathrm{Proj}_2(D_f \circ W')$ if and only if $W R^{(o)}_{\mathcal{X},\mathcal{Y}_1} W'$. Therefore, $\mathrm{Proj}_2(D_f \circ W')$ does not depend on $W' \in \hat{W}$, hence F is well defined. Corollary 11.2 also shows that $\mathrm{Proj}_2(D_f \circ W')$ does not depend on the particular choice of the injection f, hence it is canonical (i.e., it depends only on $\mathcal{X}, \mathcal{Y}_1$ and $\mathcal{Y}_2$).

On the other hand, the mapping $W \to D_f \circ W$ is a continuous mapping from $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}_1}$ to $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2}$, and Proj_2 is continuous. Therefore, the mapping $W \to \mathrm{Proj}_2(D_f \circ W)$ is a continuous mapping from $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}_1}$ to $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$. Now since $\mathrm{Proj}_2(D_f \circ W)$ depends only on the $R^{(o)}_{\mathcal{X},\mathcal{Y}_1}$-equivalence class $\hat{W}$ of W, Lemma 11.1 implies that F is continuous. Moreover, we can see from Corollary 11.2 that F is an injection.

For every closed subset B of $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}$, B is compact since $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}$ is compact, hence $F(B)$ is compact because F is continuous. This implies that $F(B)$ is closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$ since $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$ is Hausdorff (as it is metrizable). Therefore, F is a closed mapping.

Now since F is an injection that is both continuous and closed, we can deduce that F is a homeomorphism from $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}$ to $F\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}\big) \subset \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$.

We would like now to show that $F\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}\big)$ depends only on $|\mathcal{Y}_1|$, $\mathcal{X}$ and $\mathcal{Y}_2$. Let $\mathcal{Y}_1'$ be a finite set such that $|\mathcal{Y}_1| = |\mathcal{Y}_1'|$. For every $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_1'}$, let $\overline{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1'}$ be the $R^{(o)}_{\mathcal{X},\mathcal{Y}_1'}$-equivalence class of W.

Let $g : \mathcal{Y}_1' \to \mathcal{Y}_1$ be a fixed bijection from $\mathcal{Y}_1'$ to $\mathcal{Y}_1$ and let $f' = f \circ g$. Define $F' : \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1'} \to \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$ as $F'(\overline{W}) = \widetilde{D_{f'} \circ W'} = \mathrm{Proj}_2(D_{f'} \circ W')$, where $W' \in \overline{W}$. As above, F' is well defined, and it is a homeomorphism from $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1'}$ to $F'\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1'}\big)$. We want to show that $F'\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1'}\big) = F\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}\big)$. For every $\overline{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1'}$, let $W' \in \overline{W}$. We have

$$F'(\overline{W}) = \mathrm{Proj}_2(D_{f'} \circ W') = \mathrm{Proj}_2(D_f \circ (D_g \circ W')) = F\left(\widetilde{D_g \circ W'}\right) \in F\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}\big).$$

Since this is true for every $\overline{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1'}$, we deduce that $F'\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1'}\big) \subset F\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}\big)$. By exchanging the roles of $\mathcal{Y}_1$ and $\mathcal{Y}_1'$ and using the fact that $f = f' \circ g^{-1}$, we get $F\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}\big) \subset F'\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1'}\big)$. We conclude that $F\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}\big) = F'\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1'}\big)$, which means that $F\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}\big)$ depends only on $|\mathcal{Y}_1|$, $\mathcal{X}$ and $\mathcal{Y}_2$.

Finally, for every $W' \in \hat{W}$ and every $W'' \in F(\hat{W}) = \widetilde{D_f \circ W'}$, W'' is output-equivalent to $D_f \circ W'$ and $D_f \circ W'$ is output-equivalent to W' (by Lemma 11.4), hence W'' is output-equivalent to W'. □

Corollary 11.3. *If $|\mathcal{Y}_1| = |\mathcal{Y}_2|$, there exists a canonical homeomorphism from $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}$ to $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$ depending only on $\mathcal{X}, \mathcal{Y}_1$ and $\mathcal{Y}_2$.*

Proof. Let f be a bijection from $\mathcal{Y}_1$ to $\mathcal{Y}_2$. Define the mapping $F : \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1} \to \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$ as $F(\hat{W}) = \widetilde{D_f \circ W'} = \mathrm{Proj}_2(D_f \circ W')$, where $W' \in \hat{W}$ and $\mathrm{Proj}_2 : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2} \to \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$ is the projection onto the $R^{(o)}_{\mathcal{X},\mathcal{Y}_2}$-equivalence classes.

Also, define the mapping $F' : \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2} \to \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}$ as $F'(\tilde{V}) = \widehat{D_{f^{-1}} \circ V'} = \mathrm{Proj}_1(D_{f^{-1}} \circ V')$, where $V' \in \tilde{V}$ and $\mathrm{Proj}_1 : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_1} \to \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}$ is the projection onto the $R^{(o)}_{\mathcal{X},\mathcal{Y}_1}$-equivalence classes.

Proposition 11.1 shows that F and F' are well defined.

For every $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_1}$, we have:

$$F'(F(\hat{W})) \overset{(a)}{=} F'(\widetilde{D_f \circ W}) \overset{(b)}{=} D_{f^{-1}} \circ \widehat{(D_f \circ W)} = \hat{W},$$

where (a) follows from the fact that $W \in \hat{W}$ and (b) follows from the fact that $D_f \circ W \in \widetilde{D_f \circ W}$.

We can similarly show that $F(F'(\tilde{V})) = \tilde{V}$ for every $\tilde{V} \in \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$. Therefore, both F and F' are bijections. Proposition 11.1 now implies that F is a homeomorphism from $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}$ to $F\big(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}\big) = \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$. Moreover, F depends only on $\mathcal{X}, \mathcal{Y}_1$ and $\mathcal{Y}_2$. □

Corollary 11.3 allows us to identify $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}$ with $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$ whenever $|\mathcal{Y}_1| = |\mathcal{Y}_2|$. In the rest of this chapter, we identify $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ with $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ through the canonical identification, where $n = |\mathcal{Y}|$ and $[n] = \{1, \dots, n\}$.

Moreover, for every $1 \leq n \leq m$, Proposition 11.1 allows us to identify $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ with the canonical subspace of $\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ that is homeomorphic to $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. In the rest of this chapter, we consider that $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is a compact subspace of $\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$.

Intuitively, $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ has a "lower dimension" compared to $\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$. So one expects that the interior of $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ in $(\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}, \mathcal{T}^{(o)}_{\mathcal{X},[m]})$ is empty if $m > n$. The following proposition shows that this intuition is accurate.

Proposition 11.2. *If $|\mathcal{X}| \geq 2$, then for every $1 \leq n < m$, the interior of $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ in $(\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}, \mathcal{T}^{(o)}_{\mathcal{X},[m]})$ is empty.*

Proof. See Appendix 11.10.2. □

11.5 Spaces of Output-Equivalent Channels

We would like to form the space of all output-equivalent channels having the same input alphabet $\mathcal{X}$. The previous section showed that if $|\mathcal{Y}_1| = |\mathcal{Y}_2|$, there is a canonical identification between $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1}$ and $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$. This shows that if we are interested in output-equivalent channels, it is sufficient to study the spaces $\mathrm{DMC}_{\mathcal{X},[n]}$ and $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ for every $n \geq 1$. Define the space

$$\mathrm{DMC}_{\mathcal{X},*} = \coprod_{n \geq 1} \mathrm{DMC}_{\mathcal{X},[n]}\,.$$

The subscript $*$ indicates that the output alphabets of the considered channels are arbitrary but finite.

We define the output-equivalence relation $R^{(o)}_{\mathcal{X},*}$ on $\mathrm{DMC}_{\mathcal{X},*}$ as follows:

$$\forall W, W' \in \mathrm{DMC}_{\mathcal{X},*},\ \ W R^{(o)}_{\mathcal{X},*} W' \ \Leftrightarrow\ W \text{ is output-equivalent to } W'.$$

Definition 11.2. *The space of output-equivalent channels with input alphabet $\mathcal{X}$ is the quotient of the space of channels with input alphabet $\mathcal{X}$ by the output-equivalence relation:*

$$\mathrm{DMC}^{(o)}_{\mathcal{X},*} = \mathrm{DMC}_{\mathcal{X},*} / R^{(o)}_{\mathcal{X},*}.$$

For every $n \geq 1$ and every $W, W' \in \mathrm{DMC}_{\mathcal{X},[n]}$, we have $W R^{(o)}_{\mathcal{X},*} W'$ if and only if $W R^{(o)}_{\mathcal{X},[n]} W'$ by definition. Therefore, $\mathrm{DMC}_{\mathcal{X},[n]} / R^{(o)}_{\mathcal{X},*}$ can be canonically identified with $\mathrm{DMC}_{\mathcal{X},[n]} / R^{(o)}_{\mathcal{X},[n]} = \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. But since we identified $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ to its image through the canonical embedding in $\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ for every $m \geq n$, we have to make sure that these identifications are consistent with each other.

Remember that for every $m \geq n \geq 1$ and every $W \in \mathrm{DMC}_{\mathcal{X},[n]}$, we identified $\hat{W}$ with $\widetilde{D_f \circ W}$, where f is any injection from $[n]$ to $[m]$, $\hat{W}$ is the $R^{(o)}_{\mathcal{X},[n]}$-equivalence class of W and $\widetilde{D_f \circ W}$ is the $R^{(o)}_{\mathcal{X},[m]}$-equivalence class of $D_f \circ W$. Since $D_f \circ W$ is output-equivalent to W (by Lemma 11.4), W is $R^{(o)}_{\mathcal{X},*}$-equivalent to $D_f \circ W$ for every $W \in \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. We conclude that identifying $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ to its image through the canonical embedding in $\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ for every $m \geq n \geq 1$ is consistent with identifying $\mathrm{DMC}_{\mathcal{X},[n]} / R^{(o)}_{\mathcal{X},*}$ to $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ for every $n \geq 1$. Hence, we can write

$$\mathrm{DMC}^{(o)}_{\mathcal{X},*} = \bigcup_{n \geq 1} \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}.$$

For any $W, W' \in \mathrm{DMC}_{\mathcal{X},*}$, Proposition 10.2 shows that $W R^{(o)}_{\mathcal{X},*} W'$ if and only if $\mathrm{MP}_W = \mathrm{MP}_{W'}$. Therefore, for every $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$, we can define the Blackwell measure of $\hat{W}$ as $\mathrm{MP}_{\hat{W}} := \mathrm{MP}_{W'}$ for any $W' \in \hat{W}$. We also define the rank of $\hat{W}$ as $\mathrm{rank}(\hat{W}) = |\operatorname{supp}(\mathrm{MP}_{\hat{W}})|$. Due to Proposition 10.2, we have

$$\mathrm{DMC}^{(o)}_{\mathcal{X},[n]} = \{\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},*} : \ \mathrm{rank}(\hat{W}) \leq n\}.$$

A subset A of $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is said to be *rank-bounded* if there exists $n \geq 1$ such that $A \subset \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. A is *rank-unbounded* if it is not rank-bounded.

11.5.1 Natural Topologies on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$

Since $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is the quotient of $\mathrm{DMC}_{\mathcal{X},*}$ and since $\mathrm{DMC}_{\mathcal{X},*}$ was not given any topology, there is no "standard topology" on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$.

However, there are many properties that one may require from any "reasonable" topology on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$. For example, one may require the continuity of all mappings that are relevant to information theory such as capacity, mutual information, probability of error of any fixed code, optimal probability of error of a given rate and blocklength, channel sums and products, etc ... The continuity of these mappings under different topologies on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is studied in Chapter 12.

In this chapter, we focus on one particular requirement that we consider the most basic property required from any "acceptable" topology on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$:

Definition 11.3. *A topology $\mathcal{T}$ on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is said to be* natural *if it induces the quotient topology $\mathcal{T}^{(o)}_{\mathcal{X},[n]}$ on $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ for every $n \geq 1$.*

The reason why we consider such topology as natural is because $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is subset of $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ and the quotient topology $\mathcal{T}^{(o)}_{\mathcal{X},[n]}$ is the "standard" and "most natural" topology on $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. Therefore, we do not want to induce any non-standard topology on $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ by relativization.

Before discussing any particular natural topology, we would like to discuss a few properties that are common to all natural topologies.

Proposition 11.3. *Every natural topology on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is σ-compact, separable and path-connected.*

Proof. Since $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is the countable union of compact and separable subspaces (namely $\{\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}\}_{n\geq 1}$), $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is σ-compact and separable.

On the other hand, since $\bigcap_{n\geq 1} \mathrm{DMC}^{(o)}_{\mathcal{X},[n]} = \mathrm{DMC}^{(o)}_{\mathcal{X},[1]} \neq \emptyset$ and since $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is path-connected for every $n \geq 1$, the union $\mathrm{DMC}^{(o)}_{\mathcal{X},*} = \bigcup_{n\geq 1} \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is path-connected. □

Proposition 11.4. *If $|\mathcal{X}| \geq 2$ and $\mathcal{T}$ is a natural topology, every non-empty open set is rank-unbounded.*

Proof. Assume to the contrary that there exists a non-empty open set $U \in \mathcal{T}$ such that $U \subset \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ for some $n \geq 1$. $U \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n+1]}$ is open in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n+1]}$ because $\mathcal{T}$ is natural. On the other hand, $U \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n+1]} \subset U \subset \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. Proposition 11.2 now implies that $U \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n+1]} = \emptyset$. Therefore,

$$U = U \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]} \subset U \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n+1]} = \emptyset,$$

which is a contradiction. □

Corollary 11.4. *If $|\mathcal{X}| \geq 2$ and $\mathcal{T}$ is a natural topology, then for every $n \geq 1$, the interior of $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T})$ is empty.*

Proposition 11.5. *If $|\mathcal{X}| \geq 2$ and $\mathcal{T}$ is a Hausdorff natural topology, then the space $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T})$ is not a Baire space.*

Proof. Fix $n \geq 1$. Since $\mathcal{T}$ is natural, $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is a compact subset of $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T})$. But $\mathcal{T}$ is Hausdorff, so $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is a closed subset of $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T})$. Therefore, $\mathrm{DMC}^{(o)}_{\mathcal{X},*} \setminus \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is open.

On the other hand, Corollary 11.4 shows that the interior of $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ in the space $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T})$ is empty. Therefore, $\mathrm{DMC}^{(o)}_{\mathcal{X},*} \setminus \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is dense in the space $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T})$.

Now since

$$\bigcap_{n\geq 1}\left(\mathrm{DMC}^{(o)}_{\mathcal{X},*}\setminus\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}\right)=\mathrm{DMC}^{(o)}_{\mathcal{X},*}\setminus\left(\bigcup_{n\geq 1}\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}\right)=\emptyset,$$

and since $\mathrm{DMC}^{(o)}_{\mathcal{X},*}\setminus\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is open and dense in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*},\mathcal{T})$ for every $n\geq 1$, we conclude that $(\mathrm{DMC}^{(o)}_{\mathcal{X},*},\mathcal{T})$ is not a Baire space. □

Corollary 11.5. *If $|\mathcal{X}|\geq 2$, no natural topology on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ can be completely metrizable.*

Proof. The corollary follows from Proposition 11.5 and the fact that every completely metrizable topology is both Hausdorff and Baire. □

Proposition 11.6. *If $|\mathcal{X}|\geq 2$ and $\mathcal{T}$ is a Hausdorff natural topology, then the space $(\mathrm{DMC}^{(o)}_{\mathcal{X},*},\mathcal{T})$ is not locally compact anywhere, i.e., for every $\hat{W}\in\mathrm{DMC}^{(o)}_{\mathcal{X},*}$, there is no compact neighborhood of $\hat{W}$ in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*},\mathcal{T})$.*

Proof. Assume to the contrary that there exists a compact neighborhood K of $\hat{W}$. There exists an open set U such that $\hat{W}\in U\subset K$.

Since K is compact and Hausdorff, it is a Baire space. Moreover, since U is an open subset of K, U is also a Baire space.

Fix $n\geq 1$. Since the interior of $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*},\mathcal{T})$ is empty, the interior of $U\cap\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ in U is also empty. Therefore, $U\setminus\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is dense in U. On the other hand, since $\mathcal{T}$ is natural, $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is compact which implies that it is closed because $\mathcal{T}$ is Hausdorff. Therefore, $U\setminus\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is open in U. Now since

$$\bigcap_{n\geq 1}\left(U\setminus\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}\right)=U\setminus\left(\bigcup_{n\geq 1}\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}\right)=\emptyset,$$

and since $U\setminus\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is open and dense in U for every $n\geq 1$, U is not Baire, which is a contradiction. Therefore, there is no compact neighborhood of $\hat{W}$ in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*},\mathcal{T})$. □

11.5.2 Strong Topology on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$

The first natural topology that we study is the *strong topology* $\mathcal{T}^{(o)}_{s,\mathcal{X},*}$ on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$, which is the finest natural topology.

Since the spaces $\{\mathrm{DMC}_{\mathcal{X},[n]}\}_{n\geq 1}$ are disjoint and since there is no a priori way to (topologically) compare channels in $\mathrm{DMC}_{\mathcal{X},[n]}$ with channels in $\mathrm{DMC}_{\mathcal{X},[n']}$ for $n\neq n'$, the "most natural" topology that we can define on $\mathrm{DMC}_{\mathcal{X},*}$ is the disjoint union topology $\mathcal{T}_{s,\mathcal{X},*}:=\bigoplus_{n\geq 1}\mathcal{T}_{\mathcal{X},[n]}$. Clearly, the space $(\mathrm{DMC}_{\mathcal{X},*},\mathcal{T}_{s,\mathcal{X},*})$ is disconnected. Moreover, $\mathcal{T}_{s,\mathcal{X},*}$ is metrizable because it is the disjoint union of metrizable spaces. It is also σ-compact because it is the union of countably many compact spaces.

We added the subscript s to emphasize the fact that $\mathcal{T}_{s,\mathcal{X},*}$ is a strong topology (remember that the disjoint union topology is the *finest* topology that makes the canonical injections continuous).

Definition 11.4. *We define the strong topology* $\mathcal{T}^{(o)}_{s,\mathcal{X},*}$ *on* $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ *as the quotient topology* $\mathcal{T}_{s,\mathcal{X},*}/R^{(o)}_{\mathcal{X},*}$.

We call open and closed sets in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ *as strongly open and strongly closed sets respectively.*

Let $\mathrm{Proj} : \mathrm{DMC}_{\mathcal{X},*} \to \mathrm{DMC}^{(o)}_{\mathcal{X},*}$ be the projection onto the $R^{(o)}_{\mathcal{X},*}$-equivalence classes, and for every $n \geq 1$ let $\mathrm{Proj}_n : \mathrm{DMC}_{\mathcal{X},[n]} \to \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ be the projection onto the $R^{(o)}_{\mathcal{X},[n]}$-equivalence classes. Due to the identifications that we made at the beginning of Section 11.5, we have $\mathrm{Proj}(W) = \mathrm{Proj}_n(W)$ for every $W \in \mathrm{DMC}_{\mathcal{X},[n]}$. Therefore, for every $U \subset \mathrm{DMC}^{(o)}_{\mathcal{X},*}$, we have

$$\mathrm{Proj}^{-1}(U) = \coprod_{n \geq 1} \mathrm{Proj}_n^{-1}(U \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}).$$

Hence,

$$\begin{aligned}
U \in \mathcal{T}^{(o)}_{s,\mathcal{X},*} &\overset{(a)}{\Leftrightarrow} \mathrm{Proj}^{-1}(U) \in \mathcal{T}_{s,\mathcal{X},*} \\
&\overset{(b)}{\Leftrightarrow} \mathrm{Proj}^{-1}(U) \cap \mathrm{DMC}_{\mathcal{X},[n]} \in \mathcal{T}_{\mathcal{X},[n]}, \quad \forall n \geq 1 \\
&\Leftrightarrow \left(\coprod_{n' \geq 1} \mathrm{Proj}_{n'}^{-1}(U \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n']}) \right) \cap \mathrm{DMC}_{\mathcal{X},[n]} \in \mathcal{T}_{\mathcal{X},[n]}, \quad \forall n \geq 1 \\
&\Leftrightarrow \mathrm{Proj}_n^{-1}(U \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}) \in \mathcal{T}_{\mathcal{X},[n]}, \quad \forall n \geq 1 \\
&\overset{(c)}{\Leftrightarrow} U \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]} \in \mathcal{T}^{(o)}_{\mathcal{X},[n]}, \quad \forall n \geq 1,
\end{aligned}$$

where (a) and (c) follow from the properties of the quotient topology, and (b) follows from the properties of the disjoint union topology.

We conclude that $U \subset \mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is strongly open in $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ if and only if $U \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is open in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ for every $n \geq 1$. This shows that the topology on $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ that is inherited from $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ is exactly $\mathcal{T}^{(o)}_{\mathcal{X},[n]}$. Therefore, $\mathcal{T}^{(o)}_{s,\mathcal{X},*}$ is a natural topology. On the other hand, if $\mathcal{T}$ is an arbitrary natural topology and $U \in \mathcal{T}$, then $U \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is open in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ for every $n \geq 1$, so $U \in \mathcal{T}^{(o)}_{s,\mathcal{X},*}$. We conclude that $\mathcal{T}^{(o)}_{s,\mathcal{X},*}$ is the finest natural topology.

We can also characterize the strongly closed subsets of $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ in terms of the

closed sets of the $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ spaces:

$$
\begin{aligned}
&F \text{ is strongly closed in } \mathrm{DMC}^{(o)}_{\mathcal{X},*} \\
&\quad\Leftrightarrow\ \mathrm{DMC}^{(o)}_{\mathcal{X},*} \setminus F \text{ is strongly open in } \mathrm{DMC}^{(o)}_{\mathcal{X},*} \\
&\quad\Leftrightarrow\ \left(\mathrm{DMC}^{(o)}_{\mathcal{X},*} \setminus F\right) \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]} \text{ is open in } \mathrm{DMC}^{(o)}_{\mathcal{X},[n]},\ \ \forall n \geq 1 \\
&\quad\Leftrightarrow\ \mathrm{DMC}^{(o)}_{\mathcal{X},[n]} \setminus \left(F \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}\right) \text{ is open in } \mathrm{DMC}^{(o)}_{\mathcal{X},[n]},\ \ \forall n \geq 1 \\
&\quad\Leftrightarrow\ F \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]} \text{ is closed in } \mathrm{DMC}^{(o)}_{\mathcal{X},[n]},\ \ \forall n \geq 1.
\end{aligned}
$$

Since $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is metrizable for every $n \geq 1$, it is also normal. We can use this fact to prove that the strong topology on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is normal:

Lemma 11.5. $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ *is normal.*

Proof. See Appendix 11.10.3. □

The following theorem shows that the strong topology satisfies a few desirable properties.

Theorem 11.4. $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ *is a compactly generated, sequential and* T_4 *space.*

Proof. Since $(\mathrm{DMC}_{\mathcal{X},*}, \mathcal{T}_{s,\mathcal{X},*})$ is metrizable, it is sequential. Therefore, the space $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$, which is the quotient of a sequential space, is sequential.

Let us now show that $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is T_4. Fix $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$. For every $n \geq 1$, $\{\hat{W}\} \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is either $\{\hat{W}\}$ or $\emptyset$ depending on whether $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ or not. Since $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is metrizable, it is T_1 and so singletons are closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. We conclude that in all cases, $\{\hat{W}\} \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ for every $n \geq 1$. Therefore, $\{\hat{W}\}$ is strongly closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$. This shows that $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ is T_1. On the other hand, Lemma 11.5 shows that $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ is normal. This means that $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ is T_4, which implies that it is Hausdorff.

Now since $(\mathrm{DMC}_{\mathcal{X},*}, \mathcal{T}_{s,\mathcal{X},*})$ is metrizable, it is compactly generated. On the other hand, the quotient space $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ was shown to be Hausdorff. We conclude that $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ is compactly generated. □

Corollary 11.6. *If* $|\mathcal{X}| \geq 2$, $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ *is not locally compact anywhere.*

Proof. Since $\mathcal{T}^{(o)}_{s,\mathcal{X},*}$ is a natural Hausdorff topology, Proposition 11.6 implies that $\mathcal{T}^{(o)}_{s,\mathcal{X},*}$ is not locally compact anywhere. □

Although $(\mathrm{DMC}_{\mathcal{X},*}, \mathcal{T}_{s,\mathcal{X},*})$ is second-countable (because it is a σ-compact metrizable space), the quotient space $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ is not second-countable. In fact, we will show later that $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ fails to be first-countable (and hence it is

not metrizable). This is one manifestation of the strength of the topology $\mathcal{T}^{(o)}_{s,\mathcal{X},*}$. In order to show that $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ is not first-countable, we need to characterize the converging sequences in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$.

A sequence $(\hat{W}_n)_{n\geq 1}$ in $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is said to be *rank-bounded* if $\mathrm{rank}(\hat{W}_n)$ is bounded. $(\hat{W}_n)_{n\geq 1}$ is *rank-unbounded* if it is not bounded.

The following proposition shows that every rank-unbounded sequence does not converge in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$.

Proposition 11.7. *A sequence $(\hat{W}_n)_{n\geq 0}$ converges in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ if and only if there exists $m \geq 1$ such that $\hat{W}_n \in \mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ for every $n \geq 0$, and $(\hat{W}_n)_{n\geq 0}$ converges in $(\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}, \mathcal{T}^{(o)}_{\mathcal{X},[m]})$.*

Proof. Assume that a sequence $(\hat{W}_n)_{n\geq 0}$ in $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is rank-unbounded. This cannot happen unless $|\mathcal{X}| \geq 2$. In order to show that $(\hat{W}_n)_{n\geq 0}$ does not converge, it is sufficient to show that there exists a subsequence of $(\hat{W}_n)_{n\geq 0}$ which does not converge.

Let $(\hat{W}_{n_k})_{k\geq 0}$ be any subsequence of $(\hat{W}_n)_{n\geq 0}$ where the rank strictly increases, i.e., $\mathrm{rank}(W_{n_k}) < \mathrm{rank}(W_{n_{k'}})$ for every $0 \leq k < k'$. We will show that $(\hat{W}_{n_k})_{k\geq 0}$ does not converge.

Assume to the contrary that $(\hat{W}_{n_k})_{k\geq 0}$ converges to $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$. Define the set

$$A = \{\hat{W}_{n_k} : k \geq 0\} \setminus \{\hat{W}\}.$$

For every $m \geq 1$, the set $A \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ contains finitely many points. This means that $A \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ is a finite union of singletons (which are closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$), hence $A \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ is closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ for every $m \geq 1$. Therefore A is closed in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$.

Now define $U = \mathrm{DMC}^{(o)}_{\mathcal{X},*} \setminus A$. Since A is strongly closed, U is strongly open. Moreover, U contains $\hat{W}$, so U is a neighborhood of $\hat{W}$. Therefore, there exists $k_0 \geq 0$ such that $\hat{W}_{n_k} \in U$ for every $k \geq k_0$. Now since the rank of $(\hat{W}_{n_k})_{k\geq 0}$ strictly increases, we can find $k \geq k_0$ such that $\mathrm{rank}(\hat{W}_{n_k}) > \mathrm{rank}(\hat{W})$. This means that $\hat{W}_{n_k} \neq \hat{W}$ and so $\hat{W}_{n_k} \in A$. Therefore, $\hat{W}_{n_k} \notin U$ which is a contradiction.

We conclude that every converging sequence in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ must be rank-bounded.

Now let $(\hat{W}_n)_{n\geq 0}$ be a rank-bounded sequence in $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$, i.e., there exists $m \geq 1$ such that $\hat{W}_n \in \mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ for every $n \geq 0$. If $(\hat{W}_n)_{n\geq 0}$ converges in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ then it converges in $\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ since $\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ is strongly closed.

Conversely, let us assume that $(\hat{W}_n)_{n\geq 0}$ converges in $(\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}, \mathcal{T}^{(o)}_{\mathcal{X},[m]})$ to $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$. Let O be any neighborhood of $\hat{W}$ in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$. There exists a strongly open set U such that $\hat{W} \in U \subset O$. Since $U \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ is open in $(\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}, \mathcal{T}^{(o)}_{\mathcal{X},[m]})$, there exists $n_0 > 0$ such that $\hat{W}_n \in U \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ for every

$n \geq n_0$. This implies that $\hat{W}_n \in O$ for every $n \geq n_0$. Therefore $(\hat{W}_n)_{n\geq 0}$ converges to $\hat{W}$ in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$. □

Corollary 11.7. *If $|\mathcal{X}| \geq 2$, $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ is not first-countable anywhere, i.e., for every $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$, there is no countable neighborhood basis of $\hat{W}$.*

Proof. Fix $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$ and assume to the contrary that $\hat{W}$ admits a countable neighborhood basis $\{O_n\}_{n\geq 1}$ in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$. For every $n \geq 1$, let U'_n be a strongly open set such that $\hat{W} \in U'_n \subset O_n$. Define $U_n = \bigcap_{i=1}^{n} U'_i$. U_n is strongly open because it is the intersection of finitely many strongly open sets. Moreover, $U_n \subset O_m$ for every $n \geq m$.

For every $n \geq 1$, Proposition 11.4 implies that U_n (which is non-empty and strongly open) is rank-unbounded, so it cannot be contained in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. Hence there exists $\hat{W}_n \in U_n$ such that $\hat{W}_n \notin \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$.

Since $\hat{W}_n \notin \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$, we have $\mathrm{rank}(\hat{W}_n) > n$ for every $n \geq 1$. Therefore, $(\hat{W}_n)_{n\geq 1}$ is rank-unbounded. Proposition 11.7 implies that $(\hat{W}_n)_{n\geq 1}$ does not converge in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$.

Now let O be a neighborhood of $\hat{W}$ in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$. Since $\{O_n\}_{n\geq 1}$ is a neighborhood basis for $\hat{W}$, there exists $n_0 \geq 1$ such that $O_{n_0} \subset O$. For every $n \geq n_0$, we have $\hat{W}_n \in U_n \subset O_{n_0} \subset O$. This means that $(\hat{W}_n)_{n\geq 1}$ converges to $\hat{W}$ in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ which is a contradiction. Therefore, $\hat{W}$ does not admit a countable neighborhood basis in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$. □

Compact Subspaces of $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$

It is well known that a compact subset of $\mathbb{R}$ is compact if and only if it is closed and bounded. The following proposition shows that a similar statement holds for $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$.

Proposition 11.8. *A subspace of $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ is compact if and only if it is rank-bounded and strongly closed.*

Proof. If $|\mathcal{X}| = 1$, all channels are output-equivalent to each other and so $\mathrm{DMC}^{(o)}_{\mathcal{X},*} = \mathrm{DMC}^{(o)}_{\mathcal{X},[1]}$ consists of a single point. Therefore, all subsets of $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ are rank-bounded, compact and strongly closed.

Assume now that $|\mathcal{X}| \geq 2$. Let A be a subspace of $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$. If A is rank-bounded and strongly closed, then there exists $n \geq 1$ such that $A \subset \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. Since A is strongly closed, then $A = A \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ which is compact. Therefore, A is compact.

Now let A be a compact subspace of $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$. Since $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ is Hausdorff, A is strongly closed. It remains to show that A is rank-bounded.

Assume to the contrary that A is rank-unbounded. We can construct a sequence $(\hat{W}_n)_{n\geq 0}$ in A where the rank is strictly increasing, i.e., $\mathrm{rank}(\hat{W}_n) < \mathrm{rank}(\hat{W}_{n'})$ for

every $0 \leq n < n'$. Since the rank of $(\hat{W}_n)_{n\geq 0}$ is strictly increasing, every subsequence of $(\hat{W}_n)_{n\geq 0}$ is rank-unbounded. Proposition 11.7 implies that every subsequence of $(\hat{W}_n)_{n\geq 0}$ does not converge in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$. On the other hand, we have:

- A is countably compact because it is compact.
- Since A is strongly closed and since $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ is a sequential space, A is sequential.
- A is Hausdorff because $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ is Hausdorff.

Now since every countably compact sequential Hausdorff space is sequentially compact [78], A must be sequentially compact. Therefore, $(\hat{W}_n)_{n\geq 0}$ has a converging subsequence which is a contradiction. We conclude that A must be rank-bounded. □

11.5.3 The Noisiness Metric on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$

Theorem 11.3 implies that $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is metrizable for every $n \geq 1$. One might ask whether the spaces $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ are "simultaneously metrizable" in the sense that we can define a metric d_n on $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ for every $n \geq 1$ in such a way that d_n is the restriction of d_{n+1} for every $n \geq 1$. If this is the case, we can then define a metric on $\mathrm{DMC}^{(o)}_{\mathcal{X},*} = \bigcup_{n\geq 1} \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ as $d(\hat{W}, \hat{W}') = d_n(\hat{W}, \hat{W}')$ for any $n \geq 1$ satisfying $\hat{W}, \hat{W}' \in \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. In this section we will show that such metrics can be constructed.

Noisiness Metric on $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$

For every $m \geq 1$, let $\Delta_{[m]\times\mathcal{X}}$ be the space of probability distributions on $[m] \times \mathcal{X}$.

Let $\mathcal{Y}$ be a finite set and let $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$. For every $p \in \Delta_{[m]\times\mathcal{X}}$, define $P_c(p, W)$ as follows:

$$P_c(p, W) = \sup_{D\in\mathrm{DMC}_{\mathcal{Y},[m]}} \sum_{\substack{u\in[m],\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} p(u, x)W(y|x)D(u|y). \tag{11.1}$$

$P_c(p, W)$ can be interpreted as follows: Let (U, X) be a pair of random variables distributed according to p, send X through the channel W, and let Y be the output of W in such a way that $U - X - Y$ is a Markov chain. Let $\hat{U}$ be the estimate of U obtained by applying a random decoder $D \in \mathrm{DMC}_{\mathcal{Y},[m]}$. In this interpretation, p can be seen as a random encoder. The probability of correctly guessing U by using the decoder D is given by

$$\sum_{\substack{u\in[m],\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} p(u, x)W(y|x)D(u|y).$$

Therefore, $P_c(p, W)$ is the optimal probability of correctly guessing U from Y. Note that we can take the supremum in (11.1) over only deterministic channels $D \in \mathrm{DMC}_{\mathcal{Y},[m]}$ because we can always choose an optimal decoder that is deterministic.

It is well known that if W is output-degraded from W', then $P_c(p, W) \leq P_c(p, W')$ for every $p \in \Delta_{[m]\times\mathcal{X}}$ and every $m \geq 1$. It was shown in [70] that the converse is also true. Therefore, W is output-equivalent to W' if and only if $P_c(p, W) = P_c(p, W')$ for every $p \in \Delta_{[m]\times\mathcal{X}}$ and every $m \geq 1$. This shows that the quantity $P_c(p, W)$ depends only on the $R^{(o)}_{\mathcal{X},\mathcal{Y}}$-equivalence class of W. Therefore, if $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$, we can define $P_c(p, \hat{W}) := P_c(p, W')$ for any $W' \in \hat{W}$.

Define the *noisiness distance* $d^{(o)}_{\mathcal{X},\mathcal{Y}} : \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \times \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \to \mathbb{R}^+$ as follows:

$$d^{(o)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1, \hat{W}_2) = \sup_{\substack{m\geq 1,\\ p\in\Delta_{[m]\times\mathcal{X}}}} |P_c(p, \hat{W}_1) - P_c(p, \hat{W}_2)|.$$

It is easy to see that $0 \leq d^{(o)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1, \hat{W}_2) \leq 1$ for every $\hat{W}_1, \hat{W}_2 \in \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$. Moreover, we have:

- $d^{(o)}_{\mathcal{X},\mathcal{Y}}(\hat{W}, \hat{W}) = 0$ for every $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$.
- For every $\hat{W}_1, \hat{W}_2 \in \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$, if $d^{(o)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1, \hat{W}_2) = 0$, then $P_c(p, \hat{W}_1) = P_c(p, \hat{W}_2)$ for every $p \in \Delta_{[m]\times\mathcal{X}}$ and every $m \geq 1$, which implies that the channels in $\hat{W}_1$ are output-equivalent to the channels in $\hat{W}_2$, hence $\hat{W}_1 = \hat{W}_2$.
- $d^{(o)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1, \hat{W}_2) = d^{(o)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_2, \hat{W}_1)$ for every $\hat{W}_1, \hat{W}_2 \in \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$.
- For every $\hat{W}_1, \hat{W}_2, \hat{W}_3 \in \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$, we have

$$d^{(o)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1, \hat{W}_3) \leq d^{(o)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1, \hat{W}_2) + d^{(o)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_2, \hat{W}_3).$$

This shows that $d^{(o)}_{\mathcal{X},\mathcal{Y}}$ is a metric on $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$. $d^{(o)}_{\mathcal{X},\mathcal{Y}}$ is called the noisiness metric because it compares the "noisiness" of $\hat{W}_1$ with that of $\hat{W}_2$: If $P_c(p, \hat{W}_1)$ is close to $P_c(p, \hat{W}_2)$ for every random encoder p, then $\hat{W}_1$ and $\hat{W}_2$ have close "noisiness levels".

A natural question to ask is whether the metric topology on $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ that is induced by $d^{(o)}_{\mathcal{X},\mathcal{Y}}$ is the same as the quotient topology $\mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}}$ that we defined in Section 11.4.1. To answer this question, we need the following lemma.

Lemma 11.6. *For every $W_1, W_2 \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, we have:*

$$d^{(o)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1, \hat{W}_2) \leq d_{\mathcal{X},\mathcal{Y}}(W_1, W_2),$$

where $\hat{W}_1$ and $\hat{W}_2$ are the $R^{(o)}_{\mathcal{X},\mathcal{Y}}$-equivalence classes of W_1 and W_2 respectively.

Proof. See Appendix 11.10.4. □

Proposition 11.9. $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, d^{(o)}_{\mathcal{X},\mathcal{Y}})$ *and* $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}})$ *are topologically equivalent.*

Proof. Consider the projection mapping $\mathrm{Proj} : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \to \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ defined as $\mathrm{Proj}(W) = \hat{W}$, where $\hat{W}$ is the $R^{(o)}_{\mathcal{X},\mathcal{Y}}$-equivalence class of W.

Lemma 11.6 implies that Proj is a continuous mapping from $(\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}, d_{\mathcal{X},\mathcal{Y}})$ to $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, d^{(o)}_{\mathcal{X},\mathcal{Y}})$. Now since $\mathrm{Proj}(W) = \mathrm{Proj}(W')$ whenever $WR^{(o)}_{\mathcal{X},\mathcal{Y}}W'$, Lemma 11.1 implies that the identity mapping $id : \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \to \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ is continuous from $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}})$ to $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, d^{(o)}_{\mathcal{X},\mathcal{Y}})$. We have:

- For every $U \subset \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ that is open in $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, d^{(o)}_{\mathcal{X},\mathcal{Y}})$, $U = id^{-1}(U) \in \mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}}$ because id is a continuous mapping from $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}})$ to $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, d^{(o)}_{\mathcal{X},\mathcal{Y}})$.

- For every $U \in \mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}}$, the set $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \setminus U$ is closed in $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}})$ which is compact. Therefore, $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \setminus U$ is a compact subset of $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}})$. Now since id is continuous from $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}})$ to $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, d^{(o)}_{\mathcal{X},\mathcal{Y}})$, the set $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \setminus U = id(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \setminus U)$ is a compact subspace of $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, d^{(o)}_{\mathcal{X},\mathcal{Y}})$ which is Hausdorff (because it is metric). This shows that $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \setminus U$ is closed in $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, d^{(o)}_{\mathcal{X},\mathcal{Y}})$, which implies that U is open in $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, d^{(o)}_{\mathcal{X},\mathcal{Y}})$.

We conclude that $U \subset \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ is open in $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, d^{(o)}_{\mathcal{X},\mathcal{Y}})$ if and only if it is open in $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}})$. □

Corollary 11.8. $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, d^{(o)}_{\mathcal{X},\mathcal{Y}})$ *is a compact path-connected metric space.*

The reader might be wondering why we considered and studied the quotient topology $\mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}}$ while it is possible to explicitly define a metric on the space $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$. There are two reasons:

- The definition of $d^{(o)}_{\mathcal{X},\mathcal{Y}}$ does not seem to be intuitive at the first sight and it is not clear why one would adopt it as a standard metric on $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$. Just being a metric is not convincing enough. On the other hand, the existence of a natural standard topology on $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ makes the quotient topology the most natural starting point.

- If one wants to show that a mapping $f : \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \to S$ is continuous from $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, d^{(o)}_{\mathcal{X},\mathcal{Y}})$ to a topological space $(S, \mathcal{V})$, it is much easier to prove it through the quotient topology $\mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}}$ rather than proving it directly using the metric $d^{(o)}_{\mathcal{X},\mathcal{Y}}$. Therefore, it is important to show the topological equivalence between $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, d^{(o)}_{\mathcal{X},\mathcal{Y}})$ and $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}})$.

It is worth mentioning that in the proof of Proposition 11.9, the only topological property of $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}})$ that we used is its compactness. This means that we do not need Lemma 11.3 to prove Theorem 11.3. An alternative proof of Theorem 11.3 would be to show the compactness and path-connectedness by inheriting those properties from $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, and then show that $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}})$ is topologically equivalent to $(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}, d^{(o)}_{\mathcal{X},\mathcal{Y}})$ as in Proposition 11.9.

The main reason why we restricted ourselves to topological methods in Section 11.4.1 is because they might be useful if one wants to generalize our results to spaces

of non-discrete channels. It might not be easy to find an explicit metric for those spaces, or even worse, those spaces might fail to be metrizable. Therefore, one might want to prove weaker topological properties such as being Hausdorff and/or regular. In such cases, the methods of Section 11.4.1 might be useful.

Noisiness Metric on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$

For every $\hat{W}_1, \hat{W}_2 \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$, define the *noisiness metric* on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ as follows:

$$d^{(o)}_{\mathcal{X},*}(\hat{W}, \hat{W}') := d^{(o)}_{\mathcal{X},[n]}(\hat{W}, \hat{W}') \text{ where } n \geq 1 \text{ satisfies } \hat{W}, \hat{W}' \in \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}.$$

$d^{(o)}_{\mathcal{X},*}(\hat{W}, \hat{W}')$ is well defined because $d^{(o)}_{\mathcal{X},[n]}(\hat{W}, \hat{W}')$ does not depend on $n \geq 1$ as long as $\hat{W}, \hat{W}' \in \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. We can also express $d^{(o)}_{\mathcal{X},*}$ as follows:

$$d^{(o)}_{\mathcal{X},*}(\hat{W}_1, \hat{W}_2) = \sup_{\substack{m \geq 1, \\ p \in \Delta_{[m] \times \mathcal{X}}}} |P_c(p, \hat{W}_1) - P_c(p, \hat{W}_2)|.$$

It is easy to see that $d^{(o)}_{\mathcal{X},*}$ is a metric on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$. Let $\mathcal{T}^{(o)}_{\mathcal{X},*}$ be the metric topology on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ that is induced by $d^{(o)}_{\mathcal{X},*}$. We call $\mathcal{T}^{(o)}_{\mathcal{X},*}$ the *noisiness topology* on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$.

Clearly, $\mathcal{T}^{(o)}_{\mathcal{X},*}$ is natural because the restriction of $d^{(o)}_{\mathcal{X},*}$ on $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is exactly $d^{(o)}_{\mathcal{X},[n]}$, and the topology induced by $d^{(o)}_{\mathcal{X},[n]}$ is $\mathcal{T}^{(o)}_{\mathcal{X},[n]}$. If $|\mathcal{X}| \geq 2$, Proposition 11.6 and Corollary 11.5 imply that $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, d^{(o)}_{\mathcal{X},*})$ is not complete nor locally compact.

Since $\mathcal{T}^{(o)}_{s,\mathcal{X},*}$ is the finest natural topology, $\mathcal{T}^{(o)}_{s,\mathcal{X},*}$ is finer than $\mathcal{T}^{(o)}_{\mathcal{X},*}$. On the other hand, if $|\mathcal{X}| \geq 2$, $\mathcal{T}^{(o)}_{\mathcal{X},*}$ is metrizable and $\mathcal{T}^{(o)}_{s,\mathcal{X},*}$ is not (because it is not first-countable). Therefore, if $|\mathcal{X}| \geq 2$, the strong topology $\mathcal{T}^{(o)}_{s,\mathcal{X},*}$ is strictly finer than the noisiness topology $\mathcal{T}^{(o)}_{\mathcal{X},*}$.

It is worth mentioning that Propositions 11.7 and 11.8 do not hold for the space $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{\mathcal{X},*})$. It is easy to find a rank-unbounded sequence $\{\hat{W}_n\}_{n \geq 0}$ which converges in $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{\mathcal{X},*})$ to a point $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$. The set $\{\hat{W}_n : n \geq 0\} \cup \{\hat{W}\}$ is clearly compact and rank-unbounded.

11.5.4 Topologies from Blackwell Measures

We saw at the beginning of Section 11.5 that for every $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$, a Blackwell measure $\mathrm{MP}_{\hat{W}}$ on $\Delta_{\mathcal{X}}$ is defined. Moreover, Proposition 10.2 implies that $\hat{W}$ is uniquely determined by $\mathrm{MP}_{\hat{W}}$. Therefore, each $R^{(o)}_{\mathcal{X},*}$-equivalence class in $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ can be identified with its Blackwell measure. On the other hand, Proposition 10.1 shows that the collection of Blackwell measures of the channels with input alphabet $\mathcal{X}$ is the same as the collection of balanced and finitely supported meta-probability measures on $\mathcal{X}$.

Therefore, the mapping $\hat{W} \to \mathrm{MP}_{\hat{W}}$ is a bijection from $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ to $\mathcal{MP}_{bf}(\mathcal{X})$. We call this mapping *the canonical bijection* from $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ to $\mathcal{MP}_{bf}(\mathcal{X})$. Similarly, the inverse mapping is called the *canonical bijection* from $\mathcal{MP}_{bf}(\mathcal{X})$ to $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$.

Since $\Delta_{\mathcal{X}}$ is a metric space, there are many standard ways to construct topologies on $\mathcal{MP}(\mathcal{X})$. If we choose any of these standard topologies on $\mathcal{MP}(\mathcal{X})$ and then relativize it to the subspace $\mathcal{MP}_{bf}(\mathcal{X})$, we can construct topologies on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ through the canonical bijection.

We saw in Section 11.2.3 that there are three topologies that can be constructed on $\mathcal{MP}(\mathcal{X})$: The total-variation topology, the strong convergence topology, and the weak-$*$ topology. But since every measure in $\mathcal{MP}_{bf}(\mathcal{X})$ is a finitely supported measure, strong convergence and total-variation convergence are equivalent in $\mathcal{MP}_{bf}(\mathcal{X})$ (see Section 11.2.3). Therefore, it is sufficient to study the total-variation topology and the weak-$*$ topology. We will start by studying the weak-$*$ topology.

Weak-$*$ Topology

We first note that in the case of binary input channels, the weak-$*$ topology is equivalent to the topology induced by the convergence in distribution of D-densities (or L-densities, or G-densities) that was defined in [69]. Note also that the weak-$*$ topology is equivalent to the topology that is induced by the Le Cam deficiency distance [75].

Consider the topology on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ that is obtained by transporting the weak-$*$ topology from $\mathcal{MP}_{bf}(\mathcal{X})$ to $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ through the canonical bijection F_{can}, i.e., we let $U \subset \mathrm{DMC}^{(o)}_{\mathcal{X},*}$ be open if and only if $F^{-1}_{\mathrm{can}}(U)$ is weakly-$*$ open. We will call this topology *the weak-$*$ topology on* $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$.

In this section, we show that the weak-$*$ topology is the same as the noisiness topology $\mathcal{T}^{(o)}_{\mathcal{X},*}$. We will show this using the Wasserstein metric.

Since $\Delta_{\mathcal{X}}$ is complete and separable, the 1^{st}-Wasserstein distance metrizes the weak-$*$ topology [80]. Therefore, in order to show that the weak-$*$ topology and the noisiness topology $\mathcal{T}^{(o)}_{\mathcal{X},*}$ are the same, it is sufficient to show that the canonical bijection F_{can} from $(\mathcal{MP}_{bf}(\mathcal{X}), W_1)$ to $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, d^{(o)}_{\mathcal{X},*})$ is a homeomorphism.

Note that since $\Delta_{\mathcal{X}}$ is compact, the metric space $(\mathcal{MP}(\mathcal{X}), W_1)$ is compact as well [80].

