Zero-error network information theory : graphs, coding for computing and source-channel duality

Nicolas Charpenay

COLLEGE MATHS, TELECOMS
DOCTORAL INFORMATIQUE, SIGNAL
BRETAGNE SYSTEMES, ELECTRONIQUE

Université de Rennes

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE RENNES

ÉCOLE DOCTORALE N° 601
Mathématiques, Télécommunications, Informatique, Signal, Systèmes, Électronique
Spécialité : *Signal, Image, Vision*

Par
Nicolas CHARPENAY

Théorie de l'information zéro-erreur dans les réseaux : graphes, codage pour le calcul de fonctions et dualité source-canal

Thèse présentée et soutenue à Rennes, le mardi 28 novembre 2023
Unité de recherche : IRISA UMR 6074 et Centre Inria de l'Université de Rennes

Rapporteurs avant soutenance :

Mikael SKOGLUND Professeur à KTH
Michèle WIGGER Professeure à Telecom Paris

Composition du Jury :

Président :	Jean-Marie GORCE	Professeur des Universités à INSA Lyon *(à préciser après la soutenance)*
Rapporteurs :	Mikael SKOGLUND	Professeur à KTH
	Michèle WIGGER	Professeure à Telecom Paris
Examinatrice :	Derya MALAK	Maître de Conférences à EURECOM
Dir. de thèse :	Aline ROUMY	Directrice de Recherche Inria
Co-dir. de thèse :	Maël LE TREUST	Chargé de Recherche CNRS
Invité :	Tristan TOMALA	Professeur à HEC Paris

Table of Contents

Chapter 1

Résumé

Le codage de source zéro-erreur regroupe une grande variété de problèmes en théorie de l'information, où une source doit être compressée par un encodeur avec un débit à minimiser, puis doit être retrouvée par le décodeur avec probabilité d'erreur zéro. Dans tous les problèmes étudiés ici, la distribution de source est connue et le décodeur a accès à une information adjacente, i.e. une variable aléatoire corrélée à la source qui aide au décodage et à minimiser le débit. L'étude de tels problèmes trouve son utilité dans des applications demandant une garantie forte sur la qualité de reconstruction de la source, en particulier la télédétection et le calcul distribué. Dans un problème de codage de source zéro-erreur, trouver une formule pour le débit optimal permet d'établir une borne théorique sur les performances de tous les schémas de codage possibles. De plus, si cette formule est à une seule lettre (i.e. ne fait pas apparaître de limite), elle peut être facilement calculée. On a aussi l'existence d'un schéma de codage qui permet d'atteindre asymptotiquement ce débit optimal. Pour ces raisons, notre premier objectif dans ce manuscrit sera de trouver des expressions à une seule lettre pour les débits optimaux des problèmes de codage de source zéro-erreur.

La contrainte zéro-erreur diffère de la contrainte d'erreur asymptotiquement nulle, cette dernière ne demandant qu'une limite nulle pour la probabilité d'erreur au décodage quand la longueur du bloc codant tend vers l'infini. Dans les cadres où le décodeur n'a pas d'information adjacente, on peut facilement montrer que les débits optimaux sous les contraintes zéro-erreur et erreur asymptotiquement nulle coïncident. Toutefois, lorsqu'on impose la contrainte zéro-erreur quand le décodeur a une information adjacente – ce sera le cas dans les problèmes étudiés ici, la nature de certains problèmes passe de statistique à combinatoire (e.g. le problème de Slepian-Wolf zéro-erreur). Il s'ensuit que les outils adéquats pour de tels problèmes sont à la fois tirés de la théorie de l'information et de la combinatoire ; et peuvent aller d'un dispositif de correction qui donne la propriété zéro-erreur à des codes existants, à des codes zéro-erreur entièrement bâtis avec des contraintes tirées de la théorie des graphes.

