Numéro National de Thèse : TODO

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON

opérée au sein de

l'École Normale Supérieure de Lyon

École Doctorale N°512

École Doctorale en Informatique et Mathématiques de Lyon

Spécialité de doctorat : Informatique

Soutenue publiquement le 01/12/2023, par :

Paul Fermé

Channel Coding with Non-Signaling Correlations

Titre français

Devant le jury composé de :

VALJEAN Jean, Associate Professor, Università di Catania (Italie)	Rapporteur
VALJEANNE Jeanne, Assistant Professor, I.I.T. Madras (Inde)	Examinatrice
FAWZI Omar, Directeur de Recherche, Inria et ENS de Lyon	Directeur de thèse

Dédicace,
peut-être…

Résumé

Dans cette thèse, nous étudions…

Abstract

In this thesis, we study...

Remerciements/Acknowledgements

Je tiens tout d'abord à remercier mon directeur de thèse Omar Fawzi, qui m'a accompagné durant toute la durée de ma thèse et m'a gratifié de sa présence et de sa disponibilité tout au long de cette aventure.

I would also like to thank… english

Ma famille ?

Je tenais aussi à remercier mes amis…

Contents

List of Symbols

General Notations

to	TO
do	DO

Usual sets

$\mathbb{N}$	the set of nonnegative integers
$\mathbb{R}$	the set of real numbers
$[k]$	the set of integers from 1 to k, i.e. $\{1, \ldots, k\}$

Résumé substantiel en Français

Intro

Section 1

Resumé section 1

Sous-section 1

Résumé sous-section 1

Chapter 1

Introduction

TODO

Chapter 2

Background

TODO

Chapter **3**

Tight Approximation Guarantees for Concave Coverage Problems

Coverage functions are central objects of study in combinatorial optimization. Problems related to optimizing such functions arise in multiple fields, such as operations research [CFN77], machine learning [FK14], algorithmic game theory [DV15], and information theory [BF18]. The most basic covering problem is the *maximum coverage* one. In this problem, we are given subsets $T_1, \ldots, T_m$ of a universe $[n]$, along with a positive integer k, and the objective is to find a size-k subset $S \subseteq [m]$ that maximizes the coverage function $C(S) := |\bigcup_{i \in S} T_i|$. A fundamental result in the field of approximation algorithms establishes that an approximation ratio of $1 - e^{-1}$ can be achieved for this problem in polynomial-time [Hoc97] and, in fact, this approximation guarantee is tight, under the assumption that P $\neq$ NP [Fei98].

Note that in the maximum coverage problem, an element $a \in [n]$ is counted at most once in the objective, even if a appears in several selected sets. However, if we think of elements $a \in [n]$ as goods or resources, there are many settings wherein the utility indeed increases with the number of copies of a that get accumulated. Motivated, in part, by such settings, we consider a generalization of the maximum coverage problem where an element a can contribute by an amount that depends on the number of times it is covered.

Given a function $\varphi : \mathbb{N} \to \mathbb{R}_+$, an integer $k \in \mathbb{N}$, a universe of elements $[n]$, positive weights w_a for each $a \in [n]$, and subsets $T_1, \ldots, T_m \subseteq [n]$, the φ-MaxCoverage problem entails maximizing $C^{\varphi}(S) := \sum_{a \in [n]} w_a \varphi(|S|_a)$ over subsets $S \subseteq [m]$ of cardinality k; here $|S|_a = |\{i \in S : a \in T_i\}|$.

This chapter focuses on functions φ that are nondecreasing and concave (i.e., $\varphi(i+2) - \varphi(i+1) \leq \varphi(i+1) - \varphi(i)$ for $i \in \mathbb{N}$). We will also assume that the function φ is normalized in the sense that $\varphi(0) = 0$ and $\varphi(1) = 1$.[1] Our approximation guarantees are in terms of the *Poisson concavity ratio* of φ, which we define as follows:

$$\alpha_{\varphi} := \inf_{x \in \mathbb{N}^*} \frac{\mathbb{E}[\varphi(\text{Poi}(x))]}{\varphi(\mathbb{E}[\text{Poi}(x)])} = \inf_{x \in \mathbb{N}^*} \frac{\mathbb{E}[\varphi(\text{Poi}(x))]}{\varphi(x)} . \tag{3.1}$$

[1] One can always replace a generic φ to a normalized one without changing the optimal solutions through a simple affine transformation.

Here $\mathrm{Poi}(x)$ denotes a Poisson-distributed random variable with parameter x. We will write $\alpha_\varphi(x) := \frac{\mathbb{E}[\varphi(\mathrm{Poi}(x))]}{\varphi(x)}$, with $\alpha_\varphi(0) = 1$, and hence (see Proposition 3.18), $\alpha_\varphi = \min_{x\in\mathbb{N}^*} \alpha_\varphi(x) = \inf_{x\in\mathbb{R}_+} \alpha_\varphi(x)$.[2]

Our main result is that the φ-MaxCoverage problem admits an efficient α_φ-approximation algorithm, when φ is normalized nondecreasing concave, and this approximation guarantee is tight when φ grows sublinearly. Formally,

Theorem 3.1. *For any normalized nondecreasing concave function φ, there exists a α_φ-approximation algorithm for the φ-MaxCoverage problem running in polynomial time. Furthermore, for $\varphi(n) = o(n)$, it is NP-hard to approximate the φ-MaxCoverage problem within a factor better than $\alpha_\varphi + \varepsilon$, for any constant $\varepsilon > 0$.*

Before detailing the proof of the theorem, we provide a few remarks and connections to related work.

Applications and related work We can directly reduce the standard maximum coverage problem to φ-MaxCoverage by setting $\varphi(j) = \min\{j, 1\}$. In this case $\alpha_\varphi = 1 - e^{-1}$. One can also encapsulate, within our framework, the ℓ-MultiCoverage problem studied in [BFGG20] by instantiating $\varphi(j) = \min\{j, \ell\}$. In this setting, we recover the approximation ratio $\alpha_\varphi = 1 - \frac{\ell^\ell e^{-\ell}}{\ell!}$, which matches the approximation guarantee obtained in [BFGG20] (see Proposition 3.31). Note that the hardness result in [BFGG20] was based on the Unique Games Conjecture, whereas here we prove that this guarantee is tight under P $\neq$ NP.

The initial motivation for studying φ-MaxCoverage was to generalize the list-decoding problem interpretation studied in [BFGG20]. In the channel coding problem [BF18], we are given a noisy channel $W(y|x)$, defined as a conditional probability distribution, and the goal is to send a message chosen uniformly in $[k]$ through this channel. In order to do so, the sender encodes the input message in $\mathcal{X}$, and the receiver decodes back the output of the channel from $\mathcal{Y}$ to $[k]$. The objective is to find a code that maximize the probability of successfully decoding the message. In the list-decoding setting, instead of decoding the output into a single message, the decoder outputs a list of size ℓ of possible messages, where ℓ is some fixed parameter common to all lists. We consider that the decoding is successful if the initial message belongs to that list. The generalization of that problem, which we call the φ-list-decoding problem, allows decoding lists of any size. However, the decoding will be successful only with probability $\frac{\varphi(\ell)}{\ell}$ when the initial message belongs to that list, where ℓ is the size of that list. One can see that with a coverage function $\varphi(j) = \min\{j, \ell\}$, we recover the list-decoding problem. When the channel is of the form $W(y|x) = \frac{1}{t}$ for $y \in T_x$ with $|T_x| = t$ and $W(y|x) = 0$ elsewhere, the success probability of the φ-list-decoding problem for a code $S \subseteq \mathcal{X}$ can be written as $\frac{1}{kt}\sum_{y\in\mathcal{Y}} \varphi(|S|_y)$, with $|S|_y = |\{x \in S : y \in T_x\}|$, which is a particular instance of the φ-MaxCoverage problem. Therefore, for that class of channels, a tight approximation ratio α_φ follows from Theorem 3.1.

Another application of φ-MaxCoverage is in the context of multiwinner elections that entail selecting k (out of m) candidates with the objective of maximizing the cumulative utility of n voters; here, the utility of each voter $a \in [n]$ increases as more and more approved (by

[2] We require φ to be defined for nonnegative integers and will extend it over $\mathbb{R}_+$ by considering its piecewise linear extension.

a) candidates get selected. One can reduce multiwinner elections to a coverage problem by considering subset $T_i \subseteq [n]$ as the set of voters that approve of candidate $i \in [m]$ and $\varphi(j)$ as the utility that an agent achieves from j approved selections.[3] Addressing multiwinner elections in this standard utilitarian model, Dudycz et al. [DMMS20] obtain tight approximation guarantees for some well-studied classes of utilities. Specifically, the result in [DMMS20] applies to the classic *proportional approval voting rule*, which assigns a utility of $\sum_{i=1}^{j} \frac{1}{i}$ for j approved selections. This voting rule corresponds to the coverage problem with $\varphi(j) = \sum_{i=1}^{j} \frac{1}{i}$, which we denote as the Proportional Approval Voting problem (PAV for short). Section 3.3.2 shows that Theorem 3.1 holds for all the settings considered in [DMMS20] and, in fact, applies more generally. In particular, the voting version of ℓ-MultiCoverage (studied in [SFL16]) can be addressed by Theorem 3.1, but not by the result in [DMMS20]. Such a separation also arises when one truncates the proportional approval voting rule to, say, ℓ candidates, i.e., upon setting $\varphi(j) = \sum_{i=1}^{\min\{j,\ell\}} \frac{1}{i}$. Given that multiwinner elections model multiple real-world settings (e.g., committee selection [SFL16] and parliamentary proceedings [BLS17]), instantiations of φ-MaxCoverage in such social-choice contexts substantiate the applicability of our algorithmic result.

Coverage functions arise in numerous resource-allocation settings, such as sensor allocation [MW08], job scheduling, and plant location [CFN77]. The goal, broadly, in such setups is to select k subsets of resources (out of m pre-specified ones) such that the *welfare* generated by the selected resources is maximized–each resource's contribution to the welfare increases with the number of times it is selected. This problem can be cast as φ-MaxCoverage by setting n to be the number of resources, $\{T_i\}_{i\in[m]}$ as the given collection of subsets, and $\varphi(j)$ to be the welfare contribution of a resource when it is covered j times.[4] Here, we mention a specific allocation problem to highlight the relevance of studying φ beyond the standard coverage and ℓ-coverage formulations (see Section 3.3.4 for details): in the Vehicle-Target Assignment problem [Mur00, PM18] (VTA for short) the resources are n targets and covering a target j times contributes $\varphi^p(j) = \frac{1-(1-p)^j}{p}$ to the welfare; here, $p \in (0, 1)$ is a given parameter. Interestingly, we find that for this problem, the approximation ratio α_φ we obtain can outperform the *price of anarchy* (PoA), which corresponds to the approximation ratio of any method whereing the agents selfishly maximize their utilities (see Section 3.3.4 for further discussion of this point). By contrast, in the resource allocation problem with $\varphi(j) = \min\{j, \ell\}$, the price of anarchy is equal to α_φ; see [CPM19] for details. Another allocation problem studied in [PM18] corresponds to φ-MaxCoverage with $\varphi(j) = j^d$, for a given parameter $d \in (0, 1)$. We refer to this instantiation as the d-Power function.

Theorem 3.1 gives us a tight approximation bound of α_φ for all the above-mentioned applications of φ-MaxCoverage. The values of α_φ for these instantiations are listed in Table 3.1.

It is relevant to compare the approximation guarantee, α_φ, obtained here with the approximation ratio based on the notion of curvature of submodular functions. Note that if φ is nondecreasing and concave, then C^φ is submodular. One can show, via a direct

[3]Indeed, for a subset of candidates $S \subseteq [m]$, the utility of a voter $a \in [n]$ is equal to $\varphi(|S|_a)$, with $|S|_a = |\{i \in S : a \in T_i\}|$.

[4]Formally, to capture specific welfare-maximization problems in their entirety we have to a consider φ-MaxCoverage with a matroid constraint, and not just bound the number of selected subsets by k. Details pertaining to matroid constraints and the reduction appear in Section 3.1.1 and 3.3.3, respectively.

φ-MaxCoverage	$\varphi(j)$	α_φ	Derivation
MaxCoverage	$\min\{j,1\}$	$1-e^{-1}$	Prop. 3.31
ℓ-MultiCoverage	$\min\{j,\ell\}$	$1-\frac{\ell^\ell e^{-\ell}}{\ell!}$	Prop. 3.31
PAV	$\sum_{i=1}^{j}\frac{1}{i}$	$\alpha_\varphi(1)\simeq 0.7965\dots$	Prop. 3.28
PAV capped at 3	$\sum_{i=1}^{\min\{j,3\}}\frac{1}{i}$	$\alpha_\varphi(1)\simeq 0.7910\dots$	Prop. 3.21
p-VTA	$\frac{1-(1-p)^j}{p}$	$\frac{1-e^{-p}}{p}$	Prop. 3.32
0.1-VTA	$\frac{1-(1-0.1)^j}{0.1}$	$\frac{1-e^{-0.1}}{0.1}\simeq 0.9516\dots$	Prop. 3.32
0.1-VTA capped at 5	$\frac{1-(1-0.1)^{\min\{j,5\}}}{0.1}$	$\alpha_\varphi(5)\simeq 0.8470\dots$	Prop. 3.21
d-Power	j^d	$e^{-1}\sum_{k=1}^{+\infty}\frac{k^d}{k!}$	Prop. 3.33

Table 3.1 – Tight approximation ratios for particular choices of φ in the φ-MaxCoverage problem.

calculation, that for such a submodular C^φ the curvature (as defined in [CC84]) is given by $c = 1-(\varphi(m)-\varphi(m-1))$ for instances with at most m cover sets; see Proposition 3.19. Therefore, the algorithm of Sviridenko et al. [SVW17] provides an approximation ratio of $1-ce^{-1}$ for the φ-MaxCoverage problem. We note that the Poisson concavity ratio α_φ is always greater than or equal to this curvature-dependent ratio (Proposition 3.22). Specifically, for p-Vehicle-Target Assignment, it is strictly better for all $p\notin\{0,1\}$ and for ℓ-MultiCoverage, it is strictly better for all $\ell\geq 2$ as remarked in [BFGG20]. Therefore, for the setting at hand, we improve the approximation guarantee obtained in [SVW17].

Remarks on the Poisson concavity ratio α_φ. By Jensen's inequality along with the nonnegativity and concavity of φ, we have that $\alpha_\varphi\in[0,1]$. We show that α_φ can be computed numerically up to any precision $\varepsilon>0$, in time that is polynomial in $\frac{1}{\varepsilon}$. In fact, Proposition 3.17 shows that $\alpha_\varphi(x)\geq 1-\varepsilon$ for all $x\geq N_\varepsilon := \lceil\left(\frac{6}{\varepsilon}\right)^4\rceil$. Thus, we can iterate over all $x\in\{1,2,\dots,N_\varepsilon\}$ and find $\min_{x\in[N_\varepsilon]}\alpha_\varphi(x)$ up to ε precision (under reasonable assumptions on φ). This gives us a method to overall compute α_φ, up to an absolute error of 2ε: if $\alpha_\varphi\leq 1-\varepsilon$, then computing $\min_{x\in[N_\varepsilon]}\alpha_\varphi(x)$ (up to ε precision) suffices. Otherwise, if $\alpha_\varphi\geq 1-\varepsilon$, then $\alpha_\varphi(1)\leq 1$ provides the desired bound. Furthermore, we note that Proposition 3.16 shows that even if we consider $\alpha_\varphi(x)$ over all $x\in\mathbb{R}_+$, an infimum (i.e., the value of α_φ) is achieved at an integer.

Further hardness under Gap-ETH Theorem 3.1 shows that, under the assumption P $\neq$ NP, no polynomial-time algorithm can approximate φ-MaxCoverage within a better ratio than α_φ for sublinear φ. One natural question that arises is whether relaxing the running time constraint helps. More precisely, since there are $\binom{m}{k}=O(m^k)$ choices of k cover sets among the m available, a simple exhaustive search algorithm works in time $O(m^k)$. We can ask if FPT algorithms with respect to k, running in time $f(k)\cdot m^{o(k)}$ with f an arbitrary function, can do better. As in [DMMS20], we use the result of [Man20] to show in Theorem 3.14 that such algorithms cannot approximate φ-MaxCoverage within a better ratio than α_φ for sublinear φ, under the Gap-ETH hypothesis [CCK$^+$17]; see Section 3.2.3 for more details. This means that the brute-force strategy is essentially the best, if one wants to get a better approximation ratio than α_φ.

Proof techniques and organization In Section 3.1, we present our approximation algorithm for the φ-MaxCoverage. The algorithm is an application of *pipage rounding*, a technique introduced in [AS04], on a linear programming relaxation of φ-MaxCoverage. We show that the multilinear extension F^φ of C^φ is efficiently computable and thus, we can compute an integer solution x^{int} from the optimal fractional one x^* satisfying $C^\varphi(x^{\text{int}}) \geq F^\varphi(x^*)$. Using the notion of *convex order* between distributions, we show that $F^\varphi(x^*) \geq \sum_{a\in[n]} w_a \mathbb{E}[\varphi(\text{Poi}(|x^*|_a))]$, where $|x|_a = \sum_{i\in[m]:a\in T_i} x_i$. Comparing this to the value $\sum_{a\in[n]} w_a\varphi(|x^*|_a)$ taken by the linear program, we get a ratio given by the *Poisson concavity ratio* α_φ. The concavity of φ is crucial at several steps of the proof: it guarantees that the natural relaxation can be written as a linear program, it is used to relate between sums of Bernouilli random variables and a Poisson random variable via the convex order, as well as for the fact that we can restrict the infimum in the definition of α_φ to integer values of x. The generalization to matroid constraints follows in a standard way and is presented in Section 3.1.1.

In Section 3.2, we present the hardness result for φ-MaxCoverage. For this, we define a generalization of the partitioning gadget of Feige [Fei98], extending also [BFGG20]. Roughly speaking, for an integer $x_\varphi \in \mathbb{N}$, it is a collection of x_φ-covers of the set $[n]$ (an x-cover is a collection of subsets such that each element $a \in [n]$ is covered x times, or in other words, its φ-coverage is $\varphi(x)n$) that are incompatible in the sense that if we take an element from each one of these x_φ-covers, then the φ-coverage is bounded approximately by $\mathbb{E}[\varphi(\text{Poi}(x_\varphi))]n$. Then, we construct an instance of φ-MaxCoverage from an instance of the NP-hard problem *Label Cover* (as in [DMMS20]) using such a gadget with $x_\varphi \in \text{argmin}_{x\in\mathbb{N}}\alpha_\varphi(x)$. Having set up the partitioning gadget, the analysis of the reduction can be obtained by carefully generalizing the reductions of [BFGG20] and [DMMS20].

In Section 3.3, we present different domains of application of our result.

3.1 Approximation Algorithm for φ-MaxCoverage

Fix a function $\varphi : \mathbb{N} \to \mathbb{R}_+$ that is normalized, nondecreasing and concave. The φ-MaxCoverage problem is defined as follows. The input to the problem is given by positive integers n, m, t and m subsets $T_1, \ldots, T_m$ of the set $[n]$ (described as characteristic vectors), the weights $w_a \in \mathbb{Q}_+^*$ for $a \in [n]$ (described as a couple of bitstring of length t), as well as an integer $k \in \{1, \ldots, m\}$. The output is a subset $S \subseteq [m]$ of size k that maximizes $C^\varphi(S) = \sum_{a\in[n]} w_a\varphi(|S|_a)$, where $|S|_a = |\{i \in S : a \in T_i\}|$.

Note that the input to this problem can be specified using $n(m + 2t) + O(\log nmt)$ bits. To reduce the number of parameters, we will assume that t is polynomial in n and m, so that a polynomial time algorithm for this problem means an algorithm that runs in time polynomial in n and m. The counting function φ is fixed and does not depend on the instance of the problem, but for a given instance the problem only depends on the values $\varphi(0), \varphi(1), \ldots, \varphi(m)$. We assume that we have black box access to φ and to ensure that all the algorithms run in polynomial time, we assume that $\varphi(j)$ can be described with a number of bits that is polynomial in j and that this description can be computed in polynomial time.

We now describe the approximation algorithm for φ-MaxCoverage that we analyze. As described above, we follow the standard relax and round strategy, as in [BFGG20]. First, we define a natural convex relaxation.

Definition 3.1 (Relaxed program)**.**

$$\begin{aligned} \text{maximize} \quad & \sum_{a\in[n]} w_a c_a \\ \text{subject to} \quad & c_a \le \varphi(|x|_a), \forall a \in [n], \text{ with } |x|_a := \sum_{i\in[m]:a\in T_i} x_i \\ & 0 \le x_i \le 1, \forall i \in [m] \\ & \sum_{i=1}^{m} x_i = k \ . \end{aligned} \tag{3.2}$$

As previously mentioned, φ is defined on $\mathbb{R}_+$ by extending it in a piecewise linear fashion on non-integral points. As such, the constraint $c_a \le \varphi(|x|_a)$ is equivalent to m linear constraints. In fact, we can define φ_j to be the linear function $\varphi_j(t) = (\varphi(j)-\varphi(j-1))t-(j-1)\varphi(j)+j\varphi(j-1)$ for $j \in [m]$. Since φ is concave, we have that for all $t \in [0,m]$, $\varphi(t) = \min_{j\in[m]} \varphi_j(t)$. As such, the constraint $c_a \le \varphi(|x|_a)$ is equivalent to $c_a \le \varphi_j(|x|_a)$ for all $j \in [m]$ and so the program from Definition 3.1 is a linear program. Overall there are $n+m$ variables and $(n+1)m+1$ linear constraints, and by assumptions all the coefficients can be described using a number of bits that is polynomial in n and m. Hence an optimal solution of this linear program can be found in polynomial time.

Also observe that the program from Definition 3.1 is indeed a relaxation of the φ-MaxCoverage problem. To see this, given a set S of size k, consider the characteristic vector $x \in \{0,1\}^m$ defined by $x_i = 1$ if and only if $i \in S$. Then for all $a \in [n]$, we can set $c_a = \varphi(|x|_a) = \varphi(|S|_a)$, and we get an objective value of $\sum_{a\in[n]} w_a\varphi(|S|_a)$ which is exactly $C^\varphi(S)$. When solving the program from Definition 3.1, we get an optimal $x^* \in [0,1]^m$ which is in general not integral. Next, we describe a method to round it to an integral vector $x^{\text{int}} \in \{0,1\}^m$.

Rounding For a submodular function $f : \{0,1\}^m \to \mathbb{R}$, one can use pipage rounding [AS04, Von07, CCPV11] to transform, in polynomial time, any fractional solution $x \in [0,1]^m$ satisfying $\sum_{i=1}^m x_i = k$ into an integral vector $x^{\text{int}} \in \{0,1\}^m$ such that $\sum_{i=1}^m x_i^{\text{int}} = k$ and $F(x^{\text{int}}) \ge F(x)$, where F corresponds to the multilinear extension of f, provided that $F(x)$ is computable in polynomial time for a given x; see e.g., [Von07, Lemma 3.4]. The multilinear extension $F : [0,1]^m \to \mathbb{R}$ of f is defined by $F(x_1,\dots,x_m) := \mathbb{E}[f(X_1,\dots,X_m)]$, where X_i are independent random variables with $X_i \sim \text{Ber}(x_i)$, i.e., $X_i \in \{0,1\}$ with $\mathbb{P}(X_i = 1) = x_i$. Note that $F(x) = f(x)$ for an integral vector $x \in \{0,1\}^m$.

We apply this strategy to C^φ, which is shown to be submodular in Proposition 3.19, and the solution x^* of the LP relaxation from Definition 3.1. Note that overall the algorithm is polynomial time, since here $F(x)$ is computable in polynomial time for a given x (see Proposition 3.23). We now analyze the value returned by the algorithm. Using the property of pipage rounding, with the notation $X = (X_1,\dots,X_m)$ and $\text{Ber}(x) = (\text{Ber}(x_1),\dots,\text{Ber}(x_m))$, we get

$$C^\varphi(x^{\text{int}}) = \mathbb{E}_{X\sim\text{Ber}(x^{\text{int}})}[C^\varphi(X)] \ge \mathbb{E}_{X\sim\text{Ber}(x^*)}[C^\varphi(X)] \ .$$

Then it suffices to relate $\mathbb{E}_{X\sim\mathrm{Ber}(x^*)}[C^\varphi(X)]$ to the optimal value of the LP relaxation 3.1, which can only be larger than the optimal value of the φ-MaxCoverage problem.

Theorem 3.2. *Let x, c be a feasible solution of the program from Definition 3.1 and $X \sim Ber(x)$. Recalling the definition of α_φ and $\alpha_\varphi(j)$ from (3.1), we have*

$$\mathbb{E}_{X\sim Ber(x)}[C^\varphi(X)] \geq \left(\min_{j\in[m]} \alpha_\varphi(j)\right) \sum_{a\in[n]} w_a c_a \ .$$

In particular, this implies that the described polynomial time algorithm has an approximation ratio of α_φ:

$$C^\varphi(x^{int}) \geq \alpha_\varphi \sum_{a\in[n]} w_a c_a^* \geq \alpha_\varphi \max_{S\subseteq[m]:|S|=k} C^\varphi(S) \ .$$

In order to prove this theorem, we need the following lemma:

Lemma 3.3. *For φ concave, and $p \in [0,1]^m$, we have:*

$$\mathbb{E}\Big[\varphi\Big(\sum_{i=1}^m Ber(p_i)\Big)\Big] \geq \mathbb{E}\Big[\varphi\Big(Poi\Big(\sum_{i=1}^m p_i\Big)\Big)\Big] \ .$$

Proof. The notion of *convex order* discussed in [SS07] allows us to prove this result. We say that $X \leq_{\text{cx}} Y \iff \mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$ for any convex f. Thanks to Lemma 2.3 of [BFGG20], we have that for $p \in [0,1]$:

$$\mathrm{Ber}(p) \leq_{\text{cx}} \mathrm{Poi}(p) \ .$$

Since this order is preserved through convolution (Theorem 3.A.12 of [SS07]), and the fact that $\sum_{i=1}^m \mathrm{Poi}(p_i) \sim \mathrm{Poi}\left(\sum_{i=1}^m p_i\right)$, we have:

$$\sum_{i=1}^m \mathrm{Ber}(p_i) \leq_{\text{cx}} \mathrm{Poi}\left(\sum_{i=1}^m p_i\right) \ .$$

Applying this result to $-\varphi$, which is convex, concludes the proof. □

Proof of Theorem 3.2. By linearity of expectation and the fact that the weights w_a are positive, it is sufficient to show that for all $a \in [n]$:

$$\mathbb{E}[C_a^\varphi(X)] \geq \left(\min_{j\in[m]} \alpha_\varphi(j)\right) c_a \ ,$$

where $C_a^\varphi(S) := \varphi(|S|_a)$. Note that $|X|_a = \sum_{i\in[m]:a\in T_i} X_i$, and thus:

$$\begin{aligned}
\mathbb{E}[C_a^\varphi(X)] = \ & \mathbb{E}\Big[\varphi\Big(\sum_{i\in[m]:a\in T_i} X_i\Big)\Big] = \mathbb{E}\Big[\varphi\Big(\sum_{i\in[m]:a\in T_i} \mathrm{Ber}(x_i)\Big)\Big] \\
\geq \ & \mathbb{E}\Big[\varphi\Big(\mathrm{Poi}\Big(\sum_{i\in[m]:a\in T_i} x_i\Big)\Big)\Big] \text{ thanks to Lemma 5.12} \\
= \ & \mathbb{E}[\varphi(\mathrm{Poi}(|x|_a))] \geq \min\{\alpha_\varphi(\lfloor |x|_a \rfloor), \alpha_\varphi(\lceil |x|_a \rceil)\}\varphi(|x|_a) \\
& \text{thanks to Proposition 3.16} \\
\geq \ & \left(\min_{j\in[m]} \alpha_\varphi(j)\right)\varphi(|x|_a) \geq \left(\min_{j\in[m]} \alpha_\varphi(j)\right) c_a \ .
\end{aligned} \tag{3.3}$$

□

3.1.1 Generalization to Matroid Constraints

Instead of taking a cardinality constraint k on the size of the subset S, we look now at general matroid constraints on S. Specifically, as input, instead of k, we take a matroid $\mathcal{M}$ defined on $[m]$ and given by a set of linear constraints describing its base polytope $B(\mathcal{M})$. The output is a set $S \in \mathcal{M}$ that maximizes $C^{\varphi}(S)$. Note that the cardinality constraint considered above is the special case where $\mathcal{M}$ is the uniform matroid of all subsets of size at most k and the base polytope $B(\mathcal{M}) = \{x \in [0,1]^m : \sum_{i=1}^{m} x_i = k\}$.

We first note that in the order to establish Theorem 3.2, the cardinality constraint $\sum_{i=1}^{m} x_i = k$ is not used. Thus, since the pipage rounding strategy applies to matroid constraints $\mathcal{M}$ (see [Von07, Lemma 3.4]), the strategy and the analysis of its efficiency generalize immediately when applied to the following linear program:

Definition 3.2 (Relaxed program for matroid constraints)**.**

$$\begin{aligned}
\text{maximize} \quad & \sum_{a \in [n]} w_a c_a \\
\text{subject to} \quad & c_a \leq \varphi(|x|_a), \forall a \in [n] \\
& 0 \leq x_i \leq 1, \forall i \in [m] \\
& x \in B(\mathcal{M}) \quad \text{the base polytope of } \mathcal{M} \; .
\end{aligned} \tag{3.4}$$

Theorem 3.4. *Let x, c a feasible solution of the program from Definition 3.2 and $X \sim Ber(x)$. Then:*

$$\mathbb{E}_{X \sim Ber(x)}[C^{\varphi}(X)] \geq \left(\min_{j \in [m]} \alpha_{\varphi}(j) \right) \sum_{a \in [n]} w_a c_a \; .$$

In particular, this implies that the described polynomial time algorithm has an approximation ratio of α_{φ}:

$$C^{\varphi}(x^{int}) \geq \alpha_{\varphi} \sum_{a \in [n]} w_a c_a^* \geq \alpha_{\varphi} \max_{S \in \mathcal{M}} C^{\varphi}(S) \; .$$

3.2 Hardness of Approximation for φ-MaxCoverage

In this section, we establish an inapproximability bound for the φ-MaxCoverage problem with weights 1 under cardinality constraints. Throughout this section we use Γ to denote the universe of elements and, hence, an instance of the φ-MaxCoverage problem consists of Γ, along with a collection of subsets $\mathcal{F} = \{F_i \subseteq \Gamma\}_{i=1}^{m}$ and an integer k. Recall that the objective of this problem is to find a size-k subset $S \subseteq [m]$ that maximizes $C^{\varphi}(S) = \sum_{a \in \Gamma} \varphi(|S|_a)$.

We establish the following theorem in this section:

Theorem 3.5. *It is NP-hard to approximate the φ-MaxCoverage problem for $\varphi(n) = o(n)$ within a factor greater that $\alpha_{\varphi} + \varepsilon$ for any $\varepsilon > 0$.*

Our reduction is based on a problem called h-AryLabelCover, which is equivalent to the more standard GapLabelCover problem as will be shown in Appendix 3.5.3.

Definition 3.3 (h-AryLabelCover)**.** An instance $\mathcal{G} = (V, E, [L], [R], \{\pi_{e,v}\}_{e\in E, v\in e})$ of h-AryLabelCover is characterized by an h-uniform regular hypergraph (V, E) and constraints $\pi_{e,v} : [L] \to [R]$. Here, each h-uniform hyperedge represents a h-ary constraint. Additionally, for any labeling $\sigma : V \to [L]$, we have the following notions of strongly and weakly satisfied constraints:

- An edge $e = (v_1, \ldots, v_h) \in E$ is *strongly satisfied* by σ if:

$$\forall x, y \in [h], \pi_{e,v_x}(\sigma(v_x)) = \pi_{e,v_y}(\sigma(v_y)) \ .$$

- An edge $e = (v_1, \ldots, v_h) \in E$ is *weakly satisfied* by σ if:

$$\exists x \neq y \in [h], \pi_{e,v_x}(\sigma(v_x)) = \pi_{e,v_y}(\sigma(v_y)) \ .$$

Proposition 3.6 (δ, h-AryGapLabelCover)**.** *For any fixed integer $h \geq 2$ and fixed $\delta > 0$, there exists an R_0 such that for any integer $R \geq R_0$, it is NP-hard for instances $\mathcal{G} = (V, E, [L], [R], \{\pi_{e,v}\}_{e\in E, v\in e})$ of h-AryLabelCover with right alphabet $[R]$ to distinguish between:*

YES: *There exists a labeling σ that* strongly satisfies *all the edges.*

NO: *No labeling* weakly satisfies *more than δ fraction of the edges.*

3.2.1 Partitioning System

The key ingredient to prove Theorem 3.5 is a constant size combinatorial object called partitioning system, generalizing the work of Feige [Fei98] and [BFGG20]. For any set $[n]$, $\mathcal{Q} \subseteq 2^{[n]}$, we overload the definition $C^\varphi(\mathcal{Q}) := \sum_{a\in[n]} \varphi(|\mathcal{Q}|_a)$ with $|\mathcal{Q}|_a := |\{P \in \mathcal{Q} : a \in P\}|$ and $C^\varphi_a(\mathcal{Q}) := \varphi(|\mathcal{Q}|_a)$. Let us take $x_\varphi \in \text{argmin}_{x\in\mathbb{N}^*} \alpha_\varphi(x)$, thus $\alpha_\varphi = \alpha_\varphi(x_\varphi)$.

We say that $\mathcal{Q}$ is an *x-cover* of $x \in \mathbb{N}$ if every element of $[n]$ is covered x times, so $C^\varphi(\mathcal{Q}) = n\varphi(x)$.

Definition 3.4. An $([n], h, R, \varphi, \eta)$*-partitioning system* consists of R collections of subsets of $[n]$, $\mathcal{P}_1, \ldots, \mathcal{P}_R \subseteq 2^{[n]}$, that satisfy $\frac{x_\varphi n}{h} \in \mathbb{N}$, $x_\varphi \geq h$ and:

1. For every $i \in [R]$, $\mathcal{P}_i$ is a collection of h subsets $P_{i,1}, \ldots, P_{i,h} \subseteq [n]$ each of size $\frac{x_\varphi n}{h}$ which is an x_φ-cover.

2. For any $T \subseteq [R]$ and $\mathcal{Q} = \{P_{i,j(i)} : i \in T\}$ for some function $j : T \to [h]$, we have $\left|C^\varphi(\mathcal{Q}) - \psi^\varphi_{|T|,h} n\right| \leq \eta n$ where:

$$\psi^\varphi_{k,h} := \mathbb{E}\Big[\varphi\Big(\text{Bin}\left(k, \frac{x_\varphi}{h}\right)\Big)\Big] \ . \tag{3.5}$$

Remark. In particular, for any $\mathcal{Q} = \{Q_1, \dots, Q_k\}$ with Q_i of size $\frac{x_\varphi n}{h}$, we have that $C^\varphi(\mathcal{Q}) \leq n\varphi(k\frac{x_\varphi}{h})$. Indeed $C^\varphi(\mathcal{Q}) = \sum_{a\in[n]} \varphi(|\mathcal{Q}|_a)$ with $\sum_{a\in[n]} |\mathcal{Q}|_a = \sum_{i\in[k]} |Q_i| = k \cdot \frac{x_\varphi n}{h}$. By concavity of φ and Jensen's inequality, this function is maximized when all $|\mathcal{Q}|_a$ are equals, where we get $n\varphi(k\frac{x_\varphi}{h})$.

Proposition 3.7. *For every choice of $R, h \in \mathbb{N}$ with*

$$h \geq x_\varphi, \eta \in (0,1), n \geq \eta^{-2} R\varphi(R)^2 \log(20(h+1))$$

such that $\frac{x_\varphi n}{h} \in \mathbb{N}$, there exists a $([n], h, R, \varphi, \eta)$-partitioning system, which can be found in time exp($Rn\log(n)$) $\cdot$ poly(h).

The proof can be found in Appendix 3.5.4.

3.2.2 The Reduction

Proof of Theorem 3.5. Let $\varepsilon > 0$. Without loss of generality, we can assume that $\varepsilon < 1$. We show that it is NP-hard to reach an approximation greater than $\alpha_\varphi + \varepsilon$ for the φ-MaxCoverage problem, via a reduction from δ, h-AryGapLabelCover.

- $\eta = \frac{\varphi(x_\varphi)}{4x_\varphi}\varepsilon$, so $0 < \eta \leq \varepsilon < 1$,
- $h \geq x_\varphi$ such that $\left|\psi^\varphi_{h,h} - \alpha_\varphi\varphi(x_\varphi)\right| \leq \eta$ (see (3.5) for the definition of ψ^φ); such a choice exists thanks to Proposition 3.24,
- θ such that for all $x \geq \theta$, $\frac{\varphi(x)}{x} \leq \eta$, which exists since $\varphi(x) = o(x)$,
- $\xi = \frac{x_\varphi}{\theta}$,
- $\delta = \frac{\eta}{2}\frac{\xi^3}{h^2}$,
- $R \geq h$ large enough for Proposition 3.6 to hold.

Then, given an instance $\mathcal{G} = (V, E, [L], [R], \Sigma, \{\pi_{e,v}\}_{e\in E, v\in e})$ of δ, h-AryGapLabelCover, we construct an instance $(\Gamma, \mathcal{F}, k)$ of the φ-MaxCoverage problem with:

- n a large enough integer to have the existence of $([n], h, R, \varphi, \eta)$-partitioning systems using Proposition 3.7. Note that the size of these partitioning systems is independent of the size of the instance $\mathcal{G}$, and that one can find one of those in constant time, with relation to the size of the instance $\mathcal{G}$, thanks to Proposition 3.7.
- $\Gamma = [n] \times E$,
- $k = |V|$,
- Consider a $([n], h, R, \varphi, \eta)$-partitioning system, and call $\mathcal{P} = \{\mathcal{P}_1, \dots, \mathcal{P}_R\}$ the corresponding set of collections. Define sets $T^{e,v_j}_\beta = P_{\pi_{e,v_j}(\beta),j} \times \{e\}$ for $e = (v_1, \dots, v_h) \in E, j \in [h], \beta \in [L]$. Then, choose as cover sets $F^v_\beta := \bigsqcup_{e\in E: v\in e} T^{e,v}_\beta$ and take $\mathcal{F} := \{F^v_\beta, v \in V, \beta \in [L]\}$.

We will now prove that if we are in a YES instance, we have that there exists $\mathcal{T}$ of size k such that $C^{\varphi}(\mathcal{T}) \geq \varphi(x_{\varphi})|\Gamma|$ (completeness). Moreover, if we are in a NO instance, then we have that for all $\mathcal{T}$ of size $k = |V|$, $C^{\varphi}(\mathcal{T}) \leq (\alpha_{\varphi} + \varepsilon)\varphi(x_{\varphi})|\Gamma|$ (soundness). Establishing these two properties would conclude the proof. In fact, an algorithm for φ-MaxCoverage achieving a factor strictly greater than $\alpha_{\varphi} + \varepsilon$ would allow us to decide whether we have YES or a NO instance of the NP-hard problem δ, h-AryGapLabelCover.

In order to achieve this, let us define $C^{\varphi,e} := \sum_{a\in[n]\times\{e\}} C^{\varphi}_a$. In particular, $C^{\varphi} = \sum_{a\in\Gamma} C^{\varphi}_a = \sum_{e\in E} C^{\varphi,e}$. For $\mathcal{T} \subseteq \mathcal{F}$, we define the relevant part of $\mathcal{T}$ on e by:

$$\mathcal{T}_e := \{T^{e,v}_{\beta} : v \in e, \beta \in [L], F^v_{\beta} \in \mathcal{T}\} = \{F^v_{\beta} \cap ([n] \times \{e\}), F^v_{\beta} \in \mathcal{T}\} .$$

Note that $C^{\varphi,e}(\mathcal{T}) = C^{\varphi,e}(\mathcal{T}_e)$, and in particular $C^{\varphi}(\mathcal{T}) = \sum_{e\in E} C^{\varphi,e}(\mathcal{T}_e)$.

3.2.2.1 Completeness

Suppose the given h-AryLabelCover instance $\mathcal{G}$ is a YES instance. Then, there exists a labeling $\sigma : V \mapsto [L]$ which strongly satisfies all edges. Consider the collection of $|V|$ subsets $\mathcal{T} := \{F^v_{\sigma(v)} : v \in V\}$. Fix $e = (v_1, \dots, v_h) \in E$. Since e is strongly satisfied by σ, there exists $r \in [R]$ such that $\pi_{e,v_i}(\sigma(v_i)) = r$ for all $i \in [h]$. Thus, $\mathcal{T}_e = \{T^{e,v_i}_{\sigma(v_i)}\}_{i\in[h]} = \{P_{r,i} \times \{e\}\}_{i\in[h]}$ is an x_{φ}-cover of $[n] \times \{e\}$, and so $C^{\varphi,e}(\mathcal{T}_e) = n\varphi(x_{\varphi})$. Thus $C^{\varphi}(\mathcal{T}) = \sum_{e\in E} C^{\varphi,e}(\mathcal{T}_e) = |E|\varphi(x_{\varphi})n = \varphi(x_{\varphi})|\Gamma|$.

3.2.2.2 Soundness

Suppose the given h-AryLabelCover instance $\mathcal{G}$ is a NO instance. Let us prove the contrapositive of the soundness: we suppose that there exists $\mathcal{T}$ of size $k = |V|$ such that $C^{\varphi}(\mathcal{T}) > (\alpha_{\varphi} + \varepsilon)\varphi(x_{\varphi})|\Gamma|$. Let us show that there exists a labeling σ that weakly satisfies a strictly larger fraction of the edges than δ.

For every vertex $v \in V$, we define $L(v) := \{\beta \in [L] : F^v_{\beta} \in \mathcal{T}\}$ to be the candidate set of labels that can be associated with the vertex v. We extend this definition to hyperedges $e = (v_1, \dots, v_h)$ where we define $L(e) := \bigcup_{i\in[h]} L(v_i)$ to be the *multiset* of all labels associated with the edge. Note that $|\mathcal{T}_e| = |L(e)|$.

We say that $e = (v_1, \dots, v_h) \in E$ is *consistent* if and only if $\exists x \neq y \in [h], \pi_{e,v_x}(L(v_x)) \cap \pi_{e,v_y}(L(v_y)) \neq \emptyset$. We then decompose E in three parts:

- B is the set of edges $e \in E$ with $|L(e)| \geq \frac{h}{\xi}$.
- N is the set of consistent edges $e \in E$ with $|L(e)| < \frac{h}{\xi}$.
- $I = E - (B \cup N)$ is the set of inconsistent edges $e \in E$ with $|L(e)| < \frac{h}{\xi}$.

We want to show that the contribution of N is not too small, which we will use to construct a labeling weakly satisfying enough edges. This comes from the following lemmas:

Lemma 3.8. $\sum_{e\in E} |L(e)| = |E|h$

Proof. Recall that our h-uniform hypergraph is regular; call d its regular degree. In particular, we have that $d|V| = |E|h$. Note also that $\sum_{v\in V}|L(v)| = |\mathcal{T}| = |V|$. Thus:

$$\sum_{e\in E}|L(e)| = \sum_{e\in E}\sum_{v\in V:v\in e}|L(v)| = \sum_{v\in V}\sum_{e\in E:v\in e}|L(v)| = d\sum_{v\in V}|L(v)| = d|V| = |E|h\ .$$

□

Next, we bound the contribution of B:

Lemma 3.9. $\sum_{e\in B} C^{\varphi,e}(\mathcal{T}_e) \le \frac{\varepsilon}{4}\varphi(x_\varphi)|\Gamma|$.

Proof. We have:

$$\begin{aligned}
\sum_{e\in B} C^{\varphi,e}(\mathcal{T}_e) \le \quad & \sum_{e\in B} n\varphi\Big(|L(e)|\frac{x_\varphi}{h}\Big) \quad \text{by remark on Definition 3.4 and } |\mathcal{T}_e| = |L(e)| \\
\le \quad & |B|\cdot n\varphi\Big(\frac{\sum_{e\in B}|L(e)|}{|B|}\frac{x_\varphi}{h}\Big) \quad \text{by Jensen's inequality on concave } \varphi \\
\le \quad & |B|\cdot n\varphi\Big(\frac{|E|h}{|B|}\frac{x_\varphi}{h}\Big) \\
& \text{since } \varphi \text{ nondecreasing and } \sum_{e\in B}|L(e)| \le |E|h \text{ by Lemma 3.8} \\
= \quad & \frac{\varphi(\frac{|E|x_\varphi}{|B|})}{\frac{|E|x_\varphi}{|B|}} x_\varphi|\Gamma|\ .
\end{aligned} \tag{3.6}$$

We have seen that $\sum_{e\in B}|L(e)| \le |E|h$, but $\sum_{e\in B}|L(e)| \ge |B|\frac{h}{\xi}$ by definition of B, so we have that $\frac{|B|}{|E|} \le \xi$. Thus $\frac{|E|x_\varphi}{|B|} \ge \frac{x_\varphi}{\xi} = \theta$. By definition of θ, we get that $\sum_{e\in B} C^{\varphi,e}(\mathcal{T}_e) \le \eta x_\varphi|\Gamma| = \frac{\varepsilon}{4}\varphi(x_\varphi)|\Gamma|$. □

In order to bound the contribution of I, we will prove a property on inconsistent edges:

Proposition 3.10. *Let $e = (v_1,\dots,v_h) \in E$ be an inconsistent hyperedge with respect to $\mathcal{T}$. Then we have that $\left|C^{\varphi,e}(\mathcal{T}_e) - \psi^\varphi_{|L(e)|,h}n\right| \le \eta n$.*

Proof. Since e is inconsistent, $\forall x \ne y \in [h], \pi_{e,v_x}(L(v_x)) \cap \pi_{e,v_y}(L(v_y)) = \emptyset$. Therefore, for every $i \in [R]$, there is at most one $v \in e$ such that $i \in \pi_{e,v}(L(v))$, i.e., $\mathcal{T}_e$ intersects with $\mathcal{P}_i \times \{e\}$ in at most one subset. This gives us a subset $T \subseteq [R]$ and a function $j : T \to [h]$ such that $\mathcal{T}_e = \{P_{i,j(i)} \times \{e\} : i \in T\}$. As a consequence, $|T| = |\mathcal{T}_e| = |L(e)|$ and by the second condition of the partitioning system, we get the expected result. □

Now, we can bound the contribution of I:

Lemma 3.11. $\sum_{e\in I} C^{\varphi,e}(\mathcal{T}_e) \le (\alpha_\varphi + \frac{\varepsilon}{2})\varphi(x_\varphi)|\Gamma|$.

Proof. Thanks to Proposition 3.10, we have:

$$\sum_{e\in I} C^{\varphi,e}(\mathcal{T}_e) \le \sum_{e\in I}(\psi^\varphi_{|L(e)|,h} + \eta)n \le \sum_{e\in E}(\psi^\varphi_{|L(e)|,h} + \eta)n\ ,$$

since $I \subseteq E$ and $\psi^{\varphi}_{|L(e)|,h} \geq 0$. But $\sum_{e\in E} |L(e)| = |E|h$ by Lemma 3.8 and $x \mapsto \psi^{\varphi}_{x,h}$ is concave thanks to Proposition 3.26, so we can use Jensen's inequality to get $\sum_{e\in E} \psi^{\varphi}_{|L(e)|,h} \leq |E|\psi^{\varphi}_{\frac{\sum_{e\in E}|L(e)|}{|E|},h} = |E|\psi^{\varphi}_{h,h}$ and thus:

$$\sum_{e\in I} C^{\varphi,e}(\mathcal{T}_e) \leq (\psi^{\varphi}_{h,h} + \eta)n|E| \leq (\alpha_\varphi \varphi(x_\varphi) + 2\eta)|\Gamma| \ ,$$

by definition of h. This implies that the total contribution of inconsistent edges I is at most $\sum_{e\in I} C^{\varphi,e}(\mathcal{T}_e) \leq (\alpha_\varphi\varphi(x_\varphi) + 2\eta)|\Gamma| \leq (\alpha_\varphi + \frac{\varepsilon}{2})\varphi(x_\varphi)|\Gamma|$ by definition of η. □

Lemma 3.12. $\frac{|N|}{|E|} \geq \xi\eta$.

Proof. Since we have supposed that $\sum_{e\in E} C^{\varphi,e}(\mathcal{T}_e) = C^{\varphi}(\mathcal{T}) > (\alpha_\varphi + \varepsilon)\varphi(x_\varphi)|\Gamma|$, and with the help of Lemmas 3.9 and 3.11, we have that the contribution of N is:

$$\sum_{e\in N} C^{\varphi,e}(\mathcal{T}_e) > \frac{\varepsilon}{4}\varphi(x_\varphi)|\Gamma| \ .$$

However, we have that for $e \in N$ that $C^{\varphi,e}(\mathcal{T}_e) \leq n\varphi\Big(|\mathcal{T}_e|\frac{x_\varphi}{h}\Big) = n\varphi\Big(|L(e)|\frac{x_\varphi}{h}\Big) \leq n\varphi\Big(\frac{x_\varphi}{\xi}\Big) \leq \frac{nx_\varphi}{\xi}$ thanks to the remark on Definition 3.4 and the bound $|L(e)| < \frac{h}{\xi}$. This implies that:

$$\frac{|N|}{|E|} \geq \frac{\xi}{x_\varphi}\frac{\varepsilon\varphi(x_\varphi)}{4} = \xi\eta \ .$$

□

Finally, we construct a randomized labeling $\sigma : V \mapsto [L]$ as follows: for $v \in V$, if $L(v) \neq \emptyset$, set $\sigma(v)$ uniformly from $L(v)$, otherwise set it arbitrarily. We claim that in expectation, this labeling must weakly satisfy δ fraction of the hyperedges.

To see this, fix any $e = (v_1, \ldots, v_h) \in N$. Thus $\exists x \neq y \in [h], \pi_{e,v_x}(L(v_x)) \cap \pi_{e,v_y}(L(v_y)) \neq \emptyset$. Furthermore $|L(v_x)|, |L(v_y)| \leq \frac{h}{\xi}$. Thus, we have that $\pi_{e,v_x}(L(v_x)) = \pi_{e,v_y}(L(v_y))$ with probability at least $\frac{1}{|L(v_x)||L(v_y)|} \geq \left(\frac{\xi}{h}\right)^2$.

Therefore:

$$\begin{aligned}
& \mathbb{E}_\sigma\mathbb{E}_{e\sim E}[\sigma \text{ weakly satisfies } e] \\
\geq\ & \xi\eta\mathbb{E}_\sigma\mathbb{E}_{e\sim E}[\sigma \text{ weakly satisfies } e | e \in N] \quad \text{by Lemma 3.12} \\
>\ & \frac{\eta}{2}\frac{\xi^3}{h^2} = \delta \ .
\end{aligned} \tag{3.7}$$

In particular there exists some labeling σ such that $\mathbb{E}_{e\sim E}[\sigma \text{ weakly satisfies } e] > \delta$, and thus the soundness is also proved. □

3.2.3 Further Hardness under Gap-ETH

The Gap Exponential Time hypothesis states that, for some constant $\delta > 0$, there is no $2^{o(n)}$-time algorithm that, given n-variable 3-SAT formula, can distinguish whether the formula is fully satisfiable or that it is not even $(1-\delta)$-satisfiable. Gap-ETH is a standard assumption in proving FPT hardness of approximation (see e.g. [CCK+17]). Under such hypothesis, Manurangsi showed the following theorem:

Theorem 3.13 ([Man20], adapted to (δ, h)-AryGapLabelCover)**.** *Assuming Gap-ETH, for every $\delta > 0$, every $h \in \mathbb{N}, h \geq 2$ and any sufficiently large $R \in \mathbb{N}$ (depending on δ, h), no $f(k) \cdot N^{o(k)}$-time algorithm can solve (δ, h)-AryGapLabelCover with right alphabet $[R]$, where k denotes the number of vertices in h-AryLabelCover, N is the size of the instance, and f can be any function.*

Such a statement can be made in terms of the (δ, h)-AryGapLabelCover problem, since it can be shown to be equivalent to δ-Gap-Label-Cover(t, R) (see Appendix 3.5.3 for more details).

Furthermore, in the previous reduction, the constructed instance $(\Gamma, k, \mathcal{F})$ sizes are $|\Gamma| = n|E|$ (with n a constant independent of the size of the instance), $k = |V|$, and $|\mathcal{F}| = k \cdot L$. Therefore, plugin Theorem 3.13 in the previous reduction leads to the following hardness result:

Theorem 3.14. *Assuming Gap-ETH and $\varphi(n) = o(n)$, we cannot achieve an $(\alpha_\varphi + \varepsilon)$-approximation for the φ-MaxCoverage problem, even in $f(k) \cdot m^{o(k)}$-time, for any function f, with m the number of cover sets and k the cardinality constraint.*

3.3 Applications

This section shows that instantiations of φ-MaxCoverage encapsulate and generalize multiple problems from fields such as information theory [CT01], computational social choice [BCE+16] and algorithmic game theory [NRTV07].

3.3.1 Generalized List-Decoding

In the usual list-decoding problem, we are given a noisy channel $W(y|x)$ and a list size ℓ, and the goal is to send a message chosen uniformly in $[k]$ through this channel. In order to do so, the sender encodes the input message in $\mathcal{X}$, and the receiver decodes a list of size ℓ of possible messages in $[k]$ from the output of the channel in $\mathcal{Y}$. The goal is to maximize the probability of successfully decoding the message, which happens when the initial message belongs to the decoded list. We consider the generalization of the list-decoding problem that allows decoding lists of any size, with an associated cost function φ, which we call the φ-list-decoding problem. Now, the decoding will be successful only with probability $\frac{\varphi(\ell)}{\ell}$ when the initial message belongs to a decoding list of size ℓ.

Let us show that with a coverage function $\varphi(j) = \min\{j, \ell\}$, we recover the list-decoding problem. Indeed, let us consider an optimal solution $S \subseteq \mathcal{X}$, $\{L_y\}_{y \in \mathcal{Y}}$ of the φ-list-decoding, with L_y the list of possible messages when the output of the channel is y. If all lists L_y are of size ℓ, then as $\frac{\varphi(\ell)}{\ell} = 1$, we recover a solution of the list-decoding problem with the same objective value. This shows in particular that the φ-list-decoding problem has an objective optimal value at least as large as the usual list-decoding problem. Let us now consider on an arbitrary set $\{L_y\}_{y\in\mathcal{Y}}$, and let us look at a particular list L_y with $L_y \neq \ell$. Let us show that we can efficiently build another list of size ℓ with an objective value at least as good as with L_y. If $|L_y| \leq \ell$, then $\frac{\varphi(|L_y|)}{|L_y|} = 1$. Therefore, adding elements in L_y can only increase the success probability, so we can transform all the lists L_y with $|L_y| < \ell$ into lists of size ℓ without decreasing the value of that solution. If $|L_y| > \ell$, then $\frac{\varphi(|L_y|)}{|L_y|} = \frac{\ell}{|L_y|}$. Now, let us sort the elements $x \in L_y$ in a increasing order according to $\mathbb{1}_{x\in S}W(y|x)$. Let us remove the first $|L_y| - \ell$ elements of that list, and call L'_y the resulting list. We claim that replacing L_y by this reduced size list L'_y will not decrease the objective value. Indeed, we have kept the ℓ elements that contribute the most to the success probability. Therefore, the remaining success probability of having the input message in L'_y is larger than or equal to $\frac{\ell}{|L_y|}\sum_{y\in L_y}\mathbb{1}_{x\in S}W(y|x) = \frac{\varphi(|L_y|)}{|L_y|}\sum_{y\in L_y}\mathbb{1}_{x\in S}W(y|x)$. As $\frac{\varphi(|L'_y|)}{|L'_y|} = 1$, replacing L_y by L'_y does not decrease the objective value. This concludes the equivalence of the list-decoding problem with the φ-list-decoding with $\varphi(j) = \min\{j, \ell\}$.

When the channel is of the form $W(y|x) = \frac{1}{t}$ for $y \in T_x$ with $|T_x| = t$ and $W(y|x) = 0$ elsewhere, the success probability of the φ-list-decoding problem for a code $S \subseteq \mathcal{X}$ can be written as $\frac{1}{kt}\sum_{y\in\mathcal{Y}}\varphi(|S|_y)$, with $|S|_y = |\{x \in S : y \in T_x\}|$, which is a particular instance of the φ-MaxCoverage problem. Indeed, for a fixed code S, the best decoding strategy $\{L_y\}_{y\in\mathcal{Y}}$ leads to the following objective value:

$$\begin{aligned} &\max_{\{L_y\}_{y\in\mathcal{Y}}} \frac{1}{k}\sum_{x\in S}\sum_{y\in\mathcal{Y}}\frac{\varphi(|L_y|)}{|L_y|}\mathbb{1}_{x\in L_y}W(y|x) = \max_{\{L_y\}_{y\in\mathcal{Y}}} \frac{1}{k}\sum_{y\in\mathcal{Y}}\frac{\varphi(|L_y|)}{|L_y|}\sum_{x\in S}\mathbb{1}_{x\in L_y}W(y|x) \\ &= \frac{1}{kt}\sum_{y\in\mathcal{Y}}\max_{L_y}\frac{\varphi(|L_y|)}{|L_y|}\sum_{x\in S}\mathbb{1}_{x\in L_y}\mathbb{1}_{y\in T_x} = \frac{1}{kt}\sum_{y\in\mathcal{Y}}\max_{L_y}\frac{\varphi(|L_y|)}{|L_y|}|S_y \cap L_y| \; . \end{aligned} \tag{3.8}$$

We will show that $\max_{L_y}\frac{\varphi(|L_y|)}{|L_y|}|S_y \cap L_y| = \varphi(|S_y|)$, which will prove the claimed rewriting of the φ-list-decoding for those particular channels. As $\varphi(0) = 0$ and φ is concave, we have for all $0 \leq y$:

$$\frac{\varphi(y) - \varphi(0)}{y - 0} \leq \frac{\varphi(x) - \varphi(0)}{x - 0} \; ,$$

i.e. $\frac{x}{y}\varphi(y) \leq \varphi(x)$. Therfore, we have that:

$$\frac{\varphi(|L_y|)}{|L_y|}|S_y \cap L_y| \leq \varphi(|S_y \cap L_y|) \leq \varphi(|S_y|) \; ,$$

as φ is nondecreasing. Thus, $\max_{L_y}\frac{\varphi(|L_y|)}{|L_y|}|S_y \cap L_y| \leq \varphi(|S_y|)$, with the equality obtained by choosing $L_y := S_y$.

Therefore, for that class of channels, a tight approximation ratio α_φ follows from Theorem 3.1. As a consequence, the hardness part holds for the general φ-list-decoding problem, whereas finding an approximation algorithm achieving the ratio α_φ for all channels is left as an open question.

3.3.2 Multiwinner Elections

As mentioned previously, multiwinner elections (with a utilitarian model for the voters) entail selection of k (out of m) candidates that maximize the utility across n voters. Here, the utility of each voter $a \in [n]$ increases with the number of approved (by a) selections. The work of Dudycz et al. [DMMS20] study the computational complexity of such elections and, in particular, address classic voting rules in which—for a specified sequence of non-negative weights $(w_1, w_2, \ldots)$—voter a's utility is equal to $\sum_{i=1}^{j} w_i$, when she approves of j candidates among the selected ones. One can view this election exercise as a coverage problem by considering subset $T_i \subseteq [n]$ as the set of voters that approve of candidate $i \in [m]$ and $\varphi(j) = \sum_{i=1}^{j} w_i$. Indeed, for a subset of candidates $S \subseteq [m]$, the utility of a voter $a \in [n]$ is equal to $\varphi(|S|_a)$, with $|S|_a = |\{i \in S : a \in T_i\}|$.

Dudycz et al. [DMMS20] show that if the weights satisfy $w_1 \geq w_2 \geq \ldots$ (i.e., bear a diminishing returns property) along with geometric dominance ($w_i \cdot w_{i+2} \geq w_{i+1}^2$ for all $i \in \mathbb{N}^*$) and $\lim_{i\to\infty} w_i = 0$, then a tight approximation guarantee can be obtained for the election problem at hand. Note that the diminishing returns property implies that $\varphi(j) = \sum_{i=1}^{j} w_i$ is concave and $\lim_{i\to\infty} w_i = 0$ ensures that φ is sublinear (see Proposition 3.27). Hence, Theorem 3.1, together with Proposition 3.28, can be invoked to recover the result in [DMMS20] where we get $\alpha_\varphi = \alpha_\varphi(1)$. In fact, Theorem 3.1 does not require geometric dominance among the weights and, hence, applies to a broader class of voting rules. For instance, the geometric dominance property does not hold if one considers the voting weights induced by ℓ-MultiCoverage, i.e., $w_i = 1$, for $1 \leq i \leq \ell$, and $w_j = 0$ for $j > \ell$. However, using Theorem 3.1, we get that for this voting rule we can approximate the optimal utility within a factor of $\alpha_\varphi = 1 - \frac{\ell^\ell e^{-\ell}}{\ell!}$ (see Proposition 3.31). Another example of such a separation arises if one truncates the proportional approval voting. The standard proportional approval voting corresponds to $w_i = \frac{1}{i}$, for all $i \in \mathbb{N}$ (equivalently, $\varphi(j) = \sum_{i=1}^{j} \frac{1}{i}$) and falls within the purview of [DMMS20]. While the truncated version with $\varphi(j) = \sum_{i=1}^{\min\{j,\ell\}} \frac{1}{i}$, for a given threshold ℓ, does not satisfy geometric dominance, Theorem 3.1 continues to hold and provide a tight approximation ratio that can be computed numerically (see Proposition 3.21 and Table 3.1 for examples).

3.3.3 Resource Allocation in Multiagent Systems

A significant body of prior work in algorithmic game theory has addressed game-theoretic aspects of maximizing welfare among multiple (strategic) agents; see, e.g., [PM18]. Complementing such results, this section shows that the optimization problem underlying multiple welfare-maximization games can be expressed in terms of φ-MaxCoverage.

Specifically, consider a setting with n resources, k agents, and a (counting) function $\varphi : \mathbb{N} \mapsto \mathbb{R}_+$. Every agent i is endowed with a collection of resource subsets $\mathcal{A}_i = \{T_1^i, \ldots, T_{m_i}^i\} \subseteq 2^{[n]}$ (i.e., each $T_j^i \subseteq [n]$). The objective is to select a subset $A_i \in \mathcal{A}_i$, for all $i \in [k]$, so as to maximize $W^\varphi(A_1, A_2, \ldots, A_k) := \sum_{a\in[n]} w_a \, \varphi(|A|_a)$. Here, $w_a \in \mathbb{R}_+$ is a weight associated with $a \in [n]$ and $|A|_a := |\{i \in [k] : a \in A_i\}|$. We will refer to this problem as the φ-Resource Allocation problem.