Lemma 11.7. *For every $\hat{W}, \hat{W}' \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$, we have*

$$d^{(o)}_{\mathcal{X},*}(\hat{W}, \hat{W}') \leq |\mathcal{X}| \cdot W_1(\mathrm{MP}_{\hat{W}}, \mathrm{MP}_{\hat{W}'}).$$

Proof. See Appendix 11.10.5. □

Lemma 11.7 can also be expressed as follows: For every $\mathrm{MP}, \mathrm{MP}' \in \mathcal{MP}_{bf}(\mathcal{X})$, we have $d^{(o)}_{\mathcal{X},*}(F_{\mathrm{can}}(\mathrm{MP}), F_{\mathrm{can}}(\mathrm{MP}')) \leq |\mathcal{X}| \cdot W_1(\mathrm{MP}, \mathrm{MP}')$. This shows that the canonical bijection F_{can} is continuous. Therefore, the weak-$*$ topology is at least as strong as $\mathcal{T}^{(o)}_{\mathcal{X},*}$. It remains to show that F^{-1}_{can} is continuous. One approach to prove the continuity of F^{-1}_{can} is to find a lower bound of $d^{(o)}_{\mathcal{X},*}(\hat{W}, \hat{W}')$ in terms of the Wasserstein metric, but this is tedious. We will follow another approach in order to show that the canonical bijection F_{can} is a homeomorphism. We need the following proposition:

Proposition 11.10. *The weak-$*$ closure of $\mathcal{MP}_{bf}(\mathcal{X})$ is $\mathcal{MP}_b(\mathcal{X})$.*

Proof. See appendix 11.10.6. □

Theorem 11.5. *The weak-$*$ topology on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is the same as the noisiness topology $\mathcal{T}^{(o)}_{\mathcal{X},*}$.*

Proof. Let $(\overline{\mathrm{DMC}}^{(o)}_{\mathcal{X},*}, \overline{d}^{(o)}_{\mathcal{X},*})$ be a completion of $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, d^{(o)}_{\mathcal{X},*})$. Since $\mathcal{MP}_b(\mathcal{X})$ is the weak-$*$ closure of $\mathcal{MP}_{bf}(\mathcal{X})$ (Proposition 11.10), we can extend the canonical bijection $F_{\mathrm{can}} : \mathcal{MP}_{bf}(\mathcal{X}) \to \mathrm{DMC}^{(o)}_{\mathcal{X},*}$ to a mapping $\overline{F} : \mathcal{MP}_b(\mathcal{X}) \to \overline{\mathrm{DMC}}^{(o)}_{\mathcal{X},*}$ as follows:

$$\overline{F}(\mathrm{MP}) = \lim_{n\to\infty} F_{\mathrm{can}}(\mathrm{MP}_n), \tag{11.2}$$

where $(\mathrm{MP}_n)_{n\geq 0}$ is any sequence in $\mathcal{MP}_{bf}(\mathcal{X})$ that converges to $\mathrm{MP} \in \mathcal{MP}_b(\mathcal{X})$, and where the limit in (11.2) is taken inside $\overline{\mathrm{DMC}}^{(o)}_{\mathcal{X},*}$. In order to show that $\overline{F}$ is well defined, we have to make sure that the limit in (11.2) exists and that it does not depend on the sequence $(\mathrm{MP}_n)_{n\geq 0}$.

Since the sequence $(\mathrm{MP}_n)_{n\geq 0}$ converges, it is a Cauchy sequence. Therefore, for every $\epsilon > 0$ there exists $n_0 > 0$ such that for every $n_1, n_2 \geq 1$ we have $W_1(\mathrm{MP}_{n_1}, \mathrm{MP}_{n_2}) < \dfrac{\epsilon}{|\mathcal{X}|}$. By Lemma 11.7, we have

$$\begin{aligned}\overline{d}^{(o)}_{\mathcal{X},*}(F_{\mathrm{can}}(\mathrm{MP}_{n_1}), F_{\mathrm{can}}(\mathrm{MP}_{n_2})) &= d^{(o)}_{\mathcal{X},*}(F_{\mathrm{can}}(\mathrm{MP}_{n_1}), F_{\mathrm{can}}(\mathrm{MP}_{n_2})) \\ &\leq |\mathcal{X}| \cdot W_1(\mathrm{MP}_{n_1}, \mathrm{MP}_{n_2}) < \epsilon.\end{aligned}$$

Therefore, $(F_{\mathrm{can}}(\mathrm{MP}_n))_{n\geq 0}$ is a Cauchy sequence in $(\overline{\mathrm{DMC}}^{(o)}_{\mathcal{X},*}, \overline{d}^{(o)}_{\mathcal{X},*})$ which is complete, hence the limit in (11.2) exists. Now assume that $(\mathrm{MP}'_n)_{n\geq 0}$ is another sequence in $\mathcal{MP}_{bf}(\mathcal{X})$ which converges to MP. We have:

$$\begin{aligned}\lim_{n\to\infty} \overline{d}^{(o)}_{\mathcal{X},*}\big(F_{\mathrm{can}}(\mathrm{MP}_n), F_{\mathrm{can}}(\mathrm{MP}'_n)\big) &= \lim_{n\to\infty} d^{(o)}_{\mathcal{X},*}\big(F_{\mathrm{can}}(\mathrm{MP}_n), F_{\mathrm{can}}(\mathrm{MP}'_n)\big) \\ &\overset{(a)}{\leq} \lim_{n\to\infty} |\mathcal{X}| \cdot W_1(\mathrm{MP}_n, \mathrm{MP}'_n) \overset{(b)}{=} 0,\end{aligned}$$

where (a) follows from Lemma 11.7 and (b) follows from the fact that $(\mathrm{MP}_n)_{n\geq 0}$ and $(\mathrm{MP}'_n)_{n\geq 0}$ converge to the same point. Therefore, the sequences $(F_{\mathrm{can}}(\mathrm{MP}_n))_{n\geq 0}$ and $(F_{\mathrm{can}}(\mathrm{MP}'_n))_{n\geq 0}$ converge to the same point in $\overline{\mathrm{DMC}}^{(o)}_{\mathcal{X},*}$. We conclude that $\overline{F}$ is well defined.

Now fix $\mathrm{MP}, \mathrm{MP}' \in \mathcal{MP}_b(\mathcal{X})$ and let $(\mathrm{MP}_n)_{n\geq 0}$ and $(\mathrm{MP}'_n)_{n\geq 0}$ be two sequences in $\mathcal{MP}_{bf}(\mathcal{X})$ that converge to MP and MP′ respectively. We have:

$$\begin{aligned}\overline{d}^{(o)}_{\mathcal{X},*}\left(\overline{F}(\mathrm{MP}), \overline{F}(\mathrm{MP}')\right) &= \overline{d}^{(o)}_{\mathcal{X},*}\left(\lim_{n\to\infty} F_{\mathrm{can}}(\mathrm{MP}_n), \lim_{n\to\infty} F_{\mathrm{can}}(\mathrm{MP}'_n)\right) \\ &\overset{(a)}{=} \lim_{n\to\infty} \overline{d}^{(o)}_{\mathcal{X},*}(F_{\mathrm{can}}(\mathrm{MP}_n), F_{\mathrm{can}}(\mathrm{MP}'_n)) \\ &= \lim_{n\to\infty} d^{(o)}_{\mathcal{X},*}(F_{\mathrm{can}}(\mathrm{MP}_n), F_{\mathrm{can}}(\mathrm{MP}'_n)) \\ &\overset{(b)}{\leq} \lim_{n\to\infty} |\mathcal{X}| \cdot W_1(\mathrm{MP}_n, \mathrm{MP}'_n) \overset{(c)}{=} |\mathcal{X}| \cdot W_1(\mathrm{MP}, \mathrm{MP}'),\end{aligned}$$

where (a) and (c) follow from the fact that metric distances are continuous, and (b) follows from Lemma 11.7. Therefore, $\overline{F}$ is continuous from $(\mathcal{MP}_b(\mathcal{X}), W_1)$ to $(\overline{\mathrm{DMC}}^{(o)}_{\mathcal{X},*}, \overline{d}^{(o)}_{\mathcal{X},*})$. Moreover, since $\mathcal{MP}_b(\mathcal{X})$ is weakly-$*$ closed in $\mathcal{MP}(\mathcal{X})$ which is compact, $\mathcal{MP}_b(\mathcal{X})$ is compact under the weak-$*$ topology. Therefore for every weakly-$*$ closed subset A of $\mathcal{MP}_b(\mathcal{X})$, A is compact and so $\overline{F}(A)$ is compact in $(\overline{\mathrm{DMC}}^{(o)}_{\mathcal{X},*}, \overline{d}^{(o)}_{\mathcal{X},*})$ which is Hausdorff. This implies that $\overline{F}(A)$ is closed in $(\overline{\mathrm{DMC}}^{(o)}_{\mathcal{X},*}, \overline{d}^{(o)}_{\mathcal{X},*})$ for every weakly-$*$ closed subset A of $\mathcal{MP}_b(\mathcal{X})$. Therefore, $\overline{F}$ is both continuous and closed. In particular, $\overline{F}(\mathcal{MP}_b(\mathcal{X}))$ is closed in $(\overline{\mathrm{DMC}}^{(o)}_{\mathcal{X},*}, \overline{d}^{(o)}_{\mathcal{X},*})$. But $\overline{F}(\mathcal{MP}_b(\mathcal{X})) \supset \overline{F}(\mathcal{MP}_{bf}(\mathcal{X})) = F_{\mathrm{can}}(\mathcal{MP}_{bf}(\mathcal{X})) = \mathrm{DMC}^{(o)}_{\mathcal{X},*}$, and $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is dense in $(\overline{\mathrm{DMC}}^{(o)}_{\mathcal{X},*}, \overline{d}^{(o)}_{\mathcal{X},*})$. Therefore, we must have $\overline{F}(\mathcal{MP}_b(\mathcal{X})) = \overline{\mathrm{DMC}}^{(o)}_{\mathcal{X},*}$. We conclude that $\overline{F}$ is a homeomorphism from $(\mathcal{MP}_b(\mathcal{X}), W_1)$ to $(\overline{\mathrm{DMC}}^{(o)}_{\mathcal{X},*}, \overline{d}^{(o)}_{\mathcal{X},*})$.

Now since $\overline{F}\big(\mathcal{MP}_{bf}(\mathcal{X})\big) = \mathrm{DMC}^{(o)}_{\mathcal{X},*}$, the restriction of $\overline{F}$ to $\mathcal{MP}_{bf}(\mathcal{X})$ is a homeomorphism from $(\mathcal{MP}_{bf}(\mathcal{X}), W_1)$ to $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, d^{(o)}_{\mathcal{X},*})$. But the restriction of $\overline{F}$ to $\mathcal{MP}_{bf}(\mathcal{X})$ is nothing but F_{can}. We conclude that the canonical bijection is a homeomorphism from $(\mathcal{MP}_{bf}(\mathcal{X}), W_1)$ to $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, d^{(o)}_{\mathcal{X},*})$. Therefore, the weak-$*$ topology on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is the same as the noisiness topology $\mathcal{T}^{(o)}_{\mathcal{X},*}$. □

Since $(\mathcal{MP}_b(\mathcal{X}), W_1)$ is homeomorphic to $(\overline{\mathrm{DMC}}^{(o)}_{\mathcal{X},*}, \overline{d}^{(o)}_{\mathcal{X},*})$, we can interpret this by saying that $\overline{\mathrm{DMC}}^{(o)}_{\mathcal{X},*}$ is the space of all output-equivalent channels with input alphabet $\mathcal{X}$ and arbitrary output alphabet (with arbitrary cardinality). Moreover, since $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is dense in $(\overline{\mathrm{DMC}}^{(o)}_{\mathcal{X},*}, \overline{d}^{(o)}_{\mathcal{X},*})$, we can say that any channel with input alphabet $\mathcal{X}$ can be approximated in the noisiness/weak-$*$ sense by a channel having a finite output alphabet.

Total-Variation topology

The *total-variation metric distance* $d^{(o)}_{TV,\mathcal{X},*}$ on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is defined as

$$d^{(o)}_{TV,\mathcal{X},*}(\hat{W}, \hat{W}') = \|\mathrm{MP}_{\hat{W}} - \mathrm{MP}_{\hat{W}'}\|_{TV}.$$

The *total-variation topology* $\mathcal{T}^{(o)}_{TV,\mathcal{X},*}$ is the metric topology that is induced by $d^{(o)}_{TV,\mathcal{X},*}$ on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$. We will refer to the open sets (respectively, closed sets, compact sets, ...) of $\mathcal{T}^{(o)}_{TV,\mathcal{X},*}$ as TV-open (respectively, TV-closed, TV-compact, ...). The same notation is also used for open sets of $\mathcal{MP}_{bf}(\mathcal{X})$, $\mathcal{MP}_b(\mathcal{X})$ and $\mathcal{MP}(\mathcal{X})$ in the total-variation topology.

Proposition 11.11. *If $|\mathcal{X}| \geq 2$ and $n \geq 2$, then $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is not TV-compact in $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$.*

Proof. Let $p, p' \in \Delta_{\mathcal{X}}$ be such that $p \neq p'$ and $\frac{1}{2}p + \frac{1}{2}p' = \pi_{\mathcal{X}}$, where $\pi_{\mathcal{X}}$ is the uniform distribution on $\mathcal{X}$. For every $n \geq 1$, define $p_n, p'_n \in \Delta_{\mathcal{X}}$ as

$$p_n = \frac{1}{n}p + \left(1 - \frac{1}{n}\right)\pi_{\mathcal{X}},$$

and

$$p'_n = \frac{1}{n}p' + \left(1 - \frac{1}{n}\right)\pi_{\mathcal{X}}.$$

Clearly, $\frac{1}{2}p_n + \frac{1}{2}p'_n = \pi_{\mathcal{X}}$ for every $n \geq 1$.

Now let $\mathrm{MP}_n = \frac{1}{2}\delta_{p_n} + \frac{1}{2}\delta_{p'_n}$, where δ_{p_n} and $\delta_{p'_n}$ are Dirac measures centered at p_n and p'_n respectively. Clearly, MP_n is balanced and finitely supported for every $n \geq 1$. Let $\hat{W}_n = F_{\mathrm{can}}(\mathrm{MP}_n)$. We have

$$|\operatorname{supp}(\mathrm{MP}_{\hat{W}_n})| = |\operatorname{supp}(\mathrm{MP}_n)| = |\{p_n, p'_n\}| = 2.$$

Therefore, $\hat{W}_n \in \mathrm{DMC}^{(o)}_{\mathcal{X},[2]} \subset \mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ for every $n \geq 1$ and every $m \geq 2$. It is easy to see that $d^{(o)}_{TV,\mathcal{X},*}(\hat{W}_{n_1}, \hat{W}_{n_2}) = \|\mathrm{MP}_{n_1} - \mathrm{MP}_{n_2}\|_{TV} = 1$ for every $n_2 > n_1 \geq 1$. Therefore, no subsequence of $(\mathrm{MP}_n)_{n\geq 1}$ can converge. This means that $\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ is not sequentially compact for any $m \geq 2$. Now since $\mathcal{T}^{(o)}_{TV,\mathcal{X},*}$ is metrizable, we conclude that $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is not compact for any $n \geq 2$. □

Corollary 11.9. *If $|\mathcal{X}| \geq 2$, then $\mathcal{T}^{(o)}_{TV,\mathcal{X},*}$ is not a natural topology.*

Proof. If $\mathcal{T}^{(o)}_{TV,\mathcal{X},*}$ were natural, $\mathrm{DMC}^{(o)}_{\mathcal{X},[2]}$ would be compact, and this is not the case. □

Since the noisiness topology is the same as the weak-$*$ topology, $\mathcal{T}^{(o)}_{\mathcal{X},*}$ is coarser than $\mathcal{T}^{(o)}_{TV,\mathcal{X},*}$. On the other hand, since $\mathcal{T}^{(o)}_{\mathcal{X},*}$ is natural and $\mathcal{T}^{(o)}_{TV,\mathcal{X},*}$ is not, $\mathcal{T}^{(o)}_{\mathcal{X},*}$ is strictly coarser than $\mathcal{T}^{(o)}_{TV,\mathcal{X},*}$ when $|\mathcal{X}| \geq 2$.

Note that the sequence $(\mathrm{MP}_n)_{n\geq 1}$ in the proof of Proposition 11.11 converges in the strong topology because of Proposition 11.7. Therefore, $\mathcal{T}^{(o)}_{s,\mathcal{X},*}$ is not finer than $\mathcal{T}^{(o)}_{TV,\mathcal{X},*}$.

Although $\mathcal{T}^{(o)}_{TV,\mathcal{X},*}$ is not a natural topology itself, it has many properties of natural topologies.

Proposition 11.12. *If $|\mathcal{X}| \geq 2$, every non-empty TV-open subset of $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is rank-unbounded.*

Proof. Let U be a non-empty TV-open set of $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$. Let $\hat{W} \in U$ and let $\epsilon > 0$ be such that $\hat{W}' \in U$ whenever $d^{(o)}_{TV,\mathcal{X},*}(\hat{W}, \hat{W}') < \epsilon$.

Let p, p', $(p_n)_{n\geq 1}$ and $(p'_n)_{n\geq 1}$ be as in Proposition 11.11. For every $n \geq 1$, define $\mathrm{MP}_n \in \mathcal{MP}(\mathcal{X})$ as follows:

$$\mathrm{MP}_n = \left(1 - \frac{\epsilon}{4n}\right)\mathrm{MP}_{\hat{W}} + \frac{\epsilon}{8n^2} \cdot \sum_{i=1}^{n} (\delta_{p_i} + \delta_{p'_i}).$$

Clearly, MP_n is balanced and finitely supported, so $\mathrm{MP}_n \in \mathcal{MP}_{bf}(\mathcal{X})$. Moreover,

$$d^{(o)}_{TV,\mathcal{X},*}(F_{\mathrm{can}}(\mathrm{MP}_n), \hat{W}) = \|\mathrm{MP}_n - \mathrm{MP}_{\hat{W}}\|_{TV} \leq \frac{\epsilon}{2n} < \epsilon.$$

Therefore, $F_{\text{can}}(\text{MP}_n) \in U$ for every $n \geq 1$. On the other hand, $\text{supp}(\text{MP}_n) \supset \{p_i, p_i' : \ 1 \leq i \leq n\}$, which means that $|\text{supp}(\text{MP}_n)| \geq 2n$ and so $F_{\text{can}}(\text{MP}_n) \notin \text{DMC}_{\mathcal{X},[n]}^{(o)}$ for every $n \geq 1$. We conclude that U is rank-unbounded. □

Corollary 11.10. *If $|\mathcal{X}| \geq 2$, the TV-interior of $\text{DMC}_{\mathcal{X},[n]}^{(o)}$ in $\text{DMC}_{\mathcal{X},*}^{(o)}$ is empty.*

Note that the sequence $(F_{\text{can}}(\text{MP}_n))_{n\geq 1}$ in the proof of Proposition 11.12 is rank-unbounded and converges in total-variation to $\hat{W}$. On the other hand, Proposition 11.7 implies that $(F_{\text{can}}(\text{MP}_n))_{n\geq 1}$ does not converge in $(\text{DMC}_{\mathcal{X},*}^{(o)}, \mathcal{T}_{s,\mathcal{X},*}^{(o)})$. We conclude that $\mathcal{T}_{TV,\mathcal{X},*}^{(o)}$ is not finer than $\mathcal{T}_{s,\mathcal{X},*}^{(o)}$.

Although $\text{DMC}_{\mathcal{X},[n]}^{(o)}$ is not TV-compact if $|\mathcal{X}| \geq 2$ and $n \geq 2$, it is TV-complete:

Proposition 11.13. *For every $n \geq 1$, $\text{DMC}_{\mathcal{X},[n]}^{(o)}$ is TV-complete in $\text{DMC}_{\mathcal{X},*}^{(o)}$.*

Proof. Let $\mathcal{MP}_{b,n}(\mathcal{X})$ be the set of balanced meta-probability measures whose support is of size at most n:

$$\mathcal{MP}_{b,n}(\mathcal{X}) = \{\text{MP} \in \mathcal{MP}_b(\mathcal{X}) : \ |\text{supp}(\text{MP})| \leq n\}.$$

Since $(\text{DMC}_{\mathcal{X},[n]}^{(o)}, d_{TV,\mathcal{X},*}^{(o)})$ is isometric to $(\mathcal{MP}_{b,n}(\mathcal{X}), \|\cdot\|_{TV})$, and since $(\mathcal{MP}(\mathcal{X}), \|\cdot\|_{TV})$ is complete, it is sufficient to show that $\mathcal{MP}_{b,n}(\mathcal{X})$ is TV-closed in $\mathcal{MP}(\mathcal{X})$.

Let MP be in the TV-closure of $\mathcal{MP}_{b,n}(\mathcal{X})$. Since we are working in a metric space, there exists a sequence $(\text{MP}_m)_{m\geq 0}$ in $\mathcal{MP}_{b,n}(\mathcal{X})$ that TV-converges to MP. Assume that $\text{MP} \notin \mathcal{MP}_{b,n}(\mathcal{X})$. There exist $p_1, \ldots, p_{n+1} \in \Delta_{\mathcal{X}}$ that are pairwise different and which satisfy $\text{MP}(p_i) > 0$ for every $1 \leq i \leq n+1$. Since $(\text{MP}_m)_{m\geq 0}$ TV-converges to MP, there exists $m_0 \geq 0$ such that $\text{MP}_{m_0}(p_i) > 0$ for every $1 \leq i \leq n+1$. This contradicts the fact $\text{MP}_{m_0} \in \mathcal{MP}_{b,n}(\mathcal{X})$. Therefore, $\text{MP} \in \mathcal{MP}_{b,n}(\mathcal{X})$ for every MP in the TV-closure of $\mathcal{MP}_{b,n}(\mathcal{X})$. This shows that $\mathcal{MP}_{b,n}(\mathcal{X})$ is TV-closed. Therefore, $\text{DMC}_{\mathcal{X},[n]}^{(o)}$ is TV-complete in $\text{DMC}_{\mathcal{X},*}^{(o)}$. □

Proposition 11.14. *If $|\mathcal{X}| \geq 2$, $(\text{DMC}_{\mathcal{X},*}^{(o)}, \mathcal{T}_{TV,\mathcal{X},*}^{(o)})$ is neither Baire nor locally compact anywhere.*

Proof. Since $\text{DMC}_{\mathcal{X},[n]}^{(o)}$ is TV-complete, it is TV-closed. Since it also has empty TV-interior, the same techniques that were used for natural topologies in Section 11.5.1 can be applied for $\mathcal{T}_{TV,\mathcal{X},*}^{(o)}$. □

The above proposition shows that the space $(\text{DMC}_{\mathcal{X},*}^{(o)}, \mathcal{T}_{TV,\mathcal{X},*}^{(o)})$ cannot be completely metrized. Note that since the space $(\text{DMC}_{\mathcal{X},*}^{(o)}, d_{TV,\mathcal{X},*}^{(o)})$ is isometric to the space $(\mathcal{MP}_{bf}(\mathcal{X}), \|\cdot\|_{TV})$, and since $(\mathcal{MP}(\mathcal{X}), \|\cdot\|_{TV})$ is complete, the completion of $(\text{DMC}_{\mathcal{X},*}^{(o)}, d_{TV,\mathcal{X},*}^{(o)})$ is isometric to the closure of $\mathcal{MP}_{bf}(\mathcal{X})$ in $(\mathcal{MP}(\mathcal{X}), \|\cdot\|_{TV})$. It can be shown that the TV-closure of $\mathcal{MP}_{bf}(\mathcal{X})$ in $\mathcal{MP}(\mathcal{X})$ is the set of all balanced and countably supported meta-probability measures on $\mathcal{X}$. Therefore, the completion of $(\text{DMC}_{\mathcal{X},*}^{(o)}, d_{TV,\mathcal{X},*}^{(o)})$ can be thought of as the space of output-equivalent channels from $\mathcal{X}$ to a countably infinite output alphabet. This allows us to say that any channel with input alphabet $\mathcal{X}$ and a countable output alphabet can be approximated in the total-variation sense by a channel having a finite output alphabet.

11.5.5 The Natural Borel σ-algebra on $\mathrm{DMC}_{\mathcal{X},*}^{(o)}$

Let $\mathcal{T}$ be a Hausdorff natural topology on $\mathrm{DMC}_{\mathcal{X},*}^{(o)}$. Since $\mathcal{T}_{s,\mathcal{X},*}^{(o)}$ is the finest natural topology, we have $\mathcal{T} \subset \mathcal{T}_{s,\mathcal{X},*}^{(o)}$. Therefore, $\mathcal{B}(\mathcal{T}) \subset \mathcal{B}(\mathcal{T}_{s,\mathcal{X},*}^{(o)})$.

On the other hand, for every $U \in \mathcal{T}_{s,\mathcal{X},*}^{(o)}$ and every $n \geq 1$, we have $U \cap \mathrm{DMC}_{\mathcal{X},[n]}^{(o)} \in \mathcal{T}_{\mathcal{X},[n]}^{(o)}$. But $\mathcal{T}$ is a natural topology, so there must exist $U_n \in \mathcal{T}$ such that $U_n \cap \mathrm{DMC}_{\mathcal{X},[n]}^{(o)} = U \cap \mathrm{DMC}_{\mathcal{X},[n]}^{(o)}$. Since $U_n \in \mathcal{T}$, we have $U_n \in \mathcal{B}(\mathcal{T})$. Moreover, $\mathrm{DMC}_{\mathcal{X},[n]}^{(o)}$ is $\mathcal{T}$-closed (because it is compact and $\mathcal{T}$ is Hausdorff). Therefore, $\mathrm{DMC}_{\mathcal{X},[n]}^{(o)} \in \mathcal{B}(\mathcal{T})$. This implies that $U \cap \mathrm{DMC}_{\mathcal{X},[n]}^{(o)} = U_n \cap \mathrm{DMC}_{\mathcal{X},[n]}^{(o)} \in \mathcal{B}(T)$, hence

$$U = \bigcup_{n \geq 1} (U \cap \mathrm{DMC}_{\mathcal{X},[n]}^{(o)}) \in \mathcal{B}(T).$$

Since this is true for every $U \in \mathcal{T}_{s,\mathcal{X},*}^{(o)}$, we have $\mathcal{T}_{s,\mathcal{X},*}^{(o)} \subset \mathcal{B}(\mathcal{T})$ which implies that $\mathcal{B}(\mathcal{T}_{s,\mathcal{X},*}^{(o)}) \subset \mathcal{B}(\mathcal{T})$. We conclude that all Hausdorff natural topologies on $\mathrm{DMC}_{\mathcal{X},*}^{(o)}$ have the same σ-algebra. This σ-algebra deserves to be called the *natural Borel σ-algebra* on $\mathrm{DMC}_{\mathcal{X},*}^{(o)}$.

Note that for every $n \geq 1$, the inclusion mapping $i_n : \mathrm{DMC}_{\mathcal{X},[n]}^{(o)} \to \mathrm{DMC}_{\mathcal{X},*}^{(o)}$ is continuous from $(\mathrm{DMC}_{\mathcal{X},[n]}^{(o)}, \mathcal{T}_{\mathcal{X},[n]}^{(o)})$ to $(\mathrm{DMC}_{\mathcal{X},*}^{(o)}, \mathcal{T}_{s,\mathcal{X},*}^{(o)})$, hence it is measurable. Therefore, for every $B \in \mathcal{B}(\mathcal{T}_{s,\mathcal{X},*}^{(o)})$, we have $i_n^{-1}(B) = B \cap \mathrm{DMC}_{\mathcal{X},[n]}^{(o)} \in \mathcal{B}(\mathcal{T}_{\mathcal{X},[n]}^{(o)})$. In the following, we show a converse for this statement.

Fix $n \geq 1$ and let $U \in \mathcal{T}_{\mathcal{X},[n]}^{(o)}$. There exists $U' \in \mathcal{T}_{s,\mathcal{X},*}^{(o)}$ such that $U = U' \cap \mathrm{DMC}_{\mathcal{X},[n]}^{(o)}$. Since U' and $\mathrm{DMC}_{\mathcal{X},[n]}^{(o)}$ are respectively open and closed in the topology $\mathcal{T}_{s,\mathcal{X},*}^{(o)}$, they are both in its Borel σ-algebra. Therefore, $U = U' \cap \mathrm{DMC}_{\mathcal{X},[n]}^{(o)} \in \mathcal{B}(\mathcal{T}_{s,\mathcal{X},*}^{(o)})$ for every $U \in \mathcal{T}_{\mathcal{X},[n]}^{(o)}$. This means that $\mathcal{T}_{\mathcal{X},[n]}^{(o)} \subset \mathcal{B}(\mathcal{T}_{s,\mathcal{X},*}^{(o)})$ and $\mathcal{B}(\mathcal{T}_{\mathcal{X},[n]}^{(o)}) \subset \mathcal{B}(\mathcal{T}_{s,\mathcal{X},*}^{(o)})$ for every $n \geq 1$.

Assume now that $A \subset \mathrm{DMC}_{\mathcal{X},*}^{(o)}$ satisfies $A \cap \mathrm{DMC}_{\mathcal{X},[n]}^{(o)} \in \mathcal{B}(\mathcal{T}_{\mathcal{X},[n]}^{(o)})$ for every $n \geq 1$. This implies that $A \cap \mathrm{DMC}_{\mathcal{X},[n]}^{(o)} \in \mathcal{B}(\mathcal{T}_{s,\mathcal{X},*}^{(o)})$ for every $n \geq 1$, hence

$$A = \bigcup_{n \geq 1} (A \cap \mathrm{DMC}_{\mathcal{X},[n]}^{(o)}) \in \mathcal{B}(\mathcal{T}_{s,\mathcal{X},*}^{(o)}).$$

We conclude that a subset A of $\mathrm{DMC}_{\mathcal{X},*}^{(o)}$ is in the natural Borel σ-algebra if and only if $A \cap \mathrm{DMC}_{\mathcal{X},[n]}^{(o)} \in \mathcal{B}(\mathcal{T}_{\mathcal{X},[n]}^{(o)})$ for every $n \geq 1$.

11.6 Space of Input-Equivalent Channels from $\mathcal{X}$ to $\mathcal{Y}$

11.6.1 The $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}^{(i)}$ Space

Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets. Define the relation $R_{\mathcal{X},\mathcal{Y}}^{(i)}$ on $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ as follows:

$$W R_{\mathcal{X},\mathcal{Y}}^{(i)} W' \quad \Leftrightarrow \quad W \text{ is input-equivalent to } W'.$$

It is easy to see that $R^{(i)}_{\mathcal{X},\mathcal{Y}}$ is an equivalence relation on $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$. $R^{(i)}_{\mathcal{X},\mathcal{Y}}$ is called the *input-equivalence relation* on $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$.

Definition 11.5. *The space of input-equivalent channels with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ is the quotient of the space of channels from $\mathcal{X}$ to $\mathcal{Y}$ by the input-equivalence relation:*

$$\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}} = \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} / R^{(i)}_{\mathcal{X},\mathcal{Y}}.$$

We define the topology $\mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}}$ on $\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}$ as the quotient topology $\mathcal{T}_{\mathcal{X},\mathcal{Y}}/R^{(i)}_{\mathcal{X},\mathcal{Y}}$.

Due to proposition 10.4, we can define the *input-equivalence characteristic of* $\hat{W} \in \mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}$ as $\mathrm{CE}(\hat{W}) := \mathrm{CE}(W')$ for any $W' \in \hat{W}$. Define $\mathrm{co}(\hat{W}) := \mathrm{co}(\mathrm{CE}(\hat{W}))$. It is easy to see that $\mathrm{co}(\hat{W}) = \mathrm{co}(\{W'_x : \ x \in \mathcal{X}\})$ for any $W' \in \hat{W}$.

Let A and B be two sets. A *coupling* of A and B is a subset R of $A \times B$ such that

$$\{a \in A : \ \exists b \in B, \ (a,b) \in R\} = A,$$

and

$$\{b \in B : \ \exists a \in A, \ (a,b) \in R\} = B.$$

We denote the set of couplings of A and B as $\mathcal{R}(A,B)$.

We define the *similarity distance* on $\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}$ as follows:

$$\begin{aligned} d^{(i)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1, \hat{W}_2) &= \inf_{R \in \mathcal{R}(\mathrm{co}(\hat{W}_1), \mathrm{co}(\hat{W}_2))} \sup_{(P_1,P_2) \in R} \|P_1 - P_2\|_{TV} \\ &= \frac{1}{2} \inf_{R \in \mathcal{R}(\mathrm{co}(\hat{W}_1), \mathrm{co}(\hat{W}_2))} \sup_{(P_1,P_2) \in R} \sum_{y \in \mathcal{Y}} |P_1(y) - P_2(y)|. \end{aligned}$$

Proposition 11.15. $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, d^{(i)}_{\mathcal{X},\mathcal{Y}})$ *is a metric space.*

Proof. We will show that $d^{(i)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1, \hat{W}_2) = d_H\big(\mathrm{co}(\hat{W}_1), \mathrm{co}(\hat{W}_2)\big)$, where d_H is the Hausdorff metric on $\mathcal{K}(\Delta_{\mathcal{Y}})$ corresponding to the total-variation distance on $\Delta_{\mathcal{Y}}$. Define $K_1 = \mathrm{co}(\hat{W}_1)$ and $K_2 = \mathrm{co}(\hat{W}_2)$, and let $R \in \mathcal{R}(K_1, K_2)$. For every $(P_1, P_2) \in R$, we have:

$$\|P_1 - P_2\|_{TV} \geq \inf_{P'_2 \in K_2} \|P_1 - P'_2\|_{TV}.$$

Therefore,

$$\sup_{(P_1,P_2) \in R} \|P_1 - P_2\|_{TV} \geq \sup_{P'_1 \in K_1} \inf_{P'_2 \in K_2} \|P'_1 - P'_2\|_{TV}.$$

Similarly,

$$\sup_{(P_1,P_2) \in R} \|P_1 - P_2\|_{TV} \geq \sup_{P'_2 \in K_2} \inf_{P'_1 \in K_1} \|P'_1 - P'_2\|_{TV}.$$

Hence,

$$\begin{aligned} \sup_{(P_1,P_2) \in R} \|P_1 - P_2\|_{TV} &\geq \max \left\{ \sup_{P'_1 \in K_1} \inf_{P'_2 \in K_2} \|P'_1 - P'_2\|_{TV}, \ \sup_{P'_2 \in K_2} \inf_{P'_1 \in K_1} \|P'_1 - P'_2\|_{TV} \right\} \\ &= d_H(K_1, K_2). \end{aligned}$$

We conclude that

$$d^{(i)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1,\hat{W}_2)=\inf_{R\in\mathcal{R}(K_1,K_2)}\sup_{(P_1,P_2)\in R}\|P_1-P_2\|_{TV}\geq d_H(K_1,K_2).$$

Let $P_1\in K_1$. Since K_2 is compact, there exists $\tilde{P}_2(P_1)\in K_2$ such that

$$\|P_1-\tilde{P}_2(P_1)\|_{TV}=\inf_{P_2\in K_2}\|P_1-P_2\|_{TV}.$$

Similarly, for every $P_2\in K_2$, there exists $\tilde{P}_1(P_2)\in K_1$ such that $\|P_2-\tilde{P}_1(P_2)\|_{TV}=\inf_{P_1\in K_1}\|P_1-P_2\|_{TV}$. Define the coupling $R_0\in\mathcal{R}(K_1,K_2)$ as

$$R_0=\{(P_1,\tilde{P}_2(P_1)):\ P_1\in K_1\}\cup\{(\tilde{P}_1(P_2),P_2):\ P_2\in K_2\}.$$

We have:

$$\begin{aligned}d^{(i)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1,\hat{W}_2)&=\inf_{R\in\mathcal{R}(K_1,K_2)}\sup_{(P_1,P_2)\in R}\|P_1-P_2\|_{TV}\leq\sup_{(P_1,P_2)\in R_0}\|P_1-P_2\|_{TV}\\&=\max\left\{\sup_{P_1\in K_1}\|P_1-\tilde{P}_2(P_1)\|,\sup_{P_2\in K_2}\|P_2-\tilde{P}_1(P_2)\|\right\}=d_H(K_1,K_2).\end{aligned}$$

We conclude that $d^{(i)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1,\hat{W}_2)=d_H(K_1,K_2)=d_H\big(\mathrm{co}(\hat{W}_1),\mathrm{co}(\hat{W}_2)\big)$, hence $d^{(i)}_{\mathcal{X},\mathcal{Y}}$ is a metric. □

Proposition 11.16. *Let $\hat{W},\hat{W}'\in\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}$ be the $R^{(i)}_{\mathcal{X},\mathcal{Y}}$-equivalence classes of $W,W'\in\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, respectively. We have $d^{(i)}_{\mathcal{X},\mathcal{Y}}(\hat{W},\hat{W}')\leq d_{\mathcal{X},\mathcal{Y}}(W,W')$.*

Proof. Define $R_0\subset\mathrm{co}(\hat{W})\times\mathrm{co}(\hat{W}')$ as follows:

$$R_0=\left\{\left(\sum_{x\in\mathcal{X}}\lambda_xW_x,\sum_{x\in\mathcal{X}}\lambda_xW'_x\right):\ \sum_{x\in\mathcal{X}}\lambda_x=1,\ \text{and }\lambda_x\geq 0,\ \forall x\in\mathcal{X}\right\}.$$

Clearly, R_0 is a coupling of $\mathrm{co}(\hat{W})$ and $\mathrm{co}(\hat{W}')$. For every $(P_1,P_2)\in R_0$, there exists $(\lambda_x)_{x\in\mathcal{X}}\in[0,1]^{\mathcal{X}}$ such that $\sum_{x\in\mathcal{X}}\lambda_x=1$, $P_1=\sum_{x\in\mathcal{X}}\lambda_xW_x$ and $P_2=\sum_{x\in\mathcal{X}}\lambda_xW'_x$. We have:

$$\begin{aligned}\|P_1-P_2\|_{TV}&=\left\|\left(\sum_{x\in\mathcal{X}}\lambda_xW_x\right)-\left(\sum_{x\in\mathcal{X}}\lambda_xW'_x\right)\right\|_{TV}=\left\|\sum_{x\in\mathcal{X}}\lambda_x(W_x-W'_x)\right\|_{TV}\\&\leq\sum_{x\in\mathcal{X}}\lambda_x\|W_x-W'_x\|_{TV}\leq\sup_{x\in\mathcal{X}}\|W_x-W'_x\|_{TV}=d_{\mathcal{X},\mathcal{Y}}(W,W').\end{aligned}$$

Therefore,

$$\begin{aligned}d^{(i)}_{\mathcal{X},\mathcal{Y}}(\hat{W},\hat{W}')&=\inf_{R\in\mathcal{R}(\mathrm{co}(\hat{W}),\mathrm{co}(\hat{W}'))}\sup_{(P_1,P_2)\in R}\|P_1-P_2\|_{TV}\\&\leq\sup_{(P_1,P_2)\in R_0}\|P_1-P_2\|_{TV}\leq d_{\mathcal{X},\mathcal{Y}}(W,W').\end{aligned}$$

□

Theorem 11.6. *The topology induced by $d^{(i)}_{\mathcal{X},\mathcal{Y}}$ on $\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}$ is the same as the quotient topology $\mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}}$. Moreover, $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, d^{(i)}_{\mathcal{X},\mathcal{Y}})$ is compact and path-connected.*

Proof. Since $(\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}, d_{\mathcal{X},\mathcal{Y}})$ is compact and path-connected, the quotient space $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}})$ is compact and path-connected.

Define the mapping $\mathrm{Proj} : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \to \mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}$ as $\mathrm{Proj}(W) = \hat{W}$, where $\hat{W}$ is the $R^{(i)}_{\mathcal{X},\mathcal{Y}}$-equivalence class of W. Proposition 11.16 implies that Proj is a continuous mapping from $(\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}, d_{\mathcal{X},\mathcal{Y}})$ to $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, d^{(i)}_{\mathcal{X},\mathcal{Y}})$. Since $\mathrm{Proj}(W)$ depends only on $\hat{W}$, Lemma 11.1 implies that the transcendent mapping of Proj defined on the quotient space $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}})$ is continuous. But the transcendent mapping of Proj is nothing but the identity on $\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}$. Therefore, the identity mapping id on $\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}$ is a continuous mapping from $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}})$ to $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, d^{(i)}_{\mathcal{X},\mathcal{Y}})$. For every subset U of $\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}$ we have:

- If U is open in $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, d^{(i)}_{\mathcal{X},\mathcal{Y}})$, then $U = id^{-1}(U)$ is open in $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}})$.
- If U is open in $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}})$, then its complement U^c is closed in the space $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}})$ which is compact, hence U^c is compact in $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}})$. This shows that $U^c = id(U^c)$ is a compact subset of $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, d^{(i)}_{\mathcal{X},\mathcal{Y}})$. But $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, d^{(i)}_{\mathcal{X},\mathcal{Y}})$ is a metric space, so U^c is closed in $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, d^{(i)}_{\mathcal{X},\mathcal{Y}})$. Therefore, U is open in $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, d^{(i)}_{\mathcal{X},\mathcal{Y}})$.

We conclude that $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}})$ and $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, d^{(i)}_{\mathcal{X},\mathcal{Y}})$ have the same open sets. Therefore, the topology induced by $d^{(i)}_{\mathcal{X},\mathcal{Y}}$ on $\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}$ is the same as the quotient topology $\mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}}$. Now since $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}})$ is compact and path-connected, $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, d^{(i)}_{\mathcal{X},\mathcal{Y}})$ is compact and path-connected as well. □

In the rest of this chapter, we always associate $\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}$ with the similarity metric $d^{(i)}_{\mathcal{X},\mathcal{Y}}$ and the quotient topology $\mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}}$.

11.6.2 Canonical Embedding and Canonical Identification

Let $\mathcal{X}_1, \mathcal{X}_2$ and $\mathcal{Y}$ be three finite sets such that $|\mathcal{X}_1| \le |\mathcal{X}_2|$. We will show that there is a canonical embedding from $\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}$ to $\mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$. In other words, there exists an explicitly constructable compact subset A of $\mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$ such that A is homeomorphic to $\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}$. A and the homeomorphism depend only on $\mathcal{X}_1, \mathcal{X}_2$ and $\mathcal{Y}$ (this is why we say that they are canonical). Moreover, we can show that A depends only on $|\mathcal{X}_1|$, $\mathcal{X}_2$ and $\mathcal{Y}$.

Lemma 11.8. *For every $W \in \mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}}$ and every surjection f from $\mathcal{X}_2$ to $\mathcal{X}_1$, W is input-equivalent to $W \circ D_f$.*

Proof. Clearly $W \circ D_f$ is input-degraded from W. Now let f' be any mapping from $\mathcal{X}_1$ to $\mathcal{X}_2$ such that $f(f'(x_1)) = x_1$ for every $x_1 \in \mathcal{X}_1$. We have $W = W \circ (D_f \circ D_{f'}) = (W \circ D_f) \circ D_{f'}$, and so W is also input-degraded from $W \circ D_f$. □

Corollary 11.11. *For every $W, W' \in \mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}}$ and every two surjections f, g from $\mathcal{X}_2$ to $\mathcal{X}_1$, we have:*

$$W R^{(i)}_{\mathcal{X}_1,\mathcal{Y}} W' \;\Leftrightarrow\; (W \circ D_f) R^{(i)}_{\mathcal{X}_2,\mathcal{Y}} (W' \circ D_g).$$

Proof. Since W is input-equivalent to $W \circ D_f$ and W' is input-equivalent to $W' \circ D_g$, then W is input-equivalent to W' if and only if $W \circ D_f$ is input-equivalent to $W' \circ D_g$. □

For every $W \in \mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}}$, we denote the $R^{(i)}_{\mathcal{X}_1,\mathcal{Y}}$-equivalence class of W as $\hat{W}$, and for every $W \in \mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}}$, we denote the $R^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$-equivalence class of W as $\tilde{W}$.

Proposition 11.17. *Let $\mathcal{X}_1, \mathcal{X}_2$ and $\mathcal{Y}$ be three finite sets such that $|\mathcal{X}_1| \le |\mathcal{X}_2|$. Let $f : \mathcal{X}_2 \to \mathcal{X}_1$ be any fixed surjection from $\mathcal{X}_2$ to $\mathcal{X}_1$. Define the mapping $F : \mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}} \to \mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$ as $F(\hat{W}) = \widetilde{W' \circ D_f} = \mathrm{Proj}_2(W' \circ D_f)$, where $W' \in \hat{W}$ and Proj_2 is the projection onto the $R^{(i)}_{\mathcal{X},\mathcal{Y}_2}$-equivalence classes. We have:*

- *F is well defined, i.e., $F(\hat{W})$ does not depend on $W' \in \hat{W}$.*
- *F is a homeomorphism from $\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}$ to $F\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}\big) \subset \mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$.*
- *F does not depend on the surjection f. It depends only on $\mathcal{X}_1$, $\mathcal{X}_2$ and $\mathcal{Y}$, hence it is canonical.*
- *$F\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}\big)$ depends only on $|\mathcal{X}_1|$, $\mathcal{X}_2$ and $\mathcal{Y}$.*
- *For every $W' \in \hat{W}$ and every $W'' \in F(\hat{W})$, W' is input-equivalent to W''.*

Proof. Corollary 11.11 implies that $\mathrm{Proj}_2(W \circ D_f) = \mathrm{Proj}_2(W' \circ D_f)$ if and only if $W R^{(i)}_{\mathcal{X}_1,\mathcal{Y}} W'$. Therefore, $\mathrm{Proj}_2(W' \circ D_f)$ does not depend on $W' \in \hat{W}$, hence F is well defined. Corollary 11.11 also shows that $\mathrm{Proj}_2(W' \circ D_f)$ does not depend on the particular choice of the surjection f, hence it is canonical (i.e., it depends only on $\mathcal{X}_1, \mathcal{X}_2$ and $\mathcal{Y}$).

On the other hand, the mapping $W \to W \circ D_f$ is a continuous mapping from $\mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}}$ to $\mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}}$, and Proj_2 is continuous. Therefore, the mapping $W \to \mathrm{Proj}_2(W \circ D_f)$ is a continuous mapping from $\mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}}$ to $\mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$. Now since $\mathrm{Proj}_2(W \circ D_f)$ depends only on the $R^{(i)}_{\mathcal{X}_1,\mathcal{Y}}$-equivalence class $\hat{W}$ of W, Lemma 11.1 implies that the transcendent mapping of $W \to \mathrm{Proj}_2(W \circ D_f)$ that is defined on $\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}$ is continuous. Therefore, F is a continuous mapping from $(\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X}_1,\mathcal{Y}})$ to $(\mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X}_2,\mathcal{Y}})$. Moreover, we can see from Corollary 11.11 that F is an injection.

For every closed subset B of $\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}$, B is compact since $\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}$ is compact, hence $F(B)$ is compact because F is continuous. This implies that $F(B)$ is closed in

$\mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$ since $\mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$ is Hausdorff (as it is metrizable). Therefore, F is a closed mapping.

Now since F is an injection that is both continuous and closed, F is a homeomorphism from $\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}$ to $F\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}\big) \subset \mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$.

We would like now to show that $F\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}\big)$ depends only on $|\mathcal{X}_1|$, $\mathcal{X}_2$ and $\mathcal{Y}$. Let $\mathcal{X}_1'$ be a finite set such that $|\mathcal{X}_1| = |\mathcal{X}_1'|$. For every $W \in \mathrm{DMC}_{\mathcal{X}_1',\mathcal{Y}}$, let $\overline{W} \in \mathrm{DMC}^{(i)}_{\mathcal{X}_1',\mathcal{Y}}$ be the $R^{(i)}_{\mathcal{X}_1',\mathcal{Y}}$-equivalence class of W.

Let $g : \mathcal{X}_1 \to \mathcal{X}_1'$ be a fixed bijection from $\mathcal{X}_1$ to $\mathcal{X}_1'$ and let $f' = g \circ f$. Define $F' : \mathrm{DMC}^{(i)}_{\mathcal{X}_1',\mathcal{Y}} \to \mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$ as $F'(\overline{W}) = \widetilde{W' \circ D_{f'}} = \mathrm{Proj}_2(W' \circ D_{f'})$, where $W' \in \overline{W}$. As above, F' is well defined, and it is a homeomorphism from $\mathrm{DMC}^{(i)}_{\mathcal{X}_1',\mathcal{Y}}$ to $F'\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1',\mathcal{Y}}\big)$. We want to show that $F'\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1',\mathcal{Y}}\big) = F\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}\big)$. For every $\overline{W} \in \mathrm{DMC}^{(i)}_{\mathcal{X}_1',\mathcal{Y}}$, let $W' \in \overline{W}$. We have

$$F'(\overline{W}) = \mathrm{Proj}_2(W' \circ D_{f'}) = \mathrm{Proj}_2((W' \circ D_g) \circ D_f) = F\left(\widetilde{W' \circ D_g}\right) \in F\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}\big).$$

Since this is true for every $\overline{W} \in \mathrm{DMC}^{(i)}_{\mathcal{X}_1',\mathcal{Y}}$, we deduce that $F'\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1',\mathcal{Y}}\big) \subset F\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}\big)$. By exchanging the roles of $\mathcal{X}_1$ and $\mathcal{X}_1'$ and using the fact that $f = g^{-1} \circ f'$, we get $F\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}\big) \subset F'\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1',\mathcal{Y}}\big)$. We conclude that $F\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}\big) = F'\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1',\mathcal{Y}}\big)$, which means that $F\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}\big)$ depends only on $|\mathcal{X}_1|$, $\mathcal{X}_2$ and $\mathcal{Y}$.