La nature combinatoire des problèmes zéro-erreur est liée à l'incertitude de l'encodeur sur l'information adjacente du décodeur. Dans les problèmes à erreur asymptotiquement nulle, on peut supposer que la suite des réalisations de l'information adjacente du décodeur sera typique par rapport à la distribution de source, avec grande probabilité ; même si l'encodeur n'a pas accès à ces réalisations. En revanche, la contrainte zéro-erreur rend ce manque d'information plus difficile à gérer ; en particulier, il faut que les codes utilisés permettent une reconstruction parfaite de la source pour *n'importe quelle* réalisation de l'information adjacente du décodeur.

Par conséquent, nous classifions les problèmes zéro-erreur étudiés ici en trois catégories ; selon l'information que possède l'encodeur sur l'information adjacente du décodeur, et selon ce que le décodeur lui-même doit retrouver :

- Les problèmes zéro-erreur *basés sur la correction*, dans lesquels l'encodeur a accès à toutes les informations adjacentes du problème, et peut simuler chaque décodage ;
- Les problèmes zéro-erreur *basés sur les graphes*, dans lesquels l'encodeur est "moins informé" sur l'information adjacente du décodeur ;
- Les problèmes zéro-erreur *pour le calcul de fonction*, dans lesquels le décodeur doit calculer une fonction de la source et de son information adjacente ; au lieu de simplement reconstruire la source.

1.1 Problèmes zéro-erreur basés sur la correction

Le problèmes zéro-erreur *basés sur la correction* peuvent être abordés avec l'aide d'un dispositif de correction qui garantit la propriété zéro-erreur. Il consiste en l'utilisation d'un livre-code adapté à la contrainte d'erreur asymptotiquement nulle, la simulation de chaque décodage, et l'envoi de toute les suite de source s'il y a erreur dans l'un des décodages. Ce dispositif de correction a un impact négligeable sur le débit en raison de la faible probabilité d'erreur, et permet la conversion des codes à erreur asymptotiquement nulle en codes zéro-erreur. Pour cette raison, les outils que nous utilisons avec le dispositif de correction sont la typicalité, et les codes aléatoires et de parité.

Dans le premier problème de cette classe, présenté en Figure 4.1, l'information adjacente peut être présente au décodeur. Dans ce cadre, l'encodeur dispose d'une information adjacente, mais ignore si le décodeur y a accès aussi ; l'encodage est divisé en deux étapes. À la première étape, l'encodeur envoie de l'information sans savoir si l'information adjacente est présente au décodeur. Le message envoyé à la première étape doit être exploitable

indépendamment de l'éventuelle présence de l'information adjacente. De plus, un décodage zéro-erreur doit être réalisé si le décodeur a accès à l'information adjacente. À la deuxième étape, un supplément d'information est envoyé par l'encodeur dans le cas où l'information adjacente n'est pas présente au décodeur, et ce dernier doit ensuite retrouver la source avec zéro-erreur. Ces deux étapes sont modélisées par deux décodeurs différents et deux canaux parfaits. Nous voulons répondre à la question suivante : quelles sont les paires de débits réalisables pour les deux étapes d'encodage, qui permettent un décodage zéro-erreur ? Notre première contribution consiste à déterminer la région réalisable, à l'aide d'une partition en cosets des suites de source, couplée à un dispositif de correction ; prouvant de ce fait que cette région réalisable en zéro-erreur coïncide avec la région réalisable en erreur asymptotiquement nulle.