While φ-Resource Allocation does not directly reduce to φ-MaxCoverage, the next theorem shows that it corresponds to maximizing φ-coverage functions subject to a matroid constraint. Hence, invoking our result from Section 3.1.1, we obtain a tight α_φ-approximation for φ-Resource Allocation (see Appendix 3.5.5 for the proof):

Theorem 3.15. *For any normalized nondecreasing concave function φ, there exists a α_φ-approximation algorithm for φ-Resource Allocation runninng in polynomial time. Furthermore, for $\varphi(n) = o(n)$, it is NP-hard to approximate φ-Resource Allocation within a factor better than $\alpha_\varphi + \varepsilon$, for any constant $\varepsilon > 0$.*

3.3.4 Vehicle-Target Assignment

Vehicle-Target Assignment [Mur00, PM18] is another problem which highlights the applicability of coverage problems, with a concave φ. In particular, Vehicle-Target Assignment can be directly expressed as φ-Resource Allocation: the $[n]$ resources correspond to targets, the agents correspond to vehicles $i \in [k]$, each with a collection of covering choices $\mathcal{A}_i \subseteq 2^{[n]}$, and $\varphi^p(j) = \frac{1-(1-p)^j}{p}$, for a given parameter $p \in (0,1)$. As limit cases, we define $\varphi^0(j) := \lim_{p\to 0} \varphi^p(j) = j$ and $\varphi^1(j) := 1$. Since $\varphi^p(j)$ is concave, by Proposition 3.32 and Theorem 3.15, we obtain a novel tight approximation ratio of $\alpha_{\varphi^p} = \frac{1-e^{-p}}{p}$ for this problem. Also, one can look at the capped version of this problem, $\varphi_\ell^p(j) := \varphi^p(\min\{j,\ell\})$. In particular, we recover the ℓ-MultiCoverage function when $p = 0$. In Figure 3.1, we have plotted several cases of the tight approximations $\alpha_{\varphi_\ell^p}$ in function of ℓ for several values of ℓ:

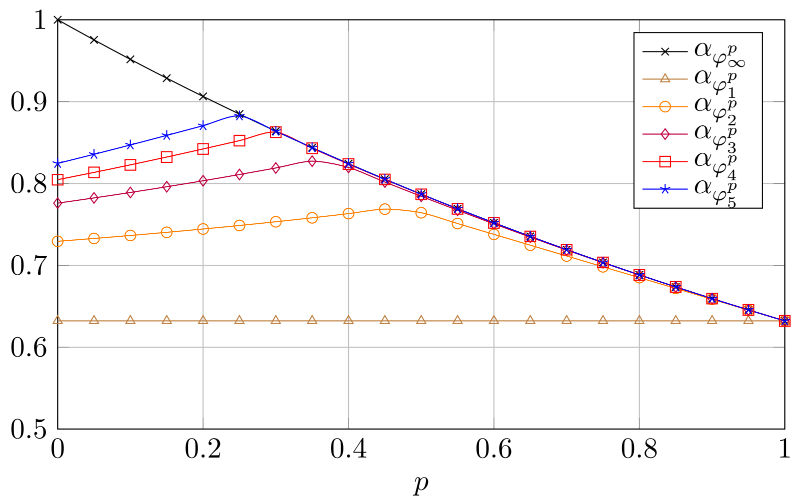


Figure 3.1 – Tight approximation ratios $\alpha_{\varphi_\ell^p}$, where ℓ is the rank of the capped version of the p-Vehicle-Target Assignment problem. When $p = 0$, we recover the ℓ-coverage problem.

Paccagnan and Marden [PM18] study the game-theoretic aspects of Vehicle-target Assignment. A key goal in [PM18] is to bound the welfare loss incurred due to strategic selection by the k vehicles, i.e., the selection of each $A_i \in \mathcal{A}_i$ by a self-interested vehicle/agent $i \in [k]$. The loss is quantified in terms of the *Price of Anarchy* (PoA). Formally, this performance metric is defined as ratio between the welfare of the worst-possible equilibria and the optimal welfare. Paccagnan and Marden [PM18] show that, for computationally tractable equilibrium concepts (in particular, for coarse correlated equilibria), tight price of anarchy bounds can be obtained via linear programs.

Note that our hardness result (Theorem 3.1) provides upper bounds on PoA of tractable equilibrium concepts–this follows from the observation that computing an equilibrium provides a specific method for finding a coverage solution. In [CPM19] and in the particular case of the ℓ-MultiCoverage problem, it is shown that this in fact an equality, i.e., PoA $= \alpha_\varphi$ if $\varphi(j) = \min\{j, \ell\}$ for all values of ℓ. However, numerically comparing the approximation ratio for Vehicle-Target Assignment, $\alpha_{\varphi^p} = \frac{1-e^{-p}}{p}$, with the optimal PoA bound, we note that α_{φ^p} can in fact be strictly greater than the PoA guarantee; see Figure 3.2.

Another form of the current problem, considered in [PM18], corresponds to $\varphi^d(j) = j^d$, for a given parameter $d \in (0,1)$. We refer to this instantiation as the d-Power function and for it obtain the approximation ratio $\alpha_{\varphi^d} = e^{-1}\sum_{k=1}^{+\infty}\frac{k^d}{k!}$ (Proposition 3.33). In this case, the question whether the inequality PoA $\leq \alpha_\varphi$ is tight remains open; see Figure 3.3.

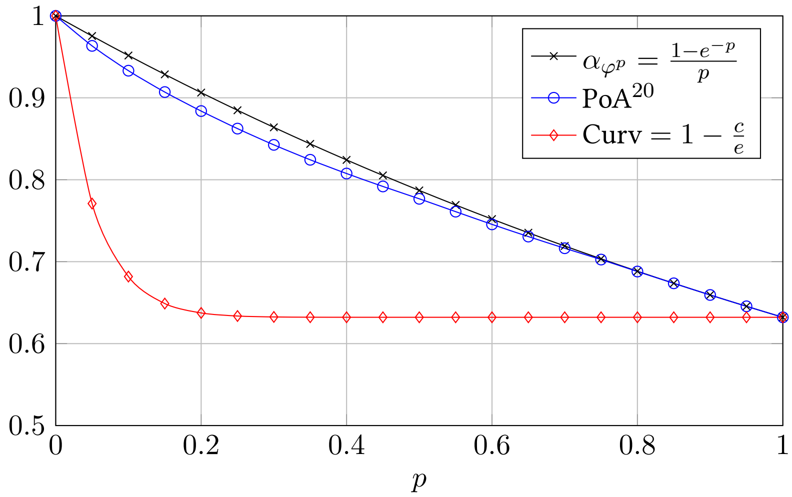


Figure 3.2 – Comparison between the PoA and α_φ for the Vehicle-Target Assignment problem. Using the linear program found in [PM18], we were able to compute the blue curve PoA^{20}, the *Price of Anarchy* of this problem for $m = 20$ players. Since the PoA only decreases when the number of players grows, this means that PoA $< \alpha_\varphi$ in that case. As a comparison, the red curve Curv depicts the general approximation ratio (see [SVW17]) obtained for submodular function with curvature c, with $c = 1 - \varphi^p(m) + \varphi^p(m-1)$ here.

3.3.5 Welfare Maximization for φ-Coverage

Maximizing (social) welfare by partitioning items among agents is a key problem in algorithmic game theory; see, e.g., the extensive work on combinatorial auctions [NRTV07]. The goal here is to partition t items among a set of k agents such that the sum of values achieved by the agents—referred to as the social welfare—is maximized. That is, one needs to partition $[t]$ into k pairwise disjoint subsets $A_1, A_2, \ldots, A_k$ with the objective of maximizing $\sum_{i=1}^k v_i(A_i)$. Here, $v_i(S)$ denotes the valuation that agent i has for a subset of items $S \subseteq [t]$.

When each agent's valuation v_i is submodular, a tight $(1 - e^{-1})$-approximation ratio is known for social welfare maximization [Von07]. This section shows that improved

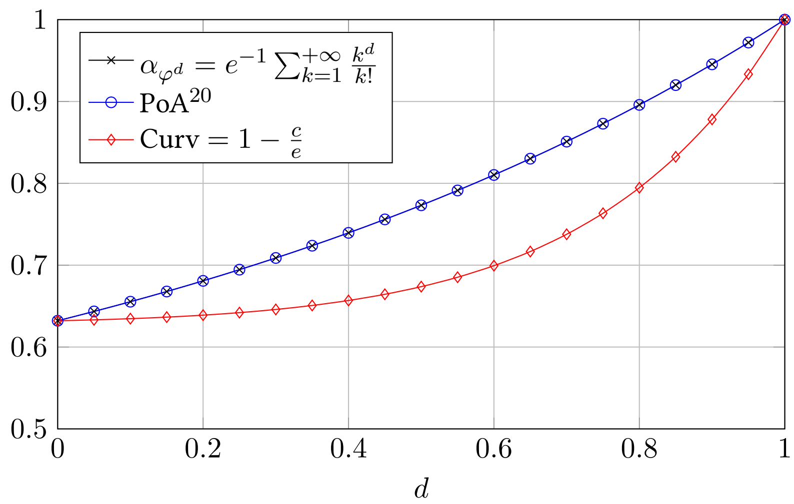


Figure 3.3 – Comparison between the PoA and α_φ for the d-Power problem. Using the linear program found in [PM18], we were able to compute the blue curve PoA^{20}, the *Price of Anarchy* of this problem for $m = 20$ players. Here, the question whether the inequality PoA $\leq \alpha_\varphi$ is tight remains open. As a comparison, the red curve Curv depicts the general approximation ratio (see [SVW17]) obtained for submodular function with curvature c, with $c = 1 - \varphi^d(m) + \varphi^d(m-1)$ here.

approximation guarantees can be achieved if, in particular, the agents' valuations are φ-coverage functions. Towards a stylized application of such valuations, consider a setting in which each "item" $b \in [t]$ represents a bundle (subset) of goods $T_b \subseteq [n]$ and the value of an agent increases with the number of copies of any good $a \in [n]$ that get accumulated. Indeed, if each agent's value for j copies of a good is $\varphi(j)$, then we have a φ-coverage function and the overall optimization problem is find a k-partition, $A_1, A_2, \dots, A_k$, of $[t]$ that maximizes $\sum_{i=1}^{k} \left(\sum_{a \in [n]} \varphi \left(|A_i|_a \right) \right)$, where $|A_i|_a := \{b \in A_i : a \in T_b\}$.

In the current setup, one can obtain an α_φ approximation ratio for social-welfare maximization by reducing this problem to φ-coverage with a matroid constraint, and applying the result from Section 3.1.1. Specifically, we can consider a partition matroid over the universe $[t] \times [k]$: for a bundle/item $b \in [t]$ and an agent $i \in [k]$, the element (b, i) in the universe represents that bundle b is assigned to agent i, i.e., $b \in A_i$. The partition-matroid constraint is imposed to ensure that each bundle b is assigned to at most one agent. Furthermore, we can create k copies of the underlying set of goods $[n]$ and set $T_{(b,i)} := \{(a, i) : a \in T_b\}$ to map the φ-coverage over the universe to the social-welfare objective. This, overall, gives us the desired α_φ approximation guarantee.

3.4 Conclusion

We have introduced the φ-MaxCoverage problem where having c copies of element a gives a value $\varphi(c)$. We have shown that when φ is normalized, nondecreasing and concave, we can obtain an approximation guarantee given by the *Poisson concavity ratio* $\alpha_\varphi := \min_{x \in \mathbb{N}} \frac{\mathbb{E}[\varphi(\mathrm{Poi}(x))]}{\varphi(\mathbb{E}[\mathrm{Poi}(x)])}$ and we showed it is tight for sublinear functions φ. The Poisson concavity ratio strictly beats the bound one gets when using the notion of curvature submodular functions, except in very special cases such as MaxCoverage where the two bounds are equal.

An interesting open question is whether there exists combinatorial algorithms that achieve this approximation ratio. As mentioned in [BFGG20], for the ℓ-MultiCoverage with $\ell \geq 2$, which is the special case where $\varphi(x) = \min\{x, \ell\}$, the simple greedy algorithm only gives a $1 - e^{-1}$ approximation ratio, which is strictly less than the ratio $\alpha_\varphi = 1 - \frac{\ell^\ell e^{-\ell}}{\ell!}$ in that case. Also, for any geometrically dominant vector $w = (\varphi(i+1) - \varphi(i))_{i \in \mathbb{N}}$ which is not p-geometric, such as Proportional Approval Voting, the greedy algorithm achieves an approximation ratio which is strictly less than α_φ (see Theorem 18 of [DMMS20]).

Another open question is whether the hardness result remains true even when $\varphi(n) \neq o(n)$. A good example is given by $\varphi(0) = 0$ and $\varphi(1+t) = 1 + (1-c)t$ with $c \in (0, 1)$. We know that the problem is hard for $c = 1$ but easy for $c = 0$. One can show that the approximation ratio achieved by our algorithm is $\alpha_\varphi = 1 - \frac{c}{e}$ in that case (which is the same approximation ratio obtained from the curvature in [SVW17]), but the tightness of this approximation ratio remains open.

3.5 Appendix

3.5.1 General Properties

In this section, we will assume that φ is specified over the nonnegative integers (i.e., $\varphi : \mathbb{N} \to \mathbb{R}_+$) and it is nondecreasing, concave, and normalized: $\varphi(0) = 0$ and $\varphi(1) = 1$. We will consider its piecewise linear extension on $\mathbb{R}_+$ by defining $\varphi(x) := \lambda\varphi(\lfloor x \rfloor)+(1-\lambda)\varphi(\lceil x \rceil)$; here, parameter $\lambda \in [0, 1]$ satisfies $x = \lambda\lfloor x \rfloor + (1 - \lambda)\lceil x \rceil$. Note that the piecewise linear extension is also nondecreasing and concave.

Proposition 3.16. *For all $x \in \mathbb{R}_+$, we have $\alpha_\varphi(x) \geq \min\{\alpha_\varphi(\lfloor x \rfloor), \alpha_\varphi(\lceil x \rceil)\}$; here, $\alpha_\varphi(0) := \lim_{x\to 0} \alpha_\varphi(x) = 1$.*

Proof. For any $x \geq 1$, consider parameter $\lambda \in [0, 1]$ such that $x = \lambda\lfloor x \rfloor + (1 - \lambda)\lceil x \rceil$. Since $x \mapsto \mathbb{E}[\varphi(\mathrm{Poi}(x))]$ is concave (Proposition 3.25), the following bound holds for all $x \geq 1$:

$$\begin{aligned}
\mathbb{E}[\varphi(\mathrm{Poi}(x))] &\geq \lambda\mathbb{E}[\varphi(\mathrm{Poi}(\lfloor x \rfloor))] + (1-\lambda)\mathbb{E}[\varphi(\mathrm{Poi}(\lceil x \rceil))] \\
&= \lambda\alpha_\varphi(\lfloor x \rfloor)\varphi(\lfloor x \rfloor) + (1-\lambda)\alpha_\varphi(\lceil x \rceil)\varphi(\lceil x \rceil) \quad \text{by definition of } \alpha_\varphi(x) \\
&\geq \min\{\alpha_\varphi(\lfloor x \rfloor), \alpha_\varphi(\lceil x \rceil)\} \left(\lambda\varphi(\lfloor x \rfloor) + (1-\lambda)\varphi(\lceil x \rceil)\right) \\
&= \min\{\alpha_\varphi(\lfloor x \rfloor), \alpha_\varphi(\lceil x \rceil)\}\varphi(x) \quad \text{since } \varphi \text{ linear between integer points.}
\end{aligned}$$

Therefore, $\alpha_\varphi(x) = \frac{\mathbb{E}[\varphi(\mathrm{Poi}(x))]}{\varphi(x)} \geq \min\{\alpha_\varphi(\lfloor x \rfloor), \alpha_\varphi(\lceil x \rceil)\}$.

Next we will show that $\alpha_\varphi(x)$ is non-increasing from 0 to 1, which implies that for $x \in [0, 1)$, we have $\alpha_\varphi(x) \geq \min\{\alpha_\varphi(\lfloor x \rfloor), \alpha_\varphi(\lceil x \rceil)\}$. Recall that φ, by definition, is linear between integers. Hence, the fact that $\varphi(0) = 0$ and $\varphi(1) = 1$, gives us $\varphi(x) = x$ for all $x \in [0, 1]$. Therefore,

$$\alpha_\varphi(x) = \frac{\mathbb{E}[\varphi(\mathrm{Poi}(x))]}{x} = e^{-x} \sum_{k=1}^{+\infty} \frac{\varphi(k)}{k} \frac{x^{k-1}}{(k-1)!} = e^{-x} \sum_{k=0}^{+\infty} \frac{\varphi(k+1)}{k+1} \frac{x^k}{k!} .$$

In particular, $\alpha_\varphi(x)$ is well-defined at 0 and $\alpha_\varphi(0) = e^{-0} \sum_{k=0}^{+\infty} \frac{\varphi(k+1)}{k+1} \frac{0^k}{k!} = 1$. Now, consider the derivative:

$$\begin{aligned}
\alpha'_\varphi(x) &= e^{-x} \left(-\sum_{k=0}^{+\infty} \frac{\varphi(k+1)}{k+1} \frac{x^k}{k!} + \sum_{k=1}^{+\infty} \frac{\varphi(k+1)}{k+1} \frac{x^{k-1}}{(k-1)!} \right) \\
&= e^{-x} \sum_{k=0}^{+\infty} \left(\frac{\varphi(k+2)}{k+2} - \frac{\varphi(k+1)}{k+1} \right) \frac{x^k}{k!} .
\end{aligned} \tag{3.9}$$

Note that $\frac{\varphi(k+2)}{k+2} - \frac{\varphi(k+1)}{k+1} = \frac{\varphi(k+2)-\varphi(0)}{(k+2)-0} - \frac{\varphi(k+1)-\varphi(0)}{(k+1)-0} \leq 0$; the last inequality follows from the concavity of φ. Hence, $\alpha'_\varphi(x) \leq 0$. That is, $\alpha_\varphi(x)$ is non-increasing from 0 to 1. □

Proposition 3.17. *For any $\varepsilon > 0$, the bound $1 - \alpha_\varphi(x) \leq \varepsilon$ holds for all $x \geq \left(\frac{6}{\varepsilon}\right)^4$.*

Proof. Write $X \sim \mathrm{Poi}(x)$ and note that $\mathbb{P}(X \le x(1-\delta(x))) \le \exp\left(-\frac{x\delta(x)^2}{2(1+\delta(x))}\right)$, for any positive function $\delta(\cdot)$ which satisfies $\delta(x) < 1$, for all $x > 1$; see, e.g., [Can17]. Therefore,

$$\begin{aligned}
\mathbb{E}[\varphi(X)] \ge e^{-x} &\sum_{k=\lceil x(1-\delta(x))\rceil}^{+\infty} \varphi(k)\frac{x^k}{k!} \quad \text{since } \varphi \text{ nonnegative} \\
&\ge \varphi(x(1-\delta(x))) \sum_{k=\lceil x(1-\delta(x))\rceil}^{+\infty} e^{-x}\frac{x^k}{k!} \quad \text{since } \varphi \text{ nondecreasing} \\
&\ge \varphi(x(1-\delta(x)))(1-\mathbb{P}(X \le x(1-\delta(x)))) \\
&\ge \varphi(x(1-\delta(x)))\left(1-\exp\left(-\frac{x\delta(x)^2}{2(1+\delta(x))}\right)\right) .
\end{aligned} \tag{3.10}$$

Next, we will show that $\frac{\varphi(x(1-\delta(x)))}{\varphi(x)} \ge 1 - \frac{\delta(x)+\frac{1}{x}}{1-\delta(x)}$. Towards this end, we will first bound $\varphi(x+y)-\varphi(x)$ in terms of $w_k^x = \varphi(x+k)-\varphi(x+k-1)$, which constitutes a non-increasing sequence (since φ is concave):

$$\varphi(x+y)-\varphi(x) \le \varphi(x+\lfloor y\rfloor+1)-\varphi(x+\lfloor y\rfloor) + \sum_{k=1}^{\lfloor y\rfloor} w_k^x \le (\lfloor y\rfloor+1)w_1^x.$$

Applying this bound to $x(1-\delta(x))$ and $x\delta(x)$ gives us

$$\begin{aligned}
1-\frac{\varphi(x(1-\delta(x)))}{\varphi(x)} = \frac{\varphi(x)-\varphi(x(1-\delta(x)))}{\varphi(x)} &\le \frac{(\lfloor x\delta(x)\rfloor+1)w_1^{x(1-\delta(x))}}{\varphi(x)} \\
&\le \frac{x\delta(x)+1}{\varphi(x)}\,\frac{\varphi(x(1-\delta(x)))}{x(1-\delta(x))} \le \frac{x\delta(x)+1}{x(1-\delta(x))} = \frac{\delta(x)+\frac{1}{x}}{1-\delta(x)}.
\end{aligned} \tag{3.11}$$

Here, $w_1^{x(1-\delta(x))} = \frac{\varphi(x(1-\delta(x))+1)-\varphi(x(1-\delta(x)))}{(x(1-\delta(x))+1)-(x(1-\delta(x)))} \le \frac{\varphi(x(1-\delta(x)))-\varphi(0)}{x(1-\delta(x))-0} = \frac{\varphi(x(1-\delta(x)))}{x(1-\delta(x))}$ follows from the concavity of φ and $\frac{\varphi(x(1-\delta(x)))}{\varphi(x)} \le 1$ from the fact that φ is nondecreasing.

Inequalities (3.10) and (3.11) lead to following upper bound on $1-\alpha_\varphi(x)$ in terms of $\delta(x)$:

$$\begin{aligned}
1-\alpha_\varphi(x) = 1-\frac{\mathbb{E}[\varphi(\mathrm{Poi}(x))]}{\varphi(x)} &\le 1-\left(1-\frac{\delta(x)+\frac{1}{x}}{1-\delta(x)}\right)\left(1-\exp\left(-\frac{x\delta(x)^2}{2(1+\delta(x))}\right)\right) \\
&\le \frac{\delta(x)+\frac{1}{x}}{1-\delta(x)} + \exp\left(-\frac{x\delta(x)^2}{2(1+\delta(x))}\right).
\end{aligned} \tag{3.12}$$

Specifically setting $\delta(x) = x^{-\frac{1}{4}}$, we have (for all $x \ge 16$): $\delta(x) \le \frac{1}{2}$, $\frac{1}{x} \le x^{-\frac{1}{4}}$, and $\exp\left(-\frac{x\delta(x)^2}{2(1+\delta(x))}\right) \le \exp\left(-\frac{\sqrt{x}}{4}\right) \le 2x^{-\frac{1}{4}}$. Hence, inequality (3.12) reduces to

$$1-\alpha_\varphi(x) \le \frac{2x^{-\frac{1}{4}}}{1-\frac{1}{2}} + 2x^{-\frac{1}{4}} \le 6x^{-\frac{1}{4}} \qquad \text{for all } x \ge 16.$$

If $\varepsilon \ge 1$, we have $1-\alpha_\varphi(x) \le 1 \le \varepsilon$. Otherwise, we have that $\left(\frac{6}{\varepsilon}\right)^4 \ge 6^4 \ge 16$. Therefore, given any $\varepsilon > 0$, for all $x \ge \left(\frac{6}{\varepsilon}\right)^4$ we have $1-\alpha_\varphi(x) \le \varepsilon$. □

Proposition 3.18. *We have that* $\alpha_\varphi = \inf_{x\in\mathbb{R}_+} \alpha_\varphi(x) = \min_{x\in\mathbb{N}^*} \alpha_\varphi(x)$.

Proof. Thanks to Proposition 3.16, we have that $\inf_{x\in\mathbb{R}^+} \alpha_\varphi(x) = \inf_{x\in\mathbb{N}^*} \alpha_\varphi(x)$, and thanks to Proposition 3.17, since $\alpha_\varphi(x) \leq 1$, we have that $\inf_{x\in\mathbb{N}^*} \alpha_\varphi(x) = \min_{x\in\mathbb{N}^*} \alpha_\varphi(x)$. □

Proposition 3.19. C^φ *is submodular, its curvature is at most* $c = 1 - (\varphi(m) - \varphi(m-1))$ *and it cannot be improved for a general instance with* m *cover sets.*

Proof. We use the following lemma which is trivial to prove:

Lemma 3.20 (Properties of $|S|_a = |\{i \in S : a \in T_i\}|$.). *We have:*

1. $|S|_a \leq |S|$,
2. $|S \cup S'|_a \leq |S|_a + |S'|_a$. *In particular, if* $S \subseteq T$ *then* $|S|_a \leq |T|_a$ *and* $|S \cup \{x\}|_a \leq |S|_a + 1$,
3. *If* $S \subseteq T$, $x \notin T$ *then* $|S|_a = |T|_a \Rightarrow |S \cup \{x\}|_a = |T \cup \{x\}|_a$.

Let us show first the submodularity of C^φ. Let $S \subseteq T \subseteq [m]$ and $x \notin T$:

$$\begin{aligned} & C^\varphi(S \cup \{x\}) - C^\varphi(S) - (C^\varphi(T \cup \{x\}) - C^\varphi(T)) = \\ = & \sum_{a\in[n]} w_a[\varphi(|S \cup \{x\}|_a) - \varphi(|S|_a) - (\varphi(|T \cup \{x\}|_a) - \varphi(|T|_a))] \,. \end{aligned} \tag{3.13}$$

Let us call $g(a) := \varphi(|S \cup \{x\}|_a) - \varphi(|S|_a) - (\varphi(|T \cup \{x\}|_a) - \varphi(|T|_a))$:

1. If $|T|_a = |S|_a$ then thanks to Lemma 3.20, we have that $|T \cup \{x\}|_a = |S \cup \{x\}|_a$, so $g(a) = 0$
2. Else, we have that $|T|_a > |S|_a$:
   a) If $|S \cup \{x\}|_a = |S|_a$, then we add elements of $T - S$ using Lemma 3.20 to get that $|T \cup \{x\}|_a = |T|_a$, so $g(a) = 0$ in that case.
   b) Else $|S \cup \{x\}|_a \neq |S|_a$. So with $|S|_a = k$, we get that $|S \cup \{x\}|_a = k + 1$ and $|T|_a > |S|_a$ so $|T|_a \geq k + 1$.
      i. If $|T \cup \{x\}|_a = |T|_a$, then $g(a) = \varphi(k + 1) - \varphi(k) \geq 0$ since φ is nondecreasing.
      ii. Else $|T \cup \{x\}|_a \neq |T|_a$ so with $|T|_a = \ell$ with $\ell \geq k + 1$, we get that $|T|_a = \ell + 1$. So we have that:

$$\begin{aligned} g(a) = & \ \varphi(k+1) - \varphi(k) - (\varphi(\ell+1) - \varphi(\ell)) \\ = & \ \frac{\varphi(k+1) - \varphi(k)}{(k+1) - k} - \frac{\varphi(\ell+1) - \varphi(\ell)}{(\ell+1) - \ell} \geq 0 \,, \end{aligned} \tag{3.14}$$

by concavity of φ: its slopes are nonincreasing.

So in all cases, we have $g(a) \geq 0$ so $C^\varphi(S \cup \{x\}) - C^\varphi(S) - (C^\varphi(T \cup \{x\}) - C^\varphi(T)) \geq 0$: C^φ is submodular.

Let us now compute its curvature:

$$c = 1 - \min_{i \in [m]} \frac{C^\varphi([m]) - C^\varphi([m] - \{i\})}{C^\varphi(\{i\}) - C^\varphi(\emptyset)} \; .$$

Let $i \in [m]$ fixed:

$$\begin{aligned} \frac{C^\varphi([m]) - C^\varphi([m] - \{i\})}{C^\varphi(\{i\}) - C^\varphi(\emptyset)} &= \frac{\sum_{a \in [n]} w_a[\varphi(|[m]|_a) - \varphi(|[m] - \{i\}|_a)]}{\sum_{a \in [n]} w_a[\varphi(|\{i\}|_a) - \varphi(|\emptyset|_a)]} \\ &= \frac{\sum_{a \in T_i} w_a[\varphi(|[m]|_a) - \varphi(|[m] - \{i\}|_a)]}{\sum_{a \in T_i} w_a} \\ &= \frac{\sum_{a \in T_i} w_a[\varphi(|[m]|_a) - \varphi(|[m]|_a - 1)]}{\sum_{a \in T_i} w_a} \; , \end{aligned} \tag{3.15}$$

since $a \in T_i$. But $|[m]|_a \leq m$ and φ concave, so $\varphi(|[m]|_a)) - \varphi(|[m]|_a - 1) \geq \varphi(m) - \varphi(m-1)$ for all $a \in [n]$. As a consequence we have that:

$$\frac{C^\varphi([m]) - C^\varphi([m] - \{i\})}{C^\varphi(\{i\}) - C^\varphi(\emptyset)} \geq \varphi(m) - \varphi(m-1) \; .$$

and this lower bound is true for its minimum over i. Thus we get that $c \leq 1 - (\varphi(m) - \varphi(m-1))$. Also one can find instances for all m such that this bound is tight: take $T_1 = \{a\}$ and $\forall j \in [m], a \in T_j$ for instance. □

Proposition 3.21. *Let $\ell \in \mathbb{N}^*$. if $\forall x \geq \ell, \varphi(x) = \varphi(\ell) + a(x - \ell)$ for some $0 \leq a \leq \varphi(\ell) - \varphi(\ell - 1)$, then $\alpha_\varphi(x)$ is nondecreasing from ℓ to $+\infty$ and:*

$$\alpha_\varphi(x) = \frac{\varphi(\ell) + a(x - \ell)}{\varphi(x)} - \frac{e^{-x}}{\varphi(x)} \left(\sum_{k=0}^{\ell} (\varphi(\ell) + a(x - \ell) - \varphi(k)) \frac{x^k}{k!} - a \frac{x^{\ell+1}}{\ell!} \right) \; .$$

In particular, $\alpha_\varphi = \min_{x \in [\ell]} \alpha_\varphi(x)$, and the argmin can be computed numerically.

Proof. One can compute a closed form value for $\alpha_\varphi(x)$ using the fact that φ is linear from ℓ:

$$\begin{aligned} \alpha_\varphi(x) = \frac{e^{-x}}{\varphi(x)} \sum_{k=0}^{+\infty} \varphi(k) \frac{x^k}{k!} &= \frac{e^{-x}}{\varphi(x)} \left(\sum_{k=0}^{\ell} \varphi(k) \frac{x^k}{k!} + \sum_{k=\ell+1}^{+\infty} (\varphi(\ell) + a(k - \ell)) \frac{x^k}{k!} \right) \\ &= \frac{\varphi(\ell) - a\ell}{\varphi(x)} + \frac{e^{-x}}{\varphi(x)} \left(\sum_{k=0}^{\ell} (\varphi(k) - \varphi(\ell) + a\ell) \frac{x^k}{k!} + ax \sum_{k=\ell+1}^{+\infty} \frac{x^{k-1}}{(k-1)!} \right) \\ &= \frac{\varphi(\ell) + a(x - \ell)}{\varphi(x)} + \frac{e^{-x}}{\varphi(x)} \left(\sum_{k=0}^{\ell} (\varphi(k) - \varphi(\ell) + a(\ell - x)) \frac{x^k}{k!} + ax \frac{x^\ell}{\ell!} \right) \; , \end{aligned} \tag{3.16}$$

and thus we get:

$$\alpha_\varphi(x) = \frac{\varphi(\ell) + a(x - \ell)}{\varphi(x)} - \frac{e^{-x}}{\varphi(x)} \left(\sum_{k=0}^{\ell} (\varphi(\ell) + a(x - \ell) - \varphi(k)) \frac{x^k}{k!} - a \frac{x^{\ell+1}}{\ell!} \right) \; .$$

Let us show that it is nondecreasing from ℓ to $+\infty$ by computing its derivative. Indeed, for $x \geq \ell$, we have that $\varphi(x) = \varphi(\ell) + a(x-\ell)$ and $\varphi'(x) = a$, so:

$$\alpha_\varphi(x) = 1 - e^{-x}\left(\sum_{k=0}^{\ell}\frac{x^k}{k!} - \frac{1}{\varphi(x)}\left(\sum_{k=0}^{\ell}\varphi(k)\frac{x^k}{k!} + a\frac{x^{\ell+1}}{\ell!}\right)\right) .$$

Thus:

$$\begin{aligned}
\alpha'_\varphi(x) &= e^{-x}\left(\sum_{k=0}^{\ell}\frac{x^k}{k!} - \frac{1}{\varphi(x)}\left(\sum_{k=0}^{\ell}\varphi(k)\frac{x^k}{k!} + a\frac{x^{\ell+1}}{\ell!}\right)\right)\\
&- e^{-x}\left(\sum_{k=0}^{\ell-1}\frac{x^k}{k!} - \frac{1}{\varphi(x)}\left(\sum_{k=0}^{\ell-1}\varphi(k+1)\frac{x^k}{k!} + a(\ell+1)\frac{x^{\ell}}{\ell!}\right)\right)\\
&+ e^{-x}\frac{a}{\varphi(x)^2}\left(\sum_{k=0}^{\ell}\varphi(k)\frac{x^k}{k!} + a\frac{x^{\ell+1}}{\ell!}\right)\\
&= e^{-x}\left(\frac{x^\ell}{\ell!} - \frac{1}{\varphi(x)}\left(\sum_{k=0}^{\ell-1}(\varphi(k)-\varphi(k+1))\frac{x^k}{k!} + \varphi(\ell)\frac{x^\ell}{\ell!} + (a(x-\ell)-a)\frac{x^\ell}{\ell!}\right)\right)\\
&- e^{-x}\left(\frac{a}{\varphi(x)^2}\left(\sum_{k=0}^{\ell}\varphi(k)\frac{x^k}{k!} + a\frac{x^{\ell+1}}{\ell!}\right)\right)\\
&= e^{-x}\left(\frac{x^\ell}{\ell!} + \frac{1}{\varphi(x)}\left(\sum_{k=0}^{\ell-1}(\varphi(k+1)-\varphi(k))\frac{x^k}{k!}\right) - \frac{\varphi(x)-a}{\varphi(x)}\frac{x^\ell}{\ell!}\right)\\
&- e^{-x}\frac{a}{\varphi(x)^2}\left(\sum_{k=0}^{\ell}\varphi(k)\frac{x^k}{k!} + a\frac{x^{\ell+1}}{\ell!}\right)\\
&= e^{-x}\left(\frac{1}{\varphi(x)}\left(\sum_{k=0}^{\ell-1}(\varphi(k+1)-\varphi(k))\frac{x^k}{k!}\right)\right)\\
&+ e^{-x}\frac{a}{\varphi(x)^2}\left(\varphi(x)\frac{x^\ell}{\ell!} - \sum_{k=0}^{\ell}\varphi(k)\frac{x^k}{k!} - a\frac{x^{\ell+1}}{\ell!}\right) .
\end{aligned} \tag{3.17}$$

If $a = 0$, then it is nonnegative since φ nondecreasing and nonnegative. Otherwise, suppose that $a > 0$. Then:

$$\begin{aligned}
\alpha_\varphi(x) &= \frac{ae^{-x}}{\varphi(x)^2}\left(\sum_{k=0}^{\ell-1}\left(\varphi(x)\frac{\varphi(k+1)-\varphi(k)}{a} - \varphi(k)\right)\frac{x^k}{k!} + (\varphi(x)-\varphi(\ell)-ax)\frac{x^\ell}{\ell!}\right)\\
&\geq \frac{ae^{-x}}{\varphi(x)^2}\left(\sum_{k=0}^{\ell-1}(\varphi(x)-\varphi(k))\frac{x^k}{k!} - a\frac{x^\ell}{(\ell-1)!}\right) ,
\end{aligned} \tag{3.18}$$

since $\frac{\varphi(k+1)-\varphi(k)}{a} \geq \frac{\varphi(k+1)-\varphi(k)}{\varphi(\ell)-\varphi(\ell-1)} \geq 1$ by concavity of φ. Thus:

$$\alpha_\varphi(x) \geq \frac{ae^{-x}}{\varphi(x)^2}\left(\sum_{k=0}^{\ell-1}(\varphi(\ell)-\varphi(k))\frac{x^k}{k!} + a\left((x-\ell)\sum_{k=0}^{\ell-1}\frac{x^k}{k!} - \frac{x^\ell}{(\ell-1)!}\right)\right) ,$$

but:

$$(x-\ell)\sum_{k=0}^{\ell-1}\frac{x^k}{k!} - \frac{x^\ell}{(\ell-1)!} = \ell\sum_{k=1}^{\ell}\frac{x^k}{k!} - \ell\sum_{k=0}^{\ell-1}\frac{x^k}{k!} - \frac{x^\ell}{(\ell-1)!} = -\ell ,$$

so:

$$\begin{aligned}
\alpha'_\varphi(x) &\geq \frac{ae^{-x}}{\varphi(x)^2}\left(\sum_{k=0}^{\ell-1}(\varphi(\ell)-\varphi(k))\frac{x^k}{k!} - a\ell\right)\\
&\geq \frac{ae^{-x}}{\varphi(x)^2}\left(\sum_{k=0}^{\ell-1}(\varphi(\ell)-\varphi(k))\frac{x^k}{k!} - (\varphi(\ell)-\varphi(\ell-1))\ell\right)\\
&\geq \frac{ae^{-x}}{\varphi(x)^2}\left(\sum_{k=0}^{\ell-1}(\varphi(\ell)-\varphi(k))\frac{x^k}{k!} - \sum_{k=0}^{\ell-1}(\varphi(\ell)-\varphi(\ell-1))\frac{x^k}{k!}\right) \qquad (3.19)\\
&\quad \text{since } \frac{x^k}{k!} \geq \frac{\ell^k}{k!} \geq 1 \text{ for } k \leq \ell\\
&= \frac{ae^{-x}}{\varphi(x)^2}\left(\sum_{k=0}^{\ell-1}(\varphi(\ell-1)-\varphi(k))\frac{x^k}{k!}\right) \geq 0 \quad \text{since } \varphi \text{ nondecreasing}\,.
\end{aligned}$$

Thus, $\alpha_\varphi(x)$ is nondecreasing from ℓ to $+\infty$, and we get that $\alpha_\varphi = \min_{x\in[\ell]} \alpha_\varphi(x)$. □

Proposition 3.22. *The Poisson concavity ratio α_φ is always greater than or equal to the curvature-dependent ratio defined in [SVW17]: if φ is linear from m with slope $1-c = \varphi(m)-\varphi(m-1)$, then we have $\alpha_\varphi \geq 1-ce^{-1}$.*

Proof. Note that by Proposition 3.19, the curvature of C^φ is equal to c, so the efficiency of the algorithm described in [SVW17] is indeed $1-ce^{-1}$. Thanks to Proposition 3.21, we have that $\alpha_\varphi = \min_{x\in[\ell]} \alpha_\varphi(x)$, so we only have to show that:

$$\min_{\ell\in[m]} \alpha_\varphi(\ell) \geq 1-ce^{-1}\,.$$

Let us denote by $\varphi^{(\ell,a)}$ the function which is equal to φ for $k \leq \ell$ and linear from ℓ with nonnegative coefficient a: $\forall k \geq \ell, \varphi^{(\ell,a)}(k) = \varphi(\ell) + a(k-\ell)$. Note that we ask that $0 \leq a \leq \varphi(\ell)-\varphi(\ell-1)$ in order to $\varphi^{(\ell,a)}$ to be nondecreasing concave, and $\ell \geq 1$.

This is done in two steps:

1. Let $1 \leq \ell \leq m$, then:

   $$\alpha_\varphi(\ell) = \frac{\mathbb{E}[\varphi(\mathrm{Poi}(\ell))]}{\varphi(\ell)} = \frac{\mathbb{E}[\varphi(\mathrm{Poi}(\ell))]}{\varphi^{(\ell,1-c)}(\ell)} \geq \frac{\mathbb{E}[\varphi^{(\ell,1-c)}(\mathrm{Poi}(\ell))]}{\varphi^{(\ell,1-c)}(\ell)} = \alpha_{\varphi^{(\ell,1-c)}}(\ell)\,,$$

   since $\varphi^{(\ell,1-c)}(x) \leq \varphi(x)$ for all x. Note that we have $\varphi(\ell)-\varphi(\ell-1) \geq \varphi(m)-\varphi(m-1) = 1-c$ by concavity of φ. So, we only have to show that for all $1 \leq \ell \leq m$, we have $\alpha^{(\ell,1-c)} := \alpha_{\varphi^{(\ell,1-c)}}(\ell) \geq 1-ce^{-1}$.

2. Let us show that $\alpha^{(\ell,1-c)} := \alpha_{\varphi^{(\ell,1-c)}}(\ell) \geq 1-ce^{-1}$ for $1 \leq \ell \leq m$.

   Using the closed-form expression of Proposition 3.21 on $\varphi^{(\ell,1-c)}$ evaluated at ℓ, one gets:

   $$\alpha^{(\ell,1-c)} = \alpha_{\varphi^{(\ell,1-c)}}(\ell) = 1 - e^{-\ell}\left(\sum_{k=0}^{\ell-1}\left(\frac{\varphi(\ell)-\varphi(k)}{\varphi(\ell)}\right)\frac{\ell^k}{k!} - \frac{1-c}{\varphi(\ell)}\frac{\ell^{\ell+1}}{\ell!}\right)\,.$$

The worst case occurs when $\varphi^{(\ell,1-c)}$ is linear between 1 and ℓ, which we call $\varphi_{\text{lin}}^{(\ell,1-c)}$. Indeed, if we call $b := \frac{\varphi(\ell)-1}{\ell-1}$, then for $1 \leq k \leq \ell$, we have that $\varphi_{\text{lin}}^{(\ell,1-c)}(k) = 1 + b(k-1)$. But:

$$\sum_{k=0}^{\ell-1} \left(\frac{\varphi(\ell) - \varphi(k)}{\varphi(\ell)} \right) \frac{\ell^k}{k!} \leq 1 + \sum_{k=1}^{\ell-1} \left(\frac{\varphi(\ell) - (1 + b(k-1))}{\varphi(\ell)} \right) \frac{\ell^k}{k!} ,$$

since $\varphi(k) \geq 1 + b(k-1)$, because $\frac{\varphi(k)-\varphi(1)}{k-1} \geq \frac{\varphi(\ell)-\varphi(1)}{\ell-1} = b$ by concavity of φ. In that case, the expression can be simplified:

$$\begin{aligned}
\alpha^{(\ell,1-c)} \geq \alpha_{\varphi_{\text{lin}}^{(\ell,1-c)}}(\ell) &= 1 - e^{-\ell} \left(1 + \sum_{k=1}^{\ell-1} \left(\frac{b(\ell-k)}{\varphi(\ell)} \right) \frac{\ell^k}{k!} - \frac{1-c}{\varphi(\ell)} \frac{\ell^{\ell+1}}{\ell!} \right) \\
&= 1 - e^{-\ell} \left(1 + \frac{b\ell}{\varphi(\ell)} \sum_{k=1}^{\ell-1} \frac{\ell^k}{k!} - \frac{b\ell}{\varphi(\ell)} \sum_{k=1}^{\ell-1} \frac{\ell^{k-1}}{(k-1)!} - \frac{1-c}{\varphi(\ell)} \frac{\ell^{\ell+1}}{\ell!} \right) \\
&= 1 - \frac{e^{-\ell}}{\varphi(\ell)} \left(\varphi(\ell) + b\ell \left(\frac{\ell^{\ell-1}}{(\ell-1)!} - 1 \right) - (1-c) \frac{\ell^{\ell+1}}{\ell!} \right) \qquad (3.20) \\
&= 1 - \frac{e^{-\ell}}{\varphi(\ell)} \left(1 + b(\ell-1) + b \left(\frac{\ell^\ell}{(\ell-1)!} - \ell \right) - (1-c) \frac{\ell^{\ell+1}}{\ell!} \right) \\
&= 1 - e^{-\ell} \frac{1 - b + (b-(1-c)) \frac{\ell^{\ell+1}}{\ell!}}{\varphi(\ell)} .
\end{aligned}$$

We have also that $b \geq \varphi(\ell) - \varphi(\ell-1) \geq 1-c$ since φ concave. As a function of $(b-(1-c))$ for c fixed, we get $g(x) := 1 - e^{-\ell} \frac{c + x\left(\frac{\ell^{\ell+1}}{\ell!} - 1\right)}{1 + (x + (1-c))(\ell-1)}$. In particular, we have that $\alpha_{\varphi_{\text{lin}}^{(\ell,(1-c))}}(\ell) = g(b-(1-c))$, since $\varphi(\ell) = 1 + b(\ell-1)$. We have that $g'(x) = -e^{-\ell} \frac{\ell\left(\frac{\ell^\ell}{\ell!} - 1\right) + (1-c)\frac{\ell^{\ell+1}}{\ell!}(\ell-1)}{(1 + (x+(1-c))(\ell-1))^2} \leq 0$, so g is nonincreasing: it is thus enough to show that $g(c) \geq 1 - ce^{-1}$ to get the result, since $\alpha^{(\ell,1-c)} \geq g(b-(1-c)) \geq g(c) \geq 1 - ce^{-1}$. But:

$$g(c) = 1 - \frac{c\frac{\ell^{\ell+1}}{\ell!}}{1 + \ell - 1} e^{-\ell} = 1 - c\frac{\ell^\ell}{\ell!} e^{-\ell} \geq 1 - ce^{-1} ,$$

since $\frac{\ell^\ell}{\ell!} e^{-\ell}$ is a decreasing sequence.

$\square$

Proposition 3.23. *Let $F(x) := \mathbb{E}_{X \sim x}[C^\varphi(X)]$ for $x \in \{0,1\}^m$. We have an explicit formula for F:*

$$F(x) = \sum_{a=1}^{n} \sum_{k=0}^{m} \Big[\frac{1}{m+1} \sum_{\ell=0}^{m} \omega_{m+1}^{-\ell k} \prod_{j \in [m] : a \in T_j} (1 + (\omega_{m+1}^\ell - 1) x_j) \Big] \varphi(k) ,$$

with $\omega_{m+1} := \exp\left(\frac{2i\pi}{m+1}\right)$. Thus, F is computable in polynomial time in n and m.

Proof. Recall that $C^{\varphi}(S) = \sum_{a=1}^{n} C_a^{\varphi}(S)$, so by linearity of expectation we can focus on $\mathbb{E}_{X\sim x}[C_a^{\varphi}(X)]$. But $C_a^{\varphi}(X) = \varphi(|X|_a)$ where $|X|_a = |\{i \in [m] : X_i = 1 \text{ and } a \in T_i\}| \in [0, m]$. Thus:

$$\mathbb{E}_{X\sim x}[C_a^{\varphi}(X)] = \sum_{k=0}^{m} \mathbb{P}_{X\sim x}(|X|_a = k)\varphi(k) \ .$$

It remains to compute the distribution of $|X|_a$. But $|X|_a = \sum_{i\in[m]:a\in T_i} X_i$ and $X_i \sim \mathrm{Ber}(x_i)$. Thus, $|X|_a \sim \mathrm{Poi\,Bin}((x_i)_{i\in[m]:a\in T_i})$, which is known as the Poisson binomial law. Thanks to [FW10], we have that:

$$\mathbb{P}_{X\sim x}(|X|_a = k) = \frac{1}{m+1}\sum_{\ell=0}^{m} \omega_{m+1}^{-\ell k} \prod_{j\in[m]:a\in T_j} (1 + (\omega_{m+1}^{\ell} - 1)x_j) \ ,$$

where $\omega_{m+1} := \exp\left(\frac{2i\pi}{m+1}\right)$, and the result is proved. □

Proposition 3.24. *We have that*

$$|\mathbb{E}[\varphi(\mathit{Bin}(n, x/n))] - \mathbb{E}[\varphi(\mathit{Poi}(x))]| \leq \frac{x\varphi(n)}{2n} + \frac{x^{n+1}}{n!} \ .$$

In particular when $\varphi(n) = o(n)$:

$$\lim_{n\to\infty} \mathbb{E}[\varphi(\mathit{Bin}(n, x_{\varphi}/n))] = \mathbb{E}[\varphi(\mathit{Poi}(x_{\varphi}))] = \alpha_{\varphi}\varphi(x_{\varphi}) \ .$$

Proof. Thanks to [BH84, TT19], we have that the total variation distance between $\mathrm{Bin}(n, x/n)$ and $\mathrm{Poi}(x)$ is bounded in the following way:

$$\Delta(\mathrm{Bin}(n, x/n), \mathrm{Poi}(x)) \leq \frac{1 - e^{-x}}{2x} n \cdot \left(\frac{x}{n}\right)^2 \leq \frac{x}{2n} \ .$$

Thus with $B \sim \mathrm{Bin}(n, x/n)$ and $P \sim \mathrm{Poi}(x)$:

$$\begin{aligned}
|\mathbb{E}[\varphi(B)] - \mathbb{E}[\varphi(P)]| &= \left|\sum_{k=0}^{+\infty} \varphi(k)\mathbb{P}(B = k) - \sum_{k=0}^{+\infty} \varphi(k)\mathbb{P}(P = k)\right| \\
&= \left|\sum_{k=0}^{+\infty} \varphi(k)(\mathbb{P}(B = k) - \mathbb{P}(P = k))\right| \\
&\leq \sum_{k=0}^{+\infty} \varphi(k)|\mathbb{P}(B = k) - \mathbb{P}(P = k)| \\
&\leq \varphi(n)\Delta(\mathrm{Bin}(n, x/n), \mathrm{Poi}(x)) + \sum_{k=n+1}^{+\infty} \varphi(k)\mathbb{P}(P = k) \qquad (3.21) \\
&\leq \frac{x\varphi(n)}{2n} + e^{-x}\sum_{k=n+1}^{+\infty} k\frac{x^k}{k!} \quad \text{since } \varphi(k) \leq k \\
&= \frac{x\varphi(n)}{2n} + xe^{-x}\sum_{k=n}^{+\infty} \frac{x^k}{k!} \\
&\leq \frac{x\varphi(n)}{2n} + \frac{x^{n+1}}{n!} \underset{n\to\infty}{\to} 0 \text{ when } \varphi(n) = o(n) \ ,
\end{aligned}$$

by a standard upper bound on the remainder of the exponential series.

□

Proposition 3.25. *The function $g : x \mapsto \mathbb{E}[\varphi(Poi(x))]$ on $\mathbb{R}^+$ is $\mathcal{C}^\infty$ nondecreasing concave.*

Proof. Since we have that $0 \leq \varphi(k) \leq k$ for $k \in \mathbb{N}$, in particular $g(x) = e^{-x} \sum_{k=0}^{+\infty} \varphi(k) \frac{x^k}{k!}$ is $\mathcal{C}^\infty$. It is thus enough to compute its first and second derivatives:

$$\begin{aligned} g'(x) = & -e^{-x} \sum_{k=0}^{+\infty} \varphi(k) \frac{x^k}{k!} + e^{-x} \sum_{k=1}^{+\infty} \varphi(k) k \frac{x^{k-1}}{k!} \\ = & -e^{-x} \sum_{k=0}^{+\infty} \varphi(k) \frac{x^k}{k!} + e^{-x} \sum_{k=0}^{+\infty} \varphi(k+1) \frac{x^k}{k!} \\ = & \ e^{-x} \sum_{k=0}^{+\infty} (\varphi(k+1) - \varphi(k)) \frac{x^k}{k!} \ . \end{aligned} \tag{3.22}$$

But $\varphi(k+1) - \varphi(k) \geq 0$ since φ nondecreasing, so $g'(x) \geq 0$ and g is nondecreasing.

The calculus of g'' is the same where we replace φ by $\psi(k) := \varphi(k+1) - \varphi(k)$ which is a nonincreasing function by concavity of φ. Thus:

$$g''(x) = e^{-x} \sum_{k=0}^{+\infty} (\psi(k+1) - \psi(k)) \frac{x^k}{k!} \leq 0 \ .$$

since $\psi(k+1) - \psi(k) \leq 0$, and so g is concave. □

Proposition 3.26. *The function $g_q : n \mapsto \mathbb{E}[\varphi(Bin(n, q))]$ defined on $\mathbb{N}$ is nondecreasing concave. As a consequence, one can uses Jensen's inequality on the piecewise linear extension of g_q which is also continuous.*

Proof. $\mathrm{Bin}(n, q) \leq_{\mathrm{st}} \mathrm{Bin}(n+1, q)$ and we have that φ is nondecreasing, so $\mathbb{E}[\varphi(\mathrm{Bin}(n, q))] \leq \mathbb{E}[\varphi(\mathrm{Bin}(n+1, q))]$, ie $g_q(n+1) - g_q(n) \geq 0$: g_q is nondecreasing.

We show then the concavity, i.e. $g_q(n+2) - g_q(n+1) \leq g_q(n+1) - g_q(n)$. Call $\psi(x) = \varphi(x+1) - \varphi(x)$ which is nonincreasing since φ concave. Let us take $X_{k,q} \sim \mathrm{Bin}(k, q)$. Then:

$$\begin{aligned} g_q(n+1) = & \ \mathbb{E}[\varphi(X_{n+1,q})] \\ = & \sum_{i=0}^{n} \mathbb{E}[\varphi(X_{n,q} + X_{1,q}) | X_{n,q} = i] \mathbb{P}(X_{n,q} = i) \\ = & \sum_{i=0}^{n} \mathbb{E}[\varphi(i + X_{1,q}) - \varphi(i)] \mathbb{P}(X_{n,q} = i) + \sum_{i=0}^{n} \varphi(i) \mathbb{P}(X_{n,q} = i) \\ = & \sum_{i=0}^{n} \mathbb{E}[\varphi(i + X_{1,q}) - \varphi(i)] \mathbb{P}(X_{n,q} = i) + g_q(n) \ . \end{aligned} \tag{3.23}$$

Thus:

$$\begin{aligned} g_q(n+1) - g_q(n) = & \sum_{i=0}^{n} \mathbb{E}[\varphi(i + X_{1,q}) - \varphi(i)]\mathbb{P}(X_{n,q} = i) \\ = & \sum_{i=0}^{n} q(\varphi(i+1) - \varphi(i))\mathbb{P}(X_{n,q} = i) \\ = & \; q\mathbb{E}[\psi(\mathrm{Bin}(n,q))] \; . \end{aligned} \tag{3.24}$$

Then thanks to the fact that $\mathrm{Bin}(n,q) \leq_{\mathrm{st}} \mathrm{Bin}(n+1,q)$ and ψ is nonincreasing, we have that $\mathbb{E}[\psi(\mathrm{Bin}(n,q))] \geq \mathbb{E}[\psi(\mathrm{Bin}(n+1,q))]$, i.e. $g_q(n+2) - g_q(n+1) \leq g_q(n+1) - g_q(n)$. $\square$

Proposition 3.27. *With $w_i := \varphi(i) - \varphi(i-1)$, we have:*

$$\lim_{i \to +\infty} w_i = 0 \iff \varphi(n) = o(n) \; .$$

Proof.

- ($\Rightarrow$) Let $\epsilon > 0$, let us find a rank N such that for $n \geq N$, $\frac{\varphi(n)}{n} \leq \epsilon$. Let N_0 the rank from which $w_i \leq \frac{\epsilon}{2}$ and N_1 the rank from which $\frac{1}{n}\sum_{i=1}^{N_0-1} w_i \leq \frac{\epsilon}{2}$. We have

$$\begin{aligned} \frac{\varphi(n)}{n} = & \; \frac{1}{n}\sum_{i=1}^{n} w_i \leq \frac{1}{n}\sum_{i=1}^{N_0-1} w_i + \frac{1}{n}\sum_{i=N_0}^{n-1} \frac{\epsilon}{2} \\ \leq & \; \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \text{ for } n \geq \max(N_0, N_1) =: N \; . \end{aligned} \tag{3.25}$$

- ($\Leftarrow$) Since $w_i = \varphi(i) - \varphi(i-1)$ is nonnegative and nonincreasing (respectively because φ is nondecreasing and concave), then the sequence w has a limit $L \geq 0$. But

$$\frac{\varphi(n)}{n} = \frac{1}{n}\sum_{i=1}^{n} w_i \geq L \; .$$

  Since the left hand side tends to 0 by hypothesis, this means that $L = 0$.

$\square$

Proposition 3.28. *If $w_i := \varphi(i) - \varphi(i-1)$ is* geometrically dominant, *i.e.* $\forall i \in \mathbb{N}^*$, $\frac{w_i}{w_{i+1}} \geq \frac{w_{i+1}}{w_{i+2}}$, *then $\alpha_\varphi = \alpha_\varphi(1)$.*

Remark. Proposition 3.28 and in particular its proof uses similar ideas to the sketch provided in [DMMS20].

Proof. Let $g(k) = \mathbb{E}[\varphi(\mathrm{Poi}(k))]$, and thus $\alpha_\varphi(k) = \frac{g(k)}{\varphi(k)}$. Let us show that for $k \in \mathbb{N}^*, \alpha_\varphi(k) \geq \alpha_\varphi(1)$, which will be enough to conclude. In order to show this, we will need the following lemmas:

Lemma 3.29. *$\forall k < i \in \mathbb{N}, w_i \geq w_{k+1}w_{i-k}$ and thus $\forall k, j \in \mathbb{N}, \varphi(k+j) - \varphi(k) \geq w_{k+1}\varphi(j)$.*

Proof. We have that:

$$w_i = \frac{w_i}{w_{i-1}} \frac{w_{i-1}}{w_{i-2}} \dots \frac{w_{i-k+1}}{w_{i-k}} w_{i-k} \; .$$

But for $j \in [k]$:

$$\frac{w_{i-j+1}}{w_{i-j}} \geq \frac{w_{(i-1)-j+1}}{w_{(i-1)-j}} \geq \dots \geq \frac{w_{(k+1)-j+1}}{w_{(k+1)-j}} \; ,$$

since w is geometrically dominant and $k+1 \leq i$. Thus applying this bound on each term of the previous product, we get:

$$w_i \geq \frac{w_{k+1}}{w_k} \frac{w_k}{w_{k-1}} \dots \frac{w_2}{w_1} w_{i-k} = \frac{w_{k+1}}{w_1} w_{i-k} = w_{k+1} w_{i-k} \; .$$

In particular, $\forall k, j \in \mathbb{N}$, we get:

$$\varphi(k+j) - \varphi(k) = \sum_{i=k+1}^{k+j} w_i \geq w_{k+1} \sum_{i=1}^{j} w_i = w_{k+1} \varphi(j) \; .$$

□

Lemma 3.30. *The piecewise linear extension on* $[1, +\infty[$ *of* w*, defined on integers by* $w(k) = w_k$*, is convex.*

Proof. We will show that $\forall k \in \mathbb{N}^*$, $w_{k+2} - w_{k+1} \geq w_{k+1} - w_k$ which implies the convexity of its piecewise linear extension on $[1, +\infty[$. For $k \in \mathbb{N}^*$ we have:

$$\frac{w_{k+1}}{w_{k+2}} - 1 \leq \frac{w_{k+1}}{w_{k+2}} \Big(\frac{w_{k+1}}{w_{k+2}} - 1 \Big) \leq \frac{w_{k+1}}{w_{k+2}} \Big(\frac{w_k}{w_{k+1}} - 1 \Big) = \frac{w_k - w_{k+1}}{w_{k+2}} \; ,$$

since w is nonnegative nonincreasing (respectively φ nondecreasing concave) and $\frac{w_{k+1}}{w_{k+2}} \leq \frac{w_k}{w_{k+1}}$ since w is geometrically dominant. Then, multiplying by $-w_{k+2} \leq 0$ gives the expected result $w_{k+2} - w_{k+1} \geq w_{k+1} - w_k$. □

We have $g(k+1) = \mathbb{E}[\varphi(\text{Poi}(k+1))] = \mathbb{E}[\varphi(\text{Poi}(k) + \text{Poi}(1))]$ since $\text{Poi}(k+1) \sim \text{Poi}(k) + \text{Poi}(1)$. Thus:

$$\begin{aligned} g(k+1) - g(k) = \;& \mathbb{E}_{X, X' \sim \text{Poi}(k), Y \sim \text{Poi}(1)} [\varphi(X+Y) - \varphi(X')] \\ = \;& \mathbb{E}_{X \sim \text{Poi}(k), Y \sim \text{Poi}(1)} [\varphi(X+Y) - \varphi(X)] \\ \geq \;& \mathbb{E}_{X \sim \text{Poi}(k), Y \sim \text{Poi}(1)} [w_{X+1} \varphi(Y)] \quad \text{by Lemma 3.29} \\ = \;& \mathbb{E}_{X \sim \text{Poi}(k)} [w(X+1)] \mathbb{E}_{Y \sim \text{Poi}(1)} [\varphi(Y)] \; , \end{aligned} \tag{3.26}$$

by independence of $w(X+1)$ and $\varphi(Y)$. Since w is convex on $[1, +\infty[$ by Lemma 3.30 and $\text{Poi}(k) + 1 \in [1, +\infty[$, we have that $\mathbb{E}[w(\text{Poi}(k)+1)] \geq w(\mathbb{E}[\text{Poi}(k)+1]) = w(k+1) = w_{k+1}$ thanks to Jensen's inequality. Note that $g(0) = \mathbb{E}[\varphi(\text{Poi}(0))] = \varphi(0) = 0$. Then:

$$g(k) = \sum_{i=0}^{k-1} g(i+1) - g(i) \geq \Big(\sum_{i=0}^{k-1} w_{i+1} \Big) \mathbb{E}[\varphi(\text{Poi}(1))] = \varphi(k) g(1) \; .$$

Therefore:

$$\alpha_\varphi(k) = \frac{g(k)}{\varphi(k)} \geq g(1) = \frac{g(1)}{\varphi(1)} = \alpha_\varphi(1)\ .$$

□

3.5.2 Calculations of α_φ

Proposition 3.31. *For $\ell \in \mathbb{N}^*$ and $\varphi(j) = \min\{j, \ell\}$, we have that $\alpha_\varphi = 1 - \frac{\ell^\ell e^{-\ell}}{\ell!}$.*

Proof. Thanks to Proposition 3.22, we have that $\alpha_\varphi = \min_{x\in\mathbb{N}^*} \alpha_\varphi(x)$. Let us compute $\mathbb{E}[\varphi(\mathrm{Poi}(x))]$:

$$\begin{aligned}
\mathbb{E}[\varphi(\mathrm{Poi}(x))] = &\ e^{-x}\sum_{k=0}^{+\infty}\varphi(k)\frac{x^k}{k!} \\
= &\ e^{-x}\sum_{k=0}^{\ell} k\frac{x^k}{k!} + e^{-x}\sum_{k=\ell+1}^{+\infty}\ell\frac{x^k}{k!} \\
= &\ e^{-x}x\sum_{k=0}^{\ell-1}\frac{x^k}{k!} + \ell e^{-x}\sum_{k=\ell+1}^{+\infty}\frac{x^k}{k!} \\
= &\ e^{-x}\Big[(x-\ell)\sum_{k=0}^{\ell-1}\frac{x^k}{k!} - \ell\frac{x^\ell}{\ell!}\Big] + \ell e^{-x}\sum_{k=0}^{+\infty}\frac{x^k}{k!} \\
= &\ \ell - e^{-x}\Big[\frac{x^\ell}{(\ell-1)!} - (x-\ell)\sum_{k=0}^{\ell-1}\frac{x^k}{k!}\Big]\ .
\end{aligned} \tag{3.27}$$

Let us show that $x \mapsto \alpha_\varphi(x)$ takes its minimum in ℓ, where we have indeed:

$$\alpha_\varphi(\ell) = \frac{1}{\ell}\Big(\ell - e^{-\ell}\Big[\frac{\ell^\ell}{(\ell-1)!} - (\ell-\ell)\sum_{k=0}^{\ell-1}\frac{\ell^k}{k!}\Big]\Big) = 1 - e^{-\ell}\frac{\ell^\ell}{\ell!}\ .$$

Thanks to proposition 3.21, $x \mapsto \alpha_\varphi(x)$ is nondecreasing from ℓ to $+\infty$. Suppose now that $\ell \geq 2$ (otherwise the result is already proved). Since $x \mapsto \alpha_\varphi(x)$ is differentiable, we have for $1 \leq x \leq \ell$:

$$
\begin{aligned}
\alpha'_\varphi(x) = & -\frac{\ell}{x^2} + e^{-x}\Big[\frac{x^{\ell-1}}{(\ell-1)!} - \sum_{k=0}^{\ell-1}\frac{x^k}{k!} + \ell\sum_{k=0}^{\ell-2}\frac{x^k}{(k+1)!} + \frac{\ell}{x}\Big] \\
- & \; e^{-x}\Big[\frac{x^{\ell-2}}{(\ell-2)!} - \sum_{k=0}^{\ell-2}\frac{x^k}{k!} + \ell\sum_{k=0}^{\ell-3}\frac{x^k}{(k+2)k!} - \frac{\ell}{x^2}\Big] \\
= & \; \frac{\ell}{x}\Big(e^{-x}\Big(1+\frac{1}{x}\Big) - \frac{1}{x}\Big) \\
+ & \; e^{-x}\Big[\Big(\frac{\ell}{\ell-1} - 1\Big)\frac{x^{\ell-2}}{(\ell-2)!} + \ell\sum_{k=0}^{\ell-3}\Big(\frac{x^k}{(k+1)!} - \frac{x^k}{(k+2)k!}\Big)\Big] \\
= & \; \frac{\ell}{x}\Big(e^{-x}\Big(1+\frac{1}{x}\Big) - \frac{1}{x}\Big) + e^{-x}\Big[\frac{x^{\ell-2}}{(\ell-1)!} + \ell\sum_{k=0}^{\ell-3}\frac{x^k}{k!}\Big(\frac{1}{k+1} - \frac{1}{k+2}\Big)\Big] \\
= & \; \frac{\ell}{x}\Big(e^{-x}\Big(1+\frac{1}{x} + \frac{x^{\ell-1}}{\ell!} + x\sum_{k=0}^{\ell-3}\frac{x^k}{k!}\frac{1}{(k+1)(k+2)}\Big) - \frac{1}{x}\Big) \\
= & \; \frac{\ell e^{-x}}{x^2}\Big(\Big(1 + x + \frac{x^\ell}{\ell!} + \sum_{k=0}^{\ell-3}\frac{x^{k+2}}{(k+2)!}\Big) - e^x\Big) \\
= & \; \frac{\ell e^{-x}}{x^2}\Big(\sum_{k=0}^{\ell}\frac{x^k}{k!} - e^x\Big) \leq 0\,.
\end{aligned}
\tag{3.28}
$$

since the partial sum of the exponential series is bounded by its total sum. Thus $\alpha_\varphi(x)$ is nonincreasing from 1 to ℓ, and nondecreasing after, so it takes indeed its minimum in ℓ and the proposition is proved. $\square$

Proposition 3.32. *For $p \in (0,1)$ and $\varphi(j) = \frac{1-(1-p)^j}{p}$, we have that $\alpha_\varphi = \frac{1-e^{-p}}{p}$.*

Proof. By definition:

$$
\begin{aligned}
\alpha_\varphi(x) = & \; \frac{\mathbb{E}[\varphi(\mathrm{Poi}(x))]}{\varphi(x)} = \frac{\sum_{k=0}^{+\infty}\varphi(k)e^{-x}\frac{x^k}{k!}}{\varphi(x)} \\
= & \; \frac{1 - e^{-x}\sum_{k=0}^{+\infty}(1-p)^k\frac{x^k}{k!}}{p\varphi(x)} \\
= & \; \frac{1 - e^{-x}e^{(1-p)x}}{p\varphi(x)} = \frac{1-e^{-px}}{p\varphi(x)}\,.
\end{aligned}
\tag{3.29}
$$

If $x \geq 1$, $\alpha_\varphi(x) = \frac{1-e^{-px}}{1-(1-p)^x} = \frac{1-e^{-px}}{1-e^{-qx}}$ with $q = \ln\Big(\frac{1}{1-p}\Big) > 0$ and:

$$
\alpha'_\varphi(x) = \frac{pe^{-px}(1-e^{-qx}) - qe^{-qx}(1-e^{-px})}{(1-e^{-qx})^2} = \frac{pe^{-px} - qe^{-qx} + (q-p)e^{-(p+q)x}}{(1-e^{-qx})^2}\,.
$$

Let us take $t = \frac{p}{q} \in (0,1)$, since $q = \ln\Big(\frac{1}{1-p}\Big) > p > 0$, $x_1 = -px$ and $x_2 = -(p+q)x$. Then by strict convexity of the exponential function, we have:

$$e^{tx_1+(1-t)x_2} < te^{x_1} + (1-t)e^{x_2} = \frac{pe^{-px} + (q-p)e^{-(p+q)x}}{q} \ .$$

But $tx_1 + (1-t)x_2 = \frac{-p^2x}{q} + \frac{-(q-p)(p+q)x}{q} = \frac{-p^2x}{q} + \frac{-(q^2x-p^2x)}{q} = -qx$, so we get $pe^{-px} - qe^{-qx} + (q-p)e^{-(p+q)x} > 0$, and $\alpha'_\varphi(x) > 0$. Thus, $\alpha_\varphi(x)$ increases from 1 to infinity and takes its minimum in 1:

$$\alpha_\varphi = \alpha_\varphi(1) = \frac{1-e^{-p}}{p} \ .$$

□

Proposition 3.33. *For $d \in (0,1)$ and $\varphi(j) = j^d$, we have that $\alpha_\varphi = e^{-1}\sum_{k=1}^{+\infty} \frac{k^d}{k!}$.*

Proof. We have for $x \geq 1$:

$$\alpha_\varphi(x) = \frac{\mathbb{E}[\mathrm{Poi}(x)^d]}{\varphi(x)} = \frac{e^{-x}\sum_{k=0}^{+\infty} k^d \frac{x^k}{k!}}{\varphi(x)} = e^{-x}\sum_{k=0}^{+\infty} k^d \frac{x^{k-d}}{k!} \ .$$

Then:

$$\begin{aligned}
\alpha'_\varphi(x) = & \ -\alpha_\varphi(x) + e^{-x}\sum_{k=1}^{+\infty}(k-d)k^d\frac{x^{k-d-1}}{k!} \\
= & \ -\alpha_\varphi(x) + e^{-x}\sum_{k=0}^{+\infty}(k+1-d)(k+1)^d\frac{x^{k-d}}{(k+1)!} \\
= & \ -\alpha_\varphi(x) + e^{-x}\Big((1-d)x^{-d} + \sum_{k=1}^{+\infty}(k+1-d)(k+1)^{d-1}\frac{x^{k-d}}{k!}\Big) \\
= & \ e^{-x}x^{-d}\Big(1-d+\sum_{k=1}^{+\infty}(\frac{k+1-d}{k+1}(k+1)^d - k^d)\frac{x^k}{k!}\Big) \ .
\end{aligned} \tag{3.30}$$

But the function $f(k) = \frac{k+1-d}{k+1}(k+1)^d - k^d$ is positive on $\mathbb{R}_+^*$, so we get that $\alpha'_\varphi(x) > 0$ for $x \geq 1$, thus $\alpha_\varphi(x)$ is increasing from 1 to $+\infty$, so $\alpha_\varphi = \alpha_\varphi(1) = e^{-1}\sum_{k=1}^{+\infty}\frac{k^d}{k!}$. □

3.5.3 NP-hardness of δ, h-AryGapLabelCover

Proof of Proposition 3.6. We reduce from the Label Cover problem described in [DMMS20] which is known to be an NP-hard problem. The main idea of this reduction is the usual equivalence between bipartite graphs and hypergraphs.