Finally, for every $W' \in \hat{W}$ and every $W'' \in F(\hat{W}) = \widetilde{W' \circ D_f}$, W'' is input-equivalent to $W' \circ D_f$ and $W' \circ D_f$ is input-equivalent to W' (by Lemma 11.8), hence W'' is input-equivalent to W'. □

Corollary 11.12. *If $|\mathcal{X}_1| = |\mathcal{X}_2|$, there exists a canonical homeomorphism from $\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}$ to $\mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$ depending only on $\mathcal{X}_1, \mathcal{X}_2$ and $\mathcal{Y}$.*

Proof. Let f be a bijection from $\mathcal{X}_2$ to $\mathcal{X}_1$. Define the mapping $F : \mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}} \to \mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$ as $F(\hat{W}) = \widetilde{W' \circ D_f} = \mathrm{Proj}_2(W' \circ D_f)$, where $W' \in \hat{W}$ and $\mathrm{Proj}_2 : \mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}} \to \mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$ is the projection onto the $R^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$-equivalence classes.

Also, define the mapping $F' : \mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}} \to \mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}$ as $F'(\tilde{V}) = \widehat{V' \circ D_{f^{-1}}} = \mathrm{Proj}_1(V' \circ D_{f^{-1}})$, where $V' \in \tilde{V}$ and $\mathrm{Proj}_1 : \mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}} \to \mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}$ is the projection onto the $R^{(i)}_{\mathcal{X}_1,\mathcal{Y}}$-equivalence classes.

Proposition 11.17 shows that F and F' are well defined.

For every $W \in \mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}}$, we have:

$$F'(F(\hat{W})) \overset{(a)}{=} F'(\widetilde{W \circ D_f}) \overset{(b)}{=} \widehat{(W \circ D_f) \circ D_{f^{-1}}} = \hat{W},$$

where (a) follows from the fact that $W \in \hat{W}$ and (b) follows from the fact that $W \circ D_f \in \widetilde{W \circ D_f}$.

We can similarly show that $F(F'(\tilde{V})) = \tilde{V}$ for every $\tilde{V} \in \mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$. Therefore, both F and F' are bijections. Proposition 11.17 now implies that F is a homeomorphism from $\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}$ to $F\big(\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}}\big) = \mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}}$. Moreover, F depends only on $\mathcal{X}_1, \mathcal{X}_2$ and $\mathcal{Y}$. □

Corollary 11.12 allows us to identify $\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}$ with $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ through the canonical homeomorphism, where $n = |\mathcal{X}|$ and $[n] = \{1, \dots, n\}$. Moreover, for every $1 \leq n \leq m$, Proposition 11.17 allows us to identify $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ with the canonical subspace of $\mathrm{DMC}^{(i)}_{[m],\mathcal{Y}}$ that is homeomorphic to $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$. In the rest of this chapter, we consider that $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ is a compact subspace of $\mathrm{DMC}^{(i)}_{[m],\mathcal{Y}}$.

Intuitively, $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ has a "lower dimension" compared to $\mathrm{DMC}^{(i)}_{[m],\mathcal{Y}}$. So one expects that the interior of $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ in $(\mathrm{DMC}^{(i)}_{[m],\mathcal{Y}}, \mathcal{T}^{(i)}_{[m],\mathcal{Y}})$ is empty if $m > n$. The following proposition shows that this intuition is accurate when $|\mathcal{Y}| \geq 3$.

Proposition 11.18. *We have:*

- *If* $|\mathcal{Y}| = 1$, *then* $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}} = \mathrm{DMC}^{(i)}_{[1],\mathcal{Y}}$ *for every* $n \geq 1$.
- *If* $|\mathcal{Y}| = 2$, *then* $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}} = \mathrm{DMC}^{(i)}_{[2],\mathcal{Y}}$ *for every* $n \geq 2$.
- *If* $|\mathcal{Y}| \geq 3$, *then for every* $1 \leq n < m$, *the interior of* $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ *in the space* $(\mathrm{DMC}^{(i)}_{[m],\mathcal{Y}}, \mathcal{T}^{(i)}_{[m],\mathcal{Y}})$ *is empty.*

Proof. See Appendix 11.10.7. □

11.7 Spaces of Input-Equivalent Channels

The previous section showed that if we are interested in input-equivalent channels, it is sufficient to study the spaces $\mathrm{DMC}_{[n],\mathcal{Y}}$ and $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ for every $n \geq 1$, where $[n] = \{1, \dots, n\}$. Define the space

$$\mathrm{DMC}_{*,\mathcal{Y}} = \coprod_{n \geq 1} \mathrm{DMC}_{[n],\mathcal{Y}},$$

where $\coprod$ is the disjoint union symbol. The subscript $*$ indicates that the input alphabets of the considered channels are arbitrary but finite. We define the input-equivalence relation $R^{(i)}_{*,\mathcal{Y}}$ on $\mathrm{DMC}_{*,\mathcal{Y}}$ as follows:

$$W R^{(i)}_{*,\mathcal{Y}} W' \quad \Leftrightarrow \quad W \text{ is input-equivalent to } W'.$$

Definition 11.6. *The space of input-equivalent channels with output alphabet* $\mathcal{Y}$ *is the quotient of the space of channels with output alphabet* $\mathcal{Y}$ *by the input-equivalence relation:*

$$\mathrm{DMC}^{(i)}_{*,\mathcal{Y}} = \mathrm{DMC}_{*,\mathcal{Y}} / R^{(i)}_{*,\mathcal{Y}}.$$

Clearly, $\mathrm{DMC}_{[n],\mathcal{Y}} / R^{(i)}_{*,\mathcal{Y}}$ can be canonically identified with $\mathrm{DMC}_{[n],\mathcal{Y}} / R^{(i)}_{[n],\mathcal{Y}} = \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$. Therefore, we can write

$$\mathrm{DMC}^{(i)}_{*,\mathcal{Y}} = \bigcup_{n \geq 1} \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}.$$

We define the *input-rank* of $\hat{W} \in \mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ as the size of its characteristic: $\mathrm{irank}(\hat{W}) = |\,\mathrm{CE}(\hat{W})|$. Due to Proposition 10.4, we have

$$\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}} = \{\hat{W} \in \mathrm{DMC}^{(i)}_{*,\mathcal{Y}} :\ \mathrm{irank}(\hat{W}) \leq n\}.$$

A subset A of $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ is said to be *rank-bounded* if there exists $n \geq 1$ such that $A \subset \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$.

11.7.1 Natural Topologies on $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$

As in Section 11.5.1, we can define natural topologies on the spaces of input-equivalent channels:

Definition 11.7. *A topology $\mathcal{T}$ on $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ is said to be* natural *if it induces the quotient topology $\mathcal{T}^{(i)}_{[n],\mathcal{Y}}$ on $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ for every $n \geq 1$.*

Proposition 11.19. *Every natural topology on $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ is σ-compact, separable and path-connected.*

Proof. We follow the same proof as in Proposition 11.3. □

Proposition 11.18 implies that if $|\mathcal{Y}| = 1$, then $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}} = \mathrm{DMC}^{(i)}_{[1],\mathcal{Y}}$, and so the only natural topology on $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ is $\mathcal{T}^{(i)}_{[1],\mathcal{Y}}$. Similarly, if $|\mathcal{Y}| = 2$, then $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}} = \mathrm{DMC}^{(i)}_{[2],\mathcal{Y}}$, and the only natural topology on $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ is $\mathcal{T}^{(i)}_{[2],\mathcal{Y}}$. In the rest of this section, we investigate the properties of natural topologies when $|\mathcal{Y}| \geq 3$.

Proposition 11.20. *If $|\mathcal{Y}| \geq 3$ and $\mathcal{T}$ is a natural topology, every non-empty open set is rank-unbounded.*

Proof. We follow the same proof as in Proposition 11.4. □

Corollary 11.13. *If $|\mathcal{Y}| \geq 3$ and $\mathcal{T}$ is a natural topology, then for every $n \geq 1$, the interior of $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ in $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T})$ is empty.*

Proposition 11.21. *If $|\mathcal{Y}| \geq 3$ and $\mathcal{T}$ is a Hausdorff natural topology, then the space $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T})$ is not a Baire space.*

Proof. We follow the same proof as in Proposition 11.5. □

Corollary 11.14. *If $|\mathcal{Y}| \geq 3$, no natural topology on $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ can be completely metrizable.*

Proof. The corollary follows from Proposition 11.21 and the fact that every completely metrizable topology is both Hausdorff and Baire. □

Proposition 11.22. *If $|\mathcal{Y}| \geq 3$ and $\mathcal{T}$ is a Hausdorff natural topology, then the space $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T})$ is not locally compact anywhere, i.e., for every $\hat{W} \in \mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$, there is no compact neighborhood of $\hat{W}$ in $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T})$.*

Proof. We follow the same proof as in Proposition 11.6. □

11.7.2 Strong Topology on $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$

The first natural topology on $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ that we study is the *strong topology* $\mathcal{T}^{(i)}_{s,*,\mathcal{Y}}$ on $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$, which is the finest natural topology.

Since the spaces $\{\mathrm{DMC}_{[n],\mathcal{Y}}\}_{n\geq 1}$ are disjoint and since there is no a priori way to (topologically) compare channels in $\mathrm{DMC}_{[n],\mathcal{Y}}$ with channels in $\mathrm{DMC}_{[n'],\mathcal{Y}}$ for $n \neq n'$, the "most natural" topology that we can define on $\mathrm{DMC}_{*,\mathcal{Y}}$ is the disjoint union topology $\mathcal{T}_{s,*,\mathcal{Y}} := \bigoplus_{n\geq 1} \mathcal{T}_{[n],\mathcal{Y}}$. Clearly, the space $(\mathrm{DMC}_{*,\mathcal{Y}}, \mathcal{T}_{s,*,\mathcal{Y}})$ is disconnected. Moreover, $\mathcal{T}_{s,*,\mathcal{Y}}$ is metrizable because it is the disjoint union of metrizable spaces. It is also σ-compact because it is the union of countably many compact spaces.

We added the subscript s to emphasize the fact that $\mathcal{T}_{s,*,\mathcal{Y}}$ is a strong topology (remember that the disjoint union topology is the *finest* topology that makes the canonical injections continuous).

Definition 11.8. *We define the strong topology $\mathcal{T}^{(i)}_{s,*,\mathcal{Y}}$ on $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ as the quotient topology $\mathcal{T}_{s,*,\mathcal{Y}}/R^{(i)}_{*,\mathcal{Y}}$.*

We call open and closed sets in $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$ as strongly open and strongly closed sets respectively.

Let $\mathrm{Proj} : \mathrm{DMC}_{*,\mathcal{Y}} \to \mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ be the projection onto the $R^{(i)}_{*,\mathcal{Y}}$-equivalence classes, and for every $n \geq 1$ let $\mathrm{Proj}_n : \mathrm{DMC}_{[n],\mathcal{Y}} \to \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ be the projection onto the $R^{(i)}_{[n],\mathcal{Y}}$-equivalence classes. Due to the identifications that we made at the beginning of Section 11.7, we have $\mathrm{Proj}(W) = \mathrm{Proj}_n(W)$ for every $W \in \mathrm{DMC}_{[n],\mathcal{Y}}$. Therefore, for every $U \subset \mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$, we have

$$\mathrm{Proj}^{-1}(U) = \coprod_{n\geq 1} \mathrm{Proj}_n^{-1}(U \cap \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}).$$

Hence,

$$\begin{aligned}
U \in \mathcal{T}^{(i)}_{s,*,\mathcal{Y}} &\overset{(a)}{\Leftrightarrow} \mathrm{Proj}^{-1}(U) \in \mathcal{T}_{s,*,\mathcal{Y}} \\
&\overset{(b)}{\Leftrightarrow} \mathrm{Proj}^{-1}(U) \cap \mathrm{DMC}_{[n],\mathcal{Y}} \in \mathcal{T}_{[n],\mathcal{Y}}, \ \ \forall n \geq 1 \\
&\Leftrightarrow \left(\coprod_{n'\geq 1} \mathrm{Proj}_{n'}^{-1}(U \cap \mathrm{DMC}^{(i)}_{[n'],\mathcal{Y}}) \right) \cap \mathrm{DMC}_{[n],\mathcal{Y}} \in \mathcal{T}_{[n],\mathcal{Y}}, \ \ \forall n \geq 1 \\
&\Leftrightarrow \mathrm{Proj}_n^{-1}(U \cap \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}) \in \mathcal{T}_{[n],\mathcal{Y}}, \ \ \forall n \geq 1 \\
&\overset{(c)}{\Leftrightarrow} U \cap \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}} \in \mathcal{T}^{(i)}_{[n],\mathcal{Y}}, \ \ \forall n \geq 1,
\end{aligned}$$

where (a) and (c) follow from the properties of the quotient topology, and (b) follows from the properties of the disjoint union topology.

We conclude that $U \subset \mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ is strongly open in $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ if and only if $U \cap \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ is open in $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ for every $n \geq 1$. This shows that the topology on $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ that is inherited from $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$ is exactly $\mathcal{T}^{(i)}_{[n],\mathcal{Y}}$. Therefore, $\mathcal{T}^{(i)}_{s,*,\mathcal{Y}}$

is a natural topology. On the other hand, if $\mathcal{T}$ is an arbitrary natural topology and $U \in \mathcal{T}$, then $U \cap \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ is open in $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ for every $n \geq 1$, so $U \in \mathcal{T}^{(i)}_{s,*,\mathcal{Y}}$. We conclude that $\mathcal{T}^{(i)}_{s,*,\mathcal{Y}}$ is the finest natural topology.

We can also characterize the strongly closed subsets of $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ in terms of the closed sets of the $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ spaces:

$$
\begin{aligned}
&F \text{ is strongly closed in } \mathrm{DMC}^{(i)}_{*,\mathcal{Y}} \\
&\quad \Leftrightarrow \quad \mathrm{DMC}^{(i)}_{*,\mathcal{Y}} \setminus F \text{ is strongly open in } \mathrm{DMC}^{(i)}_{*,\mathcal{Y}} \\
&\quad \Leftrightarrow \quad \left(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}} \setminus F\right) \cap \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}} \text{ is open in } \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}, \quad \forall n \geq 1 \\
&\quad \Leftrightarrow \quad \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}} \setminus \left(F \cap \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}\right) \text{ is open in } \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}, \quad \forall n \geq 1 \\
&\quad \Leftrightarrow \quad F \cap \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}} \text{ is closed in } \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}, \quad \forall n \geq 1.
\end{aligned}
$$

Since $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ is metrizable for every $n \geq 1$, it is also normal. We can use this fact to prove that the strong topology on $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ is normal:

Lemma 11.9. $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$ *is normal.*

Proof. We follow the same proof as in Lemma 11.5. □

The following theorem shows that the strong topology satisfies a few desirable properties.

Theorem 11.7. $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$ *is a compactly generated, sequential and* T_4 *space.*

Proof. We follow the same proof as in Theorem 11.4. □

Corollary 11.15. *If* $|\mathcal{Y}| \geq 3$, $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$ *is not locally compact anywhere.*

Proof. Since $\mathcal{T}^{(i)}_{s,*,\mathcal{Y}}$ is a natural Hausdorff topology, Proposition 11.22 implies that $\mathcal{T}^{(i)}_{s,*,\mathcal{Y}}$ is not locally compact anywhere. □

As in the case of the space of output-equivalent channels (Section 11.5.2), the space $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$ fails to be first-countable (and hence it is not metrizable) when $|\mathcal{Y}| \geq 3$. This is one manifestation of the strength of the topology $\mathcal{T}^{(i)}_{s,*,\mathcal{Y}}$. In order to show that $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$ is not first-countable, we need to characterize the converging sequences in $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$.

A sequence $(\hat{W}_n)_{n\geq 1}$ in $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ is said to be *rank-bounded* if $\mathrm{irank}(\hat{W}_n)$ is bounded. $(\hat{W}_n)_{n\geq 1}$ is *rank-unbounded* if it is not bounded.

The following proposition shows that every rank-unbounded sequence does not converge in $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$.

Proposition 11.23. *A sequence* $(\hat{W}_n)_{n\geq 0}$ *converges in* $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$ *if and only if there exists* $m \geq 1$ *such that* $\hat{W}_n \in \mathrm{DMC}^{(i)}_{[m],\mathcal{Y}}$ *for every* $n \geq 0$*, and* $(\hat{W}_n)_{n\geq 0}$ *converges in* $(\mathrm{DMC}^{(i)}_{[m],\mathcal{Y}}, \mathcal{T}^{(i)}_{[m],\mathcal{Y}})$.

Proof. We follow the same proof as in Proposition 11.7. □

Corollary 11.16. *If* $|\mathcal{Y}| \geq 3$, $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$ *is not first-countable anywhere, i.e., for every* $\hat{W} \in \mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$*, there is no countable neighborhood basis of* $\hat{W}$.

Proof. We follow the same proof as in Corollary 11.7. □

Compact Subspaces of $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$

It is well known that a compact subset of $\mathbb{R}$ is compact if and only if it is closed and bounded. The following proposition shows that a similar statement holds for $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$.

Proposition 11.24. *A subspace of* $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$ *is compact if and only if it is rank-bounded and strongly closed.*

Proof. If $|\mathcal{Y}| = 1$, $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}} = \mathrm{DMC}^{(i)}_{[1],\mathcal{Y}}$ consists of only one point, hence all subsets of $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ are rank-bounded, compact and strongly closed.

If $|\mathcal{Y}| = 2$, $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}} = \mathrm{DMC}^{(i)}_{[2],\mathcal{Y}}$ and $\mathcal{T}^{(i)}_{s,*,\mathcal{Y}} = \mathcal{T}^{(i)}_{[2],\mathcal{Y}}$, hence all subsets of $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ are rank-bounded. But $\mathrm{DMC}^{(i)}_{[2],\mathcal{Y}}$ is compact and Hausdorff. Therefore, a subset of $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ is compact if and only if it is closed in $\mathcal{T}^{(i)}_{[2],\mathcal{Y}} = \mathcal{T}^{(i)}_{s,*,\mathcal{Y}}$.

For $|\mathcal{Y}| \geq 3$, we follow the same proof as in Proposition 11.8. □

11.7.3 The Similarity Metric on the Space of Input-Equivalent Channels

We define the *similarity metric* on $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ as follows:

$$\begin{aligned} d^{(i)}_{*,\mathcal{Y}}(\hat{W}_1, \hat{W}_2) &= \min_{R \in \mathcal{R}(\mathrm{co}(\hat{W}_1), \mathrm{co}(\hat{W}_2))} \max_{(P_1,P_2)\in R} \|P_1 - P_2\|_{TV} \\ &= \frac{1}{2} \min_{R \in \mathcal{R}(\mathrm{co}(\hat{W}_1), \mathrm{co}(\hat{W}_2))} \max_{(P_1,P_2)\in R} \sum_{y\in\mathcal{Y}} |P_1(y) - P_2(y)|. \end{aligned}$$

Let $\mathcal{T}^{(i)}_{*,\mathcal{Y}}$ be the metric topology on $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ that is induced by $d^{(i)}_{*,\mathcal{Y}}$. We call $\mathcal{T}^{(i)}_{*,\mathcal{Y}}$ the *similarity topology* on $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$.

Clearly, $\mathcal{T}^{(i)}_{*,\mathcal{Y}}$ is natural because the restriction of $d^{(i)}_{*,\mathcal{Y}}$ on $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ is exactly $d^{(i)}_{[n],\mathcal{Y}}$, and the topology induced by $d^{(i)}_{[n],\mathcal{Y}}$ is $\mathcal{T}^{(i)}_{[n],\mathcal{Y}}$ (Theorem 11.6).

11.8 Space of Shannon-Equivalent Channels from $\mathcal{X}$ to $\mathcal{Y}$

11.8.1 The $\mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}}$ space

Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets. Define the relation $R^{(s)}_{\mathcal{X},\mathcal{Y}}$ on $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ as follows:

$$W R^{(s)}_{\mathcal{X},\mathcal{Y}} W' \;\Leftrightarrow\; W \text{ is Shannon-equivalent to } W'.$$

It is easy to see that $R^{(s)}_{\mathcal{X},\mathcal{Y}}$ is an equivalence relation on $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$. $R^{(s)}_{\mathcal{X},\mathcal{Y}}$ is called the *Shannon-equivalence relation* on $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$.

Definition 11.9. *The space of Shannon-equivalent channels with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ is the quotient of the space of channels from $\mathcal{X}$ to $\mathcal{Y}$ by the Shannon-equivalence relation:*

$$\mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}} = \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} / R^{(s)}_{\mathcal{X},\mathcal{Y}}.$$

We define the topology $\mathcal{T}^{(s)}_{\mathcal{X},\mathcal{Y}}$ on $\mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}}$ as the quotient topology $\mathcal{T}_{\mathcal{X},\mathcal{Y}}/R^{(s)}_{\mathcal{X},\mathcal{Y}}$.

Notation 11.1. *Let $(\mathcal{U},\mathcal{X},\mathcal{Y},\mathcal{V},l,W)$ be a BRM game. Since $\mathcal{U},\mathcal{X},\mathcal{Y}$ and $\mathcal{V}$ are implicitly determined by l and W, we write $\$_{\mathrm{opt}}(l,W)$ to denote $\$_{\mathrm{opt}}(\mathcal{U},\mathcal{X},\mathcal{Y},\mathcal{V},l,W)$ for the sake of brevity.*

Let $W, W' \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$. Theorem 10.3 shows that W' contains W if and only if $\$_{\mathrm{opt}}(l,W) \le \$_{\mathrm{opt}}(l,W')$ for every $l \in \Delta_{\mathcal{U}\times\mathcal{V}}$ and every two finite sets $\mathcal{U}$ and $\mathcal{V}$. Therefore, $W R^{(s)}_{\mathcal{X},\mathcal{Y}} W'$ if and only if $\$_{\mathrm{opt}}(l,W) = \$_{\mathrm{opt}}(l,W')$ for every $l \in \Delta_{\mathcal{U}\times\mathcal{V}}$ and every two finite sets $\mathcal{U}$ and $\mathcal{V}$. This shows that $\$_{\mathrm{opt}}(l,W)$ only depends on the $R^{(s)}_{\mathcal{X},\mathcal{Y}}$-equivalence class of W. Therefore, if $\hat{W} \in \mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}}$, we can define $\$_{\mathrm{opt}}(l,\hat{W}) := \$_{\mathrm{opt}}(l,W')$ for any $W' \in \hat{W}$.

Define the *BRM metric* $d^{(s)}_{\mathcal{X},\mathcal{Y}}$ on $\mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}}$ as follows:

$$d^{(s)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1,\hat{W}_2) = \sup_{\substack{n,m\ge 1,\\ l\in\Delta_{[n]\times[m]}}} |\$_{\mathrm{opt}}(l,\hat{W}_1) - \$_{\mathrm{opt}}(l,\hat{W}_2)|.$$

Proposition 11.25. *Let $\hat{W}_1,\hat{W}_2 \in \mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}$ be the $R^{(s)}_{\mathcal{X},\mathcal{Y}}$-equivalence classes of $W_1, W_2 \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, respectively. We have $d^{(s)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1,\hat{W}_2) \le d_{\mathcal{X},\mathcal{Y}}(W_1,W_2)$.*

Proof. See Appendix 11.10.8. □

Theorem 11.8. *The topology induced by $d^{(s)}_{\mathcal{X},\mathcal{Y}}$ on $\mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}}$ is the same as the quotient topology $\mathcal{T}^{(s)}_{\mathcal{X},\mathcal{Y}}$. Moreover, $(\mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}}, d^{(s)}_{\mathcal{X},\mathcal{Y}})$ is compact and path-connected.*

Proof. We use Proposition 11.25 and follow the same proof as in Theorem 11.6. □

Throughout this chapter, we always associate $\mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}}$ with the BRM metric $d^{(s)}_{\mathcal{X},\mathcal{Y}}$ and the quotient topology $\mathcal{T}^{(s)}_{\mathcal{X},\mathcal{Y}}$.

11.8.2 Canonical Embedding and Canonical Identification

Let $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_1$ and $\mathcal{Y}_2$ be four finite sets such that $|\mathcal{X}_1| \leq |\mathcal{X}_2|$ and $|\mathcal{Y}_1| \leq |\mathcal{Y}_2|$. We will show that there is a canonical embedding from $\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}$ to $\mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$. In other words, there exists an explicitly constructable compact subset A of $\mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$ such that A is homeomorphic to $\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}$. A and the homeomorphism depend only on $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_1$ and $\mathcal{Y}_2$ (this is why we say that they are canonical). Moreover, we can show that A depends only on $|\mathcal{X}_1|$, $|\mathcal{Y}_1|$, $\mathcal{X}_2$ and $\mathcal{Y}_2$.

Lemma 11.10. *For every $W \in \mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1}$, every surjection f from $\mathcal{X}_2$ to $\mathcal{X}_1$, and every injection g from $\mathcal{Y}_1$ to $\mathcal{Y}_2$, the channel W is Shannon-equivalent to $D_g \circ W \circ D_f$.*

Proof. Clearly W contains $D_g \circ W \circ D_f$. Now let f' be any mapping from $\mathcal{X}_1$ to $\mathcal{X}_2$ such that $f(f'(x_1)) = x_1$ for every $x_1 \in \mathcal{X}_1$, and let g' be any mapping from $\mathcal{Y}_2$ to $\mathcal{Y}_1$ such that $g'(g(y_1)) = y_1$ for every $y_1 \in \mathcal{Y}_1$. We have

$$W = (D_{g'} \circ D_g) \circ W \circ (D_f \circ D_{f'}) = D_{g'} \circ (D_g \circ W \circ D_f) \circ D_{f'},$$

and so $D_g \circ W \circ D_f$ also contains W. Therefore, W and $D_g \circ W \circ D_f$ are Shannon-equivalent. □

Corollary 11.17. *For every $W, W' \in \mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1}$, every two surjections f, f' from $\mathcal{X}_2$ to $\mathcal{X}_1$, and every two injections g, g' from $\mathcal{Y}_1$ to $\mathcal{Y}_2$, we have:*

$$W R^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1} W' \;\Leftrightarrow\; (D_g \circ W \circ D_f) R^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2} (D_{g'} \circ W' \circ D_{f'}).$$

Proof. Since W is Shannon-equivalent to $D_g \circ W \circ D_f$ and W' is Shannon-equivalent to $D_{g'} \circ W' \circ D_{f'}$, then W is Shannon-equivalent to W' if and only if $D_g \circ W \circ D_f$ is Shannon-equivalent to $D_{g'} \circ W' \circ D_{f'}$. □

For every $W \in \mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1}$, we denote the $R^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}$-equivalence class of W as $\hat{W}$, and for every $W \in \mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}_2}$, we denote the $R^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$-equivalence class of W as $\tilde{W}$.

Proposition 11.26. *Let $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_1$ and $\mathcal{Y}_2$ be four finite sets such that $|\mathcal{X}_1| \leq |\mathcal{X}_2|$ and $|\mathcal{Y}_1| \leq |\mathcal{Y}_2|$. Let $f : \mathcal{X}_2 \to \mathcal{X}_1$ be any fixed surjection from $\mathcal{X}_2$ to $\mathcal{X}_1$, and let $g : \mathcal{Y}_1 \to \mathcal{Y}_2$ be any fixed injection from $\mathcal{Y}_1$ to $\mathcal{Y}_2$. Define the mapping $F : \mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1} \to \mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$ as $F(\hat{W}) = D_g \circ \widetilde{W'} \circ D_f = \mathrm{Proj}_2(D_g \circ W' \circ D_f)$, where $W' \in \hat{W}$, $D_g \circ \widetilde{W'} \circ D_f$ is the $R^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$-equivalence class of $D_g \circ W' \circ D_f$, and Proj_2 is the projection onto the $R^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$-equivalence classes. We have:*

- *F is well defined, i.e., $F(\hat{W})$ does not depend on $W' \in \hat{W}$.*
- *F is a homeomorphism from $\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}$ to $F\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}\big) \subset \mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$.*
- *F does not depend on the surjection f nor on the injection g. It depends only on $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{Y}_1$ and $\mathcal{Y}_2$, hence it is canonical.*
- *$F\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}\big)$ depends only on $|\mathcal{X}_1|$, $|\mathcal{Y}_1|$, $\mathcal{X}_2$ and $\mathcal{Y}_2$.*
- *For every $W' \in \hat{W}$ and every $W'' \in F(\hat{W})$, W' is Shannon-equivalent to W''.*

Proof. See Appendix 11.10.9. □

Corollary 11.18. *If $|\mathcal{X}_1| = |\mathcal{X}_2|$ and $|\mathcal{Y}_1| = |\mathcal{Y}_2|$, there exists a canonical homeomorphism from $\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}$ to $\mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$ depending only on $\mathcal{X}_1, \mathcal{Y}_1, \mathcal{X}_2$ and $\mathcal{Y}_2$.*

Proof. Let f be a bijection from $\mathcal{X}_2$ to $\mathcal{X}_1$, and let g be a bijection from $\mathcal{Y}_1$ to $\mathcal{Y}_2$. Define the mapping $F : \mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1} \to \mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$ as $F(\hat{W}) = D_g \circ \widetilde{W'} \circ D_f = \mathrm{Proj}_2(D_g \circ W' \circ D_f)$, where $W' \in \hat{W}$ and $\mathrm{Proj}_2 : \mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}_2} \to \mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$ is the projection onto the $R^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$-equivalence classes.

Also, define the mapping $F' : \mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2} \to \mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}$ as

$$F'(\tilde{V}) = D_{g^{-1}} \circ \widehat{V'} \circ D_{f^{-1}} = \mathrm{Proj}_1(D_{g^{-1}} \circ V' \circ D_{f^{-1}}),$$

where $V' \in \tilde{V}$ and $\mathrm{Proj}_1 : \mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1} \to \mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}$ is the projection onto the $R^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}$-equivalence classes.

Proposition 11.26 shows that F and F' are well defined.

For every $W \in \mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1}$, we have:

$$F'(F(\hat{W})) \overset{(a)}{=} F'(D_g \circ \widetilde{W} \circ D_f) \overset{(b)}{=} \mathrm{Proj}_1(D_{g^{-1}} \circ (D_g \circ W \circ D_f) \circ D_{f^{-1}}) = \hat{W},$$

where (a) follows from the fact that $W \in \hat{W}$ and (b) follows from the fact that $D_g \circ W \circ D_f \in D_g \circ \widetilde{W} \circ D_f$.

We can similarly show that $F(F'(\tilde{V})) = \tilde{V}$ for every $\tilde{V} \in \mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$. Therefore, both F and F' are bijections. Proposition 11.26 now implies that F is a homeomorphism from $\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}$ to $F\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}\big) = \mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$. Moreover, F depends only on $\mathcal{X}_1, \mathcal{Y}_1, \mathcal{X}_2$ and $\mathcal{Y}_2$. □

Corollary 11.18 allows us to identify $\mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}}$ with $\mathrm{DMC}^{(s)}_{[n],[m]}$ through the canonical homeomorphism, where $n = |\mathcal{X}|$, $m = |\mathcal{Y}|$, $[n] = \{1, \dots, n\}$ and $[m] = \{1, \dots, m\}$. Moreover, for every $1 \le n \le n'$ and $1 \le m \le m'$, Proposition 11.26 allows us to identify $\mathrm{DMC}^{(s)}_{[n],[m]}$ with the canonical subspace of $\mathrm{DMC}^{(s)}_{[n'],[m']}$ that is homeomorphic to $\mathrm{DMC}^{(s)}_{[n],[m]}$. In the rest of this chapter, we consider that $\mathrm{DMC}^{(s)}_{[n],[m]}$ is a compact subspace of $\mathrm{DMC}^{(s)}_{[n'],[m']}$.

Conjecture 11.1. *For every $1 \le n < m$, the interior of $\mathrm{DMC}^{(s)}_{[n],[n]}$ in $\mathrm{DMC}^{(s)}_{[m],[m]}$ is empty.*

11.9 Space of Shannon-Equivalent Channels

The previous section showed that if we are interested in Shannon-equivalent channels, it is sufficient to study the spaces $\mathrm{DMC}_{[n],[m]}$ and $\mathrm{DMC}^{(s)}_{[n],[m]}$ for every $n, m \ge 1$. Define the space

$$\mathrm{DMC}_{*,*} = \coprod_{\substack{n \ge 1,\\ m \ge 1}} \mathrm{DMC}_{[n],[m]}\,.$$

The subscripts $*$ indicate that the input and output alphabets of the considered channels are arbitrary but finite. We define the Shannon-equivalence relation $R^{(s)}_{*,*}$ on $\mathrm{DMC}_{*,*}$ as follows:

$$W R^{(s)}_{*,*} W' \;\Leftrightarrow\; W \text{ is Shannon-equivalent to } W'.$$

Definition 11.10. *The space of Shannon-equivalent channels is the quotient of the space of channels by the Shannon-equivalence relation:*

$$\mathrm{DMC}^{(s)}_{*,*} = \mathrm{DMC}_{*,*} / R^{(s)}_{*,*}.$$

Clearly, $\mathrm{DMC}_{[n],[m]} / R^{(s)}_{*,*}$ can be canonically identified with $\mathrm{DMC}_{[n],[m]} / R^{(s)}_{[n],[m]} = \mathrm{DMC}^{(s)}_{[n],[m]}$ for every $n, m \geq 1$. Therefore, we can write

$$\mathrm{DMC}^{(s)}_{*,*} = \bigcup_{n,m\geq 1} \mathrm{DMC}^{(s)}_{[n],[m]} \overset{(a)}{=} \bigcup_{n\geq 1} \mathrm{DMC}^{(s)}_{[n],[n]}\,.$$

Note that (a) follows from the fact that $\mathrm{DMC}^{(s)}_{[n],[m]} \subset \mathrm{DMC}^{(s)}_{[k],[k]}$ (see Section 11.8.2), where $k = \max\{n, m\}$.

We define the *Shannon-rank* of $\hat{W} \in \mathrm{DMC}^{(s)}_{*,*}$ as:

$$\mathrm{srank}(\hat{W}) = \min\{n \geq 1 :\ \hat{W} \in \mathrm{DMC}^{(s)}_{[n],[n]}\}.$$

Clearly,

$$\mathrm{DMC}^{(s)}_{[n],[n]} = \{\hat{W} \in \mathrm{DMC}^{(s)}_{*,*} :\ \mathrm{srank}(\hat{W}) \leq n\}.$$

A subset A of $\mathrm{DMC}^{(s)}_{*,*}$ is said to be *rank-bounded* if there exists $n \geq 1$ such that $A \subset \mathrm{DMC}^{(s)}_{[n],[n]}$.

11.9.1 Natural Topologies on $\mathrm{DMC}^{(s)}_{*,*}$

As in Section 11.5.1, we can define natural topologies on the space of Shannon-equivalent channels:

Definition 11.11. *A topology $\mathcal{T}$ on $\mathrm{DMC}^{(s)}_{*,*}$ is said to be* natural *if it induces the quotient topology $\mathcal{T}^{(s)}_{[n],[m]}$ on $\mathrm{DMC}^{(s)}_{[n],[m]}$ for every $n, m \geq 1$.*

Proposition 11.27. *Every natural topology on $\mathrm{DMC}^{(s)}_{*,*}$ is σ-compact, separable and path-connected.*

Proof. We follow the same proof as in Proposition 11.3. □

Remark 11.1. *It is possible to show that if Conjecture 11.1 is true, then for every natural topology $\mathcal{T}$ on $\mathrm{DMC}^{(s)}_{*,*}$, we have:*

- *Every open set is rank-unbounded.*
- *For every $n \geq 1$, the interior of $\mathrm{DMC}^{(s)}_{[n],[n]}$ in $(\mathrm{DMC}^{(s)}_{*,*}, \mathcal{T})$ is empty.*
- *If $\mathcal{T}$ is Hausdorff, then*
  - *$(\mathrm{DMC}^{(s)}_{*,*}, \mathcal{T})$ is not a Baire space, hence no natural topology can be completely metrized.*
  - *$(\mathrm{DMC}^{(s)}_{*,*}, \mathcal{T})$ is not locally compact anywhere.*

11.9.2 Strong Topology on $\mathrm{DMC}^{(s)}_{*,*}$

Since the spaces $\{\mathrm{DMC}_{[n],[m]}\}_{n,m\geq 1}$ are disjoint and since there is no a priori way to (topologically) compare channels in $\mathrm{DMC}_{[n],[m]}$ with channels in $\mathrm{DMC}_{[n'],[m']}$ for $(n,m)\neq(n',m')$, the "most natural" topology that we can define on $\mathrm{DMC}_{*,*}$ is the disjoint union topology $\mathcal{T}_{s,*,*} := \bigoplus_{n,m\geq 1} \mathcal{T}_{[n],[m]}$. Clearly, the space $(\mathrm{DMC}_{*,*}, \mathcal{T}_{s,*,*})$ is disconnected. Moreover, $\mathcal{T}_{s,*,*}$ is metrizable because it is the disjoint union of metrizable spaces. It is also σ-compact because it is the union of countably many compact spaces.

We added the subscript s to emphasize the fact that $\mathcal{T}_{s,*,*}$ is a strong topology (remember that the disjoint union topology is the *finest* topology that makes the canonical injections continuous).

Definition 11.12. *We define the strong topology $\mathcal{T}^{(s)}_{s,*,*}$ on $\mathrm{DMC}^{(s)}_{*,*}$ as the quotient topology $\mathcal{T}_{s,*,*}/R^{(s)}_{*,*}$.*

We call open and closed sets in $(\mathrm{DMC}^{(s)}_{*,*}, \mathcal{T}^{(s)}_{s,*,*})$ as strongly open and strongly closed sets respectively.

Let $\mathrm{Proj} : \mathrm{DMC}_{*,*} \to \mathrm{DMC}^{(s)}_{*,*}$ be the projection onto the $R^{(s)}_{*,*}$-equivalence classes, and for every $n,m\geq 1$ let $\mathrm{Proj}_{n,m} : \mathrm{DMC}_{[n],[m]} \to \mathrm{DMC}^{(s)}_{[n],[m]}$ be the projection onto the $R^{(s)}_{[n],[m]}$-equivalence classes. Due to the identifications that we made at the beginning of Section 11.9, we have $\mathrm{Proj}(W) = \mathrm{Proj}_{n,m}(W)$ for every $W \in \mathrm{DMC}_{[n],[m]}$. Therefore, for every $U \subset \mathrm{DMC}^{(s)}_{*,*}$, we have

$$\mathrm{Proj}^{-1}(U) = \coprod_{n,m\geq 1} \mathrm{Proj}^{-1}_{n,m}(U \cap \mathrm{DMC}^{(s)}_{[n],[m]}).$$

Hence,

$$\begin{aligned}
U \in \mathcal{T}^{(s)}_{s,*,*} &\overset{(a)}{\Leftrightarrow} \mathrm{Proj}^{-1}(U) \in \mathcal{T}_{s,*,*} \\
&\overset{(b)}{\Leftrightarrow} \mathrm{Proj}^{-1}(U) \cap \mathrm{DMC}_{[n],[m]} \in \mathcal{T}_{[n],[m]}, \quad \forall n,m\geq 1 \\
&\Leftrightarrow \left(\coprod_{n',m'\geq 1} \mathrm{Proj}^{-1}_{n',m'}(U \cap \mathrm{DMC}^{(s)}_{[n'],[m']}) \right) \cap \mathrm{DMC}_{[n],[m]} \in \mathcal{T}_{[n],[m]}, \quad \forall n,m\geq 1 \\
&\Leftrightarrow \mathrm{Proj}^{-1}_{n,m}(U \cap \mathrm{DMC}^{(s)}_{[n],[m]}) \in \mathcal{T}_{[n],[m]}, \quad \forall n,m\geq 1 \\
&\overset{(c)}{\Leftrightarrow} U \cap \mathrm{DMC}^{(s)}_{[n],[m]} \in \mathcal{T}^{(s)}_{[n],[m]}, \quad \forall n,m\geq 1,
\end{aligned}$$

where (a) and (c) follow from the properties of the quotient topology, and (b) follows from the properties of the disjoint union topology.

We conclude that $U \subset \mathrm{DMC}^{(s)}_{*,*}$ is strongly open in $\mathrm{DMC}^{(s)}_{*,*}$ if and only if $U \cap \mathrm{DMC}^{(s)}_{[n],[m]}$ is open in $\mathrm{DMC}^{(s)}_{[n],[m]}$ for every $n,m\geq 1$. This shows that the topology on $\mathrm{DMC}^{(s)}_{[n],[m]}$ that is inherited from $(\mathrm{DMC}^{(s)}_{*,*}, \mathcal{T}^{(s)}_{s,*,*})$ is exactly $\mathcal{T}^{(s)}_{[n],[m]}$. Therefore, $\mathcal{T}^{(s)}_{s,*,*}$ is a natural topology. On the other hand, if $\mathcal{T}$ is an arbitrary natural topology and

$U \in \mathcal{T}$, then $U \cap \mathrm{DMC}^{(s)}_{[n],[m]}$ is open in $\mathrm{DMC}^{(s)}_{[n],[m]}$ for every $n, m \geq 1$, so $U \in \mathcal{T}^{(s)}_{s,*,*}$. We conclude that $\mathcal{T}^{(s)}_{s,*,*}$ is the finest natural topology.

We can also characterize the strongly closed subsets of $\mathrm{DMC}^{(s)}_{*,*}$ in terms of the closed sets of the $\mathrm{DMC}^{(s)}_{[n],[m]}$ spaces:

F is strongly closed in $\mathrm{DMC}^{(s)}_{*,*}$

$$\begin{aligned}
&\Leftrightarrow \quad \mathrm{DMC}^{(s)}_{*,*} \setminus F \text{ is strongly open in } \mathrm{DMC}^{(s)}_{*,*} \\
&\Leftrightarrow \quad \left(\mathrm{DMC}^{(s)}_{*,*} \setminus F\right) \cap \mathrm{DMC}^{(s)}_{[n],[m]} \text{ is open in } \mathrm{DMC}^{(s)}_{[n],[m]}, \quad \forall n, m \geq 1 \\
&\Leftrightarrow \quad \mathrm{DMC}^{(s)}_{[n],[m]} \setminus \left(F \cap \mathrm{DMC}^{(s)}_{[n],[m]}\right) \text{ is open in } \mathrm{DMC}^{(s)}_{[n],[m]}, \quad \forall n, m \geq 1 \\
&\Leftrightarrow \quad F \cap \mathrm{DMC}^{(s)}_{[n],[m]} \text{ is closed in } \mathrm{DMC}^{(s)}_{[n],[m]}, \quad \forall n, m \geq 1.
\end{aligned}$$

Lemma 11.11. *For every subset U of $\mathrm{DMC}^{(s)}_{*,*}$, we have:*

- *U is strongly open if and only if $U \cap \mathrm{DMC}^{(s)}_{[n],[n]}$ is open in $\mathrm{DMC}^{(s)}_{[n],[n]}$ for every $n \geq 1$.*
- *U is strongly closed if and only if $U \cap \mathrm{DMC}^{(s)}_{[n],[n]}$ is closed in $\mathrm{DMC}^{(s)}_{[n],[n]}$ for every $n \geq 1$.*

Proof. If U is strongly open then $U \cap \mathrm{DMC}^{(s)}_{[n],[m]}$ is open in $\mathrm{DMC}^{(s)}_{[n],[m]}$ for every $n, m \geq 1$. This implies that $U \cap \mathrm{DMC}^{(s)}_{[n],[n]}$ is open in $\mathrm{DMC}^{(s)}_{[n],[n]}$ for every $n \geq 1$.

Conversely, assume that $U \cap \mathrm{DMC}^{(s)}_{[n],[n]}$ is open in $\mathrm{DMC}^{(s)}_{[n],[n]}$ for every $n \geq 1$. Fix $n, m \geq 1$ and let $k = \max\{n, m\}$. We have $\mathrm{DMC}^{(s)}_{[n],[m]} \subset \mathrm{DMC}^{(s)}_{[k],[k]}$. Since $U \cap \mathrm{DMC}^{(s)}_{[k],[k]}$ is open in $\mathrm{DMC}^{(s)}_{[k],[k]}$, the set $U \cap \mathrm{DMC}^{(s)}_{[n],[m]} = (U \cap \mathrm{DMC}^{(s)}_{[k],[k]}) \cap \mathrm{DMC}^{(s)}_{[n],[m]}$ is open in $\mathrm{DMC}^{(s)}_{[n],[m]}$. Therefore, $U \cap \mathrm{DMC}^{(s)}_{[n],[m]}$ is open in $\mathrm{DMC}^{(s)}_{[n],[m]}$ for every $n, m \geq 1$, which implies that U is strongly open.

We can similarly show that U is strongly closed if and only if $U \cap \mathrm{DMC}^{(s)}_{[n],[n]}$ is closed in $\mathrm{DMC}^{(s)}_{[n],[n]}$ for every $n \geq 1$. □

Since $\mathrm{DMC}^{(s)}_{[n],[n]}$ is metrizable for every $n \geq 1$, it is also normal. We can use this fact to prove that the strong topology on $\mathrm{DMC}^{(s)}_{*,*}$ is normal:

Lemma 11.12. $(\mathrm{DMC}^{(s)}_{*,*}, \mathcal{T}^{(s)}_{s,*,*})$ *is normal.*

Proof. We follow the same proof as in Lemma 11.5. □

The following theorem shows that the strong topology satisfies a few desirable properties.

Theorem 11.9. $(\mathrm{DMC}^{(s)}_{*,*}, \mathcal{T}^{(s)}_{s,*,*})$ *is a compactly generated, sequential and T_4 space.*

Proof. We follow the same proof as in Theorem 11.4. □

Remark 11.2. *It is possible to show that if Conjecture 11.1 is true, then we have:*

- $\mathcal{T}^{(s)}_{s,*,*}$ *is not first-countable anywhere.*
- *A subset of* $\mathrm{DMC}^{(s)}_{*,*}$ *is compact in* $\mathcal{T}_{s,*,*}$ *if and only if it is rank-bounded and strongly closed.*

11.9.3 The BRM Metric on the Space of Shannon-Equivalent Channels

We define the *BRM metric* on $\mathrm{DMC}^{(s)}_{*,*}$ as follows:

$$d^{(s)}_{*,*}(\hat{W}_1, \hat{W}_2) = \sup_{\substack{n,m\geq 1,\\ l\in\Delta_{[n]\times[m]}}} |\$_{\mathrm{opt}}(l, \hat{W}_1) - \$_{\mathrm{opt}}(l, \hat{W}_2)|.$$

Let $\mathcal{T}^{(s)}_{*,*}$ be the metric topology on $\mathrm{DMC}^{(s)}_{*,*}$ that is induced by $d^{(s)}_{*,*}$. We call $\mathcal{T}^{(s)}_{*,*}$ the *BRM topology* on $\mathrm{DMC}^{(s)}_{*,*}$.

Clearly, $\mathcal{T}^{(s)}_{*,*}$ is natural because the restriction of $d^{(s)}_{*,*}$ on $\mathrm{DMC}^{(s)}_{[n],[m]}$ is exactly $d^{(s)}_{[n],[m]}$, and the topology induced by $d^{(s)}_{[n],[m]}$ is $\mathcal{T}^{(s)}_{[n],[m]}$ (Theorem 11.8).

11.10 Appendix

11.10.1 Proof of Lemma 11.3

We need the following lemma:

Lemma 11.13. *The relation* $R^{(o)}_{\mathcal{X},\mathcal{Y}}$ *is closed in* $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$.

Proof. Define the mapping $f : (\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^2 \times (\mathrm{DMC}_{\mathcal{Y},\mathcal{Y}})^2 \to (\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^4$ as:

$$f(W, W', V, V') = (W, V' \circ W', W', V \circ W).$$

f is continuous because channel composition is continuous.

Define the set $A \subset (\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^4$ as:

$$A := \{(W, W, W', W') :\ W, W' \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}\}.$$

It is easy to see that A is a closed subset of $(\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^4$. We have:

$$f^{-1}(A) = \{(W, W', V, V') \in (\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^2 \times (\mathrm{DMC}_{\mathcal{Y},\mathcal{Y}})^2 :\ V' \circ W' = W,\ V \circ W = W'\}.$$

Since f is continuous and since A is a closed subset of $(\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^4$, $f^{-1}(A)$ is a closed subset of $(\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^2 \times (\mathrm{DMC}_{\mathcal{Y},\mathcal{Y}})^2$ which is compact. Therefore, $f^{-1}(A)$ is compact.

Now define the mapping $g : (\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^2 \times (\mathrm{DMC}_{\mathcal{Y},\mathcal{Y}})^2 \to (\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^2$ as follows:

$$g(W, W', V, V') = (W, W').$$

Since g is continuous and since $f^{-1}(A)$ is compact, $g(f^{-1}(A))$ is a compact subset of $\mathrm{DMC}^2_{\mathcal{X},\mathcal{Y}}$. Now notice that

$$\begin{aligned} g(f^{-1}(A)) &= \big\{(W,W') \in (\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^2 :\ \exists V,V' \in \mathrm{DMC}_{\mathcal{Y},\mathcal{Y}},\ (W,W',V,V') \in f^{-1}(A)\big\} \\ &= \big\{(W,W') \in (\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^2 :\ \exists V,V' \in \mathrm{DMC}_{\mathcal{Y},\mathcal{Y}},\ V' \circ W' = W,\ V \circ W = W'\big\} \\ &= \big\{(W,W') \in (\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^2 :\ W \text{ is output-equivalent to } W'\big\} \\ &= \big\{(W,W') \in (\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^2 :\ W R^{(o)}_{\mathcal{X},\mathcal{Y}} W'\big\} = R^{(o)}_{\mathcal{X},\mathcal{Y}}. \end{aligned}$$

We conclude that $R^{(o)}_{\mathcal{X},\mathcal{Y}}$ is compact, hence it is also closed because $(\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^2$ is a metric space. □

Now we are ready to prove Lemma 11.3:

Let $\mathrm{Proj} : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \to \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ be defined as $\mathrm{Proj}(W) = \hat{W}$. The continuity of Proj follows from the definition of the quotient topology.