Le deuxième problème basé sur la correction est illustré en Figure 4.3 et est une généralisation du premier. Dans ce problème où l'encodeur connaît deux sources différentes et leurs informations adjacentes respectives, deux incertitudes se présentent : la première est que l'encodeur ignore quelle source sera demandée par le décodeur, et la deuxième est que l'information adjacente de la source demandée peut être présente au décodeur. L'encodage est alors divisé en trois étapes. À la première étape, l'encodeur envoie de l'information sans savoir quelle source sera demandée ; nous cherchons à envoyer l'information commune aux deux différentes sources, d'où l'appellation "réseau de Gray-Wyner" pour ce problème. À la deuxième étape, l'encodeur est informé de la source demandée par le décodeur, mais ignore toujours si ce dernier dispose d'une information adjacente ; si tel est le cas, alors un décodage zéro-erreur doit être réalisé. La troisième étape est un supplément d'information envoyé par l'encodeur dans le cas où le décodeur ne dispose pas de l'information adjacente ; le décodeur doit ensuite retrouver la source avec zéro-erreur. Pour ce problème, notre deuxième contribution consiste en une borne interne et une borne externe sur la région réalisable. Pour la borne interne, nous utilisons le schéma de codage suivant : des suites auxiliaires sont utilisées pour capturer l'information commune entre les suites de réalisations des deux sources, ensuite nous effectuons un étiquetage aléatoire de ces suites, et les ensembles de suites de source sont partitionnées en cosets. Les algorithmes de décodages sont basés sur les V-enveloppes des suites d'information adjacente, et la propriété zéro-erreur est garantie par un dispositif de correction.

1.2 Problèmes zéro-erreur basés sur les graphes

Dans les problèmes zéro-erreur *basés sur les graphes*, les contraintes zéro-erreur sont représentées par un graphe caractéristique, et déterminer le débit optimal dans de tels problèmes revient à résoudre des questions difficiles de théorie des graphes. Le problème de Slepian-Wolf zéro-erreur illustré en Figure 3.4 est l'un d'eux, son débit optimal est donné par l'entropie complémentaire $\overline{H}$ de son graphe caractéristique, et trouver une expression à une seule lettre pour $\overline{H}$ est une question ouverte. Toutefois, afin de mieux comprendre la nature du problème Slepian-Wolf zéro-erreur, nous donnons de nouveaux résultats structurels qui lient ce problème avec d'autres, en particulier la capacité zéro-erreur d'un canal C_0 (voir la Figure 3.2).

Notre étude principale concerne le problème de "linéarisation". Dans le cadre du problème Slepian-Wolf zéro-erreur, si l'encodeur est en train de réaliser plusieurs tâches de compression indépendantes avec leurs informations adjacentes respectives, alors le débit optimal est donné par $\overline{H}(\wedge\cdot)$, i.e. $\overline{H}$ d'un graphe avec une structure de produit ET. Comme illustré par Tuncel et al. dans [71], "séparer" les tâches indépendantes donne un schéma d'atteignabilité, d'où $\overline{H}(\wedge\,\cdot) \leq \sum \overline{H}(\cdot)$. Un autre cas d'intérêt est celui où l'encodeur dispose d'une information partielle $g(Y)$ sur l'information adjacente du décodeur. Le débit optimal est donné par $\overline{H}(\sqcup\,\cdot)$, i.e. $\overline{H}$ d'un graphe avec une structure d'union disjointe. "Séparer" les réalisations de la source selon les valeurs de l'information adjacente de l'encodeur donne aussi un schéma d'atteignabilité, d'où $\overline{H}(\sqcup^{P_{g(Y)}}\,\cdot) \leq \sum_z P_{g(Y)}(z)\overline{H}(\cdot)$. Lorsqu'on a égalité dans l'un ou l'autre de ces cas, on dit qu'on a "linéarisation" de $\overline{H}$.

Notre contribution consiste à prouver les équivalences des linéarisations de $C_0(\wedge\,\cdot)$, $C_0(\sqcup\cdot)$, $C(\wedge\cdot, P_V)$, $C(\sqcup\cdot, P_V)$, $\overline{H}(\wedge\cdot)$, et $\overline{H}(\sqcup\cdot)$; où $C(\cdot, P_V)$ est le débit optimal du problème de codage canal zéro-erreur où l'encodeur doit utiliser des mots de codes typiques par rapport à la distribution P_V. Par conséquent, le schéma de codage "séparé" est optimal dans le problème de codage canal zéro-erreur, si et seulement si il est optimal dans le problème de Slepian-Wolf zéro-erreur avec le même graphe caractéristique. De plus, dans chacun de ces problèmes, les optimalités des schémas de codage "séparés" respectifs pour le produit $\wedge$ et pour l'union disjointe $\sqcup$ sont équivalentes. Pour prouver cela, nous définissons les distributions atteignant la capacité d'un canal en régime zéro-erreur, et nous déterminons plusieurs résultats sur celles-ci ; cela nous permet de lier C_0 avec $C(\cdot, P_V)$ et $\overline{H}$. De plus, nous donnons des exemples et contre-exemples de linéarisation pour toutes ces équivalences. Tout d'abord, cela nous donne une formule à une seule lettre pour $\overline{H}$