Definition 3.5. A Label Cover instance $\mathcal{L} = (A, B, E, [L], [R], \{\pi_e\}_{e\in E})$ consists of a bi-regular bipartite graph (A, B, E) with right degree t, alphabet sets $[L], [R]$ and for every edge $e \in E$, a constraint $\pi_e : [L] \to [R]$. A *labeling* of $\mathcal{L}$ is a function $\sigma : A \to [L]$. We say that σ *strongly satisfies* a right vertex $v \in B$ if for every two neighbours u, u' of v, we have $\pi_{(u,v)}(\sigma(u)) = \pi_{(u',v)}(\sigma(u'))$. Moreover, we say that σ *weakly satisfies* a right vertex $v \in B$ if there exists two neighbours u, u' of v such that $\pi_{(u,v)}(\sigma(u)) = \pi_{(u',v)}(\sigma(u'))$.

Theorem 3.34 (δ-Gap-Label-Cover(t, R) from [DMMS20]). *For any fixed integer $t \geq 2$ and fixed $\delta > 0$, there exists R_0 such that for any integer $R \geq R_0$, it is NP-hard for Label Cover instances $\mathcal{L} = (A, B, E, [L], [R], \{\pi_e\}_{e\in E})$ with right degree t and right alphabet $[R]$ to distinguish between:*

YES: *There exists a labeling σ that* strongly satisfies *all the right vertices.*

NO: *No labeling* weakly satisfies *more than δ fraction of the right vertices.*

The reduction is the following. From δ-Gap-Label-Cover(t, R), we take $h = t$ and the same parameters δ, R. Given an instance $\mathcal{L} = (A, B, E, [L], [R], \{\pi_e\}_{e\in E})$, we take $\mathcal{G} = (A, E', [L], [R], \{\pi'_{e',v}\}_{e'\in E', v\in e'})$ with $E' = \{N(b), b \in B\}$ with $N(b)$ the set of neighbours of b in $\mathcal{L}$, and $\pi'_{e',v} = \pi'_{N(b),v} := \pi_{v,b}$ since $v \in N(b)$. Since (A, B, E) is bipartite and biregular, we get that our hypergraph has all hyperedges of size $h = |N(b)| = t$, and that it is regular from the regular left degree of (A, B, E). By construction, the notion of weakly and strongly satisfied is the same in both cases, as well as the labelings, and thus we have the NP-hardness of δ, h-AryGapLabelCover.

Note that both problems are in fact linearly equivalent since we could do the same reduction backwards.

□

3.5.4 Proof of existence of partitioning systems

Proof of Proposition 3.7. The existential proof is based on the probabilistic method. We take $\mathcal{P}_i$ an h-equi-sized uniform random x_φ-cover of $[n]$. Hence in the collection $\mathcal{P}_i = (P_{i,1}, \ldots, P_{i,h})$, each of the h subsets is of cardinality $\frac{x_\varphi n}{h}$. Write $\mathcal{P} = (\mathcal{P}_1, \ldots, \mathcal{P}_R)$. We have that for any $a \in [n], \mathbb{P}(a \in P_{i,j}) = \frac{x_\varphi}{h}$. Note that these events are independent for different is.

By construction, the first condition is fulfilled. Let us prove the second one.

Fix $T \subseteq [R]$ and $\mathcal{Q} := \{P_{i,j(i)} : i \in T\}$ for some function $j : T \to [h]$. We have for $a \in [n]$:

$$\mathbb{E}[C^\varphi_a(\mathcal{Q})] = \mathbb{E}[\varphi(|\mathcal{Q}|_a)] = \mathbb{E}[\varphi(\left|\{i \in T : a \in P_{i,j(i)}\}\right|)] \,.$$

But the random variables $\{X^a_i := \mathbb{1}_{a\in P_{i,j(i)}}\}_{i\in T}$ are independent and $X^a_i \sim \text{Ber}(\frac{x_\varphi}{h})$, so $X^a := \left|\{i \in T : a \in P_{i,j(i)}\}\right| = \sum_{i\in T} X^a_i \sim \text{Bin}(|T|, \frac{x_\varphi}{h})$, and thus:

$$\mathbb{E}[C^\varphi_a(\mathcal{Q})] = \mathbb{E}[\varphi(\text{Bin}(|T|, \frac{x_\varphi}{h}))] = \psi^\varphi_{|T|,h} \,.$$

Since $|\mathcal{Q}|_a \leq |\mathcal{Q}| \leq R$ and φ nondecreasing, we have $0 \leq C^\varphi_a(\mathcal{Q}) \leq \varphi(R)$. We claim that we can apply a Chernoff-Hoeffding bound on $C^\varphi(\mathcal{Q}) = \sum_{a\in[n]} C^\varphi_a(\mathcal{Q})$ and get:

$$\mathbb{P}\Big(\left|C^\varphi(\mathcal{Q}) - \psi^\varphi_{|T|,h} n\right| > \eta n\Big) \leq 2\text{exp}\Big(-2\Big(\frac{\eta}{\varphi(R)}\Big)^2 n\Big) \,.$$

The random variables $\{C^\varphi_a(\mathcal{Q})\}_{a\in[n]}$ are not independent in general. However, they are negatively associated [JDP83], and this is sufficient for the Chernoff-Hoeffding bound to hold as pointed out in [DR98], provided that $\eta \in (0,1)$. The set of random variables $\{C_a^\varphi(\mathcal{Q})\}_{a\in[n]}$ is said to be negatively associated if for any functions f and g either both increasing or both decreasing and any disjoint index sets $I, J \subseteq [n]$, we have:

$$\mathbb{E}[f(C_a^\varphi(\mathcal{Q}) : a \in I) \cdot g(C_a^\varphi(\mathcal{Q}) : a \in J)] \leq \mathbb{E}[f(C_a^\varphi(\mathcal{Q}) : a \in I)] \cdot \mathbb{E}[g(C_a^\varphi(\mathcal{Q}) : a \in J)] \,.$$

Note that $C_a^\varphi(\mathcal{Q}) = \varphi(|\mathcal{Q}|_a) = \varphi(X^a)$ is a nondecreasing function of $\{X_i^a\}_{i\in[R]}$, since φ is nondecreasing and $X^a = \sum_{i\in T} X_i^a$. Thus in order to show that $\{C_a^\varphi(\mathcal{Q})\}_{a\in[n]}$ are negatively associated, it suffices to show that $\{X_i^a\}_{i\in[R],a\in[n]}$ are negatively associated (see Proposition P_6 of [JDP83]).

For fixed $i \in [R]$, $\{X_i^a\}_{a\in[n]}$ are negatively associated because it corresponds to a permutation distribution of $(0,\dots,0,1,\dots,1)$, with $n - \frac{x_\varphi n}{h}$ zeros and $\frac{x_\varphi n}{h}$ ones, since it describes a random subset of size $\frac{x_\varphi n}{h}$ (see Definition 2.10 and Theorem 2.11 of [JDP83]). Then, using the fact that the families $\{X_i^a\}_{a\in[n]}$ are mutually independent, we obtain that $\{X_i^a\}_{i\in[R],a\in[n]}$ are negatively associated (see Property P_7 of [JDP83]). Using [DR98], this establishes the claimed Chernoff-Hoeffding bound.

Since there are at most $(h+1)^R$ choices of T and $\mathcal{Q}$, a union bound gives:

$$\mathbb{P}\Big(\exists C, \mathcal{Q} : \Big|C^\varphi(\mathcal{Q}) - \psi_{|T|,h}^\varphi n\Big| > \eta n\Big) \leq 2(h+1)^R \exp\Big(-2\Big(\frac{\eta}{\varphi(R)}\Big)^2 n\Big)\,.$$

Thus with probability at least $9/10$, we have that $\Big|C^\varphi(\mathcal{Q}) - \psi_{|T|,h}^\varphi n\Big| \leq \eta n$, since we have taken $n \geq \eta^{-2} R\varphi(R)^2 \log(20(h+1))$. So there must exists some choice of $\mathcal{P}$ that satisfies the first and second constraints of partitioning systems. Thus, we can enumerate over all choices of $\mathcal{P}$ in time $\exp(Rn\log(n)) \cdot \mathrm{poly}(h)$ to find such a partitioning system. □

3.5.5 Proof of Theorem 3.15

Proof. We show that φ-Resource Allocation corresponds to φ-MaxCoverage under a matroid constraint. Given an instance of φ-Resource Allocation, consider the partition matroid $\mathcal{M}$ on $[\sum_{i\in[k]} m_i] := [m_1]+\ldots+[m_k]$, where $(B_i)_{i\in[k]} := ([m_i])_{i\in[k]}$ is a partition of the ground set and the cardinality constraint for each i is to $d_i = 1$.

Here, $I \subseteq [\sum_{i\in[k]} m_i]$ is an independent set of the matroid iff $|I \cap B_i| \leq d_i = 1$, for all $i \in [k]$. This corresponds to each agent $i \in [k]$ selecting at most one element from the available m_i choices. In other words, we have a bijection f between tuples $(A_1,\dots,A_k) \in \mathcal{A}_1 \times \ldots \times \mathcal{A}_k$ and maximal independent sets (bases) of $\mathcal{M}$ such that $W^\varphi(A) = C^\varphi(f(A))$. Therefore, Theorem 3.4 leads to a polynomial-time α_φ-approximation algorithm for φ-Resource Allocation.

For the hardness part of the theorem, the proof is exactly the same as in Theorem 3.5, but instead of $\mathcal{F} := \{F_\beta^v, v \in V, \beta \in [L]\}$ and $k = |V|$, we take $k = |V|$ to be the number of agents and $\mathcal{A}_i := \{F_\beta^{v_i}, \beta \in [L]\}$ where $V = \{v_1,\dots,v_k\}$. Hence, instead of subsets of $\mathcal{F}$ of size k, we only consider one set $F_\beta^v \in \mathcal{F}$, for each $v \in V$. The function we maximize in the reduction remains unchanged.

To establish completeness, we note that the subset described is already of the right form and, hence, the arguments continue to hold. For proving soundness, the constraint on the shape of the subset of $\mathcal{F}$ only helps us, since it gives more constraints on the given subset from which we want to construct a labeling. Therefore, the NP-hardness follows. $\square$

Chapter 4

Multiple-Access Channel Coding with Non-Signaling Correlations

Multiple-access channels (MACs for short) are one of the simplest models of network communication settings, where two senders aim to transmit individual messages to one receiver. The capacity of such channels has been entirely characterized by the seminal works by Liao [Lia73] and Ahlswede [Ahl73] in terms of a simple single-letter formula. From the point of view of quantum information, it is natural to ask whether additional resources, such as quantum entanglement or more generally non-signaling correlations between the parties, changes the capacity region. A non-signaling correlation is a multipartite input-output box shared between parties that, as the name suggests, cannot by itself be used to send information between parties. However, non-signaling correlations such as the ones generated by measurements of entangled quantum particles, can provide an advantage for various information processing tasks and nonlocal games. The study of such correlations has given rise to the quantum information area known as nonlocality [BCP+14]. For example, in the context of channel coding, there exists classical point-to-point channels for which quantum entanglement between the sender and the receiver can increase the optimal success probability for sending one bit of information with a single use of the channel [PLM+11, BF18]. However, a well-known result [BSST99] states that for classical point-to-point channels, entanglement and even more generally non-signaling correlations do not change the capacity of the channel; see also [Mat12, BF18].

In the network setting, behavior is different. Quek and Shor showed in [QS17] the existence of two-sender two-receiver interference channels with gaps between their classical, quantum-entanglement assisted and non-signaling assisted capacity regions. Following this result, Leditzky et al. [LALS20] (see also [SLSS22]) showed that quantum entanglement shared between the two senders of a MAC can strictly enlarge the capacity region. This has been demonstrated through channels that are constructed from two-player non-local games, such as the Magic Square game [Mer90, Per90, Ara02, BBT05], by translating known gaps between classical and quantum values of games into MAC capacity gaps. Other instances of network channels for which entanglement increases the capacity region were studied in [Noe20, ND20]. This raises the following natural question: Can non-signaling correlations lead to significant gains in capacity for natural MACs? Can we find a characterization of the capacity region of the MAC when non-signaling resources between the parties are allowed?

Our Results We focus here on the MAC with two senders and we allow arbitrary tripartite non-signaling correlations between the two senders and the receiver. This is the most optimistic setting, in the sense that we only enforce the non-signaling constraints between the parties, and also the mathematically simplest setting. Even if not all non-signaling correlations are feasible within quantum theory, the setting we study here can be seen as a tractable and physically motivated outer approximation of what can be achieved with quantum theory. In fact, the quantum set is notoriously complicated and deciding membership in this set is not computable [JNV+20]. We note that very recently, Pereg et al. [PDB23] found a single-letter formula for the capacity of MACs with quantum entanglement shared between the two senders. Unfortunately, this characterization is very difficult to evaluate for any fixed channel.

We denote by $\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)$ the success probability of the best non-signaling assisted (k_1, k_2)-code for the MAC W. Contrary to the unassisted value that we denote $\mathrm{S}(W, k_1, k_2)$, $\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)$ can be formulated as a linear program; see Proposition 4.4. Furthermore, using symmetries, we have developed a linear program computing S^{NS} for a finite number of copies of a MAC W with a size growing polynomially in the number of copies; see Theorem 4.10 and Corollary 4.11. Using this result, we describe a method to derive inner bounds on the non-signaling assisted capacity region achievable with zero error; see Proposition 4.15. Applied to the binary adder channel, which maps $(x_1, x_2) \in \{0,1\}^2$ to $x_1 + x_2 \in \{0,1,2\}$, we show that the sum-rate $\frac{\log_2(72)}{4} \simeq 1.5425$ can be reached with zero error, which beats the maximum classical sum-rate capacity of $\frac{3}{2}$; see Theorem 4.16. For noisy channels, where the zero-error non-signaling assisted capacity region is trivial, we can use concatenated codes to obtain achievable points in the capacity region; see Proposition 4.18. Applied to a noisy version of the binary adder channel, we show that non-signaling assistance still improves the sum-rate capacity.

In order to find outer bounds, we define a relaxed notion of non-signaling assistance and characterize its capacity region by a single-letter expression, which is the same as the well-known expression for the capacity of the MAC (see Theorem 4.1) except that the inputs X_1 and X_2 are not required to be independent; see Theorem 4.22. This gives in particular an outer bound on the non-signaling assisted capacity region; see Corollary 4.29. The main open problem that we leave is whether this outer bound on the non-signaling capacity region is tight. We give an example of a channel for which the relaxed notion of non-signaling assistance gives a strictly larger success probability than non-signaling assistance but we do not know if such a gap can persist for the capacity region.

We also study the case where non-signaling assistance is shared only between each sender and the receiver independently. Note that no assistance is shared between the senders. We show that this capacity region is the same as the capacity region without any assistance; see Theorem 4.36 and Corollary 4.37. We note that a similar setting with independent entangled states between each sender and the receiver was studied by Hsieh et al. [HDW08]: a regularized characterization of the capacity region is obtained for any quantum MAC in this setting. It is simple to show using their result that for a classical MAC, this type of entanglement does not change the capacity region given in Theorem 4.1.

Organization In Section 4.1, we define precisely the different notions of MAC capacities: the classical capacity (i.e. without any assistance) as well as the non-signaling assisted capacity. In Section 4.2, we address computational complexity questions concerning the probability of success of the best classical coding strategy and the best non-signaling strategy for a MAC. In Section 4.3, we develop numerical methods to find inner bounds on non-signaling assisted capacity regions, and apply those to the binary adder channel and a noisy variant. In Section 4.4, we define our relaxation of non-signaling assistance, we characterize its capacity region by a single-letter formula, and apply those to the binary adder channel. Finally, in Section 4.5, we show that the capacity region with non-signaling assistance shared only between each sender and the receiver independently is the same as without assistance.

4.1 Multiple Access Channels Capacities

4.1.1 Classical Capacities

Formally, a MAC W is a conditional probability distribution depending on two inputs in $\mathcal{X}_1$ and $\mathcal{X}_2$, and an output in $\mathcal{Y}$, so $W := (W(y|x_1x_2))_{x_1\in\mathcal{X}_1,x_2\in\mathcal{X}_2,y\in\mathcal{Y}}$ such that $W(y|x_1x_2) \geq 0$ and $\sum_{y\in\mathcal{Y}} W(y|x_1x_2) = 1$. We will denote such a MAC by $W : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}$. The tensor product of two MACs $W : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}$ and $W' : \mathcal{X}'_1 \times \mathcal{X}'_2 \to \mathcal{Y}'$ is denoted by $W \otimes W' : (\mathcal{X}_1 \times \mathcal{X}'_1) \times (\mathcal{X}_2 \times \mathcal{X}'_2) \to \mathcal{Y} \times \mathcal{Y}'$ and defined by $(W \otimes W')(yy'|x_1x'_1x_2x'_2) := W(y|x_1x_2) \cdot W'(y'|x'_1x'_2)$. We denote by $W^{\otimes n}(y^n|x_1^nx_2^n) := \prod_{i=1}^n W(y_i|x_{1,i}x_{2,i})$, for $y^n := y_1 \dots y_n \in \mathcal{Y}^n$, $x_1^n := x_{1,1} \dots x_{1,n} \in \mathcal{X}_1^n$ and $x_2^n := x_{2,1} \dots x_{2,n} \in \mathcal{X}_2^n$. We will use the notation $[k] := \{1, \dots, k\}$.

The coding problem for a MAC $W : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}$ is the following: one wants to encode messages in $[k_1]$ into $\mathcal{X}_1$ and messages in $[k_2]$ into $\mathcal{X}_2$ independently, which will be given as input to the channel W. This results in a random output in $\mathcal{Y}$, which one needs to decode back into the corresponding messages in $[k_1]$ and $[k_2]$. We will call $e_1 : [k_1] \to \mathcal{X}_1$ the first encoder, $e_2 : [k_2] \to \mathcal{X}_2$ the second encoder and $d : \mathcal{Y} \to [k_1] \times [k_2]$ the decoder. This is depicted in Figure ??.

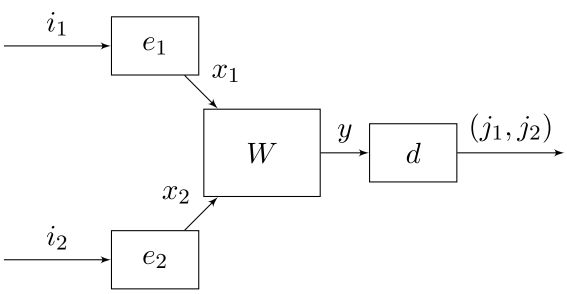


Figure 4.1 – Coding for a MAC W.

We want to maximize over all encoders e_1, e_2 and decoders d the probability of successfully encoding and decoding the messages through W, i.e. the probability that $j_1 = i_1$ and $j_2 = i_2$, given that the input messages are taken uniformly in $[k_1]$ and $[k_2]$. We call this quantity $\mathrm{S}(W, k_1, k_2)$, which is characterized by the following optimization program:

$$
\begin{aligned}
\mathrm{S}(W, k_1, k_2) := \quad \underset{e_1, e_2, d}{\text{maximize}} \quad & \frac{1}{k_1 k_2} \sum_{i_1, i_2, x_1, x_2, y} W(y|x_1x_2) e_1(x_1|i_1) e_2(x_2|i_2) d(i_1 i_2|y) \\
\text{subject to} \quad & \sum_{x_1 \in \mathcal{X}_1} e_1(x_1|i_1) = 1, \forall i_1 \in [k_1] \\
& \sum_{x_2 \in \mathcal{X}_2} e_2(x_2|i_2) = 1, \forall i_2 \in [k_2] \\
& \sum_{j_1 \in [k_1], j_2 \in [k_2]} d(j_1 j_2|y) = 1, \forall y \in \mathcal{Y} \\
& e_1(x_1|i_1), e_2(x_2|i_2), d(j_1 j_2|y) \geq 0
\end{aligned} \tag{4.1}
$$

Proof. One should note that we allow randomized encoders and decoders for generality reasons, although the value of the program is not changed as it is convex. Besides that remark, let us name $I_1, I_2, J_1, J_2, X_1, X_2, Y$ the random variables corresponding to $i_1, i_2, j_1, j_2, x_1, x_2, y$ in the coding and decoding process. Then, for given e_1, e_2, d and W, the objective value of the previous program is:

$$
\begin{aligned}
&\mathbb{P}\left(J_1 = I_1, J_2 = I_2\right) = \frac{1}{k_1 k_2} \sum_{i_1, i_2} \mathbb{P}\left(J_1 = I_1, J_2 = I_2 | I_1 = i_1, I_2 = i_2\right) \\
&= \frac{1}{k_1 k_2} \sum_{i_1, i_2, x_1, x_2} e_1(x_1|i_1) e_2(x_2|i_2) \mathbb{P}\left(J_1 = i_1, J_2 = i_2 | i_1, i_2, X_1 = x_1, X_2 = x_2\right) \\
&= \frac{1}{k_1 k_2} \sum_{i_1, i_2, x_1, x_2, y} W(y|x_1x_2) e_1(x_1|i_1) e_2(x_2|i_2) \mathbb{P}\left(J_1 = i_1, J_2 = i_2 | i_1, i_2, x_1, x_2, Y = y\right) \\
&= \frac{1}{k_1 k_2} \sum_{i_1, i_2, x_1, x_2, y} W(y|x_1x_2) e_1(x_1|i_1) e_2(x_2|i_2) d(i_1, i_2|y) \ .
\end{aligned}
$$

□

Since MACs are more general than point-to-point channels (by defining $W(y|x_1x_2) := \hat{W}(y|x_1)$ for $\hat{W}$ a point-to-point channel and looking only at its first input), computing $\mathrm{S}(W, k_1, k_2)$ is NP-hard, and it is even NP-hard to approximate $\mathrm{S}(W, k_1, k_2)$ within a better ratio than $(1 - e^{-1})$, as a consequence of the hardness result on $\mathrm{S}(W, k)$ shown in [BF18].

The (classical) capacity of a MAC, as defined for example in [CT01], can be reformulated in the following way:

Definition 4.1 (Capacity Region $\mathcal{C}(W)$ of a MAC W)**.** A rate pair (R_1, R_2) is achievable if:

$$\lim_{n \to +\infty} \mathrm{S}(W^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil) = 1 \ .$$

We define the (classical) capacity region $\mathcal{C}(W)$ as the closure of the set of all achievable rate pairs.

The capacity region $\mathcal{C}(W)$ is characterized by a single-letter formula:

Theorem 4.1 (Liao [Lia73] and Ahlswede [Ahl73])**.** *$\mathcal{C}(W)$ is the closure of the convex hull of all rate pairs (R_1, R_2) satisfying:*

$$R_1 < I(X_1 : Y|X_2) \ , \ R_2 < I(X_2 : Y|X_1) \ , \ R_1 + R_2 < I((X_1, X_2) : Y) \ ,$$

for $(X_1, X_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ following a product law $P_{X_1} \times P_{X_2}$, and $Y \in \mathcal{Y}$ the outcome of W on inputs X_1, X_2.

For the zero-error (classical) capacity, this leads to the following definition:

Definition 4.2 (Zero-Error Capacity Region $\mathcal{C}_0(W)$ of a MAC W)**.** A rate pair (R_1, R_2) is achievable with zero-error if:

$$\exists n_0 \in \mathbb{N}^*, \forall n \geq n_0, \mathrm{S}(W^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil) = 1 \ .$$

We define the zero-error (classical) capacity region $\mathcal{C}_0(W)$ as the closure of the set of all achievable rate pairs with zero-error.

We will also consider what we call the sum success probability $\mathrm{S}_{\text{sum}}(W, k_1, k_2)$, defined using $\frac{\mathbb{P}(J_1 = I_1) + \mathbb{P}(J_2 = I_2)}{2}$ rather than $\mathbb{P}\left(J_1 = I_1, J_2 = I_2\right)$ as an objective value, which leads to the following optimization program:

$$\begin{aligned}
\mathrm{S}_{\text{sum}}(W, k_1, k_2) := \quad & \underset{e_1, e_2, d_1, d_2}{\text{maximize}} && \frac{1}{2k_1k_2} \sum_{i_1, i_2, x_1, x_2, y} W(y|x_1x_2)e_1(x_1|i_1)e_2(x_2|i_2)d_1(i_1|y) \\
& + && \frac{1}{2k_1k_2} \sum_{i_1, i_2, x_1, x_2, y} W(y|x_1x_2)e_1(x_1|i_1)e_2(x_2|i_2)d_2(i_2|y) \\
& \text{subject to} && \sum_{x_1 \in \mathcal{X}_1} e_1(x_1|i_1) = 1, \forall i_1 \in [k_1] \\
& && \sum_{x_2 \in \mathcal{X}_2} e_2(x_2|i_2) = 1, \forall i_2 \in [k_2] \\
& && \sum_{j_1 \in [k_1]} d_1(j_1|y) = 1, \forall y \in \mathcal{Y} \\
& && \sum_{j_2 \in [k_2]} d_2(j_2|y) = 1, \forall y \in \mathcal{Y} \\
& && e_1(x_1|i_1), e_2(x_2|i_2), d_1(j_1|y), d_2(j_2|y) \geq 0
\end{aligned} \tag{4.2}$$

Note that we used independent decoders $d_1(j_1|y), d_2(j_2|y)$ rather than a global $d(j_1j_2|y)$ here. This does not change the value of the optimization program. Indeed, since the program is convex, an optimal solution can be found on the extremal points of the search space. Thus, if we had used the variable $d(j_1j_2|y)$, we could always take it to be a function d from $\mathcal{Y}$ to $[k_1] \times [k_2]$. Taking d_1, d_2 as the first and second coordinates of that function satisfies the equality $d(j_1j_2|y) = d_1(j_1|y)d_2(j_2|y)$, and therefore, the value of the program is the same in both cases. Note that it is also true for the program computing $\mathrm{S}(W, k_1, k_2)$.

As for the usual (joint) success probability, we can define its capacity region:

Definition 4.3 (Sum-Capacity Region $\mathcal{C}_{\text{sum}}(W)$ of a MAC W)**.** A rate pair (R_1, R_2) is sum-achievable if:

$$\lim_{n \to +\infty} \mathrm{S}_{\text{sum}}(W^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil) = 1 \ .$$

We define the sum-capacity region $\mathcal{C}_{\text{sum}}(W)$ as the closure of the set of all sum-achievable rate pairs.

However, it turns out those two notions of success define the same capacity region:

Proposition 4.2. $\mathcal{C}(W) = \mathcal{C}_{sum}(W)$

Proof. Let us focus on error probabilities rather than success ones. Call them respectively $\mathrm{E}(W, k_1, k_2) := 1 - \mathrm{S}(W, k_1, k_2)$ and $\mathrm{E}_{\text{sum}}(W, k_1, k_2) := 1 - \mathrm{S}_{\text{sum}}(W, k_1, k_2)$. Let us fix a solution e_1, d_1, e_2, d_2 of the optimization program computing $\mathrm{S}(W, k_1, k_2)$. Let us remark first that:

$$\sum_{i_1,i_2,x_1,x_2,y} W(y|x_1x_2)e_1(x_1|i_1)e_2(x_2|i_2) = k_1k_2 \ ,$$

thus, the value of the maximum error for those encoders and decoders, which we call $\mathrm{E}(W, k_1, k_2, e_1, d_1, e_2, d_2)$, is:

$$\begin{aligned}
&1 - \frac{1}{k_1k_2}\left(\sum_{i_1,i_2,x_1,x_2,y} W(y|x_1x_2)e_1(x_1|i_1)e_2(x_2|i_2)d_1(i_1|y)d_2(i_2|y)\right) \\
&= \frac{1}{k_1k_2}\sum_{i_1,i_2,x_1,x_2,y} W(y|x_1x_2)e_1(x_1|i_1)e_2(x_2|i_2) \\
&- \frac{1}{k_1k_2}\sum_{i_1,i_2,x_1,x_2,y} W(y|x_1x_2)e_1(x_1|i_1)e_2(x_2|i_2)d_1(i_1|y)d_2(i_2|y) \\
&= \frac{1}{k_1k_2}\left(\sum_{i_1,i_2,x_1,x_2,y} W(y|x_1x_2)e_1(x_1|i_1)e_2(x_2|i_2)\left[1 - d_1(i_1|y)d_2(i_2|y)\right]\right) \ .
\end{aligned} \tag{4.3}$$

Similarly, the value of the sum error for those encoder and decoders, which we call $\mathrm{E}_{\text{sum}}(W, k_1, k_2, e_1, d_1, e_2, d_2)$, is:

$$\begin{aligned}
&1 - \frac{1}{k_1k_2}\left(\sum_{i_1,i_2,x_1,x_2,y} W(y|x_1x_2)e_1(x_1|i_1)e_2(x_2|i_2)\frac{d_1(i_1|y) + d_2(i_2|y)}{2}\right) \\
&= \frac{1}{k_1k_2}\left(\sum_{i_1,i_2,x_1,x_2,y} W(y|x_1x_2)e_1(x_1|i_1)e_2(x_2|i_2)\left[1 - \frac{d_1(i_1|y) + d_2(i_2|y)}{2}\right]\right) \ .
\end{aligned} \tag{4.4}$$

However, for $x, y \in [0, 1]$, we have that:

$$1 - xy \geq \max(1 - x, 1 - y) \geq 1 - \frac{x + y}{2} \ ,$$

and:
$$1 - xy \leq (1-x) + (1-y) = 2\left(1 - \frac{x+y}{2}\right) .$$

This means that, for all e_1, d_1, e_2, d_2, we have:

$$\mathrm{E}_{\text{sum}}(W, k_1, k_2, e_1, d_1, e_2, d_2) \leq \mathrm{E}(W, k_1, k_2, e_1, d_1, e_2, d_2) \leq 2\mathrm{E}_{\text{sum}}(W, k_1, k_2, e_1, d_1, e_2, d_2) ,$$

so, maximizing over all e_1, d_1, e_2, d_2, we get:

$$\mathrm{E}_{\text{sum}}(W, k_1, k_2) \leq \mathrm{E}(W, k_1, k_2) \leq 2\mathrm{E}_{\text{sum}}(W, k_1, k_2) .$$

Thus, up to a multiplicative factor 2, the error is the same. In particular, when one of those errors tends to zero, the other one tends to zero as well. This implies that the capacity regions are the same. □

4.1.2 Non-Signaling Assisted Capacities

Three-party non-signaling assistance We now consider the case where the senders and the receiver are given non-signaling assistance. This resource, which is a theoretical but easier to manipulate generalization of quantum entanglement, can be characterized as follows. A tripartite non-signaling box is described by a joint conditional probability distribution $P(x_1x_2(j_1j_2)|i_1i_2y)$ such that the marginal from any two parties is independent from the removed party's input, i.e., we have:

$$\begin{aligned}
\forall x_2, j_1, j_2, i_1, i_2, y, i_1', \quad \sum_{x_1} P(x_1x_2(j_1j_2)|i_1i_2y) &= \sum_{x_1} P(x_1x_2(j_1j_2)|i_1'i_2y) , \\
\forall x_1, j_1, j_2, i_1, i_2, y, i_2', \quad \sum_{x_2} P(x_1x_2(j_1j_2)|i_1i_2y) &= \sum_{x_2} P(x_1x_2(j_1j_2)|i_1i_2'y) , \\
\forall x_1, x_2, i_1, i_2, y, y', \quad \sum_{j_1,j_2} P(x_1x_2(j_1j_2)|i_1i_2y) &= \sum_{j_1,j_2} P(x_1x_2(j_1j_2)|i_1i_2y') .
\end{aligned} \tag{4.5}$$

This implies that one can consider for example $P(x_1x_2|i_1i_2)$ since it does not depend on y, or even $P(x_1|i_1)$ since it does not depend on i_2, y. Then, in our coding scenario, when the senders and the receiver are given non-signaling assistance, it means that they share a tripartite non-signaling box, the behavior of which is described by P. In this case, the expression $e_1(x_1|i_1)e_2(x_2|i_2)d(j_1j_2|y)$ in (4.1) is replaced by $P(x_1x_2(j_1j_2)|i_1i_2y)$, as depicted in Figure 4.2.

The cyclicity of Figure 4.2 is at first sight counter-intuitive. Note first that P being a non-signaling box is completely independent from W: in particular, the variable y does not need to follow any law in the definition of P being a non-signaling box. Therefore, the remaining ambiguity is the apparent need to encode and decode at the same time. However, since P is a non-signaling box, we do not need to do both at the same time. Indeed, $\forall y, P(x_1x_2|i_1i_2) = P(x_1x_2|i_1i_2y)$ by the non-signaling property of P. Thus, one can get the outputs x_1, x_2 on inputs i_1, i_2 without access to y, as that knowledge won't affect the laws of x_1, x_2. Then y follows the law given by W given those x_1, x_2. Finally, given access to y, the decoding process is described by:

$$P((j_1j_2)|i_1i_2yx_1x_2) = \frac{P(x_1x_2(j_1j_2)|i_1i_2y)}{P(x_1x_2|i_1i_2y)} = \frac{P(x_1x_2(j_1j_2)|i_1i_2y)}{P(x_1x_2|i_1i_2)} ,$$

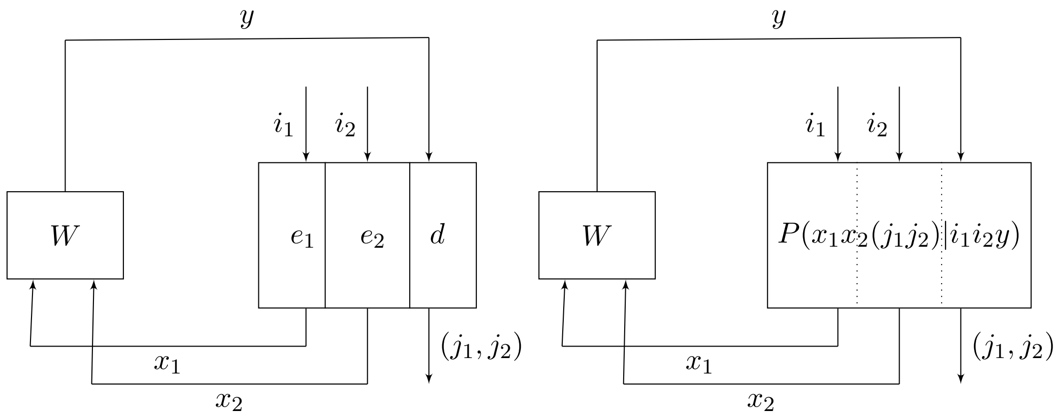


Figure 4.2 – A non-signaling box P replacing e_1, e_2 and d in the coding problem for the MAC W.

so we recover globally $P((j_1j_2)|i_1i_2yx_1x_2) \times P(x_1x_2|i_1i_2) = P(x_1x_2(j_1j_2)|i_1i_2y)$ the prescribed conditional probability. The non-signaling condition ensures that it is possible to consider the conditional probabilities of each party independently. This clarifies how one can effectively encode and then decode messages through a non-signaling box.

As in the unassisted case, we want to maximize over all non-signaling box P the probability of successfully encoding and decoding the messages through W, i.e. the probability that $j_1 = i_1$ and $j_2 = i_2$, given that the input messages are taken uniformly in $[k_1]$ and $[k_2]$. We call this quantity $\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)$, which is characterized by the following optimization program:

$$
\begin{aligned}
\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2) := \quad \underset{P}{\text{maximize}} \quad & \frac{1}{k_1k_2} \sum_{i_1,i_2,x_1,x_2,y} W(y|x_1x_2)P(x_1x_2(i_1i_2)|i_1i_2y) \\
\text{subject to} \quad & \sum_{x_1} P(x_1x_2(j_1j_2)|i_1i_2y) = \sum_{x_1} P(x_1x_2(j_1j_2)|i_1'i_2y) \\
& \sum_{x_2} P(x_1x_2(j_1j_2)|i_1i_2y) = \sum_{x_2} P(x_1x_2(j_1j_2)|i_1i_2'y) \\
& \sum_{j_1,j_2} P(x_1x_2(j_1j_2)|i_1i_2y) = \sum_{j_1,j_2} P(x_1x_2(j_1j_2)|i_1i_2y') \\
& \sum_{x_1,x_2,j_1,j_2} P(x_1x_2(j_1j_2)|i_1i_2y) = 1 \\
& P(x_1x_2(j_1j_2)|i_1i_2y) \geq 0
\end{aligned}
\tag{4.6}
$$

Since it is given as a linear program, the complexity of computing $\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)$ is polynomial in the number of variables and constraints (see for instance Section 7.1 of [GM07]), which is a polynomial in $|\mathcal{X}_1|, |\mathcal{X}_2|, |\mathcal{Y}|, k_1$ and k_2. Also, as it is easy to check that a classical strategy is a particular case of a non-signaling assisted strategy, we have that $\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2) \geq \mathrm{S}(W, k_1, k_2)$.

We have then the same definitions of capacity and zero-error capacity:

Definition 4.4 (Non-Signaling Assisted Capacity Region $\mathcal{C}^{\mathrm{NS}}(W)$ of a MAC W)**.** A rate pair (R_1, R_2) is achievable with non-signaling assistance if:

$$\lim_{n \to +\infty} \mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil) = 1 \ .$$

We define the non-signaling assisted capacity region $\mathcal{C}^{\mathrm{NS}}(W)$ as the closure of the set of all achievable rate pairs with non-signaling assistance.

Definition 4.5 (Zero-Error Non-Signaling Assisted Capacity Region $\mathcal{C}_0^{\mathrm{NS}}(W)$ of a MAC W)**.** A rate pair (R_1, R_2) is achievable with zero-error and non-signaling assistance if:

$$\exists n_0 \in \mathbb{N}^*, \forall n \geq n_0, \mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil) = 1 \ .$$

We define the zero-error non-signaling assisted capacity region $\mathcal{C}_0^{\mathrm{NS}}(W)$ as the closure of the set of all achievable rate pairs with zero-error and non-signaling assistance.

Independent non-signaling assistance One can also consider the case where non-signaling assistance is shared independently between the first sender and the receiver as well as between the second encoder and the receiver, which we call independent non-signaling assistance. The precise scenario is depicted in Figure 4.3:

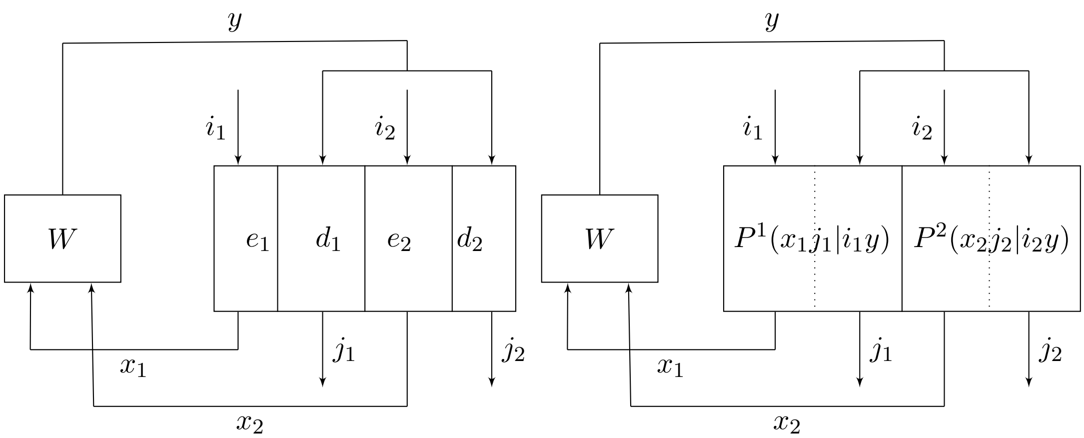


Figure 4.3 – Non-signaling boxes P^1, P^2 replacing e_1, d_1 and e_2, d_2 in the coding problem for the MAC W.

This leads to the following definition of the success probability $\mathrm{S}^{\mathrm{NS}_{\mathrm{SR}}}(W, k_1, k_2)$:

$$
\begin{aligned}
\underset{P^1,P^2}{\text{maximize}} \quad & \frac{1}{k_1k_2} \sum_{i_1,i_2,x_1,x_2,y} W(y|x_1x_2)P^1(x_1i_1|i_1y)P^2(x_2i_2|i_2y) \\
\text{subject to} \quad & \sum_{x_1} P^1(x_1j_1|i_1y) = \sum_{x_1} P^1(x_1j_1|i_1'y) \\
& \sum_{j_1} P^1(x_1j_1|i_1y) = \sum_{j_1} P^1(x_1j_1|i_1y') \\
& \sum_{x_1,j_1} P^1(x_1j_1|i_1y) = 1 \\
& \sum_{x_2} P^2(x_2j_2|i_2y) = \sum_{x_2} P^2(x_2j_2|i_2'y) \\
& \sum_{j_2} P^2(x_2j_2|i_2y) = \sum_{j_2} P^2(x_2j_2|i_2y') \\
& \sum_{x_2,j_2} P^2(x_2j_2|i_2y) = 1 \\
& P^1(x_1j_1|i_1y), P^2(x_2j_2|i_2y) \geq 0
\end{aligned}
\tag{4.7}
$$

As before, one can also consider the sum-success probability $\mathrm{S}_{\mathrm{sum}}^{\mathrm{NS}_{\mathrm{SR}}}(W, k_1, k_2)$:

$$
\begin{aligned}
\underset{P^1,P^2}{\text{maximize}} \quad & \frac{1}{2k_1k_2} \sum_{i_1,i_2,x_1,x_2,y} W(y|x_1x_2)P^1(x_1i_1|i_1y)\sum_{j_2} P^2(x_2j_2|i_2y) \\
+ \quad & \frac{1}{2k_1k_2} \sum_{i_1,i_2,x_1,x_2,y} W(y|x_1x_2)P^2(x_2i_2|i_2y)\sum_{j_1} P^1(x_1j_1|i_1y) \\
\text{subject to} \quad & \sum_{x_1} P^1(x_1j_1|i_1y) = \sum_{x_1} P^1(x_1j_1|i_1'y) \\
& \sum_{j_1} P^1(x_1j_1|i_1y) = \sum_{j_1} P^1(x_1j_1|i_1y') \\
& \sum_{x_1,j_1} P^1(x_1j_1|i_1y) = 1 \\
& \sum_{x_2} P^2(x_2j_2|i_2y) = \sum_{x_2} P^2(x_2j_2|i_2'y) \\
& \sum_{j_2} P^2(x_2j_2|i_2y) = \sum_{j_2} P^2(x_2j_2|i_2y') \\
& \sum_{x_2,j_2} P^2(x_2j_2|i_2y) = 1 \\
& P^1(x_1j_1|i_1y), P^2(x_2j_2|i_2y) \geq 0
\end{aligned}
\tag{4.8}
$$

Definition 4.6 (Independent Non-Signaling Assisted Capacity (resp. Sum-Capacity) Region $\mathcal{C}^{\mathrm{NS}_{\mathrm{SR}}}(W)$ (resp. $\mathcal{C}_{\mathrm{sum}}^{\mathrm{NS}_{\mathrm{SR}}}(W)$) of a MAC W)**.** A rate pair (R_1, R_2) is achievable (resp. sum-achievable) with independent non-signaling assistance if:

$$\lim_{n\to+\infty} \mathrm{S}^{\mathrm{NS}_{\mathrm{SR}}}(W^{\otimes n}, \lceil 2^{R_1n} \rceil, \lceil 2^{R_2n} \rceil) = 1 \ .$$

$$(\text{resp. } \lim_{n\to+\infty} \mathrm{S}_{\mathrm{sum}}^{\mathrm{NS}_{\mathrm{SR}}}(W^{\otimes n}, \lceil 2^{R_1n} \rceil, \lceil 2^{R_2n} \rceil) = 1 \ .)$$

We define the independent non-signaling assisted capacity (reps. sum-capacity) region $\mathcal{C}^{\mathrm{NS}_{\mathrm{SR}}}(W)$ (resp. $\mathcal{C}^{\mathrm{NS}_{\mathrm{SR}}}_{\text{sum}}(W)$) as the closure of the set of all achievable (resp. sum-achievable) rate pairs with independent non-signaling assistance.

However, it turns out those two notions of success define the same capacity region:

Proposition 4.3. $\mathcal{C}^{\mathrm{NS}_{\mathrm{SR}}}(W) = \mathcal{C}^{\mathrm{NS}_{\mathrm{SR}}}_{sum}(W)$

Proof. Given non-signaling boxes P^1, P^2, the maximum success probability of encoding and decoding correctly with those is given by:

$$\mathrm{S}^{\mathrm{NS}_{\mathrm{SR}}}(W, k_1, k_2, P^1, P^2) := \frac{1}{k_1 k_2} \sum_{i_1, i_2, x_1, x_2, y} W(y|x_1 x_2) P^1(x_1 i_1 | i_1 y) P^2(x_2 i_2 | i_2 y) \; .$$

This should be compared to the sum success probability of encoding and decoding correctly with those, which we call $\mathrm{S}^{\mathrm{NS}_{\mathrm{SR}}}_{\text{sum}}(W, k_1, k_2, P^1, P^2)$ and is equal to:

$$\frac{1}{k_1 k_2} \sum_{i_1, i_2, x_1, x_2, y} W(y|x_1 x_2) \frac{P^1(x_1 i_1 | i_1 y) \sum_{j_2} P^2(x_2 j_2 | i_2 y) + P^2(x_2 i_2 | i_2 y) \sum_{j_1} P^1(x_1 j_1 | i_1 y)}{2} \; .$$

Similarly to what was done in Proposition 4.2, we focus on error probabilities rather than success probabilities. We have that:

$$\frac{1}{k_1 k_2} \sum_{i_1, i_2, x_1, x_2, y} W(y|x_1 x_2) \sum_{j_1, j_2} P^1(x_1 j_1 | i_1 y) P^2(x_2 j_2 | i_2 y) = 1 \; ,$$

so we get that $\mathrm{E}^{\mathrm{NS}_{\mathrm{SR}}}(W, k_1, k_2, P^1, P^2)$ is equal to:

$$\frac{1}{k_1 k_2} \sum_{i_1, i_2, x_1, x_2, y} W(y|x_1 x_2) \left[\sum_{j_1, j_2} P^1(x_1 j_1 | i_1 y) P^2(x_2 j_2 | i_2 y) - P^1(x_1 i_1 | i_1 y) P^2(x_2 i_2 | i_2 y) \right] \; ,$$

and thus:

$$\mathrm{E}^{\mathrm{NS}_{\mathrm{SR}}}(W, k_1, k_2, P^1, P^2) = \sum_{i_1, i_2, x_1, x_2, y} W(y|x_1 x_2) \sum_{(j_1, j_2) \neq (i_1, i_2)} P^1(x_1 j_1 | i_1 y) P^2(x_2 j_2 | i_2 y) \; .$$

On the other hand, since:

$$\sum_{j_1, j_2} P^1(x_1 j_1 | i_1 y) P^2(x_2 j_2 | i_2 y) - P^1(x_1 i_1 | i_1 y) \sum_{j_2} P^2(x_2 j_2 | i_2 y) = \sum_{j_1 \neq i_1, j_2} P^1(x_1 j_1 | i_1 y) P^2(x_2 j_2 | i_2 y) \; ,$$

and:

$$\sum_{j_1,j_2} P^1(x_1j_1|i_1y)P^2(x_2j_2|i_2y) - P^2(x_2i_2|i_2y)\sum_{j_1} P^1(x_1j_1|i_1y) = \sum_{j_1,j_2\neq i_2} P^1(x_1j_1|i_1y)P^2(x_2j_2|i_2y)\,,$$

we get that $\mathrm{E}^{\mathrm{NS_{SR}}}_{\mathrm{sum}}(W,k_1,k_2,P^1,P^2)$ is equal to:

$$\begin{aligned}
&\frac{1}{k_1k_2}\sum_{i_1,i_2,x_1,x_2,y} W(y|x_1x_2)\left[\frac{\sum_{j_1\neq i_1,j_2} P^1(x_1j_1|i_1y)P^2(x_2j_2|i_2y) + \sum_{j_1,j_2\neq i_2} P^1(x_1j_1|i_1y)P^2(x_2j_2|i_2}{2}\right.\\
&= \frac{1}{k_1k_2}\sum_{i_1,i_2,x_1,x_2,y} W(y|x_1x_2)\left[\sum_{j_1\neq i_1,j_2\neq i_2} P^1(x_1j_1|i_1y)P^2(x_2j_2|i_2y) + \frac{\sum_{(j_1,j_2)\in S} P^1(x_1j_1|i_1y)P^2(x_2}{2}\right.
\end{aligned} \tag{4.9}$$

with $S := \{(j_1,i_2) : j_1 \in [k_1] - \{i_1\}\} \cup \{(i_1,j_2) : j_2 \in [k_2] - \{i_2\}\}$. However, we have that:

$$\begin{aligned}
&\sum_{j_1\neq i_1,j_2\neq i_2} P^1(x_1j_1|i_1y)P^2(x_2j_2|i_2y) + \frac{\sum_{(j_1,j_2)\in S} P^1(x_1j_1|i_1y)P^2(x_2j_2|i_2y)}{2}\\
&\leq \sum_{j_1\neq i_1,j_2\neq i_2} P^1(x_1j_1|i_1y)P^2(x_2j_2|i_2y) + \sum_{(j_1,j_2)\in S} P^1(x_1j_1|i_1y)P^2(x_2j_2|i_2y)\\
&= \sum_{(j_1,j_2)\neq(i_1,i_2)} P^1(x_1j_1|i_1y)P^2(x_2j_2|i_2y)\\
&\leq 2\left(\sum_{j_1\neq i_1,j_2\neq i_2} P^1(x_1j_1|i_1y)P^2(x_2j_2|i_2y) + \frac{\sum_{(j_1,j_2)\in S} P^1(x_1j_1|i_1y)P^2(x_2j_2|i_2y)}{2}\right)\,.
\end{aligned} \tag{4.10}$$

This implies that:

$$\mathrm{E}^{\mathrm{NS_{SR}}}_{\mathrm{sum}}(W,k_1,k_2,P^1,P^2) \leq \mathrm{E}^{\mathrm{NS_{SR}}}(W,k_1,k_2,P^1,P^2) \leq 2\mathrm{E}^{\mathrm{NS_{SR}}}_{\mathrm{sum}}(W,k_1,k_2,P^1,P^2)\,,$$

and by maximizing over all P^1 and P^2:

$$\mathrm{E}^{\mathrm{NS_{SR}}}_{\mathrm{sum}}(W,k_1,k_2) \leq \mathrm{E}^{\mathrm{NS_{SR}}}(W,k_1,k_2) \leq 2\mathrm{E}^{\mathrm{NS_{SR}}}_{\mathrm{sum}}(W,k_1,k_2)\,.$$

Thus, as before, the capacity regions are the same. □

4.2 Properties of Non-Signaling Assisted Codes

4.2.1 Symmetrization

One can prove an equivalent formulation of the linear program computing $\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2)$ with a number of variables and constraints polynomial in only $|\mathcal{X}_1|, |\mathcal{X}_2|$ and $|\mathcal{Y}|$ and independent of k_1 and k_2:

Proposition 4.4. *For a MAC $W : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}$ and $k_1, k_2 \in \mathbb{N}^*$, we have:*

$$\begin{aligned}
\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2) = \quad \underset{r, r^1, r^2, p}{\textit{maximize}} \quad & \frac{1}{k_1 k_2} \sum_{x_1, x_2, y} W(y|x_1 x_2) r_{x_1,x_2,y} \\
\textit{subject to} \quad & \sum_{x_1,x_2} r_{x_1,x_2,y} = 1 \\
& \sum_{x_1} r^1_{x_1,x_2,y} = k_1 \sum_{x_1} r_{x_1,x_2,y} \\
& \sum_{x_2} r^2_{x_1,x_2,y} = k_2 \sum_{x_2} r_{x_1,x_2,y} \\
& \sum_{x_1} p_{x_1,x_2} = k_1 \sum_{x_1} r^2_{x_1,x_2,y} \\
& \sum_{x_2} p_{x_1,x_2} = k_2 \sum_{x_2} r^1_{x_1,x_2,y} \\
& 0 \leq r_{x_1,x_2,y} \leq r^1_{x_1,x_2,y}, r^2_{x_1,x_2,y} \leq p_{x_1,x_2} \\
& p_{x_1,x_2} - r^1_{x_1,x_2,y} - r^2_{x_1,x_2,y} + r_{x_1,x_2,y} \geq 0
\end{aligned} \tag{4.11}$$

Proof. One can check that given a solution of the original program, the following choice of variables is a valid solution of the second program achieving the same objective value:

$$\begin{aligned}
r_{x_1,x_2,y} &:= \sum_{i_1,i_2} P(x_1 x_2 (i_1 i_2)|i_1 i_2 y) \ , \qquad & r^1_{x_1,x_2,y} &:= \sum_{j_1,i_1,i_2} P(x_1 x_2 (j_1 i_2)|i_1 i_2 y) \ , \\
r^2_{x_1,x_2,y} &:= \sum_{j_2,i_1,i_2} P(x_1 x_2 (i_1 j_2)|i_1 i_2 y) \ , & p_{x_1,x_2} &:= \sum_{j_1,j_2,i_1,i_2} P(x_1 x_2 (j_1 j_2)|i_1 i_2 y) \ .
\end{aligned} \tag{4.12}$$

Note that p_{x_1,x_2} is well-defined since $\sum_{j_1,j_2,i_1,i_2} P(x_1 x_2 (j_1 j_2)|i_1 i_2 y)$ is independent from y by since P is a non-signaling box.

For the other direction, given those variables, a non-signaling probability distribution $P(x_1 x_2 (j_1 j_2)|i_1 i_2 y)$ achieving the same objective value is given by, for $j_1 \neq i_1$ and $j_2 \neq i_2$:

$$\begin{aligned}
P(x_1 x_2 (i_1 i_2)|i_1 i_2 y) &:= \frac{r_{x_1,x_2,y}}{k_1 k_2} \ , \\
P(x_1 x_2 (j_1 i_2)|i_1 i_2 y) &:= \frac{r^1_{x_1,x_2,y} - r_{x_1,x_2,y}}{k_1 k_2 (k_1 - 1)} \ , \\
P(x_1 x_2 (i_1 j_2)|i_1 i_2 y) &:= \frac{r^2_{x_1,x_2,y} - r_{x_1,x_2,y}}{k_1 k_2 (k_2 - 1)} \ , \\
P(x_1 x_2 (j_1 j_2)|i_1 i_2 y) &:= \frac{p_{x_1,x_2} - r^1_{x_1,x_2,y} - r^2_{x_1,x_2,y} + r_{x_1,x_2,y}}{k_1 k_2 (k_1 - 1)(k_2 - 1)} \ .
\end{aligned} \tag{4.13}$$

□

This symmetrization can also be done for the program computing $\mathrm{S}^{\mathrm{NS}_{\mathrm{SR}}}_{\mathrm{sum}}(W, k_1, k_2)$:

Proposition 4.5.

$$
\begin{aligned}
\mathrm{S}^{\mathrm{NS}_{\mathrm{SR}}}_{sum}(W,k_1,k_2) = \quad & \underset{r^1,r^2,p^1,p^2}{\textit{maximize}} && \frac{1}{2k_1k_2}\sum_{x_1,x_2,y} W(y|x_1x_2)\left(p^2_{x_2}r^1_{x_1,y} + p^1_{x_1}r^2_{x_2,y}\right)\\
& = && \frac{1}{2}\left[\frac{1}{k_1}\sum_{x_1,y} W^1_{p^2,k_2}(y|x_1)r^1_{x_1,y} + \frac{1}{k_2}\sum_{x_2,y} W^2_{p^1,k_1}(y|x_2)r^2_{x_2,y}\right]\\
& \textit{with} && W^1_{p^2,k_2}(y|x_1) := \frac{1}{k_2}\sum_{x_2} W(y|x_1x_2)p^2_{x_2}\\
& \textit{and} && W^2_{p^1,k_1}(y|x_2) := \frac{1}{k_1}\sum_{x_1} W(y|x_1x_2)p^1_{x_1}\\
& \textit{subject to} && \sum_{x_1} r^1_{x_1,y} = 1, \sum_{x_2} r^2_{x_2,y} = 1\\
& && \sum_{x_1} p^1_{x_1} = k_1, \sum_{x_2} p^2_{x_2} = k_2\\
& && 0 \le r^1_{x_1,y} \le p^1_{x_1}, 0 \le r^2_{x_2,y} \le p^2_{x_2}
\end{aligned}
\tag{4.14}
$$

Proof. One can check that given a solution of the original program, the following choice of variables is a valid solution of the second program achieving the same objective value:

$$
\begin{aligned}
r^1_{x_1,y} &:= \sum_{i_1} P^1(x_1i_1|i_1y)\ , \quad p^1_{x_1} := \sum_{j_1,i_1} P^1(x_1j_1|i_1y)\ ,\\
r^2_{x_2,y} &:= \sum_{i_2} P^2(x_2i_2|i_2y)\ , \quad p^2_{x_2} := \sum_{j_2,i_2} P^2(x_2j_2|i_2y)\ .
\end{aligned}
\tag{4.15}
$$

Note that $p^1_{x_1}$ and $p^2_{x_2}$ are well-defined since $\sum_{j_1,i_1} P^1(x_1j_1|i_1y)$ and $\sum_{j_2,i_2} P^2(x_2j_2|i_2y)$ are independent from y since P^1 and P^2 are non-signaling boxes.