Now let A be a closed subset of $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$. We want to show that $\mathrm{Proj}(A)$ is closed.

Since A is closed in $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, the set $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \times A$ is closed in $(\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^2$. On the other hand, $R^{(o)}_{\mathcal{X},\mathcal{Y}}$ is closed in $(\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^2$ by Lemma 11.13. Therefore, $(\mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \times A) \cap R^{(o)}_{\mathcal{X},\mathcal{Y}}$ is closed in $(\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^2$ which is compact, hence the set $(\mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \times A) \cap R^{(o)}_{\mathcal{X},\mathcal{Y}}$ is compact. We have:

$$(\mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \times A) \cap R^{(o)}_{\mathcal{X},\mathcal{Y}} = \big\{(W,W') \in (\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^2 :\ W R^{(o)}_{\mathcal{X},\mathcal{Y}} W' \text{ and } W' \in A\big\}.$$

Now define the mapping $g : (\mathrm{DMC}_{\mathcal{X},\mathcal{Y}})^2 \to \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ as

$$g(W,W') = W.$$

Let $A_R := g\big((\mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \times A) \cap R^{(o)}_{\mathcal{X},\mathcal{Y}}\big)$. Since g is continuous and since $(\mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \times A) \cap R^{(o)}_{\mathcal{X},\mathcal{Y}}$ is compact, A_R is also compact. We have:

$$A_R = \big\{W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} :\ \exists W' \in A,\ W R^{(o)}_{\mathcal{X},\mathcal{Y}} W'\big\} = \mathrm{Proj}^{-1}(\mathrm{Proj}(A)).$$

Since $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ is a metric space and since A_R is compact, $\mathrm{Proj}^{-1}(\mathrm{Proj}(A)) = A_R$ is closed in $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$. On the other hand, we have $\mathrm{Proj}^{-1}\left(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \setminus \mathrm{Proj}(A)\right) = \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \setminus \mathrm{Proj}^{-1}(\mathrm{Proj}(A))$, hence $\mathrm{Proj}^{-1}\left(\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \setminus \mathrm{Proj}(A)\right)$ is open in $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, which implies that $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \setminus \mathrm{Proj}(A)$ is open in $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$. Therefore, $\mathrm{Proj}(A)$ is closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$.

11.10.2 Proof of Proposition 11.2

let $\hat{U}$ be an arbitrary non-empty open subset of $(\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}, \mathcal{T}^{(o)}_{\mathcal{X},[m]})$ and let Proj be the projection onto the $R^{(o)}_{\mathcal{X},[m]}$-equivalence classes. $\mathrm{Proj}^{-1}(\hat{U})$ is open in the metric space $(\mathrm{DMC}_{\mathcal{X},[m]}, d_{\mathcal{X},[m]})$. Therefore, there exists $W \in \mathrm{DMC}_{\mathcal{X},[m]}$ and $\epsilon > 0$ such that $\mathrm{Proj}^{-1}(\hat{U})$ contains the open ball of center W and radius ϵ.

We will show that there exists $W' \in \mathrm{DMC}_{\mathcal{X},[m]}$ such that $\mathrm{rank}(W') = m > n$ and $d_{\mathcal{X},[m]}(W, W') < \epsilon$. If $\mathrm{rank}(W) = m$, take $W' = W$.

Assume that $\mathrm{rank}(W) < m$. Let $P^o_W \in \Delta_{[m]}$, $\mathrm{Im}(W)$ and $\{W_y^{-1} : y \in \mathrm{Im}(W)\}$ be as in Section 10.2.

Let $\{v_y\}_{y\in[m]}$ be a collection of m vectors in $\mathbb{R}^{\mathcal{X}}$ such that:

- $\sum_{y\in\mathrm{Im}(W)} P^o_W(y) \cdot v_y = 0.$
- $\sum_{y\in[m]\setminus\mathrm{Im}(W)} v_y = 0.$
- For every $y \in [m]$, $\sum_{x\in\mathcal{X}} v_y(x) = 0.$
- The vectors $\{v_y\}_{y\in[m]}$ are pairwise different.

Such collection can always be found.

Let $0 < \delta, \delta' < 1$ and define $P^o_{W'} \in \mathbb{R}^{[m]}$ as follows:

$$P^o_{W'}(y) = \begin{cases} (1-\delta)P^o_W(y) & \text{if } y \in \mathrm{Im}(W), \\ \dfrac{\delta}{m - |\,\mathrm{Im}(W)|} & \text{otherwise.} \end{cases}$$

Clearly, $P^o_{W'} \in \Delta_{[m]}$ and $P^o_{W'}(y) > 0$ for every $y \in [m]$. Now for every $y \in [m]$, define $W_y'^{-1}$ as follows:

$$W_y'^{-1} = \begin{cases} (1-\delta)W_y^{-1} + \delta\pi_{\mathcal{X}} + \delta' v_y & \text{if } y \in \mathrm{Im}(W), \\ \pi_{\mathcal{X}} + \delta' v_y & \text{otherwise,} \end{cases}$$

where $\pi_{\mathcal{X}} \in \Delta_{\mathcal{X}}$ is the uniform probability distribution on $\mathcal{X}$. A simple calculation shows that $\sum_{y\in[m]} P^o_{W'}(y) W_y'^{-1} = \pi_{\mathcal{X}}$, and for every $y \in [m]$ we have $\sum_{x\in\mathcal{X}} W_y'^{-1}(x) = 1$.

Notice that for $y \in \mathrm{Im}(W)$, since $0 < \delta < 1$, $(1-\delta)W_y^{-1} + \delta\pi_{\mathcal{X}}$ lies inside the interior of the probability distribution simplex $\Delta_{\mathcal{X}}$. This means that for δ' small enough, $(1-\delta)W_y^{-1} + \delta\pi_{\mathcal{X}} + \delta' v_y \in \Delta_{\mathcal{X}}$ for every $y \in \mathrm{Im}(W)$, and $\pi_{\mathcal{X}} + \delta' v_y \in \Delta_{\mathcal{X}}$ for every $y \notin \mathrm{Im}(W)$. For every $0 < \delta < 1$, choose $\delta' := \delta'(\delta)$ so that $0 < \delta' < \delta$ and $W_y'^{-1} \in \Delta_{\mathcal{X}}$ for every $y \in [m]$.

It is easy to see that for δ small enough, $W_{y_1}'^{-1} \neq W_{y_2}'^{-1}$ for every $y_1, y_2 \in [m]$ satisfying $y_1 \neq y_2$. Define the channel $W' \in \mathrm{DMC}_{\mathcal{X},[m]}$ as follows:

$$W'(y|x) = |\mathcal{X}| P^o_{W'}(y) W_y'^{-1}(x).$$

Since $P^o_{W'}(y) > 0$ for every $y \in [m]$, we have $\mathrm{supp}(\mathrm{MP}_{W'}) = \{W_y'^{-1} : y \in [m]\}$. Therefore, there exists $\delta_0 > 0$ such for every $0 < \delta < \delta_0$, we have $\mathrm{rank}(W') = m$. On the other hand, we have $\lim_{\delta\to 0} P^o_{W'} = P^o_W$ and $\lim_{\delta\to 0} W_y'^{-1} = W_y^{-1}$ for every $y \in \mathrm{Im}(W)$. Therefore, $\lim_{\delta\to 0} W' = W$ (where the limit is taken in $(\mathrm{DMC}_{\mathcal{X},[m]}, d_{\mathcal{X},[m]})$). This shows that there exists $W' \in \mathrm{DMC}_{\mathcal{X},[m]}$ such that $\mathrm{rank}(W') = m > n$ and $d_{\mathcal{X},[m]}(W, W') < \epsilon$, which means that $W' \in \mathrm{Proj}^{-1}(\hat{U})$ and W' is not output-equivalent to any channel in $\mathrm{DMC}_{\mathcal{X},[n]}$ (see Corollary 10.1). Therefore, $\mathrm{Proj}(W') \in \hat{U}$ and $\mathrm{Proj}(W') \notin$

$\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ because W' is not output-equivalent to any channel in $\mathrm{DMC}_{\mathcal{X},[n]}$. This shows that every non-empty open subset of $\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ is not contained in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. We conclude that the interior of $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ in $\mathrm{DMC}^{(o)}_{\mathcal{X},[m]}$ is empty.

11.10.3 Proof of Lemma 11.5

Define $\mathrm{DMC}^{(o)}_{\mathcal{X},[0]} = \emptyset$, which is strongly closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$.

Let A and B be two disjoint strongly closed subsets of $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$. For every $n \geq 0$, let $A_n = A \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ and $B_n = B \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. Since A and B are strongly closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$, A_n and B_n are closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. Moreover, $A_n \cap B_n \subset A \cap B = \emptyset$.

Construct the sequences $(U_n)_{n\geq 0}, (U'_n)_{n\geq 0}, (K_n)_{n\geq 0}$ and $(K'_n)_{n\geq 0}$ recursively as follows:

$U_0 = U'_0 = K_0 = K'_0 = \emptyset \subset \mathrm{DMC}^{(o)}_{\mathcal{X},[0]}$. Since $A_0 = B_0 = \emptyset$, we have $A_0 \subset U_0 \subset K_0$ and $B_0 \subset U'_0 \subset K'_0$. Moreover, U_0 and U'_0 are open in $\mathrm{DMC}^{(o)}_{\mathcal{X},[0]}$, K_0 and K'_0 are closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},[0]}$, and $K_0 \cap K'_0 = \emptyset$.

Now let $n \geq 1$ and assume that we constructed $(U_i)_{0\leq i<n}, (U'_i)_{0\leq i<n}, (K_i)_{0\leq i<n}$ and $(K'_i)_{0\leq i<n}$ such that for every $0 \leq i < n$, we have $A_i \subset U_i \subset K_i \subset \mathrm{DMC}^{(o)}_{\mathcal{X},[i]}$, $B_i \subset U'_i \subset K'_i \subset \mathrm{DMC}^{(o)}_{\mathcal{X},[i]}$, U_i and U'_i are open in $\mathrm{DMC}^{(o)}_{\mathcal{X},[i]}$, K_i and K'_i are closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},[i]}$, and $K_i \cap K'_i = \emptyset$. Moreover, assume that $K_i \subset U_{i+1}$ and $K'_i \subset U'_{i+1}$ for every $0 \leq i < n-1$.

Let $C_n = A_n \cup K_{n-1}$ and $D_n = B_n \cup K'_{n-1}$. Since K_{n-1} and K'_{n-1} are closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n-1]}$ and since $\mathrm{DMC}^{(o)}_{\mathcal{X},[n-1]}$ is closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$, we can see that K_{n-1} and K'_{n-1} are closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. Therefore, C_n and D_n are closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. Moreover, we have

$$\begin{aligned}
C_n \cap D_n &= (A_n \cup K_{n-1}) \cap (B_n \cup K'_{n-1}) \\
&= (A_n \cap B_n) \cup (A_n \cap K'_{n-1}) \cup (K_{n-1} \cap B_n) \cup (K_{n-1} \cap K'_{n-1}) \\
&\overset{(a)}{=} \left(A_n \cap K'_{n-1} \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n-1]}\right) \cup \left(K_{n-1} \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n-1]} \cap B_n\right) \\
&= (A_{n-1} \cap K'_{n-1}) \cup (K_{n-1} \cap B_{n-1}) \subset (K_{n-1} \cap K'_{n-1}) \cup (K_{n-1} \cap K'_{n-1}) = \emptyset,
\end{aligned}$$

where (a) follows from the fact that $A_n \cap B_n = K_{n-1} \cap K'_{n-1} = \emptyset$ and the fact that $K_{n-1} \subset \mathrm{DMC}^{(o)}_{\mathcal{X},[n-1]}$ and $K'_{n-1} \subset \mathrm{DMC}^{(o)}_{\mathcal{X},[n-1]}$.

Since $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is normal (because it is metrizable), and since C_n and D_n are closed disjoint subsets of $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$, there exist two sets $U_n, U'_n \subset \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ that are open in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ and two sets $K_n, K'_n \subset \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ that are closed in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ such that $C_n \subset U_n \subset K_n$, $D_n \subset U'_n \subset K'_n$ and $K_n \cap K'_n = \emptyset$. Clearly, $A_n \subset U_n \subset K_n \subset \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$, $B_n \subset U'_n \subset K'_n \subset \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$, $K_{n-1} \subset U_n$ and $K'_{n-1} \subset U'_n$. This concludes the recursive construction.

Now define $U = \bigcup_{n\geq 0} U_n = \bigcup_{n\geq 1} U_n$ and $U' = \bigcup_{n\geq 0} U'_n = \bigcup_{n\geq 1} U'_n$. Since $A_n \subset U_n$ for

every $n \geq 1$, we have

$$\begin{aligned} A = A \cap \mathrm{DMC}^{(o)}_{\mathcal{X},*} &= A \cap \left(\bigcup_{n \geq 1} \mathrm{DMC}^{(o)}_{\mathcal{X},[n]} \right) \\ &= \bigcup_{n \geq 1} \left(A \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]} \right) = \bigcup_{n \geq 1} A_n \subset \bigcup_{n \geq 1} U_n = U. \end{aligned}$$

Moreover, for every $n \geq 1$ we have

$$\begin{aligned} U \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]} &= \left(\bigcup_{i \geq 1} U_i \right) \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]} \\ &\overset{(a)}{=} \left(\bigcup_{i \geq n} U_i \right) \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]} = \bigcup_{i \geq n} \left(U_i \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]} \right), \end{aligned}$$

where (a) follows from the fact that $U_i \subset K_i \subset U_{i+1}$ for every $i \geq 0$, which means that the sequence $(U_i)_{i \geq 1}$ is increasing.

For every $i \geq n$, we have $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]} \subset \mathrm{DMC}^{(o)}_{\mathcal{X},[i]}$ and U_i is open in $\mathrm{DMC}^{(o)}_{\mathcal{X},[i]}$, hence $U_i \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ is open in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. Therefore, $U \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]} = \bigcup_{i \geq n} \left(U_i \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]} \right)$ is open in $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$. Since this is true for every $n \geq 1$, we conclude that U is strongly open in $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$.

We can similarly show that $B \subset U'$ and that U' is strongly open in $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$. Finally, we have

$$\begin{aligned} U \cap U' &= \left(\bigcup_{n \geq 1} U_n \right) \cap \left(\bigcup_{n' \geq 1} U'_{n'} \right) = \bigcup_{n \geq 1, n' \geq 1} (U_n \cap U'_{n'}) \\ &\overset{(a)}{=} \bigcup_{n \geq 1} (U_n \cap U'_n) \subset \bigcup_{n \geq 1} (K_n \cap K'_n) = \emptyset, \end{aligned}$$

where (a) follows from the fact that for every $n \geq 1$ and every $n' \geq 1$, we have

$$U_n \cap U'_{n'} \subset U_{\max\{n,n'\}} \cap U'_{\max\{n,n'\}}$$

because $(U_n)_{n \geq 1}$ and $(U'_n)_{n \geq 1}$ are increasing. We conclude that $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ is normal.

11.10.4 Proof of Lemma 11.6

Let $W_1, W_2 \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, and let $\hat{W}_1$ and $\hat{W}_2$ be the $R^{(o)}_{\mathcal{X},\mathcal{Y}}$-equivalence classes of W_1 and W_2 respectively.

Fix $m \geq 1$, $p \in \Delta_{[m] \times \mathcal{X}}$ and $D \in \mathrm{DMC}_{\mathcal{Y},[m]}$. We have:

$$
\begin{aligned}
&\sum_{\substack{u\in[m],\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} p(u,x)W_1(y|x)D(u|y) \\
&\quad= \left(\sum_{\substack{u\in[m],\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} p(u,x)W_2(y|x)D(u|y)\right) + \sum_{\substack{u\in[m],\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} p(u,x)\cdot\big(W_1(y|x)-W_2(y|x)\big)\cdot D(u|y) \\
&\quad\le \left(\sup_{D'\in\mathrm{DMC}_{\mathcal{Y},[m]}} \sum_{\substack{u\in[m],\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} p(u,x)W_2(y|x)D'(u|y)\right) \\
&\qquad\qquad + \sum_{\substack{u\in[m],\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} p(u,x)\cdot\big(W_1(y|x)-W_2(y|x)\big)\cdot D(u|y) \\
&\quad\le P_c(p,W_2) + \sum_{\substack{u\in[m],\\ x\in\mathcal{X}}} p(u,x)\cdot \sum_{\substack{y\in\mathcal{Y}:\\ W_1(y|x)>W_2(y|x)}} \big(W_1(y|x)-W_2(y|x)\big)\cdot\left(\sum_{u'\in[m]} D(u'|y)\right) \\
&\quad= P_c(p,W_2) + \sum_{\substack{u\in[m],\\ x\in\mathcal{X}}} p(u,x)\cdot\left(\sum_{\substack{y\in\mathcal{Y}:\\ W_1(y|x)>W_2(y|x)}} \big(W_1(y|x)-W_2(y|x)\big)\right) \\
&\quad\overset{(a)}{\le} P_c(p,W_2) + \sum_{\substack{u\in[m],\\ x\in\mathcal{X}}} p(u,x)\cdot d_{\mathcal{X},\mathcal{Y}}(W_1,W_2) = P_c(p,W_2) + d_{\mathcal{X},\mathcal{Y}}(W_1,W_2),
\end{aligned}
$$

where (a) follows from the fact that

$$
\begin{aligned}
\sum_{\substack{y\in\mathcal{Y}:\\ W_1(y|x)>W_2(y|x)}} \big(W_1(y|x)-W_2(y|x)\big) &= \frac{1}{2}\sum_{y\in\mathcal{Y}}\big|W_1(y|x)-W_2(y|x)\big| \\
&\le \frac{1}{2}\sup_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\big|W_1(y|x)-W_2(y|x)\big| \\
&= d_{\mathcal{X},\mathcal{Y}}(W_1,W_2).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
P_c(p,W_1) &= \sup_{D\in\mathrm{DMC}_{\mathcal{Y},[m]}} \sum_{\substack{u\in[m],\\ x\in\mathcal{X},\\ y\in\mathcal{Y}}} p(u,x)W_1(y|x)D(u|y) \\
&\le P_c(p,W_2) + d_{\mathcal{X},\mathcal{Y}}(W_1,W_2).
\end{aligned}
$$

Similarly, we can show that $P_c(p,W_2) \le P_c(p,W_1) + d_{\mathcal{X},\mathcal{Y}}(W_1,W_2)$, hence

$$
|P_c(p,W_1) - P_c(p,W_2)| \le d_{\mathcal{X},\mathcal{Y}}(W_1,W_2).
$$

We conclude that

$$\begin{aligned} d^{(o)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1,\hat{W}_2) &= \sup_{\substack{m\geq 1,\\ p\in\Delta_{[m]\times\mathcal{X}}}} |P_c(p,\hat{W}_1)-P_c(p,\hat{W}_2)| \\ &= \sup_{\substack{m\geq 1,\\ p\in\Delta_{[m]\times\mathcal{X}}}} |P_c(p,W_1)-P_c(p,W_2)| \\ &\leq d_{\mathcal{X},\mathcal{Y}}(W_1,W_2). \end{aligned}$$

11.10.5 Proof of Lemma 11.7

Let $\gamma \in \Gamma(\mathrm{MP}_{\hat{W}}, \mathrm{MP}_{\hat{W}'})$ be a measure on $\Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}}$ that couples $\mathrm{MP}_{\hat{W}}$ and $\mathrm{MP}_{\hat{W}'}$.

Let $S = \mathrm{supp}(\mathrm{MP}_{\hat{W}})$ and $S' = \mathrm{supp}(\mathrm{MP}_{\hat{W}'})$ be the supports of $\hat{W}$ and $\hat{W}'$ respectively. Since $\mathrm{MP}_{\hat{W}}$ and $\mathrm{MP}_{\hat{W}'}$ are finitely supported, γ is also finitely supported and its support is a subset of $S \times S'$. Therefore, there exists a collection of coefficients $\alpha_{p,p'} \in [0,1]$ such that

$$\gamma = \sum_{\substack{p\in S,\\ p'\in S'}} \alpha_{p,p'}\delta_{(p,p')},$$

where $\delta_{(p,p')}$ is a Dirac measure centered at $(p,p') \in \Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}}$. Since $\mathrm{MP}_{\hat{W}}$ and $\mathrm{MP}_{\hat{W}'}$ are the marginals of γ on the first and the second factors respectively, we have $\mathrm{MP}_{\hat{W}}(p) = \sum_{p'\in S'} \alpha_{p,p'}$ for every $p \in S$. Similarly, $\mathrm{MP}_{\hat{W}'}(p') = \sum_{p\in S} \alpha_{p,p'}$ for every $p' \in S'$.

Let $\mathcal{Y} = S \times S'$ and define the channels $W, W' \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ as:

$$W(p,p'|x) = |\mathcal{X}|\alpha_{p,p'} \cdot p(x),$$

and

$$W'(p,p'|x) = |\mathcal{X}|\alpha_{p,p'} \cdot p'(x).$$

For every $x \in \mathcal{X}$, we have

$$\begin{aligned} \sum_{(p,p')\in\mathcal{Y}} W(p,p'|x) &= |\mathcal{X}| \sum_{(p,p')\in S\times S'} \alpha_{p,p'} \cdot p(x) = |\mathcal{X}| \sum_{p\in S} \mathrm{MP}_{\hat{W}}(p) \cdot p(x) \\ &= |\mathcal{X}| \int_{\Delta_{\mathcal{X}}} p(x) \cdot d\mathrm{MP}_{\hat{W}}(p) = |\mathcal{X}| \frac{1}{|\mathcal{X}|} = 1. \end{aligned}$$

Similarly, $\sum_{(p,p')\in\mathcal{Y}} W'(p,p'|x) = 1$. Therefore, W and W' are valid channels.

For every $(p,p') \in \mathcal{Y}$, we have

$$P^o_W(p,p') = \sum_{x\in\mathcal{X}} \frac{1}{|\mathcal{X}|} W(p,p'|x) = \sum_{x\in\mathcal{X}} \alpha_{p,p'} \cdot p(x) = \alpha_{p,p'}.$$

Therefore, $\mathrm{Im}(W) = \{(p,p') \in \mathcal{Y} :\ \alpha_{p,p'} > 0\}$. For every $(p,p') \in \mathrm{Im}(W)$ and every $x \in \mathcal{X}$, we have:

$$W^{-1}_{p,p'}(x) = \frac{W(p,p'|x)}{|\mathcal{X}|P^o_W(p,p')} = \frac{|\mathcal{X}|\alpha_{p,p'} \cdot p(x)}{|\mathcal{X}|\alpha_{p,p'}} = p(x),$$

hence $W^{-1}_{p,p'} = p$ for every $(p, p') \in \mathrm{Im}(W)$, which shows that $\mathrm{supp}(\mathrm{MP}_W) \subset S$. Similarly, we can show that

$$\mathrm{Im}(W') = \{(p, p') \in \mathcal{Y} : \ \alpha_{p,p'} > 0\},$$

$\mathrm{supp}(\mathrm{MP}_{W'}) \subset S'$, and for every $(p, p') \in \mathcal{Y}$, $P^o_{W'}(p, p') = \alpha_{p,p'}$ and $W'^{-1}_{p,p'} = p'$.

For every $p \in S$, we have:

$$\mathrm{MP}_W(p) = \sum_{\substack{y \in \mathrm{Im}(W),\\ W_y^{-1} = p}} P^o_W(y) = \sum_{\substack{p' \in S',\\ \alpha_{p,p'} > 0}} \alpha_{p,p'} = \sum_{p' \in S'} \alpha_{p,p'} = \mathrm{MP}_{\hat{W}}(p) > 0.$$

This shows that $\mathrm{supp}(\mathrm{MP}_W) = S = \mathrm{supp}(\mathrm{MP}_{\hat{W}})$ and $\mathrm{MP}_W(p) = \mathrm{MP}_{\hat{W}}(p)$ for every $p \in S$. Therefore, $\mathrm{MP}_W = \mathrm{MP}_{\hat{W}}$ and so W is output-equivalent to every channel in $\hat{W}$. Similarly, we can show that $\mathrm{MP}_{W'} = \mathrm{MP}_{\hat{W}'}$ and W' is output-equivalent to every channel in $\hat{W}'$.

Let $\tilde{W}$ and $\tilde{W}'$ be the $R^{(o)}_{\mathcal{X},\mathcal{Y}}$-equivalence classes of W and W' respectively. We can write $\hat{W} = \tilde{W}$ and $\hat{W}' = \tilde{W}'$ because of the canonical identification of $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ with $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$, where $n = |\mathcal{Y}|$. We have:

$$\begin{aligned}
&d^{(o)}_{\mathcal{X},*}(\hat{W}, \hat{W}') \\
&\quad = d^{(o)}_{\mathcal{X},\mathcal{Y}}(\tilde{W}, \tilde{W}') \overset{(a)}{\leq} d_{\mathcal{X},\mathcal{Y}}(W, W') = \frac{1}{2} \max_{x \in \mathcal{X}} \sum_{(p,p') \in \mathcal{Y}} |W(p, p'|x) - W'(p, p'|x)| \\
&\quad = \frac{1}{2} \max_{x \in \mathcal{X}} \sum_{\substack{p \in S,\\ p' \in S'}} \Big| |\mathcal{X}| \alpha_{p,p'} \cdot p(x) - |\mathcal{X}| \alpha_{p,p'} \cdot p'(x) \Big| = \frac{1}{2} |\mathcal{X}| \max_{x \in \mathcal{X}} \sum_{\substack{p \in S,\\ p' \in S'}} \alpha_{p,p'} \cdot |p(x) - p'(x)| \\
&\quad \leq \frac{1}{2} |\mathcal{X}| \sum_{x \in \mathcal{X}} \sum_{\substack{p \in S,\\ p' \in S'}} \alpha_{p,p'} \cdot |p(x) - p'(x)| = \frac{1}{2} |\mathcal{X}| \sum_{\substack{p \in S,\\ p' \in S'}} \alpha_{p,p'} \sum_{x \in \mathcal{X}} |p(x) - p'(x)| \\
&\quad = \frac{1}{2} |\mathcal{X}| \sum_{\substack{p \in S,\\ p' \in S'}} \alpha_{p,p'} \|p - p'\|_1 = |\mathcal{X}| \sum_{\substack{p \in S,\\ p' \in S'}} \alpha_{p,p'} d(p, p') = |\mathcal{X}| \int_{\Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}}} d(p, p') \cdot d\gamma(p, p'),
\end{aligned}$$

where (a) follows from Lemma 11.6, and $d(p, p') = \frac{1}{2}\|p - p'\|_1$ is the total-variation distance between p and p'. Therefore,

$$\begin{aligned}
d^{(o)}_{\mathcal{X},*}(\hat{W}, \hat{W}') &\leq |\mathcal{X}| \inf_{\gamma \in \Gamma(\mathrm{MP}_{\hat{W}}, \mathrm{MP}_{\hat{W}'})} \int_{\Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}}} d(p, p') \cdot d\gamma(p, p') \\
&= |\mathcal{X}| \cdot W_1(\mathrm{MP}_{\hat{W}}, \mathrm{MP}_{\hat{W}'}).
\end{aligned}$$

11.10.6 Proof of Proposition 11.10

If $|\mathcal{X}| = 1$, $\Delta_{\mathcal{X}}$ consists of a single probability distribution and $\mathcal{MP}(\mathcal{X})$ consists of a single meta-probability measure which is balanced and finitely supported, so $\mathcal{MP}(\mathcal{X}) = \mathcal{MP}_b(\mathcal{X}) = \mathcal{MP}_{bf}(\mathcal{X})$.

Now assume that $|\mathcal{X}| \geq 2$. We start by showing that $\mathcal{MP}_b(\mathcal{X})$ is weakly-$*$ closed.

For every $x \in \mathcal{X}$. Consider the mapping $f_x : \Delta_{\mathcal{X}} \to \mathbb{R}$ defined as $f_x(p) = p(x)$. Clearly, f_x is bounded and continuous. Therefore, the mapping

$$F_x : \mathcal{MP}(\mathcal{X}) \to \mathbb{R}$$

defined as

$$F_x(\mathrm{MP}) = \int_{\Delta_{\mathcal{X}}} f_x d\mathrm{MP} = \int_{\Delta_{\mathcal{X}}} p(x) \cdot d\mathrm{MP}(p)$$

is continuous in the weak-$*$ topology. Therefore, $F_x^{-1}\left(\left\{\frac{1}{|\mathcal{X}|}\right\}\right)$ is weakly-$*$ closed. It is easy to see that $\mathcal{MP}_b(\mathcal{X}) = \bigcap_{x \in \mathcal{X}} F_x^{-1}\left(\left\{\frac{1}{|\mathcal{X}|}\right\}\right)$. This proves that $\mathcal{MP}_b(\mathcal{X})$, which is the finite intersection of weakly-$*$ closed sets, is weakly-$*$ closed.

It remains to show that $\mathcal{MP}_{bf}(\mathcal{X})$ is weakly-$*$ dense in $\mathcal{MP}_b(\mathcal{X})$. We will show that for every $\epsilon > 0$ and every $\mathrm{MP} \in \mathcal{MP}_b(\mathcal{X})$, there exists $\mathrm{MP}' \in \mathcal{MP}_{bf}(\mathcal{X})$ such that $W_1(\mathrm{MP}, \mathrm{MP}') < \epsilon$.

Fix $0 < \epsilon < 1$ and let $\mathrm{MP} \in \mathcal{MP}_b(\mathcal{X})$ be any balanced meta-probability measure on $\mathcal{X}$, i.e., for every $x \in \mathcal{X}$ we have

$$\int_{\Delta_{\mathcal{X}}} p(x) d\mathrm{MP}(p) = \frac{1}{|\mathcal{X}|}.$$

Now fix $x \in \mathcal{X}$. By the definition of the Lebesgue integral, there exists a finite partition $\{B_{x,i}\}_{1 \le i \le k_x}$ of $\Delta_{\mathcal{X}}$ and a sequence of positive numbers $(b_{x,i})_{1 \le i \le k_x}$ such that for every $1 \le i \le k_x$, $B_{x,i}$ is a Borel set of $\Delta_{\mathcal{X}}$, $b_{x,i} \le p(x)$ for every $p \in B_{x,i}$, and

$$\sum_{i=1}^{k_x} b_{x,i} \mathrm{MP}(B_{x,i}) \ge \left(\int_{\Delta_{\mathcal{X}}} p(x) \cdot d\mathrm{MP}(p)\right) - \frac{\epsilon}{12|\mathcal{X}|} = \frac{1}{|\mathcal{X}|} - \frac{\epsilon}{12|\mathcal{X}|}.$$

By applying the same reasoning on the function $1 - p(x) \ge 0$, we can find a finite partition $\{C_{x,i}\}_{1 \le i \le m_x}$ of $\Delta_{\mathcal{X}}$ and a sequence of positive numbers $(c_{x,i})_{1 \le i \le m_x}$ such that for every $1 \le i \le m_x$, $C_{x,i}$ is a Borel set of $\Delta_{\mathcal{X}}$, $c_{x,i} \ge p(x)$ for every $p \in C_{x,i}$ and

$$\sum_{i=1}^{m_x} c_{x,i} \mathrm{MP}(C_{x,i}) \le \left(\int_{\Delta_{\mathcal{X}}} p(x) \cdot d\mathrm{MP}(p)\right) + \frac{\epsilon}{12|\mathcal{X}|} = \frac{1}{|\mathcal{X}|} + \frac{\epsilon}{12|\mathcal{X}|}.$$

Let d be the total-variation distance on $\Delta_{\mathcal{X}}$, i.e., $d(p, p') = \frac{1}{2}\|p - p'\|_1$. Since $\Delta_{\mathcal{X}}$ is compact, it can be covered by a finite number of open balls of radius $\frac{\epsilon}{4}$, i.e., there exist h points $p'_1, \ldots, p'_h$ such that $\Delta_{\mathcal{X}} = \bigcup_{i=1}^{h} B_{\frac{\epsilon}{4}}(p'_i) = \bigcup_{i=1}^{h} \left\{p \in \Delta_{\mathcal{X}} : \ d(p, p'_i) < \frac{\epsilon}{4}\right\}$. For every $1 \le i \le h$, define the set

$$D_i = B_{\frac{\epsilon}{4}}(p'_i) \setminus \left(\bigcup_{1 \le j < i} B_{\frac{\epsilon}{4}}(p'_j)\right).$$

Clearly, the sets $\{D_i\}_{1 \le i \le h}$ are disjoint Borel sets that cover $\Delta_{\mathcal{X}}$. Let $n = h \times \prod_{x \in \mathcal{X}} (k_x \cdot m_x)$, and let $A_1, \ldots, A_n$ be the Borel sets obtained by intersecting the sets

in the collections $\{D_1, \ldots, D_h\}$, $\{B_{x,i}\}_{1\le i\le k_x}$ and $\{C_{x,i}\}_{1\le i\le m_x}$ for every $x \in \mathcal{X}$. In other words,

$$\{A_i : 1 \le i \le n\} = \left\{ D_i \cap \bigcap_{x\in\mathcal{X}} (B_{x,i_x} \cap C_{x,j_x}) : 1 \le i \le h, \text{ and } \forall x \in \mathcal{X},\ 1 \le i_x \le k_x \text{ and } 1 \le j_x \le m_x \right\}.$$

For every $1 \le i \le n$, let $l_{x,i} = b_{x,i'}$ where i' is the unique integer satisfying $1 \le i' \le k_x$ and $A_i \subset B_{x,i'}$. Similarly, let $u_{x,i} = c_{x,i''}$ where i'' is the unique integer satisfying $1 \le i'' \le k_x$ and $A_i \subset C_{x,i''}$. Clearly, $l_{x,i} \le p(x) \le u_{x,i}$ for every $x \in A_i$. Moreover,

$$\sum_{i=1}^{n} l_{x,i}\mathrm{MP}(A_i) = \sum_{i=1}^{k_x} b_{x,i}\mathrm{MP}(B_{x,i}) \ge \frac{1}{|\mathcal{X}|} - \frac{\epsilon}{12|\mathcal{X}|},$$

and

$$\sum_{i=1}^{n} u_{x,i}\mathrm{MP}(A_i) = \sum_{i=1}^{m_x} c_{x,i}\mathrm{MP}(C_{x,i}) \le \frac{1}{|\mathcal{X}|} + \frac{\epsilon}{12|\mathcal{X}|}.$$

For every $1 \le i \le n$, choose $p_i \in A_i$ arbitrarily. Let j_i be the unique integer such that $A_i \subset D_{j_i}$. Since $D_{j_i} \subset B_{\frac{\epsilon}{4}}(p'_{j_i})$, we have $d(p, p'_{j_i}) < \frac{\epsilon}{4}$ for every $p \in A_i$. Therefore, $d(p, p_i) \le d(p, p'_{j_i}) + d(p'_{j_i}, p_i) < \frac{\epsilon}{2}$ for every $p \in A_i$.

Define the mapping $f : \Delta_{\mathcal{X}} \to \Delta_{\mathcal{X}}$ as $f(p) = p_i$ for every $p \in A_i$. Clearly, $d(p, f(p)) < \frac{\epsilon}{2}$ for every $p \in \Delta_{\mathcal{X}}$.

Now let $\mathrm{MP}_f = f_{\#}(\mathrm{MP})$, where $f_{\#}(\mathrm{MP})$ is the push-forward measure of MP by the mapping f, i.e., $\mathrm{MP}_f(B) = (f_{\#}(\mathrm{MP}))(B) = \mathrm{MP}\big(f^{-1}(B)\big)$ for every Borel set B of $\Delta_{\mathcal{X}}$. We have:

$$\mathrm{MP}_f(B) = \sum_{p_i\in B} \mathrm{MP}\big(f^{-1}(\{p_i\})\big) = \sum_{p_i\in B} \mathrm{MP}(A_i) = \sum_{p_i\in B} \alpha_i,$$

where $\alpha_i = \mathrm{MP}(A_i)$ for every $1 \le i \le n$. Therefore, MP_f is finitely supported and

$$\mathrm{supp}(\mathrm{MP}_f) \subset \{p_i : 1 \le i \le n\}.$$

Moreover, $\mathrm{MP}_f(p_i) = \alpha_i$ for every $1 \le i \le n$.

Now define the mapping $f_\times : \Delta_{\mathcal{X}} \to \Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}}$ as $f_\times(p) = (p, f(p))$, and define the measure γ_f on $\Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}}$ as the push-forward of MP by $f_\times$, i.e., $\gamma_f(B) = \mathrm{MP}(f_\times^{-1}(B))$ for every Borel set B of $\Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}}$. It is easy to see that the marginals of γ_f on the first and second factors are MP and MP_f respectively. Therefore, γ_f is a coupling between MP and MP_f, hence

$$\begin{aligned} W_1(\mathrm{MP}, \mathrm{MP}_f) &= \inf_{\gamma\in\Gamma(\mathrm{MP},\mathrm{MP}_f)} \int_{\Delta_{\mathcal{X}}\times\Delta_{\mathcal{X}}} d(p,p') \cdot d\gamma(p,p') \\ &\le \int_{\Delta_{\mathcal{X}}\times\Delta_{\mathcal{X}}} d(p,p') \cdot d\gamma_f(p,p') \overset{(a)}{=} \int_{\Delta_{\mathcal{X}}} d(p, f(p)) \cdot d\mathrm{MP}(p) \overset{(b)}{\le} \frac{\epsilon}{2}, \end{aligned}$$

where (a) follows from the fact that γ_f is the push-forward of MP by $f_\times(p) = (p, f(p))$. (b) follows from the fact that $d(p, f(p)) < \frac{\epsilon}{2}$ for every $p \in \Delta_{\mathcal{X}}$. Therefore, MP_f well approximates MP and it is finitely supported. However, MP_f may not be

balanced, so more work needs to be done in order to find a balanced and finitely supported meta-probability measure that well approximates MP.

For every $x \in \mathcal{X}$, we have:

$$\int_{\Delta_{\mathcal{X}}} p(x) \cdot d\mathrm{MP}_f(p) \overset{(a)}{=} \int_{\Delta_{\mathcal{X}}} (f(p))(x) \cdot d\mathrm{MP}(p) = \sum_{i=1}^{n} p_i(x)\mathrm{MP}(A_i)$$
$$\overset{(b)}{\geq} \sum_{i=1}^{n} l_{i,x}\mathrm{MP}(A_i) \geq \frac{1}{|\mathcal{X}|} - \frac{\epsilon}{12|\mathcal{X}|},$$

where (a) follows from the fact that MP_f is the push-forward of MP by f. (b) follows from the fact that $p_i \in A_i$ and so $p_i(x) \geq l_{i,x}$ for every $1 \leq i \leq n$. Similarly, we have

$$\int_{\Delta_{\mathcal{X}}} p(x) \cdot d\mathrm{MP}_f(p) = \sum_{i=1}^{n} p_i(x)\mathrm{MP}(A_i) \overset{(c)}{\leq} \sum_{i=1}^{n} u_{i,x}\mathrm{MP}(A_i) \leq \frac{1}{|\mathcal{X}|} + \frac{\epsilon}{12|\mathcal{X}|},$$

where (c) follows from the fact that $p_i \in A_i$ and so $p_i(x) \leq u_{i,x}$ for every $1 \leq i \leq n$. We conclude that for every $x \in \mathcal{X}$, we have

$$\left| \pi_{\mathcal{X}}(x) - \int_{\Delta_{\mathcal{X}}} p(x) \cdot d\mathrm{MP}_f(p) \right| \leq \frac{\epsilon}{12|\mathcal{X}|},$$

where $\pi_{\mathcal{X}}$ is the uniform distribution on $\mathcal{X}$. Define $\tilde{p} \in \Delta_{\mathcal{X}}$ as:

$$\tilde{p} = \int_{\Delta_{\mathcal{X}}} p \cdot d\mathrm{MP}_f(p).$$

For every $x \in \mathcal{X}$, define

$$p'(x) = \frac{6(\pi_{\mathcal{X}}(x) - \tilde{p}(x))}{\epsilon} + \tilde{p}(x).$$

Clearly, $\sum_{x \in \mathcal{X}} p'(x) = 1$. Moreover,

$$p'(x) = \frac{6(\pi_{\mathcal{X}}(x) - \tilde{p}(x))}{\epsilon} + \tilde{p}(x)$$
$$\overset{(a)}{\geq} \frac{6\left(\pi_{\mathcal{X}}(x) - \pi_{\mathcal{X}}(x) - \frac{\epsilon}{12|\mathcal{X}|}\right)}{\epsilon} + \frac{1}{|\mathcal{X}|} - \frac{\epsilon}{12|\mathcal{X}|} = \frac{1}{2|\mathcal{X}|} - \frac{\epsilon}{12|\mathcal{X}|} \geq 0,$$

Where (a) follows from the fact that $|\pi_{\mathcal{X}}(x) - \tilde{p}(x)| \leq \frac{\epsilon}{12|\mathcal{X}|}$. We conclude that $p' \in \Delta_{\mathcal{X}}$. Now define the meta-probability measure MP' as follows:

$$\mathrm{MP}' = \frac{\epsilon}{6} \cdot \delta_{p'} + \left(1 - \frac{\epsilon}{6}\right) \mathrm{MP}_f,$$

where $\delta_{p'}$ is a Dirac measure centered at p'.

For every $x \in \mathcal{X}$, we have

$$\begin{aligned}\int_{\Delta_{\mathcal{X}}} p(x) \cdot d\mathrm{MP}'(p) &= \frac{\epsilon}{6} \cdot p'(x) + \left(1 - \frac{\epsilon}{6}\right) \int_{\Delta_{\mathcal{X}}} p(x) \cdot d\mathrm{MP}_f(p) \\ &= \frac{\epsilon}{6} \cdot p'(x) + \left(1 - \frac{\epsilon}{6}\right) \cdot \tilde{p}(x) \\ &= \pi_{\mathcal{X}}(x) - \tilde{p}(x) + \frac{\epsilon}{6} \cdot \tilde{p}(x) + \left(1 - \frac{\epsilon}{6}\right) \tilde{p}(x) = \pi_{\mathcal{X}}(x).\end{aligned}$$

Therefore, MP' is balanced and finitely supported. Moreover,

$$\begin{aligned}W_1(\mathrm{MP}, \mathrm{MP}') &\leq W_1(\mathrm{MP}, \mathrm{MP}_f) + W_1(\mathrm{MP}_f, \mathrm{MP}') \overset{(a)}{\leq} \frac{\epsilon}{2} + \|\mathrm{MP}_f - \mathrm{MP}'\|_{TV} \\ &= \frac{\epsilon}{2} + \left\|\mathrm{MP}_f - \left(1 - \frac{\epsilon}{6}\right) \mathrm{MP}_f - \frac{\epsilon}{6} \cdot \delta_{p'}\right\|_{TV} \\ &\leq \frac{\epsilon}{2} + \left\|\frac{\epsilon}{6} \cdot \mathrm{MP}_f\right\|_{TV} + \left\|\frac{\epsilon}{6}\delta_{p'}\right\|_{TV} = \frac{\epsilon}{2} + \frac{\epsilon}{6} + \frac{\epsilon}{6} < \epsilon,\end{aligned}$$

where (a) follows from the fact that the 1^{st} Wasserstein metric is upper bounded by the total-variation multiplied by the diameter of $\Delta_{\mathcal{X}}$ (which is equal to 1 in our case) [80]. We conclude that $\mathcal{MP}_{bf}(\mathcal{X})$ is dense in $\mathcal{MP}_b(\mathcal{X})$ which is weakly-$*$ closed. Therefore, $\mathcal{MP}_b(\mathcal{X})$ is the weak-$*$ closure of $\mathcal{MP}_{bf}(\mathcal{X})$.

11.10.7 Proof of Proposition 11.18

If $|\mathcal{Y}| = 1$, then $\Delta_{\mathcal{Y}}$ contains only one point and so $|\operatorname{CE}(W)| = 1$ for every $W \in \mathrm{DMC}_{[n],\mathcal{Y}}$ and every $n \geq 1$. Therefore, $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}} = \mathrm{DMC}^{(i)}_{[1],\mathcal{Y}}$ for every $n \geq 1$.

If $|\mathcal{Y}| = 2$, then $\Delta_{\mathcal{Y}}$ is a one dimensional segment. Therefore, there are at most two convex-extreme points for any finite subset of $\Delta_{\mathcal{Y}}$. This means that $|\operatorname{CE}(W)| \leq 2$ for every $W \in \mathrm{DMC}_{[n],\mathcal{Y}}$ and every $n \geq 2$. Therefore, $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}} = \mathrm{DMC}^{(i)}_{[2],\mathcal{Y}}$ for every $n \geq 2$.

Now assume that $|\mathcal{Y}| \geq 3$. Let $\hat{U}$ be an arbitrary non-empty open subset of $(\mathrm{DMC}^{(i)}_{[m],\mathcal{Y}}, \mathcal{T}^{(i)}_{[m],\mathcal{Y}})$ and let Proj be the projection onto the $R^{(i)}_{[m],\mathcal{Y}}$-equivalence classes. $\mathrm{Proj}^{-1}(\hat{U})$ is open in the metric space $(\mathrm{DMC}_{[m],\mathcal{Y}}, d_{[m],\mathcal{Y}})$. Let $\hat{W} \in \hat{U}$ and define $r = \operatorname{irank}(\hat{W})$. Let $P_1, \ldots, P_r \in \Delta_{\mathcal{Y}}$ be such that $\operatorname{CE}(\hat{W}) = \{P_1, \ldots, P_r\}$. Define the channel $W \in \mathrm{DMC}_{[m],\mathcal{Y}}$ as follows:

$$W(y|i) = \begin{cases} P_i(y) & \text{if } 1 \leq i < r, \\ P_r(y) & \text{if } r \leq i \leq m. \end{cases}$$

Clearly $\operatorname{CE}(W) = \operatorname{CE}(\hat{W})$ and so $W \in \hat{W}$ which implies that $W \in \mathrm{Proj}^{-1}(\hat{U})$. Since $\mathrm{Proj}^{-1}(\hat{U})$ is open in the metric space $(\mathrm{DMC}_{[m],\mathcal{Y}}, d_{[m],\mathcal{Y}})$, there exists $\epsilon > 0$ such that $\mathrm{Proj}^{-1}(\hat{U})$ contains the open ball of center W and radius ϵ.

We will show that there exists $W' \in \mathrm{DMC}_{[m],\mathcal{Y}}$ such that $\operatorname{irank}(W') = m > n$ and $d_{[m],\mathcal{Y}}(W, W') < \epsilon$. If $r = \operatorname{irank}(W) = m$, take $W' = W$.

Assume that $r = \operatorname{irank}(W) < m$. Since $|\mathcal{Y}| \geq 3$, the dimension of $\Delta_{\mathcal{Y}}$ is at least 2. Therefore, we can find $P_{r+1} \in \Delta_{\mathcal{Y}}$ such that $\|P_r - P_{r+1}\|_{TV} < \epsilon$ and $\operatorname{CE}(\{P_1, \ldots, P_{r+1}\}) = \{P_1, \ldots, P_{r+1}\}$. By repeating this procedure $m - r$ times, we

obtain $P_{r+1}, \dots, P_m \in \Delta\mathcal{Y}$ such that $\|P_r - P_i\|_{TV} < \epsilon$ for every $r+1 \le i \le m$, and $\mathrm{CE}(\{P_1, \dots, P_m\}) = \{P_1, \dots, P_m\}$. Define the channel $W' \in \Delta_{[m],\mathcal{Y}}$ as:

$$W'(y|i) = P_i(y).$$

We have $\mathrm{CE}(W') = \mathrm{CE}(\{P_1, \dots, P_m\}) = \{P_1, \dots, P_m\}$. Therefore, $\mathrm{irank}(W') = m$. Moreover,

$$d_{[m],\mathcal{Y}}(W, W') = \max_{1 \le i \le m} \|W_i - W'_i\|_{TV} = \max_{r+1 \le i \le m} \|P_r - P_i\|_{TV} < \epsilon.$$

This means that $W' \in \mathrm{Proj}^{-1}(\hat{U})$ and W' is not input-equivalent to any channel in $\mathrm{DMC}_{[n],\mathcal{Y}}$ (see Proposition 10.4). Therefore, $\mathrm{Proj}(W') \in \hat{U}$ and $\mathrm{Proj}(W') \notin \mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ because W' is not input-equivalent to any channel in $\mathrm{DMC}_{[n],\mathcal{Y}}$. This shows that every non-empty open subset of $\mathrm{DMC}^{(i)}_{[m],\mathcal{Y}}$ is not contained in $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$. We conclude that the interior of $\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}$ in $\mathrm{DMC}^{(i)}_{[m],\mathcal{Y}}$ is empty.