dans des cas où elle manquait ; en particulier pour les produits de graphes parfaits, qui ne sont pas nécessairement parfaits. Enfin, les contre-exemples développés illustrent que les schémas de codage “séparés” ne sont pas toujours optimaux: dans le problème de Slepian-Wolf zéro-erreur, le débit peut strictement décroître quand on compresse ensemble des sources indépendantes.

1.3 Problèmes zéro-erreur pour le calcul de fonction

Les problèmes zéro-erreur *pour le calcul de fonction* sont une généralisation des problèmes basés sur les graphes, et font aussi usage des graphes caractéristiques. Toutefois, la fonction à retrouver impacte aussi les outils adéquats de la théorie des graphes à utiliser. Dans le problème illustré en Figure 6.2, le décodeur doit retrouver une fonction f de la source et de son information adjacente. L'encodeur ne connaît pas les réalisations de cette dernière, mais en observe une version déterministiquement dégradée, représentée par la fonction g.

Notre première contribution est de donner une expression asymptotique pour le débit optimal de ce problème. Notre deuxième contribution est de formuler une hypothèse que nous appelons “information adjacente partagée deux-à-deux” qui nous permet, lorsqu'elle est satisfaite, d'obtenir une expression à une seule lettre pour le débit optimal. Cette hypothèse est satisfaite si chaque paire de symboles de source “partage” au moins un symbole d'information adjacente pour tout résultat de g. Cette condition a des interprétations en termes de théorie des graphes, car les formules à une seule lettre que nous trouvons émanent de la structure particulière du graphe caractéristique : ce dernier est une union disjointe de produits OU. De plus, ce résultat est d'intérêt pratique car il couvre toutes les instances où la distribution de source est à support plein, sans aucune hypothèse sur f, g. Enfin, nous donnons une interprétation de cette condition, en termes de plus mauvais débit optimal dans un problème auxiliaire de Slepian-Wolf zéro-erreur.

1.4 Organisation du manuscrit

Ce manuscrit est organisé comme suit. Dans le Chapitre 3, nous détaillons la présentation de chacun des problèmes étudiés et l'état de l'art pour chacun d'eux, ainsi que les définitions utilisées dans ce manuscrit. Dans le Chapitre 4, le Chapitre 5, et le Chapitre 6, nous présentons respectivement nos résultats pour les problèmes zéro-erreur *basés sur la*

correction, *basés sur les graphes*, et *pour le calcul de fonction*. Les preuves de ces résultats se trouvent respectivement en Appendice A, Appendice B, et Appendice C.

GLOSSARY

For reader's convenience we give here a table of the notations and concepts used throughout this manuscript.