For the other direction, given those variables, non-signaling probability distributions $P^1(x_1j_1|i_1y)$ and $P^2(x_2j_2|i_2y)$ achieving the same objective value are given by, for $j_1 \neq i_1$ and $j_2 \neq i_2$:

$$
\begin{aligned}
P^1(x_1i_1|i_1y) &:= \frac{r^1_{x_1,y}}{k_1}\ ,\\
P^1(x_1j_1|i_1y) &:= \frac{p^1_{x_1,y} - r^1_{x_1,y}}{k_1(k_1-1)}\ ,\\
P^2(x_2i_2|i_2y) &:= \frac{r^2_{x_2,y}}{k_2}\ ,\\
P^2(x_2j_2|i_2y) &:= \frac{p^2_{x_2,y} - r^2_{x_2,y}}{k_2(k_2-1)}\ .
\end{aligned}
\tag{4.16}
$$

□

4.2.2 Properties of $\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2)$, $\mathcal{C}^{\mathrm{NS}}(W)$ and $\mathcal{C}^{\mathrm{NS}}_0(W)$

Definition 4.7. We say that a conditional probability distribution $Q(a^n|x^n)$ defined on $(\mathcal{A}_1 \times \ldots \times \mathcal{A}_n) \times (\mathcal{X}_1 \times \ldots \times \mathcal{X}_n)$ and $Q'(a'^n|x'^n)$ is *non-signaling* if for all $a^n, x^n, \hat{x}^n$, we have

$$\forall i \in [n], \sum_{a_i} Q(a_1 \dots a_i \dots a_n | x_1 \dots x_i \dots x_n) = \sum_{a_i} Q(a_1 \dots a_i \dots a_n | x_1 \dots \hat{x}_i \dots x_n) \ .$$

Definition 4.8. Let $Q(a^n|x^n)$ be a conditional probability distribution defined on $(\mathcal{A}_1 \times \ldots \times \mathcal{A}_n) \times (\mathcal{X}_1 \times \ldots \times \mathcal{X}_n)$ and $Q'(a'^n|x'^n)$ defined on $(\mathcal{A}'_1 \times \ldots \times \mathcal{A}'_n) \times (\mathcal{X}'_1 \times \ldots \times \mathcal{X}'_n)$. We define $P := Q \otimes Q'$ the tensor product conditional probability distribution defined on $((\mathcal{A}_1 \times \mathcal{A}'_1) \times \ldots \times (\mathcal{A}_n \times \mathcal{A}'_n)) \times ((\mathcal{X}_1 \times \mathcal{X}'_1) \times \ldots \times (\mathcal{X}_n \times \mathcal{X}'_n))$ by $P(a_1a'_1 \dots a_na'_n | x_1x'_1 \dots x_nx'_n) := Q(a^n|x^n) \cdot Q'(a'^n|x'^n)$.

Lemma 4.6. *If both Q and Q' are non-signaling, then $P = Q \otimes Q'$ is non-signaling.*

Proof. Let $a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n \in \mathcal{A}_1 \times \ldots \times \mathcal{A}_{i-1} \times \mathcal{A}_{i+1} \times \ldots \times \mathcal{A}_n$, $a'_1, \dots, a'_{i-1}, a'_{i+1}, \dots, a'_n \in \mathcal{A}'_1 \times \ldots \times \mathcal{A}'_{i-1} \times \mathcal{A}'_{i+1} \times \ldots \times \mathcal{A}'_n$, $x_1, \dots, x_n \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$ and $x'_1, \dots, x'_n \in \mathcal{X}'_1 \times \ldots \times \mathcal{X}'_n$. We have:

$$\begin{aligned}
&\sum_{a_ia'_i} P(a_1a'_1 \dots a_ia'_i \dots a_na'_n | x_1x'_1 \dots x_ix'_i \dots x_nx'_n) \\
&= \sum_{a_ia'_i} Q(a_1 \dots a_i \dots a_n | x_1 \dots x_i \dots x_n) \cdot Q'(a'_1 \dots a'_i \dots a'_n | x'_1 \dots x'_i \dots x'_n) \\
&= \left(\sum_{a_i} Q(a_1 \dots a_i \dots a_n | x_1 \dots x_i \dots x_n) \right) \cdot \left(\sum_{a_i} Q'(a'_1 \dots a'_i \dots a'_n | x'_1 \dots x'_i \dots x'_n) \right) \\
&= \left(\sum_{a_i} Q(a_1 \dots a_i \dots a_n | x_1 \dots \hat{x}_i \dots x_n) \right) \cdot \left(\sum_{a_i} Q'(a'_1 \dots a'_i \dots a'_n | x'_1 \dots \hat{x}'_i \dots x'_n) \right) \\
&\quad \text{since } Q \text{ and } Q' \text{ are non-signaling} \\
&\sum_{a_ia'_i} P(a_1a'_1 \dots a_ia'_i \dots a_na'_n | x_1x'_1 \dots \hat{x}_i\hat{x}'_i \dots x_nx'_n) \ ,
\end{aligned} \tag{4.17}$$

so P is non-signaling. ☐

Proposition 4.7. *For a MAC $W : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}$ and $k_1, k_2 \in \mathbb{N}^*$, we have:*

1. $\frac{1}{k_1k_2} \leq \mathrm{S}^{\mathrm{NS}}(W, k_1, k_2) \leq 1$.
2. $\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2) \leq \min\left(\frac{|\mathcal{X}_1|}{k_1}, \frac{|\mathcal{X}_2|}{k_2}, \frac{|\mathcal{Y}|}{k_1k_2}\right)$.
3. *If $k'_1 \leq k_1$ and $k'_2 \leq k_2$, then $\mathrm{S}^{\mathrm{NS}}(W, k'_1, k'_2) \geq \mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)$.*
4. *For any MAC $W' : \mathcal{X}'_1 \times \mathcal{X}'_2 \to \mathcal{Y}'$ and $k_1, k_2 \in \mathbb{N}^*$, we have $\mathrm{S}^{\mathrm{NS}}(W \otimes W', k_1k'_1, k_2k'_2) \geq \mathrm{S}^{\mathrm{NS}}(W, k_1, k_2) \cdot \mathrm{S}^{\mathrm{NS}}(W', k'_1, k'_2)$. In particular, for any positive integer n, $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, k_1^n, k_2^n) \geq \left[\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)\right]^n$ and $\mathrm{S}^{\mathrm{NS}}(W \otimes W', k_1, k_2) \geq \mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)$.*

Proof. 1. Let us first show that $\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2) \geq \frac{1}{k_1k_2}$. Take $p_{x_1,x_2} := \frac{k_1k_2}{|\mathcal{X}_1||\mathcal{X}_2|}$, $r^1_{x_1,x_2,y} := \frac{p_{x_1,x_2}}{k_2}$, $r^2_{x_1,x_2,y} := \frac{p_{x_1,x_2}}{k_1}$ and $r_{x_1,x_2,y} := \frac{p_{x_1,x_2}}{k_1k_2} = \frac{1}{|\mathcal{X}_1||\mathcal{X}_2|}$. One can easily check that it is indeed a valid solution of the linear program computing $\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2)$. Thus we have:

$$\begin{aligned}\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2) &\geq \frac{1}{k_1k_2}\sum_{x_1,x_2,y} W(y|x_1x_2)r_{x_1,x_2,y} = \frac{1}{k_1k_2}\sum_{x_1,x_2}\frac{1}{|\mathcal{X}_1||\mathcal{X}_2|}\sum_y W(y|x_1x_2)\\ &= \frac{1}{k_1k_2}\sum_{x_1,x_2}\frac{1}{|\mathcal{X}_1||\mathcal{X}_2|} = \frac{1}{k_1k_2}\ .\end{aligned} \tag{4.18}$$

Furthermore, in order to show that it is at most 1, let us consider an optimal solution of $\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2)$. We have:

$$\begin{aligned}\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2) &= \frac{1}{k_1k_2}\sum_{x_1,x_2,y} W(y|x_1x_2)r_{x_1,x_2,y} \leq \frac{1}{k_1k_2}\sum_{x_1,x_2,y} W(y|x_1x_2)p_{x_1,x_2}\\ &= \frac{1}{k_1k_2}\sum_{x_1,x_2}p_{x_1,x_2}\sum_y W(y|x_1x_2) = \frac{1}{k_1k_2}\sum_{x_1,x_2}p_{x_1,x_2} = 1\ ,\end{aligned} \tag{4.19}$$

since $\sum_{x_1,x_2}p_{x_1,x_2} = k_1\sum_{x_1,x_2}r^2_{x_1,x_2,y} = k_1k_2\sum_{x_1,x_2}r_{x_1,x_2,y} = k_1k_2$.

2. First let us show that $\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2) \leq \frac{|\mathcal{X}_1|}{k_1}$ (the case $\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2) \leq \frac{|\mathcal{X}_2|}{k_2}$ is symmetric):

$$\begin{aligned}\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2) &= \frac{1}{k_1k_2}\sum_{x_1,x_2,y} W(y|x_1x_2)r_{x_1,x_2,y} \leq \frac{1}{k_1k_2}\sum_{x_1,x_2,y} W(y|x_1x_2)r^2_{x_1,x_2,y}\\ &\leq \frac{1}{k_1k_2}\sum_{x_2,y}\left(\sum_{x_1'}W(y|x_1'x_2)\right)\cdot\left(\sum_{x_1}r^2_{x_1,x_2,y}\right) \quad \text{since nonnegative terms.}\\ &= \frac{1}{k_1k_2}\sum_{x_2,y}\left(\sum_{x_1'}W(y|x_1'x_2)\right)\cdot\left(\frac{1}{k_1}\sum_{x_1}p_{x_1,x_2}\right)\\ &= \frac{1}{k_1^2k_2}\sum_{x_1,x_2}p_{x_1,x_2}\sum_{x_1'}\left(\sum_y W(y|x_1'x_2)\right) = \frac{|\mathcal{X}_1|}{k_1^2k_2}\sum_{x_1,x_2}p_{x_1,x_2} = \frac{|\mathcal{X}_1|}{k_1}\ .\end{aligned} \tag{4.20}$$

Let us show now that $\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2) \leq \frac{|\mathcal{Y}|}{k_1k_2}$:

$$\begin{aligned}\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2) &= \frac{1}{k_1k_2}\sum_{x_1,x_2,y} W(y|x_1x_2)r_{x_1,x_2,y} \leq \frac{1}{k_1k_2}\sum_y\left(\max_{x_1,x_2}W(y|x_1x_2)\right)\sum_{x_1,x_2}r_{x_1,x_2,y}\\ &\leq \frac{1}{k_1k_2}\sum_y\sum_{x_1,x_2}r_{x_1,x_2,y} = \frac{|\mathcal{Y}|}{k_1k_2}\ .\end{aligned} \tag{4.21}$$

3. Let us assume that $k_1' \leq k_1$ and that $k_2' = k_2$, since this latter case will follow by symmetry. Consider an optimal solution of $\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2) = \frac{1}{k_1}\sum_{i_1\in[k_1]}f(i_1)$ with:

$$f(i_1) := \frac{1}{k_2}\sum_{x_1,x_2,y,i_2} W(y|x_1x_2)P(x_1x_2(i_1i_2)|i_1i_2y)\ ,$$

and P non-signaling. Let us consider $S \in \underset{S' \subseteq [k_1]:|S'|=k'_1}{\operatorname{argmax}} \sum_{i_1 \in S'} f(i_1)$. Then, by construction, we have that $\frac{1}{k'_1} \sum_{i_1 \in S} f(i_1) \geq \frac{1}{k_1} \sum_{i_1 \in [k_1]} f(i_1) = \mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)$, since we have taken the average of the k'_1 largest values of the sum.

Let us define the strategy P' on the smallest set $\mathcal{X}_1 \times \mathcal{X}_2 \times (S \times [k_2]) \times S \times [k_2] \times \mathcal{Y}$:

$$
\begin{aligned}
&P'(x_1x_2(j_1j_2)|i_1i_2y) := P(x_1x_2(j_1j_2)|i_1i_2y) + C(x_1x_2j_2|i_1i_2y) \ , \\
&\text{with } C(x_1x_2j_2|i_1i_2y) := \frac{1}{k'_1} \sum_{j'_1 \in [k_1]-S} P(x_1x_2(j'_1j_2)|i_1i_2y) \ .
\end{aligned}
\tag{4.22}
$$

P' is a correct conditional probability distribution. Indeed, it is nonnegative by construction, and we have that:

$$
\begin{aligned}
&\sum_{x_1,x_2,j_1 \in S, j_2} P'(x_1x_2(j_1j_2)|i_1i_2y) = \sum_{x_1,x_2,j_1 \in S, j_2} P(x_1x_2(j_1j_2)|i_1i_2y) + \sum_{x_1,x_2,j_1 \in S, j_2} C(x_1x_2j_2|i_1i_2y) \\
&= \sum_{x_1,x_2j_2} \sum_{j_1 \in S} P(x_1x_2(j_1j_2)|i_1i_2y) + \sum_{x_1,x_2,j_2} \sum_{j_1 \in S} \frac{1}{k'_1} \sum_{j'_1 \in [k_1]-S} P(x_1x_2(j'_1j_2)|i_1i_2y) \\
&= \sum_{x_1,x_2,j_2} \sum_{j_1 \in S} P(x_1x_2(j_1j_2)|i_1i_2y) + \sum_{x_1,x_2,j_2} \sum_{j'_1 \in [k_1]-S} P(x_1x_2(j'_1j_2)|i_1i_2y) \\
&= \sum_{x_1,x_2,j_1,j_2} P(x_1x_2(j_1j_2)|i_1i_2y) = 1 \ .
\end{aligned}
\tag{4.23}
$$

Let us show that P' is non-signaling:

a) First with x_1:

$$
\begin{aligned}
\sum_{x_1} P'(x_1x_2(j_1j_2)|i_1i_2y) &= \sum_{x_1} P(x_1x_2(j_1j_2)|i_1i_2y) + \sum_{x_1} C(x_1x_2j_2|i_1i_2y) \\
&= \sum_{x_1} P(x_1x_2(j_1j_2)|i_1i_2y) + \frac{1}{k'_1} \sum_{j'_1 \in [k_1]-S} \sum_{x_1} P(x_1x_2(j'_1j_2)|i_1i_2y) \\
&= \sum_{x_1} P(x_1x_2(j_1j_2)|i'_1i_2y) + \frac{1}{k'_1} \sum_{j'_1 \in [k_1]-S} \sum_{x_1} P(x_1x_2(j'_1j_2)|i'_1i_2y) \\
&\text{since } P \text{ is non-signaling.} \\
&= \sum_{x_1} P'(x_1x_2(j_1j_2)|i'_1i_2y) \ .
\end{aligned}
\tag{4.24}
$$

b) Then with x_2:

$$\begin{aligned}\sum_{x_2} P'(x_1x_2(j_1j_2)|i_1i_2y) &= \sum_{x_2} P(x_1x_2(j_1j_2)|i_1i_2y) + \sum_{x_2} C(x_1x_2j_2|i_1i_2y)\\ &= \sum_{x_2} P(x_1x_2(j_1j_2)|i_1i_2y) + \frac{1}{k_1'} \sum_{j_1' \in [k_1]-S} \sum_{x_2} P(x_1x_2(j_1'j_2)|i_1i_2y)\\ &= \sum_{x_2} P(x_1x_2(j_1j_2)|i_1i_2'y) + \frac{1}{k_1'} \sum_{j_1' \in [k_1]-S} \sum_{x_2} P(x_1x_2(j_1'j_2)|i_1i_2'y)\\ &\text{since } P \text{ is non-signaling.}\\ &= \sum_{x_2} P'(x_1x_2(j_1j_2)|i_1i_2'y)\ .\end{aligned} \tag{4.25}$$

c) Finally with (j_1j_2):

$$\begin{aligned}\sum_{j_1 \in S, j_2} P'(x_1x_2(j_1j_2)|i_1i_2y) &= \sum_{j_2}\sum_{j_1 \in S} P(x_1x_2(j_1j_2)|i_1i_2y) + \sum_{j_2}\sum_{j_1 \in S} C(x_1x_2j_2|i_1i_2y)\\ = \sum_{j_2}\sum_{j_1 \in S} P(x_1x_2(j_1j_2)|i_1i_2y) &+ \sum_{j_2}\sum_{j_1 \in S} \frac{1}{k_1'} \sum_{j_1' \in [k_1]-S} P(x_1x_2(j_1'j_2)|i_1i_2y)\\ = \sum_{j_2}\sum_{j_1 \in S} P(x_1x_2(j_1j_2)|i_1i_2y) &+ \sum_{j_2} \sum_{j_1' \in [k_1]-S} P(x_1x_2(j_1'j_2)|i_1i_2y)\\ = \sum_{j_1,j_2} P(x_1x_2(j_1j_2)|i_1i_2y)&\\ = \sum_{j_1,j_2} P(x_1x_2(j_1j_2)|i_1i_2y') &\text{ since } P \text{ is non-signaling.}\\ = \sum_{j_1 \in S, j_2} P'(x_1x_2(j_1j_2)|i_1i_2y')\ .&\end{aligned} \tag{4.26}$$

Thus P' is a correct solution of the program computing $\mathrm{S}^{\mathrm{NS}}(W, k_1', k_2)$, and it leads to the value:

$$\begin{aligned}\mathrm{S}^{\mathrm{NS}}(W, k_1', k_2) &\geq \frac{1}{k_1'k_2} \sum_{x_1,x_2,y,i_1 \in S,i_2} W(y|x_1x_2)P'(x_1x_2(i_1i_2)|i_1i_2y)\\ &\geq \frac{1}{k_1'k_2} \sum_{x_1,x_2,y,i_1 \in S,i_2} W(y|x_1x_2)P(x_1x_2(i_1i_2)|i_1i_2y)\\ &= \frac{1}{k_1'} \sum_{i_1 \in S} f(i_1) \geq \frac{1}{k_1} \sum_{i_1 \in [k_1]} f(i_1) = \mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)\ .\end{aligned} \tag{4.27}$$

4. Consider optimal non-signaling probability distributions P and P' reaching respectively the values $\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)$ and $\mathrm{S}^{\mathrm{NS}}(W', k_1', k_2')$. Then by Lemma 4.6, $P \otimes P'$ is a non-signaling probability distribution on $(\mathcal{X}_1 \times \mathcal{X}_1') \times (\mathcal{X}_2 \times \mathcal{X}_2') \times (([k_1] \times [k_1']) \times ([k_2] \times [k_2'])) \times ([k_1] \times [k_1']) \times ([k_2] \times [k_2']) \times (\mathcal{Y} \times \mathcal{Y}')$, which is trivially in bijection with $(\mathcal{X}_1 \times \mathcal{X}_1') \times (\mathcal{X}_2 \times \mathcal{X}_2') \times ([k_1k_1'] \times [k_2k_2']) \times [k_1k_1'] \times [k_2k_2'] \times (\mathcal{Y} \times \mathcal{Y}')$. This gives a valid solution of the program computing $\mathrm{S}^{\mathrm{NS}}(W \otimes W', k_1k_1', k_2k_2')$. Thus, we get that $\mathrm{S}^{\mathrm{NS}}(W \otimes W', k_1k_1', k_2k_2')$ is larger than or equal

to:

$$\begin{aligned}
&\sum_{x_1x_1',x_2x_2',yy',i_1i_1',i_2i_2'} (W \otimes W')\left(yy'|x_1x_1'x_2x_2'\right)(P \otimes P')\left(x_1x_1'x_2x_2'(i_1i_1'i_2i_2')|i_1i_1',i_2i_2'yy'\right)\\
&= \sum_{x_1x_1',x_2x_2',yy',i_1i_1',i_2i_2'} \left(W(y|x_1x_2) \cdot W'(y'|x_1'x_2')\right)\left(P(x_1x_2(i_1i_2)|i_1i_2y) \cdot P'(x_1'x_2'(i_1'i_2')|i_1'i_2'y')\right)\\
&= \left(\sum_{i_1,i_2,x_1,x_2,y} W(y|x_1x_2)P(x_1x_2(i_1i_2)|i_1i_2y)\right) \cdot \left(\sum_{x_1',x_2',y',i_1',i_2'} W'(y'|x_1'x_2')P'(x_1'x_2'(i_1'i_2')|i_1'i_2'y')\right)\\
&= \mathrm{S}^{\mathrm{NS}}(W,k_1,k_2) \cdot \mathrm{S}^{\mathrm{NS}}(W',k_1',k_2') \ .
\end{aligned} \tag{4.28}$$

In particular, applying this n times on the same MAC W gives the first corollary, and the second one comes from the fact that $\mathrm{S}^{\mathrm{NS}}(W \otimes W',k_1,k_2) \geq \mathrm{S}^{\mathrm{NS}}(W,k_1,k_2) \cdot \mathrm{S}^{\mathrm{NS}}(W',1,1) = \mathrm{S}^{\mathrm{NS}}(W,k_1,k_2)$, since $\mathrm{S}^{\mathrm{NS}}(W',1,1) = 1$ by the first property of Proposition 4.7.

□

Corollary 4.8. *1. $\mathcal{C}^{\mathrm{NS}}(W)$ is convex.*

2. If (R_1,R_2) is achievable with non-signaling assistance, then $R_1 \leq \log_2|\mathcal{X}_1|$, $R_2 \leq \log_2|\mathcal{X}_2|$ and $R_1+R_2 \leq \log_2|\mathcal{Y}|$.

3. If (R_1,R_2) is achievable with non-signaling assistance, then for all $R_i' \leq R_i$, (R_1',R_2') is achievable with non-signaling assistance.

Proof. 1. Let (R_1,R_2) and $(\tilde{R}_1,\tilde{R}_2)$, two pairs of rational rates achievable with non-signaling assistance for W, ie:

$$\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{R_1n}\rceil, \lceil 2^{R_2n}\rceil) \underset{n\to+\infty}{\to} 1 \text{ and } \mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{\tilde{R}_1n}\rceil, \lceil 2^{\tilde{R}_2n}\rceil) \underset{n\to+\infty}{\to} 1 \ .$$

Let $\lambda \in (0,1)$ rational and define $R_{\lambda,i} := \lambda \cdot R_i + (1-\lambda)\cdot\tilde{R}_i$, let us show that $(R_{\lambda,1},R_{\lambda,2})$ is achievable with non-signaling assistance. Let us call respectively $k_i := 2^{R_i}$, $\tilde{k}_i := 2^{\tilde{R}_i}$, $k_{\lambda,i} := 2^{R_{\lambda,i}} = k_i^{\lambda} \cdot k_i^{(1-\lambda)}$.

We have $R_{\lambda,i}n = \lambda \cdot R_in + (1-\lambda)\cdot\tilde{R}_in = (\lambda n)\cdot R_i + (1-\lambda)n\cdot\tilde{R}_i$. This is the idea of *time-sharing*: for λn copies of the MAC, we use the strategy with rate (R_1,R_2) and for the $(1-\lambda)n$ other copies of the MAC, we use the strategy with rate $(\tilde{R}_1,\tilde{R}_2)$. There exists some n such that $\lambda n, (1-\lambda)n, \lambda nR_i, (1-\lambda)n\tilde{R}_i$ are integers, since everything is rational. This implies that $k_i^{\lambda n}, \tilde{k}_i^{(1-\lambda)n}, k_{\lambda,i}^n$ are integers. Thus, thanks to the fourth property of Proposition 4.7, we have:

$$\begin{aligned}
\mathrm{S}^{\mathrm{NS}}(W^{\otimes n},k_{\lambda,1}^n,k_{\lambda,2}^n) &\geq \mathrm{S}^{\mathrm{NS}}(W^{\otimes(\lambda n)},k_1^{\lambda n},k_2^{\lambda n}) \cdot \mathrm{S}^{\mathrm{NS}}(W^{\otimes((1-\lambda)n)},\tilde{k}_1^{(1-\lambda)n},\tilde{k}_2^{(1-\lambda)n})\\
&\underset{n\to+\infty}{\to} 1\cdot 1 = 1 \ .
\end{aligned} \tag{4.29}$$

Thus in particular, since we have $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n},k_{\lambda,1}^n,k_{\lambda,2}^n) \leq 1$, we get that $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n},k_{\lambda,1}^n,k_{\lambda,2}^n) \underset{n\to+\infty}{\to} 1$, so $(R_{\lambda,1},R_{\lambda,2})$ is achievable with non-signaling assistance. Finally, since $\mathcal{C}^{\mathrm{NS}}(W)$ is defined as the closure of achievable rates with non-signaling assistance, we get that $\mathcal{C}^{\mathrm{NS}}(W)$ is convex.

2. By the second property of Proposition 4.7, we have that $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, k_1^n, k_2^n) \leq \frac{|\mathcal{X}_1^n|}{k_1^n}$. In particular, if one takes $R_1 > \log_2 |\mathcal{X}_1|$, then $k_1 > |\mathcal{X}_1|$ and we get that $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, k_1^n, k_2^n) \leq \left(\frac{|\mathcal{X}_1|}{k_1}\right)^n \underset{n\to+\infty}{\to} 0$, so $R_1 > \log_2 |\mathcal{X}_1|$ is not achievable with non-signaling assistance. Symmetrically, $R_2 > \log_2 |\mathcal{X}_2|$ is not achievable with non-signaling assistance.

   Furthermore, if one takes $R_1+R_2 > \log_2 |\mathcal{Y}|$, then in particular $k_1k_2 > |\mathcal{Y}|$, so by the second property of Proposition 4.7, $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, k_1^n, k_2^n) \leq \frac{|\mathcal{Y}^n|}{k_1^n k_2^n} = \left(\frac{|\mathcal{Y}|}{k_1 k_2}\right)^n \underset{n\to+\infty}{\to} 0$. Thus, $R_1 + R_2 > \log_2 |\mathcal{Y}|$ is not achievable with non-signaling assistance.

3. Since (R_1, R_2) is achievable with non-signaling assistance, we have $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil) \underset{n\to+\infty}{\to}$ 1. But, for all positive integer n, we have that $\lceil 2^{nR_1'} \rceil \leq \lceil 2^{nR_1} \rceil$ and $\lceil 2^{nR_2'} \rceil \leq \lceil 2^{nR_2} \rceil$, so by the third property of Proposition 4.7, we have that $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{nR_1'} \rceil, \lceil 2^{nR_2'} \rceil) \geq \mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{nR_1} \rceil, \lceil 2^{nR_2} \rceil)$. Thus $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{nR_1'} \rceil, \lceil 2^{nR_2'} \rceil \underset{n\to+\infty}{\to} 1$ since it is upper bounded by 1, and so (R_1', R_2') is achievable with non-signaling assistance.

□

Proposition 4.9. *$\mathcal{C}_0^{\mathrm{NS}}(W)$ is the closure of the set of rate pairs (R_1, R_2) such that:*

$$\exists n \in \mathbb{N}^*, \mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil) = 1 \ .$$

Proof. It is clear that if (R_1, R_2) is such that $\exists n_0 \in \mathbb{N}^*, \forall n \geq n_0, \mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil) = 1$, then in particular $\exists n \in \mathbb{N}^*, \mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil) = 1$. So, $\mathcal{C}_0^{\mathrm{NS}}(W)$, which is the closure of the former rate pairs, is in particular included in the closure of the latter rate pairs.

For the other inclusion, consider a rate pair (R_1, R_2) and let us assume that there exists some positive integer n such that $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil) = 1$. Let us show that for any (R_1', R_2') such that $R_1' < R_1$ and $R_2' < R_2$:

$$\exists n_0 \in \mathbb{N}^*, \forall n \geq n_0, \mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{R_1' n} \rceil, \lceil 2^{R_2' n} \rceil) = 1 \ ,$$

which is enough to conclude, since we consider only closure of such sets.

First, for all positive integer m, we have that $\mathrm{S}^{\mathrm{NS}}(W^{\otimes nm}, \lceil 2^{R_1 nm} \rceil, \lceil 2^{R_2 nm} \rceil) = 1$. By the fourth property of Proposition 4.7, we have that $\mathrm{S}^{\mathrm{NS}}((W^{\otimes n})^{\otimes m}, \lceil 2^{R_1 n} \rceil^m, \lceil 2^{R_2 n} \rceil^m) \geq \left[\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil)\right]^m = 1$, so $\mathrm{S}^{\mathrm{NS}}((W^{\otimes n})^{\otimes m}, \lceil 2^{R_1 n} \rceil^m, \lceil 2^{R_2 n} \rceil^m) = 1$ since $\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2) \leq 1$ by the first property of Proposition 4.7. But $(W^{\otimes n})^{\otimes m} = W^{\otimes nm}$, and $\lceil 2^{R_1 n} \rceil^m \geq \lceil 2^{R_1 nm} \rceil, \lceil 2^{R_2 n} \rceil^m \geq \lceil 2^{R_2 nm} \rceil$, so by the third property of Proposition 4.7, we have $\mathrm{S}^{\mathrm{NS}}(W^{\otimes nm}, \lceil 2^{R_1 nm} \rceil, \lceil 2^{R_2 nm} \rceil) \geq 1$, so $\mathrm{S}^{\mathrm{NS}}(W^{\otimes nm}, \lceil 2^{R_1 nm} \rceil, \lceil 2^{R_2 nm} \rceil) = 1$.

Then, consider some $r \in \{0, \dots, n-1\}$. By the fourth property of Proposition 4.7, we have that:

$$\begin{aligned} \mathrm{S}^{\mathrm{NS}}(W^{\otimes (nm+r)}, \lceil 2^{R_1 nm} \rceil, \lceil 2^{R_2 nm} \rceil) &= \mathrm{S}^{\mathrm{NS}}(W^{\otimes nm} \otimes W^{\otimes r}, \lceil 2^{R_1 nm} \rceil, \lceil 2^{R_2 nm} \rceil) \\ &\geq \mathrm{S}^{\mathrm{NS}}(W^{\otimes nm}, \lceil 2^{R_1 nm} \rceil, \lceil 2^{R_2 nm} \rceil) = 1 \ , \end{aligned} \tag{4.30}$$

so $\mathrm{S}^{\mathrm{NS}}(W^{\otimes (nm+r)}, \lceil 2^{R_1 nm} \rceil, \lceil 2^{R_2 nm} \rceil) = 1$. But $\lceil 2^{R_1 nm} \rceil = \lceil 2^{\frac{R_1 nm}{nm+r}(nm+r)} \rceil = \lceil 2^{\frac{R_1}{1+\delta}(nm+r)} \rceil$ with $\delta = \frac{r}{nm} \leq \frac{1}{m}$, and symmetrically $\lceil 2^{R_1 nm} \rceil = \lceil 2^{\frac{R_1}{1+\delta}(nm+r)} \rceil$. Thus in particular, for all

$R'_1 \leq \frac{R_1}{1+\frac{1}{m}}$ and $R'_2 \leq \frac{R_2}{1+\frac{1}{m}}$, we have that for all $n' \geq nm, \mathrm{S}^{\mathrm{NS}}(W^{\otimes n'}, \lceil 2^{R'_1 n'} \rceil, \lceil 2^{R'_2 n'} \rceil) = 1$. So for any (R'_1, R'_2) such that $R'_1 < R_1$ and $R'_2 < R_2$, there is large enough m such that $R'_1 \leq \frac{R_1}{1+\frac{1}{m}}$ and $R'_2 \leq \frac{R_2}{1+\frac{1}{m}}$, and thus we get the expected property on (R'_1, R'_2) for $n_0 := nm$. □

4.2.3 Linear Program with Reduced Size for Structured Channels

Although $\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)$ can be computed in polynomial time in W, k_1 and k_2, a channel of the form $W^{\otimes n}$ has exponential size in n. Thus, the linear program for $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, k_1, k_2)$ grows exponentially with n. However, using the invariance of $W^{\otimes n}$ under permutations, one can find a much smaller linear program computing $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, k_1, k_2)$.

Definition 4.9. Let G a group acting on $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}$. We say that a MAC $W : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}$ is *invariant under the action of* G if:

$$\forall g \in G, W(g \cdot y | g \cdot x_1 g \cdot x_2) = W(y|x_1 x_2) \ .$$

In particular, for $W^{\otimes n} : \mathcal{X}_1^n \times \mathcal{X}_2^n \to \mathcal{Y}^n$, the symmetric group $G := S_n$ acts in a natural way in any set $\mathcal{A}$ raised to power n. So for $\sigma \in S_n$, we have that:

$$W^{\otimes n}(\sigma \cdot y^n | \sigma \cdot x_1^n \sigma \cdot x_2^n) = \prod_{i=1}^{n} W(y_{\sigma(i)} | x_{1,\sigma(i)} x_{2,\sigma(i)}) = \prod_{i=1}^{n} W(y_i | x_{1,i} x_{2,i}) = W^{\otimes n}(y^n | x_1^n x_2^n) \ ,$$

and so $W^{\otimes n}$ is invariant under the action of S_n.

Let $\mathcal{Z} := \{\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, \mathcal{X}_1 \times \mathcal{Y}, \mathcal{X}_2 \times \mathcal{Y}, \mathcal{X}_1 \times \mathcal{X}_2, \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}\}$. Let us call $\mathcal{O}_G(\mathcal{A})$ the set of orbits of $\mathcal{A}$ under the action of G. Then, one can find an equivalent smaller linear program for $\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)$:

Theorem 4.10. *Let* $W : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}$ *a MAC invariant under the action of* G. *Let us name systematically* $w \in \mathcal{O}_G(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}), u \in \mathcal{O}_G(\mathcal{X}_1 \times \mathcal{X}_2), u^1 \in \mathcal{O}_G(\mathcal{X}_1), u^2 \in \mathcal{O}_G(\mathcal{X}_2), v^1 \in \mathcal{O}_G(\mathcal{X}_1 \times \mathcal{Y}), v^2 \in \mathcal{O}_G(\mathcal{X}_2 \times \mathcal{Y}), v \in \mathcal{O}_G(\mathcal{Y})$. *We will also call* $z_{\mathcal{A}}$ *the projection of* $z \in \mathcal{O}_G(\mathcal{B})$ *on* $\mathcal{A}$, *for* $\mathcal{A}, \mathcal{B} \in \mathcal{Z}$ *and* $\mathcal{A}$ *projection of* $\mathcal{B}$; *note that* $z_{\mathcal{A}} \in \mathcal{O}_G(\mathcal{A})$, *since by definition of the action, the projection of an orbit is an orbit. Let us finally call* $W(w) := W(y|x_1 x_2)$ *for any* $(x_1, x_2, y) \in w$, *which is well-defined since* W *is invariant* *under G. We have that $\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2)$ is the solution of the following linear program:*

$$\begin{aligned}
\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2) = \quad \underset{r,r^1,r^2,p}{\textit{maximize}} \quad & \frac{1}{k_1k_2} \sum_{w \in \mathcal{O}_G(\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{Y})} W(w) r_w \\
\textit{subject to} \quad & \sum_{w : w_{\mathcal{Y}} = v} r_w = |v|, \forall v \in \mathcal{O}_G(\mathcal{Y}) \\
& \sum_{w : w_{\mathcal{X}_2\mathcal{Y}} = v^2} r^1_w = k_1 \sum_{w : w_{\mathcal{X}_2\mathcal{Y}} = v^2} r_w,\ \forall v^2 \in \mathcal{O}_G(\mathcal{X}_2 \times \mathcal{Y}) \\
& \sum_{w : w_{\mathcal{X}_1\mathcal{Y}} = v^1} r^2_w = k_2 \sum_{w : w_{\mathcal{X}_1\mathcal{Y}} = v^1} r_w,\ \forall v^1 \in \mathcal{O}_G(\mathcal{X}_1 \times \mathcal{Y}) \\
& \sum_{u : u_{\mathcal{X}_2} = v^2_{\mathcal{X}_2}} p_u = \frac{|v^2_{\mathcal{X}_2}|}{|v^2|} k_1 \sum_{w : w_{\mathcal{X}_2\mathcal{Y}} = v^2} r^2_w,\ \forall v^2 \in \mathcal{O}_G(\mathcal{X}_2 \times \mathcal{Y}) \\
& \sum_{u : u_{\mathcal{X}_1} = v^1_{\mathcal{X}_1}} p_u = \frac{|v^1_{\mathcal{X}_1}|}{|v^1|} k_2 \sum_{w : w_{\mathcal{X}_1\mathcal{Y}} = v^1} r^1_w,\ \forall v^1 \in \mathcal{O}_G(\mathcal{X}_1 \times \mathcal{Y}) \\
& 0 \leq r_w \leq r^1_w, r^2_w \leq \frac{|w|}{|w_{\mathcal{X}_1\mathcal{X}_2}|} p_{w_{\mathcal{X}_1\mathcal{X}_2}},\ \forall w \in \mathcal{O}_G(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) \\
& \frac{|w|}{|w_{\mathcal{X}_1\mathcal{X}_2}|} p_{w_{\mathcal{X}_1\mathcal{X}_2}} - r^1_w - r^2_w + r_w \geq 0,\ \forall w \in \mathcal{O}_G(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y})\ .
\end{aligned} \tag{4.31}$$

Corollary 4.11. *For a channel $W : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}$, $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n},k_1,k_2)$ is the solution of a linear program of size bounded by $O\left(n^{|\mathcal{X}_1|\cdot|\mathcal{X}_2|\cdot|\mathcal{Y}|-1}\right)$, thus it can be computed in polynomial time in n.*

Proof. We use the linear program obtained in Theorem 4.10 with $G := S_n$ acting on $W^{\otimes n}$ as described before. The number of variables and constraints is linear in the number of orbits of the action of S^n on the different sets $\mathcal{A} \in \mathcal{Z}$, where here $\mathcal{Z} = \{\mathcal{X}_1^n, \mathcal{X}_2^n, \mathcal{Y}^n, \mathcal{X}_1^n \times \mathcal{Y}^n, \mathcal{X}_2^n \times \mathcal{Y}^n, \mathcal{X}_1^n \times \mathcal{X}_2^n, \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}^n\}$. For example, for $\mathcal{A} \in \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}^n$, we have that:

$$|\mathcal{O}_{S_n}(\mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}^n)| = \binom{n + |\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}| - 1}{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}| - 1} \leq (n + |\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}| - 1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|-1}\ .$$

So the number of variables and constraints is $O(n^{|\mathcal{X}_1|\cdot|\mathcal{X}_2|\cdot|\mathcal{Y}|-1})$. Note also that all the numbers occurring this linear program are integers or fractions of integers, with those integers ranging in $[(|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|)^n]$, thus of size $O(n\log(|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|))$. So the size of this linear program is bounded by $O(n^{|\mathcal{X}_1|\cdot|\mathcal{X}_2|\cdot|\mathcal{Y}|-1})$, and thus $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n},k_1,k_2)$ can be computed in polynomial time in n; see for instance Section 7.1 of [GM07]. □

In order to prove Theorem 4.10, we will need several lemmas. For all of them, $\mathcal{A}$ and $\mathcal{B}$ will denote finite sets on which a group G is acting, and x^G will denote the orbit of x under G:

Lemma 4.12. *Let $\tau \in \mathcal{O}_G(\mathcal{A} \times \mathcal{B})$, and call $\nu := \tau_{\mathcal{A}}$ and $\mu := \tau_{\mathcal{B}}$. For $x \in \nu$, let us call $B^x_\tau := \{y : (x,y) \in \tau\}$. Then, $|B^x_\tau| = |B^{x'}_\tau| =: c^\nu_\tau$ for any $x, x' \in \nu$, and furthermore, we* *have that* $c_\tau^\nu = \frac{|\tau|}{|\nu|}$*. Symmetrically, the same occurs for* $A_\tau^y := \{x : (x,y) \in \tau\}$ *with* $y \in \mu$*, where one gets that* $|A_\tau^y| = |A_\tau^{y'}| =: c_\tau^\mu = \frac{|\tau|}{|\mu|}$ *for* $y, y' \in \mu$*.*

Proof. Let $x, x' \in \nu$. Thus there exists $g \in G$ such that $x' = g \cdot x$. Let:

$$\begin{array}{cccc} f & : & B_\tau^x & \rightarrow & B_\tau^{x'} \\ & & y & \mapsto & g \cdot y \;. \end{array}$$

First, f is well defined. Indeed, if $y \in B_\tau^x = \{y : (x,y) \in \tau\}$, then $g \cdot y \in \{y : (g \cdot x, y) \in \tau\} = B_\tau^{x'}$, since $\tau \in \mathcal{O}_G(\mathcal{A} \times \mathcal{B})$. Let us show that f is injective. If $g \cdot y = g \cdot y'$, then $g^{-1} \cdot (g \cdot y) = (g^{-1}g) \cdot y = y$, $g^{-1} \cdot (g \cdot y') = y'$, so $y = y'$. Thus we get that $|B_\tau^x| \leq |B_\tau^{x'}|$. By a symmetric argument with x' replacing x and g^{-1} replacing g, we get that $|B_\tau^{x'}| \leq |B_\tau^x|$, and so $|B_\tau^x| = |B_\tau^{x'}| =: c_\tau^\nu$.

Furthermore, $\{B_\tau^x\}_{x \in \nu}$ is a partition of τ, so $\sum_{x \in \nu} |B_\tau^x| = |\nu| c_\tau^\nu = |\tau|$, and thus $c_\tau^\nu = \frac{|\tau|}{|\nu|}$. □

Lemma 4.13. *For any* $(x,y) \in \mathcal{A} \times \mathcal{B}$ *and* $v_{(x,y)^G}$ *variable indexed by orbits of* $\mathcal{A} \times \mathcal{B}$*, let us define the variable* $v_{x,y} := \frac{v_{(x,y)^G}}{|(x,y)^G|}$*. We have:*

$$\sum_{x \in \mathcal{A}} v_{x,y} = \frac{1}{|y^G|} \sum_{\tau \in \mathcal{O}_G(\mathcal{A} \times \mathcal{B}) : \tau_\mathcal{B} = y^G} v_\tau, \forall y \in \mathcal{B} \;.$$

Proof.

$$\begin{aligned} \sum_{x \in \mathcal{A}} v_{x,y} &= \sum_{\tau \in \mathcal{O}_G(\mathcal{A} \times \mathcal{B}) : \tau_\mathcal{B} = y^G} \sum_{x \in \mathcal{A} : (x,y) \in \tau} v_{x,y} \\ &= \sum_{\tau \in \mathcal{O}_G(\mathcal{A} \times \mathcal{B}) : \tau_\mathcal{B} = y^G} \sum_{x \in \mathcal{A} : (x,y) \in \tau} \frac{v_\tau}{|\tau|} \quad \text{since } (x,y)^G = \tau \\ &= \sum_{\tau \in \mathcal{O}_G(\mathcal{A} \times \mathcal{B}) : \tau_\mathcal{B} = y^G} c_\tau^{y^G} \frac{v_\tau}{|\tau|} \quad \text{by Lemma 4.12, since } y \in \tau_\mathcal{B} \\ &= \sum_{\tau \in \mathcal{O}_G(\mathcal{A} \times \mathcal{B}) : \tau_\mathcal{B} = y^G} \frac{|\tau|}{|y^G|} \frac{v_\tau}{|\tau|} = \frac{1}{|y^G|} \sum_{\tau \in \mathcal{O}_G(\mathcal{A} \times \mathcal{B}) : \tau_\mathcal{B} = y^G} v_\tau \;. \end{aligned} \tag{4.32}$$

□

Lemma 4.14. *For any* $\tau \in \mathcal{O}_G(\mathcal{A} \times \mathcal{B})$*,* $\mu \in \mathcal{O}_G(\mathcal{B})$ *and* $v_{x,y}$ *variable indexed by elements of* $\mathcal{A} \times \mathcal{B}$*, let us define* $v_\tau := \sum_{(x,y) \in \tau} v_{x,y}$*. We have:*

$$\sum_{\tau \in \mathcal{O}_G(\mathcal{A} \times \mathcal{B}) : \tau_\mathcal{B} = \mu} v_\tau = \sum_{y \in \mu} \sum_{x \in \mathcal{A}} v_{x,y} \;.$$

Proof.

$$\sum_{\tau \in \mathcal{O}_G(\mathcal{A} \times \mathcal{B}) : \tau_\mathcal{B} = \mu} v_\tau = \sum_{\tau \in \mathcal{O}_G(\mathcal{A} \times \mathcal{B}) : \tau_\mathcal{B} = \mu} \sum_{(x,y) \in \tau} v_{x,y} = \sum_{y \in \mu} \sum_{x \in \mathcal{A}} v_{x,y} \;.$$

□

Proof of Theorem 4.10. Let $r_{x_1,x_2,y}, r^1_{x_1,x_2,y}, r^2_{x_1,x_2,y}, p_{x_1,x_2}$ a feasible solution of the program defined in Proposition 4.4, and call $S := \frac{1}{k_1k_2}\sum_{x_1,x_2,y} W(y|x_1x_2)r_{x_1,x_2,y}$ its value. Define:

$$\begin{aligned} r_w &:= \sum_{(x_1,x_2,y)\in w} r_{x_1,x_2,y} \ , \quad r^1_w := \sum_{(x_1,x_2,y)\in w} r^1_{x_1,x_2,y} \ , \\ r^2_w &:= \sum_{(x_1,x_2,y)\in w} r^2_{x_1,x_2,y} \ , \qquad p_u := \sum_{(x_1,x_2)\in u} p_{x_1,x_2} \ . \end{aligned} \tag{4.33}$$

Let us show that r_w, r^1_w, r^2_w, p_u is a feasible solution of the program defined in Theorem 4.10, and that its value $S^* := \frac{1}{k_1k_2}\sum_w W(w)r_w = S$.

First, we have $S^* = S$. Indeed:

$$\begin{aligned} S^* &= \frac{1}{k_1k_2}\sum_w W(w)r_w = \frac{1}{k_1k_2}\sum_w W(w)\sum_{(x_1,x_2,y)\in w} r_{x_1,x_2,y} \\ &= \frac{1}{k_1k_2}\sum_w \sum_{(x_1,x_2,y)\in w} W(y|x_1x_2)r_{x_1,x_2,y} \quad \text{since } W(w) = W(y|x_1x_2) \text{ for all } (x_1,x_2,y)\in w \\ &= \frac{1}{k_1k_2}\sum_{x_1,x_2,y} W(y|x_1x_2)r_{x_1,x_2,y} = S \ . \end{aligned} \tag{4.34}$$

Then, all the constraints are satisfied. Indeed, thanks to Lemma 4.14, we have for the first constraint:

$$\sum_{w:w_{\mathcal{Y}}=v} r_w = \sum_{y\in v}\sum_{x_1,x_2} r_{x_1,x_2,y} = \sum_{y\in v} 1 = |v| \ .$$

For the second constraint (and symmetrically for the third constraint), we have:

$$\sum_{w:w_{\mathcal{X}_2\mathcal{Y}}=v^2} r^1_w = \sum_{(x_2,y)\in v^2}\sum_{x_1} r^1_{x_1,x_2,y} = \sum_{(x_2,y)\in v^2} k_1\sum_{x_1} r_{x_1,x_2,y} = k_1 \sum_{w:w_{\mathcal{X}_2\mathcal{Y}}=v^2} r_w \ .$$

For the fourth (and symmetrically for the fifth), we have:

$$\begin{aligned} \sum_{w:w_{\mathcal{X}_2\mathcal{Y}}=v^2} r^2_w &= \sum_{(x_2,y)\in v^2}\sum_{x_1} r^2_{x_1,x_2,y} = \sum_{(x_2,y)\in v^2}\frac{1}{k_1}\sum_{x_1} p_{x_1,x_2} = \frac{1}{k_1}\sum_{x_2\in v^2_{\mathcal{X}_2}}\sum_{y:(x_2,y)\in v^2}\sum_{x_1} p_{x_1,x_2} \\ &= \frac{1}{k_1}\sum_{x_2\in v^2_{\mathcal{X}_2}}\frac{|v^2|}{|v^2_{\mathcal{X}_2}|}\sum_{x_1} p_{x_1,x_2} \quad \text{thanks to Lemma 4.12} \\ &= \frac{1}{k_1}\frac{|v^2|}{|v^2_{\mathcal{X}_2}|}\sum_{u:u_{\mathcal{X}_2}=v^2_{\mathcal{X}_2}} p_u \ . \end{aligned} \tag{4.35}$$

Finally for the last constraints, we only need to compute:

$$\sum_{(x_1,x_2,y)\in w} p_{x_1,x_2} = \sum_{(x_1,x_2)\in w_{\mathcal{X}_1\mathcal{X}_2}}\sum_{y:(x_1,x_2,y)\in w} p_{x_1,x_2} = \sum_{(x_1,x_2)\in w_{\mathcal{X}_1\mathcal{X}_2}}\frac{|w|}{|w_{\mathcal{X}_1\mathcal{X}_2}|}p_{x_1,x_2} = \frac{|w|}{|w_{\mathcal{X}_1\mathcal{X}_2}|}p_{w_{\mathcal{X}_1\mathcal{X}_2}} \ ,$$

which implies that the linear inequalities on $p_{x_1,x_2}, r_{x_1,x_2,y}, r^1_{x_1,x_2,y}, r^2_{x_1,x_2,y}$ get transposed respectively to the values $\frac{|w|}{|w_{\mathcal{X}_1\mathcal{X}_2}|}p_{w_{\mathcal{X}_1\mathcal{X}_2}}, r_w, r^1_w, r^2_w$. Indeed, for instance, one has for any x_1, x_2, y that $p_{x_1,x_2} - r^1_{x_1,x_2,y} - r^2_{x_1,x_2,y} + r_{x_1,x_2,y} \geq 0$. Thus for some orbit w:

$$\sum_{(x_1,x_2,y)\in w} \left(p_{x_1,x_2} - r^1_{x_1,x_2,y} - r^2_{x_1,x_2,y} + r_{x_1,x_2,y}\right) \geq 0 \ ,$$

and then $\frac{|w|}{|w_{\mathcal{X}_1\mathcal{X}_2}|}p_{w_{\mathcal{X}_1\mathcal{X}_2}} - r^1_w - r^2_w + r_w \geq 0$, which was what we wanted to show.

Now let us consider a feasible solution r_w, r^1_w, r^2_w, p_u of the program defined in Theorem 4.10, with a value $S^* := \frac{1}{k_1k_2}\sum_w W(w)r_w$. Define:

$$\begin{aligned} r_{x_1,x_2,y} &:= \frac{r_{(x_1,x_2,y)^G}}{|(x_1,x_2,y)^G|} \ , \quad r^1_{x_1,x_2,y} := \frac{r^1_{(x_1,x_2,y)^G}}{|(x_1,x_2,y)^G|} \ , \\ r^2_{x_1,x_2,y} &:= \frac{r^2_{(x_1,x_2,y)^G}}{|(x_1,x_2,y)^G|} \ , \qquad p_{x_1,x_2} := \frac{p_{(x_1,x_2)^G}}{|(x_1,x_2)^G|} \ . \end{aligned} \tag{4.36}$$

Let us show that $r_{x_1,x_2,y}, r^1_{x_1,x_2,y}, r^2_{x_1,x_2,y}, p_{x_1,x_2}$ is a feasible solution of the program defined in Proposition 4.4, and that its value $S := \frac{1}{k_1k_2}\sum_{x_1,x_2,y} W(y|x_1x_2)r_{x_1,x_2,y} = S^*$.

First we have $S = S^*$. Indeed:

$$\begin{aligned} S &= \frac{1}{k_1k_2}\sum_{x_1,x_2,y} W(y|x_1x_2)r_{x_1,x_2,y} = \frac{1}{k_1k_2}\sum_{x_1,x_2,y} W(y|x_1x_2)\frac{r_{(x_1,x_2,y)^G}}{|r_{(x_1,x_2,y)^G}|} \\ &= \frac{1}{k_1k_2}\sum_{w}\sum_{(x_1,x_2,y)\in w} W(y|x_1x_2)\frac{r_w}{|w|} = \frac{1}{k_1k_2}\sum_{w}\sum_{(x_1,x_2,y)\in w} W(w)\frac{r_w}{|w|} \\ &= \frac{1}{k_1k_2}\sum_{w}|w|W(w)\frac{r_w}{|w|} = \frac{1}{k_1k_2}\sum_{w}W(w)r_w = S^* \ . \end{aligned} \tag{4.37}$$

Then, all the constraints are satisfied. Indeed, thanks to Lemma 4.13, we have for the first constraint:

$$\sum_{x_1,x_2} r_{x_1,x_2,y} = \frac{1}{|y^G|}\sum_{w:w_{\mathcal{Y}}=y^G} r_w = \frac{|y^G|}{|y^G|} = 1 \ .$$

For the second constraint (and symmetrically for the third constraint), we have:

$$\sum_{x_1} r^1_{x_1,x_2,y} = \frac{1}{|(x_2,y)^G|}\sum_{w:w_{\mathcal{X}_2\mathcal{Y}}=(x_2,y)^G} r^1_w = \frac{k_1}{|(x_2,y)^G|}\sum_{w:w_{\mathcal{X}_2\mathcal{Y}}=(x_2,y)^G} r_w = k_1\sum_{x_1} r_{x_1,x_2,y} \ .$$

For the fourth (and symmetrically for the fifth), we have:

$$\begin{aligned}\sum_{x_1} r^2_{x_1,x_2,y} &= \frac{1}{|(x_2,y)^G|}\sum_{w:w_{\mathcal{X}_2\mathcal{Y}}=(x_2,y)^G} r^2_w = \frac{1}{|(x_2,y)^G|}\frac{1}{k_1}\frac{|(x_2,y)^G|}{|(x_2,y)^G_{\mathcal{X}_2}|}\sum_{u:u_{\mathcal{X}_2}=(x_2,y)^G_{\mathcal{X}_2}} p_u\\ &= \frac{1}{k_1}\frac{1}{|(x_2,y)^G_{\mathcal{X}_2}|}\sum_{u:u_{\mathcal{X}_2}=(x_2,y)^G_{\mathcal{X}_2}} p_u = \frac{1}{k_1}\frac{1}{|x_2^G|}\sum_{u:u_{\mathcal{X}_2}=x_2^G} p_u \text{ since } (x_2,y)^G_{\mathcal{X}_2}=x_2^G\\ &= \frac{1}{k_1}\sum_{x_1} p_{x_1,x_2}\ .\end{aligned} \tag{4.38}$$

Finally, to conclude with the last constraints, one has only to see that for any x_1, x_2, y:

$$\frac{|(x_1,x_2,y)^G|}{|(x_1,x_2,y)^G_{\mathcal{X}_1\mathcal{X}_2}|}p_{(x_1,x_2,y)^G_{\mathcal{X}_1\mathcal{X}_2}} = \frac{|(x_1,x_2,y)^G|}{|(x_1,x_2)^G|}p_{(x_1,x_2)^G} = |(x_1,x_2,y)^G|p_{x_1,x_2}\ ,$$

which implies that the linear inequalities on $\frac{|w|}{|w_{\mathcal{X}_1\mathcal{X}_2}|}p_{w_{\mathcal{X}_1\mathcal{X}_2}}, r_w, r^1_w, r^2_w$ get transposed respectively to the values $p_{x_1,x_2}, r_{x_1,x_2,y}, r^1_{x_1,x_2,y}, r^2_{x_1,x_2,y}$. Indeed, for instance, one has for any w that $\frac{|w|}{|w_{\mathcal{X}_1\mathcal{X}_2}|}p_{w_{\mathcal{X}_1\mathcal{X}_2}} - r^1_w - r^2_w + r_w \geq 0$. But for any $(x_1,x_2,y)\in w$, one has that $r_{x_1,x_2,y} = \frac{r_w}{|w|}, r^1_{x_1,x_2,y} = \frac{r^1_w}{|w|}, r^2_{x_1,x_2,y} = \frac{r^2_w}{|w|}$. Thanks to the previous inequality, we have that $p_{x_1,x_2} = \frac{p_{w_{\mathcal{X}_1\mathcal{X}_2}}}{|w_{\mathcal{X}_1\mathcal{X}_2}|}$, and thus:

$$p_{x_1,x_2} - r^1_{x_1,x_2,y} - r^2_{x_1,x_2,y} + r_{x_1,x_2,y} = \frac{p_{w_{\mathcal{X}_1\mathcal{X}_2}}}{|w_{\mathcal{X}_1\mathcal{X}_2}|} - \frac{r^1_w}{|w|} - \frac{r^2_w}{|w|} + \frac{r_w}{|w|} \geq 0\ ,$$

which was what we wanted to show. □

4.3 Non-Signaling Achievability Bounds

4.3.1 Zero-Error Non-Signaling Assisted Achievable Rate Pairs

We will now present a numerical method to find efficiently inner bounds on $\mathcal{C}_0^{\mathrm{NS}}(W)$. Thanks to Corollary 4.11, we know how to decide in polynomial time in n, k_1, k_2 whether $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, k_1, k_2) = 1$. However, by Proposition 4.9, if $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, k_1, k_2) = 1$, then we have that $\left(\frac{k_1}{n}, \frac{k_2}{n}\right) \in \mathcal{C}_0^{\mathrm{NS}}(W)$, which describes a way of computing achievable points for that capacity region. More precisely, this leads to the following result:

Proposition 4.15 (Inner Bounds on $\mathcal{C}_0^{\mathrm{NS}}(W)$)**.** *Let us define the zero-error non-signaling assisted n-shots capacity region $\mathcal{C}_{0,\leq n}^{\mathrm{NS}}(W)$ in the following way:*

$$\mathcal{C}_{0,\leq n}^{\mathrm{NS}}(W) := \left\{\left(\frac{k_1}{n}, \frac{k_2}{n}\right) : \mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, k_1, k_2) = 1\right\}\ .$$

Then, we have that $\forall n \in \mathbb{N}, \mathcal{C}_{0,\leq n}^{\mathrm{NS}}(W) \subseteq \mathcal{C}_0^{\mathrm{NS}}(W)$, and that one can decide in polynomial time in n, k_1, k_2 if $\left(\frac{k_1}{n}, \frac{k_2}{n}\right) \in \mathcal{C}_{0,\leq n}^{\mathrm{NS}}(W)$.

This implies that we can find efficiently achievable rate pairs for MACs.

Application to the binary adder channel The binary adder channel W_{BAC} is the following MAC:

$$\forall x_1, x_2 \in \{0,1\}, \forall y \in \{0,1,2\}, W_{\mathrm{BAC}}(y|x_1x_2) := \delta_{y,x_1+x_2} \,.$$

Its classical capacity region $\mathcal{C}(W_{\mathrm{BAC}})$ is well known and consists of all (R_1, R_2) such that $R_1 \leq 1, R_2 \leq 1, R_1 + R_2 \leq \frac{3}{2}$, as a consequence of Theorem 4.1. Its zero-error classical capacity $\mathcal{C}_0(W_{\mathrm{BAC}})$ is not yet characterized. A lot of work has been done in finding outer and inner bounds on this region [Lin69, vT78, KL78, Jr.78, KLWY83, vdBvT85, BB98, UL98, AB99, MÖ05, OS15]. To date, the best lower bound on the sum-rate capacity is $\log_2(240/6) \simeq 1.3178$ [MÖ05].

Thanks to Proposition 4.15, we were able to compute the regions $\mathcal{C}^{\mathrm{NS}}_{0,\leq n}(W)$ for n going up to 7, which led to Figure 4.4. The code can be found on GitHub. It uses Mosek linear programming solver [AA00].

Note that the linear program from Theorem 4.10 has still a large number of variables and constraints although polynomial in n. Specifically, for $n = 2$, it has 244 variables and 480 constraints; for $n = 3$, it has 1112 variables and 2054 constraints; for $n = 7$, it has 95592 variables and 162324 constraints; finally, for $n = 8$, it has 226911 variables and 383103 constraints.

The first noticeable result coming from these curves is that the zero-error non-signaling assisted sum-rate capacity beats with only 7 copies the classical sum-rate capacity of $\frac{3}{2}$, even without a zero-error constraint, with a value of $\frac{2\log_2(42)}{7} \simeq 1.5406$, coming from the fact that $\mathrm{S}^{\mathrm{NS}}(W_{\mathrm{BAC}}^{\otimes 7}, 42, 42) = 1$ and Proposition 4.9. This implies that $\mathcal{C}_0^{\mathrm{NS}}(W_{\mathrm{BAC}})$ has larger sum-rate pairs than $\mathcal{C}(W_{\mathrm{BAC}})$, and that $\mathcal{C}^{\mathrm{NS}}(W_{\mathrm{BAC}})$ is strictly larger than $\mathcal{C}(W_{\mathrm{BAC}})$. This sum-rate can even be increased up to $\frac{\log_2(72)}{4} \simeq 1.5425$, since we have computed $\mathrm{S}^{\mathrm{NS}}(W_{\mathrm{BAC}}^{\otimes 8}, 72, 72) = 1$, which is the largest number of copies we have been able to manage with our efficient version of the linear program from Theorem 4.10. This should be compared with the upper bound on the non-signaling assisted sum-rate capacity coming from Proposition 4.23, which is $\log_2(3) \simeq 1.5850$ for $R_1 = R_2$.

Another surprising property is the speed at which one obtains efficient zero-error non-signaling assisted codes compared to classical zero-error codes. Indeed, with only three copies of the binary adder channel, one gets that $\mathrm{S}^{\mathrm{NS}}(W_{\mathrm{BAC}}^{\otimes 3}, 4, 5) = 1$, which corresponds to a sum-rate of $\frac{2+\log_2(5)}{3} \simeq 1.4406$, which already largely beats the best known zero-error achieved sum-rate of $\log_2(240/6) \simeq 1.3178$ [MÖ05]. These results are summarized in the following theorem:

Theorem 4.16. *We have that* $\left(\frac{\log_2(72)}{8}, \frac{\log_2(72)}{8}\right) \in \mathcal{C}_0^{\mathrm{NS}}(W_{\mathrm{BAC}})$ *but* $\left(\frac{\log_2(72)}{8}, \frac{\log_2(72)}{8}\right) \notin \mathcal{C}(W_{\mathrm{BAC}})$*, and as a consequence, we have that* $\mathcal{C}(W_{\mathrm{BAC}}) \subsetneq \mathcal{C}^{\mathrm{NS}}(W_{\mathrm{BAC}})$.