11.10.8 Proof of Proposition 11.25

Fix $n, m \ge 1$ and let $l \in \Delta_{[n]\times[m]}$. Define $\mathcal{G}_1 = ([n], \mathcal{X}, \mathcal{Y}, [m], l, W_1)$ and $\mathcal{G}_2 = ([n], \mathcal{X}, \mathcal{Y}, [m], l, W_2)$. For every $S \in \mathcal{S}_{[n],\mathcal{X},\mathcal{Y},[m]}$, we have:

$$\begin{aligned}
&\hat{\$}(S, \mathcal{G}_1) \\
&= \frac{1}{n} \sum_{u \in [n]} \hat{\$}(u, S, \mathcal{G}_1) = \frac{1}{n} \sum_{u \in [n]} \sum_{i=1}^{n_S} \alpha_S(i) \sum_{y \in \mathcal{Y}} W_1\big(y\big|f_{i,S}(u)\big) l\big(u, g_{i,S}(y)\big) \\
&= \left(\frac{1}{n} \sum_{u \in [n]} \sum_{i=1}^{n_S} \alpha_S(i) \sum_{y \in \mathcal{Y}} W_2\big(y\big|f_{i,S}(u)\big) l\big(u, g_{i,S}(y)\big) \right) \\
&\qquad + \frac{1}{n} \sum_{u \in [n]} \sum_{i=1}^{n_S} \alpha_S(i) \sum_{y \in \mathcal{Y}} \Big(W_1\big(y\big|f_{i,S}(u)\big) - W_2\big(y\big|f_{i,S}(u)\big) \Big) l\big(u, g_{i,S}(y)\big) \\
&\le \hat{\$}(S, \mathcal{G}_2) \\
&\quad + \sum_{i=1}^{n_S} \frac{\alpha_S(i)}{n} \sum_{u \in [n]} \sum_{\substack{y \in \mathcal{Y}, \\ W_1(y|f_{i,S}(u)) \ge W_2(y|f_{i,S}(u))}} \Big(W_1\big(y\big|f_{i,S}(u)\big) - W_2\big(y\big|f_{i,S}(u)\big) \Big) l\big(u, g_{i,S}(y)\big) \\
&\overset{(a)}{\le} \hat{\$}(S, \mathcal{G}_2) + \sum_{i=1}^{n_S} \frac{\alpha_S(i)}{n} \sum_{u \in [n]} \sum_{\substack{y \in \mathcal{Y}, \\ W_1(y|f_{i,S}(u)) \ge W_2(y|f_{i,S}(u))}} \Big(W_1\big(y\big|f_{i,S}(u)\big) - W_2\big(y\big|f_{i,S}(u)\big) \Big),
\end{aligned}$$

where (a) follows from the fact that $l(u, g_{i,S}(y)) \le 1$ (because $l \in \Delta_{[n]\times[m]}$). Therefore,

$$\begin{aligned}
\hat{\$}(S,\mathcal{G}_1) &\le \hat{\$}(S,\mathcal{G}_2) + \sum_{i=1}^{n_S} \frac{\alpha_S(i)}{n} \sum_{u\in[n]} \frac{1}{2} \sum_{y\in\mathcal{Y}} \left| W_1\big(y\big|f_{i,S}(u)\big) - W_2\big(y\big|f_{i,S}(u)\big)\right| \\
&\le \hat{\$}(S,\mathcal{G}_2) + \sum_{i=1}^{n_S} \frac{\alpha_S(i)}{n} \sum_{u\in[n]} \max_{x\in\mathcal{X}} \frac{1}{2} \sum_{y\in\mathcal{Y}} |W_1(y|x) - W_2(y|x)| \\
&= \hat{\$}(S,\mathcal{G}_2) + \sum_{i=1}^{n_S} \frac{\alpha_S(i)}{n} \sum_{u\in[n]} d_{\mathcal{X},\mathcal{Y}}(W_1,W_2) = \hat{\$}(S,\mathcal{G}_2) + d_{\mathcal{X},\mathcal{Y}}(W_1,W_2) \\
&\le d_{\mathcal{X},\mathcal{Y}}(W_1,W_2) + \sup_{S'\in\mathcal{S}_{[n],\mathcal{X},\mathcal{Y},[m]}} \hat{\$}(S',\mathcal{G}_2) = d_{\mathcal{X},\mathcal{Y}}(W_1,W_2) + \$_{\rm opt}(\mathcal{G}_2).
\end{aligned}$$

We conclude that

$$\$_{\rm opt}(\mathcal{G}_1) = \sup_{S\in\mathcal{S}_{[n],\mathcal{X},\mathcal{Y},[m]}} \hat{\$}(S,\mathcal{G}_1) \le \$_{\rm opt}(\mathcal{G}_2) + d_{\mathcal{X},\mathcal{Y}}(W_1,W_2),$$

hence

$$\$_{\rm opt}(\mathcal{G}_1) - \$_{\rm opt}(\mathcal{G}_2) \le d_{\mathcal{X},\mathcal{Y}}(W_1,W_2).$$

We can show similarly that $\$_{\rm opt}(\mathcal{G}_2) - \$_{\rm opt}(\mathcal{G}_2) \le d_{\mathcal{X},\mathcal{Y}}(W_1,W_2)$. Therefore,

$$\begin{aligned}
|\$_{\rm opt}(l,\hat{W}_1) - \$_{\rm opt}(l,\hat{W}_2)| &= |\$_{\rm opt}(l,W_1) - \$_{\rm opt}(l,W_2)| \\
&= |\$_{\rm opt}(\mathcal{G}_1) - \$_{\rm opt}(\mathcal{G}_2)| \le d_{\mathcal{X},\mathcal{Y}}(W_1,W_2).
\end{aligned}$$

We conclude that

$$d^{(s)}_{\mathcal{X},\mathcal{Y}}(\hat{W}_1,\hat{W}_2) = \sup_{\substack{n,m\ge 1,\\ l\in\Delta_{[n]\times[m]}}} |\$_{\rm opt}(l,\hat{W}_1) - \$_{\rm opt}(l,\hat{W}_2)| \le d_{\mathcal{X},\mathcal{Y}}(W_1,W_2).$$

11.10.9 Proof of Proposition 11.26

Corollary 11.17 implies that $\mathrm{Proj}_2(D_g \circ W \circ D_f) = \mathrm{Proj}_2(D_g \circ W' \circ D_f)$ if and only if $W R^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1} W'$. Therefore, $\mathrm{Proj}_2(D_g \circ W' \circ D_f)$ does not depend on $W' \in \hat{W}$, hence F is well defined. Corollary 11.17 also shows that $\mathrm{Proj}_2(D_g \circ W' \circ D_f)$ does not depend on the particular choice of the surjection f or the injection g, hence it is canonical (i.e., it depends only on $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_1$ and $\mathcal{Y}_2$).

On the other hand, the mapping $W \to D_g \circ W \circ D_f$ is a continuous mapping from $\mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1}$ to $\mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}_2}$, and Proj_2 is continuous. Therefore, the mapping $W \to \mathrm{Proj}_2(D_g \circ W \circ D_f)$ is a continuous mapping from $\mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1}$ to $\mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$. Now since $\mathrm{Proj}_2(D_g \circ W \circ D_f)$ depends only on the $R^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}$-equivalence class $\hat{W}$ of W, Lemma 11.1 implies that the transcendent mapping of $\hat{W} \to \mathrm{Proj}_2(D_g \circ W \circ D_f)$ that is defined on $\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}$ is continuous. Therefore, F is a continuous mapping from $(\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}, \mathcal{T}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1})$ to $(\mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}, \mathcal{T}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2})$. Moreover, we can see from Corollary 11.17 that F is an injection.

For every closed subset B of $\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}$, B is compact since $\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}$ is compact, hence $F(B)$ is compact because F is continuous. This implies that $F(B)$ is closed

in $\mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$ since $\mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$ is Hausdorff (as it is metrizable). Therefore, F is a closed mapping.

Now since F is an injection that is both continuous and closed, F is a homeomorphism from $\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}$ to $F\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}\big) \subset \mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$.

We would like now to show that $F\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}\big)$ depends only on $|\mathcal{X}_1|$, $|\mathcal{Y}_1|$, $\mathcal{X}_2$ and $\mathcal{Y}_2$. Let $\mathcal{X}_1'$ and $\mathcal{Y}_1'$ be two finite sets such that $|\mathcal{X}_1| = |\mathcal{X}_1'|$ and $|\mathcal{Y}_1| = |\mathcal{Y}_1'|$. For every $W \in \mathrm{DMC}_{\mathcal{X}_1',\mathcal{Y}_1'}$, let $\overline{W} \in \mathrm{DMC}^{(s)}_{\mathcal{X}_1',\mathcal{Y}_1'}$ be the $R^{(s)}_{\mathcal{X}_1',\mathcal{Y}_1'}$-equivalence class of W.

Let $f' : \mathcal{X}_1 \to \mathcal{X}_1'$ be a fixed bijection from $\mathcal{X}_1$ to $\mathcal{X}_1'$ and let $f'' = f' \circ f$. Also, let $g' : \mathcal{Y}_1' \to \mathcal{Y}_1$ be a fixed bijection from $\mathcal{Y}_1'$ to $\mathcal{Y}_1$ and let $g'' = g \circ g'$. Define $F' : \mathrm{DMC}^{(s)}_{\mathcal{X}_1',\mathcal{Y}_1'} \to \mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$ as $F'(\overline{W}) = D_{g''} \circ \widetilde{W'} \circ D_{f''} = \mathrm{Proj}_2(D_{g''} \circ W' \circ D_{f''})$, where $W' \in \overline{W}$. As above, F' is well defined, and it is a homeomorphism from $\mathrm{DMC}^{(s)}_{\mathcal{X}_1',\mathcal{Y}_1'}$ to $F'\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1',\mathcal{Y}_1'}\big)$. We want to show that $F'\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1',\mathcal{Y}_1'}\big) = F\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}\big)$. For every $\overline{W} \in \mathrm{DMC}^{(s)}_{\mathcal{X}_1',\mathcal{Y}_1'}$, let $W' \in \overline{W}$. We have

$$\begin{aligned} F'(\overline{W}) = \mathrm{Proj}_2(D_{g''} \circ W' \circ D_{f''}) &= \mathrm{Proj}_2(D_g \circ (D_{g'} \circ W' \circ D_{f'}) \circ D_f) \\ &= F\left(D_{g'} \circ \widehat{W'} \circ D_{f'}\right) \in F\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}\big). \end{aligned}$$

Since this is true for every $\overline{W} \in \mathrm{DMC}^{(s)}_{\mathcal{X}_1',\mathcal{Y}_1'}$, we deduce that $F'\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1',\mathcal{Y}_1'}\big) \subset F\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}\big)$. By exchanging the roles of $(\mathcal{X}_1, \mathcal{Y}_1)$ and $(\mathcal{X}_1', \mathcal{Y}_1')$ and using the fact that $f = f'^{-1} \circ f''$ and $g = g'' \circ g'^{-1}$, we get $F\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}\big) \subset F'\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1',\mathcal{Y}_1'}\big)$. We conclude that $F\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}\big) = F'\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1',\mathcal{Y}_1'}\big)$, which means that $F\big(\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1}\big)$ depends only on $|\mathcal{X}_1|$, $|\mathcal{Y}_1|$, $\mathcal{X}_2$ and $\mathcal{Y}_2$.

Finally, for every $W' \in \hat{W}$ and every $W'' \in F(\hat{W}) = D_g \circ \widetilde{W'} \circ D_f$, W'' is Shannon-equivalent to $D_g \circ W' \circ D_f$ and $D_g \circ W' \circ D_f$ is Shannon-equivalent to W' (by Lemma 11.10), hence W'' is Shannon-equivalent to W'.

Continuity of Channel Parameters and Operations

12

Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets and let W be a fixed channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. It is well known that the input-output mutual information is continuous on the simplex of input probability distributions. Many other parameters that depend on the input probability distribution were shown to be continuous on the simplex in [28].

Polyanskiy studied in [81] the continuity of the Neyman-Pearson function for a binary hypothesis test that arises in the analysis of channel codes. He showed that for arbitrary input and output alphabets, this function is continuous in the input distribution under the total-variation topology. He also showed that under some regularity assumptions, this function is continuous in the weak-$*$ topology.

If $\mathcal{X}$ and $\mathcal{Y}$ are finite sets, the space of channels with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ can naturally be endowed with the topology of the Euclidean metric, or any other equivalent metric. It is well known that the channel capacity is continuous in this topology. If $\mathcal{X}$ and $\mathcal{Y}$ are arbitrary, one can construct a topology on the space of channels using the weak-$*$ topology on the output alphabet. It was shown in [72] that the capacity is lower semi-continuous in this topology.

The continuity results that are mentioned in the previous paragraph do not take into account the equivalence between channels. In [69], output-equivalent binary-input channels were identified with their L-density (i.e., the density of log-likelihood ratios). The space of output-equivalent binary-input channels was endowed with the topology of convergence in distribution of L-densities. Since the symmetric-capacity and the Bhattacharyya parameter can be written as an integral of a continuous function with respect to the L-density [69], it immediately follows that these parameters are continuous in the L-density topology.

In this chapter[1], we study the continuity of several channel parameters and operations under the topologies that were defined in Chapter 11. In Section 12.1, we introduce the preliminaries for this chapter. In Section 12.2, we introduce the channel parameters and operations that we investigate in this chapter. In Section 12.3, we study the continuity of the channel parameters and operations on the

[1]The material of this chapter is based on [63, 64, 65, 66, 82, 83].

spaces of output-equivalent channels. In Section 12.4, we study the continuity of the channel parameters and operations on the spaces of input-equivalent channels. In Section 12.5, we study the continuity of the channel parameters and operations on the spaces of Shannon-equivalent channels.

12.1 Preliminaries

12.1.1 Topological Notations

A topological space $(T,\mathcal{U})$ is said to be *contractible*[2] to $x_0 \in T$ if there exists a continuous mapping $H : T \times [0,1] \to T$ such that $H(x,0) = x$ and $H(x,1) = x_0$ for every $x \in T$, where $[0,1]$ is endowed with the Euclidean topology. $(T,\mathcal{U})$ is *strongly contractible* to $x_0 \in T$ if we also have $H(x_0,t) = x_0$ for every $t \in [0,1]$.

Intuitively, T is contractible if it can be "continuously shrinked" to a single point x_0. If this "continuous shrinking" can be done without moving x_0, T is strongly contractible.

The following lemma is useful to show the continuity of many functions.

Lemma 12.1. *Let $(S,\mathcal{V})$ and $(T,\mathcal{U})$ be two compact topological spaces and let $f : S \times T \to \mathbb{R}$ be a continuous function on $S \times T$. For every $s \in S$ and every $\epsilon > 0$, there exists a neighborhood V_s of s such that for every $s' \in V_s$, we have*

$$\sup_{t \in T} |f(s',t) - f(s,t)| \leq \epsilon.$$

Proof. See Appendix 12.6.1. □

12.1.2 Quotient Topology

Let $(T,\mathcal{U})$ and $(S,\mathcal{V})$ be two topological spaces and let R be an equivalence relation on T. Consider the equivalence relation R' on $T \times S$ defined as $(x_1,y_1)R'(x_2,y_2)$ if and only if $x_1 R x_2$ and $y_1 = y_2$. A natural question to ask is whether the canonical bijection between $\big((T/R) \times S, (\mathcal{U}/R) \otimes \mathcal{V}\big)$ and $\big((T \times S)/R', (\mathcal{U} \otimes \mathcal{V})/R'\big)$ is a homeomorphism. It turns out that this is not the case in general. The following theorem, which is widely used in algebraic topology, provides a sufficient condition:

Theorem 12.1. *[84] If $(S,\mathcal{V})$ is locally compact and Hausdorff, then the canonical bijection between $\big((T/R) \times S, (\mathcal{U}/R) \otimes \mathcal{V}\big)$ and $\big((T \times S)/R', (\mathcal{U} \otimes \mathcal{V})/R'\big)$ is a homeomorphism.*

Corollary 12.1. *Let $(T,\mathcal{U})$ and $(S,\mathcal{V})$ be two topological spaces, and let R_T and R_S be two equivalence relations on T and S respectively. Define the equivalence relation R on $T \times S$ as $(x_1,y_1)R(x_2,y_2)$ if and only if $x_1 R_T x_2$ and $y_1 R_S y_2$. If $(S,\mathcal{V})$ and $(T/R_T, \mathcal{U}/R_T)$ are locally compact and Hausdorff, then the canonical bijection between $\big((T/R_T) \times (S/R_S), (\mathcal{U}/R_T) \otimes (\mathcal{V}/R_S)\big)$ and $\big((T \times S)/R, (\mathcal{U} \otimes \mathcal{V})/R\big)$ is a homeomorphism.*

[2] Contractibility is a very strong notion of connectedness: Every contractible space is path-connected and simply connected. Moreover, all its homotopy, homology and cohomology groups of order ≥ 1 are zero.

Proof. We just need to apply Theorem 12.1 twice. Define the equivalence relation R'_T on $T \times S$ as follows: $(x_1, y_1) R'_T (x_2, y_2)$ if and only if $x_1 R_T x_2$ and $y_1 = y_2$. Since $(S, \mathcal{V})$ is locally compact and Hausdorff, Theorem 12.1 implies that the canonical bijection from $\big((T/R_T) \times S, (\mathcal{U}/R_T) \otimes \mathcal{V}\big)$ to $\big((T \times S)/R'_T, (\mathcal{U} \otimes \mathcal{V})/R'_T\big)$ is a homeomorphism. Let us identify these two spaces through the canonical bijection.

Now define the equivalence relation R'_S on $(T/R_T) \times S$ as follows:

$$(\hat{x}_1, y_1) R'_S (\hat{x}_2, y_2) \text{ if and only if } \hat{x}_1 = \hat{x}_2 \text{ and } y_1 R_S y_2.$$

Since $(T/R_T, \mathcal{U}/R_T)$ is locally compact and Hausdorff, Theorem 12.1 implies that the canonical bijection from $\big((T/R_T) \times (S/R_S), (\mathcal{U}/R_T) \otimes (\mathcal{V}/R_S)\big)$ to $\big(((T/R_T) \times S)/R'_S, ((\mathcal{U}/R_T) \otimes \mathcal{V})/R'_S\big)$ is a homeomorphism.

Since we identified $\big((T/R_T) \times S, (\mathcal{U}/R_T) \otimes \mathcal{V}\big)$ and $\big((T \times S)/R'_T, (\mathcal{U} \otimes \mathcal{V})/R'_T\big)$ through the canonical bijection (which is a homeomorphism), R'_S can be seen as an equivalence relation on $\big((T \times S)/R'_T, (\mathcal{U} \otimes \mathcal{V})/R'_T\big)$. It is easy to see that the canonical bijection from $\big(\big((T \times S)/R'_T\big)/R'_S, \big((\mathcal{U} \otimes \mathcal{V})/R'_T\big)/R'_S\big)$ to $\big((T \times S)/R, (\mathcal{U} \otimes \mathcal{V})/R\big)$ is a homeomorphism. We conclude that the canonical bijection from $\big((T/R_T) \times (S/R_S), (\mathcal{U}/R_T) \otimes (\mathcal{V}/R_S)\big)$ to $\big((T \times S)/R, (\mathcal{U} \otimes \mathcal{V})/R\big)$ is a homeomorphism. □

12.1.3 Measure-Theoretic Notations

The push-forward probability measure

Let P be a probability measure on (M, Σ), and let $f : M \to M'$ be a measurable mapping from (M, Σ) to another measurable space (M', Σ'). The *push-forward probability measure of P by f* is the probability measure $f_\# P$ on (M', Σ') defined as $(f_\# P)(A') = P(f^{-1}(A'))$ for every $A' \in \Sigma'$.

A measurable mapping $g : M' \to \mathbb{R}$ is integrable with respect to $f_\# P$ if and only if $g \circ f$ is integrable with respect to P. Moreover,

$$\int_{M'} g \cdot d(f_\# P) = \int_M (g \circ f) \cdot dP.$$

The mapping $f_\#$ from $\mathcal{P}(M, \Sigma)$ to $\mathcal{P}(M', \Sigma')$ is continuous if these spaces are endowed with the total-variation topology:

$$\begin{aligned} \|f_\# P - f_\# P'\|_{TV} &= \sup_{A' \in \Sigma'} |(f_\# P)(A') - (f_\# P')(A')| \\ &= \sup_{A' \in \Sigma'} |P(f^{-1}(A')) - P'(f^{-1}(A'))| \\ &\leq \sup_{A \in \Sigma} |P(A) - P'(A)| \leq \|P - P'\|_{TV}. \end{aligned}$$

Products of probability measures

We denote the product of two measurable spaces (M_1, Σ_1) and (M_2, Σ_2) as $(M_1 \times M_2, \Sigma_1 \otimes \Sigma_2)$. If $P_1 \in \mathcal{P}(M_1, \Sigma_1)$ and $P_2 \in \mathcal{P}(M_2, \Sigma_2)$, we denote the product of P_1 and P_2 as $P_1 \times P_2$.

If $\mathcal{P}(M_1, \Sigma_1)$, $\mathcal{P}(M_2, \Sigma_2)$ and $\mathcal{P}(M_1 \times M_2, \Sigma_1 \otimes \Sigma_2)$ are endowed with the total-variation topology, then the mapping $(P_1, P_2) \to P_1 \times P_2$ is a continuous mapping (see Appendix 12.6.2).

Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two subsets of $\mathcal{P}(M_1, \Sigma_1)$ and $\mathcal{P}(M_2, \Sigma_2)$, respectively. We define the tensor product of $\mathcal{A}_1$ and $\mathcal{A}_2$ as follows:

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \{P_1 \times P_2 : \ P_1 \in \mathcal{A}_1, \ P_2 \in \mathcal{A}_2\} \subset \mathcal{P}(M_1 \times M_2, \Sigma_1 \otimes \Sigma_2).$$

12.1.4 Random Mappings

Let M and M' be two arbitrary sets and let Σ' be a σ-algebra on M'. A *random mapping* from M to (M', Σ') is a mapping R from M to $\mathcal{P}(M', \Sigma')$. For every $x \in M$, $R(x)$ can be interpreted as the probability distribution of the random output given that the input is x.

Let Σ be a σ-algebra on M. We say that R is a *measurable random mapping* from (M, Σ) to (M', Σ') if the mapping $R_B : M \to \mathbb{R}$ defined as $R_B(x) = (R(x))(B)$ is measurable for every $B \in \Sigma'$.

Note that this definition of measurability is consistent with the measurability of ordinary mappings: Let f be a mapping from M to M' and let $D_f : M \to \mathcal{P}(M', \Sigma')$ be the random mapping defined as $D_f(x) = \delta_{f(x)}$ for every $x \in M$, where $\delta_{f(x)} \in \mathcal{P}(M', \Sigma')$ is a Dirac measure centered at $f(x)$. We have:

$$\begin{aligned}
D_f \text{ is measurable} &\Leftrightarrow (D_f)_B \text{ is measurable}, \ \forall B \in \Sigma' \\
&\Leftrightarrow ((D_f)_B)^{-1}(B') \in \Sigma, \ \forall B' \in \mathcal{B}(\mathbb{R}), \ \forall B \in \Sigma' \\
&\overset{(a)}{\Leftrightarrow} ((D_f)_B)^{-1}(\{1\}) \in \Sigma, \ \forall B \in \Sigma' \\
&\overset{(b)}{\Leftrightarrow} f^{-1}(B) \in \Sigma, \ \forall B \in \Sigma' \\
&\Leftrightarrow f \text{ is measurable},
\end{aligned}$$

where (a) and (b) follow from the fact that $((D_f)_B)(x)$ is either 1 or 0 depending on whether $f(x) \in B$ or not.

Let P be a probability measure on (M, Σ) and let R be a measurable random mapping from (M, Σ) to (M', Σ'). The *push-forward probability measure of P by R* is the probability measure $R_\# P$ on (M', Σ') defined as:

$$(R_\# P)(B) = \int_M R_B \cdot dP, \ \forall B \in \Sigma'.$$

Note that this definition is consistent with the push-forward of ordinary mappings: If f and D_f are as above, then for every $B \in \Sigma'$, we have

$$((D_f)_\# P)(B) = \int_M (D_f)_B \cdot dP = \int_M (\mathbb{1}_B \circ f) \cdot dP = \int_{M'} \mathbb{1}_B \cdot d(f_\# P) = (f_\# P)(B).$$

Proposition 12.1. *Let R be a measurable random mapping from (M, Σ) to (M', Σ'). If $g : M' \to \mathbb{R}^+ \cup \{+\infty\}$ is a Σ'-measurable mapping, then the mapping $x \to \int_{M'} g(y) \cdot d(R(x))(y)$ is a measurable mapping from (M, Σ) to $\mathbb{R}^+ \cup \{+\infty\}$. Moreover, for every $P \in \mathcal{P}(M, \Sigma)$, we have*

$$\int_{M'} g \cdot d(R_\# P) = \int_M \left(\int_{M'} g(y) \cdot d(R(x))(y) \right) dP(x).$$

Proof. See Appendix 12.6.3. □

Corollary 12.2. *If $g : M' \to \mathbb{R}$ is bounded and Σ'-measurable, then the mapping*

$$x \to \int_{M'} g(y) \cdot d(R(x))(y)$$

is bounded and Σ-measurable. Moreover, for every $P \in \mathcal{P}(M, \Sigma)$, we have

$$\int_{M'} g \cdot d(R_{\#}P) = \int_{M} \left(\int_{M'} g(y) \cdot d(R(x))(y) \right) dP(x).$$

Proof. Write $g = g^+ - g^-$ (where $g^+ = \max\{g, 0\}$ and $g^- = \max\{-g, 0\}$), and use the fact that every bounded measurable function is integrable over any probability distribution. □

Lemma 12.2. *For every measurable random mapping R from (M, Σ) to (M', Σ'), the push-forward mapping $R_{\#}$ is continuous from $\mathcal{P}(M, \Sigma)$ to $\mathcal{P}(M', \Sigma')$ under the total-variation topology.*

Proof. See Appendix 12.6.4. □

Lemma 12.3. *Let $\mathcal{U}$ be a Polish[3] topology on M, and let $\mathcal{U}'$ be an arbitrary topology on M'. Let R be a measurable random mapping from $(M, \mathcal{B}(M))$ to $(M', \mathcal{B}(M'))$. Moreover, assume that R is a continuous mapping from $(M, \mathcal{U})$ to $\mathcal{P}(M', \mathcal{B}(M'))$ when the latter space is endowed with the weak-$*$ topology. Under these assumptions, the push-forward mapping $R_{\#}$ is continuous from $\mathcal{P}(M, \mathcal{B}(M))$ to $\mathcal{P}(M', \mathcal{B}(M'))$ under the weak-$*$ topology.*

Proof. See Appendix 12.6.4. □

12.1.5 Meta-Probability Measures

Let $\mathcal{X}$ be a finite set. The following lemma is useful to show the continuity of functions that are defined on the set $\mathcal{MP}(\mathcal{X})$ of meta-probability measures on $\mathcal{X}$.

Lemma 12.4. *Let $(S, \mathcal{V})$ be a compact topological space and let $f : S \times \Delta_{\mathcal{X}} \to \mathbb{R}$ be a continuous function on $S \times \Delta_{\mathcal{X}}$. The mapping $F : S \times \mathcal{MP}(\mathcal{X}) \to \mathbb{R}$ defined as*

$$F(s, \mathrm{MP}) = \int_{\Delta_{\mathcal{X}}} f(s, p) \cdot d\mathrm{MP}(p)$$

is continuous, where $\mathcal{MP}(\mathcal{X})$ is endowed with the weak-$*$ topology.

Proof. See Appendix 12.6.5. □

Let f be a mapping from a finite set $\mathcal{X}$ to another finite set $\mathcal{X}'$. f induces a push-forward mapping $f_{\#}$ taking probability distributions in $\Delta_{\mathcal{X}}$ to probability distributions in $\Delta_{\mathcal{X}'}$. $f_{\#}$ is continuous because $\Delta_{\mathcal{X}}$ and $\Delta_{\mathcal{X}'}$ are endowed with the total-variation distance. $f_{\#}$ in turn induces another push-forward mapping taking meta-probability measures in $\mathcal{MP}(\mathcal{X})$ to meta-probability measures in $\mathcal{MP}(\mathcal{X}')$.

[3] This assumption can be dropped. We assumed that $\mathcal{U}$ is Polish just to avoid working with Moore-Smith nets.

We denote this mapping as $f_{\#\#}$ and we call it *the meta-push-forward mapping* induced by f. Since $f_\#$ is a continuous mapping from $\Delta_{\mathcal{X}}$ to $\Delta_{\mathcal{X}'}$, $f_{\#\#}$ is a continuous mapping from $\mathcal{MP}(\mathcal{X})$ to $\mathcal{MP}(\mathcal{X}')$ under both the weak-$*$ and the total-variation topologies.

Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be two finite sets. Let $\mathrm{Mul} : \Delta_{\mathcal{X}_1} \times \Delta_{\mathcal{X}_2} \to \Delta_{\mathcal{X}_1 \times \mathcal{X}_2}$ be defined as $\mathrm{Mul}(p_1, p_2) = p_1 \times p_2$. For every $\mathrm{MP}_1 \in \mathcal{MP}(\mathcal{X}_1)$ and $\mathrm{MP}_2 \in \mathcal{MP}(\mathcal{X}_2)$, we define the *tensor product* of MP_1 and MP_2 as $\mathrm{MP}_1 \otimes \mathrm{MP}_2 = \mathrm{Mul}_\#(\mathrm{MP}_1 \times \mathrm{MP}_2) \in \mathcal{MP}(\mathcal{X}_1 \times \mathcal{X}_2)$.

Note that since $\Delta_{\mathcal{X}_1}$, $\Delta_{\mathcal{X}_2}$ and $\Delta_{\mathcal{X}_1 \times \mathcal{X}_2}$ are endowed with the total-variation topology, $\mathrm{Mul}(p_1, p_2) = p_1 \times p_2$ is a continuous mapping from $\Delta_{\mathcal{X}_1} \times \Delta_{\mathcal{X}_2}$ to $\Delta_{\mathcal{X}_1 \times \mathcal{X}_2}$. Therefore, $\mathrm{Mul}_\#$ is a continuous mapping from $\mathcal{P}(\Delta_{\mathcal{X}_1} \times \Delta_{\mathcal{X}_2})$ to $\mathcal{P}(\Delta_{\mathcal{X}_1 \times \mathcal{X}_2}) = \mathcal{MP}(\mathcal{X}_1 \times \mathcal{X}_2)$ under both the weak-$*$ and the total-variation topologies. On the other hand, Appendices 12.6.2 and 12.6.6 imply that the mapping $(\mathrm{MP}_1, \mathrm{MP}_2) \to \mathrm{MP}_1 \times \mathrm{MP}_2$ from $\mathcal{MP}(\mathcal{X}_1) \times \mathcal{MP}(\mathcal{X}_2)$ to $\mathcal{P}(\Delta_{\mathcal{X}_1} \times \Delta_{\mathcal{X}_2})$ is continuous under both the weak-$*$ and the total-variation topologies. We conclude that the tensor product is continuous under both these topologies.

12.2 Channel Parameters and Operations

12.2.1 Useful Parameters

Let $\Delta_{\mathcal{X}}$ be the space of probability distributions on $\mathcal{X}$. For every $p \in \Delta_{\mathcal{X}}$ and every $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, define $I(p, W)$ as the mutual information $I(X;Y)$, where X is distributed as p and Y is the output of W when X is the input. The *capacity* of W is defined as $C(W) = \sup_{p \in \Delta_{\mathcal{X}}} I(p, W)$.

For every $p \in \Delta_{\mathcal{X}}$, *the error probability of the MAP decoder of* W *under prior* p is defined as:

$$P_e(p, W) = 1 - \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} \{p(x)W(y|x)\}.$$

Clearly, $0 \le P_e(p, W) \le 1$.

For every $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, define the *Bhattacharyya parameter* of W as:

$$Z(W) = \begin{cases} \dfrac{1}{|\mathcal{X}| \cdot (|\mathcal{X}| - 1)} \displaystyle\sum_{\substack{x_1, x_2 \in \mathcal{X}, \\ x_1 \neq x_2}} \sum_{y \in \mathcal{Y}} \sqrt{W(y|x_1)W(y|x_2)} & \text{if } |\mathcal{X}| \ge 2, \\ 0 & \text{if } |\mathcal{X}| = 1. \end{cases}$$

It is easy to see that $0 \le Z(W) \le 1$.

As we saw in Proposition 5.1, we have $\frac{1}{4} Z(W)^2 \le P_e(\pi_{\mathcal{X}}, W) \le (|\mathcal{X}| - 1) Z(W)$, where $\pi_{\mathcal{X}}$ is the uniform distribution on $\mathcal{X}$.

An (n, M)-*encoder* on the alphabet $\mathcal{X}$ is a mapping $\mathcal{E} : \mathcal{M} \to \mathcal{X}^n$ such that $|\mathcal{M}| = M$. The set $\mathcal{M}$ is the *message set* of $\mathcal{E}$, n is the *blocklength* of $\mathcal{E}$, M is the *size* of $\mathcal{E}$, and $\frac{1}{n} \log_2 M$ is the *rate* of $\mathcal{E}$. We denote the size M of $\mathcal{E}$ as $|\mathcal{E}|$. Moreover, for every $x_1^n \in \mathcal{X}^n$, we write $x_1^n \in \mathcal{E}$ if and only if there exists $m \in \mathcal{M}$ such that $x_1^n = \mathcal{E}(m)$.

The *error probability of the ML decoder for the encoder $\mathcal{E}$ when it is used for a channel $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$* is given by:

$$\begin{aligned} P_{e,\mathcal{E}}(W) &= 1 - \frac{1}{M} \sum_{y_1^n \in \mathcal{Y}^n} \max_{m \in \mathcal{M}} \left\{ \prod_{i=1}^{n} W(y_i|\mathcal{E}_i(m)) \right\} \\ &= 1 - \frac{1}{|\mathcal{E}|} \sum_{y_1^n \in \mathcal{Y}^n} \max_{x_1^n \in \mathcal{E}} \left\{ \prod_{i=1}^{n} W(y_i|x_i) \right\}, \end{aligned}$$

where $(\mathcal{E}_1(m), \ldots, \mathcal{E}_n(m)) = \mathcal{E}(m)$.

The *optimal error probability of (n, M)-encoders for a channel W* is given by:

$$P_{e,n,M}(W) = \min_{\substack{\mathcal{E} \text{ is an} \\ (n,M)\text{-encoder}}} P_{e,\mathcal{E}}(W).$$

Let $\mathcal{D} : \mathcal{Y}^n \to \mathcal{M}$ be a decoder on $\mathcal{Y}$. The *probability of error of $\mathcal{D}$ under ML-encoding for W* is given by:

$$P_{e,\mathcal{D}}(W) = 1 - \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \max_{x_1^n \in \mathcal{X}^n} \left\{ \sum_{\substack{y_1^n \in \mathcal{Y}^n: \\ \mathcal{D}(y_1^n)=m}} \prod_{i=1}^{n} W(y_i|x_i) \right\}.$$

The following proposition shows that all the above parameters are continuous:

Proposition 12.2. *We have:*

- $I : \Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \to \mathbb{R}^+$ *is continuous, concave in p, and convex in W.*
- $C : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \to \mathbb{R}^+$ *is continuous and convex.*
- $P_e : \Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \to [0,1]$ *is continuous, concave in p and concave in W.*
- $Z : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \to [0,1]$ *is continuous.*
- *For every encoder $\mathcal{E}$ on $\mathcal{X}$, $P_{e,\mathcal{E}} : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \to [0,1]$ is continuous.*
- *For every decoder $\mathcal{D}$ on $\mathcal{Y}$, $P_{e,\mathcal{D}} : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \to [0,1]$ is continuous.*
- *For every $n, M > 0$, the mapping $P_{e,n,M} : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \to [0,1]$ is continuous.*

Proof. These facts are well known, especially the continuity of I, its concavity in p, and its convexity in W [3]. Since C is the supremum of a family of mappings that are convex in W, it is also convex in W. For a proof of the continuity of C, see Appendix 12.6.7. The continuity of Z, P_e, $P_{e,\mathcal{E}}$ and $P_{e,\mathcal{D}}$ follows immediately from their definitions. Moreover, since $P_{e,n,M}$ is the minimum of a finite number of continuous mappings, it is continuous. The concavity of P_e in p and in W can also be easily seen from the definition. □

12.2.2 Channel Operations

If $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ and $V \in \mathrm{DMC}_{\mathcal{Y},\mathcal{Z}}$, we define the composition $V \circ W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Z}}$ of W and V as follows:

$$(V \circ W)(z|x) = \sum_{y \in \mathcal{Y}} V(z|y)W(y|x), \quad \forall x \in \mathcal{X},\ \forall z \in \mathcal{Z}.$$

For every function $f : \mathcal{X} \to \mathcal{Y}$, define the *deterministic channel* $D_f \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ as follows:

$$D_f(y|x) = \begin{cases} 1 & \text{if } y = f(x), \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that if $f : \mathcal{X} \to \mathcal{Y}$ and $g : \mathcal{Y} \to \mathcal{Z}$, then $D_g \circ D_f = D_{g \circ f}$.

For every two channels $W_1 \in \mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1}$ and $W_2 \in \mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}_2}$, define the *channel sum* $W_1 \oplus W_2 \in \mathrm{DMC}_{\mathcal{X}_1 \coprod \mathcal{X}_2, \mathcal{Y}_1 \coprod \mathcal{Y}_2}$ of W_1 and W_2 as:

$$(W_1 \oplus W_2)(y, i|x, j) = \begin{cases} W_i(y|x) & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

$W_1 \oplus W_2$ arises when the transmitter has two channels W_1 and W_2 at his disposal and he can use exactly one of them at each channel use. It is an easy exercise to check that $2^{C(W_1 \oplus W_2)} = 2^{C(W_1)} + 2^{C(W_2)}$.

We define the *channel product* $W_1 \otimes W_2 \in \mathrm{DMC}_{\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{Y}_1 \times \mathcal{Y}_2}$ of W_1 and W_2 as:

$$(W_1 \otimes W_2)(y_1, y_2|x_1, x_2) = W_1(y_1|x_1)W_2(y_2|x_2).$$

$W_1 \otimes W_2$ arises when the transmitter has two channels W_1 and W_2 at his disposal and he uses both of them at each channel use. It is an easy exercise to check that $C(W_1 \otimes W_2) = C(W_1) + C(W_2)$, or equivalently $2^{C(W_1 \otimes W_2)} = 2^{C(W_1)} \cdot 2^{C(W_2)}$. Channel sums and products were first introduced by Shannon in [67].

For every $W_1 \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_1}$, $W_2 \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2}$ and every $0 \le \alpha \le 1$, we define the α-interpolation $[\alpha W_1, (1-\alpha)W_2] \in \mathrm{DMC}_{\mathcal{X}, \mathcal{Y}_1 \coprod \mathcal{Y}_2}$ between W_1 and W_2 as:

$$[\alpha W_1, (1-\alpha)W_2](y, i\big|x) = \begin{cases} \alpha W_1(y|x) & \text{if } i = 1, \\ (1-\alpha)W_2(y|x) & \text{if } i = 2. \end{cases}$$

Channel interpolation arises when a channel behaves as W_1 with probability α and as W_2 with probability $1-\alpha$. The transmitter has no control on which behavior the channel chooses, but on the other hand, the receiver knows which one was chosen. Channel interpolations were used in [85] to construct interpolations between polar codes and Reed-Muller codes.

Now fix a binary operation $*$ on $\mathcal{X}$. For every $W \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, define $W^- \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}^2}$ and $W^+ \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}^2 \times \mathcal{X}}$ as:

$$W^-(y_1, y_2|u_1) = \frac{1}{|\mathcal{X}|} \sum_{u_2 \in \mathcal{X}} W(y_1|u_1 * u_2)W(y_2|u_2),$$

and

$$W^+(y_1, y_2, u_1|u_2) = \frac{1}{|\mathcal{X}|} W(y_1|u_1 * u_2)W(y_2|u_2).$$

These operations generalize Arıkan's polarization transformations [2].

Proposition 12.3. *We have:*

- *The mapping* $(W, V) \to V \circ W$ *from* $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \times \mathrm{DMC}_{\mathcal{Y},\mathcal{Z}}$ *to* $\mathrm{DMC}_{\mathcal{X},\mathcal{Z}}$ *is continuous.*
- *The mapping* $(W_1, W_2) \to W_1 \oplus W_2$ *from* $\mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}_2}$ *to* $\mathrm{DMC}_{\mathcal{X}_1 \coprod \mathcal{X}_2, \mathcal{Y}_1 \coprod \mathcal{Y}_2}$ *is continuous.*
- *The mapping* $(W_1, W_2) \to W_1 \otimes W_2$ *from* $\mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}_2}$ *to* $\mathrm{DMC}_{\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{Y}_1 \times \mathcal{Y}_2}$ *is continuous.*
- *The mapping* $(W_1, W_2, \alpha) \to [\alpha W_1, (1-\alpha)W_2]$ *from* $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}_1} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2} \times [0,1]$ *to* $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}_1 \coprod \mathcal{Y}_2}$ *is continuous.*
- *For any binary operation* $*$ *on* $\mathcal{X}$*, the mapping* $W \to W^-$ *from* $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ *to* $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}^2}$ *is continuous.*
- *For any binary operation* $*$ *on* $\mathcal{X}$*, the mapping* $W \to W^+$ *from* $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ *to* $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}^2 \times \mathcal{X}}$ *is continuous.*

Proof. The continuity immediately follows from the definitions. □

12.3 Continuity on the Spaces of Output-Equivalent Channels

12.3.1 Continuity on $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$

It is well known that with the exception of $P_{e,\mathcal{D}}$, all the parameters defined in Section 12.2.1 depend only on the $R^{(o)}_{\mathcal{X},\mathcal{Y}}$-equivalence class of W. Therefore, we can define those parameters for any $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ through the transcendent mapping (defined in Lemma 11.1). The following proposition shows that those parameters are continuous on $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$:

Proposition 12.4. *We have:*

- $I : \Delta_{\mathcal{X}} \times \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \to \mathbb{R}^+$ *is continuous and concave in* p*.*
- $C : \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \to \mathbb{R}^+$ *is continuous.*
- $P_e : \Delta_{\mathcal{X}} \times \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \to [0,1]$ *is continuous and concave in* p*.*
- $Z : \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \to [0,1]$ *is continuous.*
- *For every encoder* $\mathcal{E}$ *on* $\mathcal{X}$*,* $P_{e,\mathcal{E}} : \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \to [0,1]$ *is continuous.*
- *For every* $n, M > 0$*, the mapping* $P_{e,n,M} : \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}} \to [0,1]$ *is continuous.*

Proof. Since the corresponding parameters are continuous on $\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ (Proposition 12.2), Lemma 11.1 implies that they are continuous on $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$. The only cases that need a special treatment are those of I and Z. We will only prove the continuity of I since the proof of continuity of Z is similar.

Define the relation R on $\Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ as

$$(p_1, W_1)R(p_2, W_2) \;\Leftrightarrow\; p_1 = p_2 \text{ and } W_1 R^{(o)}_{\mathcal{X},\mathcal{Y}} W_2.$$

It is easy to see that $I(p, W)$ depends only on the R-equivalence class of (p, W). Since I is continuous on $\Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, Lemma 11.1 implies that the transcendent mapping of I is continuous on $(\Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}})/R$. On the other hand, since $\Delta_{\mathcal{X}}$ is locally compact, Theorem 12.1 implies that $(\Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}})/R$ can be identified with $\Delta_{\mathcal{X}} \times (\mathrm{DMC}_{\mathcal{X},\mathcal{Y}}/R^{(o)}_{\mathcal{X},\mathcal{Y}}) = \Delta_{\mathcal{X}} \times \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ and the two spaces have the same topology. Therefore, I is continuous on $\Delta_{\mathcal{X}} \times \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$. □

With the exception of channel composition, all the channel operations that were defined in Section 12.2.2 can also be "quotiented". We just need to realize that the output-equivalence class of the resulting channel depends only on the output-equivalence classes of the channels that were used in the operation. Let us illustrate this in the case of channel sums:

Let $W_1, W_1' \in \mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1}$ and $W_2, W_2' \in \mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}_2}$ and assume that W_1 is degraded from W_1' and W_2 is degraded from W_2'. There exists $V_1 \in \mathrm{DMC}_{\mathcal{Y}_1,\mathcal{Y}_1}$ and $V_2 \in \mathrm{DMC}_{\mathcal{Y}_2,\mathcal{Y}_2}$ such that $W_1 = V_1 \circ W_1'$ and $W_2 = V_2 \circ W_2'$. It is easy to see that $W_1 \oplus W_2 = (V_1 \oplus V_2) \circ (W_1' \oplus W_2')$, which shows that $W_1 \oplus W_2$ is degraded from $W_1' \oplus W_2'$. This was proved by Shannon in [10].

Therefore, if W_1 is output-equivalent to W_1' and W_2 is output-equivalent to W_2', then $W_1 \oplus W_2$ is output-equivalent to $W_1' \oplus W_2'$. This allows us to define the channel sum for every $\hat{W}_1 \in \mathrm{DMC}^{(o)}_{\mathcal{X}_1,\mathcal{Y}_1}$ and every $\overline{W}_2 \in \mathrm{DMC}^{(o)}_{\mathcal{X}_2,\mathcal{Y}_2}$ as $\hat{W}_1 \oplus \overline{W}_2 = \widetilde{W_1' \oplus W_2'} \in \mathrm{DMC}^{(o)}_{\mathcal{X}_1 \coprod \mathcal{X}_2, \mathcal{Y}_1 \coprod \mathcal{Y}_2}$ for any $W_1' \in \hat{W}_1$ and any $W_2' \in \overline{W}_2$, where $\widetilde{W_1' \oplus W_2'}$ is the $R^{(o)}_{\mathcal{X}_1 \coprod \mathcal{X}_2, \mathcal{Y}_1 \coprod \mathcal{Y}_2}$-equivalence class of $W_1' \oplus W_2'$.

With the exception of channel composition, we can "quotient" all the channel operations of Section 12.2.2 in a similar fashion. Moreover, we can show that they are continuous:

Proposition 12.5. *We have:*

- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \oplus \overline{W}_2$ *from* $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,\mathcal{Y}_2}$ *to* $\mathrm{DMC}^{(o)}_{\mathcal{X}_1 \coprod \mathcal{X}_2, \mathcal{Y}_1 \coprod \mathcal{Y}_2}$ *is continuous.*
- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \otimes \overline{W}_2$ *from* $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,\mathcal{Y}_2}$ *to* $\mathrm{DMC}^{(o)}_{\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{Y}_1 \times \mathcal{Y}_2}$ *is continuous.*
- *The mapping* $(\hat{W}_1, \overline{W}_2, \alpha) \to [\alpha \hat{W}_1, (1-\alpha)\overline{W}_2]$ *from* $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1} \times \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2} \times [0,1]$ *to* $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_1 \coprod \mathcal{Y}_2}$ *is continuous.*
- *For any binary operation* $*$ *on* $\mathcal{X}$*, the mapping* $\hat{W} \to \hat{W}^-$ *from* $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ *to* $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}^2}$ *is continuous.*

- *For any binary operation $*$ on $\mathcal{X}$, the mapping $\hat{W} \to \hat{W}^+$ from $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}}$ to $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}^2\times\mathcal{X}}$ is continuous.*

Proof. We only prove the continuity of the channel sum because the proof of continuity of the other operations is similar.

Let $\mathrm{Proj} : \mathrm{DMC}_{\mathcal{X}_1 \coprod \mathcal{X}_2, \mathcal{Y}_1 \coprod \mathcal{Y}_2} \to \mathrm{DMC}^{(o)}_{\mathcal{X}_1 \coprod \mathcal{X}_2, \mathcal{Y}_1 \coprod \mathcal{Y}_2}$ be the projection onto the $R^{(o)}_{\mathcal{X}_1 \coprod \mathcal{X}_2, \mathcal{Y}_1 \coprod \mathcal{Y}_2}$-equivalence classes. Define the mapping $f : \mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}_2} \to \mathrm{DMC}^{(o)}_{\mathcal{X}_1 \coprod \mathcal{X}_2, \mathcal{Y}_1 \coprod \mathcal{Y}_2}$ as $f(W_1, W_2) = \mathrm{Proj}(W_1 \oplus W_2)$. Clearly, f is continuous.

Now define the equivalence relation R on $\mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}_2}$ as:

$$(W_1, W_2) R (W_1', W_2') \;\Leftrightarrow\; W_1 R^{(o)}_{\mathcal{X}_1,\mathcal{Y}_1} W_1' \text{ and } W_2 R^{(o)}_{\mathcal{X}_2,\mathcal{Y}_2} W_2'.$$

The discussion before the proposition shows that $f(W_1, W_2) = \mathrm{Proj}(W_1 \oplus W_2)$ depends only on the R-equivalence class of (W_1, W_2). Lemma 11.1 now shows that the transcendent map of f defined on $(\mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}_2})/R$ is continuous.

Since $(\mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}_2})/R$ can be identified with $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,\mathcal{Y}_2}$, we can define f on $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,\mathcal{Y}_2}$ through this identification. Moreover, since $\mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1}$ and $\mathrm{DMC}^{(o)}_{\mathcal{X}_2,\mathcal{Y}_2}$ are locally compact and Hausdorff, Corollary 12.1 implies that the canonical bijection between $(\mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}_2})/R$ and the space $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,\mathcal{Y}_2}$ is a homeomorphism.