$\mathbb{N}^\star$, $\mathbb{R}$	set of positive integers, real numbers
$\mathcal{M}_{n,k}(\mathcal{A})$	set of matrices of size $n \times k$ with coefficients in $\mathcal{A}$
$\lvert\mathcal{S}\rvert$, $\mathcal{S}^c$	cardinality, complement of the set $\mathcal{S}$
Im	image set of a mapping
Ker	kernel of a matrix
$\{0,1\}^*$	set of binary words
$\ell(\cdot)$	length of a word
x^n	sequence $(x_1, ..., x_n)$
$\Delta(\mathcal{X})$	probability distributions over $\mathcal{X}$
P_X	probability distribution of the random variable X
$\operatorname{supp} P_X$	support of P_X
$\mathrm{Unif}(\cdot)$	uniform distribution
$P_{X,Y}$	joint distribution of (X, Y)
$P_{X\lvert Y}$	conditional distribution of X knowing Y
$P_X \otimes P_Y$	product distribution
$\cdot \to \cdot \to \cdot$	Markov chain
T_{x^n}	type of x^n $\left(\frac{1}{n}\middle\lvert\{t \leq n \mid x_t = x'\}\middle\rvert\right)_{x' \in \mathcal{X}}$
$\Delta_n(\mathcal{X})$	set of types of sequences from $\mathcal{X}^n$
$\mathcal{T}^n_\epsilon(P_X)$	typical set $\left\{x^n \in \mathcal{X}^n \;\middle\vert\; \lVert T_{x^n} - P_X\rVert_\infty \leq \epsilon\right\}$
$\mathbb{P}(\cdot)$	probability of an event
$\mathbb{E}[\,\cdot\,]$	expected value of a random variable
$H(\cdot)$, $H(\cdot \mid \cdot)$	entropy, conditional entropy
$I(\cdot\,;\cdot)$	mutual information
$I(\cdot\,;\cdot \mid \cdot)$	conditional mutual information

C_n, K_n, N_n	cycle, complete, empty graph with n vertices	
ϕ_e, ϕ_d	encoding, decoding function	
$\wedge$	AND product	Definition 3.2.6
α	independence number	Definition 3.2.7
C_0	zero-error capacity	Definition 3.2.8
$(\mathcal{V}, \mathcal{E}, P_V)$	probabilistic graph	Definition 3.2.5
χ	chromatic number	Definition 3.3.4
$\overline{H}$	complementary graph entropy	Definition 3.3.6
H_χ	chromatic entropy	Definition 3.3.8
$\mathcal{T}_0^n(Q_X)$	type class for the type Q_X	Definition 4.2.2
$\mathcal{T}_V(x^n)$	V-shell of the sequence x^n	Definition 4.2.2
$G[S]$	subgraph of G induced by $\mathcal{S}$	Definition 3.3.5
$C(\cdot, P_V)$	zero-error capacity relative to P_V	Definition 5.1.1
$\sqcup$	disjoint union	Definition 5.1.8
$\overline{G}$	graph complement of G	Definition 5.5.1
ω	clique number	Definition 5.5.1
H_κ	Körner graph entropy	Definition 5.5.3
$\vee$	OR product	Definition 6.3.3
$\simeq$	isomorphic (probabilistic) graphs	Definition B.2.6

A set of words $\mathcal{W} \subseteq \{0,1\}^*$ is *prefix-free* if for all $w, w' \in \mathcal{W}$, w is not a prefix of w' and *vice-versa*.

Chapter 2

Introduction

Zero-error source coding encompasses a variety of problems from information theory, where a source is compressed by an encoder with rate to be minimized, then has to be retrieved at the decoder with probability of error zero. In all the problems studied here, the source distribution is known and the decoder has access to a side information, i.e. a random variable correlated to the source that helps minimizing the rate and decoding. Studying such settings is useful in applications that require a strong guarantee on the quality of reconstruction of the source, such as remote sensing networks and distributed computing. In a zero-error source coding problem, finding a formula for the optimal rate means finding the strongest theoretic limit on all compression schemes. Furthermore, if the formula is single-letter, then it becomes easily computable. One also has the existence of an optimal zero-error compression scheme that asymptotically achieves this optimal rate. Therefore, finding single-letter expressions for optimal rates in zero-error source coding problems is our main objective in this manuscript.

The zero-error constraint differs from the vanishing-error constraint, the latter only requires the probability of error to go to zero when the block length of the code goes to infinity. In the settings where the decoder does not have a side information, it can be easily shown that the optimal rates in the vanishing error and zero-error regimes coincide. However, when imposing the zero-error constraint when the decoder has a side information, as in the settings studied here, many problems change from a statistical nature to a combinatorial one (e.g. the Slepian-Wolf setting). Therefore, the adequate tools for zero-error problems are drawn from both information theory and combinatorics; and vary from a correcting device that strengthens vanishing-error codes into zero-error ones, to zero-error codes entirely built with graph-theoretic constraints.