Proof. Since $2^{8\frac{\log_2(72)}{8}} = 72$ and numerically $\mathrm{S}^{\mathrm{NS}}(W_{\mathrm{BAC}}^{\otimes 8}, 72, 72) = 1$ thanks to Corollary 4.11, we get that $\left(\frac{\log_2(72)}{8}, \frac{\log_2(72)}{8}\right) \in \mathcal{C}_0^{\mathrm{NS}}(W_{\mathrm{BAC}})$ by Proposition 4.9. However, $\frac{\log_2(72)}{8} + \frac{\log_2(72)}{8} > \frac{3}{2}$ so $\left(\frac{\log_2(72)}{8}, \frac{\log_2(72)}{8}\right) \notin \mathcal{C}(W_{\mathrm{BAC}})$ by Theorem 4.1 applied to W_{BAC}. Since $\mathcal{C}(W_{\mathrm{BAC}}) \subseteq \mathcal{C}^{\mathrm{NS}}(W_{\mathrm{BAC}})$ and $\mathcal{C}_0^{\mathrm{NS}}(W_{\mathrm{BAC}}) \subseteq \mathcal{C}^{\mathrm{NS}}(W_{\mathrm{BAC}})$, we thus get that $\mathcal{C}(W_{\mathrm{BAC}}) \subsetneq \mathcal{C}^{\mathrm{NS}}(W_{\mathrm{BAC}})$. □

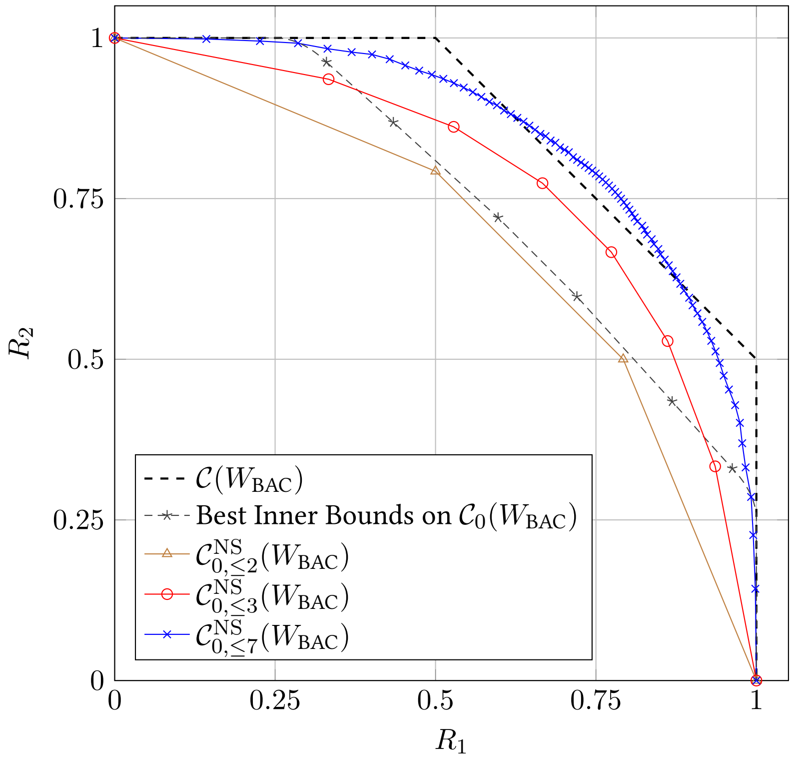


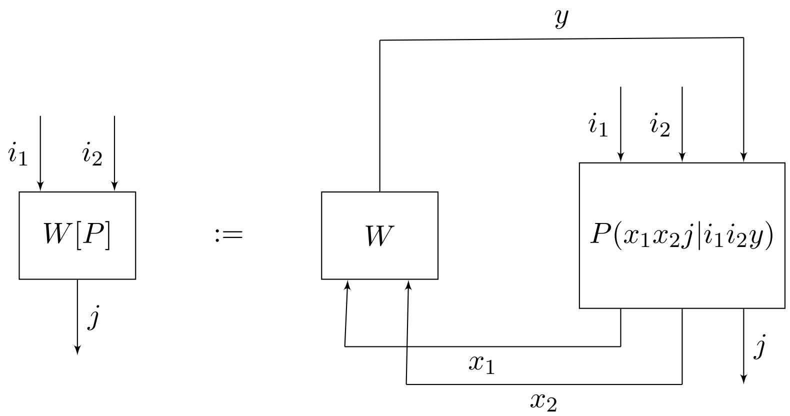
Figure 4.4 – Capacity regions of the binary adder channel W_{BAC}. The black dashed curve depicts the classical capacity region $\mathcal{C}(W_{\mathrm{BAC}})$, whereas the grey dashed curve shows the best known inner bound border on the zero-error classical capacity region $\mathcal{C}_0(W_{\mathrm{BAC}})$, made from results by [MÖ05, vdBvT85, KLWY83]; see [MÖ05] for a description of this border. On the other hand, the continuous curves depict the best zero-error non-signaling assisted achievable rate pairs for respectively 2, 3 and 7 copies of the binary adder channel.

4.3.2 Non-Signaling Assisted Achievable Rate Pairs with Non-Zero Error

We have analyzed the non-signaling assisted capacity region through zero-error strategies and applied it to the BAC. However, if some noise is added to that channel, its zero-error non-signaling assisted capacity region becomes trivial (see Proposition 4.19). Thus, the previous method fails to find significant inner bounds on the non-signaling assisted capacity region of noisy MACs.

In this section, we use concatenated codes to obtain achievable rate pairs, and apply it to a noisy version of the BAC:

Definition 4.10 (Concatenated Codes)**.** Given a MAC W and a non-signaling assisted code P, define $W[P] : [k_1]\times[k_2] \to [\ell]$ with $W[P](j|i_1i_2) := \sum_{x_1,x_2,y} W(y|x_1x_2)P(x_1x_2j|i_1i_2y)$:



Note that $W[P]$ is a MAC since $W[P](j|i_1i_2) \geq 0$ and:

$$\begin{aligned}\sum_j W[P](j|i_1i_2) &= \sum_{x_1,x_2,y} W(y|x_1x_2) \sum_j P(x_1x_2j|i_1i_2y) \\ &= \sum_{x_1,x_2} \left(\sum_y W(y|x_1x_2) \right) P(x_1x_2|i_1i_2) \text{ since } P \text{ is non-signaling} \\ &= \sum_{x_1,x_2} P(x_1x_2|i_1i_2) = 1\ .\end{aligned} \tag{4.39}$$

The following proposition states that combining a classical code to a non-signaling strategy leads to inner bounds on the non-signaling assisted capacity region of a MAC:

Proposition 4.17. *If P is a non-signaling assisted code for the MAC W, we have that $\mathcal{C}(W[P]) \subseteq \mathcal{C}^{\mathrm{NS}}(W)$.*

Proof. Let $(R_1, R_2) \in \mathcal{C}(W[P])$. Then, by definition, we have that:

$$\lim_{n \to +\infty} \mathrm{S}(W[P]^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil) = 1\ .$$

Let us fix $\varepsilon > 0$. For a large enough N, we have $\mathrm{S}(W[P]^{\otimes N}, \lceil 2^{R_1 N} \rceil, \lceil 2^{R_2 N} \rceil) \geq 1 - \varepsilon$. Let us call $\ell_1 := \lceil 2^{R_1 N} \rceil$ and $\ell_2 := \lceil 2^{R_2 N} \rceil$. Thus, there exists encoders $e_1 : [\ell_1] \to [k_1], e_2 : [\ell_2] \to [k_2]$ and a decoder $d : [\ell] \to [\ell_1] \times [\ell_2]$ such that:

$$\frac{1}{\ell_1 \ell_2} \sum_{i_1,i_2,j} W[P](j|i_1i_2) \sum_{a_1 \in [\ell_1], a_2 \in [\ell_2]} e_1(i_1|a_1) e_2(i_2|a_2) d(a_1a_2|j) \geq 1 - \varepsilon\ .$$

In particular, we have:

$$\frac{1}{\ell_1 \ell_2} \sum_{x_1,x_2,y} W(y|x_1x_2) \left[\sum_{i_1,i_2,j,a_1,a_2} P(x_1x_2j|i_1i_2y) e_1(i_1|a_1) e_2(i_2|a_2) d(a_1a_2|j) \right] \geq 1 - \varepsilon\ .$$

Let us define $\hat{P}(x_1x_2(b_1b_2)|a_1a_2y) := \sum_{i_1,i_2,j} P(x_1x_2j|i_1i_2y) e_1(i_1|a_1) e_2(i_2|a_2) d(b_1b_2|j)$. Then, one can easily check that $\hat{P}$ is non-signaling, and thus:

$$\mathrm{S}^{\mathrm{NS}}(W^{\otimes N}, \ell_1, \ell_2) \geq \frac{1}{\ell_1 \ell_2} \sum_{x_1, x_2, y} W(y|x_1 x_2) \sum_{a_1, a_2} \hat{P}(x_1 x_2 (a_1, a_2)|a_1 a_2 y) \geq 1 - \varepsilon \,.$$

This implies that $\lim_{n \to +\infty} \mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil) = 1$, i.e. $(R_1, R_2) \in \mathcal{C}^{\mathrm{NS}}(W)$. □

Thanks to Proposition 4.17, we have for any non-signaling assisted code P, $\mathcal{C}(W^{\otimes n}[P]) \subseteq \mathcal{C}^{\mathrm{NS}}(W^{\otimes n})$. But if $(R_1, R_2) \in \mathcal{C}^{\mathrm{NS}}(W^{\otimes n})$, we have that $(\frac{R_1}{n}, \frac{R_2}{n}) \in \mathcal{C}^{\mathrm{NS}}(W)$. Thus, applying Theorem 4.1 on $W^{\otimes n}[P]$ leads to inner bounds on $\mathcal{C}^{\mathrm{NS}}(W)$:

Proposition 4.18 (Inner Bounds on $\mathcal{C}^{\mathrm{NS}}(W)$)**.** *For any number of copies n, number of inputs $k_1 \in [|\mathcal{X}_1|^n]$ and $k_2 \in [|\mathcal{X}_2|^n]$, non-signaling assisted codes P on inputs in $[k_1], [k_2]$ for $W^{\otimes n}$, and distributions q_1, q_2 on $[k_1], [k_2]$, we have that the following (R_1, R_2) are in $\mathcal{C}^{\mathrm{NS}}(W)$:*

$$R_1 \leq \frac{I(I_1 : J|I_2)}{n} \,,\; R_2 \leq \frac{I(I_2 : J|I_1)}{n} \,,\; R_1 + R_2 \leq \frac{I((I_1, I_2) : J)}{n} \,,$$

for $(I_1, I_2) \in [k_1] \times [k_2]$ following the product law $q_1 \times q_2$, and $J \in [\ell]$ the outcome of $W^{\otimes n}[P]$ on inputs I_1, I_2. In particular, the corner points of this capacity region are given by:

$$\left(\frac{I(I_1 : J|I_2)}{n}, \frac{I(I_2 : J)}{n} \right) \text{ and } \left(\frac{I(I_1 : J)}{n}, \frac{I(I_2 : J|I_1)}{n} \right) .$$

Proof. The achievable region comes from the previous discussion. We just need to prove that the corner points are of the given form. If $R_1 = \frac{I(I_1:J|I_2)}{n}$, constraints on R_2 and $R_1 + R_2$ leads to a maximum $R_2 = \min\left(\frac{I(I_2:J|I_1)}{n}, \frac{I((I_1,I_2):J)}{n} - \frac{I(I_1:J|I_2)}{n} \right)$. However, $I((I_1, I_2) : J) - I(I_1 : J|I_2) = I(I_2 : J)$ by the chain rule. We only need to show that $I(I_2 : J) \leq I(I_2 : J|I_1)$ and the proof will be complete, since the other corner point is symmetric. We have:

$$I(I_2 : J) = H(I_2) - H(I_2|J) = H(I_2|I_1) - H(I_2|J) \leq H(I_2|I_1) - H(I_2|JI_1) = I(I_2 : J|I_1) \,,$$

the second equality coming from the fact that I_1 and I_2 are independent, and the inequality coming from the fact that $H(A|BC) \leq H(A|B)$ for any A, B, C. □

Application to the Noisy Binary Adder Channel We will now apply this strategy to a noisy version of the BAC. We will consider flip errors $\varepsilon_1, \varepsilon_2$ of inputs x_1, x_2 on W_{BAC}, which leads to the following definition of $W_{\mathrm{BAC},\varepsilon_1,\varepsilon_2}$:

$$\begin{aligned} \forall y, x_1, x_2, W_{\mathrm{BAC},\varepsilon_1,\varepsilon_2}(y|x_1 x_2) := {} & (1 - \varepsilon_1)(1 - \varepsilon_2) W_{\mathrm{BAC}}(y|x_1 x_2) \\ & + \varepsilon_1 (1 - \varepsilon_2) W_{\mathrm{BAC}}(y|\overline{x_1} x_2) \\ & + (1 - \varepsilon_1) \varepsilon_2 W_{\mathrm{BAC}}(y|x_1 \overline{x_2}) \\ & + \varepsilon_1 \varepsilon_2 W_{\mathrm{BAC}}(y|\overline{x_1}\,\overline{x_2}) \,. \end{aligned} \tag{4.40}$$

First, let us note that the zero-error non-signaling assisted capacity region of $W_{\mathrm{BAC},\varepsilon_1,\varepsilon_2}$ is trivial for $\varepsilon \in (0, 1)$:

Proposition 4.19. *If $\varepsilon_1, \varepsilon_2 \in (0, 1)$, then $\mathcal{C}_0^{\mathrm{NS}}(W_{BAC,\varepsilon_1,\varepsilon_2}) = \{(0, 0)\}$.*

Proof. If $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, k_1, k_2) = 1$, then $\forall y^n, x_1^n, x_2^n : W^{\otimes n}(y^n|x_1^n x_2^n) > 0 \implies r_{x_1^n,x_2^n,y^n} = p_{x_1^n,x_2^n}$. Indeed, we have for an optimal p, r that:

$$\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, k_1, k_2) = \frac{1}{k_1 k_2} \sum_{x_1^n,x_2^n,y^n} W^{\otimes n}(y^n|x_1^n x_2^n) r_{x_1^n,x_2^n,y^n} \leq \frac{1}{k_1 k_2} \sum_{x_1^n,x_2^n,y^n} W^{\otimes n}(y^n|x_1^n x_2^n) p_{x_1^n,x_2^n} = 1\,,$$

which implies the previous statement. But, for $W^{\otimes n}_{\mathrm{BAC},\varepsilon_1,\varepsilon_2}$, one can easily check that for all y^n, x_1^n, x_2^n, $W^{\otimes n}(y^n|x_1^n x_2^n) > 0$ since $\varepsilon_1, \varepsilon_2 \in (0,1)$. Indeed, you just have to flip the inputs to a valid preimage of the output. Thus if $\mathrm{S}^{NS}(W^{\otimes n}_{\mathrm{BAC},\varepsilon_1,\varepsilon_2}, k_1, k_2) = 1$, we have that $\forall y^n, x_1^n, x_2^n, r_{x_1^n,x_2^n,y^n} = p_{x_1^n,x_2^n}$. In particular, this implies that $\sum_{x_1^n,x_2^n} r_{x_1^n,x_2^n,y^n} = \sum_{x_1^n,x_2^n} p_{x_1^n,x_2^n}$, therefore $1 = k_1 k_2$, so $k_1 = 1$ and $k_2 = 1$. Thus $\mathrm{S}^{NS}(W^{\otimes n}, 2^{nR_1}, 2^{nR_2}) = 1$ implies that $(R_1, R_2) = (0,0)$. $\square$

We have then applied the numerical method described in Proposition 4.18 to $W_{\mathrm{BAC},\varepsilon_1,\varepsilon_2}$ for the symmetric case $\varepsilon_1 = \varepsilon_2 = \varepsilon := 10^{-3}$. Since it is hard to go through all non-signaling assisted codes P and product distributions q_1, q_2, we have applied the heuristic of using non-signaling assisted codes obtained while optimizing $\mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, k_1, k_2)$ in the symmetrized linear program. We have combined them with uniform q_1, q_2, as the form of those non-signaling assisted codes coming from our optimization program is symmetric. We have evaluated the achievable corner points for all $k_1, k_2 \leq 2^n$ for $n \leq 5$ copies which led to Figure 4.5: Compared to the noiseless binary adder channel, we can first notice that the classical capacity region is slightly smaller, with a classical sum-rate capacity of 1.478 at most. On the other hand, although the zero-error non-signaling assisted capacity of $W_{\mathrm{BAC},\varepsilon,\varepsilon}$ is completely trivial, we have with our concatenated codes strategy found significant rate pairs achievable with non-signaling assistance. In particular, we have reached a non-signaling assisted sum-rate capacity of 1.493 which beats the best classical sum-rate capacity. Thus, it shows that non-signaling assistance can improve the capacity of the noisy binary adder channel as well.

4.4 Relaxed Non-Signaling Assisted Capacity Region and Outer Bounds

A natural question that arises when studying the strength of non-signaling assistance is whether a result similar to Theorem 4.1 can be found to describe by a single-letter formula the non-signaling assisted capacity region of MACs. In particular, dropping the constraint that (X_1, X_2) is in product form in Theorem 4.1 seems to be a particularly good candidate to characterize the non-signaling assisted capacity region of MACs, as this looks quite similar to allowing correlations between parties.

We have not been able to show the equivalence between this region and the non-signaling assisted capacity region; however, it turns out to be equivalent to the capacity region defined by a slight relaxation of non-signaling assistance, which we call $\mathrm{S}^{\overline{\mathrm{NS}}}(W, k_1, k_2)$. In particular, this will give us the best known outer bound on the non-signaling capacity.

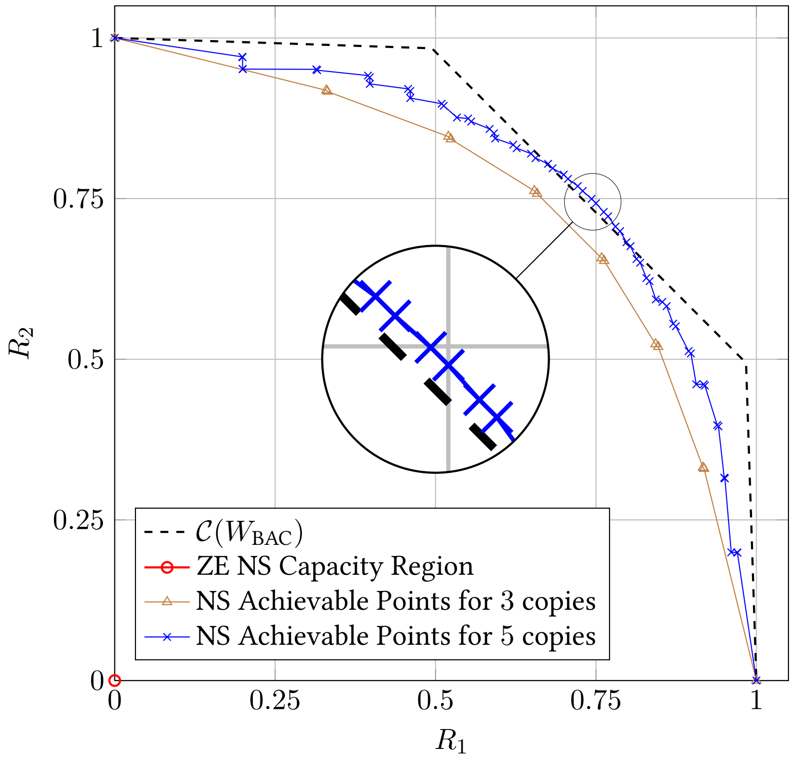


Figure 4.5 – Capacity regions of the noisy binary adder channel $W_{\mathrm{BAC},\varepsilon,\varepsilon}$ for $\varepsilon = 10^{-3}$. The black dashed curve depicts the classical capacity region $\mathcal{C}(W_{\mathrm{BAC},\varepsilon,\varepsilon})$ which was found numerically using Theorem 4.1. The red point depicts the zero-error non-signaling assisted capacity region (Proposition 4.19). The blue curve depicts achievable non-signaling assisted rates pairs obtained from $\mathcal{C}(W_{\mathrm{BAC},\varepsilon,\varepsilon}^{\otimes 5}[P])$ through the numerical method described in Proposition 4.18.

Definition 4.11.

$$\begin{aligned} \mathrm{S}^{\overline{\mathrm{NS}}}(W,k_1,k_2) := \quad \underset{r,p}{\text{maximize}} \quad & \frac{1}{k_1k_2}\sum_{x_1,x_2,y} W(y|x_1x_2)r_{x_1,x_2,y} \\ \text{subject to} \quad & \sum_{x_1,x_2} r_{x_1,x_2,y} \leq 1 \\ & \sum_{x_1,x_2} p_{x_1,x_2} = k_1k_2 \\ & \sum_{x_1} p_{x_1,x_2} \geq k_1 \sum_{x_1} r_{x_1,x_2,y} \\ & \sum_{x_2} p_{x_1,x_2} \geq k_2 \sum_{x_2} r_{x_1,x_2,y} \\ & 0 \leq r_{x_1,x_2,y} \leq p_{x_1,x_2} \end{aligned} \tag{4.41}$$

The following proposition shows that this is indeed a relaxation of the non-signaling constraint.

Proposition 4.20. $\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2) \leq \mathrm{S}^{\overline{\mathrm{NS}}}(W,k_1,k_2)$.

Proof. Let us take a solution $(p_{x_1,x_2}, r_{x_1,x_2,y}, r^1_{x_1,x_2,y}, r^2_{x_1,x_2,y})_{x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, y \in \mathcal{Y}}$ of the linear program computing $\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)$. Let us show that $(p_{x_1,x_2}, r_{x_1,x_2,y})_{x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, y \in \mathcal{Y}}$ is a solution of the linear program computing $\mathrm{S}^{\overline{\mathrm{NS}}}(W, k_1, k_2)$ with a same objective value, from which the proposition follows.

They have indeed the same value, since the definition which is the same for both programs depends only on $r_{x_1,x_2,y}$. Let us show that all constraints are satisfied for $(p_{x_1,x_2}, r_{x_1,x_2,y})_{x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, y \in \mathcal{Y}}$. We have $\sum_{x_1,x_2} r_{x_1,x_2,y} = 1 \leq 1$ so the first constraint is satisfied. We have then that:

$$\sum_{x_1,x_2} p_{x_1,x_2} = k_1 \sum_{x_1,x_2} r^2_{x_1,x_2,y} = k_1 k_2 \sum_{x_1,x_2} r_{x_1,x_2,y} = k_1 k_2 \ ,$$

so the second constraint is satisfied.

For the third constraint (and symmetrically the fourth constraint), we have:

$$\sum_{x_1} p_{x_1,x_2} = k_1 \sum_{x_1} r^2_{x_1,x_2,y} \geq k_1 \sum_{x_1} r_{x_1,x_2,y} \ .$$

Finally, we have directly $0 \leq r_{x_1,x_2,y} \leq p_{x_1,x_2}$, so the last constraint is satisfied. □

We can now introduce the relaxed non-signaling assisted capacity region $\mathcal{C}^{\overline{\mathrm{NS}}}(W)$:

Definition 4.12 ($\mathcal{C}^{\overline{\mathrm{NS}}}(W)$)**.** A rate pair (R_1, R_2) is achievable with relaxed non-signaling assistance if:

$$\lim_{n \to +\infty} \mathrm{S}^{\overline{\mathrm{NS}}}(W^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil) = 1 \ .$$

We define $\mathcal{C}^{\overline{\mathrm{NS}}}(W)$ as the closure of the convex hull of the set of all achievable rate pairs with relaxed non-signaling assistance.

Remark. One could show as in the non-relaxed case that $\mathcal{C}^{\overline{\mathrm{NS}}}(W)$ is convex without taking the convex hull in its definition.

A direct property that follows from this definition and Proposition 4.20 is the fact that the non-signaling assisted capacity region is included in the relaxed non-signaling assisted capacity region.

Corollary 4.21. $\mathcal{C}^{\mathrm{NS}}(W) \subseteq \mathcal{C}^{\overline{\mathrm{NS}}}(W)$.

We present now the main result of this section, the characterization of $\mathcal{C}^{\overline{\mathrm{NS}}}(W)$ by a single-letter formula.

Theorem 4.22 (Characterization of $\mathcal{C}^{\overline{\mathrm{NS}}}(W)$)**.** *$\mathcal{C}^{\overline{\mathrm{NS}}}(W)$ is the closure of the convex hull of all rate pairs (R_1, R_2) satisfying:*

$$R_1 < I(X_1 : Y | X_2) \ , \ R_2 < I(X_2 : Y | X_1) \ , \ R_1 + R_2 < I((X_1, X_2) : Y) \ ,$$

for (X_1, X_2) following some law $P_{X_1 X_2}$ on $\mathcal{X}_1 \times \mathcal{X}_2$, and $Y \in \mathcal{Y}$ the outcome of W on inputs X_1, X_2.

Remark. Note that the only difference with the classical capacity region of MACs in Theorem 4.1 is that the joint distribution of X_1 and X_2 does not have any product form constraints here.

The proof of Theorem 4.22 will be divided in Proposition 4.28 (outer bound part) and Proposition 4.32 (achievability part). But first, let us apply these results to the binary adder channel.

Application to the Binary Adder Channel Let us determine the relaxed non-signaling assisted capacity of the binary adder channel which will be plotted in Figure 4.6.

Proposition 4.23. *$\mathcal{C}^{\overline{\mathrm{NS}}}(W_{BAC})$ has the following description:*

$$\mathcal{C}^{\overline{\mathrm{NS}}}(W_{BAC}) = \bigcup_{q \in \left[\frac{1}{2}, \frac{2}{3}\right]} \left\{(R_1, R_2) : R_1 \leq h\left(q\right), R_2 \leq h\left(q\right), R_1 + R_2 \leq q + h\left(q\right)\right\} .$$

Remark. Note that for $q = \frac{1}{2}$, the bound becomes $R_1 \leq 1, R_2 \leq 1, R_1 + R_2 \leq \frac{3}{2}$ and when $q = \frac{2}{3}$ the bound becomes $R_1 \leq \log_2(3) - \frac{2}{3}, R_2 \leq \log_2(3) - \frac{2}{3}, R_1 + R_2 \leq \log_2(3)$.

Proof. We use the characterization of $\mathcal{C}^{\overline{\mathrm{NS}}}$ provided by Theorem 4.22.

Let us consider an arbitrary $P_{X_1X_2} = (p_{00}, p_{01}, p_{10}, p_{11})$. First, we have that $I((X_1, X_2) : Y) = H(Y) - H(Y|X_1X_2) = H(Y)$ since Y is a deterministic function of (X_1, X_2). Then, we have that $I(X_1 : Y|X_2) = H(Y|X_2) - H(Y|X_1X_2) = H(Y|X_2)$ for the same reason. Furthermore, given X_2, Y is a deterministic function of X_1, so we have $I(X_1 : Y|X_2) = H(Y|X_2) - H(Y|X_1X_2) = H(X_1|X_2)$. Symmetrically we have as well $I(X_2 : Y|X_1) = H(X_2|X_1)$. In all:

$$\mathcal{C}^{\overline{\mathrm{NS}}}(W_{\mathrm{BAC}}) = \bigcup_{P_{X_1X_2}} \left\{(R_1, R_2) : R_1 \leq H(X_1|X_2), R_2 \leq H(X_2|X_1), R_1 + R_2 \leq H(X_1 + X_2)\right\}$$

Let us call $B_1(P_{X_1X_2}) := H(X_1|X_2), B_2(P_{X_1X_2}) := H(X_2|X_1), B_{12}(P_{X_1X_2}) := H(X_1 + X_2)$ the three bounds. Let us call $P_{\overline{X}_1\overline{X}_2} = (p_{11}, p_{10}, p_{01}, p_{00})$. One can notice that:

$$\begin{aligned} B_1(P_{\overline{X}_1\overline{X}_2}) &= H(\overline{X}_1|\overline{X}_2) = H(1 - X_1|1 - X_2) = H(X_1|X_2) = B_1(P_{X_1X_2}) , \\ B_2(P_{\overline{X}_1\overline{X}_2}) &= H(1 - X_2|1 - X_1) = H(X_2|X_1) = B_2(P_{X_1X_2}) , \\ B_{12}(P_{\overline{X}_1\overline{X}_2}) &= H(\overline{X}_1 + \overline{X}_2) = H(1 - X_1 + 1 - X_2) \\ &= H(2 - (X_1 + X_2)) = H(X_1 + X_2) = B_{12}(P_{X_1X_2}) . \end{aligned} \tag{4.42}$$

Since $B_{12}(P_{X_1X_2}) = H(X_1 + X_2) = H(p_{00}, p_{11}, p_{01} + p_{10})$, it is concave in $P_{X_1X_2}$ as H is concave and $(p_{00}, p_{11}, p_{01} + p_{10})$ is linear in $P_{X_1X_2}$. Also, $B_1(P_{X_1X_2}) = H(X_1|X_2) = -D(P_{X_1X_2}||I \otimes P_{X_2})$ is concave in $P_{X_1X_2}$ as the divergence D is jointly convex and $I \otimes P_{X_2}$ is linear in $P_{X_1X_2}$. By symmetry, $B_2(P_{X_1X_2})$ is as well concave in $P_{X_1X_2}$. Let us consider any of those three bounds, which we call B. We have by concavity of B and the fact that $B(P_{X_1X_2}) = B(P_{\overline{X}_1\overline{X}_2})$:

$$B(P_{X_1X_2}) = \frac{B(P_{X_1X_2}) + B(P_{\overline{X}_1\overline{X}_2})}{2} \leq B\left(\frac{P_{X_1X_2} + P_{\hat{X}_1\hat{X}_2}}{2}\right) = B\left(\frac{q}{2}, \frac{1-q}{2}, \frac{1-q}{2}, \frac{q}{2}\right) ,$$

with $q = p_{00} + p_{11}$. This holds for the three bounds at the same time, so we can restrict ourselves to the distributions of the form $\left(\frac{q}{2}, \frac{1-q}{2}, \frac{1-q}{2}, \frac{q}{2}\right)$ for some $q \in [0,1]$, i.e., $P_{X_1X_2}(0,0) = P_{X_1X_2}(1,1) = \frac{q}{2}$ and $P_{X_1X_2}(0,1) = P_{X_1X_2}(1,0) = \frac{1-q}{2}$.
We have $P_Y(0) = P_Y(2) = \frac{q}{2}$ and $P_Y(1) = 1-q$, so:

$$\begin{aligned} B_{12}(P_{X_1X_2}) = H(Y) &= -q\log\left(\frac{q}{2}\right) - (1-q)\log(1-q) \\ &= -q\left(\log(q) - 1\right) - (1-q)\log(1-q) \\ &= q + h\left(q\right) \ . \end{aligned} \tag{4.43}$$

We have $P_{X_2}(0) = P_{X_1X_2}(0,0) + P_{X_1X_2}(1,0) = \frac{q}{2} + \frac{1-q}{2} = \frac{1}{2}$ so $P_{X_2}(1) = \frac{1}{2}$. Thus:

$$B_1(P_{X_1X_2}) = H(X_1|X_2) = \frac{1}{2}H(X_1|X_2 = 0) + \frac{1}{2}H(X_1|X_2 = 1) \ .$$

We have $P_{X_1|X_2=0}(0) = \frac{P_{X_1X_2}(0,0)}{P_{X_2}(0)} = q$ so $H(X_1|X_2 = 0) = h\left(q\right)$. On the other hand, we have $P_{X_1|X_2=1}(1) = \frac{P_{X_1X_2}(1,1)}{P_{X_2}(1)} = q$ so we get as well $H(Y|X_2 = 1) = h\left(q\right)$, and $B_1(P_{X_1X_2}) = H(X_1|X_2) = h\left(q\right)$. Symmetrically, we also get $B_2(P_{X_1X_2}) = h\left(q\right)$. Therefore, we get that $\mathcal{C}^{\overline{\mathrm{NS}}}(W_{\mathrm{BAC}})$ is the closure of the convex hull of:

$$\bigcup_{q\in[0,1]} \left\{(R_1,R_2) : R_1 < h\left(q\right), R_2 < h\left(q\right), R_1 + R_2 < q + h\left(q\right)\right\} \ .$$

However this set is already convex, so we have:

$$\mathcal{C}^{\overline{\mathrm{NS}}}(W_{\mathrm{BAC}}) = \bigcup_{q\in[0,1]} \left\{(R_1,R_2) : R_1 \leq h\left(q\right), R_2 \leq h\left(q\right), R_1 + R_2 \leq q + h\left(q\right)\right\} \ .$$

Finally, we can restrict ourselves to $q \in \left[\frac{1}{2}, \frac{2}{3}\right]$, since h is increasing from 0 to $\frac{1}{2}$ (thus $q \mapsto q + h\left(q\right)$ as well), and the fact that $q \mapsto q + h\left(q\right)$ achieves its maximum for $q = \frac{2}{3}$ with $\frac{2}{3} + h\left(\frac{2}{3}\right) = \log_2(3)$ and then decreases (whereas h is decreasing from $\frac{1}{2}$ to 1), which completes the proof. □

As before, one can also define a symmetrized version of the relaxed linear program computing the value $\mathrm{S}^{\overline{\mathrm{NS}}}(W^{\otimes n}, k_1, k_2)$ in polynomial time in n and compute the zero-error n-shots capacity region by looking at the rates where $\mathrm{S}^{\overline{\mathrm{NS}}}(W^{\otimes n}, k_1, k_2) = 1$. We have computed this up to 7 copies of the binary adder channel, which led to Figure 4.6:

The first noticeable result coming from these curves is that the values $\mathrm{S}^{\overline{\mathrm{NS}}}$ and S^{NS} differ. While the highest sum-rate of $\frac{2\log_2(42)}{7} \simeq 1.5406$ is achieved on 7 copies of the binary adder channel with zero-error and non-signaling assistance, coming from the fact that $\mathrm{S}^{\mathrm{NS}}(W_{\mathrm{BAC}}^{\otimes 7}, 42, 42) = 1$, we have that $\mathrm{S}^{\overline{\mathrm{NS}}}(W_{\mathrm{BAC}}^{\otimes 7}, 44, 44) = 1 > \mathrm{S}^{\mathrm{NS}}(W_{\mathrm{BAC}}^{\otimes 7}, 44, 44) \simeq 0.9581$ which implies that a sum-rate of $\frac{2\log_2(44)}{7} \simeq 1.5598$ is achieved on 7 copies of the binary adder channel with zero-error and relaxed non-signaling assistance. It also

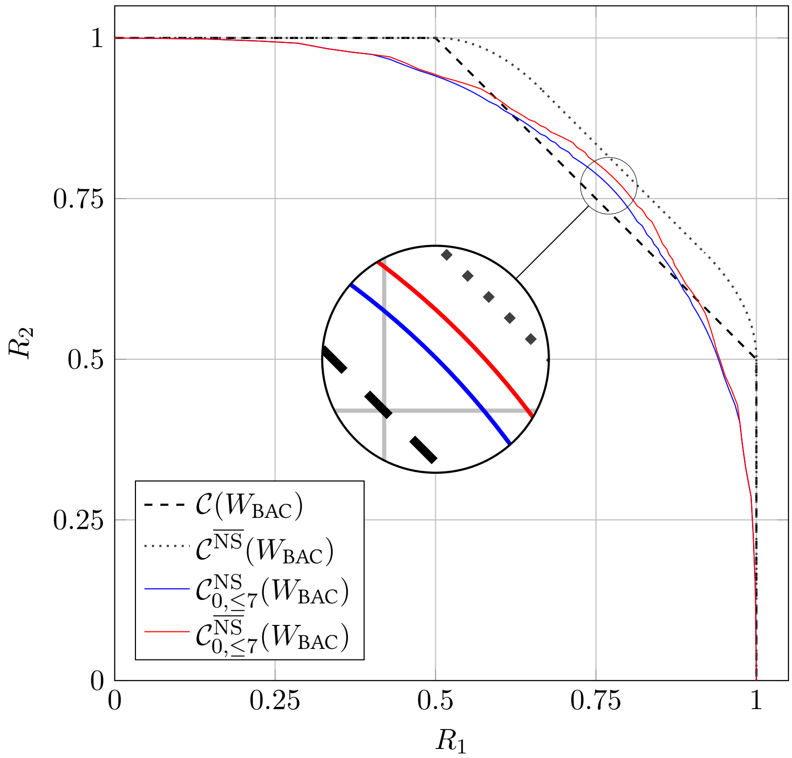


Figure 4.6 – Comparison of relaxed and regular non-signaling assisted capacity regions of the binary adder channel. The black dashed curve depicts the classical capacity region $\mathcal{C}(W_{\mathrm{BAC}})$, whereas the grey dotted curve depicts the relaxed non-signaling assisted capacity region $\mathcal{C}^{\overline{\mathrm{NS}}}(W_{\mathrm{BAC}})$ as described in Proposition 4.23. In particular, the curved corners are obtained by taking $R_1 = h(R_2)$ for $R_2 \in \left[\frac{1}{2}, \frac{2}{3}\right]$ and symmetrically by switching the roles played by R_1 and R_2. The continuous blue (respectively red) curve depicts the zero-error (respectively relaxed) non-signaling assisted achievable rate pairs for 7 copies of the binary adder channel.

largely beats the best found sum-rate of $\frac{\log_2(72)}{4} \simeq 1.5425$ achieved on 8 copies with the regular version. However the fact that the non-signaling assisted capacity region is strictly contained in the relaxed one is still open, as the same rates could potentially be achieved by the cost of using more copies of the channel.

4.4.1 Outer Bound Part of Theorem 4.22

In order to prove Proposition 4.28, we use a connection between hypothesis testing and relaxed non-signaling assisted codes as established in [Mat12] for point-to-point channels.

Definition 4.13 (Hypothesis Testing)**.** Given distributions $P^{(0)}$ and $P^{(1)}$ on the same space C, we define $\beta_{1-\varepsilon}(P^{(0)}, P^{(1)})$ to be the minimum type II error $\sum_{r \in C} T_r P^{(1)}(r)$ that can be achieved by statistical tests T which give a type I error no greater than ε, i.e. $\sum_{r \in C} T_r P^{(0)}(r) \geq 1 - \varepsilon$.

In other words, we have that:

$$\begin{aligned}\beta_{1-\varepsilon}(P^{(0)}, P^{(1)}) = \quad & \underset{T_r}{\text{minimize}} && \sum_{r\in C} T_r P^{(1)}(r) \\ & \text{subject to} && \sum_{r \in C} T_r P^{(0)}(r) \geq 1-\varepsilon \\ &&& 0 \leq T_r \leq 1 \ .\end{aligned} \tag{4.44}$$

Lemma 4.24. *For any relaxed non-signaling assisted code* $(p_{x_1,x_2}, r_{x_1,x_2,y})_{x_1\in\mathcal{X}_1,x_2\in\mathcal{X}_2,y\in\mathcal{Y}}$ *with* (k_1,k_2) *messages and a probability of success* $1-\varepsilon$*, if* $P_{X_1X_2}(x_1,x_2) = \frac{p_{x_1,x_2}}{k_1k_2}$ *and* $Y \in \mathcal{Y}$ *is the outcome of* W *on inputs* X_1, X_2*, we have:*

$$\begin{aligned}\beta_{1-\varepsilon}\left(P_{X_1X_2Y}, P_{X_1X_2}\times P_{Y|X_2}\right) &\leq \frac{1}{k_1} \\ \beta_{1-\varepsilon}\left(P_{X_1X_2Y}, P_{X_1X_2}\times P_{Y|X_1}\right) &\leq \frac{1}{k_2} \\ \beta_{1-\varepsilon}\left(P_{X_1X_2Y}, P_{X_1X_2}\times P_{Y}\right) &\leq \frac{1}{k_1k_2} \ .\end{aligned} \tag{4.45}$$

Remark. These three bounds are actually achieved with the same statistical test.

Proof. This result is a direct generalization of Theorem 9 in [Mat12] for point-to-point channels, itself a generalization of Theorem 27 in [PPV10] without non-signaling assistance.

Let us name $W_0 := W$ and W_1 a MAC yet to be defined. The coding strategy described by $r_{x_1,x_2,y}$ and p_{x_1,x_2} leads to a probability of success on channel $i \in \{0,1\}$ is given by:

$$\begin{aligned}1-\varepsilon_i &= \frac{1}{k_1k_2}\sum_{x_1,x_2,y} r_{x_1,x_2,y} W_i(y|x_1x_2) \\ &= \sum_{x_1,x_2,y:p_{x_1,x_2}>0} \frac{r_{x_1,x_2,y}}{p_{x_1,x_2}} W_i(y|x_1x_2)\frac{p_{x_1,x_2}}{k_1k_2} \qquad \text{since } 0 \leq r_{x_1,x_2,y} \leq p_{x_1,x_2} \\ &= \sum_{x_1,x_2,y} T_{x_1,x_2,y} W_i(y|x_1x_2)\frac{p_{x_1,x_2}}{k_1k_2} \ ,\end{aligned} \tag{4.46}$$

with $T_{x_1,x_2,y} := \frac{r_{x_1,x_2,y}}{p_{x_1,x_2}}$ if $p_{x_1,x_2} > 0$, and $T_{x_1,x_2,y} := 0$ otherwise.

If now Y is the output of the channel W_i, the joint distribution of X_1, X_2, Y is given by $P^{(i)}_{X_1X_2Y}(x_1,x_2,y) = W_i(y|x_1x_2)P_{X_1X_2}(x_1,x_2) = W_i(y|x_1x_2)\frac{p_{x_1,x_2}}{k_1k_2}$.

On the other hand, we have that for all x_1,x_2,y, $0 \leq T_{x_1,x_2,y} \leq 1$ since $0 \leq r_{x_1,x_2,y} \leq p_{x_1,x_2}$. So we get that:

$$1-\varepsilon_i = \sum_{x_1,x_2,y} T_{x_1,x_2,y} P^{(i)}_{X_1X_2Y}(x_1,x_2,y) \ .$$

Since $\sum_{x_1,x_2,y} T_{x_1,x_2,y}P^{(0)}_{X_1X_2Y}(x_1,x_2,y) \geq 1-\varepsilon_0$ and $0 \leq T_{x_1,x_2,y} \leq 1$, we have:

$$\beta_{1-\varepsilon_0}(P^{(0)}, P^{(1)}) \leq \sum_{x_1,x_2,y} T_{x_1,x_2,y} P^{(1)}_{X_1X_2Y}(x_1,x_2,y) = 1-\varepsilon_1 \ .$$

Let us now consider three general cases, depending on the fact that W_1 does not depend on x_1, x_2 or both: $W_1(y|x_1x_2) := Q^{(1)}(y|x_2)$; $W_1(y|x_1x_2) := Q^{(2)}(y|x_1)$; $W_1(y|x_1x_2) := Q^{(0)}(y)$. These will give respectively the three bounds we want.

First, let us consider the case where $W_1(y|x_1x_2) := Q^{(1)}(y|x_2)$ (the second case where $W_1(y|x_1x_2) := Q^{(2)}(y|x_1)$ being symmetric), we have that:

$$\begin{aligned} 1 - \varepsilon_1 &= \sum_{x_1,x_2,y} T_{x_1,x_2,y} Q^{(1)}(y|x_2) \frac{p_{x_1,x_2}}{k_1k_2} = \frac{1}{k_1k_2} \sum_{x_2,y} Q^{(1)}(y|x_2) \sum_{x_1} T_{x_1,x_2,y} p_{x_1,x_2} \\ &= \frac{1}{k_1k_2} \sum_{x_2,y} Q^{(1)}(y|x_2) \sum_{x_1} r_{x_1,x_2,y} \leq \frac{1}{k_1k_2} \sum_{x_2,y} Q^{(1)}(y|x_2) \frac{1}{k_1} \sum_{x_1} p_{x_1,x_2} \\ &= \frac{1}{k_1} \sum_{x_1,x_2} \frac{p_{x_1,x_2}}{k_1k_2} \sum_{y} Q^{(1)}(y|x_2) = \frac{1}{k_1} \sum_{x_1,x_2} \frac{p_{x_1,x_2}}{k_1k_2} = \frac{1}{k_1} \ . \end{aligned} \tag{4.47}$$

For the third case, when $W_1(y|x_1x_2) := Q^{(0)}(y)$, we have:

$$\begin{aligned} 1 - \varepsilon_1 &= \sum_{x_1,x_2,y} T_{x_1,x_2,y} Q^{(0)}(y) \frac{p_{x_1,x_2}}{k_1k_2} = \frac{1}{k_1k_2} \sum_{y} Q^{(0)}(y) \sum_{x_1,x_2} T_{x_1,x_2,y} p_{x_1,x_2} \\ &= \frac{1}{k_1k_2} \sum_{y} Q^{(0)}(y) \sum_{x_1,x_2} r_{x_1,x_2,y} \leq \frac{1}{k_1k_2} \sum_{y} Q^{(0)}(y) = \frac{1}{k_1k_2} \ . \end{aligned} \tag{4.48}$$

In those three cases, we have respectively $P^{(1)}_{X_1X_2Y} = P_{X_1X_2} \times Q^{(1)}_{Y|X_2}; P_{X_1X_2} \times Q^{(2)}_{Y|X_1}; P_{X_1X_2} \times Q^{(0)}_{Y}$. Specializing those cases with $Q^{(1)}_{Y|X_2} := P_{Y|X_2}; Q^{(2)}_{Y|X_1} := P_{Y|X_1}; Q^{(0)}_{Y} := P_Y$ and using the fact that $\beta_{1-\varepsilon_0}\left(P^{(0)}, P^{(1)}\right) \leq 1 - \varepsilon_1$ concludes the proof. □

Lemma 4.25. *For any relaxed non-signaling assisted code* $(p_{x_1,x_2}, r_{x_1,x_2,y})_{x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, y \in \mathcal{Y}}$ *with* (k_1, k_2) *messages and a probability of success* $1 - \varepsilon$*, if* $P_{X_1X_2}(x_1, x_2) = \frac{p_{x_1,x_2}}{k_1k_2}$ *and* $Y \in \mathcal{Y}$ *is the outcome of* W *on inputs* X_1, X_2*, we have:*

$$\begin{aligned} \log(k_1) &\leq \frac{I(X_1 : Y|X_2) + h(\varepsilon)}{1 - \varepsilon} \ , \\ \log(k_2) &\leq \frac{I(X_2 : Y|X_1) + h(\varepsilon)}{1 - \varepsilon} \ , \\ \log(k_1) + \log(k_2) &\leq \frac{I((X_1, X_2) : Y) + h(\varepsilon)}{1 - \varepsilon} \ . \end{aligned} \tag{4.49}$$

Proof. Thanks to Lemma 4.24, with the fact that $P_{X_1X_2} = P_{X_1|X_2} \times P_{X_2} = P_{X_2|X_1} \times P_{X_1}$, we have already:

$$\begin{aligned} &\beta_{1-\varepsilon}\left(P_{X_1X_2Y}, \left(P_{X_1|X_2} \times P_{Y|X_2}\right) \times P_{X_2}\right) \leq \frac{1}{k_1} \\ &\beta_{1-\varepsilon}\left(P_{X_1X_2Y}, \left(P_{X_2|X_1} \times P_{Y|X_1}\right) \times P_{X_1}\right) \leq \frac{1}{k_2} \\ &\beta_{1-\varepsilon}\left(P_{X_1X_2Y}, P_{X_1X_2} \times P_Y\right) \leq \frac{1}{k_1k_2} \end{aligned} \tag{4.50}$$

Following the steps of section G in [PPV10], since any hypothesis test is a binary-output transformation, by data-processing inequality for divergence, we have that:

$$\begin{aligned}
&d\left(1-\varepsilon||\beta_{1-\varepsilon}\left(P_{X_1X_2Y},\left(P_{X_1|X_2}\times P_{Y|X_2}\right)\times P_{X_2}\right)\right)\\
&=d\left(\beta_{1-\varepsilon}\left(P_{X_1X_2Y},P_{X_1X_2Y}\right)||\beta_{1-\varepsilon}\left(P_{X_1X_2Y},\left(P_{X_1|X_2}\times P_{Y|X_2}\right)\times P_{X_2}\right)\right)\\
&\leq D\left(P_{X_1X_2Y}||\left(P_{X_1|X_2}\times P_{Y|X_2}\right)\times P_{X_2}\right)=I(X_1:Y|X_2)
\end{aligned} \tag{4.51}$$

where the binary divergence $d(a||b):=a\log\left(\frac{a}{b}\right)+(1-a)\log\left(\frac{1-a}{1-b}\right)$ and satisfies, $d(a||b)\geq -h(a)-a\log(b)$ and thus:

$$\log\left(\frac{1}{b}\right)\leq\frac{d(a||b)+h(a)}{a}=\frac{d(a||b)+h(1-a)}{a},$$

This leads to:

$$\log(k_1)\leq\frac{1}{\log\left(\left(\beta_{1-\varepsilon}\left(P_{X_1X_2Y},\left(P_{X_1|X_2}\times P_{Y|X_2}\right)\times P_{X_2}\right)\right)\right)}\leq\frac{I(X_1:Y|X_2)+h(\varepsilon)}{1-\varepsilon}.$$

Similarly for the two other inequalities, since $D\left(P_{X_1X_2Y}||\left(P_{X_2|X_1}\times P_{Y|X_1}\right)\times P_{X_1}\right)=I(X_2:Y|X_1)$ and $D\left(P_{X_1X_2Y}||P_{X_1X_2}\times P_Y\right)=I((X_1,X_2):Y)$, we get:

$$\begin{aligned}
\log(k_1)&\leq\frac{I(X_1:Y|X_2)+h(\varepsilon)}{1-\varepsilon},\\
\log(k_2)&\leq\frac{I(X_2:Y|X_1)+h(\varepsilon)}{1-\varepsilon},\\
\log(k_1)+\log(k_2)&\leq\frac{I((X_1,X_2):Y)+h(\varepsilon)}{1-\varepsilon}.
\end{aligned} \tag{4.52}$$

□

In order to show additivity of the outer bound, we use the following lemma.

Lemma 4.26. *For any distribution $P_{X_1^nX_2^n}$ of (X_1^n,X_2^n), if $Y^n\in\mathcal{Y}^n$ is the outcome of W^n on inputs X_1^n,X_2^n, we have:*

$$\begin{aligned}
I(X_1^n:Y^n|X_2^n)&\leq\sum_{i=1}^n I(X_{1,i}:Y_i|X_{2,i})\\
I(X_2^n:Y^n|X_1^n)&\leq\sum_{i=1}^n I(X_{2,i}:Y_i|X_{1,i})\\
I((X_1^n,X_2^n):Y^n)&\leq\sum_{i=1}^n I((X_{1,i},X_{2,i}):Y_i).
\end{aligned} \tag{4.53}$$

Proof. Consider n copies of the MAC W. Let us write $X_{1,-i} := X_{1,1} \dots X_{1,i-1} X_{1,i+1} \dots X_{1,n}$ and $Z^n := Z_1 \dots Z_n$. We have:

$$\begin{aligned} I(X_1^n : Y^n | X_2^n) = I(X_1^n : Y^n | X_2^n) \\ &= \sum_{i=1}^{n} I(X_1^n : Y_i | X_2^n Y^{i-1}) \text{ by the chain rule} \\ &= \sum_{i=1}^{n} I(X_{1,i} : Y_i | X_2^n Y^{i-1}) + \sum_{i=1}^{n} I(X_{1,-i} : Y_i | X_2^n Y^{i-1} X_{1,i}) \\ &= \sum_{i=1}^{n} I(X_{1,i} : Y_i | X_2^n Y^{i-1}) \, , \end{aligned} \tag{4.54}$$

where the last equality comes from Lemma 4.27. As a result,

$$\begin{aligned} I(X_1^n : Y^n | X_2^n) &= \sum_{i=1}^{n} H(Y_i | X_2^n Y^{i-1}) - H(Y_i | X_2^n Y^{i-1} X_{1,i}) \\ &= \sum_{i=1}^{n} H(Y_i | X_2^n Y^{i-1}) - H(Y_i | X_{2,i} X_{1,i}) \text{ since } X_{2,-i} Y^{i-1} \to (X_{1,i}, X_{2,i}) \to Y_i \text{ Markov} \\ &\leq \sum_{i=1}^{n} H(Y_i | X_{2,i}) - H(Y_i | X_{2,i} X_{1,i}) = \sum_{i=1}^{n} I(X_{1,i} : Y_i | X_{2,i}) \, . \end{aligned} \tag{4.55}$$

Symmetrically by switching the roles of X_1 and X_2, we get the second part of Lemma 4.26.

For the sum-rate case:

$$\begin{aligned} I((X_1^n, X_2^n) : Y^n) &= \sum_{i=1}^{n} I((X_1^n, X_2^n) : Y_i | Y^{i-1}) \text{ by the chain rule} \\ &= \sum_{i=1}^{n} I((X_{1,i}, X_{2,i}) : Y_i | Y^{i-1}) + \sum_{i=1}^{n} I((X_{1,-i}, X_{2,-i}) : Y_i | Y^{i-1} X_{1,i} X_{2,i}) \\ &= \sum_{i=1}^{n} I((X_{1,i}, X_{2,i}) : Y_i | Y^{i-1}) \text{ since } (X_{1,-i}, X_{2,-i}) \to Y^{i-1} X_{1,i} X_{2,i} \to Y_i \text{ Markov} \\ &= \sum_{i=1}^{n} H(Y_i | Y^{i-1}) - H(Y_i | Y^{i-1} X_{1,i} X_{2,i}) \\ &= \sum_{i=1}^{n} H(Y_i | Y^{i-1}) - H(Y_i | X_{1,i} X_{2,i}) \text{ since } Y^{i-1} \to (X_{1,i}, X_{2,i}) \to Y_i \text{ Markov chain} \\ &\leq \sum_{i=1}^{n} H(Y_i) - H(Y_i | X_{2,i} X_{1,i}) = \sum_{i=1}^{n} I((X_{1,i}, X_{2,i}) : Y_i) \, . \end{aligned}$$

□

We next prove a technical lemma that was used in the previous proof.

Lemma 4.27. *For any distribution $P_{X_1^n X_2^n}$ of (X_1^n, X_2^n), if $Y^n \in \mathcal{Y}^n$ is the outcome of W^n on inputs X_1^n, X_2^n, we have:*

$$I(X_{1,-i} : Y_i | X_{1,i} X_2^n Y^{i-1}) = 0 \, .$$

Proof. Let us show that, conditioned on any particular instance of $X_{1,i} = x_{i,1}, X_2^n = x_2^n, Y_1^{i-1} = y^{i-1}$, $X_{1,-i}$ and Y_i are independent.

We have:

$$\mathbb{P}\left(Y_i = y_i | X_{1,i} = x_{i,1}, X_2^n = x_2^n, Y_1^{i-1} = y^{i-1}\right) = \mathbb{P}\left(Y_i = y_i | X_{1,i} = x_{i,1}, X_{2,i} = x_{2,i}\right) = W(y_i | x_{1,i} x_{2,i}) ,$$

by definition of the law of Y_i. On the other hand, we have that:

$$\mathbb{P}\left(X_1^n = x_1^n, X_2^n = x_2^n, Y^n = y_n\right) = \mathbb{P}\left(X_1^n = x_1^n, X_2^n = x_2^n\right) \prod_{j=1}^{n} W(y_j | x_{1,j} x_{2,j}) .$$

Thus, we have:

$$\begin{aligned} \mathbb{P}\left(X_{1,i} = x_{i,1}, X_2^n = x_2^n, Y_1^{i-1} = y^{i-1}\right) &= \sum_{x_{1,-i}, x_2^n, y_i, \ldots, y_n} \mathbb{P}\left(X_1^n = x_1^n, X_2^n = x_2^n\right) \prod_{j=1}^{n} W(y_j | x_{1,j} x_{2,j}) \\ &= \sum_{x_{1,-i}, x_2^n} \mathbb{P}\left(X_1^n = x_1^n, X_2^n = x_2^n\right) \prod_{j=1}^{i-1} W(y_j | x_{1,j} x_{2,j}) \prod_{j=i}^{n} \left(\sum_{y_j} W(y_j | x_{1,j} x_{2,j}) \right) \\ &= \sum_{x_{1,-i}, x_2^n} \mathbb{P}\left(X_1^n = x_1^n, X_2^n = x_2^n\right) \prod_{j=1}^{i-1} W(y_j | x_{1,j} x_{2,j}) . \end{aligned} \tag{4.56}$$

And then:

$$\mathbb{P}\left(X_{1,-i} = x_{1,-i} | X_{1,i} = x_{i,1}, X_2^n = x_2^n, Y_1^{i-1} = y^{i-1}\right) = \frac{\sum_{x_2^n} \mathbb{P}\left(X_1^n = x_1^n, X_2^n = x_2^n\right) \prod_{j=1}^{i-1} W(y_j | x_{1,j} x_{2,j})}{\sum_{x_{1,-i}, x_2^n} \mathbb{P}\left(X_1^n = x_1^n, X_2^n = x_2^n\right) \prod_{j=1}^{i-1} W(y_j | x_{1,j} x_{2,j})}$$

But:

$$\begin{aligned} &\mathbb{P}\left(X_{1,-i} = x_{1,-i}, Y_i = y_i | X_{1,i} = x_{i,1}, X_2^n = x_2^n, Y_1^{i-1} = y^{i-1}\right) \\ &= \frac{\sum_{x_2^n} \mathbb{P}\left(X_1^n = x_1^n, X_2^n = x_2^n\right) \prod_{j=1}^{i} W(y_j | x_{1,j} x_{2,j})}{\sum_{x_{1,-i}, x_2^n} \mathbb{P}\left(X_1^n = x_1^n, X_2^n = x_2^n\right) \prod_{j=1}^{i-1} W(y_j | x_{1,j} x_{2,j})} \\ &= \mathbb{P}\left(X_{1,-i} = x_{1,-i} | X_{1,i} = x_{i,1}, X_2^n = x_2^n, Y_1^{i-1} = y^{i-1}\right) W(y_i | x_{1,i} x_{2,i}) \\ &= \mathbb{P}\left(X_{1,-i} = x_{1,-i} | X_{1,i} = x_{i,1}, X_2^n = x_2^n, Y_1^{i-1} = y^{i-1}\right) \mathbb{P}\left(Y_i = y_i | X_{1,i} = x_{i,1}, X_2^n = x_2^n, Y_1^{i-1} = y^{i-1}\right) . \end{aligned} \tag{4.57}$$

Thus, conditioned on any particular instance of $X_{1,i} = x_{i,1}, X_2^n = x_2^n, Y_1^{i-1} = y^{i-1}$, $X_{1,-i}$ and Y_i are independent, and so $I(X_{1,-i} : Y_i | X_{1,i} X_2^n Y^{i-1}) = 0$. □

Combining the previous results gives the desired outer bound.

Proposition 4.28 (Outer bound part of Theorem 4.22)**.** *If a rate pair is achievable with relaxed non-signaling assistance then it is in the closure of the convex hull of all* (R_1, R_2) *satisfying:*

$$R_1 < I(X_1 : Y | X_2) \ , \ R_2 < I(X_2 : Y | X_1) \ , \ R_1 + R_2 < I((X_1, X_2) : Y) \ ,$$

for (X_1, X_2) *following some law* $P_{X_1 X_2}$ *on* $\mathcal{X}_1 \times \mathcal{X}_2$*, and* $Y \in \mathcal{Y}$ *the outcome of* W *on inputs* X_1, X_2*.*

Proof. Consider (R_1, R_2) achievable with relaxed non-signaling assistance: we have a sequence of relaxed non-signaling assisted codes for n copies of the MAC W with $k_1 = 2^{nR_1}, k_2 = 2^{nR_2}$ messages and an error probability $\varepsilon_n \underset{n\to+\infty}{\to} 0$, along with associated distributions of $X_1^n X_2^n Y^n$.

Thus combining Lemma 4.25 and Lemma 4.26, we have that:

$$\begin{aligned}
R_1 &\leq \frac{1}{n}\frac{I(X_1^n : Y^n|X_2^n) + h(\varepsilon_n)}{1-\varepsilon_n} \leq \frac{1}{n}\frac{\sum_{i=1}^n I(X_{1,i} : Y_i|X_{2,i}) + h(\varepsilon_n)}{1-\varepsilon_n} \ , \\
R_2 &\leq \frac{1}{n}\frac{I(X_2^n : Y^n|X_1^n) + h(\varepsilon_n)}{1-\varepsilon_n} \leq \frac{1}{n}\frac{\sum_{i=1}^n I(X_{2,i} : Y_i|X_{1,i}) + h(\varepsilon_n)}{1-\varepsilon_n} \ , \\
R_1 + R_2 &\leq \frac{1}{n}\frac{I((X_1^n, X_2^n) : Y^n) + h(\varepsilon_n)}{1-\varepsilon_n} \leq \frac{1}{n}\frac{\sum_{i=1}^n I((X_{1,i}, X_{2,i}) : Y_i) + h(\varepsilon_n)}{1-\varepsilon_n} \ .
\end{aligned} \tag{4.58}$$

Then let us consider some random variable Q uniform on $[n]$ and independent from (X_1^n, X_2^n, Y^n). Then we can write:

$$\sum_{i=1}^n I(X_{1,i} : Y_i|X_{2,i}) = \sum_{i=1}^n I(X_{1,i} : Y_i|X_{2,i}, Q = i) = nI(X_{1,Q} : Y_Q|X_{2,Q}, Q) \ .$$

Since Y_Q conditioned on $X_{1,Q}$ and $X_{2,Q}$ still follows the law of the MAC $W(y|x_1x_2)$, we can take $X_1 = X_{1,Q}, X_2 = X_{2,Q}$, and then the output of the channel Y satisfies $Y = Y_Q$, and thus we obtain:

$$R_1 \leq \frac{I(X_1 : Y|X_2, Q) + \frac{h(\varepsilon_n)}{n}}{1-\varepsilon_n} \ .$$

Doing this similarly on the other conditional mutual informations, we get:

$$\begin{aligned}
R_1 &\leq \frac{I(X_1 : Y|X_2, Q) + \frac{h(\varepsilon_n)}{n}}{1-\varepsilon_n} \ , \\
R_2 &\leq \frac{I(X_2 : Y|X_1, Q) + \frac{h(\varepsilon_n)}{n}}{1-\varepsilon_n} \ , \\
R_1 + R_2 &\leq \frac{I((X_1, X_2) : Y|Q) + \frac{h(\varepsilon_n)}{n}}{1-\varepsilon_n} \ .
\end{aligned} \tag{4.59}$$

By taking the limit as n goes to infinity, since the limit of ε_n is 0, then the limit of $\frac{h(\varepsilon_n)}{n}$ is 0 as well and we get that (R_1, R_2) must be in the set of rate pairs such that:

$$\begin{aligned}
R_1 &\leq I(X_1 : Y|X_2, Q) \ , \\
R_2 &\leq I(X_2 : Y|X_1, Q) \ , \\
R_1 + R_2 &\leq I((X_1, X_2) : Y|Q) \ ,
\end{aligned} \tag{4.60}$$

for some uniform Q in a finite set, (X_1, X_2) any joint law depending on Q, and Y the output of W on inputs (X_1, X_2).

Finally, in order to show that this is the right region, one has only to see that the corner points of this region, such as for instance $(I(X_1 : Y|Q), I(X_2 : Y|X_1, Q))$, are finite convex combination of the points $(I(X_1 : Y|Q = q), I(X_2 : Y|X_1, Q = q))$ which are all in the capacity region of the theorem by taking $(X_1X_2) \sim P_{X_1X_2|Q=q}$. This implies that (R_1, R_2) is in the convex hull of that region, so we can drop the random variable Q and the proof is completed. □

The main consequence of that outer bound on the relaxed non-signaling assisted capacity region is that it holds also for the non-signaling assisted capacity region thanks to Corollary 4.21:

Corollary 4.29 (Outer Bound on the Non-Signaling Assisted Capacity Region)**.** *If a rate pair is achievable with non-signaling assistance, then it is in the closure of the convex hull of all (R_1, R_2) satisfying:*

$$R_1 < I(X_1 : Y|X_2)\ ,\ R_2 < I(X_2 : Y|X_1)\ ,\ R_1 + R_2 < I((X_1, X_2) : Y)\ ,$$

for (X_1, X_2) following any law $P_{X_1X_2}$ on $\mathcal{X}_1 \times \mathcal{X}_2$, and $Y \in \mathcal{Y}$ the outcome of W on inputs X_1, X_2.