Now since the mapping f on $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,\mathcal{Y}_2}$ is just the channel sum, we conclude that the mapping $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \oplus \overline{W}_2$ from $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,\mathcal{Y}_2}$ to $\mathrm{DMC}^{(o)}_{\mathcal{X}_1 \coprod \mathcal{X}_2, \mathcal{Y}_1 \coprod \mathcal{Y}_2}$ is continuous. □

12.3.2 Continuity in the Strong Topology

The following lemma provides a way to check whether a mapping defined on the space $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ is continuous:

Lemma 12.5. *Let $(S, \mathcal{V})$ be an arbitrary topological space. A mapping $f : \mathrm{DMC}^{(o)}_{\mathcal{X},*} \to S$ is continuous on the space $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ if and only if it is continuous on $(\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}, \mathcal{T}^{(o)}_{\mathcal{X},[n]})$ for every $n \geq 1$.*

Proof.

$$\begin{aligned} f \text{ is continuous on } &(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*}) \\ &\Leftrightarrow\; f^{-1}(V) \in \mathcal{T}^{(o)}_{s,\mathcal{X},*} \;\; \forall V \in \mathcal{V} \\ &\Leftrightarrow\; f^{-1}(V) \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]} \in \mathcal{T}^{(o)}_{\mathcal{X},[n]} \;\; \forall n \geq 1,\ \forall V \in \mathcal{V} \\ &\Leftrightarrow\; f \text{ is continuous on } (\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}, \mathcal{T}^{(o)}_{\mathcal{X},[n]}) \;\; \forall n \geq 1. \end{aligned}$$

□

Since the channel parameters I, C, P_e, Z, $P_{e,\mathcal{E}}$ and $P_{e,n,M}$ are defined on $\mathrm{DMC}^{(o)}_{\mathcal{X},[l]}$ for every $l \geq 1$ (see Section 12.3.1), they are also defined on $\mathrm{DMC}^{(o)}_{\mathcal{X},*} =$

$\bigcup_{l\geq 1} \mathrm{DMC}^{(o)}_{\mathcal{X},[l]}$. The following proposition shows that those parameters are continuous in the strong topology:

Proposition 12.6. *Let $\mathcal{U}_{\mathcal{X}}$ be the standard topology on $\Delta_{\mathcal{X}}$. We have:*

- $I : \Delta_{\mathcal{X}} \times \mathrm{DMC}^{(o)}_{\mathcal{X},*} \to \mathbb{R}^+$ *is continuous on* $(\Delta_{\mathcal{X}} \times \mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{U}_{\mathcal{X}} \otimes \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ *and concave in p.*
- $C : \mathrm{DMC}^{(o)}_{\mathcal{X},*} \to \mathbb{R}^+$ *is continuous on* $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$.
- $P_e : \Delta_{\mathcal{X}} \times \mathrm{DMC}^{(o)}_{\mathcal{X},*} \to [0,1]$ *is continuous on* $(\Delta_{\mathcal{X}} \times \mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{U}_{\mathcal{X}} \otimes \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ *and concave in p.*
- $Z : \mathrm{DMC}^{(o)}_{\mathcal{X},*} \to [0,1]$ *is continuous on* $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$.
- *For every encoder $\mathcal{E}$ on $\mathcal{X}$, $P_{e,\mathcal{E}} : \mathrm{DMC}^{(o)}_{\mathcal{X},*} \to [0,1]$ is continuous on* $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$.
- *For every $n, M > 0$, the mapping $P_{e,n,M} : \mathrm{DMC}^{(o)}_{\mathcal{X},*} \to [0,1]$ is continuous on* $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$.

Proof. The continuity of $C, Z, P_{e,C}$ and $P_{e,n,M}$ immediately follows from Proposition 12.4 and Lemma 12.5. Since the proofs of continuity of I and Z are similar, we only prove the continuity for I.

Due to the distributivity of the product with respect to disjoint unions, we have

$$\Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},*} = \coprod_{n\geq 1} \left(\Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},[n]}\right),$$

and

$$\mathcal{U}_{\mathcal{X}} \otimes \mathcal{T}_{s,\mathcal{X},*} = \bigoplus_{n\geq 1} \left(\mathcal{U}_{\mathcal{X}} \otimes \mathcal{T}_{\mathcal{X},[n]}\right).$$

Therefore, $(\Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},*}, \mathcal{U}_{\mathcal{X}} \otimes \mathcal{T}_{s,\mathcal{X},*})$ is the disjoint union of the spaces $(\Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},[n]})_{n\geq 1}$. Moreover, I is continuous on $\Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},[n]}$ for every $n \geq 1$. We conclude that I is continuous on $(\Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},*}, \mathcal{U}_{\mathcal{X}} \otimes \mathcal{T}_{s,\mathcal{X},*})$.

Define the relation R on $\Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},*}$ as follows: $(p_1, W_1) R (p_2, W_2)$ if and only if $p_1 = p_2$ and $W_1 R^{(o)}_{\mathcal{X},*} W_2$. Since $I(p, W)$ depends only on the R-equivalence class of (p, W), Lemma 11.1 shows that the transcendent map of I is a continuous mapping from $\left((\Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},*})/R, (\mathcal{U}_{\mathcal{X}} \otimes \mathcal{T}_{s,\mathcal{X},*})/R\right)$ to $\mathbb{R}^+$. On the other hand, since $\Delta_{\mathcal{X}}$ is locally compact and Hausdorff, Theorem 12.1 implies that $\left((\Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},*})/R, (\mathcal{U}_{\mathcal{X}} \otimes \mathcal{T}_{s,\mathcal{X},*})/R\right)$ can be identified with $\left(\Delta_{\mathcal{X}} \times (\mathrm{DMC}_{\mathcal{X},*}/R^{(o)}_{\mathcal{X},*}), \mathcal{U}_{\mathcal{X}} \otimes (\mathcal{T}_{s,\mathcal{X},*}/R^{(o)}_{\mathcal{X},*})\right) = (\Delta_{\mathcal{X}} \times \mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{U}_{\mathcal{X}} \otimes \mathcal{T}^{(o)}_{s,\mathcal{X},*})$. Therefore, I is continuous on $(\Delta_{\mathcal{X}} \times \mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{U}_{\mathcal{X}} \otimes \mathcal{T}^{(o)}_{s,\mathcal{X},*})$. □

It is also possible to extend the definition of all the channel operations that were defined in Section 12.3.1 to $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$. Moreover, it is possible to show that many channel operations are continuous in the strong topology:

Proposition 12.7. *Assume that all output-equivalent channel spaces are endowed with the strong topology. We have:*

- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \oplus \overline{W}_2$ *from* $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,*} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,\mathcal{Y}_2}$ *to* $\mathrm{DMC}^{(o)}_{\mathcal{X}_1 \coprod \mathcal{X}_2,*}$ *is continuous.*
- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \otimes \overline{W}_2$ *from* $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,*} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,\mathcal{Y}_2}$ *to the space* $\mathrm{DMC}^{(o)}_{\mathcal{X}_1 \times \mathcal{X}_2,*}$ *is continuous.*
- *The mapping* $(\hat{W}_1, \overline{W}_2, \alpha) \to [\alpha\hat{W}_1, (1-\alpha)\overline{W}_2]$ *from* $\mathrm{DMC}^{(o)}_{\mathcal{X},*} \times \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2} \times [0,1]$ *to* $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ *is continuous.*
- *For any binary operation* $*$ *on* $\mathcal{X}$*, the mapping* $\hat{W} \to \hat{W}^-$ *from* $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ *to* $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ *is continuous.*
- *For any binary operation* $*$ *on* $\mathcal{X}$*, the mapping* $\hat{W} \to \hat{W}^+$ *from* $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ *to* $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ *is continuous.*

Proof. We only prove the continuity of the channel interpolation because the proof of the continuity of other operations is similar.

Let $\mathcal{U}$ be the standard topology on $[0,1]$. Due to the distributivity of the product with respect to disjoint unions, we have:

$$\mathrm{DMC}_{\mathcal{X},*} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2} \times [0,1] = \coprod_{n \geq 1} (\mathrm{DMC}_{\mathcal{X},[n]} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2} \times [0,1]),$$

and

$$\mathcal{T}_{s,\mathcal{X},*} \otimes \mathcal{T}_{\mathcal{X},\mathcal{Y}_2} \otimes \mathcal{U} = \bigoplus_{n \geq 1} \left(\mathcal{T}_{\mathcal{X},[n]} \otimes \mathcal{T}_{\mathcal{X},\mathcal{Y}_2} \otimes \mathcal{U}\right).$$

Therefore, the space $\mathrm{DMC}_{\mathcal{X},*} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2} \times [0,1]$ is the topological disjoint union of the spaces $(\mathrm{DMC}_{\mathcal{X},[n]} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2} \times [0,1])_{n \geq 1}$.

For every $n \geq 1$, let Proj_n be the projection onto the $R^{(o)}_{\mathcal{X},[n] \coprod \mathcal{Y}_2}$-equivalence classes and let i_n be the canonical injection from $\mathrm{DMC}^{(o)}_{\mathcal{X},[n] \coprod \mathcal{Y}_2}$ to $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$.

Define the mapping $f : \mathrm{DMC}_{\mathcal{X},*} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2} \times [0,1] \to \mathrm{DMC}^{(o)}_{\mathcal{X},*}$ as

$$f(W_1, W_2, \alpha) = i_n(\mathrm{Proj}_n([\alpha W_1, (1-\alpha)W_2])) = [\alpha\hat{W}_1, (1-\alpha)\overline{W}_2],$$

where n is the unique integer satisfying $W_1 \in \mathrm{DMC}_{\mathcal{X},[n]}$. $\hat{W}_1$ and $\overline{W}_2$ are the $R^{(o)}_{\mathcal{X},[n]}$ and $R^{(o)}_{\mathcal{X},\mathcal{Y}_2}$-equivalence classes of W_1 and W_2 respectively.

Due to Proposition 12.3 and due to the continuity of Proj_n and i_n, the mapping f is continuous on $\mathrm{DMC}_{\mathcal{X},[n]} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2} \times [0,1]$ for every $n \geq 1$. Therefore, f is continuous on $(\mathrm{DMC}_{\mathcal{X},*} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2} \times [0,1], \mathcal{T}_{s,\mathcal{X},*} \otimes \mathcal{T}_{\mathcal{X},\mathcal{Y}_2} \otimes \mathcal{U})$.

Let R' be the equivalence relation defined on $\mathrm{DMC}_{\mathcal{X},*} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2}$ as follows: $(W_1, W_2)R'(W_1', W_2')$ if and only if $W_1 R^{(o)}_{\mathcal{X},*} W_1'$ and $W_2 R^{(o)}_{\mathcal{X},\mathcal{Y}_2} W_2'$. Also, define the equivalence relation R on $\mathrm{DMC}_{\mathcal{X},*} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2} \times [0,1]$ as follows:

$$(W_1, W_2, \alpha)R(W_1', W_2', \alpha') \text{ if and only if } (W_1, W_2)R'(W_1', W_2') \text{ and } \alpha = \alpha'.$$

Since $f(W_1, W_2, \alpha)$ depends only on the R-equivalence class of (W_1, W_2, α), Lemma 11.1 implies that the transcendent mapping of f is continuous on the space $(\mathrm{DMC}_{\mathcal{X},*} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2} \times [0,1])/R$.

Since $[0,1]$ is Hausdorff and locally compact, Theorem 12.1 implies that the canonical bijection from the space $(\mathrm{DMC}_{\mathcal{X},*} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2} \times [0,1])/R$ to the space $\big((\mathrm{DMC}_{\mathcal{X},*} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2})/R'\big) \times [0,1])$ is a homeomorphism. On the other hand, since $(\mathrm{DMC}_{\mathcal{X},*}, \mathcal{T}_{s,\mathcal{X},*})$ and $\mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2} = \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2}/R^{(o)}_{\mathcal{X},\mathcal{Y}_2}$ are Hausdorff and locally compact, Corollary 12.1 implies that the canonical bijection from $\mathrm{DMC}^{(o)}_{\mathcal{X},*} \times \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2}$ to $(\mathrm{DMC}_{\mathcal{X},*} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}_2})/R'$ is a homeomorphism. We conclude that the channel interpolation is continuous on $(\mathrm{DMC}^{(o)}_{\mathcal{X},*} \times \mathrm{DMC}^{(o)}_{\mathcal{X},\mathcal{Y}_2} \times [0,1], \mathcal{T}^{(o)}_{s,\mathcal{X},*} \otimes \mathcal{T}^{(o)}_{\mathcal{X},\mathcal{Y}} \otimes \mathcal{U})$. □

Corollary 12.3. $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ *is strongly contractible to every point in* $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$.

Proof. Fix $\hat{W}_0 \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$. Define the mapping $H : \mathrm{DMC}^{(o)}_{\mathcal{X},*} \times [0,1] \to \mathrm{DMC}^{(o)}_{\mathcal{X},*}$ as $H(\hat{W}, \alpha) = [\alpha \hat{W}_0, (1-\alpha)\hat{W}]$. H is continuous by Proposition 12.7. We also have $H(\hat{W}, 0) = \hat{W}$ and $H(\hat{W}, 1) = \hat{W}_0$ for every $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$. Moreover, $H(\hat{W}_0, \alpha) = \hat{W}_0$ for every $0 \le \alpha \le 1$. Therefore, $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{s,\mathcal{X},*})$ is strongly contractible to every point in $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$. □

The reader might be wondering why channel operations such as the channel sum were not shown to be continuous on the whole space $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,*} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,*}$ instead of the smaller space $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,*} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,\mathcal{Y}_2}$. The reason is because we cannot apply Corollary 12.1 to $\mathrm{DMC}_{\mathcal{X}_1,*} \times \mathrm{DMC}_{\mathcal{X}_2,*}$ and $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,*} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,*}$ since neither $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,*}$ nor $\mathrm{DMC}^{(o)}_{\mathcal{X}_2,*}$ is locally compact (under the strong topology).

One potential method to show the continuity of the channel sum on the whole space $(\mathrm{DMC}^{(o)}_{\mathcal{X}_1,*} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,*}, \mathcal{T}^{(o)}_{s,\mathcal{X}_1,*} \otimes \mathcal{T}^{(o)}_{s,\mathcal{X}_2,*})$ is as follows: Let R be the equivalence relation on $\mathrm{DMC}_{\mathcal{X}_1,*} \times \mathrm{DMC}_{\mathcal{X}_2,*}$ defined as $(W_1, W_2) R (W_1', W_2')$ if and only if $W_1 R^{(o)}_{\mathcal{X}_1,*} W_1'$ and $W_2 R^{(o)}_{\mathcal{X}_2,*} W_2'$. We can identify $(\mathrm{DMC}_{\mathcal{X}_1,*} \times \mathrm{DMC}_{\mathcal{X}_2,*})/R$ with $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,*} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,*}$ through the canonical bijection. Using Lemma 11.1, it is easy to see that the mapping $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \oplus \overline{W}_2$ is continuous from the space $\big(\mathrm{DMC}^{(o)}_{\mathcal{X}_1,*} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,*}, (\mathcal{T}_{s,\mathcal{X}_1,*} \otimes \mathcal{T}_{s,\mathcal{X}_2,*})/R\big)$ to $(\mathrm{DMC}^{(o)}_{\mathcal{X}_1 \coprod \mathcal{X}_2,*}, \mathcal{T}^{(o)}_{s,\mathcal{X}_1 \coprod \mathcal{X}_2,*})$.

It was shown in [79] that the topology $(\mathcal{T}_{s,\mathcal{X}_1,*} \otimes \mathcal{T}_{s,\mathcal{X}_2,*})/R$ is homeomorphic to $\kappa(\mathcal{T}^{(o)}_{s,\mathcal{X}_1,*} \otimes \mathcal{T}^{(o)}_{s,\mathcal{X}_2,*})$ through the canonical bijection, where $\kappa(\mathcal{T}^{(o)}_{s,\mathcal{X}_1,*} \otimes \mathcal{T}^{(o)}_{s,\mathcal{X}_2,*})$ is the coarsest topology that is both compactly generated and finer than $\mathcal{T}^{(o)}_{s,\mathcal{X}_1,*} \otimes \mathcal{T}^{(o)}_{s,\mathcal{X}_2,*}$. Therefore, the mapping $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \oplus \overline{W}_2$ is continuous on the space $\big(\mathrm{DMC}^{(o)}_{\mathcal{X}_1,*} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,*}, \kappa(\mathcal{T}^{(o)}_{s,\mathcal{X}_1,*} \otimes \mathcal{T}^{(o)}_{s,\mathcal{X}_2,*})\big)$. This means that if $\mathcal{T}^{(o)}_{s,\mathcal{X}_1,*} \otimes \mathcal{T}^{(o)}_{s,\mathcal{X}_2,*}$ is compactly generated, we will have $\mathcal{T}^{(o)}_{s,\mathcal{X}_1,*} \otimes \mathcal{T}^{(o)}_{s,\mathcal{X}_2,*} = \kappa(\mathcal{T}^{(o)}_{s,\mathcal{X}_1,*} \otimes \mathcal{T}^{(o)}_{s,\mathcal{X}_2,*})$ and so the channel sum will be continuous on $(\mathrm{DMC}^{(o)}_{\mathcal{X}_1,*} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,*}, \mathcal{T}^{(o)}_{s,\mathcal{X}_1,*} \otimes \mathcal{T}^{(o)}_{s,\mathcal{X}_2,*})$. Note that although $\mathcal{T}^{(o)}_{s,\mathcal{X}_1,*}$ and $\mathcal{T}^{(o)}_{s,\mathcal{X}_2,*}$ are compactly generated, their product $\mathcal{T}^{(o)}_{s,\mathcal{X}_1,*} \otimes \mathcal{T}^{(o)}_{s,\mathcal{X}_2,*}$ might not be compactly generated.

12.3.3 Continuity in the Noisiness/Weak-$*$ and the Total-Variation Topologies

We need to express the channel parameters and operations in terms of the Blackwell measures.

Channel Parameters

The following proposition shows that many channel parameters can be expressed as an integral of a continuous function with respect to the Blackwell measure:

Proposition 12.8. *If $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$, then:*

- *For every $p \in \Delta_{\mathcal{X}}$, we have*

$$I(p, \hat{W}) = H(p) - |\mathcal{X}| \cdot \int_{\Delta_{\mathcal{X}}} \left(\sum_{x \in \mathcal{X}} p(x)p'(x) \log_2 \frac{p(x)p'(x)}{\sum_{x'} p(x')p'(x')} \right) \cdot d\mathrm{MP}_{\hat{W}}(p'),$$

  where $H(p)$ is the entropy of p. Note that we adopt the standard convention that $0 \log_2 \frac{0}{0} = 0$.

- *For every $p \in \Delta_{\mathcal{X}}$, we have*

$$P_e(p, \hat{W}) = 1 - |\mathcal{X}| \int_{\Delta_{\mathcal{X}}} \max_{x \in \mathcal{X}} \{p(x) \times p'(x)\} \cdot d\mathrm{MP}_{\hat{W}}(p').$$

- *If $|\mathcal{X}| \geq 2$, we have*

$$Z(\hat{W}) = \frac{1}{|\mathcal{X}| - 1} \sum_{\substack{x, x' \in \mathcal{X}, \\ x \neq x'}} \int_{\Delta_{\mathcal{X}}} \sqrt{p(x)p(x')} \cdot d\mathrm{MP}_{\hat{W}}(p).$$

- *For every (n, M)-encoder $\mathcal{E}$ on $\mathcal{X}$, we have*

$$P_{e,\mathcal{E}}(\hat{W}) = 1 - \frac{|\mathcal{X}|^n}{|\mathcal{E}|} \int_{\Delta^n_{\mathcal{X}}} \max_{x_1^n \in \mathcal{E}} \left\{ \prod_{i=1}^{n} p_i(x_i) \right\} d\mathrm{MP}^n_{\hat{W}}(p_1^n),$$

  where $\mathrm{MP}^n_{\hat{W}}$ is the product measure on $\Delta^n_{\mathcal{X}}$ obtained by multiplying $\mathrm{MP}_{\hat{W}}$ with itself n times.

Proof. By choosing any representative channel $W \in \hat{W}$ and replacing $W(y|x)$ by $|\mathcal{X}| P^o_W(y) W_y^{-1}(x)$ in the definitions of the channel parameters, all the above formulas

immediately follow. Let us show how this works for P_e:

$$\begin{aligned}
P_e(p,\hat{W}) = P_e(p,W) &\overset{(a)}{=} 1 - \sum_{y\in\mathrm{Im}(W)} \max_{x\in\mathcal{X}}\{p(x)W(y|x)\} \\
&= 1 - \sum_{y\in\mathrm{Im}(W)} \max_{x\in\mathcal{X}}\left\{p(x)\cdot|\mathcal{X}|\cdot P_W^o(y)W_y^{-1}(x)\right\} \\
&= 1 - |\mathcal{X}| \sum_{y\in\mathrm{Im}(W)} \max_{x\in\mathcal{X}}\{p(x)W_y^{-1}(x)\}\cdot P_W^o(y) \\
&= 1 - |\mathcal{X}| \int_{\Delta_\mathcal{X}} \max_{x\in\mathcal{X}}\{p(x)p'(x)\}\cdot d\mathrm{MP}_W(p') \\
&= 1 - |\mathcal{X}| \int_{\Delta_\mathcal{X}} \max_{x\in\mathcal{X}}\{p(x)p'(x)\}\cdot d\mathrm{MP}_{\hat{W}}(p'),
\end{aligned}$$

where (a) is true because $W(y|x) = 0$ for $y \notin \mathrm{Im}(W)$. □

Proposition 12.9. *Let $\mathcal{U}_\mathcal{X}$ be the standard topology on $\Delta_\mathcal{X}$. We have:*

- $I : \Delta_\mathcal{X} \times \mathrm{DMC}_{\mathcal{X},*}^{(o)} \to \mathbb{R}^+$ *is continuous on* $(\Delta_\mathcal{X} \times \mathrm{DMC}_{\mathcal{X},*}^{(o)}, \mathcal{U}_\mathcal{X} \otimes \mathcal{T}_{\mathcal{X},*}^{(o)})$ *and concave in p.*
- $C : \mathrm{DMC}_{\mathcal{X},*}^{(o)} \to \mathbb{R}^+$ *is continuous on* $(\mathrm{DMC}_{\mathcal{X},*}^{(o)}, \mathcal{T}_{\mathcal{X},*}^{(o)})$.
- $P_e : \Delta_\mathcal{X} \times \mathrm{DMC}_{\mathcal{X},*}^{(o)} \to [0,1]$ *is continuous on* $(\Delta_\mathcal{X} \times \mathrm{DMC}_{\mathcal{X},*}^{(o)}, \mathcal{U}_\mathcal{X} \otimes \mathcal{T}_{\mathcal{X},*}^{(o)})$ *and concave in p.*
- $Z : \mathrm{DMC}_{\mathcal{X},*}^{(o)} \to [0,1]$ *is continuous on* $(\mathrm{DMC}_{\mathcal{X},*}^{(o)}, \mathcal{T}_{\mathcal{X},*}^{(o)})$.
- *For every encoder $\mathcal{E}$ on $\mathcal{X}$, $P_{e,\mathcal{E}} : \mathrm{DMC}_{\mathcal{X},*}^{(o)} \to [0,1]$ is continuous on* $(\mathrm{DMC}_{\mathcal{X},*}^{(o)}, \mathcal{T}_{\mathcal{X},*}^{(o)})$.
- *For every $n, M > 0$, the mapping $P_{e,n,M} : \mathrm{DMC}_{\mathcal{X},*}^{(o)} \to [0,1]$ is continuous on* $(\mathrm{DMC}_{\mathcal{X},*}^{(o)}, \mathcal{T}_{\mathcal{X},*}^{(o)})$.

Proof. We associate the space $\mathcal{MP}(\mathcal{X})$ with the weak-$*$ topology. Define the mapping

$$\overline{I} : \Delta_\mathcal{X} \times \mathcal{MP}(\mathcal{X}) \to \mathbb{R}^+$$

as follows:

$$\overline{I}(p,\mathrm{MP}) = H(p) - |\mathcal{X}| \cdot \int_{\Delta_\mathcal{X}} \left(\sum_{x\in\mathcal{X}} p(x)p'(x) \log_2 \frac{p(x)p'(x)}{\sum_{x'} p(x')p'(x')} \right) \cdot d\mathrm{MP}(p'),$$

Lemma 12.4 implies that $\overline{I}$ is continuous. On the other hand, Proposition 12.8 shows that $I(p,\hat{W}) = \overline{I}(p,\mathrm{MP}_{\hat{W}})$. Therefore, I is continuous on $(\Delta_\mathcal{X} \times \mathrm{DMC}_{\mathcal{X},*}^{(o)}, \mathcal{U}_\mathcal{X} \otimes \mathcal{T}_{\mathcal{X},*}^{(o)})$. We can prove the continuity of P_e and Z similarly.

Now define the mapping $\overline{C} : \mathcal{MP}(\mathcal{X}) \to \mathbb{R}$ as

$$\overline{C}(\mathrm{MP}) = \sup_{p\in\Delta_\mathcal{X}} \overline{I}(p,\mathrm{MP}).$$

Fix $\mathrm{MP} \in \mathcal{MP}(\mathcal{X})$ and let $\epsilon > 0$. Since $\mathcal{MP}(\mathcal{X})$ is compact (under the weak-$*$ topology), Lemma 12.1 implies the existence of a weakly-$*$ open neighborhood U_{MP} of MP such that $|\overline{I}(p, \mathrm{MP}) - \overline{I}(p, \mathrm{MP}')| < \epsilon$ for every $\mathrm{MP}' \in U_{\mathrm{MP}}$ and every $p \in \Delta_{\mathcal{X}}$. Therefore, for every $\mathrm{MP}' \in U_{\mathrm{MP}}$ and every $p \in \Delta_{\mathcal{X}}$, we have

$$\overline{I}(p, \mathrm{MP}) < \overline{I}(p, \mathrm{MP}') + \epsilon \leq \overline{C}(\mathrm{MP}') + \epsilon,$$

hence,

$$\overline{C}(\mathrm{MP}) = \sup_{p \in \Delta_{\mathcal{X}}} \overline{I}(p, \mathrm{MP}) \leq \overline{C}(\mathrm{MP}') + \epsilon.$$

Similarly, we can show that $\overline{C}(\mathrm{MP}') \leq \overline{C}(\mathrm{MP}) + \epsilon$. This shows that $|\overline{C}(\mathrm{MP}') - \overline{C}(\mathrm{MP})| \leq \epsilon$ for every $\mathrm{MP}' \in U_{\mathrm{MP}}$. Therefore, $\overline{C}$ is continuous. But $C(\hat{W}) = \overline{C}(\mathrm{MP}_{\hat{W}})$, so C is continuous on $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{\mathcal{X},*})$.

Now let $\mathcal{E}$ be an (n, M)-encoder on $\mathcal{X}$. For every $0 \leq i \leq n$, define the mapping $f_i : \Delta^i_{\mathcal{X}} \times \mathcal{MP}(\mathcal{X}) \to \mathbb{R}$ backward-recursively as follows:

- $f_n(p_1^n, \mathrm{MP}) = \max\limits_{x_1^n \in \mathcal{E}} \left\{ \prod\limits_{i=1}^{n} p_i(x_i) \right\}$.
- For every $0 \leq i < n$, define

$$f_i(p_1^i, \mathrm{MP}) = \int_{\Delta_{\mathcal{X}}} f_{i+1}(p_1^{i+1}, \mathrm{MP}) \cdot d\mathrm{MP}(p_{i+1}).$$

Clearly, f_n is continuous. Now let $0 \leq i < n$ and assume that f_{i+1} is continuous. If we let $S = \Delta^i_{\mathcal{X}} \times \mathcal{MP}(\mathcal{X})$, Lemma 12.4 implies that the mapping

$$F_i : \Delta^i_{\mathcal{X}} \times \mathcal{MP}(\mathcal{X}) \times \mathcal{MP}(\mathcal{X}) \to \mathbb{R}$$

defined as

$$F_i(p_1^i, \mathrm{MP}, \mathrm{MP}') = \int_{\Delta_{\mathcal{X}}} f_{i+1}(p_1^{i+1}, \mathrm{MP}) \cdot d\mathrm{MP}'(p_{i+1})$$

is continuous. But $f_i(p_1^i, \mathrm{MP}) = F_i(p_1^i, \mathrm{MP}, \mathrm{MP})$, so f_i is also continuous. Therefore, f_0 is continuous. By noticing that $P_{e,\mathcal{E}}(\hat{W}) = 1 - \dfrac{|\mathcal{X}|^n}{|\mathcal{E}|} f_0(\mathrm{MP}_{\hat{W}})$, we conclude that $P_{e,\mathcal{E}}$ is continuous on $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{\mathcal{X},*})$. Moreover, since $P_{e,n,M}$ is the minimum of a finite family of continuous mappings, it is continuous. □

It is worth mentioning that Proposition 12.6 can be shown from Proposition 12.9 because the noisiness topology is coarser than the strong topology.

Corollary 12.4. *All the mappings in Proposition 12.9 are also continuous if we replace the noisiness topology $\mathcal{T}^{(o)}_{\mathcal{X},*}$ with the total-variation topology $\mathcal{T}^{(o)}_{TV,\mathcal{X},*}$.*

Proof. This is true because $\mathcal{T}^{(o)}_{TV,\mathcal{X},*}$ is finer than $\mathcal{T}^{(o)}_{\mathcal{X},*}$. □

Channel Operations

In the following, we show that we can express the channel operations in terms of Blackwell measures. We have all the tools to achieve this for the channel sum, channel product and channel interpolation. In order to express the channel polarization transformations in terms of the Blackwell measures, we need to introduce new definitions.

Let $\mathcal{X}$ be a finite set and let $*$ be a binary operation on a finite set $\mathcal{X}$. We say that $*$ is *uniformity-preserving* if the mapping $(a, b) \to (a*b, b)$ is a bijection from $\mathcal{X}^2$ to itself. For every $a, b \in \mathcal{X}$, we denote the unique element $c \in \mathcal{X}$ satisfying $c * b = a$ as $c = a/^* b$. Note that $/^*$ is a binary operation and it is uniformity-preserving. $/^*$ is called the *right-inverse* of $*$. We saw in Chapter 3 that a binary operation is polarizing if and only if it is uniformity-preserving and its right-inverse is strongly ergodic.

Binary operations that are not uniformity-preserving are not interesting for polarization theory because they do not preserve the symmetric capacity (see Remark 3.1). Therefore, we will only focus on polarization transformations that are based on uniformity-preserving operations.

Let $*$ be a fixed uniformity-preserving operation on $\mathcal{X}$. Define the mapping $C^{-,*} : \Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}} \to \Delta_{\mathcal{X}}$ as

$$(C^{-,*}(p_1, p_2))(u_1) = \sum_{u_2 \in \mathcal{X}} p_1(u_1 * u_2) p_2(u_2).$$

The probability distribution $C^{-,*}(p_1, p_2)$ can be interpreted as follows: Let X_1 and X_2 be two independent random variables in $\mathcal{X}$ that are distributed as p_1 and p_2 respectively, and let (U_1, U_2) be the random pair in $\mathcal{X}^2$ defined as $(U_1, U_2) = (X_1/^* X_2, X_2)$, or equivalently $(X_1, X_2) = (U_1 * U_2, U_2)$. $C^{-,*}(p_1, p_2)$ is the probability distribution of U_1.

Clearly, $C^{-,*}$ is continuous. Therefore, the push-forward mapping $C^{-,*}_{\#}$ is continuous from $\mathcal{P}(\Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}})$ to $\mathcal{P}(\Delta_{\mathcal{X}}) = \mathcal{MP}(\mathcal{X})$ under both the weak-$*$ and the total-variation topologies (see Section 12.1.5). For every $\mathrm{MP}_1, \mathrm{MP}_2 \in \mathcal{MP}(\mathcal{X})$, we define the $(-, *)$-convolution of MP_1 and MP_2 as:

$$(\mathrm{MP}_1, \mathrm{MP}_2)^{-,*} = C^{-,*}_{\#}(\mathrm{MP}_1 \times \mathrm{MP}_2) \in \mathcal{MP}(\mathcal{X}).$$

Since the product of meta-probability measures is continuous under both the weak-$*$ and the total-variation topologies (Appendices 12.6.2 and 12.6.6), the $(-, *)$-convolution is also continuous under these topologies.

For every $p_1, p_2 \in \Delta_{\mathcal{X}}$ and every $u_1 \in \mathrm{supp}(C^{-,*}(p_1, p_2))$, define $C^{+,u_1,*}(p_1, p_2) \in \Delta_{\mathcal{X}}$ as

$$(C^{+,u_1,*}(p_1, p_2))(u_2) = \frac{p_1(u_1 * u_2) p_2(u_2)}{(C^{-,*}(p_1, p_2))(u_1)}.$$

The probability distribution $C^{+,u_1,*}(p_1, p_2)$ can be interpreted as follows: If X_1, X_2, U_1 and U_2 are as above, $C^{+,u_1,*}(p_1, p_2)$ is the conditional probability distribution of U_2 given $U_1 = u_1$.

Define the mapping $C^{+,*} : \Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}} \to \mathcal{P}(\Delta_{\mathcal{X}}) = \mathcal{MP}(\mathcal{X})$ as follows:

$$C^{+,*}(p_1, p_2) = \sum_{u_1 \in \mathrm{supp}(C^{-,*}(p_1,p_2))} (C^{-,*}(p_1, p_2))(u_1) \cdot \delta_{C^{+,u_1,*}(p_1,p_2)},$$

where $\delta_{C^{+,u_1,*}(p_1,p_2)}$ is a Dirac measure centered at $C^{+,u_1,*}(p_1,p_2)$.

If X_1, X_2, U_1 and U_2 are as above, $C^{+,*}(p_1,p_2)$ is the meta-probability measure that describes the possible conditional probability distributions of U_2 that are seen by someone having knowledge of U_1. Clearly, $C^{+,*}$ is a random mapping from $\Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}}$ to $\Delta_{\mathcal{X}}$. In Appendix 12.6.8, we show that $C^{+,*}$ is a measurable random mapping. We also show in Appendix 12.6.8 that $C^{+,*}$ is a continuous mapping from $\Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}}$ to $\mathcal{MP}(\mathcal{X})$ when the latter space is endowed with the weak-$*$ topology. Lemmas 12.2 and 12.3 now imply that the push-forward mapping $C^{+,*}_{\#}$ is continuous under both the weak-$*$ and the total-variation topologies.

For every $\mathrm{MP}_1, \mathrm{MP}_2 \in \mathcal{MP}(\mathcal{X})$, we define the $(+,*)$-convolution of MP_1 and MP_2 as:

$$(\mathrm{MP}_1, \mathrm{MP}_2)^{+,*} = C^{+,*}_{\#}(\mathrm{MP}_1 \times \mathrm{MP}_2) \in \mathcal{MP}(\mathcal{X}).$$

Since the product of meta-probability measures is continuous under both the weak-$*$ and the total-variation topologies (Appendices 12.6.2 and 12.6.6), the $(+,*)$-convolution is also continuous under these topologies.

Proposition 12.10. *We have:*

- *For every $\hat{W}_1 \in \mathrm{DMC}^{(o)}_{\mathcal{X}_1,*}$ and $\overline{W}_2 \in \mathrm{DMC}^{(o)}_{\mathcal{X}_2,*}$, we have:*

$$\mathrm{MP}_{\hat{W}_1 \oplus \overline{W}_2} = \frac{|\mathcal{X}_1|}{|\mathcal{X}_1| + |\mathcal{X}_2|} \mathrm{MP}'_{\hat{W}_1} + \frac{|\mathcal{X}_2|}{|\mathcal{X}_1| + |\mathcal{X}_2|} \mathrm{MP}'_{\overline{W}_2},$$

  where $\mathrm{MP}'_{\hat{W}_1}$ (respectively $\mathrm{MP}'_{\hat{W}_2}$) is the meta-push-forward of $\mathrm{MP}_{\hat{W}_1}$ (respectively $\mathrm{MP}_{\hat{W}_2}$) by the canonical injection from $\mathcal{X}_1$ (respectively $\mathcal{X}_2$) to $\mathcal{X}_1 \coprod \mathcal{X}_2$.

- *For every $\hat{W}_1 \in \mathrm{DMC}^{(o)}_{\mathcal{X}_1,*}$ and $\overline{W}_2 \in \mathrm{DMC}^{(o)}_{\mathcal{X}_2,*}$, we have:*

$$\mathrm{MP}_{\hat{W}_1 \otimes \overline{W}_2} = \mathrm{MP}_{\hat{W}_1} \otimes \mathrm{MP}_{\overline{W}_2}.$$

- *For every $\alpha \in [0,1]$ and every $\hat{W}_1, \hat{W}_2 \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$, we have*

$$\mathrm{MP}_{[\alpha \hat{W}_1, (1-\alpha)\hat{W}_2]} = \alpha \mathrm{MP}_{\hat{W}_1} + (1-\alpha)\mathrm{MP}_{\hat{W}_2}.$$

- *For every uniformity-preserving binary operation $*$ on $\mathcal{X}$, and every $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$, we have*

$$\mathrm{MP}_{\hat{W}^-} = (\mathrm{MP}_{\hat{W}}, \mathrm{MP}_{\hat{W}})^{-,*}.$$

- *For every uniformity-preserving binary operation $*$ on $\mathcal{X}$, and every $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$, we have*

$$\mathrm{MP}_{\hat{W}^+} = (\mathrm{MP}_{\hat{W}}, \mathrm{MP}_{\hat{W}})^{+,*}.$$

Proof. See Appendix 12.6.9. □

Note that the polarization transformation formulas in Proposition 12.10 generalize the formulas given by Raginsky in [86] for binary-input channels.

Proposition 12.11. *Assume that all output-equivalent channel spaces are endowed with the noisiness/weak-$*$ or the total-variation topology. We have:*

- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \oplus \overline{W}_2$ *from* $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,*} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,*}$ *to* $\mathrm{DMC}^{(o)}_{\mathcal{X}_1 \coprod \mathcal{X}_2,*}$ *is continuous.*

- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \otimes \overline{W}_2$ *from* $\mathrm{DMC}^{(o)}_{\mathcal{X}_1,*} \times \mathrm{DMC}^{(o)}_{\mathcal{X}_2,*}$ *to* $\mathrm{DMC}^{(o)}_{\mathcal{X}_1 \times \mathcal{X}_2,*}$ *is continuous.*

- *The mapping* $(\hat{W}_1, \overline{W}_2, \alpha) \to [\alpha\hat{W}_1, (1-\alpha)\overline{W}_2]$ *from* $\mathrm{DMC}^{(o)}_{\mathcal{X},*} \times \mathrm{DMC}^{(o)}_{\mathcal{X},*} \times [0,1]$ *to* $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ *is continuous.*

- *For every uniformity-preserving binary operation* $*$ *on* $\mathcal{X}$*, the mapping* $\hat{W} \to \hat{W}^-$ *from* $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ *to* $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ *is continuous.*

- *For every uniformity-preserving binary operation* $*$ *on* $\mathcal{X}$*, the mapping* $\hat{W} \to \hat{W}^+$ *from* $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ *to* $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ *is continuous.*

Proof. The proposition directly follows from Proposition 12.10 and the fact that all the meta-probability measure operations that are involved in the formulas are continuous under both the weak-$*$ and the total-variation topologies. □

Corollary 12.5. *Both* $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{\mathcal{X},*})$ *and* $(\mathrm{DMC}^{(o)}_{\mathcal{X},*}, \mathcal{T}^{(o)}_{TV,\mathcal{X},*})$ *are strongly contractible to every point in* $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$.

Proof. We can use the same proof of Corollary 12.3. □

12.4 Continuity on the Spaces of Input-Equivalent Channels

12.4.1 Channel Parameters

Since input-degradedness is a particular case of the Shannon ordering [10], we can easily see that if W and W' are input-equivalent, then $C(W) = C(W')$ and $P_{e,n,M}(W) = P_{e,n,M}(W')$ for every $n \geq 1$ and every $M \geq 1$. Therefore, for every $\hat{W} \in \mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$, we can define $C(\hat{W}) := C(W')$ for any $W' \in \hat{W}$. We can define $P_{e,n,M}(\hat{W})$ similarly. Moreover, due to Proposition 10.5, we can also define $P_{e,\mathcal{D}}(\hat{W})$ for any decoder $\mathcal{D}$ on the output alphabet $\mathcal{Y}$.

Proposition 12.12. *Let* $\mathcal{X}$ *and* $\mathcal{Y}$ *be two finite sets. We have:*

- $C : \mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}} \to \mathbb{R}^+$ *is continuous on* $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}})$.

- *For every* $n \geq 1$ *and every* $M \geq 1$*, the mapping* $P_{e,n,M} : \mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}} \to [0,1]$ *is continuous on* $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}})$.

- *For every decoder* $\mathcal{D}$ *on* $\mathcal{Y}$*, the mapping* $P_{e,\mathcal{D}} : \mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}} \to [0,1]$ *is continuous on* $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}})$.

Proof. Since $C : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \to \mathbb{R}^+$ is continuous, and since $C(W)$ depends only on the $R^{(i)}_{\mathcal{X},\mathcal{Y}}$-equivalence class of W, Lemma 11.1 implies that $C : \mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}} \to \mathbb{R}^+$ is continuous on $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}})$. We can show the continuity of $P_{e,n,M}$ and $P_{e,\mathcal{D}}$ on $(\mathrm{DMC}^{(i)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(i)}_{\mathcal{X},\mathcal{Y}})$ similarly. □

The following lemma provides a way to check whether a mapping defined on $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$ is continuous:

Lemma 12.6. *Let $(S, \mathcal{V})$ be an arbitrary topological space. A mapping $f : \mathrm{DMC}^{(i)}_{*,\mathcal{Y}} \to S$ is continuous on the space $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$ if and only if it is continuous on $(\mathrm{DMC}^{(i)}_{[n],\mathcal{Y}}, \mathcal{T}^{(i)}_{[n],\mathcal{Y}})$ for every $n \geq 1$.*

Proof. Same proof as Lemma 12.5. □

Proposition 12.13. *Let $\mathcal{Y}$ be a finite set. We have:*

- *$C : \mathrm{DMC}^{(i)}_{*,\mathcal{Y}} \to \mathbb{R}^+$ is continuous on $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$.*
- *For every $n \geq 1$ and every $M \geq 1$, the mapping $P_{e,n,M} : \mathrm{DMC}^{(i)}_{*,\mathcal{Y}} \to [0,1]$ is continuous on $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$.*
- *For every decoder $\mathcal{D}$ on $\mathcal{Y}$, the mapping $P_{e,\mathcal{D}} : \mathrm{DMC}^{(i)}_{*,\mathcal{Y}} \to [0,1]$ is continuous on $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}})$.*

Proof. The proposition follows from Proposition 12.12 and Lemma 12.6. □

12.4.2 Channel Operations

Channel sums and products can be "quotiented" by the input-equivalence relation. We just need to realize that the input-equivalence class of the resulting channel depends only on the input-equivalence classes of the channels that were used in the operation. Let us illustrate this in the case of channel sums:

Let $W_1, W_1' \in \mathrm{DMC}_{\mathcal{X}_1,\mathcal{Y}_1}$ and $W_2, W_2' \in \mathrm{DMC}_{\mathcal{X}_2,\mathcal{Y}_2}$ and assume that W_1 is input-degraded from W_1' and W_2 is input-degraded from W_2'. There exists $V_1' \in \mathrm{DMC}_{\mathcal{X}_1,\mathcal{X}_1}$ and $V_2' \in \mathrm{DMC}_{\mathcal{X}_2,\mathcal{X}_2}$ such that $W_1 = W_1' \circ V_1'$ and $W_2 = W_2' \circ V_2'$. It is easy to see that $W_1 \oplus W_2 = (W_1' \oplus W_2') \circ (V_1' \oplus V_2')$, which shows that $W_1 \oplus W_2$ is input-degraded from $W_1' \oplus W_2'$.

Therefore, if W_1 is input-equivalent to W_1' and W_2 is input-equivalent to W_2', then $W_1 \oplus W_2$ is input-equivalent to $W_1' \oplus W_2'$. This allows us to define the channel sum for every $\hat{W}_1 \in \mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}_1}$ and every $\overline{W}_2 \in \mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}_2}$ as $\hat{W}_1 \oplus \overline{W}_2 = \widetilde{W_1' \oplus W_2'} \in \mathrm{DMC}^{(i)}_{\mathcal{X}_1 \coprod \mathcal{X}_2, \mathcal{Y}_1 \coprod \mathcal{Y}_2}$ for any $W_1' \in \hat{W}_1$ and any $W_2' \in \overline{W}_2$, where $\widetilde{W_1' \oplus W_2'}$ is the $R^{(i)}_{\mathcal{X}_1 \coprod \mathcal{X}_2, \mathcal{Y}_1 \coprod \mathcal{Y}_2}$-equivalence class of $W_1' \oplus W_2'$. We can define the product on the quotient spaces similarly.

Proposition 12.14. *We have:*

- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \oplus \overline{W}_2$ *from* $\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}_2}$ *to* $\mathrm{DMC}^{(i)}_{\mathcal{X}_1 \coprod \mathcal{X}_2, \mathcal{Y}_1 \coprod \mathcal{Y}_2}$ *is continuous.*

- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \otimes \overline{W}_2$ *from* $\mathrm{DMC}^{(i)}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}_2}$ *to* $\mathrm{DMC}^{(i)}_{\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{Y}_1 \times \mathcal{Y}_2}$ *is continuous.*

Proof. Same proof as Proposition 12.5. □

Proposition 12.15. *Assume that all spaces of input-equivalent channels are endowed with the strong topology. We have:*

- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \oplus \overline{W}_2$ *from* $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}_1} \times \mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}_2}$ *to* $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}_1 \coprod \mathcal{Y}_2}$ *is continuous.*

- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \otimes \overline{W}_2$ *from* $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}_1} \times \mathrm{DMC}^{(i)}_{\mathcal{X}_2,\mathcal{Y}_2}$ *to* $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}_1 \times \mathcal{Y}_2}$ *is continuous.*

Proof. Same proof as Proposition 12.7. □

As in the case of output-equivalent channels[4], the continuity of channel sums and products on the whole space $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}_1} \times \mathrm{DMC}^{(i)}_{*,\mathcal{Y}_2}, \mathcal{T}^{(i)}_{s,*,\mathcal{Y}_1} \otimes \mathcal{T}^{(i)}_{s,*,\mathcal{Y}_2})$ can be shown by proving that $\mathcal{T}^{(i)}_{s,*,\mathcal{Y}_1} \otimes \mathcal{T}^{(i)}_{s,*,\mathcal{Y}_2}$ is compactly generated. Note that although $\mathcal{T}^{(i)}_{s,*,\mathcal{Y}_1}$ and $\mathcal{T}^{(i)}_{s,*,\mathcal{Y}_2}$ are compactly generated, their product $\mathcal{T}^{(i)}_{s,*,\mathcal{Y}_1} \otimes \mathcal{T}^{(i)}_{s,*,\mathcal{Y}_2}$ might not be compactly generated.

Proposition 12.16. *Let* $\mathcal{Y}_1$ *and* $\mathcal{Y}_2$ *be two finite sets. Let* $\hat{W}_1 \in \mathrm{DMC}^{(i)}_{*,\mathcal{Y}_1}$ *and* $\overline{W}_2 \in \mathrm{DMC}^{(i)}_{*,\mathcal{Y}_2}$*. We have:*

$$\mathrm{co}(\hat{W}_1 \oplus \overline{W}_2) = \bigcup_{0 \le \lambda \le 1} \Big((1-\lambda)\phi_{1\#}(\mathrm{co}(\hat{W}_1)) + \lambda \phi_{2\#}(\mathrm{co}(\hat{W}_2)) \Big),$$

where $\phi_{1\#}$ *and* $\phi_{2\#}$ *are the push-forwards by the canonical injections from* $\mathcal{Y}_1$ *and* $\mathcal{Y}_2$ *to* $\mathcal{Y}_1 \coprod \mathcal{Y}_2$ *respectively. On the other hand,*

$$\mathrm{co}(\hat{W}_1 \otimes \overline{W}_2) = \mathrm{co}\Big(\mathrm{co}(\hat{W}_1) \otimes \mathrm{co}(\hat{W}_2) \Big).$$

Proof. See Appendix 12.6.10. □

Proposition 12.17. *Assume that all spaces of input-equivalent channels are endowed with the similarity topology. We have:*

- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \oplus \overline{W}_2$ *from* $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}_1} \times \mathrm{DMC}^{(i)}_{*,\mathcal{Y}_2}$ *to* $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}_1 \coprod \mathcal{Y}_2}$ *is continuous.*

- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \otimes \overline{W}_2$ *from* $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}_1} \times \mathrm{DMC}^{(i)}_{*,\mathcal{Y}_2}$ *to* $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}_1 \times \mathcal{Y}_2}$ *is continuous.*

Proof. See Appendix 12.6.11. □

[4] See the discussion after Corollary 12.3.