4.4.2 Achievability Part of Theorem 4.22

In order to construct the relaxed non-signaling assisted code for achievability, we will need the notions of jointly and conditional typical sets. We will consider the following typical sets defined in Chapter 2.5 of [GK11]: $\mathcal{T}_\varepsilon^n(X_1, X_2, Y)$, $\mathcal{T}_\varepsilon^n(X_1, X_2)$, $\mathcal{T}_\varepsilon^n(Y)$, $\mathcal{T}_\varepsilon^n(X_1|x_2^n)$, $\mathcal{T}_\varepsilon^n(X_2|x_1^n)$, $\mathcal{T}_\varepsilon^n(X_1|x_2^n, y^n)$, $\mathcal{T}_\varepsilon^n(X_2|x_1^n, y^n)$, $\mathcal{T}_\varepsilon^n(X_1, X_2|y^n)$. Recall that:

Definition 4.14 (Typical set and conditional typical set)**.** We have the following definitions:

1. $\mathcal{T}_\varepsilon^n(X_1, X_2) := \{(x_1^n, x_2^n) : |\pi(x_1, x_2|x_1^n, x_2^n) - P_{X_1X_2}(x_1, x_2)| \leq \varepsilon P_{X_1X_2}(x_1, x_2)$ for all $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}$ where $\pi(x_1, x_2|x_1^n, x_2^n) := \frac{|\{i : (x_{1,i}, x_{2,i}) = (x_1, x_2)\}|}{n}$. This definition generalizes for any t-uple of variables.

2. $\forall y^n \in \mathcal{T}_\varepsilon^n(Y), \mathcal{T}_\varepsilon^n(X_1, X_2|y^n) := \{(x_1^n, x_2^n) : (x_1^n, x_2^n, y^n) \in \mathcal{T}_\varepsilon^n(X_1, X_2, Y)\}$

A crucial property of such typical sets is the typical average lemma:

Lemma 4.30 (Typical Average Lemma [GK11])**.** *Let $(x_1^n, x_2^n) \in \mathcal{T}_\varepsilon^n(X_1, X_2)$. Then for any nonnegative function g on $\mathcal{X}_1 \times \mathcal{X}_2$:*

$$(1 - \varepsilon)\mathbb{E}[g(X_1, X_2)] \leq \frac{1}{n}\sum_{i=1}^{n} g(x_{1,i}, x_{2,i}) \leq (1 + \varepsilon)\mathbb{E}[g(X_1, X_2)]\ .$$

In particular, with this tool, we can derive the following properties:

Lemma 4.31 (Properties of typical sets [GK11])**.** *We have, among others, the following statements about typical sets:*

1. $\forall (x_1^n, x_2^n) \in \mathcal{T}_\varepsilon^n(X_1, X_2), 2^{-n(1+\varepsilon)H(X_1,X_2)} \leq P_{X_1^nX_2^n}(x_1^n, x_2^n) \leq 2^{-n(1-\varepsilon)H(X_1,X_2)}$.

2. $\lim\limits_{n\to+\infty} \mathbb{P}\left((X_1^n, X_2^n) \in \mathcal{T}_\varepsilon^n(X_1, X_2)\right) = 1$.

3. $|\mathcal{T}_\varepsilon^n(X_1, X_2)| \leq 2^{n(1+\varepsilon)H(X_1,X_2)}$.

4. *For n sufficiently large,* $|\mathcal{T}_\varepsilon^n(X_1, X_2)| \geq (1-\varepsilon)2^{n(1-\varepsilon)H(X_1,X_2)}$.

5. *If* $(x_1^n, x_2^n) \in \mathcal{T}_\varepsilon^n(X_1, X_2)$ *then* $x_1^n \in \mathcal{T}_\varepsilon^n(X_1)$ *and* $x_2^n \in \mathcal{T}_\varepsilon^n(X_2)$.

6. $\forall y^n \in \mathcal{T}_\varepsilon^n(Y), \mathcal{T}_\varepsilon^n(X_1, X_2|y^n) \subseteq \mathcal{T}_\varepsilon^n(X_1, X_2)$.

7. $\forall (x_1^n, x_2^n, y^n) \in \mathcal{T}_\varepsilon^n(X_1, X_2, Y), 2^{-n(1+\varepsilon)H(X_1,X_2|Y)} \leq P_{X_1^nX_2^nY^n}(x_1^n, x_2^n|y^n) \leq 2^{-n(1-\varepsilon)H(X_1,X_2|Y)}$.

8. $\forall y^n \in \mathcal{T}_\varepsilon^n(Y), |\mathcal{T}_\varepsilon^n(X_1, X_2|y^n)| \leq 2^{n(1+\varepsilon)H(X_1,X_2|Y)}$.

9. *For $\varepsilon' < \varepsilon$ and n sufficiently large, we get* $\forall y^n \in \mathcal{T}_{\varepsilon'}^n(Y), |\mathcal{T}_\varepsilon^n(X_1, X_2|y^n)| \geq (1-\varepsilon)2^{n(1-\varepsilon)H(X_1,X_2|Y)}$.

Proof. We reproduce the proof of the last statement here to emphasize on the fact that there is an n_0 such that for all $n \geq n_0$ and for all $y^n \in \mathcal{T}_{\varepsilon'}^n(Y)$, the property holds.

For any $\varepsilon > \varepsilon' > 0$, let us show that there exists n such that we have:

$$\forall y^n \in \mathcal{T}_{\varepsilon'}^n(Y), \mathbb{P}\left((X_1^n, X_2^n, y^n) \in \mathcal{T}_\varepsilon^n(X_1, X_2, Y)\right) \geq 1 - \varepsilon \ ,$$

where X_1^n, X_2^n are drawn from the distribution $P_{X_1^nX_2^n|Y^n=y^n}$. This will imply the statement. Indeed, we have that:

$$\begin{aligned}\mathbb{P}\left((X_1^n, X_2^n, y^n) \in \mathcal{T}_\varepsilon^n(X_1, X_2, Y)\right) &= \sum_{(x_1^n,x_2^n)\in\mathcal{T}_\varepsilon^n(X_1,X_2|y^n)} P_{X_1^nX_2^n|Y^n}(x_1^n, x_2^n|y^n) \\ &\leq |\mathcal{T}_\varepsilon^n(X_1, X_2|y^n)|2^{-n(1-\varepsilon)H(X_1,X_2|Y)} \ ,\end{aligned} \tag{4.61}$$

since $P_{X_1^nX_2^n|Y^n}(x_1^n, x_2^n|y^n) \leq 2^{-n(1-\varepsilon)H(X_1,X_2|Y)}$ as $(x_1^n, x_2^n, y^n) \in \mathcal{T}_\varepsilon^n(X_1, X_2, Y)$. Thus, we have that $|\mathcal{T}_\varepsilon^n(X_1, X_2|y^n)| \geq (1-\varepsilon)2^{n(1-\varepsilon)H(X_1,X_2|Y)}$. In order to prove our result, we take the proof in Appendix 2A of [GK11]. We take $y^n \in \mathcal{T}_{\varepsilon'}^n(\mathcal{Y})$ and $(X_1^n, X_2^n) \sim P_{X_1^nX_2^n|Y^n}(x_1^n, x_2^n|y^n) = \prod_{i=1}^n P_{X_1X_2|Y}(x_{1,i}, x_{2,i}|y_i)$. Applied to our choice of variables, we have the following result :

$$\begin{aligned}&\mathbb{P}\left((X_1^n, X_2^n, y^n) \notin \mathcal{T}_\varepsilon^n(X_1, X_2, Y)\right) \\ &= \mathbb{P}\left(\exists (x_1, x_2, y) : |\pi(x_1, x_2, y|X_1^n, X_2^n, y^n) - P_{X_1X_2Y}(x_1, x_2, y)| > \varepsilon P_{X_1X_2Y}(x_1, x_2, y)\right) \\ &\leq \sum_{x_1,x_2,y} \mathbb{P}\left(|\pi(x_1, x_2, y|X_1^n, X_2^n, y^n) - P_{X_1X_2Y}(x_1, x_2, y)| > \varepsilon P_{X_1X_2Y}(x_1, x_2, y)\right) \text{ by union bound,} \\ &\leq \sum_{x_1,x_2,y} \mathbb{P}\left(\frac{\pi(x_1, x_2, y|X_1^n, X_2^n, y^n)}{\pi(y|y^n)} > \frac{1+\varepsilon}{1+\varepsilon'}P_{X_1X_2|Y}(x_1, x_2|y)\right) \\ &+ \sum_{x_1,x_2,y} \mathbb{P}\left(\frac{\pi(x_1, x_2, y|X_1^n, X_2^n, y^n)}{\pi(y|y^n)} < \frac{1-\varepsilon}{1-\varepsilon'}P_{X_1X_2|Y}(x_1, x_2|y)\right) \text{ by calculations of [GK11].}\end{aligned} \tag{4.62}$$

However, since $\varepsilon' < \varepsilon$, we have $\frac{1+\varepsilon}{1+\varepsilon'} > 1$ and $\frac{1-\varepsilon}{1-\varepsilon'} < 1$. We will show that for all x_1, x_2, y with $P_Y(y) > 0$, we have $\frac{\pi(x_1,x_2,y|X_1^n,X_2^n,y^n)}{\pi(y|y^n)} \underset{n\to+\infty}{\to} P_{X_1X_2|Y}(x_1, x_2|y)$ in probability, with a convergence rate independent from $y^n \in \mathcal{T}^n_{\varepsilon'}(Y)$, which will be enough to conclude the proof.

Let us fix some x_1, x_2, y with $P_Y(y) > 0$. Since $y^n \in \mathcal{T}^n_{\varepsilon'}(Y)$, we have in particular $(1-\varepsilon')P_Y(y) \leq \pi(y|y^n) \leq (1+\varepsilon')P_Y(y)$. Thus $N := |\{i : y_i = y\}| = n\pi(y|y^n) \geq (1-\varepsilon')P_Y(y)n$. Then we have:

$$\frac{\pi(x_1,x_2,y|X_1^n,X_2^n,y^n)}{\pi(y|y^n)} = \frac{1}{N}\sum_{i \in S} Z_i \text{ with } Z_i := \mathbb{1}_{(X_{1,i},X_{2,i})=(x_1,x_2)} \text{ and } S := \{i : y_i = y\}\,.$$

Thus, all Z_i with $i \in S$ are independent and follow the same law:

$$Z_i := \begin{cases} 1 & \text{with probability } P_{X_1X_2|Y}(x_1,x_2|y) \\ 0 & \text{otherwise} \end{cases}$$

Furthermore, we have $\mathbb{E}[Z_i] = P_{X_1X_2|Y}(x_1,x_2|y)$, and all Z_i have the same variance $\sigma^2_{x_1,x_2|y} < +\infty$ (depending only on X_1, X_2, Y, x_1, x_2, y). Thus we can apply Chebyshev inequality:

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{i\in S} Z_i - P_{X_1X_2|Y}(x_1,x_2|y)\right| \geq \eta\right) \leq \frac{\sigma^2_{x_1,x_2|y}}{N\eta^2}\,.$$

However, since $N \geq (1-\varepsilon')P_Y(y)n$, we get:

$$\mathbb{P}\left(\left|\frac{\pi(x_1,x_2,y|X_1^n,X_2^n,y^n)}{\pi(y|y^n)} - P_{X_1X_2|Y}(x_1,x_2|y)\right| \geq \eta\right) \leq \frac{\sigma^2_{x_1,x_2|y}}{\eta^2(1-\varepsilon')P_Y(y)n} \underset{n\to+\infty}{\to} 0\,.$$

Thus, we have $\frac{\pi(x_1,x_2,y|X_1^n,X_2^n,y^n)}{\pi(y|y^n)} \underset{n\to+\infty}{\to} P_{X_1X_2|Y}(x_1,x_2|y)$ in probability with a convergence rate independent from $y^n \in \mathcal{T}^n_{\varepsilon'}(Y)$. □

Proposition 4.32 (Achievability part of Theorem 4.22). *If a rate pair is in the closure of the convex hull of (R_1, R_2) satisfying:*

$$R_1 < I(X_1 : Y|X_2)\;,\; R_2 < I(X_2 : Y|X_1)\;,\; R_1 + R_2 < I((X_1,X_2) : Y)\;,$$

for (X_1, X_2) following some law $P_{X_1X_2}$ on $\mathcal{X}_1 \times \mathcal{X}_2$, and $Y \in \mathcal{Y}$ the outcome of W on inputs X_1, X_2, then it is in $\mathcal{C}^{\overline{\mathrm{NS}}}(W)$.

Proof. Let us fix $\varepsilon, \varepsilon' \in (0,1)$ such that $\varepsilon' < \varepsilon \leq \frac{1}{2}$. Let $n \in \mathbb{N}$ which will be chosen large enough during the proof.

We consider n independent random variables $(X_{1,i}X_{2,i}Y_i) \sim P_{X_1X_2Y}$, with $P_{X_1X_2Y}(x_{1,i},x_{2,i},y_i) = W(y_i|x_{1,i}x_{2,i})P_{X_1X_2}(x_{1,i},x_{2,i})$. We call $P_{X_1^nX_2^nY^n}$ the law of their product. We have then $P_{X_1^nX_2^n}(x_1^n,x_2^n) := \prod_{i=1}^n P_{X_1X_2}(x_{1,i},x_{2,i})$. If $\hat{Y}$ is the output of $W^{\otimes n}$ on $X_1^nX_2^n$, we have that:

$$P_{X_1^n X_2^n \hat{Y}}(x_1^n, x_2^n, y^n) = W^{\otimes n}(y^n | x_1^n x_2^n) P_{X_1^n X_2^n}(x_1^n, x_2^n) = W^{\otimes n}(y^n | x_1^n x_2^n) \prod_{i=1}^{n} P_{X_1 X_2}(x_{1,i}, x_{2,i})$$
$$= \prod_{i=1}^{n} W(y_i | x_{1,i} x_{2,i}) P_{X_1 X_2}(x_{1,i}, x_{2,i}) = \prod_{i=1}^{n} P_{X_1 X_2 Y}(x_{1,i}, x_{2,i}, y_i) \,. \tag{4.63}$$

Thus, $\hat{Y}$ follows the product law of Y_i, i.e. $\hat{Y} = Y^n$.

Let us consider C_1, C_2, C_3 some positive numbers independent from n and ε which we will define later, $k_1 = 2^{nR_1}$, $k_2 = 2^{nR_2}$ integers with (R_1, R_2) positive rates such that:

$$\begin{aligned} R_1 &\leq I(X_1 : Y | X_2) - \frac{1}{n} - C_1 \varepsilon \,, \\ R_2 &\leq I(X_2 : Y | X_1) - \frac{1}{n} - C_2 \varepsilon \,, \\ R_1 + R_2 &\leq I((X_1, X_2) : Y) - \frac{1}{n} - C_3 \varepsilon \,. \end{aligned} \tag{4.64}$$

We define a solution of $\mathrm{S}^{\overline{\mathrm{NS}}}(W^{\otimes n}, 2^{nR_1}, 2^{nR_2})$ in the following way:

$$p_{x_1^n, x_2^n} := \begin{cases} \frac{2^{n(R_1+R_2)} P_{X_1^n X_2^n}(x_1^n, x_2^n)}{\sum_{(x_1^n, x_2^n) \in \mathcal{T}_\varepsilon^n(X_1, X_2)} P_{X_1^n X_2^n}(x_1^n, x_2^n)} & \text{if } (x_1^n, x_2^n) \in \mathcal{T}_\varepsilon^n(X_1, X_2) \\ 0 & \text{otherwise} \end{cases}$$

and

$$r_{x_1^n, x_2^n, y^n} := \begin{cases} p_{x_1^n, x_2^n} & \text{if } (x_1^n, x_2^n, y^n) \in \mathcal{T}_{\varepsilon'}^n(X_1, X_2, Y) \\ 0 & \text{otherwise} \end{cases}$$

By construction, the constraint $0 \leq r_{x_1^n, x_2^n, y^n} \leq p_{x_1^n, x_2^n}$ is satisfied. We have also that:

$$\sum_{x_1^n, x_2^n} p_{x_1^n, x_2^n} = \sum_{(x_1^n, x_2^n) \in \mathcal{T}_\varepsilon^n(X_1, X_2)} \frac{2^{n(R_1+R_2)} P_{X_1^n X_2^n}(x_1^n, x_2^n)}{\sum_{(x_1^n, x_2^n) \in \mathcal{T}_\varepsilon^n(X_1, X_2)} P_{X_1^n X_2^n}(x_1^n, x_2^n)} = 2^{n(R_1+R_2)} = k_1 k_2 \,.$$

If $(x_1^n, x_2^n, y^n) \in \mathcal{T}_{\varepsilon'}^n(X_1, X_2, Y)$, we have that $(x_1^n, x_2^n) \in \mathcal{T}_{\varepsilon'}^n(X_1, X_2) \subseteq \mathcal{T}_\varepsilon^n(X_1, X_2)$, so in that case:

$$r_{x_1^n, x_2^n, y^n} = \frac{2^{n(R_1+R_2)} P_{X_1^n X_2^n}(x_1^n, x_2^n)}{\sum_{(x_1^n, x_2^n) \in \mathcal{T}_\varepsilon^n(X_1, X_2)} P_{X_1^n X_2^n}(x_1^n, x_2^n)} \,.$$

If $y^n \notin \mathcal{T}_{\varepsilon'}^n(Y)$, then for all (x_1^n, x_2^n), $(x_1^n, x_2^n, y^n) \notin \mathcal{T}_{\varepsilon'}^n(X_1, X_2, Y)$, so $\sum_{x_1^n, x_2^n} r_{x_1^n, x_2^n, y^n} = 0 \leq 1$ in that case.

Otherwise, if $y^n \in \mathcal{T}_{\varepsilon'}^n(Y)$, then:

$$\begin{aligned}
\sum_{x_1^n,x_2^n} r_{x_1^n,x_2^n,y^n} &= 2^{n(R_1+R_2)} \frac{\sum_{(x_1^n,x_2^n)\in\mathcal{T}_{\varepsilon'}^n(X_1,X_2|y^n)} P_{X_1^nX_2^n}(x_1^n,x_2^n)}{\sum_{(x_1^n,x_2^n)\in\mathcal{T}_{\varepsilon}^n(X_1,X_2)} P_{X_1^nX_2^n}(x_1^n,x_2^n)}\\
&\leq 2^{n(R_1+R_2)} \frac{\sum_{(x_1^n,x_2^n)\in\mathcal{T}_{\varepsilon}^n(X_1,X_2|y^n)} P_{X_1^nX_2^n}(x_1^n,x_2^n)}{\sum_{(x_1^n,x_2^n)\in\mathcal{T}_{\varepsilon}^n(X_1,X_2)} P_{X_1^nX_2^n}(x_1^n,x_2^n)}\\
&\leq 2^{n(R_1+R_2)} \frac{2^{-n(1-\varepsilon)H(X_1,X_2)}}{2^{-n(1+\varepsilon)H(X_1,X_2)}} \frac{|\mathcal{T}_{\varepsilon}^n(X_1,X_2|y^n)|}{|\mathcal{T}_{\varepsilon}^n(X_1,X_2)|} \text{ since } (x_1^n,x_2^n)\in\mathcal{T}_{\varepsilon}^n(X_1,X_2)\\
&= 2^{n(R_1+R_2+2\varepsilon H(X_1,X_2))} \frac{|\mathcal{T}_{\varepsilon}^n(X_1,X_2|y^n)|}{|\mathcal{T}_{\varepsilon}^n(X_1,X_2)|}\ .
\end{aligned} \tag{4.65}$$

But $|\mathcal{T}_{\varepsilon}^n(X_1,X_2|y^n)| \leq 2^{n(1+\varepsilon)H(X_1,X_2|Y)}$ and for a large enough n we have that $|\mathcal{T}_{\varepsilon}^n(X_1,X_2)| \geq (1-\varepsilon)2^{n(1-\varepsilon)H(X_1,X_2)} \geq 2^{n\left((1-\varepsilon)H(X_1,X_2)-\frac{1}{n}\right)}$, so in that case:

$$\sum_{x_1^n,x_2^n} r_{x_1^n,x_2^n,y^n} \leq 2^{n(R_1+R_2+2\varepsilon H(X_1,X_2))} \frac{2^{n(1+\varepsilon)H(X_1,X_2|Y)}}{2^{n\left((1-\varepsilon)H(X_1,X_2)-\frac{1}{n}\right)}} = 2^{n\left(R_1+R_2-I(X_1,X_2:Y)+\frac{1}{n}+C_3\varepsilon\right)} \leq 1\,,$$

since $I(X_1,X_2:Y) = H(X_1,X_2) - H(X_1,X_2|Y)$ and $R_1+R_2 \leq I(X_1,X_2:Y) - \frac{1}{n} - C_3\varepsilon$, with $C_3 := H(X_1,X_2|Y) + 3H(X_1,X_2)$.

Let us focus on the constraint $\sum_{x_1^n} p_{x_1^n,x_2^n} \geq k_1 \sum_{x_1^n} r_{x_1^n,x_2^n,y^n}$ (the symmetric constraint $\sum_{x_2^n} p_{x_1^n,x_2^n} \geq k_2 \sum_{x_2^n} r_{x_1^n,x_2^n,y^n}$ will be achieved for symmetric reasons).

Let us fix (x_2^n,y^n). If $(x_2^n,y^n) \notin \mathcal{T}_{\varepsilon'}^n(X_2,Y)$, then for all x_1^n, $(x_1^n,x_2^n,y^n) \notin \mathcal{T}_{\varepsilon'}^n(X_1,X_2,Y)$, thus $r_{x_1^n,x_2^n,y^n} = 0$ and the constraint is fulfilled. Let us assume that $(x_2^n,y^n) \in \mathcal{T}_{\varepsilon'}^n(X_2,Y)$. Since $r_{x_1^n,x_2^n,y^n} > 0$ implies that $(x_1^n,x_2^n,y^n) \in \mathcal{T}_{\varepsilon'}^n(X_1,X_2,Y)$, we have that:

$$\sum_{x_1^n} r_{x_1^n,x_2^n,y^n} = \sum_{x_1^n\in\mathcal{T}_{\varepsilon'}^n(X_1|x_2^n,y^n)} r_{x_1^n,x_2^n,y^n} = \sum_{x_1^n\in\mathcal{T}_{\varepsilon'}^n(X_1|x_2^n,y^n)} p_{x_1^n,x_2^n}\ .$$

Thus:

$$\begin{aligned}
\frac{\sum_{x_1^n} p_{x_1^n,x_2^n}}{k_1\sum_{x_1^n} r_{x_1^n,x_2^n,y^n}} &\geq \frac{1}{k_1} \frac{\sum_{x_1\in\mathcal{T}_{\varepsilon}^n(X_1|x_2^n)} P_{X_1^nX_2^n}(x_1^n,x_2^n)}{\sum_{x_1\in\mathcal{T}_{\varepsilon'}^n(X_1|x_2^n,y^n)} P_{X_1^nX_2^n}(x_1^n,x_2^n)} \geq \frac{1}{k_1} \frac{\sum_{x_1\in\mathcal{T}_{\varepsilon}^n(X_1|x_2^n)} P_{X_1^nX_2^n}(x_1^n,x_2^n)}{\sum_{x_1\in\mathcal{T}_{\varepsilon}^n(X_1|x_2^n,y^n)} P_{X_1^nX_2^n}(x_1^n,x_2^n)}\\
&\geq \frac{1}{k_1} \frac{2^{-n(1+\varepsilon)H(X_1,X_2)}}{2^{-n(1-\varepsilon)H(X_1,X_2)}} \frac{|\mathcal{T}_{\varepsilon}^n(X_1|x_2^n)|}{|\mathcal{T}_{\varepsilon}^n(X_1|x_2^n,y^n)|} \geq 2^{n(-R_1-2\varepsilon H(X_1,X_2))} \frac{|\mathcal{T}_{\varepsilon}^n(X_1|x_2^n)|}{|\mathcal{T}_{\varepsilon}^n(X_1|x_2^n,y^n)|}\ .
\end{aligned} \tag{4.66}$$

But $|\mathcal{T}_{\varepsilon}^n(X_1|x_2^n,y^n)| \leq 2^{n(1+\varepsilon)H(X_1|X_2Y)}$ and for a large enough n we have $\forall x_2^n \in \mathcal{T}_{\varepsilon'}^n(X_2), |\mathcal{T}_{\varepsilon}^n(X_1|x_2^n)| \geq (1-\varepsilon)2^{n(1-\varepsilon)H(X_1|X_2)} \geq 2^{n\left((1-\varepsilon)H(X_1|X_2)-\frac{1}{n}\right)}$, so we get with $C_1 := 2H(X_1,X_2) + H(X_1|X_2Y) + H(X_1|X_2)$ (symmetrically $C_2 := 2H(X_1,X_2) + H(X_2|X_1Y) + H(X_2|X_1)$):

$$\frac{\sum_{x_1^n} p_{x_1^n,x_2^n}}{k_1\sum_{x_1^n} r_{x_1^n,x_2^n,y^n}} \geq 2^{n\left(H(X_1|X_2)-\frac{1}{n}-H(X_1|X_2Y)-R_1-C_1\varepsilon\right)} = 2^{n\left(I(X_1:Y|X_2)-\frac{1}{n}-C_1\varepsilon-R_1\right)} \geq 1\,.$$

For a large enough n, all constraints are satisfied, thus $(p_{x_1^n,x_2^n}, r_{x_1^n,x_2^n,y^n})$ is a valid solution. Then:

$$\begin{aligned}
\mathrm{S}^{\overline{\mathrm{NS}}}(W^{\otimes n}, 2^{nR_1}, 2^{nR_2}) &\geq \frac{1}{2^{n(R_1+R_2)}} \sum_{(x_1^n,x_2^n,y^n) \in \mathcal{T}_{\varepsilon'}^n(X_1,X_2,Y)} W^{\otimes n}(y^n|x_1^n x_2^n) r_{x_1^n,x_2^n,y^n} \\
&= \frac{1}{2^{n(R_1+R_2)}} \sum_{(x_1^n,x_2^n,y^n) \in \mathcal{T}_{\varepsilon'}^n(X_1,X_2,Y)} \frac{P_{X_1^n X_2^n Y^n}(x_1^n,x_2^n,y^n)}{P_{X_1^n X_2^n}(x_1^n,x_2^n)} \frac{2^{n(R_1+R_2)} P_{X_1^n X_2^n}(x_1^n,x_2^n)}{\sum_{(x_1^n,x_2^n)\in\mathcal{T}_{\varepsilon}^n(X_1,X_2)} P_{X_1^n X_2^n}(x_1^n,x_2^n)} \\
&= \frac{\sum_{(x_1^n,x_2^n,y^n) \in \mathcal{T}_{\varepsilon'}^n(X_1,X_2,Y)} P_{X_1^n X_2^n Y^n}(x_1^n,x_2^n,y^n)}{\sum_{(x_1^n,x_2^n)\in\mathcal{T}_{\varepsilon}^n(X_1,X_2)} P_{X_1^n X_2^n}(x_1^n,x_2^n)} \underset{n\to+\infty}{\to} 1 \,,
\end{aligned} \tag{4.67}$$

since typical sets cover asymptotically the whole probability mass. Thus, since $\mathrm{S}^{\overline{\mathrm{NS}}}(W^{\otimes n}, 2^{nR_1}, 2^{nR_2}) \leq 1$, we get that $\mathrm{S}^{\overline{\mathrm{NS}}}(W^{\otimes n}, 2^{nR_1}, 2^{nR_2}) \underset{n\to+\infty}{\to} 1$. Thus for sufficiently large n we can achieve a rate pair arbitrarily close to the outer bound. Finally, since $\mathcal{C}^{\overline{\mathrm{NS}}}(W)$ is closed and convex, a rate pair that is in the closure of the convex hull of the initial region is also in $\mathcal{C}^{\overline{\mathrm{NS}}}(W)$, and thus the proof is completed. □

4.5 Independent Non-Signaling Assisted Capacity Region

The goal of this section is to show that independent non-signaling assistance does not change the capacity region of a MAC W, i.e. that $\mathcal{C}^{\mathrm{NS}_{\mathrm{SR}}}(W) = \mathcal{C}(W)$. In order to prove this result, we will need some properties in the one-sender one-receiver case from [BF18]. Specifically, let us first recall the definition of the maximum success probability $\mathrm{S}(W,k)$ of transmitting k messages using the channel W:

$$\begin{aligned}
\mathrm{S}(W,k) := \quad \underset{e,d}{\text{maximize}} \quad & \frac{1}{k} \sum_{i,x,y} W(y|x) e(x|i) d(i|y) \\
\text{subject to} \quad & \sum_{x\in\mathcal{X}} e(x|i) = 1, \forall i \in [k] \\
& \sum_{j\in[k]} d(j|y) = 1, \forall y \in \mathcal{Y} \\
& e(x|i), d(j|y) \geq 0
\end{aligned} \tag{4.68}$$

Then, the following characterization of $\mathrm{S}(W,k)$ can be derived:

Proposition 4.33 (Proposition 2.1 of [BF18]). $\mathrm{S}(W,k) = \frac{1}{k} \max_{S\subseteq X : |S|\leq k} f_W(S)$ *with* $f_W(S) := \sum_{y\in Y} \max_{x\in S} W(y|x)$.

As in the MAC scenario, one can consider non-signaling assistance shared between the sender and the receiver, which leads to the following maximum success probability $\mathrm{S}^{\mathrm{NS}}(W,k)$:

$$\begin{aligned} \mathrm{S}^{\mathrm{NS}}(W,k) := \quad & \underset{P}{\text{maximize}} && \frac{1}{k}\sum_{i,x,y} W(y|x)P(xi|iy) \\ & \text{subject to} && \sum_x P(xj|iy) = \sum_x P(xj|i'y) \\ &&& \sum_j P(xj|iy) = \sum_j P(xj|iy') \\ &&& \sum_{x,j} P(xj|iy) = 1 \\ &&& P(xj|iy) \geq 0 \end{aligned} \tag{4.69}$$

A symmetrization can also be done to simplify the expression of the linear program defining $\mathrm{S}^{\mathrm{NS}}(W,k)$:

Proposition 4.34 (Appendix A of [BF18])**.**

$$\begin{aligned} \mathrm{S}^{\mathrm{NS}}(W,k) = \quad & \underset{r,p}{\textit{maximize}} && \frac{1}{k}\sum_{x,y} W(y|x)r_{x,y} \\ & \textit{subject to} && \sum_x r_{x,y} = 1 \\ &&& \sum_x p_x = k \\ &&& 0 \leq r_{x,y} \leq p_x \end{aligned} \tag{4.70}$$

Finally, the main tool we will use from [BF18] is the following random coding technique, which describes how to find a classical code with a success probability close to the non-signaling assisted one:

Theorem 4.35 (Theorem 3.1 of [BF18])**.** *Given a solution r,p of the program computing $\mathrm{S}^{\mathrm{NS}}(W,k)$, we have that:*

$$\mathbb{E}_S\left[\frac{f_W(S)}{\ell}\right] \geq \frac{k}{\ell}\left(1-\left(1-\frac{1}{k}\right)^{\ell}\right)\cdot\frac{1}{k}\sum_{x,y}W(y|x)r_{x,y},$$

for the multiset S obtained by choosing ℓ elements of $\mathcal{X}$ independently according to the distribution $\left(\frac{p_x}{k}\right)_{x\in\mathcal{X}}$.

We can now state our result on independent non-signaling assistance, which says that even in one-shot scenarios, the success probability with and without that assistance are close:

Theorem 4.36. *For any ℓ_1,k_1,ℓ_2,k_2:*

$$\min\left(\frac{k_1}{\ell_1}\left(1-\left(1-\frac{1}{k_1}\right)^{\ell_1}\right),\frac{k_2}{\ell_2}\left(1-\left(1-\frac{1}{k_2}\right)^{\ell_2}\right)\right)\mathrm{S}_{\mathit{sum}}^{\mathrm{NS}_{\mathrm{SR}}}(W,k_1,k_2) \leq \mathrm{S}_{\mathit{sum}}(W,\ell_1,\ell_2)\,.$$

In particular, this will imply that the capacity regions are the same:

Corollary 4.37. $\mathcal{C}^{\mathrm{NS}_{\mathrm{SR}}}(W) = \mathcal{C}(W)$.

Proof. We will show that $\mathcal{C}_{\text{sum}}^{\text{NS}_{\text{SR}}}(W) = \mathcal{C}_{\text{sum}}(W)$, which is enough to conclude thanks to Proposition 4.2 and Proposition 4.3. We apply Theorem 4.36 on the MAC $W^{\otimes n}$.

Let us fix $k_1 = 2^{nR_1}, k_2 = 2^{nR_2}$ and $\ell_1 = \frac{2^{nR_1}}{n}, \ell_2 = \frac{2^{nR_2}}{n}$. Since:

$$\frac{k}{\ell}\left(1-\left(1-\frac{1}{k}\right)^{\ell}\right) \geq \frac{k}{\ell}\left(1-e^{-\frac{\ell}{k}}\right) \geq 1-\frac{\ell}{2k} \ ,$$

and $1-\frac{\ell_1}{2k_1} = 1-\frac{\ell_2}{2k_2} = 1-\frac{1}{2n}$, we get:

$$\left(1-\frac{1}{2n}\right)\mathrm{S}_{\text{sum}}^{\text{NS}_{\text{SR}}}(W^{\otimes n}, 2^{nR_1}, 2^{nR_1}) \leq \mathrm{S}_{\text{sum}}\left(W^{\otimes n}, \frac{2^{nR_1}}{n}, \frac{2^{nR_2}}{n}\right) \ .$$

As $1-\frac{1}{2n}$ tends to 1 when n tends to infinity, we get that $\forall \varepsilon > 0, \exists N \in \mathbb{N}, \forall n \geq N$:

$$(1-\varepsilon)\mathrm{S}_{\text{sum}}^{\text{NS}_{\text{SR}}}(W^{\otimes n}, 2^{nR_1}, 2^{nR_1}) \leq \mathrm{S}_{\text{sum}}(W^{\otimes n}, 2^{n(R_1-\frac{\log(n)}{n})}, 2^{n(R_2-\frac{\log(n)}{n})}) \ .$$

Thus, if $\lim\limits_{n\to+\infty} \mathrm{S}_{\text{sum}}^{\text{NS}_{\text{SR}}}(W^{\otimes n}, 2^{nR_1}, 2^{nR_1}) = 1$, we have that for all $R_1' < R_1$ and $R_2' < R_2$:

$$\lim_{n\to+\infty} \mathrm{S}_{\text{sum}}(W^{\otimes n}, 2^{nR_1'}, 2^{nR_1'}) \geq 1-\varepsilon \ .$$

Since this is true for all $\varepsilon > 0$, we get in fact that $\lim\limits_{n\to+\infty} \mathrm{S}_{\text{sum}}(W^{\otimes n}, 2^{nR_1'}, 2^{nR_1'}) = 1$. This implies that $\mathcal{C}_{\text{sum}}^{\text{NS}_{\text{SR}}}(W) \subseteq \mathcal{C}_{\text{sum}}(W)$, and thus that the capacity regions are equal as the other inclusion is always satisfied. $\square$

In order to prove Theorem 4.36, we will need the following lemma:

Lemma 4.38. *If S_1, S_2 are classical codes (i.e. multisets with elements in $\mathcal{X}_1, \mathcal{X}_2$) of size ℓ_1, ℓ_2:*

$$\mathrm{S}_{\mathit{sum}}(W, \ell_1, \ell_2) \geq \frac{1}{2}\left(\frac{f_{W^1_{S_2,\ell_2}}(S_1)}{\ell_1} + \frac{f_{W^2_{S_1,\ell_1}}(S_2)}{\ell_2}\right) \ ,$$

where $W^1_{S_2,\ell_2}$ is the channel defined by $W^1_{S_2,\ell_2}(y|x_1) = \frac{1}{\ell_2}\sum_{i_2=1}^{\ell_2} W(y|x_1 S_2^{i_2})$ and similarly for $W^2_{S_1,\ell_1}$.

Proof. Let us define $e_1(x_1|i_1) := \mathbb{1}_{S_1^{i_1}=x_1}$ and $e_2(x_2|i_2) := \mathbb{1}_{S_2^{i_2}=x_2}$. Then for fixed y, let us take $j_1^y \in \text{argmax}_{i_1}\{\sum_{i_2=1}^{\ell_2} W(y|S_1^{i_1}S_2^{i_2})\}$, $j_2^y \in \text{argmax}_{i_2}\{\sum_{i_1=1}^{\ell_1} W(y|S_1^{i_1}S_2^{i_2})\}$ and then define $d(j_1|y) := \mathbb{1}_{j_1=j_1^y}$, $d(j_2|y) := \mathbb{1}_{j_2=j_2^y}$. We have then:

$$\begin{aligned}
&\mathrm{S}_{\text{sum}}(W,\ell_1,\ell_2) \geq \frac{1}{\ell_1\ell_2}\sum_{i_1,i_2,x_1,x_2,y} W(y|x_1x_2)\mathbb{1}_{S_1^{i_1}=x_1}\mathbb{1}_{S_2^{i_2}=x_2}\frac{\mathbb{1}_{i_1=i_1^y}+\mathbb{1}_{i_2=i_2^y}}{2} \\
&= \frac{1}{\ell_1\ell_2}\sum_{i_1,i_2,y} W(y|S_1^{i_1}S_2^{i_2})\frac{\mathbb{1}_{i_1=i_1^y}+\mathbb{1}_{i_2=i_2^y}}{2} = \frac{1}{\ell_1\ell_2}\sum_y \frac{1}{2}\left(\sum_{i_2} W(y|S_1^{i_1^y}S_2^{i_2}) + \sum_{i_1} W(y|S_1^{i_1}S_2^{i_2^y})\right) \\
&= \frac{1}{\ell_1\ell_2}\sum_y \frac{1}{2}\left(\max_{i_1}\sum_{i_2} W(y|S_1^{i_1}S_2^{i_2}) + \max_{i_2}\sum_{i_1} W(y|S_1^{i_1}S_2^{i_2})\right) \\
&= \frac{1}{2}\left(\frac{\sum_y \max_{i_1}\left[\frac{1}{\ell_2}\sum_{i_2} W(y|S_1^{i_1}S_2^{i_2})\right]}{\ell_1} + \frac{\sum_y \max_{i_2}\left[\frac{1}{\ell_1}\sum_{i_1} W(y|S_1^{i_1}S_2^{i_2})\right]}{\ell_2}\right) \\
&= \frac{1}{2}\left(\frac{\sum_y \max_{i_1} W^1_{S_2,\ell_2}(y|S_1^{i_1})}{\ell_1} + \frac{\sum_y \max_{i_2} W^2_{S_1,\ell_1}(y|S_2^{i_2})}{\ell_2}\right) = \frac{1}{2}\left(\frac{f_{W^1_{S_2,\ell_2}}(S_1)}{\ell_1} + \frac{f_{W^2_{S_1,\ell_1}}(S_2)}{\ell_2}\right) .
\end{aligned} \tag{4.71}$$

□

We have now all the tools to prove Theorem 4.36:

Proof of Theorem 4.36. Let us consider an optimal solution r^1, r^2, p^1, p^2 of the program of Proposition 4.5 computing $\mathrm{S}_{\text{sum}}^{\mathrm{NS}_{\mathrm{SR}}}(W,k_1,k_2)$.

Let us fix some multiset S_2 with elements in $\mathcal{X}_2$ of size ℓ_2. Note that r^1 and p^1 are a feasible solution of the program of Proposition 4.34 computing $\mathrm{S}^{\mathrm{NS}}(W^1_{S_2,\ell_2},k_1)$. As a result, we can apply Theorem 4.35 and get the following statement. For the multiset S_1 obtained by choosing ℓ_1 elements of $\mathcal{X}_1$ independently according to the distribution $\left(\frac{p^1_{x_1}}{k_1}\right)_{x_1\in\mathcal{X}_1}$, we have:

$$\mathbb{E}_{S_1}\left[\frac{f_{W^1_{S_2,\ell_2}}(S_1)}{\ell_1}\right] \geq \frac{k_1}{\ell_1}\left(1-\left(1-\frac{1}{k_1}\right)^{\ell_1}\right)\cdot\frac{1}{k_1}\sum_{x_1,y} W^1_{S_2,\ell_2}(y|x_1)r^1_{x_1,y} .$$

Now, let S_2 be the multiset obtained by choosing ℓ_2 elements of $\mathcal{X}_2$ independently according to the distribution $\left(\frac{p^2_{x_2}}{k_2}\right)_{x_2\in\mathcal{X}_2}$. We have:

$$\begin{aligned}
&\mathbb{E}_{S_2}\left[\frac{1}{k_1}\sum_{x_1,y} W^1_{S_2,\ell_2}(y|x_1)r^1_{x_1,y}\right] = \mathbb{E}_{S_2}\left[\frac{1}{k_1}\sum_{x_1,y}\frac{1}{\ell_2}\sum_{i_2=1}^{\ell_2} W(y|x_1S_2^{i_2})r^1_{x_1,y}\right] \\
&= \frac{1}{\ell_2}\sum_{i_2=1}^{\ell_2}\mathbb{E}_{X_2^{i_2}\sim\frac{p^2_{x_2}}{k_2}}\left[\frac{1}{k_1}\sum_{x_1,y} W(y|x_1X_2^{i_2})r^1_{x_1,y}\right] = \mathbb{E}_{X_2\sim\frac{p^2_{x_2}}{k_2}}\left[\frac{1}{k_1}\sum_{x_1,y} W(y|x_1X_2)r^1_{x_1,y}\right] \\
&= \frac{1}{k_1}\sum_{x_1,x_2,y}\frac{p^2_{x_2}}{k_2}W(y|x_1x_2)r^1_{x_1,y} = \frac{1}{k_1}\sum_{x_1,y} W^1_{p^2,k_2}(y|x_1)r^1_{x_1,y} .
\end{aligned} \tag{4.72}$$

Thus in all, we have:

$$\begin{aligned}
\mathbb{E}_{S_2}\left[\mathbb{E}_{S_1}\left[\frac{f_{W^1_{S_2,\ell_2}}(S_1)}{\ell_1}\right]\right] &\geq \mathbb{E}_{S_2}\left[\frac{k_1}{\ell_1}\left(1-\left(1-\frac{1}{k_1}\right)^{\ell_1}\right)\cdot\frac{1}{k_1}\sum_{x_1,y}W^1_{S_2,\ell_2}(y|x_1)r^1_{x_1,y}\right]\\
&= \frac{k_1}{\ell_1}\left(1-\left(1-\frac{1}{k_1}\right)^{\ell_1}\right)\cdot\mathbb{E}_{S_2}\left[\frac{1}{k_1}\sum_{x_1,y}W^1_{S_2,\ell_2}(y|x_1)r^1_{x_1,y}\right]\\
&\geq \frac{k_1}{\ell_1}\left(1-\left(1-\frac{1}{k_1}\right)^{\ell_1}\right)\cdot\frac{1}{k_1}\sum_{x_1,y}W^1_{p^2,k_2}(y|x_1)r^1_{x_1,y}\ ,
\end{aligned} \tag{4.73}$$

and symmetrically for $\mathbb{E}_{S_1}\left[\mathbb{E}_{S_2}\left[\frac{f_{W^2_{S_1,\ell_1}}(S_2)}{\ell_2}\right]\right]$. Since there exists classical codes S_1^*, S_2^* such that:

$$\frac{1}{2}\left(\frac{f_{W^1_{S_2^*,\ell_2}}(S_1^*)}{\ell_1}+\frac{f_{W^2_{S_1^*,\ell_1}}(S_2^*)}{\ell_2}\right) \geq \mathbb{E}_{S_1,S_2}\left[\frac{1}{2}\left(\frac{f_{W^1_{S_2,\ell_2}}(S_1)}{\ell_1}+\frac{f_{W^2_{S_1,\ell_1}}(S_2)}{\ell_2}\right)\right]\ ,$$

by applying Lemma 4.38, we get:

$$\begin{aligned}
&\mathrm{S_{sum}}(W,\ell_1,\ell_2) \geq \frac{1}{2}\left(\frac{f_{W^1_{S_2^*,\ell_2}}(S_1^*)}{\ell_1}+\frac{f_{W^2_{S_1^*,\ell_1}}(S_2^*)}{\ell_2}\right) \geq \mathbb{E}_{S_1,S_2}\left[\frac{1}{2}\left(\frac{f_{W^1_{S_2,\ell_2}}(S_1)}{\ell_1}+\frac{f_{W^2_{S_1,\ell_1}}(S_2)}{\ell_2}\right)\right]\\
&= \frac{1}{2}\left(\mathbb{E}_{S_2}\left[\mathbb{E}_{S_1}\left[\frac{f_{W^1_{S_2,\ell_2}}(S_1)}{\ell_1}\right]\right]+\mathbb{E}_{S_1}\left[\mathbb{E}_{S_2}\left[\frac{f_{W^2_{S_1,\ell_1}}(S_2)}{\ell_2}\right]\right]\right)\\
&\geq \frac{1}{2}\left(\frac{k_1}{\ell_1}\left(1-\left(1-\frac{1}{k_1}\right)^{\ell_1}\right)\cdot\frac{1}{k_1}\sum_{x_1,y}W^1_{p^2}(y|x_1)r^1_{x_1,y}+\frac{k_2}{\ell_2}\left(1-\left(1-\frac{1}{k_2}\right)^{\ell_2}\right)\cdot\frac{1}{k_2}\sum_{x_2,y}W^2_{p^1}(y|x_2)r^2_{x_2,y}\right)\\
&\geq \min\left(\frac{k_1}{\ell_1}\left(1-\left(1-\frac{1}{k_1}\right)^{\ell_1}\right),\frac{k_2}{\ell_2}\left(1-\left(1-\frac{1}{k_2}\right)^{\ell_2}\right)\right)\mathrm{S}^{\mathrm{NS_{SR}}}_{\mathrm{sum}}(W,k_1,k_2)\ .
\end{aligned} \tag{4.74}$$

□

Remark. In the whole proof of Theorem 4.36, as well as the properties it depends on, we have never used the fact that the output of the channel y was the same for both decoders d_1 and d_2. This implies that the result also holds for interference channels, i.e. two-sender two-receiver channels $W(y_1y_2|x_2x_2)$. Specifically, non-signaling assistance shared between the first sender and the first receiver and independently shared between the second sender and the second receiver does not change the capacity region of interference channels.

4.6 Conclusion

In this chapter, we have studied the impact of non-signaling assistance on the capacity of multiple-access channels. We have developed an efficient linear program computing the success probability of the best non-signaling assisted code for a finite number of copies of a multiple-access channel. In particular, this gives lower bounds on the zero-error non-signaling assisted capacity of multiple-access channels. Applied to the binary adder channel, these results were used to prove that a sum-rate of $\frac{\log_2(72)}{4} \simeq 1.5425$ can be reached with zero error, which beats the maximum classical sum-rate capacity of $\frac{3}{2}$. For noisy channels, we have developed a technique giving lower bounds through the use of concatenated codes. Applied to the noisy binary adder channel, this technique was used to show that non-signaling assistance still improves the sum-rate capacity. We have also found an outer bound on the non-signaling assisted capacity region through a relaxed notion of non-signaling assistance, whose capacity region was characterized by a single-letter formula. Finally, we have shown that independent non-signaling assistance does not change the capacity region.

Our results suggest that quantum entanglement may also increase the capacity of such channels. However, even for the binary adder channel, this question remains open. One could also ask if such efficient methods to compute the best non-signaling assisted codes can be extended to Gaussian multiple-access channels. Finally, establishing a single-letter formula for the non-signaling assisted capacity of multiple-access channels is the main open question left here. It remains open even for the binary adder channel. Proving that non-signaling assistance and relaxed non-signaling assistance coincide asymptotically would directly answer this question and show that the capacity region is described in Theorem 4.22.

Chapter 5

Broadcast Channel Coding with Non-Signaling Correlations

Broadcast channels, introduced by Cover in [Cov72], describe the simple network communication setting where one sender aims to transmit individual messages to two receivers. Contrary to one-way channels [Sha48] or multiple-access channels [Lia73, Ahl73], the capacity region of broadcast channels is known only for particular classes such as the degraded [Ber73, Gal74, AK75], deterministic [Mar77, Pin78] and semi-deterministic [GP80]. Only inner bounds [Cov75, vdM75, Mar79] and outer bounds [Sat78, Mar79, NG07, GN20] on the capacity region are known in the general setting.

On the one hand, from the point of view of quantum information, it is natural to ask whether additional resources, such as quantum entanglement or more generally non-signaling correlations between the parties, changes the capacity region. A non-signaling correlation is a multipartite input-output box shared between parties that, as the name suggests, cannot by itself be used to send information between parties. However, non-signaling correlations such as the ones generated by measurements of entangled quantum particles, can provide an advantage for various information processing tasks and nonlocal games. The study of such correlations has given rise to the quantum information area known as nonlocality [BCP+14]. For example, in the context of channel coding, there exists classical point-to-point channels for which quantum entanglement between the sender and the receiver can increase the optimal success probability for sending one bit of information with a single use of the channel [PLM+11, BF18]. However, a well-known result [BSST99] states that for classical point-to-point channels, entanglement and even more generally non-signaling correlations do not change the capacity of the channel; see also [Mat12, BF18].

In the network setting, behavior is different. Quek and Shor showed in [QS17] the existence of two-sender two-receiver interference channels with gaps between their classical, quantum-entanglement assisted and non-signaling assisted capacity regions. Following this result, Leditzky et al. [LALS20, SLSS22] showed that quantum entanglement shared between the two senders of a multiple access channel can strictly enlarge the capacity region. More specifically, a general investigation of non-signaling resources on multiple-access channel coding was done in [FF22], where it was notably proved that non-signaling advantage occurs even for a simple textbook multiple-access channel: the binary adder channel. A single-letter formula characterization of the quantum-entanglement assisted capacity region of multiple-access channels was later found in [PDB23]. The influence of nonlocal resources on broadcast channel coding has been comparably less studied, the main known result being that quantum entanglement shared between the decoders does not change the capacity region [PDB21].

On the other hand, from an algorithmic point of view, a crucial question in information theory is the complexity of the channel coding problem, which entails maximizing the success probability that can be achieved by sending a fixed number of messages over a channel. However, solving exactly this problem is out of reach, as it is NP-hard to find optimal codes. Therefore, the natural question that arises is the approximability of such a task. For point-to-point channels, Barman and Fawzi found in [BF18] an $(1-e^{-1})$-approximation algorithm running in polynomial time. They showed also that it is NP-hard to approximate the channel coding problem for any strictly better ratio. For ℓ-list-decoding, where the decoder is allowed to output a list of ℓ guesses, a polynomial-time approximation algorithm achieving a $1-\frac{\ell^\ell e^{-\ell}}{\ell!}$ ratio was found in [BFGG20], and it was shown to be NP-hard to do better in [DMMS20]. For multiple-access channel coding, the complexity of the problem can be linked to the bipartite densest κ-subgraph problem [FKP01], which is conjectured to be NP-hard to approximate within any constant ratio [AAM$^+$11]. However, so far, the approximability of channel coding has not been addressed for broadcast channels.

In the point-to-point scenario studied in [BF18], the existence of a constant-ratio approximation algorithm is linked to the equality of the capacity regions with and without non-signaling assistance. This is due to the fact the channel coding problem with non-signaling assistance becomes a linear program, thus computable in polynomial time. It is even equal to the natural linear relaxation of the channel coding problem, which is very common strategy towards approximating an integer linear program. Therefore, showing that this approximation strategy guarantees a constant ratio is the main ingredient towards showing the equality of the capacity regions with and without non-signaling assistance. This link has not been studied yet for broadcast channels. This raises the following questions: Does the capacity region of the broadcast channel change when non-signaling resources between parties are allowed? What is the best approximability ratio of the broadcast channel coding problem? How those two questions are related?

Our Results In this chapter, we first extend the result by [PDB21] in a natural way. We prove that non-signaling resource shared between the decoders does not change the capacity region; see Theorem 5.4. More significantly, we study the influence of sharing a non-signaling resource between the three parties. Contrary to the previous case, this turns the broadcast channel coding problem, which is NP-hard, into a linear program, thus solvable in polynomial time.

We describe a $(1-e^{-1})^2$-approximation algorithm running in polynomial time for the broadcast channel coding problem limited to the class of deterministic channels. This is achieved through a graph interpretation of the problem, where one aims at partitioning a bipartite graph into k_1 and k_2 parts, such that the resulting quotient graph is the densest possible; see Proposition 5.7 and Theorem 5.8. Using the ideas coming from this algorithm, we show that for the class of deterministic channels, non-signaling resource shared between the three parties does not change the capacity region; see Theorem 5.10 and Corollary 5.11.

On the other direction, we consider the subproblem of broadcast channel coding where the number of messages one decoder is responsible of is maximum. This subproblem can be interpreted as a social welfare maximization problem. In the theory of fair division [BT96, Mou03], social welfare maximization entails partitioning a set of goods among agents in order to maximize the sum of their utilities. The social welfare problem has been extensively studied through black box approach [BN05], which led to a precise analysis of achievable approximation ratio as well as hardness results [DS06, MSV08], depending on the class of utility functions considered and the type of black box access to them. We refine the hardness result for the class of fractionally sub-additive utility functions to the subclass coming from the broadcast channel coding subproblem interpretation. Specifically, we show that in the value query model, we cannot achieve a better approximation ratio than $\Omega\left(\frac{1}{\sqrt{m}}\right)$ in polynomial time, with m the size of one of the outputs of the channel: see Theorem 5.16. Following the previous discussion on the links between constant ratio approximation algorithms and non-signaling capacity regions, this hardness result is a first step towards showing that sharing a non-signaling resource between the three parties of a broadcast channel can enlarge its capacity region.

Organization In Section 5.1, we define precisely the different versions of the broadcast channel coding problem depending on the choice of objective value, and show that they all lead to the same capacity region. In Section 5.2, we define the different non-signaling assisted versions of the broadcast channel coding problem. In particular, when shared only between the decoders, we show that the capacity region does not change. In Section 5.3, we address both algorithmic aspects and capacity considerations of deterministic broadcast channels. Specifically, we describe a $(1-e^{-1})^2$-approximation algorithm running in polynomial time for that class, and we show that the capacity region for that class is the same with or without non-signaling assistance. Finally, in Section 5.4, we show that in the value query model, we cannot achieve a better approximation ratio than $\Omega\left(\frac{1}{\sqrt{m}}\right)$ in polynomial time for the general broadcast channel coding problem, with m the size of one of the outputs of the channel.

5.1 Broadcast Channel Coding

5.1.1 Broadcast Channels

Formally, a broadcast channel is given by a conditional probability distribution on input $\mathcal{X}$ and two outputs $\mathcal{Y}_1$ and $\mathcal{Y}_2$, so $W := (W(y_1, y_2|x))_{y_1 \in \mathcal{Y}_1, y_2 \in \mathcal{Y}_2, x \in \mathcal{X}}$, with $W(y_1y_2|x) \geq 0$ and such that $\sum_{y_1 \in \mathcal{Y}_1, y_2 \in \mathcal{Y}_2} W(y_1y_2|x) = 1$. We define its marginals W_1 and W_2 respectively by $W_1(y_1|x) := \sum_{y_2 \in \mathcal{Y}_2} W(y_1y_2|x)$ and $W_2(y_2|x) := \sum_{y_1 \in \mathcal{Y}_1} W(y_1y_2|x)$. We will denote such a broadcast channel by $W : \mathcal{X} \to \mathcal{Y}_1 \times \mathcal{Y}_2$. The tensor product of two broadcast channels $W : \mathcal{X} \to \mathcal{Y}_1 \times \mathcal{Y}_2$ and $W' : \mathcal{X}' \to \mathcal{Y}_1' \times \mathcal{Y}_2'$ is denoted by $W \otimes W' : \mathcal{X} \times \mathcal{X}' \to (\mathcal{Y}_1 \times \mathcal{Y}_1') \times (\mathcal{Y}_2 \times \mathcal{Y}_2')$ and defined by $(W \otimes W')(y_1y_1'y_2y_2'|xx') := W(y_1y_2|x) \cdot W'(y_1'y_2'|x')$. We define $W^{\otimes n}(y_1^n y_2^n|x^n) := \prod_{i=1}^n W(y_{1,i}y_{2,i}|x_i)$, for $y_1^n := y_{1,1} \dots y_{1,n} \in \mathcal{Y}_1^n$ and $y_2^n := y_{2,1} \dots y_{2,n} \in \mathcal{Y}_2^n$ and $x^n := x_1 \dots x_n \in \mathcal{X}^n$. We will use the notation $[k] := \{1, \dots, k\}$.

The coding problem for a broadcast channel $W : \mathcal{X} \to \mathcal{Y}_1 \times \mathcal{Y}_2$ is the following: one wants to encode a couple of messages belonging to $[k_1] \times [k_2]$ into $\mathcal{X}$, which will be given as input to the channel W. This results in two random outputs in $\mathcal{Y}_1$ and $\mathcal{Y}_2$, which one needs to decode back into the corresponding messages in $[k_1]$ and $[k_2]$. We will call

$e : [k_1] \times [k_2] \to \mathcal{X}$ the encoder, $d_1 : \mathcal{Y}_1 \to [k_1]$ the first decoder and $d_2 : \mathcal{Y}_2 \to [k_2]$ the second decoder. This is depicted in Figure ??.

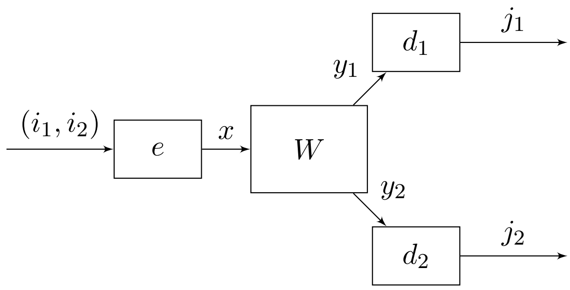


Figure 5.1 – Coding for a broadcast channel W.

We will call $\mathrm{p}_1(W, e, d_1)$ (resp. $\mathrm{p}_2(W, e, d_2)$) the probability of successfully decoding the first (resp. second) message, i.e. that $j_1 = i_1$ (resp. $j_2 = i_2$), given that the encoder is e and the decoder is d_1 (resp. d_2). We will also consider $\mathrm{p}(W, e, d_1, d_2)$, the probability of successfully decoding both messages, i.e. that $j_1 = i_1$ and $j_2 = i_2$, given that the encoder is e and the decoders are d_1, d_2.

We aim to find the best encoder and decoders according to some figure of merit. However, to do so, we need a one dimensional real value objective to optimize. This leads to two different quantities.

5.1.2 The Sum Success Probability $\mathrm{S}_{\mathrm{sum}}(W, k_1, k_2)$

We will focus first on maximizing $\frac{\mathrm{p}_1(W,e,d_1)+\mathrm{p}_2(W,e,d_2)}{2}$ over all encoders e and decoders d_1, d_2. We will call $\mathrm{S}_{\mathrm{sum}}(W, k_1, k_2)$ the resulting maximum sum probability of successfully encoding and decoding the messages through W, given that the input message couple is taken uniformly in $[k_1] \times [k_2]$. $\mathrm{S}_{\mathrm{sum}}(W, k_1, k_2)$ is the solution of the following optimization program:

$$
\begin{aligned}
\mathrm{S}_{\mathrm{sum}}(W, k_1, k_2) := \quad \underset{e, d_1, d_2}{\text{maximize}} \quad & \frac{1}{k_1 k_2} \sum_{i_1, i_2, x, y_1, y_2} W(y_1 y_2 | x) e(x | i_1 i_2) \frac{d_1(i_1 | y_1) + d_2(i_2 | y_2)}{2} \\
\text{subject to} \quad & \sum_{x \in \mathcal{X}} e(x | i_1 i_2) = 1, \forall i_1 \in [k_1], i_2 \in [k_2] \\
& \sum_{i_1 \in [k_1]} d_1(y_1 | i_1) = 1, \forall y_1 \in \mathcal{Y}_1 \\
& \sum_{i_2 \in [k_2]} d_2(y_2 | i_2) = 1, \forall y_2 \in \mathcal{Y}_2 \\
& e(x | i_1 i_2), d_1(y_1 | i_1), d_2(y_2 | i_2) \geq 0
\end{aligned} \tag{5.1}
$$

Proof. One should note that we allow in fact non-deterministic encoders and decoders for generality reasons, although the value of the program is not changed as it is convex. Besides that remark, let us name $I_1, I_2, J_1, J_2, X, Y_1, Y_2$ the random variables corresponding to

$i_1, i_2, j_1, j_2, x, y_1, y_2$ in the coding and decoding process. Then, given e, d_1, d_2 and W, the objective value of the previous program comes from:

$$\begin{aligned}
\mathrm{p}_1(W, e, d_1) = \mathbb{P}\left(J_1 = I_1\right) &= \frac{1}{k_1 k_2} \sum_{i_1, i_2} \mathbb{P}\left(J_1 = i_1 | I_1 = i_1, I_2 = i_2\right) \\
&= \frac{1}{k_1 k_2} \sum_{i_1, i_2, x} e(x|i_1 i_2) \mathbb{P}\left(J_1 = i_1 | I_1 = i_1, I_2 = i_2, X = x\right) \\
&= \frac{1}{k_1 k_2} \sum_{i_1, i_2, x, y_1, y_2} W(y_1 y_2 | x) e(x|i_1 i_2) \mathbb{P}\left(J_1 = i_1 | I_1 = i_1, I_2 = i_2, X = x, Y_1 = y_1, Y_2 = y_2\right) \\
&= \frac{1}{k_1 k_2} \sum_{i_1, i_2, x, y_1, y_2} W(y_1 y_2 | x) e(x|i_1 i_2) d_1(i_1 | y_1) \ ,
\end{aligned} \tag{5.2}$$

and symmetrically for $\mathrm{p}_2(W, e, d_2)$, which leads to the announced objective value. $\square$

One can rewrite this optimization program in a more convenient way, proving that $\mathrm{S}_{\mathrm{sum}}(W, k_1, k_2)$ depends only on the marginals of W:

Proposition 5.1.

$$\begin{aligned}
\mathrm{S}_{sum}(W, k_1, k_2) = \quad \underset{e, d_1, d_2}{\mathit{maximize}} \quad & \frac{1}{2k_1 k_2} \sum_{i_1, x, y_1} W_1(y_1|x) d_1(i_1|y_1) \sum_{i_2} e(x|i_1 i_2) \\
+ \quad & \frac{1}{2k_1 k_2} \sum_{i_2, x, y_2} W_2(y_2|x) d_2(i_2|y_2) \sum_{i_1} e(x|i_1 i_2) \\
\mathit{subject\ to} \quad & \sum_{x \in \mathcal{X}} e(x|i_1 i_2) = 1, \forall i_1 \in [k_1], i_2 \in [k_2] \\
& \sum_{i \in [k_1]} d_1(y_1|i_1) = 1, \forall y_1 \in \mathcal{Y}_1 \\
& \sum_{i \in [k_2]} d_2(y_2|i_2) = 1, \forall y_2 \in \mathcal{Y}_2 \\
& e(x|i_1 i_2), d_1(y_1|i_1), d_2(y_2|i_2) \geq 0
\end{aligned} \tag{5.3}$$

Proof. It follows from the definitions $W_1(y_1|x) := \sum_{y_2} W(y_1 y_2|x)$ and $W_2(y_2|x) := \sum_{y_1} W(y_1 y_2|x)$. $\square$

Since broadcast channels are more general than one-way channels (by defining $W_1(y_1|x) := \hat{W}(y_1|x)$ for $\hat{W}$ a one-way channel and taking $W_2(y_2|x) = \frac{1}{|\mathcal{Y}_2|}$ a completely trivial channel), computing a single value $\mathrm{S}_{\mathrm{sum}}(W, k_1, k_2)$ is NP-hard, and it is even NP-hard to approximate $\mathrm{S}_{\mathrm{sum}}(W, k_1, k_2)$ within a better factor than $(1 - e^{-1})$, as a consequence of the hardness result on $\mathrm{S}(W, k)$ proved in [BF18].

5.1.3 The Joint Success Probability $\mathrm{S}(W, k_1, k_2)$

We will now focus on maximizing $\mathrm{p}(W, e, d_1, d_2)$ over all encoders e and decoders d_1, d_2. We will call $\mathrm{S}(W, k_1, k_2)$ the resulting maximum probability of successfully encoding and decoding the messages through W, given that the input message couple is taken uniformly in $[k_1] \times [k_2]$. $\mathrm{S}(W, k_1, k_2)$ is the solution of the following optimization program:

$$\begin{aligned} \mathrm{S}(W,k_1,k_2) := \quad \underset{e,d_1,d_2}{\text{maximize}} \quad & \frac{1}{k_1 k_2} \sum_{i_1,i_2,x,y_1,y_2} W(y_1 y_2|x) e(x|i_1 i_2) d_1(i_1|y_1) d_2(i_2|y_2) \\ \text{subject to} \quad & \sum_{x \in \mathcal{X}} e(x|i_1 i_2) = 1, \forall i_1 \in [k_1], i_2 \in [k_2] \\ & \sum_{i_1 \in [k_1]} d_1(y_1|i_1) = 1, \forall y_1 \in \mathcal{Y}_1 \\ & \sum_{i_2 \in [k_2]} d_2(y_2|i_2) = 1, \forall y_2 \in \mathcal{Y}_2 \\ & e(x|i_1 i_2), d_1(y_1|i_1), d_2(y_2|i_2) \geq 0 \end{aligned} \tag{5.4}$$

The proof is the same as in the sum probability scenario. The objective values of those two optimization programs are not the same, but $\mathrm{S}(W, k_1, k_2)$ and $\mathrm{S}_{\text{sum}}(W, k_1, k_2)$ still characterize the same capacity region [Wil90]. Let us recall first the definition of their capacity regions:

Definition 5.1 (Capacity Region $\mathcal{C}(W)$ (resp. $\mathcal{C}_{\text{sum}}(W)$) of a broadcast channel W)**.** A rate pair (R_1, R_2) is achievable (resp. sum-achievable) if:

$$\lim_{n \to +\infty} \mathrm{S}(W^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil) = 1 \ .$$

$$(\text{resp. } \lim_{n \to +\infty} \mathrm{S}_{\text{sum}}(W^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil) = 1) \ .$$

We define the joint (resp. sum) (classical) capacity region $\mathcal{C}(W)$ (resp. $\mathcal{C}_{\text{sum}}(W)$) as the closure of the set of all achievable (resp. sum-achievable) rate pairs.

Proposition 5.2. *For any broadcast channel W, $\mathcal{C}(W) = \mathcal{C}_{sum}(W)$.*

Proof. Let us focus on error probabilities rather than success ones. Call them respectively $\mathrm{E}(W, k_1, k_2) := 1 - \mathrm{S}(W, k_1, k_2)$ and $\mathrm{E}_{\text{sum}}(W, k_1, k_2) := 1 - \mathrm{S}_{\text{sum}}(W, k_1, k_2)$. Let us fix a solution e, d_1, d_2 of the optimization program computing $\mathrm{S}(W, k_1, k_2)$. Let us remark first that:

$$\sum_{i_1,i_2,x,y_1,y_2} W(y_1 y_2|x) e(x|i_1 i_2) = k_1 k_2 \ ,$$

thus, the value of the maximum error for those encoder and decoders is:

$$\begin{aligned}
&\mathrm{E}(W,k_1,k_2,e,d_1,d_2) := 1 - \frac{1}{k_1k_2}\left(\sum_{i_1,i_2,x,y_1,y_2} W(y_1y_2|x)e(x|i_1i_2)d_1(i_1|y_1)d_2(i_2|y_2)\right)\\
&= \frac{1}{k_1k_2}\left(\sum_{i_1,i_2,x,y_1,y_2} W(y_1y_2|x)e(x|i_1i_2) - \sum_{i_1,i_2,x,y_1,y_2} W(y_1y_2|x)e(x|i_1i_2)d_1(i_1|y_1)d_2(i_2|y_2)\right)\\
&= \frac{1}{k_1k_2}\left(\sum_{i_1,i_2,x,y_1,y_2} W(y_1y_2|x)e(x|i_1i_2)\left[1 - d_1(i_1|y_1)d_2(i_2|y_2)\right]\right) .
\end{aligned} \tag{5.5}$$

Similarly, the value of the sum error for those encoder and decoders is:

$$\begin{aligned}
\mathrm{E}_{\text{sum}}(W,k_1,k_2,e,d_1,d_2) &:= 1 - \frac{1}{k_1k_2}\left(\sum_{i_1,i_2,x,y_1,y_2} W(y_1y_2|x)e(x|i_1i_2)\frac{d_1(i_1|y_1)+d_2(i_2|y_2)}{2}\right)\\
&= \frac{1}{k_1k_2}\left(\sum_{i_1,i_2,x,y_1,y_2} W(y_1y_2)e(x|i_1i_2)\left[1 - \frac{d_1(i_1|y_1)+d_2(i_2|y_2)}{2}\right]\right) .
\end{aligned} \tag{5.6}$$

However, for $x, y \in [0,1]$, we have that:

$$1 - xy \geq \max\left(1-x, 1-y\right) \geq 1 - \frac{x+y}{2} ,$$

and:

$$1 - xy \leq (1-x) + (1-y) = 2\left(1 - \frac{x+y}{2}\right) .$$

This means that, for all e, d_1, d_2, we have:

$$\mathrm{E}_{\text{sum}}(W,k_1,k_2,e,d_1,d_2) \leq \mathrm{E}(W,k_1,k_2,e,d_1,d_2) \leq 2\mathrm{E}_{\text{sum}}(W,k_1,k_2,e,d_1,d_2) ,$$

so, maximizing over all e, d_1, d_2, we get:

$$\mathrm{E}_{\text{sum}}(W,k_1,k_2) \leq \mathrm{E}(W,k_1,k_2) \leq 2\mathrm{E}_{\text{sum}}(W,k_1,k_2) .$$

Thus, up to a multiplicative factor 2, the error is the same. In particular, when one of those errors tends to zero, the other one tends to zero as well. This implies that the capacity regions are the same. □

5.2 Non-Signaling Assistance

In this section, we will consider the broadcast channel coding problem with additional resources, in order to determine how these resources affect its success probabilities as well as the capacity regions that can be defined from them.

5.2.1 Non-Signaling Assistance between the Decoders

Here, we consider the case where the receivers are given non-signaling assistance. This resource, which is a theoretical but easier to manipulate generalization of quantum entanglement, can be characterized as follows. A non-signaling box $d(j_1j_2|y_1y_2)$ is any joint conditional probability distribution such that the marginal from one party is independent from the other party's input, i.e. we have:

$$\begin{aligned} \forall j_1, y_1, y_2, y_2', \quad \sum_{j_2} d(j_1j_2|y_1y_2) &= \sum_{j_1} d(j_1j_2|y_1y_2') \ , \\ \forall j_2, y_1, y_2, y_1', \quad \sum_{j_1} d(j_1j_2|y_1y_2) &= \sum_{j_1} d(j_1j_2|y_1'y_2) \ . \end{aligned} \tag{5.7}$$

Thus, when receivers are given non-signaling assistance, it means that the product $d_1(j_1|y_1)d_2(j_2|y_2)$ is replaced by the non-signaling box $d(j_1j_2|y_1y_2)$. We define the joint and sum success probabilities $\mathrm{S}^{\mathrm{NS_{dec}}}(W, k_1, k_2)$ and $\mathrm{S}^{\mathrm{NS_{dec}}}_{\text{sum}}(W, k_1, k_2)$ according to this:

$$\begin{aligned} \mathrm{S}^{\mathrm{NS_{dec}}}(W, k_1, k_2) :=\ & \underset{e,d_1,d_2}{\text{maximize}} && \frac{1}{k_1k_2} \sum_{i_1,i_2,x,y_1,y_2} W(y_1y_2|x)e(x|i_1i_2)d(i_1i_2|y_1y_2) \\ \Big(\text{resp. } \mathrm{S}^{\mathrm{NS_{dec}}}_{\text{sum}}(W, k_1, k_2) :=\ & \underset{e,d_1,d_2}{\text{maximize}} && \frac{1}{2k_1k_2} \sum_{i_1,i_2,x,y_1,y_2} W(y_1y_2|x)e(x|i_1i_2) \sum_{j_2} d(i_1j_2|y_1y_2) \\ & + && \frac{1}{2k_1k_2} \sum_{i_1,i_2,x,y_1,y_2} W(y_1y_2|x)e(x|i_1i_2) \sum_{j_1} d(j_1i_2|y_1y_2)\Big) \\ & \text{subject to} && \sum_x e(x|i_1i_2) = 1 \\ & && \sum_{j_2} d(j_1j_2|y_1y_2) = \sum_{j_1} d(j_1j_2|y_1y_2') \\ & && \sum_{j_1} d(j_1j_2|y_1y_2) = \sum_{j_1} d(j_1j_2|y_1'y_2) \\ & && \sum_{j_1,j_2} d(j_1j_2|y_1y_2) = 1 \\ & && e(x|i_1i_2), d(j_1j_2|y_1y_2) \geq 0 \end{aligned} \tag{5.8}$$

The corresponding capacity regions $\mathcal{C}^{\mathrm{NS_{dec}}}(W)$ and $\mathcal{C}^{\mathrm{NS_{dec}}}_{\text{sum}}(W)$ are defined as before:

Definition 5.2 (Capacity Region $\mathcal{C}^{\mathrm{NS_{dec}}}(W)$ (resp. $\mathcal{C}^{\mathrm{NS_{dec}}}_{\text{sum}}(W)$) of a broadcast channel W)**.** A rate pair (R_1, R_2) is achievable (resp. sum-achievable) with non-signaling assistance between the decoders if:

$$\lim_{n\to+\infty} \mathrm{S}^{\mathrm{NS_{dec}}}(W^{\otimes n}, \lceil 2^{R_1n} \rceil, \lceil 2^{R_2n} \rceil) = 1 \ .$$

$$(\text{resp. } \lim_{n\to+\infty} \mathrm{S}^{\mathrm{NS_{dec}}}_{\text{sum}}(W^{\otimes n}, \lceil 2^{R_1n} \rceil, \lceil 2^{R_2n} \rceil) = 1) \ .$$

We define the joint (resp. sum) capacity region with non-signaling assistance between the decoders $\mathcal{C}^{\mathrm{NS_{dec}}}(W)$ (resp. $\mathcal{C}^{\mathrm{NS_{dec}}}_{\text{sum}}(W)$) as the closure of the set of all rate pairs achievable (resp. sum-achievable) with non-signaling assistance between the decoders.

The objective of this section is to show that sharing non-signaling assistance between the decoders does not change the capacity regions of a broadcast channel. In order to do so, we will also show that sum and joint capacity regions with non-signaling assistance between the decoders are the same. This is an extension of the result by [PDB21] where it has been proved to be the case for the weaker resource that is quantum entanglement.

Proposition 5.3. *For any broadcast channel W, $\mathcal{C}_{sum}^{\mathrm{NS}_{dec}}(W) = \mathcal{C}^{\mathrm{NS}_{dec}}(W)$*

Proof. Given an encoder e and a non-signaling decoding box d, the maximum success probability of encoding and decoding correctly with those is given by:

$$\mathrm{S}^{\mathrm{NS}_{\mathrm{dec}}}(W, k_1, k_2, e, d) := \frac{1}{k_1 k_2} \sum_{i_1, i_2, x, y_1, y_2} W(y_1 y_2|x) e(x|i_1 i_2) d(i_1 i_2|y_1 y_2) \ .$$

This should be compared to the sum success probability of encoding and decoding correctly with those:

$$\mathrm{S}_{\mathrm{sum}}^{\mathrm{NS}_{\mathrm{dec}}}(W, k_1, k_2, e, d) := \frac{1}{k_1 k_2} \sum_{i_1, i_2, x, y_1, y_2} W(y_1 y_2|x) e(x|i_1 i_2) \left[\frac{\sum_{j_2} d(i_1 j_2|y_1 y_2) + \sum_{j_1} d(j_1 i_2|y_1 y_2)}{2} \right] .$$

Similarly to what was done in Proposition 5.2, we focus on the error probabilities rather than success probabilities. This leads again to:

$$\mathrm{E}^{\mathrm{NS}_{\mathrm{dec}}}(W, k_1, k_2, e, d) = \frac{1}{k_1 k_2} \sum_{i_1, i_2, x, y_1, y_2} W(y_1 y_2|x) e(x|i_1 i_2) \left[1 - d(i_1 i_2|y_1 y_2) \right] \ ,$$

and:

$$\mathrm{E}_{\mathrm{sum}}^{\mathrm{NS}_{\mathrm{dec}}}(W, k_1, k_2, e, d) = \frac{1}{k_1 k_2} \sum_{i_1, i_2, x, y_1, y_2} W(y_1 y_2|x) e(x|i_1 i_2) \left[\frac{1 - \sum_{j_2} d(i_1 j_2|y_1 y_2)}{2} + \frac{1 - \sum_{j_1} d(j_1 i_2|y_1 y_2)}{2} \right] .$$

But we have that:

$$1 - d(i_1 i_2|y_1 y_2) \geq \max \left(1 - \sum_{j_2} d(i_1 j_2|y_1 y_2), 1 - \sum_{j_1} d(j_1 i_2|y_1 y_2) \right) \geq \frac{1 - \sum_{j_2} d(i_1 j_2|y_1 y_2)}{2} + \frac{1 - \sum_{j_1} d(j_1 i_2|y}{2}$$

since $d(j_1 j_2|y_1 y_2) \in [0, 1]$, and we have that:

$$1 - \sum_{j_2} d(i_1 j_2|y_1 y_2) + 1 - \sum_{j_1} d(j_1 i_2|y_1 y_2) = 1 - d(i_1 i_2|y_1 y_2) + 1 - \sum_{(j_1, j_2) \in S} d(j_1 j_2|y_1 y_2) \geq 1 - d(i_1 i_2|y_1 y_2) \ ,$$

with $S := \{(i_1, j_2) : j_2 \in [k_2] - \{i_2\}\} \sqcup \{(j_1, i_2) : j_1 \in [k_1] - \{i_1\}\}$.

Thus, this implies that:

$$\mathrm{E}_{\mathrm{sum}}^{\mathrm{NS}_{\mathrm{dec}}}(W, k_1, k_2, e, d) \leq \mathrm{E}^{\mathrm{NS}_{\mathrm{dec}}}(W, k_1, k_2, e, d) \leq 2\mathrm{E}_{\mathrm{sum}}^{\mathrm{NS}_{\mathrm{dec}}}(W, k_1, k_2, e, d) \ ,$$

and by maximizing over all e and d:

$$\mathrm{E}_{\mathrm{sum}}^{\mathrm{NS}_{\mathrm{dec}}}(W, k_1, k_2) \leq \mathrm{E}^{\mathrm{NS}_{\mathrm{dec}}}(W, k_1, k_2) \leq 2\mathrm{E}_{\mathrm{sum}}^{\mathrm{NS}_{\mathrm{dec}}}(W, k_1, k_2) \ .$$

As before, this implies that the capacity regions are the same. ☐

Theorem 5.4. *For any broadcast channel W, $\mathcal{C}(W) = \mathcal{C}^{\mathrm{NS}_{dec}}(W)$*

Proof. In the sum scenario, since the objective function does not depend on the product $d_1(j_1|y_1)d_2(j_2|y_2)$ but only on the marginals $d_1(j_1|y_1)$ and $d_2(j_2|y_2)$, the non-signaling box won't give additional decoding power. Indeed, for any encoder e and non-signaling decoding box d, we have that:

$$\begin{aligned}\mathrm{S}_{\text{sum}}^{\mathrm{NS}_{\text{dec}}}(W, k_1, k_2, e, d) := &\frac{1}{2k_1k_2} \sum_{i_1,x,y_1} W_1(y_1|x) \left(\sum_{j_2} d(i_1j_2|y_1y_2) \right) \sum_{i_2} e(x|i_1i_2) \\ &+ \frac{1}{2k_1k_2} \sum_{i_2,x,y_2} W_2(y_2|x) \left(\sum_{j_1} d(j_1i_2|y_1y_2) \right) \sum_{i_1} e(x|i_1i_2) \ .\end{aligned} \tag{5.9}$$

Thus, by choosing $d_1(j_1|y_1) := \sum_{j_2} d(j_1j_2|y_1y_2)$ and $d_2(j_2|y_2) := \sum_{j_1} d(j_1j_2|y_1y_2)$, which are well-defined since d is a non-signaling box, one gets that $\mathrm{S}_{\text{sum}}(W, k_1, k_2, e, d_1, d_2) = \mathrm{S}_{\text{sum}}^{\mathrm{NS}_{\text{dec}}}(W, k_1, k_2, e, d)$. By optimizing over all e and d, one gets $\mathrm{S}_{\text{sum}}^{\mathrm{NS}_{\text{dec}}}(W, k_1, k_2) \leq \mathrm{S}_{\text{sum}}(W, k_1, k_2)$. Since the inequality is obvious in the other direction, as $d(j_1j_2|y_1y_2) := d_1(j_1|y_1)d_2(j_2|y_2)$ is always a non-signaling box, we have that $\mathrm{S}_{\text{sum}}(W, k_1, k_2) = \mathrm{S}_{\text{sum}}^{\mathrm{NS}_{\text{dec}}}(W, k_1, k_2)$. This implies in particular that the capacity regions are the same, i.e. $\mathcal{C}_{\text{sum}}(W) = \mathcal{C}_{\text{sum}}^{\mathrm{NS}_{\text{dec}}}(W)$

Finally, since $\mathcal{C}(W) = \mathcal{C}_{\text{sum}}(W)$ by Proposition 5.2 and $\mathcal{C}_{\text{sum}}^{\mathrm{NS}_{\text{dec}}}(W) = \mathcal{C}^{\mathrm{NS}_{\text{dec}}}(W)$ by Proposition 5.3, we get that $\mathcal{C}(W) = \mathcal{C}^{\mathrm{NS}_{\text{dec}}}(W)$. □

5.2.2 Full Non-Signaling Assistance

In this section, we will consider the case where the sender and the receivers are given non-signaling assistance. This means that a three-party non-signaling box $P(xj_1j_2|(i_1i_2)y_1y_2)$ will replace the product $e(x|i_1i_2)d_1(j_1|y_1)d_2(j_2|y_2)$ in the previous objective values. A joint conditional probability $P(xj_1j_2|(i_1i_2)y_1y_2)$ is a non-signaling box if the marginal from any two parties is independent from the removed party's input:

$$\begin{aligned}\forall j_1, j_2, i_1, i_2, y_1, y_2, i_1', i_2', \quad &\sum_x P(xj_1j_2|(i_1i_2)y_1y_2) = \sum_x P(xj_1j_2|(i_1'i_2')y_1y_2) \ , \\ \forall x, j_2, i_1, i_2, y_1, y_2, y_1', \quad &\sum_{j_1} P(xj_1j_2|(i_1i_2)y_1y_2) = \sum_{j_1} P(xj_1j_2|(i_1i_2)y_1'y_2) \ , \\ \forall x, j_1, i_1, i_2, y_1, y_2, y_2', \quad &\sum_{j_2} P(xj_1j_2|(i_1i_2)y_1y_2) = \sum_{j_2} P(xj_1j_2|(i_1i_2)y_1y_2') \ .\end{aligned} \tag{5.10}$$

The scenario is depicted in Figure 5.2.

The cyclicity of Figure 5.2 is at first sight counter-intuitive. Note first that P being a non-signaling box is completely independent from W: in particular, the variables y_1, y_2 do not need to follow any laws in the definition of P being a non-signaling box. Therefore, the remaining ambiguity is the apparent need to encode and decode at the same time. However, since P is a non-signaling box, we won't need to do both at the same time, although the global correlation between the sender and the receivers will be characterized only by $P(xj_1j_2|(i_1i_2)y_1y_2)$. Indeed, $\forall y_1, y_2, P(x|(i_1i_2)) = P(x|(i_1i_2)y_1y_2)$ by the non-signaling property of P. Thus, one can get the output x on input (i_1i_2) without access to y_1, y_2, as

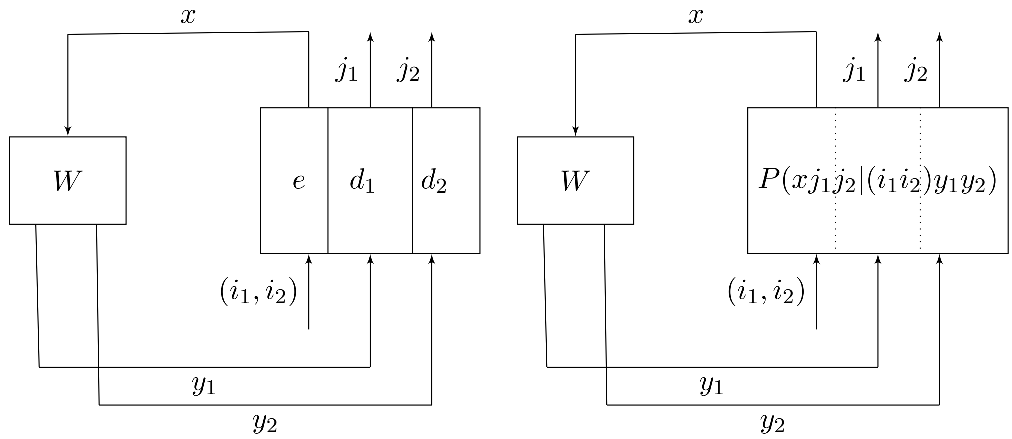


Figure 5.2 – A non-signaling box P replacing e, d_1, d_2 in the coding problem for the broadcast channel W.

that knowledge won't affect the law of x. Then (y_1, y_2) follows the law given by W given that x. Finally, given access to y_1, y_2, the decoding process is described by:

$$P(j_1j_2|(i_1i_2)y_1y_2x) = \frac{P(xj_1j_2|(i_1i_2)y_1y_2)}{P(x|(i_1i_2)y_1y_2)} = \frac{P(xj_1j_2|(i_1i_2)y_1y_2)}{P(x|(i_1i_2))} \ ,$$

so we recover $P(j_1j_2|(i_1i_2)y_1y_2x)) \times P(x|(i_1i_2)) = P(xj_1j_2|(i_1i_2)y_1y_2)$, and therefore, the process is not cyclic. Non-signaling boxes define exactly the conditional probability distributions where it is possible to consider the conditional probabilities of each party independently. This clarifies how one can effectively encode and then decode messages through a non-signaling box.

We will call the maximum sum success probability $\mathrm{S}^{\mathrm{NS}}_{\mathrm{sum}}(W, k_1, k_2)$, which is given by the following linear program, where the constraints translate precisely the fact that P is a non-signaling box:

$$\begin{aligned}
\mathrm{S}^{\mathrm{NS}}_{\mathrm{sum}}(W, k_1, k_2) := \quad \underset{P}{\text{maximize}} \quad & \frac{1}{2k_1k_2} \sum_{i_1,x,y_1} W_1(y_1|x) \sum_{i_2,j_2} P(xi_1j_2|(i_1i_2)y_1y_2) \\
+ \quad & \frac{1}{2k_1k_2} \sum_{i_2,x,y_2} W_2(y_2|x) \sum_{i_1,j_1} P(xj_1i_2|(i_1i_2)y_1y_2) \\
\text{subject to} \quad & \sum_{x} P(xj_1j_2|(i_1i_2)y_1y_2) = \sum_{x} P(xj_1j_2|(i'_1i'_2)y_1y_2) \\
& \sum_{j_1} P(xj_1j_2|(i_1i_2)y_1y_2) = \sum_{j_1} P(xj_1j_2|(i_1i_2)y'_1y_2) \\
& \sum_{j_2} P(xj_1j_2|(i_1i_2)y_1y_2) = \sum_{j_2} P(xj_1j_2|(i_1i_2)y_1y'_2) \\
& \sum_{x,j_1,j_2} P(xj_1j_2|(i_1i_2)y_1y_2) = 1 \\
& P(xj_1j_2|(i_1i_2)y_1y_2) \geq 0
\end{aligned} \tag{5.11}$$

Since it is given as a linear program, the complexity of computing $\mathrm{S}^{\mathrm{NS}}_{\mathrm{sum}}(W, k_1, k_2)$ is polynomial in the number of variables and constraints (see for instance Section 7.1 of [GM07]), which is a polynomial in $|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|, k_1$ and k_2.

Similarly, we define the maximum joint success probability $\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)$ in the following way:

$$\begin{aligned}
\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2) := \quad \underset{P}{\text{maximize}} \quad & \frac{1}{k_1 k_2} \sum_{i_1, i_2, x, y_1, y_2} W(y_1 y_2|x) P(x i_1 i_2|(i_1 i_2) y_1 y_2) \\
\text{subject to} \quad & \sum_{x} P(x j_1 j_2|(i_1 i_2) y_1 y_2) = \sum_{x} P(x j_1 j_2|(i'_1 i'_2) y_1 y_2) \\
& \sum_{j_1} P(x j_1 j_2|(i_1 i_2) y_1 y_2) = \sum_{j_1} P(x j_1 j_2|(i_1 i_2) y'_1 y_2) \\
& \sum_{j_2} P(x j_1 j_2|(i_1 i_2) y_1 y_2) = \sum_{j_2} P(x j_1 j_2|(i_1 i_2) y_1 y'_2) \\
& \sum_{x, j_1, j_2} P(x j_1 j_2|(i_1 i_2) y_1 y_2) = 1 \\
& P(x j_1 j_2|(i_1 i_2) y_1 y_2) \geq 0
\end{aligned} \tag{5.12}$$

We can rewrite both these programs in more convenient and smaller linear programs:

Proposition 5.5.

$$\begin{aligned}
\mathrm{S}^{\mathrm{NS}}_{sum}(W, k_1, k_2) = \quad \underset{p, r, r^1, r^2}{\textit{maximize}} \quad & \frac{1}{2 k_1 k_2} \left(\sum_{x, y_1} W_1(y_1|x) r^1_{x, y_1} + \sum_{x, y_2} W_2(y_2|x) r^2_{x, y_2} \right) \\
\textit{subject to} \quad & \sum_{x} r_{x, y_1, y_2} = 1 \\
& \sum_{x} r^1_{x, y_1} = k_2 \\
& \sum_{x} r^2_{x, y_2} = k_1 \\
& \sum_{x} p_x = k_1 k_2 \\
& 0 \leq r_{x, y_1, y_2} \leq r^1_{x, y_1}, r^2_{x, y_2} \leq p_x \\
& p_x - r^1_{x, y_1} - r^2_{x, y_2} + r_{x, y_1, y_2} \geq 0
\end{aligned} \tag{5.13}$$

$$
\begin{aligned}
\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2) = \quad \underset{p,r,r^1,r^2}{\textit{maximize}} \quad & \frac{1}{k_1 k_2} \sum_{x,y_1,y_2} W(y_1 y_2|x) r_{x,y_1,y_2} \\
\textit{subject to} \quad & \sum_x r_{x,y_1,y_2} = 1 \\
& \sum_x r^1_{x,y_1} = k_2 \\
& \sum_x r^2_{x,y_2} = k_1 \\
& \sum_x p_x = k_1 k_2 \\
& 0 \leq r_{x,y_1,y_2} \leq r^1_{x,y_1}, r^2_{x,y_2} \leq p_x \\
& p_x - r^1_{x,y_1} - r^2_{x,y_2} + r_{x,y_1,y_2} \geq 0
\end{aligned}
\tag{5.14}
$$

Proof. One can check that given a solution of the original program, the following choice of variables is a valid solution of the second program achieving the same objective value:

$$
\begin{aligned}
r_{x,y_1,y_2} &:= \sum_{i_1,i_2} P(x i_1 i_2|(i_1 i_2) y_1 y_2) \ , \\
r^1_{x,y_1} &:= \sum_{j_2,i_1,i_2} P(x i_1 j_2|(i_1 i_2) y_1 y_2) \ , \\
r^2_{x,y_2} &:= \sum_{j_1,i_1,i_2} P(x j_1 i_2|(i_1 i_2) y_1 y_2) \ , \\
p_x &:= \sum_{j_1,j_2,i_1,i_2} P(x j_1 j_2|(i_1 i_2) y_1 y_2) \ .
\end{aligned}
\tag{5.15}
$$

For the other direction, given those variables, a non-signaling probability distribution $P(xj_1j_2|(i_1i_2)y_1y_2)$ is given by, for $j_1 \neq i_1$ and $j_2 \neq i_2$:

$$
\begin{aligned}
P(x i_1 i_2|(i_1 i_2) y_1 y_2) &= \frac{r_{x,y_1,y_2}}{k_1 k_2} \ , \\
P(x j_1 i_2|(i_1 i_2) y_1 y_2) &= \frac{r^2_{x,y_2} - r_{x,y_1,y_2}}{k_1 k_2 (k_1 - 1)} \ , \\
P(x i_1 j_2|(i_1 i_2) y_1 y_2) &= \frac{r^1_{x,y_1} - r_{x,y_1,y_2}}{k_1 k_2 (k_2 - 1)} \ , \\
P(x j_1 j_2|(i_1 i_2) y_1 y_2) &= \frac{p_x - r^1_{x,y_1} - r^2_{x,y_2} + r_{x,y_1,y_2}}{k_1 k_2 (k_1 - 1)(k_2 - 1)} \ .
\end{aligned}
\tag{5.16}
$$

□

As before, one can define the capacity regions of broadcast channels with non-signaling assistance:

Definition 5.3 (Capacity Region $\mathcal{C}^{\mathrm{NS}}(W)$ (resp. $\mathcal{C}^{\mathrm{NS}}_{\mathrm{sum}}(W)$) of a broadcast channel W)**.** A rate pair (R_1, R_2) is achievable with (full) non-signaling assistance if:

$$\lim_{n \to +\infty} \mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{R_1 n} \rceil, \lceil 2^{R_2 n} \rceil) = 1 \ .$$

$$\left(\text{resp. } \lim_{n\to+\infty} \mathrm{S}^{\mathrm{NS}}_{\mathrm{sum}}(W^{\otimes n}, \lceil 2^{R_1 n}\rceil, \lceil 2^{R_2 n}\rceil) = 1\right) .$$

We define the joint (resp. sum) non-signaling assisted capacity region $\mathcal{C}^{\mathrm{NS}}(W)$ (resp. $\mathcal{C}^{\mathrm{NS}}_{\mathrm{sum}}(W)$) as the closure of the set of all rate pairs achievable with (full) non-signaling assistance.

Proposition 5.6. *For any broadcast channel W, $\mathcal{C}^{\mathrm{NS}}(W) = \mathcal{C}^{\mathrm{NS}}_{sum}(W)$.*

Proof. Let us show that:

$$2\mathrm{S}^{\mathrm{NS}}_{\mathrm{sum}}(W, k_1, k_2) - 1 \leq \mathrm{S}^{\mathrm{NS}}(W, k_1, k_2) \leq \mathrm{S}^{\mathrm{NS}}_{\mathrm{sum}}(W, k_1, k_2) .$$

This will imply in particular that:

$$\lim_{n\to+\infty} \mathrm{S}^{\mathrm{NS}}(W^{\otimes n}, \lceil 2^{R_1 n}\rceil, \lceil 2^{R_2 n}\rceil) = 1 \iff \lim_{n\to+\infty} \mathrm{S}^{\mathrm{NS}}_{\mathrm{sum}}(W^{\otimes n}, \lceil 2^{R_1 n}\rceil, \lceil 2^{R_2 n}\rceil) = 1 ,$$

thus define the same capacity region.

Let us consider an optimal solution $p_x, r_{x,y_1,y_2}, r^1_{x,y_1}, r^2_{x,y_2}$ of the program computing $\mathrm{S}^{\mathrm{NS}}_{\mathrm{sum}}(W, k_1, k_2)$. We have:

$$\mathrm{S}^{\mathrm{NS}}_{\mathrm{sum}}(W, k_1, k_2) = \frac{1}{k_1 k_2}\left(\sum_{x,y_1,y_2} W(y_1y_2|x)\frac{r^1_{x,y_1} + r^2_{x,y_2}}{2}\right) .$$

However $r^1_{x,y_1} + r^2_{x,y_2} \leq p_x + r_{x,y_1,y_2}$ so we get that:

$$\begin{aligned}\mathrm{S}^{\mathrm{NS}}_{\mathrm{sum}}(W, k_1, k_2) &\leq \frac{1}{2k_1k_2}\left(\sum_{x,y_1,y_2} W(y_1y_2|x)\left(p_x + r_{x,y_1,y_2}\right)\right) = \frac{1}{2} + \frac{1}{2}\left[\frac{1}{k_1k_2}\left(\sum_{x,y_1,y_2} W(y_1y_2|x) r_{x,y_1,y_2}\right.\right. \\ &\leq \frac{1}{2} + \frac{1}{2}\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2) ,\end{aligned} \tag{5.17}$$

since $p_x, r_{x,y_1,y_2}, r^1_{x,y_1}, r^2_{x,y_2}$ is a valid solution of the program computing $\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)$.

On the other hand, consider now $p_x, r_{x,y_1,y_2}, r^1_{x,y_1}, r^2_{x,y_2}$ an optimal solution of the program computing $\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2)$. We have that $r_{x,y_1,y_2} \leq r^1_{x,y_1}, r^2_{x,y_2}$ so we have that $r_{x,y_1,y_2} \leq \frac{r^1_{x,y_1}+r^2_{x,y_2}}{2}$ and thus:

$$\begin{aligned}\mathrm{S}^{\mathrm{NS}}(W, k_1, k_2) &= \frac{1}{k_1k_2}\left(\sum_{x,y_1,y_2} W(y_1y_2|x) r_{x,y_1,y_2}\right) \leq \frac{1}{k_1k_2}\left(\sum_{x,y_1,y_2} W(y_1y_2|x)\frac{r^1_{x,y_1} + r^2_{x,y_2}}{2}\right) \\ &\leq \mathrm{S}^{\mathrm{NS}}_{\mathrm{sum}}(W, k_1, k_2) ,\end{aligned} \tag{5.18}$$

since $p_x, r_{x,y_1,y_2}, r^1_{x,y_1}, r^2_{x,y_2}$ is a valid solution of the program computing $\mathrm{S}^{\mathrm{NS}}_{\mathrm{sum}}(W, k_1, k_2)$.

This prove the inequalities $2\mathrm{S}^{\mathrm{NS}}_{\mathrm{sum}}(W, k_1, k_2) - 1 \leq \mathrm{S}^{\mathrm{NS}}(W, k_1, k_2) \leq \mathrm{S}^{\mathrm{NS}}_{\mathrm{sum}}(W, k_1, k_2)$, and thus concludes the proof. $\square$

5.3 Approximation of Deterministic Broadcast Channel Coding

In this section, we will address the question of the approximability of $\mathrm{S}(W, k_1, k_2)$, in the restricted scenario of a deterministic broadcast channel W. Specifically, we study the problem of finding a code $e : [k_1] \times [k_2] \to \mathcal{X}$, $d_1 : \mathcal{Y}_1 \to [k_1]$, $d_2 : \mathcal{Y}_2 \to [k_2]$ that maximizes the program computing $\mathrm{S}(W, k_1, k_2)$. Note that the restriction to deterministic codes does not affect the value of the objective of the program which is convex, and that the problem is as hard as finding any code maximizing the program computing $\mathrm{S}(W, k_1, k_2)$, as a deterministic code with a better or equal value can be retrieved easily from any code.

We say that W is deterministic if $\forall x, y_1, y_2, W(y_1y_2|x) \in \{0, 1\}$. We can then define $(W_1(x), W_2(x))$ as the only pair (y_1, y_2) such that $W(y_1y_2|x) = 1$, which exists uniquely as W is a conditional probability distribution. Thus, the deterministic broadcast channel coding problem can be defined in the following way:

Definition 5.4 (DetBCC)**.** Given a deterministic channel W and integers k_1 and k_2, the deterministic broadcast channel coding problem, which we call DetBCC, entails maximizing

$$\mathrm{S}(W, k_1, k_2, e, d_1, d_2) := \frac{1}{k_1k_2} \sum_{i_1, i_2} \mathbb{1}_{d_1(W_1(e(i_1i_2)))=i_1} \mathbb{1}_{d_2(W_2(e(i_1i_2)))=i_2}$$

over all functions $e : [k_1] \times [k_2] \to \mathcal{X}$, $d_1 : \mathcal{Y}_1 \to [k_1]$, $d_2 : \mathcal{Y}_2 \to [k_2]$.

5.3.1 Reformulation as a Bipartite Graph Problem

In this subsection, we will reformulate DetBCC as a bipartite graph problem. But first, let us introduce some notations:

Definition 5.5 (Graph notations)**.** Consider a bipartite graph $G = (V_1 \sqcup V_2, E \subseteq V_1 \times V_2)$:

1. $G^{\mathcal{P}_1,\mathcal{P}_2}$, the quotient of G by partitions $\mathcal{P}_1, \mathcal{P}_2$ of respectively V_1, V_2, is defined by:

$$G^{\mathcal{P}_1,\mathcal{P}_2} := (\mathcal{P}_1 \sqcup \mathcal{P}_2, \{(p_1, p_2) \in \mathcal{P}_1 \times \mathcal{P}_2 : E \cap (p_1 \times p_2) \neq \emptyset\}) \ .$$

2. $e_G(\mathcal{P}_1, \mathcal{P}_2) := |E^{G^{\mathcal{P}_1,\mathcal{P}_2}}|$ is the number of edges of $G^{\mathcal{P}_1,\mathcal{P}_2}$.
3. $N_G^{\mathcal{P}_1,\mathcal{P}_2}(p) := N_{G^{\mathcal{P}_1,\mathcal{P}_2}}(p)$ is the set of neighbors of p in the graph $G^{\mathcal{P}_1,\mathcal{P}_2}$.
4. Similarly, $\deg_G^{\mathcal{P}_1,\mathcal{P}_2}(p) := \deg_{G^{\mathcal{P}_1,\mathcal{P}_2}}(p)$ is the degree, i.e. the number of neighbors, of p in the graph $G^{\mathcal{P}_1,\mathcal{P}_2}$.
5. We will use V_1, V_2 in previous notations when we do not partition on the left and right part respectively (or identify those to trivial partitions in singletons). For instance, $G^{V_1,V_2} = G$.
6. We will use the notations $e(\mathcal{P}_1, \mathcal{P}_2)$, $N_{\mathcal{P}_1,\mathcal{P}_2}(p)$ and $\deg_{\mathcal{P}_1,\mathcal{P}_2}(p)$ when the graph G considered is clear from the context.

Now, let us remark that a deterministic channel W, up to a permutation of elements of $\mathcal{X}$, is characterized by the following bipartite graph:

Definition 5.6 (Bipartite Graph G_W associated to the deterministic channel W)**.**

$$G_W := (\mathcal{Y}_1 \sqcup \mathcal{Y}_2, E = \{(y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2 : \exists x \in \mathcal{X}, y_1 = W_1(x) \text{ and } y_2 = W_2(x)\}) \ .$$

Indeed, permuting the elements of $\mathcal{X}$ does not change G_W nor $\mathrm{S}(W, k_1, k_2)$. As a consequence, up to a multiplicative factor $k_1 k_2$, we will show that DetBCC is equivalent to the following bipartite graph problem:

Definition 5.7 (DensestQuotientGraph)**.** Given a bipartite graph $G = (V_1 \sqcup V_2, E)$ and integers k_1, k_2, the problem DensestQuotientGraph entails maximizing $e_G(\mathcal{P}_1, \mathcal{P}_2)$, the number of edges of the quotient graph of G by $\mathcal{P}_1, \mathcal{P}_2$, over all partitions $\mathcal{P}_1$ of V_1 in k_1 parts and $\mathcal{P}_2$ of V_2 in k_2 parts.

Proposition 5.7. *Given a deterministic channel W and integers k_1, k_2, it is equivalent to solve DetBCC on W, k_1, k_2 or DensestQuotientGraph on G_W, k_1, k_2. That is to say, given an optimal solution of one of those problems, one can efficiently construct an optimal solution of the other. Furthermore, their optimal values satisfy $k_1 k_2$DetBCC$(W, k_1, k_2) =$ DensestQuotientGraph(G_W, k_1, k_2).*

Proof. Consider an optimal solution e, d_1, d_2 of DetBCC. Note that d_1 defines a partition $\mathcal{P}_1$ of $\mathcal{Y}_1$ in k_1 parts and d_2 defines a partition $\mathcal{P}_2$ of $\mathcal{Y}_2$ in k_2 parts, with $\mathcal{P}_b^{i_b} := \{y_b \in \mathcal{Y}_b : d_b(y) = i_b\}$ for $b \in \{1, 2\}$. Then we have:

$$\begin{aligned} k_1 k_2 \mathrm{S}(W, k_1, k_2, e, d_1, d_2) &= \sum_{i_1, i_2} \mathbb{1}_{i_1 = d_1(W_1(e(i_1,i_2)))} \mathbb{1}_{i_2 = d_2(W_2(e(i_1,i_2)))} \\ &= \sum_{i_1, i_2} \mathbb{1}_{W_1(e(i_1,i_2)) \in \mathcal{P}_1^{i_1}} \mathbb{1}_{W_2(e(i_1,i_2)) \in \mathcal{P}_2^{i_2}} \ . \end{aligned} \tag{5.19}$$

However, since we consider an optimal solution, we have that:

$$\mathbb{1}_{W_1(e(i_1,i_2)) \in \mathcal{P}_1^{i_1}} \mathbb{1}_{W_2(e(i_1,i_2)) \in \mathcal{P}_2^{i_2}} = \max_{x \in \mathcal{X}} \mathbb{1}_{W_1(x) \in \mathcal{P}_1^{i_1}} \mathbb{1}_{W_2(x) \in \mathcal{P}_2^{i_2}} \ ,$$

as $e(i_1, i_2)$ appears only here in the objective value. Thus:

$$\begin{aligned} k_1 k_2 \mathrm{S}(W, k_1, k_2, e, d_1, d_2) &= \sum_{i_1, i_2} \max_{x \in \mathcal{X}} \mathbb{1}_{W_1(x) \in \mathcal{P}_1^{i_1}} \mathbb{1}_{W_2(x) \in \mathcal{P}_2^{i_2}} \\ &= \sum_{i_1, i_2} \mathbb{1}_{\exists (y_1, y_2) \in E^{G_W} : y_1 \in \mathcal{P}_1^{i_1} \text{ and } y_2 \in \mathcal{P}_2^{i_2}} \\ &= \sum_{i_1, i_2} \mathbb{1}_{E^{G_W} \cap \left(\mathcal{P}_1^{i_1} \times \mathcal{P}_2^{i_2}\right) \neq \emptyset} \\ &= e_{G_W}(\mathcal{P}_1, \mathcal{P}_2) \ , \end{aligned} \tag{5.20}$$

which proves that given an optimal solution of DetBCC, one can efficiently construct a solution $\mathcal{P}_1, \mathcal{P}_2$ of DensestQuotientGraph such that:

$$e_{G_W}(\mathcal{P}_1, \mathcal{P}_2) = k_1 k_2 \text{DetBCC}(W, k_1, k_2) \ .$$

For the other direction, consider an optimal solution $\mathcal{P}_1, \mathcal{P}_2$ of DensestQuotientGraph. We have as before that:

$$e_{G_W}(\mathcal{P}_1, \mathcal{P}_2) = \sum_{i_1, i_2} \max_{x \in \mathcal{X}} \mathbb{1}_{W_1(x) \in \mathcal{P}_1^{i_1}} \mathbb{1}_{W_2(x) \in \mathcal{P}_2^{i_2}} .$$

Now, let us define $e(i_1, i_2) \in \text{argmax}_{x \in \mathcal{X}} \mathbb{1}_{W_1(x) \in \mathcal{P}_1^{i_1}} \mathbb{1}_{W_2(x) \in \mathcal{P}_2^{i_2}}$ and $d_b(y_b)$ the index of the part of $\mathcal{P}_b$ where y_b lies, for $b \in \{1, 2\}$. With those definitions, we get again that:

$$\max_{x \in \mathcal{X}} \mathbb{1}_{W_1(x) \in \mathcal{P}_1^{i_1}} \mathbb{1}_{W_2(x) \in \mathcal{P}_2^{i_2}} = \mathbb{1}_{W_1(e(i_1,i_2)) \in \mathcal{P}_1^{i_1}} \mathbb{1}_{W_2(e(i_1,i_2)) \in \mathcal{P}_2^{i_2}} = \mathbb{1}_{i_1 = d_1(W_1(e(i_1,i_2)))} \mathbb{1}_{i_2 = d_2(W_2(e(i_1,i_2)))} ,$$

and thus we have:

$$\begin{aligned} e_{G_W}(\mathcal{P}_1, \mathcal{P}_2) &= \sum_{i_1, i_2} \max_{x \in \mathcal{X}} \mathbb{1}_{W_1(x) \in \mathcal{P}_1^{i_1}} \mathbb{1}_{W_2(x) \in \mathcal{P}_2^{i_2}} \\ &= \sum_{i_1, i_2} \mathbb{1}_{i_1 = d_1(W_1(e(i_1,i_2)))} \mathbb{1}_{i_2 = d_2(W_2(e(i_1,i_2)))} \\ &= k_1 k_2 \mathrm{S}(W, k_1, k_2, e, d_1, d_2) , \end{aligned} \tag{5.21}$$

which proves that given an optimal solution of DensestQuotientGraph, one can efficiently construct a solution e, d_1, d_2 of DetBCC such that:

$$k_1 k_2 \mathrm{S}(W, k_1, k_2, e, d_1, d_2) = \textsc{DensestQuotientGraph}(G_W, k_1, k_2) .$$

In particular, this implies that their optimal objective values satisfy $k_1 k_2 \textsc{DetBCC}(W, k_1, k_2) = \textsc{DensestQuotientGraph}(G_W, k_1, k_2)$. Therefore, the solutions of both problems constructed throughout the proof are in fact optimal. □

Remark. Note that all bipartite graphs can be written as G_W for some deterministic broadcast channel W, with W unique up to a permutation of $\mathcal{X}$.

5.3.2 Approximation Algorithm for DensestQuotientGraph

In this section, we will sort out how hard is DensestQuotientGraph, and thanks to Proposition 5.7, how hard is it to solve DetBCC.

Theorem 5.8. *There exists a polynomial-time $(1 - e^{-1})^2$-approximation algorithm for DensestQuotientGraph. Furthermore, it is NP-hard to solve exactly DensestQuotientGraph.*

Corollary 5.9. *There exists a polynomial-time $(1 - e^{-1})^2$-approximation algorithm for DetBCC.*

The approximation algorithm is a two-step process. First, we consider the problem of maximizing $\sum_{i_2=1}^{k_2} \min\left(k_1, \deg_{V_1, \mathcal{P}_2}(\mathcal{P}_2^{i_2})\right)$ over all partitions $\mathcal{P}_2$ of V_2 in k_2 parts. We will show that this is a special case of the submodular welfare problem, which can be approximated within a factor $1 - e^{-1}$ in polynomial time [Von08]. We then choose the partition $\mathcal{P}_1$ on V_1 in k_1 parts uniformly at random. This partition couple will give an objective value $e(\mathcal{P}_1, \mathcal{P}_2)$ within a $(1 - e^{-1})^2$ factor from the optimal solution in expectation.

Proof of Theorem 5.8. Consider first the hardness result. Let us show that the decision version of DensestQuotientGraph is NP-complete. It is in NP, the certificate being the two partitions and the selection of edges between those partitions. It is NP-hard as one of its particular cases is the SetSplitting problem (see for instance [GJ79]), in the case where $k_1 = 2$ and $k_2 = |V_2|$, by interpreting the neighbors of $v_2 \in V_2$ as a set covering elements of V_1.

We will show nonetheless that this problem can be approximated within a factor $(1-e^{-1})^2$ in polynomial time. First we consider the case where $k_2 = |V_2|$. We can then always assume that the right partition is $\mathcal{P}_2 := \{\{v_2\} : v_2 \in V_2\}$, which leads necessarily to a greater or equal number of edges in the quotient graph that with any other right partition. So, in that setting, we need only to find a partition of V_1 in k_1 parts maximizing the number of edges between vertices in the right part and the quotient of the left vertices.

First, one can note that the maximum value we can get is upper bounded by $\sum_{v_2 \in V_2} \min(k_1, \deg(v_2))$. Indeed, each vertex of v_2 van be connected at most to the k_1 parts of V_1, so its contribution is bounded by k_1, and there needs to be an edge to each part it is connected, so its contribution is also bounded by $\deg(v_2)$. Let us show that if we take a partition $\mathcal{P}_1$ of V_1 uniformly at random, we get:

$$\mathbb{E}_{\mathcal{P}_1}[e(\mathcal{P}_1, V_2)] \geq \left(1 - \left(1 - \frac{1}{k_1}\right)^{k_1}\right) \sum_{v_2 \in V_2} \min(k_1, \deg(v_2)) \geq (1-e^{-1}) \max_{\mathcal{P}_1} e(\mathcal{P}_1, V_2)\,.$$

We have that $e(\mathcal{P}_1, V_2) = \sum_{v_2 \in V_2} \deg_{\mathcal{P}_1, V_2}(v_2)$, so by linearity of expectation $\mathbb{E}_{\mathcal{P}_1}[e(\mathcal{P}_1, V_2)] = \sum_{v_2 \in V_2} \mathbb{E}_{\mathcal{P}_1}[\deg_{\mathcal{P}_1, V_2}(v_2)]$. However $\deg_{\mathcal{P}_1, V_2}(v_2) = |\{i_1 \in [k_1] : N(v_2) \cap \mathcal{P}_1^{i_1} \neq \emptyset\}|$. Recall also that for any v_1, $\mathbb{P}\left(v_1 \in \mathcal{P}_1^{i_1}\right) = \frac{1}{k_1}$ since the partition is taken uniformly at random. Thus, we get:

$$\begin{aligned}
\mathbb{E}_{\mathcal{P}_1}[\deg_{\mathcal{P}_1, V_2}(v_2)] &= \mathbb{E}_{\mathcal{P}_1}\left[|\{i_1 \in [k_1] : N(v_2) \cap \mathcal{P}_1^{i_1} \neq \emptyset\}|\right] = \mathbb{E}_{\mathcal{P}_1}\left[\sum_{i_1=1}^{k_1} \mathbb{1}_{N(v_2)\cap\mathcal{P}_1^{i_1}\neq\emptyset}\right] \\
&= \sum_{i_1=1}^{k_1} \mathbb{E}_{\mathcal{P}_1}\left[\mathbb{1}_{N(v_2)\cap\mathcal{P}_1^{i_1}\neq\emptyset}\right] = \sum_{i_1=1}^{k_1} \mathbb{P}\left(N(v_2) \cap \mathcal{P}_1^{i_1} \neq \emptyset\right) \\
&= \sum_{i_1=1}^{k_1} \left(1 - \mathbb{P}\left(N(v_2) \cap \mathcal{P}_1^{i_1} = \emptyset\right)\right) = \sum_{i_1=1}^{k_1} \left(1 - \prod_{v_1 \in N(v_2)} \mathbb{P}\left(v_1 \notin \mathcal{P}_1^{i_1}\right)\right) \\
&= \sum_{i_1=1}^{k_1} \left(1 - \prod_{v_1 \in N(v_2)} \mathbb{P}\left(v_1 \notin \mathcal{P}_1^{i_1}\right)\right) = k_1 \left(1 - \left(1 - \frac{1}{k_1}\right)^{\deg(v_2)}\right)\,,
\end{aligned} \tag{5.22}$$

since $\mathbb{P}\left(v_1 \notin \mathcal{P}_1^{i_1}\right) = 1 - \frac{1}{k_1}$ and $|N(v_2)| = \deg(v_2)$. So, in all:

$$\mathbb{E}_{\mathcal{P}_1}[e(\mathcal{P}_1, V_2)] = \sum_{v_2 \in V_2} \mathbb{E}_{\mathcal{P}_1}[\deg_{\mathcal{P}_1, V_2}(v_2)] = k_1 \sum_{v_2 \in V_2} \left(1 - \left(1 - \frac{1}{k_1}\right)^{\deg(v_2)}\right)\,.$$

However, the function $f : x \mapsto 1 - \left(1 - \frac{1}{k_1}\right)^x$ is nondecreasing concave with $f(0) = 0$, so $\frac{f(x)}{x} \geq \frac{f(y)}{y}$ for $x \leq y$. In particular, we have that:

$$f(\min(k_1, \deg(v_2))) \geq \frac{\min(k_1, \deg(v_2))}{k_1} f(k_1) \ ,$$

and thus:

$$\begin{aligned}
\mathbb{E}_{\mathcal{P}_1}[e(\mathcal{P}_1, V_2)] &\geq k_1 \sum_{v_2 \in V_2} \left(1 - \left(1 - \frac{1}{k_1}\right)^{\min(k_1, \deg(v_2))}\right) \\
&\geq k_1 \frac{\sum_{v_2 \in V_2} \min(k_1, \deg(v_2))}{k_1} \left(1 - \left(1 - \frac{1}{k_1}\right)^{k_1}\right) \\
&\geq \left(1 - \left(1 - \frac{1}{k_1}\right)^{k_1}\right) \sum_{v_2 \in V_2} \min\left(k_1, \deg(v_2)\right) \\
&\geq (1 - e^{-1}) \max_{\mathcal{P}_1} e(\mathcal{P}_1, V_2) \ ,
\end{aligned} \tag{5.23}$$

Let us now consider the general case with k_2 unconstrained. We apply the previous discussion on the graph $G^{V_1, \mathcal{P}_2}$ for some fixed partition $\mathcal{P}_2$ of V_2. Since $e_{G^{V_1, \mathcal{P}_2}}(\mathcal{P}_1, \mathcal{P}_2) = e(\mathcal{P}_1, \mathcal{P}_2)$, we have the upper bound:

$$\max_{\mathcal{P}_1} e(\mathcal{P}_1, \mathcal{P}_2) \leq \sum_{i_2=1}^{k_2} \min\left(k_1, \deg_{V_1, \mathcal{P}_2}(\mathcal{P}_2^{i_2})\right) \ ,$$

and the previous algorithm gives us a partition $\mathcal{P}_1$ of V_1 such that:

$$e(\mathcal{P}_1, \mathcal{P}_2) \geq (1 - e^{-1}) \sum_{i_2=1}^{k_2} \min\left(k_1, \deg_{V_1, \mathcal{P}_2}(\mathcal{P}_2^{i_2})\right) \ .$$

Therefore, let us focus on the following optimization problem:

$$\max_{\mathcal{P}_2} \sum_{i_2=1}^{k_2} \min\left(k_1, \deg_{V_1, \mathcal{P}_2}(\mathcal{P}_2^{i_2})\right) \ ,$$

We will give a $(1 - e^{-1})$ approximation algorithm running in polynomial time for this problem. In all, this will allow us to get in polynomial time a partition pair $(\mathcal{P}_1, \mathcal{P}_2)$ such that:

$$\begin{aligned}
e(\mathcal{P}_1, \mathcal{P}_2) &\geq (1 - e^{-1}) \sum_{i_2=1}^{k_2} \min\left(k_1, \deg_{V_1, \mathcal{P}_2}(\mathcal{P}_2^{i_2})\right) \\
&\geq (1 - e^{-1})^2 \max_{\mathcal{P}_2} \sum_{i_2=1}^{k_2} \min\left(k_1, \deg_{V_1, \mathcal{P}_2}(\mathcal{P}_2^{i_2})\right) \\
&\geq (1 - e^{-1})^2 \max_{\mathcal{P}_1, \mathcal{P}_2} e(\mathcal{P}_1, \mathcal{P}_2) \ .
\end{aligned} \tag{5.24}$$

The problem $\max_{\mathcal{P}_2} \sum_{i_2=1}^{k_2} \min\left(k_1, \deg_{V_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2})\right)$ is a particular instance of the submodular welfare problem discussed in [Von08]. Note first that $\deg_{V_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2}) = \deg_{V_1,\{\mathcal{P}_2^{i_2},V_2-\mathcal{P}_2^{i_2}\}}(\mathcal{P}_2^{i_2})$, as the degree of $\mathcal{P}_2^{i_2}$ does not depend on the rest of the partition $\mathcal{P}_2$. Then, $h(S_2) := \min\left(k_1, \deg_{V_1,\{S_2,V_2-S_2\}}(S_2)\right)$, for $S_2 \subseteq V_2$, is a nondecreasing submodular function, as $S_2 \mapsto \deg_{V_1,\{S_2,V_2-S_2\}}(S_2)$ is a nondecreasing submodular function on V_2 and $\min(k_1,\cdot)$ is nondecreasing concave. Thus, we want to maximize $\sum_{i_2=1}^{k_2} h(S_{i_2})$ where $(S_{i_2})_{i_2\in[k_2]}$ is a partition of items in V_2 among k_2 bidders. It is a particular case of the submodular welfare problem where each nondecreasing submodular utility weight is the same for all bidders and equal to h. Thus, thanks to [Von08], there exists a $(1-e^{-1})$ polynomial-time approximation of $\max_{\mathcal{P}_2} \sum_{i_2=1}^{k_2} \min\left(k_1, \deg_{V_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2})\right)$. □

5.3.3 Deterministic Non-Signaling Assisted Capacity Region

Thanks to Theorem 5.8 and Proposition 5.7, there exists a constant factor approximation algorithm for the broadcast channel coding problem running in polynomial time. We aim to show here that that the non-signaling assisted value is linked by a constant factor to the unassisted one. Indeed, the hope is that the non-signaling assisted program is linked to the linear relaxation of the unassisted problem, thus is likely a good approximation since the broadcast channel coding problem can be approximated in polynomial time.

This turns out to be true, and will be proved through the following theorem:

Theorem 5.10. *If W is a deterministic broadcast channel, then for all $\ell_1 \leq k_1$ and $\ell_2 \leq k_2$:*

$$\left(1 - \frac{k_1^{k_1}e^{-k_1}}{k_1!}\right)\left(1-\left(1-\frac{1}{\ell_1}\right)^{k_1}\right)\left(1-\left(1-\frac{1}{\ell_2}\right)^{k_2}\right)\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2) \leq \mathrm{S}(W,\ell_1,\ell_2)\ .$$

Corollary 5.11. *For any deterministic broadcast channel W, $\mathcal{C}^{\mathrm{NS}}(W) = \mathcal{C}(W)$.*

Proof. We apply Theorem 5.10 on the deterministic broadcast channel $W^{\otimes n}$.

We fix $k_1 = 2^{nR_1}, k_2 = 2^{nR_2}$ and $\ell_1 = \frac{2^{nR_1}}{n}, \ell_2 = \frac{2^{nR_2}}{n}$. Since $1-\left(1-\frac{1}{\ell}\right)^k \geq 1-e^{-\frac{k}{\ell}}$, we get:

$$\left(1 - \frac{k_1^{k_1}e^{-k_1}}{k_1!}\right)\left(1-e^{-n}\right)^2\mathrm{S}^{\mathrm{NS}}(W^{\otimes n},2^{nR_1},2^{nR_1}) \leq \mathrm{S}\left(W^{\otimes n},\frac{2^{nR_1}}{n},\frac{2^{nR_2}}{n}\right)\ .$$

As $\left(1 - \frac{k_1^{k_1}e^{-k_1}}{k_1!}\right)\left(1-e^{-n}\right)^2$ tends to 1 when n tends to infinity, we get that $\forall \varepsilon, \exists N \in \mathbb{N}, \forall n \geq N$:

$$(1-\varepsilon)\mathrm{S}^{\mathrm{NS}}(W^{\otimes n},2^{nR_1},2^{nR_1}) \leq \mathrm{S}(W^{\otimes n},2^{n(R_1-\frac{\log(n)}{n})},2^{n(R_2-\frac{\log(n)}{n})})\ .$$

Thus, if $\lim_{n\to+\infty} \mathrm{S}^{\mathrm{NS}}(W^{\otimes n},2^{nR_1},2^{nR_1}) = 1$, we have that for all $R_1' < R_1$ and $R_2' < R_2$:

$$\lim_{n\to+\infty} \mathrm{S}(W^{\otimes n},2^{nR_1'},2^{nR_1'}) \geq 1-\varepsilon\ .$$

Since this is true for all $\varepsilon > 0$, we get in fact that $\lim_{n \to +\infty} \mathrm{S}(W^{\otimes n}, 2^{nR_1'}, 2^{nR_1'}) = 1$. This implies that $\mathcal{C}^{\mathrm{NS}}(W) \subseteq \mathcal{C}(W)$, and thus that the capacity regions are equal as the other inclusion is always satisfied.

□

Let us now prove the main result:

Proof of Theorem 5.10. The proof will be done in three parts. We will work on the graph G_W:

1. First, we prove that for any partition $\mathcal{P}_2$ of $\mathcal{Y}_2$ in ℓ_2 parts:

$$\mathrm{S}(W, \ell_1, \ell_2) \geq \left(1 - \left(1 - \frac{1}{\ell_1}\right)^{k_1}\right) \frac{\sum_{i_2=1}^{\ell_2} \min\left(k_1, \deg_{\mathcal{Y}_1, \mathcal{P}_2}(\mathcal{P}_2^{i_2})\right)}{k_1 \ell_2} ,$$

2. Then, we show that there exists a partition $\mathcal{P}_2$ such that:

$$\frac{\sum_{i_2=1}^{\ell_2} \min\left(k_1, \deg_{\mathcal{Y}_1, \mathcal{P}_2}(\mathcal{P}_2^{i_2})\right)}{k_1 \ell_2} \geq \left(1 - \frac{k_1^{k_1} e^{-k_1}}{k_1!}\right) \left(1 - \left(1 - \frac{1}{\ell_2}\right)^{k_2}\right) \frac{\min\left(k_1 k_2, \sum_{y_1} \min(k_2, \deg(y_1))\right)}{k_1 k_2}$$

3. Finally, we prove that:

$$\frac{\min\left(k_1 k_2, \sum_{y_1} \min(k_2, \deg(y_1))\right)}{k_1 k_2} \geq \mathrm{S}^{\mathrm{NS}}(W, k_1, k_2) .$$

By combining these three inequalities, we get precisely the claimed result.