12.5 Continuity on the Space of Shannon-Equivalent Channels

12.5.1 Channel parameters

For every $W \in \mathrm{DMC}_{*,*}$, $C(W)$ depends only on the Shannon-equivalence class of W [10]. Therefore, for every $\hat{W} \in \mathrm{DMC}^{(s)}_{*,*}$, we can define $C(\hat{W}) := C(W')$ for any $W' \in \hat{W}$. We can define $P_{e,n,M}(\hat{W})$ similarly.

Proposition 12.18. *Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets. We have:*

- $C : \mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}} \to \mathbb{R}^+$ *is continuous on* $(\mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(s)}_{\mathcal{X},\mathcal{Y}})$.
- *For every $n \geq 1$ and every $M \geq 1$, the mapping $P_{e,n,M} : \mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}} \to [0,1]$ is continuous on* $(\mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(s)}_{\mathcal{X},\mathcal{Y}})$.

Proof. Since $C : \mathrm{DMC}_{\mathcal{X},\mathcal{Y}} \to \mathbb{R}^+$ is continuous, and since $C(W)$ depends only on the $R^{(s)}_{\mathcal{X},\mathcal{Y}}$-equivalence class of W, Lemma 11.1 implies that $C : \mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}} \to \mathbb{R}^+$ is continuous on $(\mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(s)}_{\mathcal{X},\mathcal{Y}})$. We can show the continuity of $P_{e,n,M}$ on $(\mathrm{DMC}^{(s)}_{\mathcal{X},\mathcal{Y}}, \mathcal{T}^{(s)}_{\mathcal{X},\mathcal{Y}})$ similarly. □

The following lemma provides a way to check whether a mapping defined on $(\mathrm{DMC}^{(s)}_{*,*}, \mathcal{T}^{(s)}_{s,*,*})$ is continuous:

Lemma 12.7. *Let $(S, \mathcal{V})$ be an arbitrary topological space. A mapping $f : \mathrm{DMC}^{(s)}_{*,*} \to S$ is continuous on $(\mathrm{DMC}^{(s)}_{*,*}, \mathcal{T}^{(s)}_{s,*,*})$ if and only if it is continuous on the space $(\mathrm{DMC}^{(s)}_{[n],[n]}, \mathcal{T}^{(s)}_{[n],[n]})$ for every $n \geq 1$.*

Proof.

$$\begin{aligned} f \text{ is continuous on } (\mathrm{DMC}^{(s)}_{*,*}, \mathcal{T}^{(s)}_{s,*,*}) &\Leftrightarrow f^{-1}(V) \in \mathcal{T}^{(s)}_{s,*,*}, \ \forall V \in \mathcal{V} \\ &\Leftrightarrow f^{-1}(V) \cap \mathrm{DMC}^{(s)}_{[n],[n]} \in \mathcal{T}^{(s)}_{[n],[n]}, \ \forall n \geq 1, \ \forall V \in \mathcal{V} \\ &\Leftrightarrow f \text{ is continuous on } (\mathrm{DMC}^{(s)}_{[n],[n]}, \mathcal{T}^{(s)}_{[n],[n]}), \ \forall n \geq 1. \end{aligned}$$

□

Proposition 12.19. *We have:*

- $C : \mathrm{DMC}^{(s)}_{*,*} \to \mathbb{R}^+$ *is continuous on* $(\mathrm{DMC}^{(s)}_{*,*}, \mathcal{T}^{(s)}_{s,*,*})$.
- *For every $n \geq 1$ and every $M \geq 1$, the mapping $P_{e,n,M} : \mathrm{DMC}^{(s)}_{*,*} \to [0,1]$ is continuous on* $(\mathrm{DMC}^{(s)}_{*,*}, \mathcal{T}^{(s)}_{s,*,*})$.

Proof. The proposition follows from Proposition 12.18 and Lemma 12.7. □

12.5.2 Channel operations

Channel sums and products can be "quotiented" by the Shannon-equivalence relation. We just need to realize that the Shannon-equivalence class of the resulting channel depends only on the Shannon-equivalence classes of the channels that were used in the operation [10].

Proposition 12.20. *We have:*

- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \oplus \overline{W}_2$ *from* $\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$ *to* $\mathrm{DMC}^{(s)}_{\mathcal{X}_1 \coprod \mathcal{X}_2, \mathcal{Y}_1 \coprod \mathcal{Y}_2}$ *is continuous.*
- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \otimes \overline{W}_2$ *from* $\mathrm{DMC}^{(s)}_{\mathcal{X}_1,\mathcal{Y}_1} \times \mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$ *to* $\mathrm{DMC}^{(s)}_{\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{Y}_1 \times \mathcal{Y}_2}$ *is continuous.*

Proof. Same proof as Proposition 12.5. □

Proposition 12.21. *Assume that the space* $\mathrm{DMC}^{(s)}_{*,*}$ *is endowed with the strong topology. We have:*

- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \oplus \overline{W}_2$ *from* $\mathrm{DMC}^{(s)}_{*,*} \times \mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$ *to* $\mathrm{DMC}^{(s)}_{*,*}$ *is continuous.*
- *The mapping* $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \otimes \overline{W}_2$ *from* $\mathrm{DMC}^{(s)}_{*,*} \times \mathrm{DMC}^{(s)}_{\mathcal{X}_2,\mathcal{Y}_2}$ *to* $\mathrm{DMC}^{(s)}_{*,*}$ *is continuous.*

Proof. Same proof as Proposition 12.7. □

As in the case of the space of output-equivalent channels[5], we can show the continuity of channel sums and products on $(\mathrm{DMC}^{(s)}_{*,*} \times \mathrm{DMC}^{(s)}_{*,*}, \mathcal{T}^{(s)}_{s,*,*} \otimes \mathcal{T}^{(s)}_{s,*,*})$ by proving that $\mathcal{T}^{(s)}_{s,*,*} \otimes \mathcal{T}^{(s)}_{s,*,*}$ is compactly generated. Note that although $\mathcal{T}^{(s)}_{s,*,*}$ and $\mathcal{T}^{(s)}_{s,*,*}$ are compactly generated, their product $\mathcal{T}^{(s)}_{s,*,*} \otimes \mathcal{T}^{(s)}_{s,*,*}$ might not be compactly generated.

12.6 Appendix

12.6.1 Proof of Lemma 12.1

Fix $\epsilon > 0$ and let $(s,t) \in S \times T$. Since f is continuous, there exists a neighborhood $O_{s,t}$ of (s,t) in $S \times T$ such that for every $(s',t') \in O_{s,t}$, we have $|f(s',t') - f(s,t)| < \frac{\epsilon}{2}$. Moreover, since products of open sets form a base for the product topology, there exists an open neighborhood $V_{s,t}$ of s in $(S, \mathcal{V})$ and an open neighborhood $U_{s,t}$ of t in T such that $V_{s,t} \times U_{s,t} \subset O_{s,t}$.

Since $(S, \mathcal{V})$ and $(T, \mathcal{U})$ are compact, the product space is also compact. On the other hand, we have $\bigcup_{(s,t) \in S \times T} V_{s,t} \times U_{s,t} = S \times T$ so $\{V_{s,t} \times U_{s,t}\}_{(s,t) \in S \times T}$ is an open cover of $S \times T$. Therefore, there exist $s_1, \ldots, s_n \in S$ and $t_1, \ldots, t_n \in T$ such that $\bigcup_{i=1}^{n} V_{s_i,t_i} \times U_{s_i,t_i} = S \times T$.

[5] See the discussion after Corollary 12.3.

Now fix $s \in S$ and define $V_s = \bigcap_{\substack{1 \le i \le n, \\ s \in V_{s_i,t_i}}} V_{s_i,t_i}$. Since V_s is the intersection of finitely many open sets containing s, V_s is an open neighborhood of s in $(S, \mathcal{V})$. Let $s' \in V_s$ and $t \in T$. Since $\bigcup_{i=1}^{n} V_{s_i,t_i} \times U_{s_i,t_i} = S \times T$, there exists $1 \le i \le n$ such that $(s,t) \in V_{s_i,t_i} \times U_{s_i,t_i} \subset O_{s_i,t_i}$. Since $s \in V_{s_i,t_i}$, we have $V_s \subset V_{s_i,t_i}$ and so $s' \in V_{s_i,t_i}$. Therefore, $(s',t) \in V_{s_i,t_i} \times U_{s_i,t_i} \subset O_{s_i,t_i}$, hence

$$|f(s',t) - f(s,t)| \le |f(s',t) - f(s_i,t_i)| + |f(s_i,t_i) - f(s,t)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

But this is true for every $t \in T$. Therefore,

$$\sup_{t \in T} |f(s',t) - f(s,t)| \le \epsilon.$$

12.6.2 Continuity of the Product of Measures

For every subset A of $M_1 \times M_2$ and every $x_1 \in M_1$, define $A_2^{x_1} = \{x_2 \in M_2 : (x_1,x_2) \in A\}$. Similarly, for every $x_2 \in M_2$, define $A_1^{x_2} = \{x_1 \in M_1 : (x_1,x_2) \in A\}$. Let $P_1, P_1' \in \mathcal{P}(M_1, \Sigma_1)$ and $P_2, P_2' \in \mathcal{P}(M_2, \Sigma_2)$. We have:

$$\begin{aligned}
&\|P_1 \times P_2 - P_1' \times P_2'\|_{TV} \\
&\quad = \sup_{A \in \Sigma_1 \otimes \Sigma_2} |(P_1 \times P_2)(A) - (P_1' \times P_2')(A)| \\
&\quad \le \sup_{A \in \Sigma_1 \otimes \Sigma_2} \Big\{ \big|(P_1 \times P_2)(A) - (P_1' \times P_2)(A)\big| + \big|(P_1' \times P_2)(A) - (P_1' \times P_2')(A)\big| \Big\} \\
&\quad = \sup_{A \in \Sigma_1 \otimes \Sigma_2} \Bigg\{ \left| \int_{M_2} P_1(A_1^{x_2}) \cdot dP_2(x_2) - \int_{M_2} P_1'(A_1^{x_2}) \cdot dP_2(x_2) \right| \\
&\qquad\qquad + \left| \int_{M_1} P_2(A_2^{x_1}) \cdot dP_1'(x_1) - \int_{M_1} P_2'(A_2^{x_1}) \cdot dP_1'(x_1) \right| \Bigg\} \\
&\quad \le \sup_{A \in \Sigma_1 \otimes \Sigma_2} \left\{ \int_{M_2} \big|P_1(A_1^{x_2}) - P_1'(A_1^{x_2})\big| \cdot dP_2(x_2) + \int_{M_1} \big|P_2(A_2^{x_1}) - P_2'(A_2^{x_1})\big| \cdot dP_1'(x_1) \right\} \\
&\quad \le \int_{M_2} \left(\sup_{A_1 \in \Sigma_1} \big|P_1(A_1) - P_1'(A_1)\big| \right) dP_2 + \int_{M_1} \left(\sup_{A_2 \in \Sigma_2} \big|P_2(A_2) - P_2'(A_2)\big| \right) dP_1' \\
&\quad = \|P_1 - P_1'\|_{TV} + \|P_2 - P_2'\|_{TV}.
\end{aligned}$$

This shows that the product of measures is continuous under the total-variation topology.

12.6.3 Proof of Proposition 12.1

Define the mapping $G : M \to \mathbb{R}^+ \cup \{+\infty\}$ as follows:

$$G(x) = \int_{M'} g(y) d(R(x))(y).$$

For every $n \ge 0$, define the mapping $g_n : M' \to \mathbb{R}^+$ as follows:

$$g_n(y) = \frac{1}{2^n} \big\lfloor 2^n \times \min\{n, g(y)\} \big\rfloor.$$

Clearly, for every $y \in M'$ we have:

- $g_n(y) \leq g(y)$ for all $n \geq 0$.
- $g_n(y) \leq g_{n+1}(y)$ for all $n \geq 0$.
- $\lim_{n\to\infty} g_n(y) = g(y)$.

Moreover, for every fixed $n \geq 0$, we have:

- g_n is Σ'-measurable.
- g_n takes values in $\left\{\frac{i}{2^n} : \ 0 \leq i \leq n2^n\right\}$.

For every $0 \leq i \leq n2^n$, let $B_{i,n} = \{y \in M' : \ g_n(y) = \frac{i}{2^n}\}$. Since g_n is Σ'-measurable, we have $B_{i,n} \in \Sigma'$ for every $0 \leq i \leq n2^n$. Now for every $n \geq 0$, define the mapping $G_n : M \to \mathbb{R} \cup \{+\infty\}$ as follows:

$$\begin{aligned} G_n(x) &= \int_{M'} g_n(y) d(R(x))(y) = \int_{M'} \left(\sum_{i=0}^{n2^n} \frac{i}{2^n} \mathbb{1}_{B_{i,n}}(y) \right) d(R(x))(y) \\ &= \sum_{i=0}^{n2^n} \frac{i}{2^n} (R(x))(B_{i,n}) = \sum_{i=0}^{n2^n} \frac{i}{2^n} R_{B_{i,n}}(x). \end{aligned}$$

Since the random mapping R is measurable and since $B_{i,n} \in \Sigma'$, the mapping $R_{B_{i,n}}$ is Σ-measurable for every $0 \leq i \leq n2^n$. Therefore, G_n is Σ-measurable for every $n \geq 0$. Moreover, for every $x \in \Sigma$, we have:

$$\lim_{n\to\infty} G_n(x) = \lim_{n\to\infty} \int_{M'} g_n(y) d(R(x))(y) \overset{(a)}{=} \int_{M'} g(y) d(R(x))(y) = G(x),$$

where (a) follows from the monotone convergence theorem. We conclude that G is Σ-measurable because it is the point-wise limit of Σ-measurable functions. On the other hand, we have

$$\begin{aligned} \int_{M'} g_n \cdot d(R_{\#}P) &= \sum_{i=0}^{n2^n} \frac{i}{2^n} (R_{\#}P)(B_{i,n}) = \sum_{i=0}^{n2^n} \frac{i}{2^n} \int_M R_{B_{i,n}}(x) \cdot dP(x) \\ &= \sum_{i=0}^{n2^n} \frac{i}{2^n} \int_M (R(x))(B_{i,n}) \cdot dP(x) \\ &= \sum_{i=0}^{n2^n} \frac{i}{2^n} \int_M \left(\int_{M'} \mathbb{1}_{B_{i,n}}(y) \cdot d(R(x))(y) \right) dP(x) \\ &= \int_M \left(\int_{M'} \left(\sum_{i=0}^{n2^n} \frac{i}{2^n} \mathbb{1}_{B_{i,n}}(y) \right) d(R(x))(y) \right) dP(x) \\ &= \int_M \left(\int_{M'} g_n(y) d(R(x))(y) \right) dP(x) = \int_M G_n \cdot dP. \end{aligned}$$

Therefore,

$$\int_{M'} g \cdot d(R_{\#}P) \overset{(a)}{=} \lim_{n\to\infty} \int_{M'} g_n \cdot d(R_{\#}P) = \lim_{n\to\infty} \int_M G_n \cdot dP \overset{(b)}{=} \int_M G \cdot dP,$$

where (a) and (b) follow from the monotone convergence theorem.

12.6.4 Continuity of the Push-Forward by a Random Mapping

Let R be a measurable random mapping from (M,Σ) to (M',Σ'). Let $P_1,P_2 \in \mathcal{P}(M,\Sigma)$. Define the signed measure $\mu = P_1 - P_2$ and let $\{\mu^+,\mu^-\}$ be the Jordan measure decomposition of μ. It is easy to see that $\|P_1-P_2\|_{TV} = \mu^+(M) = \mu^-(M)$. For every $B \in \Sigma'$, we have:

$$\begin{aligned}(R_\#(P_1))(B) - (R_\#(P_2))(B) &= \int_M R_B \cdot dP_1 - \int_M R_B \cdot dP_2 = \int_M R_B \cdot d(P_1-P_2)\\ &= \int_M R_B \cdot d(\mu^+ - \mu^-) \leq \int_M R_B \cdot d\mu^+\\ &\leq \|R_B\|_\infty \cdot \mu^+(M) \overset{(a)}{\leq} \mu^+(M) = \|P_1-P_2\|_{TV},\end{aligned}$$

where (a) follows from the fact that $|R_B(x)| = |(R(x))(B)| \leq 1$ for every $x \in M$. We can similarly show that

$$(R_\#(P_2))(B) - (R_\#(P_1))(B) \leq \|R_B\|_\infty \cdot \mu^-(M) \leq \|P_1-P_2\|_{TV}.$$

Therefore,

$$\|R_\#(P_1) - R_\#(P_2)\|_{TV} = \sup_{B\in\Sigma'} |(R_\#(P_1))(B) - (R_\#(P_2))(B)| \leq \|P_1-P_2\|_{TV}.$$

This shows that the push-forward mapping $R_\#$ from $\mathcal{P}(M,\Sigma)$ to $\mathcal{P}(M',\Sigma')$ is continuous under the total-variation topology. This concludes the proof of Lemma 12.2.

Now assume that $\mathcal{U}$ is a Polish topology on M and $\mathcal{U}'$ is an arbitrary topology on M'. Let R be measurable random mapping from $(M,\mathcal{B}(M))$ to $(M',\mathcal{B}(M'))$. Moreover, assume that R is a continuous mapping from $(M,\mathcal{U})$ to $\mathcal{P}(M',\mathcal{B}(M'))$ when the latter space is endowed with the weak-$*$ topology. Let $(P_n)_{n\geq 0}$ be a sequence of probability measures in $\mathcal{P}(M,\mathcal{B}(M))$ that weakly-$*$ converges to $P \in \mathcal{P}(M,\mathcal{B}(M))$.

Let $g : M' \to \mathbb{R}$ be a bounded and continuous mapping. Define the mapping $G : M \to \mathbb{R}$ as follows:

$$G(x) = \int_{M'} g(y) \cdot d(R(x))(y).$$

For every sequence $(x_n)_{n\geq 0}$ converging to x in M, the sequence $(R(x_n))_{n\geq 0}$ weakly-$*$ converges to $R(x)$ in $\mathcal{P}(M',\mathcal{B}(M'))$ because of the continuity of R. This implies that the sequence $(G(x_n))_{n\geq 0}$ converges to $G(x)$. Since $\mathcal{U}$ is a Polish topology (hence metrizable and sequential [78]), this shows that G is a bounded and continuous mapping from $(M,\mathcal{U})$ to $\mathbb{R}$. Therefore, we have:

$$\lim_{n\to\infty} \int_{M'} g \cdot d(R_\# P_n) \overset{(a)}{=} \lim_{n\to\infty} \int_M G \cdot dP_n \overset{(b)}{=} \int_M G \cdot dP \overset{(c)}{=} \int_{M'} g \cdot d(R_\# P),$$

where (a) and (c) follow from Corollary 12.2, and (b) follows from the fact that $(P_n)_{n\geq 0}$ weakly-$*$ converges to P. This shows that $(R_\# P_n)_{n\geq 0}$ weakly-$*$ converges to $R_\# P$. Now since $\mathcal{U}$ is Polish, the weak-$*$ topology on $\mathcal{P}(M,\mathcal{B}(M))$ is metrizable [80], hence it is sequential [78]. This shows that the push-forward mapping $R_\#$ from $\mathcal{P}(M,\mathcal{B}(M))$ to $\mathcal{P}(M',\mathcal{B}(M'))$ is continuous under the weak-$*$ topology.

12.6.5 Proof of Lemma 12.4

For every $s \in S$, define the mapping $f_s : \Delta_{\mathcal{X}} \to \mathbb{R}$ as $f_s(p) = f(s,p)$. Clearly f_s is continuous for every $s \in S$. Therefore, the mapping $F_s : \mathcal{MP}(\mathcal{X}) \to \mathbb{R}$ defined as

$$F_s(\mathrm{MP}) = \int_{\Delta_{\mathcal{X}}} f_s \cdot d\mathrm{MP}$$

is continuous in the weak-$*$ topology of $\mathcal{MP}(\mathcal{X})$.

Fix $\epsilon > 0$ and let $(s, \mathrm{MP}) \in S \times \mathcal{MP}(\mathcal{X})$. Since F_s is continuous, there exists a weakly-$*$ open neighborhood $U_{s,\mathrm{MP}}$ of MP such that $|F_s(\mathrm{MP}') - F_s(\mathrm{MP})| < \frac{\epsilon}{2}$ for every $\mathrm{MP}' \in U_{s,\mathrm{MP}}$. On the other hand, Lemma 12.1 implies the existence of an open neighborhood V_s of s in $(S, \mathcal{V})$ such that for every $s' \in V_s$ we have

$$\sup_{p \in \Delta_{\mathcal{X}}} |f(s',p) - f(s,p)| \leq \frac{\epsilon}{2}.$$

Clearly $V_s \times U_{s,\mathrm{MP}}$ is an open neighborhood of (s, MP) in $S \times \mathcal{MP}(\mathcal{X})$. For every $(s', \mathrm{MP}') \in V_s \times U_{s,\mathrm{MP}}$, we have

$$\begin{aligned}
|F(s',\mathrm{MP}') - F(s,\mathrm{MP})| &\leq |F(s',\mathrm{MP}') - F(s,\mathrm{MP}')| + |F(s,\mathrm{MP}') - F(s,\mathrm{MP})| \\
&= \left| \int_{\Delta_{\mathcal{X}}} \big(f(s',p) - f(s,p)\big) \cdot d\mathrm{MP}'(p) \right| + |F_s(\mathrm{MP}') - F_s(\mathrm{MP})| \\
&< \left(\int_{\Delta_{\mathcal{X}}} |f(s',p) - f(s,p)| \cdot d\mathrm{MP}'(p) \right) + \frac{\epsilon}{2} \overset{(a)}{\leq} \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,
\end{aligned}$$

where (a) follows from the fact that MP' is a meta-probability measure and $|f(s',p) - f(s',p)| \leq \frac{\epsilon}{2}$ for every $p \in \Delta_{\mathcal{X}}$. We conclude that F is continuous.

12.6.6 Weak-$*$ Continuity of the Product of Meta-Probability Measures

Let $(\mathrm{MP}_{1,n})_{n \geq 0}$ and $(\mathrm{MP}_{2,n})_{n \geq 0}$ be two sequences that weakly-$*$ converge to MP_1 and MP_2 in $\mathcal{MP}(\mathcal{X}_1)$ and $\mathcal{MP}(\mathcal{X}_2)$ respectively. Let $f : \Delta_{\mathcal{X}_1} \times \Delta_{\mathcal{X}_2} \to \mathbb{R}$ be a continuous and bounded mapping. Define the mapping $F : \Delta_{\mathcal{X}_1} \times \mathcal{MP}(\mathcal{X}_2)$ as follows:

$$F(p_1, \mathrm{MP}_2') = \int_{\Delta_{\mathcal{X}_2}} f(p_1,p_2) d\mathrm{MP}_2'(p_2).$$

Fix $\epsilon > 0$. Since $f(p_1,p_2)$ is continuous, Lemma 12.4 implies that F is continuous. Therefore, the mapping $p_1 \to F(p_1, \mathrm{MP}_2)$ is continuous on $\Delta_{\mathcal{X}_1}$, which implies that it is also bounded because $\Delta_{\mathcal{X}_1}$ is compact. Therefore,

$$\lim_{n \to \infty} \int_{\Delta_{\mathcal{X}_1}} F(p_1, \mathrm{MP}_2) d\mathrm{MP}_{1,n}(p_1) = \int_{\Delta_{\mathcal{X}_1}} F(p_1, \mathrm{MP}_2) d\mathrm{MP}_1(p_1)$$

because $(\mathrm{MP}_{1,n})_{n \geq 0}$ weakly-$*$ converges to MP_1. This means that there exists $n_1 \geq 0$ such that for every $n \geq n_1$, we have

$$\left| \int_{\Delta_{\mathcal{X}_1}} F(p_1, \mathrm{MP}_2) d\mathrm{MP}_{1,n}(p_1) - \int_{\Delta_{\mathcal{X}_1}} F(p_1, \mathrm{MP}_2) d\mathrm{MP}_1(p_1) \right| < \frac{\epsilon}{2}.$$

On the other hand, since F is continuous and since $\mathcal{MP}(\mathcal{X}_2)$ is compact under the weak-$*$ topology [80], Lemma 12.1 implies the existence of a weakly-$*$ open neighborhood U_{MP_2} of MP_2 such that $|F(p_1, \mathrm{MP}_2') - F(p_1, \mathrm{MP}_2)| \leq \frac{\epsilon}{2}$ for every $\mathrm{MP}_2' \in U_{\mathrm{MP}_2}$ and every $p_1 \in \Delta_{\mathcal{X}_1}$. Moreover, since $\mathrm{MP}_{2,n}$ weakly-$*$ converges to MP_2, there exists $n_2 \geq 0$ such that $\mathrm{MP}_{2,n} \in U_{\mathrm{MP}_2}$ for every $n \geq n_2$.

Therefore, for every $n \geq \max\{n_1, n_2\}$, we have

$$
\begin{aligned}
&\left| \int_{\Delta_{\mathcal{X}_1}} \left(\int_{\Delta_{\mathcal{X}_2}} f(p_1,p_2) d\mathrm{MP}_{2,n}(p_2) \right) d\mathrm{MP}_{1,n}(p_1) - \int_{\Delta_{\mathcal{X}_1}} \left(\int_{\Delta_{\mathcal{X}_2}} f(p_1,p_2) d\mathrm{MP}_{2}(p_2) \right) d\mathrm{MP}_{1}(p_1) \right| \\
&\leq \left| \int_{\Delta_{\mathcal{X}_1}} \left(\int_{\Delta_{\mathcal{X}_2}} f(p_1,p_2) d\mathrm{MP}_{2,n}(p_2) \right) d\mathrm{MP}_{1,n}(p_1) - \int_{\Delta_{\mathcal{X}_1}} \left(\int_{\Delta_{\mathcal{X}_2}} f(p_1,p_2) d\mathrm{MP}_{2}(p_2) \right) d\mathrm{MP}_{1,n}(p_1) \right| \\
&\quad + \left| \int_{\Delta_{\mathcal{X}_1}} \left(\int_{\Delta_{\mathcal{X}_2}} f(p_1,p_2) d\mathrm{MP}_{2}(p_2) \right) d\mathrm{MP}_{1,n}(p_1) - \int_{\Delta_{\mathcal{X}_1}} \left(\int_{\Delta_{\mathcal{X}_2}} f(p_1,p_2) d\mathrm{MP}_{2}(p_2) \right) d\mathrm{MP}_{1}(p_1) \right| \\
&= \left| \int_{\Delta_{\mathcal{X}_1}} \left(F(p_1, \mathrm{MP}_{2,n}) - F(p_1, \mathrm{MP}_2) \right) d\mathrm{MP}_{1,n}(p_1) \right| \\
&\qquad\qquad + \left| \int_{\Delta_{\mathcal{X}_1}} F(p_1, \mathrm{MP}_2) d\mathrm{MP}_{1,n}(p_1) - \int_{\Delta_{\mathcal{X}_1}} F(p_1, \mathrm{MP}_2) d\mathrm{MP}_{1}(p_1) \right| \\
&< \int_{\Delta_{\mathcal{X}_1}} |F(p_1, \mathrm{MP}_{2,n}) - F(p_1, \mathrm{MP}_2)| \, d\mathrm{MP}_{1,n}(p_1) + \frac{\epsilon}{2} \overset{(a)}{\leq} \int_{\Delta_{\mathcal{X}_1}} \frac{\epsilon}{2} \cdot d\mathrm{MP}_{1,n}(p_1) + \frac{\epsilon}{2} = \epsilon,
\end{aligned}
$$

where (a) follows from the fact $\mathrm{MP}_{2,n} \in U_{\mathrm{MP}_2}$ for every $n \geq n_2$. Therefore,

$$
\begin{aligned}
\lim_{n\to\infty} \int_{\Delta_{\mathcal{X}_1} \times \Delta_{\mathcal{X}_2}} & f \cdot d(\mathrm{MP}_{1,n} \times \mathrm{MP}_{2,n}) \\
&\overset{(a)}{=} \lim_{n\to\infty} \int_{\Delta_{\mathcal{X}_1}} \left(\int_{\Delta_{\mathcal{X}_2}} f(p_1,p_2) d\mathrm{MP}_{2,n}(p_2) \right) d\mathrm{MP}_{1,n}(p_1) \\
&= \int_{\Delta_{\mathcal{X}_1}} \left(\int_{\Delta_{\mathcal{X}_2}} f(p_1,p_2) d\mathrm{MP}_{2}(p_2) \right) d\mathrm{MP}_{1}(p_1) \\
&\overset{(b)}{=} \int_{\Delta_{\mathcal{X}_1} \times \Delta_{\mathcal{X}_2}} f \cdot d(\mathrm{MP}_1 \times \mathrm{MP}_2),
\end{aligned}
$$

where (a) and (b) follow from Fubini's theorem. We conclude that $(\mathrm{MP}_{1,n} \times \mathrm{MP}_{2,n})_{n\geq 0}$ weakly-$*$ converges to $(\mathrm{MP}_1 \times \mathrm{MP}_2)_{n\geq 0}$. Therefore the product of meta-probability measures is weakly-$*$ continuous.

12.6.7 Continuity of the Capacity

Since the mapping I is continuous, and since the space $\Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ is compact, the mapping I is uniformly continuous, i.e., for every $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that for every $(p_1, W_1), (p_2, W_2) \in \Delta_{\mathcal{X}} \times \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$, if $\|p_1 - p_2\|_1 := \sum_{x\in\mathcal{X}} |p_1(x) - p_2(x)| < \delta(\epsilon)$ and $d_{\mathcal{X},\mathcal{Y}}(W_1, W_2) < \delta(\epsilon)$, then

$$
|I(p_1, W_1) - I(p_2, W_2)| < \epsilon.
$$

Let $W_1, W_2 \in \mathrm{DMC}_{\mathcal{X},\mathcal{Y}}$ be such that $d_{\mathcal{X},\mathcal{Y}}(W_1, W_2) < \delta(\epsilon)$. For every $p \in \Delta_{\mathcal{X}}$, we have $\|p - p\|_1 = 0 < \delta(\epsilon)$ so we must have $|I(p, W_1) - I(p, W_2)| < \epsilon$. Therefore,

$$I(p, W_1) < I(p, W_2) + \epsilon \leq \sup_{p' \in \Delta_{\mathcal{X}}} I(p', W_2) + \epsilon = C(W_2) + \epsilon.$$

Therefore,

$$C(W_1) = \sup_{p \in \Delta_{\mathcal{X}}} I(p, W_1) \leq C(W_2) + \epsilon.$$

Similarly, we can show that $C(W_2) \leq C(W_1) + \epsilon$. This implies that $|C(W_1) - C(W_2)| \leq \epsilon$, hence C is continuous.

12.6.8 Measurability and Continuity of $C^{+,*}$

Let us first show that the random mapping $C^{+,*}$ is measurable. We need to show that the mapping $C_B^{+,*} : \Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}} \to \mathbb{R}$ is measurable for every $B \in \mathcal{B}(\Delta_{\mathcal{X}})$, where

$$C_B^{+,*}(p_1, p_2) = (C^{+,*}(p_1, p_2))(B), \quad \forall p_1, p_2 \in \Delta_{\mathcal{X}}.$$

For every $u_1 \in \mathcal{X}$, define the set

$$A_{u_1} = \{(p_1, p_2) \in \Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}} : \ (C^{-,*}(p_1, p_2))(u_1) > 0\}.$$

Clearly, A_{u_1} is open in $\Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}}$ (and so it is measurable). The mapping $C^{+,u_1,*}$ is defined on A_{u_1} and it is clearly continuous. Therefore, for every $B \in \mathcal{B}(\Delta_{\mathcal{X}})$, $(C^{+,u_1,*})^{-1}(B)$ is measurable. We have:

$$\begin{aligned} C_B^{+,*}(p_1, p_2) = (C^{+,*}(p_1, p_2))(B) &= \sum_{\substack{u_1 \in \mathrm{supp}(C^{-,*}(p_1,p_2)),\\ C^{+,u_1,*}(p_1,p_2) \in B}} (C^{-,*}(p_1, p_2))(u_1) \\ &= \sum_{\substack{u_1 \in \mathcal{X},\\ (p_1,p_2) \in A_{u_1},\\ C^{+,u_1,*}(p_1,p_2) \in B}} (C^{-,*}(p_1, p_2))(u_1) \\ &\overset{(a)}{=} \sum_{u_1 \in \mathcal{X}} (C^{-,*}(p_1, p_2))(u_1) \cdot \mathbb{1}_{(C^{+,u_1,*})^{-1}(B)}(p_1, p_2), \end{aligned}$$

where (a) follows from the fact that $(p_1, p_2) \in (C^{+,u_1,*})^{-1}(B)$ if and only if $(p_1, p_2) \in A_{u_1}$ and $C^{+,u_1,*}(p_1, p_2) \in B$. This shows that $C_B^{+,*}$ is measurable for every $B \in \mathcal{B}(\Delta_{\mathcal{X}})$. Therefore, $C^{+,*}$ is a measurable random mapping.

Let $(p_{1,n}, p_{2,n})_{n \geq 0}$ be a converging sequence to (p_1, p_2) in $\Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}}$. Since $C^{-,*}$ is continuous, we have $\lim_{n \to \infty} (C^{-,*}(p_{1,n}, p_{2,n}))(u_1) = (C^{-,*}(p_1, p_2))(u_1)$ for every $u_1 \in \mathcal{X}$. Therefore, for every $u_1 \in \mathrm{supp}(C^{-,*}(p_1, p_2))$, there exists $n_{u_1} \geq 0$ such that for every $n \geq n_{u_1}$, we have $C^{-,*}(p_{1,n}, p_{2,n}) > 0$. Let $n_0 = \max\{n_{u_1} : \ u_1 \in \mathrm{supp}(C^{-,*}(p_1, p_2))\}$. For every $n \geq n_0$, we have

$$\mathrm{supp}(C^{-,*}(p_1, p_2)) \subset \mathrm{supp}(C^{-,*}(p_{1,n}, p_{2,n})).$$

Therefore, for every continuous and bounded mapping $g : \Delta_{\mathcal{X}} \to \mathbb{R}$, we have

$$\begin{aligned}
&\lim_{n\to\infty} \int_{\Delta_{\mathcal{X}}} g \cdot d(C^{+,*}(p_{1,n}, p_{2,n})) \\
&\quad = \lim_{n\to\infty} \sum_{u_1 \in \mathrm{supp}(C^{-,*}(p_{1,n},p_{2,n}))} g(C^{+,u_1,*}(p_{1,n}, p_{2,n})) \cdot (C^{-,*}(p_{1,n}, p_{2,n}))(u_1) \\
&\quad \overset{(a)}{=} \lim_{n\to\infty} \sum_{u_1 \in \mathrm{supp}(C^{-,*}(p_1,p_2))} g(C^{+,u_1,*}(p_{1,n}, p_{2,n})) \cdot (C^{-,*}(p_{1,n}, p_{2,n}))(u_1) \\
&\quad \overset{(b)}{=} \sum_{u_1 \in \mathrm{supp}(C^{-,*}(p_1,p_2))} g(C^{+,u_1,*}(p_1, p_2)) \cdot (C^{-,*}(p_1, p_2))(u_1) \\
&\quad = \int_{\Delta_{\mathcal{X}}} g \cdot d(C^{+,*}(p_1, p_2)),
\end{aligned}$$

where (b) follows from the continuity of g and $C^{-,*}$, and the continuity of $C^{+,u_1,*}$ on A_{u_1} for every $u_1 \in \mathcal{X}$. (a) follows from the fact that:

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{\substack{u_1 \in \mathrm{supp}(C^{-,*}(p_{1,n},p_{2,n})),\\ u_1 \notin \mathrm{supp}(C^{-,*}(p_1,p_2))}} \left| g(C^{+,u_1,*}(p_{1,n}, p_{2,n})) \cdot (C^{-,*}(p_{1,n}, p_{2,n}))(u_1) \right| \\
&\qquad \le \|g\|_\infty \lim_{n\to\infty} \sum_{\substack{u_1 \in \mathrm{supp}(C^{-,*}(p_{1,n},p_{2,n})),\\ u_1 \notin \mathrm{supp}(C^{-,*}(p_1,p_2))}} (C^{-,*}(p_{1,n}, p_{2,n}))(u_1) \\
&\qquad = \|g\|_\infty \lim_{n\to\infty} \left(1 - \sum_{u_1 \in \mathrm{supp}(C^{-,*}(p_1,p_2))} (C^{-,*}(p_{1,n}, p_{2,n}))(u_1) \right) \\
&\qquad = \|g\|_\infty \left(1 - \sum_{u_1 \in \mathrm{supp}(C^{-,*}(p_1,p_2))} (C^{-,*}(p_1, p_2))(u_1) \right) = 0.
\end{aligned}$$

We conclude that the mapping $C^{+,*}$ is a continuous mapping from $\Delta_{\mathcal{X}} \times \Delta_{\mathcal{X}}$ to $\mathcal{MP}(\mathcal{X})$ when the latter space is endowed with the weak-$*$ topology.

12.6.9 Proof of Proposition 12.10

Let $\hat{W}_1 \in \mathrm{DMC}^{(o)}_{\mathcal{X}_1,*}$ and $\overline{W}_2 \in \mathrm{DMC}^{(o)}_{\mathcal{X}_2,*}$. Fix $W_1 \in \hat{W}_1$ and $W_2 \in \overline{W}_2$ and let $\mathcal{Y}_1$ and $\mathcal{Y}_2$ be the output alphabets of W_1 and W_2 respectively. We may assume without loss of generality that $\mathrm{Im}(W_1) = \mathcal{Y}_1$ and $\mathrm{Im}(W_2) = \mathcal{Y}_2$.

Let $y \in \mathcal{Y}_1$. We have

$$\begin{aligned}
P^o_{W_1 \oplus W_2}(y) &= \frac{1}{|\mathcal{X}_1 \coprod \mathcal{X}_2|} \sum_{x \in \mathcal{X}_1 \coprod \mathcal{X}_2} (W_1 \oplus W_2)(y|x) \\
&= \frac{1}{|\mathcal{X}_1| + |\mathcal{X}_2|} \sum_{x \in \mathcal{X}_1} W_1(y|x) = \frac{|\mathcal{X}_1|}{|\mathcal{X}_1| + |\mathcal{X}_2|} P^o_{W_1}(y) > 0.
\end{aligned}$$

For every $x \in \mathcal{X}_1$, we have

$$(W_1 \oplus W_2)^{-1}_y(x) = \frac{(W_1 \oplus W_2)(y|x)}{(|\mathcal{X}_1| + |\mathcal{X}_2|) P^o_{W_1}(y)} = \frac{W_1(y|x)}{|\mathcal{X}_1| P^o_{W_1}(y)} = (W_1)^{-1}_y(x).$$

On the other hand, for every $x \in \mathcal{X}_2$, we have

$$(W_1 \oplus W_2)_y^{-1}(x) = \frac{(W_1 \oplus W_2)(y|x)}{(|\mathcal{X}_1| + |\mathcal{X}_2|)P^o_{W_1}(y)} = 0.$$

Therefore $(W_1 \oplus W_2)_y^{-1} = \phi_{1\#}(W_1)_y^{-1}$, where ϕ_1 is the canonical injection from $\mathcal{X}_1$ to $\mathcal{X}_1 \coprod \mathcal{X}_2$.

Similarly, for every $y \in \mathcal{Y}_2$, we have $P^o_{W_1 \oplus W_2}(y) = \dfrac{|\mathcal{X}_2|}{|\mathcal{X}_1| + |\mathcal{X}_2|} P^o_{W_1}(y) > 0$ and $(W_1 \oplus W_2)_y^{-1} = \phi_{2\#}(W_2)_y^{-1}$, where ϕ_2 is the canonical injection from $\mathcal{X}_2$ to $\mathcal{X}_1 \coprod \mathcal{X}_2$. For every $B \in \mathcal{B}(\Delta_{\mathcal{X}_1 \coprod \mathcal{X}_2})$, we have:

$$\begin{aligned}
&\mathrm{MP}_{W_1 \oplus W_2}(B) \\
&\quad = \sum_{\substack{y \in \mathcal{Y}_1 \coprod \mathcal{Y}_2, \\ (W_1 \oplus W_2)_y^{-1} \in B}} P^o_{W_1 \oplus W_2}(y) \\
&\quad = \left(\sum_{\substack{y \in \mathcal{Y}_1, \\ \phi_{1\#}(W_1)_y^{-1} \in B}} \frac{|\mathcal{X}_1|}{|\mathcal{X}_1| + |\mathcal{X}_2|} P^o_{W_1}(y) \right) + \left(\sum_{\substack{y \in \mathcal{Y}_2, \\ \phi_{2\#}(W_2)_y^{-1} \in B}} \frac{|\mathcal{X}_2|}{|\mathcal{X}_1| + |\mathcal{X}_2|} P^o_{W_2}(y) \right) \\
&\quad = \frac{|\mathcal{X}_1|}{|\mathcal{X}_1| + |\mathcal{X}_2|} \mathrm{MP}_{W_1}\big((\phi_{1\#})^{-1}(B)\big) + \frac{|\mathcal{X}_2|}{|\mathcal{X}_1| + |\mathcal{X}_2|} \mathrm{MP}_{W_2}\big((\phi_{2\#})^{-1}(B)\big) \\
&\quad = \frac{|\mathcal{X}_1|}{|\mathcal{X}_1| + |\mathcal{X}_2|} (\phi_{1\#\#} \mathrm{MP}_{W_1})(B) + \frac{|\mathcal{X}_2|}{|\mathcal{X}_1| + |\mathcal{X}_2|} (\phi_{2\#\#} \mathrm{MP}_{W_2})(B).
\end{aligned}$$

Therefore,

$$\mathrm{MP}_{\hat{W}_1 \oplus \overline{W}_2} = \frac{|\mathcal{X}_1|}{|\mathcal{X}_1| + |\mathcal{X}_2|} \phi_{1\#\#} \mathrm{MP}_{\hat{W}_1} + \frac{|\mathcal{X}_2|}{|\mathcal{X}_1| + |\mathcal{X}_2|} \phi_{2\#\#} \mathrm{MP}_{\overline{W}_2}.$$

This shows the first formula of Proposition 12.10.

For every $y = (y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2$, we have

$$\begin{aligned}
P^o_{W_1 \otimes W_2}(y) &= \sum_{(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2} \frac{1}{|\mathcal{X}_1 \times \mathcal{X}_2|} (W_1 \otimes W_2)(y_1, y_2 | x_1, x_2) \\
&= \sum_{\substack{x_1 \in \mathcal{X}_2, \\ x_2 \in \mathcal{X}_2}} \frac{W_1(y_1|x_1)}{|\mathcal{X}_1|} \cdot \frac{W_2(y_2|x_2)}{|\mathcal{X}_2|} = P^o_{W_1}(y_1) P^o_{W_2}(y_2) > 0.
\end{aligned}$$

For every $x = (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$, we have

$$\begin{aligned}
(W_1 \otimes W_2)_y^{-1}(x) &= \frac{(W_1 \otimes W_2)(y|x)}{|\mathcal{X}_1 \times \mathcal{X}_2| P^o_{W_1 \otimes W_2}(y)} = \frac{W_1(y_1|x_1)}{|\mathcal{X}_1| P^o_{W_1}(y_1)} \cdot \frac{W_2(y_2|x_2)}{|\mathcal{X}_2| P^o_{W_2}(y_2)} \\
&= (W_1)_{y_1}^{-1}(x_1) \cdot (W_2)_{y_2}^{-1}(x_2) = \big((W_1)_{y_1}^{-1} \times (W_2)_{y_2}^{-1}\big)(x).
\end{aligned}$$

For every $B \in \mathcal{B}(\Delta_{\mathcal{X}_1 \times \mathcal{X}_2})$, we have

$$\begin{aligned}
\mathrm{MP}_{W_1 \otimes W_2}(B) &= \sum_{\substack{y \in \mathcal{Y}_1 \times \mathcal{Y}_2, \\ (W_1 \otimes W_2)_y^{-1} \in B}} P^o_{W_1 \otimes W_2}(y) = \sum_{\substack{y \in \mathcal{Y}_1 \times \mathcal{Y}_2, \\ (W_1)_{y_1}^{-1} \times (W_2)_{y_2}^{-1} \in B}} P^o_{W_1}(y_1) P^o_{W_2}(y_2) \\
&= \sum_{\substack{y \in \mathcal{Y}_1 \times \mathcal{Y}_2, \\ \mathrm{Mul}\left((W_1)_{y_1}^{-1}, (W_2)_{y_2}^{-1}\right) \in B}} P^o_{W_1}(y_1) P^o_{W_2}(y_2) \\
&= (\mathrm{MP}_{W_1} \times \mathrm{MP}_{W_2})(\mathrm{Mul}^{-1}(B)) = \left(\mathrm{Mul}_{\#}(\mathrm{MP}_{W_1} \times \mathrm{MP}_{W_2})\right)(B) \\
&= (\mathrm{MP}_{W_1} \otimes \mathrm{MP}_{W_2})(B).
\end{aligned}$$

Therefore,

$$\mathrm{MP}_{\hat{W}_1 \otimes \overline{W}_2} = \mathrm{MP}_{\hat{W}_1} \otimes \mathrm{MP}_{\overline{W}_2}.$$

This shows the second formula of Proposition 12.10.

Now let $\alpha \in [0,1]$ and $\hat{W}_1, \hat{W}_2 \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$. Fix $W_1 \in \hat{W}_1$ and $W_2 \in \hat{W}_2$ and let $\mathcal{Y}_1$ and $\mathcal{Y}_2$ be the output alphabets of W_1 and W_2 respectively. We may assume without loss of generality that $\mathrm{Im}(W_1) = \mathcal{Y}_1$ and $\mathrm{Im}(W_2) = \mathcal{Y}_2$. Let $W = [\alpha W_1, (1-\alpha) W_2]$. If $\alpha = 0$, then W is output-equivalent to W_2 and $\mathrm{MP}_W = \mathrm{MP}_{W_2} = \alpha \,\mathrm{MP}_{W_1} + (1 - \alpha)\,\mathrm{MP}_{W_2}$. If $\alpha = 1$, then W is output-equivalent to W_1 and $\mathrm{MP}_W = \mathrm{MP}_{W_1} = \alpha \,\mathrm{MP}_{W_1} + (1-\alpha)\,\mathrm{MP}_{W_2}$.

Assume now that $0 < \alpha < 1$. For every $y \in \mathcal{Y}_1$, we have:

$$P^o_W(y) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} W(y|x) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \alpha \cdot W_1(y|x) = \alpha P^o_{W_1}(y) > 0.$$

For every $x \in \mathcal{X}$, we have:

$$W_y^{-1}(x) = \frac{W(y|x)}{|\mathcal{X}| P^o_W(y)} = \frac{\alpha W_1(y|x)}{|\mathcal{X}| \alpha P^o_{W_1}(y)} = (W_1)_y^{-1}(x).$$

Similarly, for every $y \in \mathcal{Y}_2$, we have $P^o_W(y) = (1-\alpha) P^o_{W_2}(y) > 0$ and $W_y^{-1} = (W_2)_y^{-1}$. Therefore,

$$\begin{aligned}
\mathrm{MP}_W &= \sum_{y \in \mathcal{Y}_1 \coprod \mathcal{Y}_2} P^o_W(y) \cdot \delta_{W_y^{-1}} \\
&= \left(\sum_{y \in \mathcal{Y}_1} \alpha P^o_{W_1}(y) \cdot \delta_{(W_1)_y^{-1}} \right) + \left(\sum_{y \in \mathcal{Y}_2} (1-\alpha) P^o_{W_2}(y) \cdot \delta_{(W_2)_y^{-1}} \right) \\
&= \alpha \mathrm{MP}_{W_1} + (1-\alpha) \mathrm{MP}_{W_2}.
\end{aligned}$$

Therefore,

$$\mathrm{MP}_{[\alpha \hat{W}_1, (1-\alpha)\hat{W}_2]} = \alpha \mathrm{MP}_{\hat{W}_1} + (1-\alpha) \mathrm{MP}_{\hat{W}_2}.$$

This shows the third formula of Proposition 12.10.

Now let $\hat{W} \in \mathrm{DMC}^{(o)}_{\mathcal{X},*}$ and let $*$ be a uniformity-preserving binary operation on $\mathcal{X}$. Fix $W \in \hat{W}$ and let $\mathcal{Y}$ be the output alphabet of W. We may assume without loss of generality that $\mathrm{Im}(W) = \mathcal{Y}$.

Let U_1, U_2 be two independent random variables uniformly distributed in $\mathcal{X}$. Let $X_1 = U_1 * U_2$ and $X_2 = U_2$. Send X_1 and X_2 through two independent copies of W and let Y_1 and Y_2 be the output respectively.