1. This part shares a lot of similarities with the proof of Theorem 5.8, which we will adapt to this particular situation. Let us show that if we take a partition $\mathcal{P}_1$ of $\mathcal{Y}_1$ of size ℓ_1 uniformly at random, we get, for some fixed $\mathcal{P}_2$ of size ℓ_2:

$$\mathbb{E}_{\mathcal{P}_1}[e_{G_W}(\mathcal{P}_1, \mathcal{P}_2)] \geq \frac{\ell_1}{k_1} \left(1 - \left(1 - \frac{1}{\ell_1}\right)^{k_1}\right) \sum_{i_2=1}^{\ell_2} \min\left(k_1, \deg_{\mathcal{Y}_1, \mathcal{P}_2}(\mathcal{P}_2^{i_2})\right) .$$

   Then, since $\ell_1 \ell_2 \mathrm{S}(W, \ell_1, \ell_2) = \underset{\mathcal{P}_1 \text{ in } \ell_1 \text{ parts}, \mathcal{P}_2 \text{ in } \ell_2 \text{ parts}}{\text{maximize}} e_{G_W}(\mathcal{P}_1, \mathcal{P}_2)$ by Proposition 5.7, this will imply that:

$$\mathrm{S}(W, \ell_1, \ell_2) \geq \frac{1}{\ell_1 \ell_2} \mathbb{E}_{\mathcal{P}_1}[e_{G_W}(\mathcal{P}_1, \mathcal{P}_2)] \geq \left(1 - \left(1 - \frac{1}{\ell_1}\right)^{k_1}\right) \frac{\sum_{i_2=1}^{\ell_2} \min\left(k_1, \deg_{\mathcal{Y}_1, \mathcal{P}_2}(\mathcal{P}_2^{i_2})\right)}{k_1 \ell_2} .$$

   We have that $e_{G_W}(\mathcal{P}_1, \mathcal{P}_2) = \sum_{i_2=1}^{\ell_2} \deg_{\mathcal{P}_1, \mathcal{P}_2}(\mathcal{P}_2^{i_2})$, so by linearity of expectation, we have that $\mathbb{E}_{\mathcal{P}_1}[e_{G_W}(\mathcal{P}_1, \mathcal{P}_2)] = \sum_{i_2=1}^{\ell_2} \mathbb{E}_{\mathcal{P}_1}[\deg_{\mathcal{P}_1, \mathcal{P}_2}(\mathcal{P}_2^{i_2})]$, so we will focus on the contribution of one particular $\mathcal{P}_2^{i_2}$.

Then, we have that $\deg_{\mathcal{P}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2}) = |\{i_1 \in [\ell_1] : N_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2}) \cap \mathcal{P}_1^{i_1} \neq \emptyset\}|$. Recall that $\mathbb{P}\left(v_1 \in \mathcal{P}_1^{i_1}\right) = \frac{1}{\ell_1}$ for any v_1 since the partition is taken uniformly at random. Thus:

$$
\begin{aligned}
\mathbb{E}_{\mathcal{P}_1}[\deg_{\mathcal{P}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2})] &= \mathbb{E}_{\mathcal{P}_1}\left[|\{i_1 \in [\ell_1] : N_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2}) \cap \mathcal{P}_1^{i_1} \neq \emptyset\}|\right] = \mathbb{E}_{\mathcal{P}_1}\left[\sum_{i_1=1}^{\ell_1} \mathbb{1}_{N_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2})\cap\mathcal{P}_1^{i_1}}\right. \\
&= \sum_{i_1=1}^{\ell_1} \mathbb{E}_{\mathcal{P}_1}\left[\mathbb{1}_{N_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2})\cap\mathcal{P}_1^{i_1}\neq\emptyset}\right] = \sum_{i_1=1}^{\ell_1} \mathbb{P}\left(N_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2}) \cap \mathcal{P}_1^{i_1} \neq \emptyset\right) \\
&= \sum_{i_1=1}^{\ell_1} \left(1 - \mathbb{P}\left(N_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2}) \cap \mathcal{P}_1^{i_1} = \emptyset\right)\right) = \sum_{i_1=1}^{\ell_1}\left(1 - \prod_{v_1 \in N(\mathcal{P}_2^{i_2})} \mathbb{P}\left(v_1 \notin \mathcal{P}_1^{i}\right.\right. \\
&= \ell_1\left(1 - \left(1 - \frac{1}{\ell_1}\right)^{\deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2})}\right) .
\end{aligned}
\tag{5.25}
$$

So, in all we have that:

$$
\mathbb{E}_{\mathcal{P}_1}[e_{G_W}(\mathcal{P}_1,\mathcal{P}_2)] = \sum_{i_2=1}^{\ell_2} \mathbb{E}_{\mathcal{P}_1}[\deg_{\mathcal{P}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2})] = \ell_1 \sum_{i_2=1}^{\ell_2}\left(1 - \left(1 - \frac{1}{\ell_1}\right)^{\deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2})}\right) .
$$

However the function $f : x \mapsto 1 - \left(1 - \frac{1}{\ell_1}\right)^x$ is nondecreasing concave with $f(0) = 0$, so $\frac{f(x)}{x} \geq \frac{f(y)}{y}$ for $x \leq y$. In particular, we have that:

$$
f(\min(k_1, \deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2}))) \geq \frac{\min(k_1, \deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2})))}{k_1} f(k_1) ,
$$

and thus:

$$
\begin{aligned}
\mathbb{E}_{\mathcal{P}_1}[e_{G_W}(\mathcal{P}_1,\mathcal{P}_2)] &\geq \ell_1 \sum_{i_2=1}^{\ell_2}\left(1 - \left(1 - \frac{1}{\ell_1}\right)^{\min(k_1,\deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2}))}\right) \\
&\geq \ell_1 \frac{\sum_{i_2=1}^{\ell_2} \min(k_1, \deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2}))}{k_1}\left(1 - \left(1 - \frac{1}{\ell_1}\right)^{k_1}\right) \\
&= \frac{\ell_1}{k_1}\left(1 - \left(1 - \frac{1}{\ell_1}\right)^{k_1}\right)\sum_{i_2=1}^{\ell_2} \min\left(k_1, \deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2})\right) ,
\end{aligned}
\tag{5.26}
$$

which concludes the first part of the proof.

2. Let us take $\mathcal{P}_2$ a partition of $\mathcal{Y}_2$ of size ℓ_2 uniformly at random, and let us prove that

$$
\mathbb{E}\left[\sum_{i_2=1}^{\ell_2} \min\left(k_1, \deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2})\right)\right]
$$

is greater than or equal to

$$\frac{\ell_2}{k_2}\left(1-\frac{k_1^{k_1}e^{-k_1}}{k_1!}\right)\left(1-\left(1-\frac{1}{\ell_2}\right)^{k_2}\right)\min\left(k_1k_2,\sum_{y_1}\min(k_2,\deg(y_1))\right) .$$

First, $\sum_{i_2=1}^{\ell_2}\min\left(k_1,\deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2})\right) = \sum_{i_2=1}^{\ell_2}\varphi(\deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2}))$ with $\varphi(j) := \min(k_1,j)$ which is a concave function. We will use several tools introduced in [BFF21] to handle those concave functions in approximation algorithms. Specifically, we have that the Poisson concavity ratio $\alpha_\varphi := \inf_{x\in\mathbb{R}^+}\frac{\mathbb{E}[\varphi(\mathrm{Poi}(x))]}{\varphi(x))} = 1-\frac{k_1^{k_1}e^{-k_1}}{k_1!}$ for that particular function. We will also need the following lemma, whose proof is based on convex order:

Lemma 5.12 (Lemma 2.2 of [BFF21])**.** *For φ concave, and $p\in[0,1]^m$, we have:*

$$\mathbb{E}\Big[\varphi\Big(\sum_{i=1}^m \mathit{Ber}(p_i)\Big)\Big] \geq \mathbb{E}\Big[\varphi\Big(\mathit{Poi}\Big(\sum_{i=1}^m p_i\Big)\Big)\Big] .$$

Let us find the law of $\deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2})$:

$$\begin{aligned}\deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2}) &= \sum_{y_1}\mathbb{1}_{N(y_1)\cap\mathcal{P}_2^{i_2}\neq\emptyset} = \sum_{y_1}\left(1-\mathbb{1}_{N(y_1)\cap\mathcal{P}_2^{i_2}=\emptyset}\right) = \sum_{y_1}\left(1-\mathbb{1}_{\forall y_2\in N(y_1),y_2\notin\mathcal{P}_2^{i_2}}\right)\\ &= \sum_{y_1}\mathrm{Ber}\left(1-\left(1-\frac{1}{\ell_2}\right)^{\deg_(y_1)}\right)\end{aligned} \tag{5.27}$$

Thus:

$$\begin{aligned}\mathbb{E}\left[\varphi(\deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2}))\right] &= \mathbb{E}\left[\varphi\left(\sum_{y_1}\mathrm{Ber}\left(1-\left(1-\frac{1}{\ell_2}\right)^{\deg(y_1)}\right)\right)\right]\\ &\geq \mathbb{E}\left[\varphi\left(\mathrm{Poi}\left(\sum_{y_1}\left(1-\left(1-\frac{1}{\ell_2}\right)^{\deg(y_1)}\right)\right)\right)\right] \text{ by Lemma 5.12}\\ &\geq \alpha_\varphi\varphi\left(\sum_{y_1}\left(1-\left(1-\frac{1}{\ell_2}\right)^{\deg(y_1)}\right)\right) \text{ by definition of } \alpha_\varphi.\end{aligned} \tag{5.28}$$

But:

$$\begin{aligned}\sum_{y_1}\left(1-\left(1-\frac{1}{\ell_2}\right)^{\deg(y_1)}\right) &\geq \sum_{y_1}\left(1-\left(1-\frac{1}{\ell_2}\right)^{\min(k_2,\deg(y_1)))}\right)\\ &\geq \left(1-\left(1-\frac{1}{\ell_2}\right)^{k_2}\right)\frac{1}{k_2}\sum_{y_1}\min\left(k_2,\deg(y_1)\right) ,\end{aligned} \tag{5.29}$$

as before. Since φ is concave and $\varphi(0) = 0$, we have in particular that for all $0 \leq c \leq 1$ and $x \in \mathbb{R}$, $\varphi(cx) \geq c\varphi(x)$. We know also that φ is nondecreasing. This implies that:

$$\begin{aligned} \varphi\left(\sum_{y_1}\left(1-\left(1-\frac{1}{\ell_2}\right)^{\deg(y_1)}\right)\right) &\geq \varphi\left(\left(1-\left(1-\frac{1}{\ell_2}\right)^{k_2}\right)\frac{1}{k_2}\sum_{y_1}\min\left(k_2,\deg(y_1)\right)\right) \\ &\geq \left(1-\left(1-\frac{1}{\ell_2}\right)^{k_2}\right)\varphi\left(\frac{1}{k_2}\sum_{y_1}\min\left(k_2,\deg(y_1)\right)\right) , \end{aligned} \tag{5.30}$$

as $0 \leq 1-\left(1-\frac{1}{\ell_2}\right)^{k_2} \leq 1$. Thus:

$$\begin{aligned} \mathbb{E}\left[\varphi(\deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2}))\right] &\geq \alpha_\varphi\left(1-\left(1-\frac{1}{\ell_2}\right)^{k_2}\right)\min\left(k_1,\frac{1}{k_2}\sum_{y_1}\min\left(k_2,\deg(y_1)\right)\right) \\ &= \frac{1}{k_2}\left(1-\frac{k_1^{k_1}e^{-k_1}}{k_1!}\right)\left(1-\left(1-\frac{1}{\ell_2}\right)^{k_2}\right)\min\Bigg(k_1k_2,\sum_{y_1}\min\big(k_2,\deg(y \end{aligned} \tag{5.31}$$

since $\alpha_\varphi = 1-\frac{k_1^{k_1}e^{-k_1}}{k_1!}$.

Finally, $\mathbb{E}\left[\sum_{i_2=1}^{\ell_2}\min\left(k_1,\deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2})\right)\right] = \sum_{i_2=1}^{\ell_2}\mathbb{E}\left[\varphi(\deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2}))\right]$, so we get that

$$\mathbb{E}\left[\sum_{i_2=1}^{\ell_2}\min\left(k_1,\deg_{\mathcal{Y}_1,\mathcal{P}_2}(\mathcal{P}_2^{i_2})\right)\right]$$

is larger than

$$\frac{\ell_2}{k_2}\left(1-\frac{k_1^{k_1}e^{-k_1}}{k_1!}\right)\left(1-\left(1-\frac{1}{\ell_2}\right)^{k_2}\right)\min\left(k_1k_2,\sum_{y_1}\min\left(k_2,\deg(y_1)\right)\right) .$$

Thus, in particular, there exists some partition $\mathcal{P}_2$ that satisfies the same inequality, which concludes the second part of the proof.

3. Let us consider an optimal solution $r_{x,y_1,y_2}, p_x, r^1_{x,y_1}, r^2_{x,y_2}$ of the program computing $\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2)$, so that $\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2) = \frac{1}{k_1k_2}\sum_x r_{x,W_1(x),W_2(x)}$.

   a) It comes directly from $r_{x,y_1,y_2} \leq p_x$ that:

   $$\sum_x r_{x,W_1(x),W_2(x)} \leq \sum_x p_x = k_1k_2 .$$

   b) $\sum_x r_{x,W_1(x),W_2(x)} = \sum_{y_1}\sum_{x:W_1(x)=y_1} r_{x,y_1,W_2(x)}$ and we have that:

      i. $\sum_{x:W_1(x)=y_1} r_{x,y_1,W_2(x)} \leq \sum_{x:W_1(x)=y_1} 1 = \deg(y_1)$,

      ii. $\sum_{x:W_1(x)=y_1} r_{x,y_1,W_2(x)} \leq \sum_{x:W_1(x)=y_1} r^1_{x,y_1} \leq \sum_x r^1_{x,y_1} = k_2$, so $\sum_{x:W_1(x)=y_1} r_{x,y_1,W_2(x)} \leq \min(k_2, \deg(y_1))$, and thus:

$$\sum_x r_{x,W_1(x),W_2(x)} \leq \sum_{y_1} \min(k_2, \deg(y_1)) \; .$$

In all, we get that:

$$\mathrm{S}^{\mathrm{NS}}(W,k_1,k_2) = \frac{1}{k_1k_2}\sum_x r_{x,W_1(x),W_2(x)} \leq \frac{\min\left(k_1k_2, \sum_{y_1}\min(k_2,\deg(y_1))\right)}{k_1k_2} \; ,$$

which concludes the third and last part of the proof.

□

5.4 Hardness of Approximation of Broadcast Channel Coding

The goal of this section is to show that the general broadcast channel coding problem cannot be approximated in polynomial time within a $\Omega(1)$ factor, under reasonable hardness assumptions. It will be a good insight that non-signaling assistance will enlarge the capacity region of the channel as discussed in the introduction.

Formally, one would want to show that it is NP-hard to approximate this problem within a $\Omega(1)$ factor in polynomial time. As a first step towards this goal, we will prove a $\Omega\left(\frac{1}{\sqrt{m}}\right)$-approximation hardness in the value query model.

First, let us introduce formally the problem:

Definition 5.8 (BCC)**.** Given a channel W and integers k_1 and k_2, the broadcast channel coding problem, which we call BCC, entails maximizing

$$\mathrm{S}(W,k_1,k_2,e,d_1,d_2) := \frac{1}{k_1k_2}\sum_{y_1,y_2,i_1,i_2} W(y_1y_2|e(i_1,i_2))\mathbb{1}_{d_1(y_1)=i_1,d_2(y_2)=i_2}$$

over all functions $e:[k_1]\times[k_2]\to\mathcal{X}$, $d_1:\mathcal{Y}_1\to[k_1]$, $d_2:\mathcal{Y}_2\to[k_2]$.

As in the deterministic case, we restrict ourselves to deterministic encoders and decoders, which does not change the value nor the hardness of the problem. Also, it can be equivalently stated in terms of partitions corresponding to d_1, d_2 as:

Proposition 5.13 (Equivalent formulation of BCC)**.** *Given a channel W and integers k_1 and k_2, the broadcast channel coding problem, which we call BCC, entails maximizing:*

$$\frac{1}{k_1k_2}\sum_{i_1,i_2}\max_x \sum_{y_1\in\mathcal{P}_1^{i_1},y_2\in\mathcal{P}_1^{i_2}} W(y_1y_2|x) \; ,$$

over all partitions $\mathcal{P}_1$ of $\mathcal{Y}_1$ in k_1 parts and $\mathcal{P}_2$ of $\mathcal{Y}_2$ in k_2 parts.

5.4.1 Social Welfare Reformulation

The social welfare maximization problem is defined as follows: given a set M of m items as well as k bidders with their associated utilities $\left(v_i : 2^M \to \mathbb{R}_+\right)_{i\in[k]}$, the goal is to partition M between the bidders to maximize the sum of their utilities. Formally, we want to compute:

$$\underset{\mathcal{P} \text{ partition in } k \text{ parts of } M}{\text{maximize}} \sum_{i=1}^{k} v_i\left(\mathcal{P}^i\right) .$$

Let us show that the subproblem of BCC restricted to $k_2 = |\mathcal{Y}_2|$ can be reformulated as a particular instance of the social welfare maximization problem. In that case, it is easy to see that $\mathcal{P}_2 = (\{y_2\})_{y_2\in\mathcal{Y}_2}$ is always an optimal solution. Indeed, for any partition $\mathcal{P}_2$, we have:

$$\begin{aligned}\frac{1}{k_1|\mathcal{Y}_2|}\sum_{i_1,i_2}\max_x \sum_{y_1\in\mathcal{P}_1^{i_1},y_2\in\mathcal{P}_2^{i_2}} W(y_1y_2|x) &\leq \frac{1}{k_1|\mathcal{Y}_2|}\sum_{i_1,i_2}\sum_{y_2\in\mathcal{P}_2^{i_2}}\max_x \sum_{y_1\in\mathcal{P}_1^{i_1}} W(y_1y_2|x)\\ &= \frac{1}{k_1|\mathcal{Y}_2|}\sum_{i_1}\sum_{y_2\in\mathcal{Y}_2}\max_x \sum_{y_1\in\mathcal{P}_1^{i_1}} W(y_1y_2|x) .\end{aligned} \tag{5.32}$$

Therefore, the objective function becomes:

$$\mathrm{S}^1(W,k_1,\mathcal{P}_1) := \frac{1}{k_1}\sum_{i_1=1}^{k_1} f_W^1(\mathcal{P}_1^{i_1}) \text{ with } f_W^1(S_1) := \frac{1}{|\mathcal{Y}_2|}\sum_{y_2}\max_x \sum_{y_1\in S_1} W(y_1y_2|x) .$$

Hence, up to a multiplicative factor k_1, maximizing $\mathrm{S}^1(W,k_1,\mathcal{P}_1)$ over all partitions $\mathcal{P}_1$ of size k_1 is a particular case of the social welfare maximization problem with a common utility f_W^1 for all k_1 bidders.

5.4.2 Value Query Hardness

Let us first introduce the value query model. As described in [DS06, MSV08], a value query to a utility v asks for the value of some input set $S \subseteq M$, and gets as response $v(S) \in \mathbb{R}_+$. In the value query model, we aim at solving the social welfare maximization problem accessing the data only through value queries to $(v_i)_{i\in[k]}$.

This is more restricted than using any algorithm, but in such a model, it is possible to show unconditional lower bounds on the number of queries needed to solve a given problem within an approximation rate. In the case of the social welfare maximization problem with XOS utility functions, the approximation rate achievable in polynomial time has been proved in [DS06, MSV08] to be of the order of $\Theta\left(\frac{1}{\sqrt{m}}\right)$. Specifically, in [DS06], a $\Omega\left(\frac{1}{m^{\frac{1}{2}}}\right)$-approximation in polynomial time was given, and in [MSV08], it has been shown that any $\Omega\left(\frac{1}{m^{\frac{1}{2}-\varepsilon}}\right)$-approximation for $\varepsilon > 0$ requires an exponential number of value queries. We will adapt their proof in the particular case of one common XOS utility function of the form f^1_W for some broadcast channel W. But first, let us introduce the definition of XOS functions and prove that f^1_W is one of those.

Definition 5.9. A linear valuation function (also known as additive) is a set function $a : 2^M \to \mathbb{R}_+$ that assigns a non-negative value to every singleton $\{j\}$ for $j \in M$, and for all $S \subseteq M$ it holds that $a(S) = \sum_{j \in S} a(\{j\})$.

A fractionally sub-additive function (XOS) is a set function $f : 2^M \to \mathbb{R}_+$, for which there is a finite set of linear valuation functions $A = \{a_1, \dots, a_\ell\}$ such that $f(S) = \max_{i \in [\ell]} a_i(S)$ for every $S \subseteq M$.

Remark. Note that the size of A is not bounded in the definition.

Proposition 5.14. *f^1_W is XOS.*

Proof.

$$\begin{aligned}
f^1_W(S) &= \frac{1}{|\mathcal{Y}_2|} \sum_{y_2} \max_x \sum_{y_1 \in S_1} W(y_1y_2|x) = \max_{\lambda : \mathcal{Y}_2 \to \mathcal{X}} a_\lambda(S) \text{ , where} \\
a_\lambda(S) &= \frac{1}{|\mathcal{Y}_2|} \sum_{y_2} \sum_{y_1 \in S} W(y_1y_2|\lambda(y_2)) = \sum_{y_1 \in S} \left[\frac{1}{|\mathcal{Y}_2|} \sum_{y_2} W(y_1y_2|m(y_2)) \right] \qquad (5.33) \\
&= \sum_{y_1 \in S} a_\lambda(\{y_1\})) \text{ with } a_\lambda(\{y_1\})) = \frac{1}{|\mathcal{Y}_2|} \sum_{y_2} W(y_1y_2|\lambda(y_2)) \in \mathbb{R}_+
\end{aligned}$$

So f^1_W is the maximum of the set of a_λ for $\lambda \in \mathcal{X}^{\mathcal{Y}_1}$, which are linear valuation functions, thus f^1_W is XOS. □

In order to prove our hardness result, we will need the following version of the Chernoff-Hoeffding bound with negatively associated random variables, which is a weaker notion of independence where tail bounds are still valid (see [DR98] for further discussion on negatively associated random variables):

Proposition 5.15 (Chernoff-Hoeffding bound)**.** *Let $X_1, \dots, X_m$ be negatively associated Bernouilli random variables of parameter p. Then for $0 < \varepsilon \leq \frac{1}{2}$:*

$$\mathbb{P}\left(\frac{1}{m} \sum_{i=1}^m X_i > (1+\varepsilon)p \right) \leq e^{-\frac{pm\varepsilon^2}{4}} .$$

Proof. Usual proofs of the Chernoff-Hoeffding bound work in the same way with negatively associated variables as pointed out by [DR98]. So, one obtain as in the original proof [Hoe63] that:

$$\mathbb{P}\left(\frac{1}{m} \sum_{i=1}^m X_i > (1+\varepsilon)p \right) \leq e^{-D((1+\varepsilon)p||p)m} ,$$

with $D(x||y) := x \ln\left(\frac{x}{y}\right) + (1-x)\ln\left(\frac{1-x}{1-y}\right)$ the Kullback–Leibler divergence between Bernoulli distributed random variables with parameters x and y. Using the bound $D((1+\varepsilon)p||p) \geq \frac{\varepsilon^2 p}{4}$ for $0 < \varepsilon < \frac{1}{2}$ concludes the proof. □

Let us now state the value query hardness of approximation of the broadcast channel problem:

Theorem 5.16. *In the value query model, for any fixed $\varepsilon > 0$, a $\Omega\left(\frac{1}{m^{\frac{1}{2}-\varepsilon}}\right)$-approximation algorithm for the broadcast channel coding problem on W, k_1, k_2, restricted to the case of $|\mathcal{Y}_2| = k_2$ and $m = |\mathcal{Y}_1| = k_1^2$, requires exponentially many value queries to f_W^1.*

Remark. As our problem is a particular instance of the social welfare maximization problem with XOS functions, the polynomial-time $\Omega\left(\frac{1}{m^{\frac{1}{2}}}\right)$-approximation from [DS06] works also here.

Proof. The proof is inspired by Theorem 3.1 of [MSV08]. We will show using probabilistic arguments that any $\Omega\left(\frac{1}{m^{\frac{1}{2}-\varepsilon}}\right)$-approximation algorithm requires an exponential number of value queries. Let us fix a small constant $\delta > 0$. We choose $k_1 \in \mathbb{N}$ as the number of messages (the bidders) and the output space $\mathcal{Y}_1 := [m]$ with $m := k_1^2$ (the items). Then, we choose uniformly at random an equi-partition of $\mathcal{Y}_1$ in k_1 parts of size k_1, which we name $T_1, \dots, T_{k_1}$.

Let us define now $\mathcal{Y}_2 := [m + k_1 + 1]$. We take $\mathcal{X} := \mathcal{Y}_2 = [m + 1 + k_1]$ as well. We can now define our broadcast channel W, with some positive constant C to be fixed later to guarantee that W is a conditional probability distribution. Let us define its value for $y_2 = 1$:

$$W(y_1 1|x) := C \times \begin{cases} m^{2\delta} \mathbb{1}_{y_1 = x} & \text{when } 1 \leq x \leq m \\ \frac{1}{m^{\frac{1}{2}-\delta}} & \text{when } x = m+1 \\ \mathbb{1}_{y_1 \in T_j} & \text{when } 1 \leq j := x - (m+1) \leq k_1 \end{cases}$$

Then, we define other y_2 inputs as translations of $W(y_1 1|x)$. Specifically, we define:

$$W(y_1 y_2|x) := W(y_1 1|t_{y_2 - 1}(x)) \text{ with } t_s(x) := 1 + [(x - 1 + s) \mod (m + k_1 + 1)] \,.$$

All coefficients are nonnegative. So W will be a channel if for all x, $\sum_{y_1,y_2} W(y_1 y_2|x) = 1$. However, one has, for some fixed x_0:

$$\begin{aligned} \sum_{y_1,y_2} W(y_1 y_2|x_0) &= \sum_{y_1}\sum_{y_2} W(y_1 y_2|x_0) = \sum_{y_1}\sum_{y_2} W(y_1 1|t_{y_2-1}(x_0)) = \sum_{y_1}\sum_{x} W(y_1 1|x) \\ &= C \sum_{y_1} \left[\sum_{1 \leq i \leq m} m^{2\delta} \mathbb{1}_{y_1 = i} + \frac{1}{m^{\frac{1}{2}-\delta}} + \sum_{1 \leq j \leq k_1} \mathbb{1}_{y_1 \in T_j} \right] \\ &= C \left[\sum_{1 \leq i \leq m} m^{2\delta} + m \times \frac{1}{m^{\frac{1}{2}-\delta}} + \sum_{1 \leq j \leq k_1} k_1 \right] \\ &= 1 \text{ by choosing } C = \frac{1}{m^{1+2\delta} + m^{\frac{1}{2}+\delta} + m}, \text{ which does not depend on } x_0 \,. \end{aligned} \tag{5.34}$$

Thus, we have defined a correct instance of our problem. Note that on this instance, we have:

$$\begin{aligned}
f_W^1(S) &= \frac{1}{|\mathcal{Y}_2|}\sum_{y_2}\max_x \sum_{y_1\in S} W(y_1y_2|x) = \sum_{y_2}\max_x \sum_{y_1\in S} W(y_11|t_{y_2-1}(x)) \\
&= \frac{m+k_1+1}{|\mathcal{Y}_2|}\max_x \sum_{y_1\in S} W(y_11|x) \text{ since } t_{y_2-1} \text{ bijection} \\
&= \frac{C(m+k_1+1)}{|\mathcal{Y}_2|} \times \max \begin{cases} m^{2\delta}|\{i\}\cap S| & \text{for } 1\leq i\leq m \\ \frac{1}{m^{\frac{1}{2}-\delta}}|S| & \\ |T_j\cap S| & \text{for } 1\leq j\leq k_1 \end{cases}
\end{aligned} \tag{5.35}$$

Let us also consider an alternate broadcast channel W', with the only difference that $\mathbb{1}_{y_1\in T_j}$ is replaced by $\frac{1}{m^{\frac{1}{2}}}$, for $j\in[k_1]$. For that channel, the constant C remains the same (since $\sum_j\sum_{y_1}\mathbb{1}_{y_1\in T_j} = k_1\times k_1 = k_1\times m\times\frac{1}{k_1} = \sum_j\sum_{y_1}\frac{1}{m^{\frac{1}{2}}}$), so we get that:

$$\begin{aligned}
f_{W'}^1(S) &= \frac{C(m+k_1+1)}{|\mathcal{Y}_2|} \times \max \begin{cases} m^{2\delta}|\{i\}\cap S| & \text{for } 1\leq i\leq m \\ \frac{1}{m^{\frac{1}{2}-\delta}}|S| & \\ \frac{1}{m^{\frac{1}{2}}}|S| & \text{for } 1\leq j\leq k_1 \end{cases} \\
&= \frac{C(m+k_1+1)}{|\mathcal{Y}_2|} \times \max \begin{cases} m^{2\delta}|\{i\}\cap S| & \text{for } 1\leq i\leq m \\ \frac{1}{m^{\frac{1}{2}-\delta}}|S| & \end{cases}
\end{aligned} \tag{5.36}$$

since $\frac{1}{m^{\frac{1}{2}}}|S|\leq\frac{1}{m^{\frac{1}{2}-\delta}}|S|$. Let us consider normalized versions $v(S) := \frac{|\mathcal{Y}_2|}{C(m+k_1+1}f_W^1(S)$ and $v'(S) := \frac{|\mathcal{Y}_2|}{C(m+k_1+1}f_{W'}^1(S)$, so distinguishing between v and v' is the same as distinguishing between f_W^1 and $f_{W'}^1$. We will prove that it takes an exponential number of value queries to distinguish between v and v'. On the one hand, one can easily show that the maximum value of the social welfare problem with v' is $(k_1-1)m^{2\delta}+\frac{1}{m^{\frac{1}{2}-\delta}}(m-(k_1-1)) = O(m^{\frac{1}{2}+2\delta})$, obtained taking (k_1-1) singletons as the first components of the partition (the bidders), giving the rest of $\mathcal{Y}_1$ (the items) to the last. On the other hand, the maximum value of the social welfare problem with v is $k_1\times k_1 = m$, obtained with the partition $T_1,\dots,T_{k_1}$. The fact that it requires an exponential number of value queries to distinguish between the two situations will imply that one cannot get an approximation rate better than $\Omega\left(\frac{1}{m^{\frac{1}{2}-2\delta}}\right)$ in less than an exponential number of value queries.

We will now prove that distinguishing between v and v' requires an exponential number of value queries. Note first that $v(\emptyset)=v'(\emptyset)=0$, so we do not need to consider empty sets.

Let us fix some non-empty set $S\subseteq[m]$. Let us define the random boolean variables $X_j^i := \mathbb{1}_{i\in T_j}$ for $j\in[k_1]$ and $i\in[m]$. By construction of the random equi-partition $T_1,\dots,T_{k_1}$, $(X_j^i)_{i\in[m]}$ is a permutation distribution (see Definition 2.10 of [JDP83]) of $(0,\dots,0,1,\dots,1)$ with $m-k_1$ zeros and k_1 ones, each X_j^i following a Bernouilli law of parameter $p:=\frac{1}{k_1}$. Thus it is negatively associated (by Theorem 2.11 of [JDP83]), and the sub-family $(X_j^i)_{i\in S}$ is negatively associated as well. Note in particular that $|T_j\cap S| =$

$\sum_{i \in S} X_j^i$ is a sum of negatively associated Bernouilli variables of the same parameter p, so Chernoff-Hoeffding bound as depicted in Proposition 5.15 holds.

Let us first assume that S is of size $0 < |S| \leq m^{\frac{1}{2}+\delta}$. Then, we have that $\frac{1}{m^{\frac{1}{2}-\delta}}|S| \leq m^{2\delta}$, so we get that $v'(S) = m^{2\delta}$. On the other hand, we have that:

$$v(S) = \max \begin{cases} m^{2\delta} \\ |T_j \cap S| \text{ for } 1 \leq j \leq k_1 \end{cases} \tag{5.37}$$

Thus, $v(S)$ is different from $v'(S)$ if and only if $\exists j \in k_1, |T_j \cap S| > m^{2\delta}$. But, we have:

$$\begin{aligned}
&\mathbb{P}\left(\exists j \in k_1, |T_j \cap S| > m^{2\delta}\right) \leq \sum_{j \in [k_1]} \mathbb{P}\left(|T_j \cap S| > m^{2\delta}\right) \text{ by union bound} \\
&= \sum_{j \in [k_1]} \mathbb{P}\left(\sum_{i \in S} X_j^i > m^{2\delta}\right) = \sum_{j \in [k_1]} \mathbb{P}\left(\frac{1}{|S|}\sum_{i \in S} X_j^i > \left(1 + \frac{\frac{m^{2\delta}}{|S|} - 1}{p}\right) p\right) \\
&\leq \sum_{j \in [k_1]} \mathbb{P}\left(\frac{1}{|S|}\sum_{i \in S} X_j^i > \left(1 + \frac{\frac{m^{2\delta}}{|S|}}{p}\right) p\right) \\
&\leq \sum_{j \in [k_1]} \exp\left(-\frac{p|S|}{4}\left(\frac{m^{2\delta}}{p|S|}\right)^2\right) \text{ by Chernoff-Hoeffding bound as depicted in Proposition 5.15} \\
&= \sum_{j \in [k_1]} \exp\left(-\frac{1}{4p|S|}m^{4\delta}\right) \leq \sum_{j \in [k_1]} \exp\left(-\frac{m^{-\delta}}{4}m^{4\delta}\right) \text{ since } \frac{1}{p|S|} = \frac{k_1}{|S|} \geq m^{-\delta} \\
&= m^{\frac{1}{2}} e^{-\frac{m^{3\delta}}{4}} .
\end{aligned} \tag{5.38}$$

Thus, this event occurs with exponentially small probability (on the choice of the partition $T_1, \ldots, T_{k_1}$).

Let us now study the case of S of size $|S| > m^{\frac{1}{2}+\delta}$. Then, we have that $\frac{1}{m^{\frac{1}{2}-\delta}}|S| > m^{2\delta}$, so we get that $v'(S) = \frac{1}{m^{\frac{1}{2}-\delta}}|S|$. On the other hand, we have that:

$$v(S) = \max \begin{cases} \frac{1}{m^{\frac{1}{2}-\delta}}|S| \\ |T_j \cap S| \text{ for } 1 \leq j \leq k_1 \end{cases} \tag{5.39}$$

Thus, $v(S)$ is different from $v'(S)$ if and only if $\exists j \in [k_1], |T_j \cap S| > \frac{1}{m^{\frac{1}{2}-\delta}}|S|$. But, we have:

$$\begin{aligned}
&\mathbb{P}\left(\exists j \in [k_1], |T_j \cap S| > \frac{1}{m^{\frac{1}{2}-\delta}}|S|\right) \leq \sum_{j\in[k_1]} \mathbb{P}\left(|T_j \cap S| > \frac{1}{m^{\frac{1}{2}-\delta}}|S|\right) \text{ by union bound}\\
&= \sum_{j\in[k_1]} \mathbb{P}\left(\sum_{i\in S} X_j^i > \frac{1}{m^{\frac{1}{2}-\delta}}|S|\right) = \sum_{j\in[k_1]} \mathbb{P}\left(\frac{1}{|S|}\sum_{i\in S} X_j^i > \left(1 + \frac{\frac{1}{|S|m^{\frac{1}{2}-\delta}}|S| - 1}{p}\right)p\right)\\
&\leq \sum_{j\in[k_1]} \mathbb{P}\left(\frac{1}{|S|}\sum_{i\in S} X_j^i > \left(1 + \frac{\frac{1}{m^{\frac{1}{2}-\delta}}}{p}\right)p\right)\\
&= \sum_{j\in[k_1]} \mathbb{P}\left(\frac{1}{|S|}\sum_{i\in S} X_j^i > \left(1 + m^{\delta}\right)p\right) \text{ since } p = \frac{1}{k_1} = \frac{1}{m^{\frac{1}{2}}}\\
&\leq \sum_{j\in[k_1]} \exp\left(-\frac{p|S|}{4}m^{2\delta}\right) \text{ by Chernoff-Hoeffding bound as depicted in Proposition 5.15}\\
&\leq \sum_{j\in[k_1]} \exp\left(-\frac{m^{\delta}}{4}m^{2\delta}\right) \text{ since } p|S| = \frac{|S|}{k_1} \geq m^{\delta}\\
&= m^{\frac{1}{2}}e^{-\frac{m^{3\delta}}{4}}\ .
\end{aligned} \tag{5.40}$$

Thus, this event occurs with exponentially small probability as well. We have then that for all set S, $\mathbb{P}\left(v(S) \neq v'(S)\right) \leq p_{\text{leak}} := m^{\frac{1}{2}}e^{-\frac{m^{3\delta}}{4}}$, which is an exponentially small bound that does not depend on S.

Hence, for every set S, only with exponentially small probability p_{leak} can one distinguish between v and v'. For some fixed algorithm $\mathcal{A}$, let us consider the sequence L of queries made by $\mathcal{A}$ before it is able to distinguish between v and v': $L := (S_1, \dots, S_n)$, with $v(S_i) = v'(S_i)$ for $i \in [n]$ and $v'(S_{n+1}) > v(S_{n+1})$. L is independent from $T_1, \dots T_{k_1}$ as no information from this partition is leaked before S_{n+1}. Thus, for such an algorithm to be correct, it should work for any equi-partition $T_1, \dots T_{k_1}$. We have:

$$\mathbb{P}\left(\exists i \in [n] : v(S_i) \neq v'(S_i)\right) \leq \sum_{i=1}^{n} \mathbb{P}\left(v(S_i) \neq v'(S_i)\right) = np_{\text{leak}} \text{ by union bound.}$$

In particular, this implies that:

$$\mathbb{P}\left(\forall i \in [n] : v(S_i) = v'(S_i)\right) \geq 1 - np_{\text{leak}}\ .$$

So, if $1 - np_{\text{leak}} > 0$, i.e. $n < \frac{1}{p_{\text{leak}}}$, then there exists some equi-partition $T_1, \dots T_{k_1}$ such that our algorithm outputs a sequence L of queries of length n before being able to distinguish between v and v'. In particular, we can take $n = \frac{1}{2p_{\text{leak}}}$ so that L is of exponential size. Hence, for any algorithm $\mathcal{A}$, there exists some equi-partition $T_1, \dots T_{k_1}$ such that $\mathcal{A}$ needs an exponential number of value queries to distinguish between v and v'. This concludes the proof of the theorem for any deterministic algorithm.

Finally, the hardness result holds also for randomized algorithms. Indeed, let us call $\mathcal{A}_s$, the running algorithm conditioned on its random bits being s. $\mathcal{A}_s$ is deterministic so the previous proof holds: with high probability p, the sequence of $\lfloor \frac{1-p}{p_{leak}} \rfloor$ queries does not reveal anything to distinguish between v and v', although it is of exponential size in m. Then, averaging over all its random bitstrings, the same result holds, as p_{leak} is independent from the equi-partition $T_1, \ldots, T_{k_1}$. □

5.4.3 Limitations of the Model

The main weakness of the previous result is that it highly relies on the restriction that one has access to the data only through value queries. Indeed, if one has access to the full data, it is possible to read the partition $T_1, \ldots, T_{k_1}$ which gives the optimal solution directly. This weakness comes from the fact that our utility function f_W^1 can be described by polynomial-size data, as it is characterized by a broadcast channel W, whereas one usually consider any XOS functions, in particular with some having inherently an exponential-size defining set of linear valuation functions.

On the other hand, one can also remark that when f_W^1 is written as a maximum of linear valuation functions, then that defining set of linear valuation functions is of exponential size. Hence, we think that it is still relevant to study the value query complexity of such a family of functions, as it is not clear how one could recover the partition in polynomial time from this exponential-size set of linear valuations without any additional information.

5.5 Conclusion

In this chapter, we have studied several algorithmic aspects and non-signaling assisted capacity regions of broadcast channels. We have shown that when non-signaling assistance is shared only between the decoders, the capacity region does not change. For the class of deterministic broadcast channels, we have described a $(1-e^{-1})^2$-approximation algorithm running in polynomial time, and we have shown that the capacity region for that class is the same with or without non-signaling assistance. Finally, we have shown that in the value query model, we cannot achieve a better approximation ratio than $\Omega\left(\frac{1}{\sqrt{m}}\right)$ in polynomial time for the general broadcast channel coding problem, with m the size of one of the outputs of the channel.

Our results suggest that non-signaling assistance could improve the capacity region of general broadcast channels, which is left as a major open question. An intermediate result would be to show that it is NP-hard to approximate the broadcast channel coding problem within any constant, strengthening our hardness result without relying on the value query model. Finally, one could also try to develop approximations algorithms for other subclasses of broadcast channels, such as semi-deterministic or degraded ones. This could be a crucial step towards showing that the capacity region for those classes is the same with or without non-signaling assistance.

Conclusion

TODO

Bibliography

[AA00] Erling D Andersen and Knud D Andersen. The Mosek interior point optimizer for linear programming: an implementation of the homogeneous algorithm. In *High performance optimization*, pages 197–232. Springer, 2000. doi:10.1007/978-1-4757-3216-0_8. Citations: § 69

[AAM+11] Noga Alon, Sanjeev Arora, Rajsekar Manokaran, Dana Moshkovitz, and Omri Weinstein. Inapproximability of densest κ-subgraph from average case hardness. *Unpublished manuscript*, 1, 2011. URL: https://www.tau.ac.il/~nogaa/PDFS/dks8.pdf. Citations: § 98

[AB99] Rudolf Ahlswede and Vladimir B. Balakirsky. Construction of uniquely decodable codes for the two-user binary adder channel. *IEEE Trans. Inf. Theory*, 45(1):326–330, 1999. doi:10.1109/18.746834. Citations: § 69

[Ahl73] Rudolf Ahlswede. Multi-way communication channels. In *2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971*, pages 23–52. Publishing House of the Hungarian Academy of Science, 1973. Citations: § 43, 47, and 97

[AK75] Rudolf Ahlswede and János Körner. Source coding with side information and a converse for degraded broadcast channels. *IEEE Trans. Inf. Theory*, 21(6):629–637, 1975. doi:10.1109/TIT.1975.1055469. Citations: § 97

[Ara02] PK Aravind. A simple demonstration of Bell's theorem involving two observers and no probabilities or inequalities. *CoRR*, abs/quant-ph/0206070, 2002. arXiv:quant-ph/0206070. Citations: § 43

[AS04] Alexander A. Ageev and Maxim Sviridenko. Pipage rounding: A new method of constructing algorithms with proven performance guarantee. *J. Comb. Optim.*, 8(3):307–328, 2004. doi:10.1023/B:JOCO.0000038913.96607.c2. Citations: § 9 and 10

[BB98] Shraga I. Bross and Ian F. Blake. Upper bound for uniquely decodable codes in a binary input N-user adder channel. *IEEE Trans. Inf. Theory*, 44(1):334–340, 1998. doi:10.1109/18.651062. Citations: § 69

[BBT05] Gilles Brassard, Anne Broadbent, and Alain Tapp. Quantum pseudo-telepathy. *Foundations of Physics*, 35(11):1877–1907, 2005. doi:10.1007/s10701-005-7353-4. Citations: § 43

[BCE+16] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016. doi:10.1017/CBO9781107446984. Citations: § 18

[BCP$^+$14] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell nonlocality. *Rev. Mod. Phys.*, 86:419–478, Apr 2014. doi:10.1103/RevModPhys.86.419. Citations: § 43 and 97

[Ber73] Patrick P. Bergmans. Random coding theorem for broadcast channels with degraded components. *IEEE Trans. Inf. Theory*, 19(2):197–207, 1973. doi:10.1109/TIT.1973.1054980. Citations: § 97

[BF18] Siddharth Barman and Omar Fawzi. Algorithmic aspects of optimal channel coding. *IEEE Trans. Inf. Theory*, 64(2):1038–1045, 2018. doi:10.1109/TIT.2017.2696963. Citations: § 5, 6, 43, 46, 90, 91, 97, 98, and 101

[BFF21] Siddharth Barman, Omar Fawzi, and Paul Fermé. Tight approximation guarantees for concave coverage problems. In Markus Bläser and Benjamin Monmege, editors, *38th International Symposium on Theoretical Aspects of Computer Science, STACS 2021, March 16-19, 2021, Saarbrücken, Germany (Virtual Conference)*, volume 187 of *LIPIcs*, pages 9:1–9:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.STACS.2021.9. Citations: § 119

[BFGG20] Siddharth Barman, Omar Fawzi, Suprovat Ghoshal, and Emirhan Gürpinar. Tight approximation bounds for maximum multi-coverage. In Daniel Bienstock and Giacomo Zambelli, editors, *Integer Programming and Combinatorial Optimization - 21st International Conference, IPCO 2020, London, UK, June 8-10, 2020, Proceedings*, volume 12125 of *Lecture Notes in Computer Science*, pages 66–77. Springer, 2020. doi:10.1007/978-3-030-45771-6_6. Citations: § 6, 8, 9, 11, 13, 24, and 98

[BH84] Andrew D Barbour and Peter Hall. On the rate of Poisson convergence. *Mathematical Proceedings of the Cambridge Philosophical Society*, 95(3):473–480, 1984. doi:10.1017/S0305004100061806. Citations: § 32

[BLS17] Markus Brill, Jean-François Laslier, and Piotr Skowron. Multiwinner approval rules as apportionment methods. In Satinder P. Singh and Shaul Markovitch, editors, *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, pages 414–420. AAAI Press, 2017. doi:10.1177/0951629818775518. Citations: § 7

[BN05] Liad Blumrosen and Noam Nisan. On the computational power of iterative auctions. In John Riedl, Michael J. Kearns, and Michael K. Reiter, editors, *Proceedings 6th ACM Conference on Electronic Commerce (EC-2005), Vancouver, BC, Canada, June 5-8, 2005*, pages 29–43. ACM, 2005. doi:10.1145/1064009.1064013. Citations: § 99

[BSST99] Charles H Bennett, Peter W Shor, John A Smolin, and Ashish V Thapliyal. Entanglement-assisted classical capacity of noisy quantum channels. *Physical Review Letters*, 83(15):3081, 1999. doi:10.1103/PhysRevLett.83.3081. Citations: § 43 and 97

[BT96] Steven J. Brams and Alan D. Taylor. *Fair division - from cake-cutting to dispute resolution.* Cambridge University Press, 1996. Citations: § 99

[Can17] Clément L Canonne. A short note on Poisson tail bounds. *Retrieved from the website:* http://www.cs.columbia.edu/~ccanonne, 2017. Citations: § 26

[CC84] Michele Conforti and Gérard Cornuéjols. Submodular set functions, matroids and the greedy algorithm: Tight worst-case bounds and some generalizations of the Rado-Edmonds theorem. *Discret. Appl. Math.*, 7(3):251–274, 1984. doi:10.1016/0166-218X(84)90003-9. Citations: § 8

[CCK$^+$17] Parinya Chalermsook, Marek Cygan, Guy Kortsarz, Bundit Laekhanukit, Pasin Manurangsi, Danupon Nanongkai, and Luca Trevisan. From Gap-ETH to FPT-inapproximability: Clique, dominating set, and more. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berke-*

ley, CA, USA, October 15-17, 2017, pages 743–754. IEEE Computer Society, 2017. doi:10.1109/FOCS.2017.74. Citations: § 8 and 18

[CCPV11] Gruia Călinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM J. Comput.*, 40(6):1740–1766, 2011. doi:10.1137/080733991. Citations: § 10

[CFN77] Gérard Cornuéjols, Marshall L Fisher, and George L Nemhauser. Exceptional paper—location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. *Management science*, 23(8):789–810, 1977. doi:10.1287/mnsc.23.8.789. Citations: § 5 and 7

[Cov72] Thomas M. Cover. Broadcast channels. *IEEE Trans. Inf. Theory*, 18(1):2–14, 1972. doi:10.1109/TIT.1972.1054727. Citations: § 97

[Cov75] Thomas M. Cover. An achievable rate region for the broadcast channel. *IEEE Transactions on Information Theory*, 21(4):399–404, 1975. doi:10.1109/TIT.1975.1055418. Citations: § 97

[CPM19] Rahul Chandan, Dario Paccagnan, and Jason R. Marden. Optimal mechanisms for distributed resource-allocation. *CoRR*, abs/1911.07823, 2019. arXiv:1911.07823. Citations: § 7 and 22

[CT01] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 2001. doi:10.1002/0471200611. Citations: § 18 and 46

[DMMS20] Szymon Dudycz, Pasin Manurangsi, Jan Marcinkowski, and Krzysztof Sornat. Tight approximation for proportional approval voting. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 276–282. ijcai.org, 2020. doi:10.24963/ijcai.2020/39. Citations: § 7, 8, 9, 20, 24, 34, 38, 39, and 98

[DR98] Devdatt P. Dubhashi and Desh Ranjan. Balls and bins: A study in negative dependence. *Random Struct. Algorithms*, 13(2):99–124, 1998. doi:10.7146/brics.v3i25.20006. Citations: § 40 and 123

[DS06] Shahar Dobzinski and Michael Schapira. An improved approximation algorithm for combinatorial auctions with submodular bidders. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2006, Miami, Florida, USA, January 22-26, 2006*, pages 1064–1073. ACM Press, 2006. URL: http://dl.acm.org/citation.cfm?id=1109557.1109675. Citations: § 99, 122, and 124

[DV15] Shaddin Dughmi and Jan Vondrák. Limitations of randomized mechanisms for combinatorial auctions. *Games Econ. Behav.*, 92:370–400, 2015. doi:10.1016/j.geb.2014.01.007. Citations: § 5

[Fei98] Uriel Feige. A threshold of ln n for approximating set cover. *J. ACM*, 45(4):634–652, 1998. doi:10.1145/285055.285059. Citations: § 5, 9, and 13

[FF22] Omar Fawzi and Paul Fermé. Beating the sum-rate capacity of the binary adder channel with non-signaling correlations. In *IEEE International Symposium on Information Theory, ISIT 2022, Espoo, Finland, June 26 - July 1, 2022*, pages 2750–2755. IEEE, 2022. doi:10.1109/ISIT50566.2022.9834699. Citations: § 97

[FK14] Vitaly Feldman and Pravesh Kothari. Learning coverage functions and private release of marginals. In Maria-Florina Balcan, Vitaly Feldman, and Csaba Szepesvári, editors, *Proceedings of The 27th Conference on Learning Theory, COLT 2014, Barcelona, Spain, June 13-15, 2014*, volume 35 of *JMLR Workshop and Conference Proceedings*, pages 679–702. JMLR.org, 2014. URL: http://proceedings.mlr.press/v35/feldman14a.html. Citations: § 5

[FKP01] Uriel Feige, Guy Kortsarz, and David Peleg. The dense k-subgraph problem. *Algorithmica*, 29(3):410–421, 2001. doi:10.1007/s004530010050. Citations: § 98

[FW10] Manuel Fernández and Stuart Williams. Closed-form expression for the Poisson-binomial probability density function. *IEEE Transactions on Aerospace and Electronic Systems*, 46(2):803–817, 2010. doi:10.1109/TAES.2010.5461658. Citations: § 32

[Gal74] R.G. Gallager. Capacity and coding for degraded broadcast channels. *Problemy Peredachi Informatsii*, 10(3):3–14, 1974. Citations: § 97

[GJ79] M. R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979. Citations: § 114

[GK11] Abbas El Gamal and Young-Han Kim. *Network Information Theory*. Cambridge University Press, 2011. doi:10.1017/CBO9781139030687. Citations: § 85 and 86

[GM07] Bernd Gärtner and Jirí Matousek. *Understanding and using linear programming*. Universitext. Springer, 2007. doi:10.1007/978-3-540-30717-4. Citations: § 50, 64, and 108

[GN20] Amin Gohari and Chandra Nair. New outer bounds for the two-receiver broadcast channel. In *IEEE International Symposium on Information Theory, ISIT 2020, Los Angeles, CA, USA, June 21-26, 2020*, pages 1492–1497. IEEE, 2020. doi:10.1109/ISIT44484.2020.9174003. Citations: § 97

[GP80] Sergei Izrail'evich Gel'fand and Mark Semenovich Pinsker. Capacity of a broadcast channel with one deterministic component. *Problemy Peredachi Informatsii*, 16(1):24–34, 1980. Citations: § 97

[HDW08] Min-Hsiu Hsieh, Igor Devetak, and Andreas J. Winter. Entanglement-assisted capacity of quantum multiple-access channels. *IEEE Trans. Inf. Theory*, 54(7):3078–3090, 2008. doi:10.1109/TIT.2008.924726. Citations: § 44

[Hoc97] Dorit S. Hochbaum. Approximation algorithms for NP-hard problems. *SIGACT News*, 28(2):40–52, 1997. doi:10.1145/261342.571216. Citations: § 5

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi:10.1080/01621459.1963.10500830. Citations: § 123

[JDP83] Kumar Joag-Dev and Frank Proschan. Negative association of random variables with applications. *The Annals of Statistics*, pages 286–295, 1983. doi:10.1214/aos/1176346079. Citations: § 39, 40, and 125

[JNV+20] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. MIP*=RE. *CoRR*, abs/2001.04383, 2020. URL: https://arxiv.org/abs/2001.04383, arXiv:2001.04383. Citations: § 44

[Jr.78] E. J. Weldon Jr. Coding for a multiple-access channel. *Inf. Control.*, 36(3):256–274, 1978. doi:10.1016/S0019-9958(78)90312-1. Citations: § 69

[KL78] Tadao Kasami and Shu Lin. Bounds on the achievable rates of block coding for a memoryless multiple-access channel. *IEEE Trans. Inf. Theory*, 24(2):187–197, 1978. doi:10.1109/TIT.1978.1055860. Citations: § 69

[KLWY83] Tadao Kasami, Shu Lin, Victor K.-W. Wei, and Saburo Yamamura. Graph theoretic approaches to the code construction for the two-user multiple-access binary adder channel. *IEEE Trans. Inf. Theory*, 29(1):114–130, 1983. doi:10.1109/TIT.1983.1056614. Citations: § 69, 70, and 139

[LALS20] Felix Leditzky, Mohammad A Alhejji, Joshua Levin, and Graeme Smith. Playing games with multiple access channels. *Nature communications*, 11(1):1–5, 2020. doi:10.1038/s41467-020-15240-w. Citations: § 43 and 97

[Lia73] Henry Herng-Jiunn Liao. Multiple access channels (Ph.D. Thesis abstr.). *IEEE Trans. Inf. Theory*, 19(2):253, 1973. doi:10.1109/TIT.1973.1054960. Citations: § 43, 47, and 97

[Lin69] Bernt Lindström. Determination of two vectors from the sum. *Journal of Combinatorial Theory*, 6(4):402–407, 1969. doi:10.1016/S0021-9800(69)80038-4. Citations: § 69

[Man20] Pasin Manurangsi. Tight running time lower bounds for strong inapproximability of maximum k-coverage, unique set cover and related problems (via t-wise agreement testing theorem). In Shuchi Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 62–81. SIAM, 2020. doi:10.1137/1.9781611975994.5. Citations: § 8 and 18

[Mar77] Katalin Marton. The capacity region of deterministic broadcast channels. In *Trans. Int. Symp. Inform. Theory*, pages 243–248, 1977. Citations: § 97

[Mar79] Katalin Marton. A coding theorem for the discrete memoryless broadcast channel. *IEEE Trans. Inf. Theory*, 25(3):306–311, 1979. doi:10.1109/TIT.1979.1056046. Citations: § 97

[Mat12] William Matthews. A linear program for the finite block length converse of Polyanskiy-Poor-Verdú via nonsignaling codes. *IEEE Trans. Inf. Theory*, 58(12):7036–7044, 2012. doi:10.1109/TIT.2012.2210695. Citations: § 43, 78, 79, and 97

[Mer90] N David Mermin. Simple unified form for the major no-hidden-variables theorems. *Physical review letters*, 65(27):3373, 1990. doi:10.1103/PhysRevLett.65.3373. Citations: § 43

[MÖ05] M. Mattas and Patric R. J. Östergård. A new bound for the zero-error capacity region of the two-user binary adder channel. *IEEE Trans. Inf. Theory*, 51(9):3289–3291, 2005. doi:10.1109/TIT.2005.853309. Citations: § 69, 70, and 139

[Mou03] Hervé Moulin. *Fair division and collective welfare.* MIT Press, 2003. Citations: § 99

[MSV08] Vahab S. Mirrokni, Michael Schapira, and Jan Vondrák. Tight information-theoretic lower bounds for welfare maximization in combinatorial auctions. In Lance Fortnow, John Riedl, and Tuomas Sandholm, editors, *Proceedings 9th ACM Conference on Electronic Commerce (EC-2008), Chicago, IL, USA, June 8-12, 2008*, pages 70–77. ACM, 2008. doi:10.1145/1386790.1386805. Citations: § 99, 122, and 124

[Mur00] Robert A Murphey. Target-based weapon target assignment problems. In *Nonlinear assignment problems*, pages 39–53. Springer, 2000. doi:10.1007/978-1-4757-3155-2_3. Citations: § 7 and 21

[MW08] Jason R. Marden and Adam Wierman. Distributed welfare games with applications to sensor coverage. In *Proceedings of the 47th IEEE Conference on Decision and Control, CDC 2008, December 9-11, 2008, Cancún, Mexico*, pages 1708–1713. IEEE, 2008. doi:10.1109/CDC.2008.4738800. Citations: § 7

[ND20] Janis Nötzel and Stephen Diadamo. Entanglement-enabled communication for the internet of things. In Mohammad S. Obaidat, Kuei-Fang Hsiao, Petros Nicopolitidis, and Daniel Cascado Caballero, editors, *International Conference on Computer, Information and Telecommunication Systems, CITS 2020, Hangzhou, China, October 5-7, 2020*, pages 1–6. IEEE, 2020. doi:10.1109/CITS49457.2020.9232550. Citations: § 43

[NG07] Chandra Nair and Abbas El Gamal. An outer bound to the capacity region of the broadcast channel. *IEEE Trans. Inf. Theory*, 53(1):350–355, 2007. doi:10.1109/TIT.2006.887492. Citations: § 97

[Noe20] Janis Noetzel. Entanglement-enabled communication. *IEEE J. Sel. Areas Inf. Theory*, 1(2):401–415, 2020. doi:10.1109/jsait.2020.3017121. Citations: § 43

[NRTV07] Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani, editors. *Algorithmic Game Theory.* Cambridge University Press, 2007. doi:10.1017/CBO9780511800481. Citations: § 18 and 22

[OS15] Or Ordentlich and Ofer Shayevitz. A VC-dimension-based outer bound on the zero-error capacity of the binary adder channel. In *IEEE International Symposium on Information Theory, ISIT 2015, Hong Kong, China, June 14-19, 2015*, pages 2366–2370. IEEE, 2015. doi:10.1109/ISIT.2015.7282879. Citations: § 69

[PDB21] Uzi Pereg, Christian Deppe, and Holger Boche. Quantum broadcast channels with cooperating decoders: An information-theoretic perspective on quantum repeaters. In *IEEE International Symposium on Information Theory, ISIT 2021, Melbourne, Australia, July 12-20, 2021*, pages 772–777. IEEE, 2021. doi:10.1109/ISIT45174.2021.9518284. Citations: § 98 and 105

[PDB23] Uzi Pereg, Christian Deppe, and Holger Boche. The multiple-access channel with entangled transmitters. *CoRR*, abs/2303.10456, 2023. arXiv:2303.10456, doi:10.48550/arXiv.2303.10456. Citations: § 44 and 98

[Per90] Asher Peres. Incompatible results of quantum measurements. *Physics Letters A*, 151(3-4):107–108, 1990. doi:10.1016/0375-9601(90)90172-K. Citations: § 43

[Pin78] Mark Semenovich Pinsker. Capacity of noiseless broadcast channel. *Problemy Peredachi Informatsii*, 14(2):28–34, 1978. Citations: § 97

[PLM+11] R. Prevedel, Y. Lu, W. Matthews, R. Kaltenbaek, and K. J. Resch. Entanglement-enhanced classical communication over a noisy classical channel. *Phys. Rev. Lett.*, 106:110505, Mar 2011. doi:10.1103/PhysRevLett.106.110505. Citations: § 43 and 97

[PM18] Dario Paccagnan and Jason R Marden. Utility design for distributed resource allocation–part II: Applications to submodular, covering, and supermodular problems. *CoRR*, abs/1807.01343, 2018. arXiv:1807.01343. Citations: § 7, 20, 21, 22, 23, and 139

[PPV10] Yury Polyanskiy, H. Vincent Poor, and Sergio Verdú. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory*, 56(5):2307–2359, 2010. doi:10.1109/TIT.2010.2043769. Citations: § 79 and 81

[QS17] Yihui Quek and Peter W. Shor. Quantum and superquantum enhancements to two-sender, two-receiver channels. *Phys. Rev. A*, 95:052329, May 2017. doi:10.1103/PhysRevA.95.052329. Citations: § 43 and 97

[Sat78] Hiroshi Sato. An outer bound to the capacity region of broadcast channels (corresp.). *IEEE Trans. Inf. Theory*, 24(3):374–377, 1978. doi:10.1109/TIT.1978.1055883. Citations: § 97

[SFL16] Piotr Skowron, Piotr Faliszewski, and Jérôme Lang. Finding a collective set of items: From proportional multirepresentation to group recommendation. *Artif. Intell.*, 241:191–216, 2016. doi:10.1016/j.artint.2016.09.003. Citations: § 7

[Sha48] Claude E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3):379–423, 1948. doi:10.1002/j.1538-7305.1948.tb01338.x. Citations: § 97

[SLSS22] Akshay Seshadri, Felix Leditzky, Vikesh Siddhu, and Graeme Smith. On the separation of correlation-assisted sum capacities of multiple access channels. In *IEEE International Symposium on Information Theory, ISIT 2022, Espoo, Finland, June 26 - July 1, 2022*, pages 2756–2761. IEEE, 2022. doi:10.1109/ISIT50566.2022.9834459. Citations: § 43 and 97

[SS07] Moshe Shaked and J George Shanthikumar. *Stochastic orders*. Springer Science & Business Media, 2007. doi:10.1007/978-0-387-34675-5. Citations: § 11

[SVW17] Maxim Sviridenko, Jan Vondrák, and Justin Ward. Optimal approximation for submodular and supermodular optimization with bounded curvature. *Math. Oper. Res.*, 42(4):1197–1218, 2017. doi:10.1287/moor.2016.0842. Citations: § 8, 22, 23, 24, 30, and 139

[TT19] Wenpin Tang and Fengmin Tang. The Poisson binomial distribution–old & new. *CoRR*, abs/1908.10024, 2019. arXiv:1908.10024. Citations: § 32

[UL98] Rüdiger Urbanke and Quinn Li. The zero-error capacity region of the 2-user synchronous BAC is strictly smaller than its Shannon capacity region. In *1998 Information*

Theory Workshop (Cat. No. 98EX131), page 61. IEEE, 1998. doi:10.1109/ITW.1998.706434. Citations: § 69

[vdBvT85] P. A. B. M. Coebergh van den Braak and Henk C. A. van Tilborg. A family of good uniquely decodable code pairs for the two-access binary adder channel. *IEEE Trans. Inf. Theory*, 31(1):3–9, 1985. doi:10.1109/TIT.1985.1057004. Citations: § 69, 70, and 139

[vdM75] Edward C. van der Meulen. Random coding theorems for the general discrete memoryless broadcast channel. *IEEE Trans. Inf. Theory*, 21(2):180–190, 1975. doi:10.1109/TIT.1975.1055347. Citations: § 97

[Von07] Jan Vondrák. Submodularity in Combinatorial Optimization. *Univerzita Karlova, Matematicko-Fyzikální Fakulta*, 2007. URL: https://dspace.cuni.cz/bitstream/handle/20.500.11956/13738/140038775.pdf. Citations: § 10, 12, and 22

[Von08] Jan Vondrák. Optimal approximation for the submodular welfare problem in the value oracle model. In Cynthia Dwork, editor, *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 67–74. ACM, 2008. doi:10.1145/1374376.1374389. Citations: § 113 and 116

[vT78] Henk C. A. van Tilborg. An upper bound for codes in a two-access binary erasure channel (Corresp.). *IEEE Trans. Inf. Theory*, 24(1):112–116, 1978. doi:10.1109/TIT.1978.1055814. Citations: § 69

[Wil90] F.M.J. Willems. The maximal-error and average-error capacity region of the broadcast channel are identical : A direct proof. *Problems of Control and Information Theory*, 19(4):339–347, December 1990. URL: https://research.tue.nl/nl/publications/the-maximal-error-and-average-error-capacity-region-of-the-broadc. Citations: § 102

List of Figures

List of Tables