For every $(y_1, y_2) \in \mathcal{Y}^2$, we have

$$P^o_{W^-}(y_1, y_2) = P_{Y_1,Y_2}(y_1, y_2) = P_{Y_1}(y_1)P_{Y_2}(y_2) = P^o_W(y_1)P^o_W(y_2) > 0.$$

For every $u_1 \in \mathcal{X}$, we have:

$$\begin{aligned}(W^-)^{-1}_{y_1,y_2}(u_1) &= P_{U_1|Y_1,Y_2}(u_1|y_1, y_2) = \sum_{u_2\in\mathcal{X}_2} P_{U_1,U_2|Y_1,Y_2}(u_1, u_2|y_1, y_2)\\ &= \sum_{u_2\in\mathcal{X}_2} P_{X_1,X_2|Y_1,Y_2}(u_1 * u_2, u_2|y_1, y_2)\\ &= \sum_{u_2\in\mathcal{X}_2} P_{X_1|Y_1}(u_1 * u_2|y_1)P_{X_2|Y_2}(u_2|y_2)\\ &= \sum_{u_2\in\mathcal{X}_2} W^{-1}_{y_1}(u_1 * u_2)W^{-1}_{y_2}(u_2) = \big(C^{-,*}(W^{-1}_{y_1}, W^{-1}_{y_2})\big)(u_1).\end{aligned}$$

For every $B \in \mathcal{B}(\Delta_{\mathcal{X}})$, we have

$$\begin{aligned}\mathrm{MP}_{W^-}(B) &= \sum_{\substack{y\in\mathcal{Y}^2,\\ (W^-)^{-1}_y\in B}} P^o_{W^-}(y) = \sum_{\substack{(y_1,y_2)\in\mathcal{Y}^2,\\ C^{-,*}(W^{-1}_{y_1},W^{-1}_{y_2})\in B}} P^o_{W_1}(y_1)P^o_{W_2}(y_2)\\ &= (\mathrm{MP}_W \times \mathrm{MP}_W)\big((C^{-,*})^{-1}(B)\big)\\ &= \big(C^{-,*}_{\#}(\mathrm{MP}_W \times \mathrm{MP}_W)\big)(B) = (\mathrm{MP}_W, \mathrm{MP}_W)^{-,*}(B).\end{aligned}$$

Therefore,

$$\mathrm{MP}_{\hat{W}^-} = (\mathrm{MP}_{\hat{W}}, \mathrm{MP}_{\hat{W}})^{-,*}.$$

This shows the forth formula of Proposition 12.10.

For every $(y_1, y_2, u_1) \in \mathcal{Y}^2 \times \mathcal{X}$, we have:

$$\begin{aligned}P^o_{W^+}(y_1, y_2, u_1) &= P_{Y_1,Y_2,U_1}(y_1, y_2, y_1) = P_{Y_1,Y_2}(y_1, y_2)P_{U_1|Y_1,Y_2}(u_1|y_1, y_2)\\ &= P^o_W(y_1)P^o_W(y_2) \cdot \big(C^{-,*}(W^{-1}_{y_1}, W^{-1}_{y_2})\big)(u_1).\end{aligned}$$

Therefore,

$$\mathrm{Im}(W^+) = \bigcup_{(y_1,y_2)\in\mathcal{Y}^2} \{(y_1, y_2)\} \times \mathrm{supp}(C^{-,*}(W^{-1}_{y_1}, W^{-1}_{y_2})).$$

For every $(y_1, y_2, u_1) \in \mathrm{Im}(W^+)$, we have:

$$\begin{aligned}(W^+)^{-1}_{y_1,y_2,u_1}(u_2) &= P_{U_2|Y_1,Y_2,U_1}(u_2|y_1, y_2, u_1) = \frac{P_{U_1,U_2|Y_1,Y_2}(u_1, u_2|y_1, y_2)}{P_{U_1|Y_1,Y_2}(u_1|y_1, y_2)}\\ &= \frac{P_{X_1|Y_1}(u_1 * u_2|y_1)P_{X_2|Y_2}(u_2|y_2)}{\big(C^{-,*}(W^{-1}_{y_1}, W^{-1}_{y_2})\big)(u_1)} = \frac{W^{-1}_{y_1}(u_1 * u_2)W^{-1}_{y_2}(u_2)}{\big(C^{-,*}(W^{-1}_{y_1}, W^{-1}_{y_2})\big)(u_1)}\\ &= \big(C^{+,u_1,*}(W^{-1}_{y_1}, W^{-1}_{y_2})\big)(u_2).\end{aligned}$$

For every $B \in \mathcal{B}(\Delta_{\mathcal{X}})$, we have

$$
\begin{aligned}
&\mathrm{MP}_{W^+}(B) \\
&\quad= \sum_{(y_1,y_2)\in\mathcal{Y}^2} \sum_{\substack{u_1\in\mathrm{supp}(C^{-,*}(W_{y_1}^{-1},W_{y_2}^{-1}),\\ C^{+,u_1,*}(W_{y_1}^{-1},W_{y_2}^{-1})\in B}} P_W^o(y_1)P_W^o(y_2)\cdot\big(C^{-,*}(W_{y_1}^{-1},W_{y_2}^{-1})\big)(u_1) \\
&\quad= \sum_{(y_1,y_2)\in\mathcal{Y}^2} P_W^o(y_1)P_W^o(y_2) \sum_{\substack{u_1\in\mathrm{supp}(C^{-,*}(W_{y_1}^{-1},W_{y_2}^{-1}),\\ C^{+,u_1,*}(W_{y_1}^{-1},W_{y_2}^{-1})\in B}} \big(C^{-,*}(W_{y_1}^{-1},W_{y_2}^{-1})\big)(u_1) \\
&\quad= \sum_{(y_1,y_2)\in\mathcal{Y}^2} P_W^o(y_1)P_W^o(y_2)\big(C^{+,*}(W_{y_1}^{-1},W_{y_2}^{-1})\big)(B) \\
&\quad= \sum_{(y_1,y_2)\in\mathcal{Y}^2} P_W^o(y_1)P_W^o(y_2)(C_B^{+,*}(W_{y_1}^{-1},W_{y_2}^{-1}) \\
&\quad= \int_{\Delta_{\mathcal{X}}\times\Delta_{\mathcal{X}}} C_B^{+,*}(p_1,p_2)\cdot d(\mathrm{MP}_W\times\mathrm{MP}_W)(p_1,p_2) \\
&\quad= \big(C_{\#}^{+,*}(\mathrm{MP}_W\times\mathrm{MP}_W)\big)(B) = (\mathrm{MP}_W,\mathrm{MP}_W)^{+,*}(B).
\end{aligned}
$$

Therefore,

$$\mathrm{MP}_{\hat{W}^+} = (\mathrm{MP}_{\hat{W}},\mathrm{MP}_{\hat{W}})^{+,*}.$$

This shows the fifth and last formula of Proposition 12.10.

12.6.10 Proof of Proposition 12.16

Fix $W_1 \in \hat{W}_1$ and $W_2 \in \overline{W}_2$, and let $\mathcal{X}_1$ and $\mathcal{X}_2$ be the input alphabets of W_1 and W_2 respectively.

For every $x_1 \in \mathcal{X}_1$, we have $(W_1 \oplus W_2)_{x_1} = \phi_{1\#}(W_1)_{x_1}$. Similarly, for every $x_2 \in \mathcal{X}_2$, we have $(W_1 \oplus W_2)_{x_2} = \phi_{2\#}(W_2)_{x_2}$. Therefore,

$$
\begin{aligned}
&\mathrm{co}(\hat{W}_1 \oplus \overline{W}_2) \\
&\quad= \mathrm{co}\left(\left\{(W_1\oplus W_2)_x :\ x\in\mathcal{X}_1 \coprod \mathcal{X}_2\right\}\right) \\
&\quad= \mathrm{co}(\{(W_1\oplus W_2)_{x_1} :\ x_1\in\mathcal{X}_1\}\cup\{(W_1\oplus W_2)_{x_2} :\ x_2\in\mathcal{X}_2\}) \\
&\quad= \mathrm{co}(\{\phi_{1\#}(W_1)_{x_1} :\ x_1\in\mathcal{X}_1\}\cup\{\phi_{2\#}(W_2)_{x_2} :\ x_2\in\mathcal{X}_2\}) \\
&\quad= \bigcup_{0\le\lambda\le1}\Big((1-\lambda)\,\mathrm{co}(\{\phi_{1\#}(W_1)_{x_1} :\ x_1\in\mathcal{X}_1\}) + \lambda\,\mathrm{co}(\{\phi_{2\#}(W_2)_{x_2} :\ x_2\in\mathcal{X}_2\})\Big) \\
&\quad= \bigcup_{0\le\lambda\le1}\Big((1-\lambda)\phi_{1\#}\big(\mathrm{co}(\{(W_1)_{x_1} :\ x_1\in\mathcal{X}_1\})\big) + \lambda\phi_{2\#}\big(\mathrm{co}(\{(W_2)_{x_2} :\ x_2\in\mathcal{X}_2\})\big)\Big) \\
&\quad= \bigcup_{0\le\lambda\le1}\Big((1-\lambda)\phi_{1\#}(\mathrm{co}(W_1)) + \lambda\phi_{2\#}(\mathrm{co}(W_2))\Big) \\
&\quad= \bigcup_{0\le\lambda\le1}\Big((1-\lambda)\phi_{1\#}(\mathrm{co}(\hat{W}_1)) + \lambda\phi_{2\#}(\mathrm{co}(\overline{W}_2))\Big).
\end{aligned}
$$

For every $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$, we have $(W_1 \otimes W_2)_{(x_1,x_2)} = (W_1)_{x_1} \times (W_2)_{x_2}$. Therefore,

$$\begin{aligned}
\mathrm{co}(\hat{W}_1 \otimes \overline{W}_2) &= \mathrm{co}(\{(W_1 \otimes W_2)_{(x_1,x_2)} : \ (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2\}) \\
&= \mathrm{co}(\{(W_1)_{x_1} \times (W_2)_{x_2} : \ (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2\}) \\
&= \mathrm{co}(\{(W_1)_{x_1} : \ x_1 \in \mathcal{X}_1\} \otimes \{(W_2)_{x_2} : \ x_2 \in \mathcal{X}_2\}) \\
&= \mathrm{co}\Big(\mathrm{co}\big(\{(W_1)_{x_1} : \ x_1 \in \mathcal{X}_1\}\big) \otimes \mathrm{co}\big(\{(W_2)_{x_2} : \ x_2 \in \mathcal{X}_2\}\big)\Big) \\
&= \mathrm{co}\Big(\mathrm{co}(W_1) \otimes \mathrm{co}(W_2)\Big) = \mathrm{co}\Big(\mathrm{co}(\hat{W}_1) \otimes \mathrm{co}(\overline{W}_2)\Big).
\end{aligned}$$

12.6.11 Proof of Proposition 12.17

Fix $\hat{W}_1, \hat{W}_1' \in \mathrm{DMC}^{(i)}_{*,\mathcal{Y}_1}$ and $\overline{W}_2, \overline{W}_2' \in \mathrm{DMC}^{(i)}_{*,\mathcal{Y}_2}$. Let $R_1 \in \mathcal{R}(\mathrm{co}(\hat{W}_1), \mathrm{co}(\hat{W}_1'))$ and $R_2 \in \mathcal{R}(\mathrm{co}(\overline{W}_2), \mathrm{co}(\overline{W}_2'))$. Fix $0 \leq \lambda \leq 1$, $(P_1, P_1') \in R_1$ and $(P_2, P_2') \in R_2$. Let $P = (1-\lambda)\phi_{1\#}P_1 + \lambda\phi_{2\#}P_2$ and $P' = (1-\lambda)\phi_{1\#}P_1' + \lambda\phi_{2\#}P_2'$, where $\phi_{1\#}$ and $\phi_{2\#}$ are the push-forwards by the canonical injections from $\mathcal{Y}_1$ and $\mathcal{Y}_2$ to $\mathcal{Y}_1 \coprod \mathcal{Y}_2$ respectively. We have:

$$\begin{aligned}
\|P - P'\|_{TV} &= \big\|\big((1-\lambda)\phi_{1\#}P_1 + \lambda\phi_{2\#}P_2\big) - \big((1-\lambda)\phi_{1\#}P_1' + \lambda\phi_{2\#}P_2'\big)\big\|_{TV} \\
&\leq (1-\lambda)\|\phi_{1\#}P_1 - \phi_{1\#}P_1'\|_{TV} + \lambda\|\phi_{2\#}P_2 - \phi_{2\#}P_2'\|_{TV} \\
&= (1-\lambda)\|P_1 - P_1'\|_{TV} + \lambda\|P_2 - P_2'\|_{TV} \\
&\leq \|P_1 - P_1'\|_{TV} + \|P_2 - P_2'\|_{TV}.
\end{aligned} \tag{12.1}$$

Proposition 12.16 shows that

$$\mathrm{co}(\hat{W}_1 \oplus \overline{W}_2) = \bigcup_{0\leq\lambda\leq 1} \Big((1-\lambda)\phi_{1\#}(\mathrm{co}(\hat{W}_1)) + \lambda\phi_{2\#}(\mathrm{co}(\hat{W}_2))\Big),$$

and

$$\mathrm{co}(\hat{W}_1' \oplus \overline{W}_2') = \bigcup_{0\leq\lambda\leq 1} \Big((1-\lambda)\phi_{1\#}(\mathrm{co}(\hat{W}_1')) + \lambda\phi_{2\#}(\mathrm{co}(\hat{W}_2'))\Big).$$

Define $R \subset \mathrm{co}(\hat{W}_1 \oplus \overline{W}_2) \times \mathrm{co}(\hat{W}_1' \oplus \overline{W}_2')$ as follows:

$$\begin{aligned}
R = \Big\{ &\big((1-\lambda)\phi_{1\#}P_1 + \lambda\phi_{2\#}P_2, (1-\lambda)\phi_{1\#}P_1' + \lambda\phi_{2\#}P_2'\big) : \\
&\qquad 0 \leq \lambda \leq 1, (P_1, P_1') \in R_1, (P_2, P_2') \in R_2\Big\}.
\end{aligned}$$

It is easy to see that R is a coupling of $\mathrm{co}(\hat{W}_1 \oplus \overline{W}_2)$ and $\mathrm{co}(\hat{W}_1' \oplus \overline{W}_2')$. We have:

$$\begin{aligned}
d^{(i)}_{*,\mathcal{Y}_1 \coprod \mathcal{Y}_2}(\hat{W}_1 \oplus \overline{W}_2, \hat{W}_1' \oplus \overline{W}_2') &\leq \sup_{(P,P')\in R} \|P - P'\|_{TV} \\
&\overset{(a)}{\leq} \sup_{(P_1,P_1')\in R_1} \|P_1 - P_1'\|_{TV} + \sup_{(P_2,P_2')\in R_2} \|P_2 - P_2'\|_{TV},
\end{aligned}$$

where (a) follows from (12.1). Since this is true for every $R_1 \in \mathcal{R}(\mathrm{co}(\hat{W}_1), \mathrm{co}(\hat{W}_1'))$ and every $R_2 \in \mathcal{R}(\mathrm{co}(\hat{W}_2), \mathrm{co}(\hat{W}_2'))$, we conclude that

$$\begin{aligned} d^{(i)}_{*,\mathcal{Y}_1 \coprod \mathcal{Y}_2}&(\hat{W}_1 \oplus \overline{W}_2, \hat{W}_1' \oplus \overline{W}_2') \\ &\leq \inf_{R_1 \in \mathcal{R}(\mathrm{co}(\hat{W}_1), \mathrm{co}(\hat{W}_1'))} \sup_{(P_1, P_1') \in R_1} \|P_1 - P_1'\|_{TV} \\ &\qquad + \inf_{R_2 \in \mathcal{R}(\mathrm{co}(\hat{W}_2), \mathrm{co}(\hat{W}_2'))} \sup_{(P_2, P_2') \in R_2} \|P_2 - P_2'\|_{TV} \\ &= d^{(i)}_{*,\mathcal{Y}_1}(\hat{W}_1, \hat{W}_1') + d^{(i)}_{*,\mathcal{Y}_2}(\hat{W}_2, \hat{W}_2'). \end{aligned}$$

This shows that the mapping $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \oplus \overline{W}_2$ from $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}_1} \times \mathrm{DMC}^{(i)}_{*,\mathcal{Y}_2}$ to $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}_1 \coprod \mathcal{Y}_2}$ is continuous in the similarity topology.

Fix again $R_1 \in \mathcal{R}(\mathrm{co}(\hat{W}_1), \mathrm{co}(\hat{W}_1'))$ and $R_2 \in \mathcal{R}(\mathrm{co}(\overline{W}_2), \mathrm{co}(\overline{W}_2'))$. Let $\lambda_1, \ldots, \lambda_k \geq 0$ be such that $\sum_{i=1}^k \lambda_i = 1$. Let $(P_{1,1}, P_{1,1}'), \ldots, (P_{1,k}, P_{1,k}') \in R_1$ and $(P_{2,1}, P_{2,1}')$, $\ldots, (P_{2,k}, P_{2,k}') \in R_2$. Define $P = \sum_{i=1}^k \lambda_i P_{1,i} \times P_{2,i}$ and $P' = \sum_{i=1}^k \lambda_i P_{1,i}' \times P_{2,i}'$. We have:

$$\begin{aligned} \|P - P'\|_{TV} &= \left\| \left(\sum_{i=1}^k \lambda_i P_{1,i} \times P_{2,i} \right) - \left(\sum_{i=1}^k \lambda_i P_{1,i}' \times P_{2,i}' \right) \right\|_{TV} \\ &\leq \sum_{i=1}^k \lambda_i \|(P_{1,i} \times P_{2,i}) - (P_{1,i}' \times P_{2,i}')\|_{TV} \\ &\overset{(a)}{\leq} \sum_{i=1}^k \lambda_i \Big(\|P_{1,i} - P_{1,i}'\|_{TV} + \|P_{2,i} - P_{2,i}'\|_{TV} \Big) \\ &\leq \sup_{(P_1, P_1') \in R_1} \|P_1 - P_1'\|_{TV} + \sup_{(P_2, P_2') \in R_2} \|P_2 - P_2'\|_{TV}, \end{aligned} \tag{12.2}$$

where (a) follows from Appendix 12.6.2. Proposition 12.16 shows that

$$\mathrm{co}(\hat{W}_1 \otimes \overline{W}_2) = \mathrm{co}\Big(\mathrm{co}(\hat{W}_1) \otimes \mathrm{co}(\overline{W}_2) \Big),$$

and

$$\mathrm{co}(\hat{W}_1' \otimes \overline{W}_2') = \mathrm{co}\Big(\mathrm{co}(\hat{W}_1') \otimes \mathrm{co}(\overline{W}_2') \Big).$$

Define $R \subset \mathrm{co}(\hat{W}_1 \otimes \overline{W}_2) \times \mathrm{co}(\hat{W}_1' \otimes \overline{W}_2')$ as follows:

$$\begin{aligned} R = \Bigg\{ \left(\sum_{i=1}^k \lambda_i P_{1,i} \times P_{2,i}, \sum_{i=1}^k \lambda_i P_{1,i}' \times P_{2,i}' \right) : k \geq 1,\ \lambda_1, \ldots, \lambda_k \geq 0,\ \sum_{i=1}^k \lambda_i = 1, \\ (P_{1,1}, P_{1,1}'), \ldots, (P_{1,k}, P_{1,k}') \in R_1, \\ (P_{2,1}, P_{2,1}'), \ldots, (P_{2,k}, P_{2,k}') \in R_2 \Bigg\}. \end{aligned}$$

It is easy to see that R is a coupling of $\mathrm{co}(\hat{W}_1 \otimes \overline{W}_2)$ and $\mathrm{co}(\hat{W}'_1 \otimes \overline{W}'_2)$. We have:

$$\begin{aligned} d^{(i)}_{*,\mathcal{Y}_1\times\mathcal{Y}_2}(\hat{W}_1 \otimes \overline{W}_2, \hat{W}'_1 \otimes \overline{W}'_2) &\leq \sup_{(P,P')\in R} \|P - P'\|_{TV} \\ &\overset{(a)}{\leq} \sup_{(P_1,P'_1)\in R_1} \|P_1 - P'_1\|_{TV} + \sup_{(P_2,P'_2)\in R_2} \|P_2 - P'_2\|_{TV}, \end{aligned}$$

where (a) follows from (12.2). Since this is true for every $R_1 \in \mathcal{R}(\mathrm{co}(\hat{W}_1), \mathrm{co}(\hat{W}'_1))$ and every $R_2 \in \mathcal{R}(\mathrm{co}(\hat{W}_2), \mathrm{co}(\hat{W}'_2))$, we conclude that

$$\begin{aligned} &d^{(i)}_{*,\mathcal{Y}_1\times\mathcal{Y}_2}(\hat{W}_1 \otimes \overline{W}_2, \hat{W}'_1 \otimes \overline{W}'_2) \\ &\quad\leq \inf_{R_1\in\mathcal{R}(\mathrm{co}(\hat{W}_1),\mathrm{co}(\hat{W}'_1))} \sup_{(P_1,P'_1)\in R_1} \|P_1 - P'_1\|_{TV} \\ &\qquad\qquad + \inf_{R_2\in\mathcal{R}(\mathrm{co}(\hat{W}_2),\mathrm{co}(\hat{W}'_2))} \sup_{(P_2,P'_2)\in R_2} \|P_2 - P'_2\|_{TV} \\ &\quad= d^{(i)}_{*,\mathcal{Y}_1}(\hat{W}_1, \hat{W}'_1) + d^{(i)}_{*,\mathcal{Y}_2}(\hat{W}_2, \hat{W}'_2). \end{aligned}$$

This shows that the mapping $(\hat{W}_1, \overline{W}_2) \to \hat{W}_1 \otimes \overline{W}_2$ from $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}_1} \times \mathrm{DMC}^{(i)}_{*,\mathcal{Y}_2}$ to $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}_1 \coprod \mathcal{Y}_2}$ is continuous in the similarity topology.

13 Conclusion of Part II

In this chapter, we summarize the main contributions of the second part of this thesis. Furthermore, we briefly discuss some open problems and possible future directions in the channel ordering topic.

13.1 Characterization of Various Channel Orderings

In Chapter 10, we introduced the input-degradedness ordering of communication channels, and provided several characterizations for this ordering. We showed that if W is input-degraded from W', then any decoder that is good for W is also good for W'. We also studied the Shannon ordering of communication channels, and provided a characterization of it that is similar to the Blackwell-Sherman-Stein (BSS) theorem.

The output-degradedness ordering has been applied in network information theory, e.g., in the context of broadcast channels [87, 88, 89]. It is not clear whether input-degradedness can play a similar role for multiple-access channels.

As we explained in Chapter 10, the output-equivalence class of a channel can be identified by its Blackwell measure [68]. Similarly, the input-equivalence class of a channel can be identified by its input-equivalence characteristic (see Proposition 10.4). Finding a canonical representation[1] of the Shannon-equivalence class of a channel remains an open problem.

13.2 Topological Structures on DMC Spaces

13.2.1 Spaces of Output-Equivalent Channels

The fact that the noisiness and weak-$*$ topologies are the same gives us more freedom in proving theorems. Statements that might be difficult to prove using the weak-$*$ formulation might be easier to prove using the noisiness formulation. For example,

[1] By canonical representation, we mean a mathematical object S_W that is computable from the channel W, and which satisfies: $S_W = S_{W'}$ if and only if W is Shannon-equivalent to W'.

the convergence of the polarization process is slightly easier to prove in the noisiness formulation.

The strong topology is too strong to be adopted as the "standard natural topology". However, it can still be useful because it is relatively easy to work with as it has a quotient formulation. Moreover, since it is finer than the noisiness/weak-$*$ topology, many statements that are true for the strong topology are also true for coarser topologies, e.g., any sequence that converges in the strong topology also converges in the noisiness/weak-$*$ one.

Although the total variation topology is not natural, it can still be useful because it is finer than the noisiness/weak-$*$ topology.

Many interesting questions remain open: Are all natural topologies Hausdorff? Can we find more topological properties that are common for all natural topologies? Is there a coarsest natural topology? Is there a natural topology that is coarser than the noisiness/weak-$*$ one?

Finding meaningful measures on $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ might be challenging. One might be tempted to require that the measure of $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ should be zero because it is "finite dimensional" whereas $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ is "infinite dimensional". On the other hand, if $\mathrm{DMC}^{(o)}_{\mathcal{X},[n]}$ has a zero measure for every $n \geq 1$, the whole space $\mathrm{DMC}^{(o)}_{\mathcal{X},*}$ will have a zero measure because it is a countable union of these subspaces. Nevertheless, statements such as "the property X is true for almost all channels" can still make sense. One possible definition of null-sets is as follows: for every set A in the natural Borel σ-algebra, we say that A is a null-set if and only if there exists $n_0 \geq 1$ such that $P_n\left(\mathrm{Proj}_n^{-1}(A \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]})\right) = 0$ for every $n \geq n_0$, where Proj_n is the projection onto the $R^{(o)}_{\mathcal{X},[n]}$-equivalence classes and P_n is the uniform probability measure on $\mathrm{DMC}_{\mathcal{X},[n]} \equiv (\Delta_{[n]})^{\mathcal{X}}$. Another possible definition, which is weaker, is to say that A is a null-set if and only if $\lim_{n\to\infty} P_n\left(\mathrm{Proj}_n^{-1}(A \cap \mathrm{DMC}^{(o)}_{\mathcal{X},[n]})\right) = 0$.

13.2.2 Spaces of Input-Equivalent Channels

Since $\mathcal{T}^{(i)}_{*,\mathcal{Y}}$ is a natural topology, it is not completely metrizable because of Corollary 11.14 (assuming that $|\mathcal{Y}| \geq 3$). Therefore, the metric space $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, d^{(i)}_{*,\mathcal{Y}})$ is not complete. In contrast with the case of output-equivalence[2], the completion of $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, d^{(i)}_{*,\mathcal{Y}})$ does not represent the space of all input-equivalent channels with output alphabet $\mathcal{Y}$ and arbitrary input alphabet (with arbitrary cardinality). It is possible to show that the completion of $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, d^{(i)}_{*,\mathcal{Y}})$ represents all the channels $W : \mathcal{X} \longrightarrow \mathcal{Y}$ for which $\mathrm{co}(W) := \mathrm{co}(\{W_x : \ x \in \mathcal{X}\})$ is closed in $\Delta_{\mathcal{Y}}$. Therefore, not every channel with output alphabet $\mathcal{Y}$ can be approximated (in the similarity metric sense) by a channel with finite input alphabet.

Is it possible to find a metric d on $\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}$ whose induced topology is natural, and such that the completion of $(\mathrm{DMC}^{(i)}_{*,\mathcal{Y}}, d)$ represents the space of input-equivalent channels with output alphabet $\mathcal{Y}$ and arbitrary input-alphabet (with arbitrary cardinality)?

[2]See the discussion after the proof of Theorem 11.5.

Some of the questions of Section 13.2.1 can also be asked for the spaces of input-equivalent channels: Are all natural topologies Hausdorff? Can we find more topological properties that are common for all natural topologies? Is there a coarsest natural topology? Is there a natural topology that is coarser than the similarity one?

13.2.3 Space of Shannon-Equivalent Channels

From Remark 11.1, we can see that if Conjecture 11.1 is true, then $\mathcal{T}_{*,*}^{(s)}$ is not completely metrizable. A natural question to ask is: What does the completion of $(\mathrm{DMC}_{*,*}^{(s)}, d_{*,*}^{(s)})$ represent?

Some of the questions of Section 13.2.1 can also be asked for the space of Shannon-equivalent channels: Are all natural topologies Hausdorff? Can we find more topological properties that are common for all natural topologies? Is there a coarsest natural topology? Is there a natural topology that is coarser than the BRM one?

13.3 Continuity of Channel Parameters and Operations

In Chapter 12, we studied the continuity of many channel parameters and operations under various topologies on the space of output-equivalent channels, the space of input-equivalent channels, and the space of Shannon-equivalent channels. As we mentioned in the introduction, the continuity of channel parameters and operations might be helpful in the following two problems:

1. If a parameter (such as the optimal probability of error of a given code) is difficult to compute for a channel W, one can approximate it by computing the same parameter for a sequence of channels $(W_n)_{n\geq 0}$ that converges to W in some topology where the parameter is continuous.
2. The study of robustness of a communication system against the imperfect specification of the channel.

Many continuity-related problems remain open:

- The continuity of the channel sum and the channel product on the whole product space $(\mathrm{DMC}_{\mathcal{X}_1,*}^{(o)} \times \mathrm{DMC}_{\mathcal{X}_2,*}^{(o)}, \mathcal{T}_{s,\mathcal{X}_1,*}^{(o)} \otimes \mathcal{T}_{s,\mathcal{X}_2,*}^{(o)})$. As we mentioned in Section 12.3.2, it is sufficient to prove that the product topology $\mathcal{T}_{s,\mathcal{X}_1,*}^{(o)} \otimes \mathcal{T}_{s,\mathcal{X}_2,*}^{(o)}$ is compactly generated.
- The continuity of the channel parameters C, $P_{e,n,M}$ and $P_{e,\mathcal{D}}$ in the similarity topology $\mathcal{T}_{*,\mathcal{Y}}^{(i)}$.
- The continuity of the channel sum and the channel product on the whole product space $(\mathrm{DMC}_{*,\mathcal{Y}_1}^{(i)} \times \mathrm{DMC}_{*,\mathcal{Y}_2}^{(i)}, \mathcal{T}_{s,*,\mathcal{Y}_1}^{(i)} \otimes \mathcal{T}_{s,*,\mathcal{Y}_2}^{(i)})$. As we explained in Section 12.4.2, it is sufficient to prove that the product topology $\mathcal{T}_{s,*,\mathcal{Y}_1}^{(i)} \otimes \mathcal{T}_{s,*,\mathcal{Y}_2}^{(i)}$ is compactly generated.
- The continuity of the channel parameters C and $P_{e,n,M}$ in the BRM topology $\mathcal{T}_{*,*}^{(s)}$.

- The continuity of the channel sum and the channel product on the whole product space $(\mathrm{DMC}^{(s)}_{*,*} \times \mathrm{DMC}^{(s)}_{*,*}, \mathcal{T}^{(s)}_{s,*,*} \otimes \mathcal{T}^{(s)}_{s,*,*})$. As we explained in Section 12.5.2, it is sufficient to prove that the product topology $\mathcal{T}^{(s)}_{s,*,*} \otimes \mathcal{T}^{(s)}_{s,*,*}$ is compactly generated.
- The continuity of the channel sum and the channel product in the BRM topology.

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, July 1948.

[2] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *Information Theory, IEEE Transactions on*, vol. 55, no. 7, pp. 3051–3073, 2009.

[3] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: John Wiley & Sons, 2006.

[4] E. Şaşoğlu, E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," in *Information Theory Workshop, 2009. ITW 2009. IEEE*, 2009, pp. 144–148.

[5] W. Park and A. Barg, "Polar codes for q-ary channels," *Information Theory, IEEE Transactions on*, vol. 59, no. 2, pp. 955–969, 2013.

[6] A. G. Sahebi and S. S. Pradhan, "Multilevel channel polarization for arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 7839–7857, Dec 2013.

[7] E. Şaşoğlu, "Polar codes for discrete alphabets," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, 2012, pp. 2137–2141.

[8] E. Şaşoğlu, E. Telatar, and E. M. Yeh, "Polar codes for the two-user multiple-access channel," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6583–6592, Oct 2013.

[9] E. Abbe and E. Telatar, "Polar codes for the m-user multiple access channel," *Information Theory, IEEE Transactions on*, vol. 58, no. 8, pp. 5437–5448, Aug. 2012.

[10] C. Shannon, "A note on a partial ordering for communication channels," *Inform. Contr.*, vol. 1, pp. 390–397, 1958.

[11] D. Blackwell, "Comparison of experiments," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, 1951, pp. 93–102.

[12] S. Sherman, "On a theorem of Hardy, Littlewood, Polya, and Blackwell," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 37, no. 12, pp. 826–831, 1951.

[13] C. Stein, "Notes on a seminar on theoretical statistics. I. comparison of experiments," *Report, University of Chicago*, 1951.

[14] K. Martin, "Topology in information theory in topology," *Theoretical Computer Science*, vol. 405, no. 1-2, pp. 75–87, 2008, computational Structures for Modelling Space, Time and Causality.

[15] R. Nasser, "Ergodic theory meets polarization I: A foundation of polarization theory," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 2451–2455.

[16] ——, "An ergodic theory of binary operations, part I: Key properties," *IEEE Transactions on Information Theory*, vol. 62, no. 12, pp. 6931–6952, Dec 2016.

[17] R. Nasser and E. Telatar, "Polarization theorems for arbitrary DMCs," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, 2013, pp. 1297–1301.

[18] R. Nasser, "An ergodic theory of binary operations, part II: Applications to polarization," *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 1063–1083, Feb 2017.

[19] E. Arıkan and E. Telatar, "On the rate of channel polarization," in *2009 IEEE International Symposium on Information Theory*, June 2009, pp. 1493–1495.

[20] R. Nasser and E. Telatar, "Polar codes for arbitrary DMCs and arbitrary MACs," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 2917–2936, June 2016.

[21] R. Nasser, "Ergodic theory meets polarization II: A foundation of polarization theory for MACs," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 2456–2460.

[22] E. Arikan, "Polar coding for the slepian-wolf problem based on monotone chain rules," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012, pp. 566–570.

[23] S. Önay, "Successive cancellation decoding of polar codes for the two-user binary-input MAC," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, July 2013, pp. 1122–1126.

[24] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Transactions on Information Theory*, vol. 11, no. 1, pp. 3–18, January 1965.

[25] R. L. Dobrushin, "Mathematical problems in the Shannon theory of optimal coding of information," in *Proc. 4th Berkeley Symp. Mathematics, Statistics, and Probability*, vol. 1, 1961, pp. 211–252.

[26] V. Strassen, "Asymptotische abschätzungen in Shannon's informationstheorie," in *Trans. 3rd Prague Conf. Inf. Theory*, 1962, pp. 689–723.

[27] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *Information Theory, IEEE Transactions on*, vol. 55, no. 11, pp. 4947–4966, November 2009.

[28] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[29] S. H. Hassani, K. Alishahi, and R. L. Urbanke, "Finite-length scaling for polar codes," *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 5875–5898, Oct 2014.

[30] M. Mondelli, S. H. Hassani, and R. L. Urbanke, "Scaling exponent of list decoders with applications to polar codes," *IEEE Transactions on Information Theory*, vol. 61, no. 9, pp. 4838–4851, Sept 2015.

[31] A. Fazeli and A. Vardy, "On the scaling exponent of binary polarization kernels," in *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sept 2014, pp. 797–804.

[32] M. Mondelli, S. H. Hassani, and R. L. Urbanke, "Unified scaling of polar codes: Error exponent, scaling exponent, moderate deviations, and error floors," *IEEE Transactions on Information Theory*, vol. 62, no. 12, pp. 6698–6712, Dec 2016.

[33] E. Şaşoğlu, "Polar Coding Theorems for Discrete Systems," Ph.D. dissertation, IC, Lausanne, 2011. [Online]. Available: http://library.epfl.ch/theses/?nr=5219

[34] R. Nasser and E. Telatar, "Fourier analysis of MAC polarization," *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 3600–3620, June 2017.

[35] ——, "Fourier analysis of MAC polarization," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1427–1431.

[36] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the gaussian multiple-access channel," *Information Theory, IEEE Transactions on*, vol. 42, no. 2, pp. 364–375, Mar 1996.

[37] A. Grant, B. Rimoldi, R. Urbanke, and P. Whiting, "Rate-splitting multiple access for discrete memoryless channels," *Information Theory, IEEE Transactions on*, vol. 47, no. 3, pp. 873–890, Mar 2001.

[38] S. Hassani, R. Mori, T. Tanaka, and R. Urbanke, "Rate-dependent analysis of the asymptotic behavior of channel polarization," *Information Theory, IEEE Transactions on*, vol. 59, no. 4, pp. 2267–2276, April 2013.

[39] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015.

[40] J. Forney, G.D., "Exponential error bounds for erasure, list, and decision feedback schemes," *Information Theory, IEEE Transactions on*, vol. 14, no. 2, pp. 206–220, Mar 1968.

[41] R. Nasser, "Erasure schemes using generalized polar codes: Zero-undetected-error capacity and performance trade-offs," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 820–824.

[42] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, E. Şaşoğlu, and R. L. Urbanke, "Reed-Muller codes achieve capacity on erasure channels," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4298–4316, July 2017.

[43] M. M. Wilde and S. Guha, "Polar codes for classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 1175–1187, Feb 2013.

[44] C. Hirche, C. Morgan, and M. M. Wilde, "Polar codes in network quantum information theory," *IEEE Transactions on Information Theory*, vol. 62, no. 2, pp. 915–924, Feb 2016.

[45] R. Nasser and J. M. Renes, "Polar codes for arbitrary classical-quantum channels and arbitrary cq-MACs," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 281–285.

[46] ——, "Polar codes for arbitrary classical-quantum channels and arbitrary cq-MACs," *submitted*, 2017. [Online]. Available: http://arxiv.org/abs/1701.03397

[47] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. New York, NY, USA: Cambridge University Press, 2011.

[48] P. Sen, "Achieving the Han-Kobayashi inner bound for the quantum interference channel by sequential decoding," *arXiv:1109.0802*, September 2011.

[49] H. Barnum and E. Knill, "Reversing quantum dynamics with near-optimal quantum and classical fidelity," *Journal of Mathematical Physics*, vol. 43, no. 5, pp. 2097–2106, 2002. [Online]. Available: http://aip.scitation.org/doi/abs/10.1063/1.1459754

[50] A. Winter, "The capacity of the quantum multiple-access channel," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 3059–3065, Nov 2001.

[51] T. Ogawa and H. Nagaoka, "Making good codes for classical-quantum channel coding via quantum hypothesis testing," *IEEE Transactions on Information Theory*, vol. 53, no. 6, pp. 2261–2266, June 2007.

[52] A. S. Holevo, "Reliability function of general classical-quantum channel," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 2256–2261, Sep 2000.

[53] M. Tomamichel, "A framework for non-asymptotic quantum information theory," *arXiv:1406.2943*, 2012.

[54] R. Konig, R. Renner, and C. Schaffner, "The operational meaning of min- and max-entropy," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 4337–4347, Sept 2009.

[55] H. Weyl, "Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung)," *Mathematische Annalen*, vol. 71, pp. 441–479, 1912. [Online]. Available: http://eudml.org/doc/158545

[56] R. Bhatia, *Positive Definite Matrices*, ser. Princeton Series in Applied Mathematics. Princeton University Press, 2009.

[57] S. Korada, E. Şaşoğlu, and R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *Information Theory, IEEE Transactions on*, vol. 56, no. 12, pp. 6253–6264, Dec 2010.

[58] R. Mori and T. Tanaka, "Source and channel polarization over finite fields and Reed-Solomon matrices," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2720–2736, May 2014.

[59] ——, "Channel polarization on q-ary discrete memoryless channels by arbitrary kernels," in *2010 IEEE International Symposium on Information Theory*, June 2010, pp. 894–898.

[60] N. Presman, O. Shapira, S. Litsyn, T. Etzion, and A. Vardy, "Binary polarization kernels from code decompositions," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2227–2239, May 2015.

[61] R. Nasser, "Topological structures on DMC spaces," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 3175–3179.

[62] ——, "Topological structures on DMC spaces," *submitted*, 2017. [Online]. Available: http://arxiv.org/abs/1701.04467

[63] ——, "On the input-degradedness and input-equivalence between channels," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 2453–2457.

[64] ——, "On the input-degradedness and input-equivalence between channels," *submitted*, 2017. [Online]. Available: http://arxiv.org/abs/1702.00727

[65] ——, "A characterization of the Shannon ordering of communication channels," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 2448–2452.

[66] ——, "A characterization of the Shannon ordering of communication channels," *submitted*, 2017. [Online]. Available: http://arxiv.org/abs/1705.01394

[67] C. Shannon, "The zero error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, September 1956.

[68] E. Torgersen, *Comparison of Statistical Experiments*, ser. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1991.

[69] T. Richardson and R. Urbanke, *Modern Coding Theory*. New York, NY, USA: Cambridge University Press, 2008.

[70] F. Buscemi, "Degradable channels, less noisy channels, and quantum statistical morphisms: An equivalence relation," *Probl. Inf. Transm.*, vol. 52, no. 3, pp. 201–213, Jul. 2016.

[71] D. Du and P. Pardalos, *Minimax and Applications*, ser. Nonconvex Optimization and Its Applications. Springer US, 2013.

[72] H. Schwarte, "On weak convergence of probability measures, channel capacity and code error probabilities," *IEEE Transactions on Information Theory*, vol. 42, no. 5, pp. 1549–1551, Sep 1996.

[73] V. Rathi and R. Urbanke, "Density evolution, thresholds and the stability condition for non-binary LDPC codes," *IEE Proceedings - Communications*, vol. 152, no. 6, pp. 1069–1074, Dec 2005.

[74] A. Bennatan and D. Burshtein, "Design and analysis of nonbinary LDPC codes for arbitrary discrete-memoryless channels," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 549–583, Feb 2006.

[75] L. Cam and G. Yang, *Asymptotics in Statistics: Some Basic Concepts*, ser. Springer Series in Statistics. Springer New York, 2000.

[76] J. Kelley, *General Topology*, ser. Graduate Texts in Mathematics. Springer New York, 1975.

[77] J. Munkres, *Topology*, ser. Featured Titles for Topology Series. Prentice Hall, Incorporated, 2000.

[78] S. Franklin, "Spaces in which sequences suffice," *Fundamenta Mathematicae*, vol. 57, no. 1, pp. 107–115, 1965.

[79] N. E. Steenrod, "A convenient category of topological spaces." *Michigan Math. J.*, vol. 14, no. 2, pp. 133–152, 05 1967.

[80] C. Villani, *Topics in Optimal Transportation*, ser. Graduate studies in mathematics. American Mathematical Society, 2003.

[81] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2576–2595, May 2013.

[82] R. Nasser, "Continuity of channel parameters and operations under various DMC topologies," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 3185–3189.

[83] ——, "Continuity of channel parameters and operations under various DMC topologies," *submitted*, 2017. [Online]. Available: http://arxiv.org/abs/1701.04466

[84] R. Engelking, *General topology*, ser. Monografie matematyczne. PWN, 1977.

[85] M. Mondelli, S. H. Hassani, and R. L. Urbanke, "From polar to Reed-Muller codes: A technique to improve the finite-length performance," *IEEE Transactions on Communications*, vol. 62, no. 9, pp. 3084–3091, Sept 2014.

[86] M. Raginsky, "Channel polarization and Blackwell measures," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 56–60.

[87] T. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 2–14, Jan 1972.

[88] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Transactions on Information Theory*, vol. 19, no. 2, pp. 197–207, March 1973.

[89] R. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredaci Informaccii*, vol. 10, no. 3, pp. 3–14, July-Sept 1974.

Curriculum Vitae

Education

École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

- **Ph.D. in Communication and Computer Sciences, 2012 – 2017**
  Dissertation Title: *"Polarization and Channel Ordering: Characterizations and Topological Structures"*.
  Thesis Advisor: *Prof. Emre Telatar*.

- **MSc in Communication Systems, 2010 – 2012**
  GPA: 5.86/6.
  Thesis Title: *"Aspects of Computational Lithography"*.
  Advisors: *Dr. Paul Hurley (IBM) and Prof. Martin Vetterli (EPFL)*.

Lebanese University, Lebanon

- **Diplôme d'Ingénieur in Electrical and Electronics Engineering, 2005 – 2010**
  GPA: 85.3/100.

- **MSc in Mathematics, 2008 – 2010**
  GPA: 84.2/100.
  Thesis Title: *"Oriented Paths in n-Chromatic Digraphs"*.
  Advisor: *Prof. Amine El-Sahili*.

- **BSc in Mathematics, 2005 – 2008**
  GPA: 90.2/100.

Research Experience

Visiting Graduate Student, September 2016 – January 2017

- *Quantum Information Theory Laboratory at ETH Zürich, Switzerland*.

- Advisor: *Prof. Renato Renner*.

Internship (Master thesis), February 2012 – August 2012

- *IBM Research, Zürich, Switzerland*.

- Advisors: *Dr. Paul Hurley (IBM) and Prof. Martin Vetterli (EPFL).*

Research Assistant, September 2011 – January 2012

- *Information Theory Laboratory (LTHI) at EPFL, Lausanne, Switzerland.*
- Advisor: *Prof. Emre Telatar.*

Teaching Experience

Teaching Assistant at EPFL

- Principles of Digital Communications (Spring 2014, Spring 2016, Spring 2017).
- Advanced Probability (Fall 2015 – 2016).
- Statistical Signal Processing (Spring 2015).
- Information Theory and Coding (Fall 2013 – 2014, Fall 2014 – 2015).

Honors

- Finalist for the IEEE Jack Keil Wolf ISIT Student Paper Award (2015).
- EPFL I&C Doctoral Fellowship (2012 – 2013).
- Excellence scholarship for graduate studies from the Lebanese University (2011 – 2014).
- Ranked first among all graduated students of the Faculty of Engineering at the Lebanese University (2010).
- Ranked first among all graduated Math students of the Faculty of Sciences at the Lebanese University (BSc 2008 and MSc 2010).

Publications

Journal Papers and Preprints

(J1) R. Nasser, "Topological Structures on DMC Spaces", *arXiv: 1701.04467, Submitted*, 2017.

(J2) R. Nasser, "Continuity of Channel Parameters and Operations under Various DMC Topologies", *arXiv: 1701.04466, Submitted*, 2017.

(J3) R. Nasser, "On the Input-Degradedness and Input-Equivalence Between Channels", *arXiv: 1702.00727, Submitted*, 2017.

(J4) R. Nasser, "A Characterization of the Shannon Ordering of Communication Channels", *arXiv: 1705.01394, Submitted*, 2017.

(J5) R. Nasser, J.M. Renes, "Polar Codes for Arbitrary Classical-Quantum Channels and Arbitrary cq-MACs", *arXiv: 1701.03397, Submitted*, 2017.

(J6) R. Nasser, E. Telatar, "Fourier Analysis of MAC Polarization", *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 3600–3620, June 2017.

(J7) R. Nasser, "An Ergodic Theory of Binary Operations – Part II: Applications to Polarization", *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 1063–1083, Feb. 2017.

(J8) R. Nasser, "An Ergodic Theory of Binary Operations – Part I: Key Properties", *IEEE Transactions on Information Theory*, vol. 62, no. 12, pp. 6931–6952, Dec. 2016.

(J9) R. Nasser, E. Telatar, "Polar Codes for Arbitrary DMCs and Arbitrary MACs", *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 2917–2936, June 2016.

Conference Papers

(C1) R. Nasser, "Topological Structures on DMC Spaces", *IEEE International Symposium on Information Theory (ISIT)*, Aachen, 2017, pp. 3175–3179.

(C2) R. Nasser, "Continuity of Channel Parameters and Operations under Various DMC Topologies", *IEEE International Symposium on Information Theory (ISIT)*, Aachen, 2017, pp. 3185–3189.

(C3) R. Nasser, "On the Input-Degradedness and Input-Equivalence Between Channels", *IEEE International Symposium on Information Theory (ISIT)*, Aachen, 2017, pp. 2453–2457.

(C4) R. Nasser, "A Characterization of the Shannon Ordering of Communication Channels", *IEEE International Symposium on Information Theory (ISIT)*, Aachen, 2017, pp. 2448–2452.

(C5) R. Nasser, J.M. Renes, "Polar Codes for Arbitrary Classical-Quantum Channels and Arbitrary cq-MACs", *IEEE International Symposium on Information Theory (ISIT)*, Aachen, 2017, pp. 281–285.

(C6) R. Nasser, "Erasure Schemes Using Generalized Polar Codes: Zero-Undetected-Error Capacity and Performance Trade-offs", *IEEE International Symposium on Information Theory (ISIT)*, Barcelona, 2016, pp. 820–824.

(C7) E. Najm, R. Nasser, "Age of Information: The Gamma Awakening", *IEEE International Symposium on Information Theory (ISIT)*, Barcelona, 2016, pp. 2574–2578.

(C8) R. Nasser, E. Telatar, "Fourier Analysis of MAC Polarization", *IEEE International Symposium on Information Theory (ISIT)*, Hong Kong, 2015, pp. 1427–1431.

(C9) R. Nasser, "Ergodic Theory Meets Polarization I: A Foundation of Polarization Theory", *IEEE International Symposium on Information Theory (ISIT)*, Hong Kong, 2015, pp. 2451–2455.

(C10) R. Nasser, "Ergodic Theory Meets Polarization II: A Foundation of Polarization Theory for MACs", *IEEE International Symposium on Information Theory (ISIT)*, Hong Kong, 2015, pp. 2456–2460.

(C11) R. Nasser, E. Telatar, "Polarization Theorems for Arbitrary DMCs", *IEEE International Symposium on Information Theory (ISIT)*, Istanbul, 2013, pp. 1297–1301.

(C12) R. Nasser, P. Hurley, "On the Accuracy of Different Fourier Transforms of VLSI Designs", *in SPIE Advanced Lithography*. International Society for Optics and Photonics, 2013, pp. 868 319–868 319.

Patents

(P1) P. Droz, P. Hurley, R. Nasser, J. Paki, "Polygon Recovery for VLSI Mask Correction", *US Patent 8,819,600*.

Invited Talks

(T1) "A Foundation of Polarization Theory", *IBM Research, Zürich*, September 2017.

(T2) "Polar Coding: A Technique to Achieve Capacity", *CAMS Workshop on Coding Theory, American University of Beirut (AUB)*, April 2014.