The combinatorial nature of zero-error problems is tied to the encoder's lack of knowledge of the decoder's side information. In vanishing-error problems one can assume that the decoder's side information sequence will be typical w.r.t. the source distribution with high probability; even if the encoder does not have access to it. However, the zero-error

constraint makes the encoder's uncertainty on decoder's side information harder to be dealt with; in particular it is required to build codes that enable for a perfect reconstruction of the source for *any* realization of the decoder's side information.

For this reason, we classify the zero-error problems studied here into three categories, depending on what the encoder knows about the decoder's side information and what the decoder has to retrieve:

- The *correction-based* zero-error problems, in which the encoder has access to all side information and can simulate every decoding.
- The *graph-based* zero-error problems, in which the encoder is "less informed" about the realizations of decoder's side information.
- The *coding for computing* zero-error problems, in which the decoder has to compute a function of the source and its side information; instead of only recovering the source.

2.1 Correction-based zero-error problems

The *correction-based* problems can be tackled with a correcting device that guarantees the zero-error property. It consists in using a vanishing-error codebook, simulating each decoding, and sending all the source sequences if any decoding fails. This correcting device has a negligible impact on the rate because of the low probability of error, and allows for the conversion of vanishing error coding schemes into zero-error ones. For this reason, the tools we use are typicality, random coding and parity check codes, with correcting device.

In the first problem from this class, which is depicted in Figure 4.1, side information may be present at the decoder. In this setting, the encoder has access to a side information but does not know whether the decoder has access to it; and the encoding is divided into two steps. In the first step, the encoder sends information without knowing whether side information is available at the decoder. The message sent in the first step must be exploitable independently from the presence of the side information; furthermore, a zero-error decoding must be done if the decoder has access to the side information. In the second step, an information complement is sent by the encoder in the case where the decoder has no side information, and the decoder must be able to retrieve the source with zero-error. The two steps are modeled by two decoders and two noiseless channels. We aim at answering the following question: what are the feasible pairs of rates for the two encoding steps, that allow for a zero-error decoding? Our first contribution consists

in determining the feasible rate region, thanks to a coset partition of source sequences coupled with a correcting device; therefore showing that it coincides with the feasible rate region in the vanishing error regime.

The second correction-based zero-error problem is depicted in Figure 4.3 and is a generalization of the first one, in which there are two uncertainties: firstly, the encoder has access to two sources and their respective side information but does not know which one will be requested by the decoder; and secondly, the side information may be present at the decoder. The encoding is divided into three steps. In the first step, the encoder sends information without knowing which source is requested; we aim at sending the common information between the two sources, hence the "Gray-Wyner network" name for this problem. In the second step, the encoder knows which source is requested but does not know if the decoder has access to the side information; if the latter is present at the decoder, then a zero-error decoding must be done. The third step is an information complement sent by the encoder in the case where the decoder does not have the side information; and the decoder must be able to retrieve the source with zero-error. In this setting, our second contribution consists in an inner and an outer bound on the feasible rate region. For the inner bound we use the following coding scheme: auxiliary sequences are used to capture the common information between the two sources sequences, then we do a random binning on them; and the sets of source sequences are partitioned into cosets. Decoding algorithms are based on V-shells of side information sequences, and the zero-error property is ensured by a correcting device.

2.2 Graph-based zero-error problems

In *graph-based* problems, zero-error constraints are captured by a characteristic graph, and determining the optimal rate requires to solve hard graph-theoretic questions. The zero-error Slepian-Wolf problem depicted in Figure 3.4 is one of them, its optimal rate is given by the complementary graph entropy $\overline{H}$ of the characteristic graph, and finding a single-letter formula for $\overline{H}$ is an open question. However, in order to understand better the nature of the zero-error Slepian-Wolf problem, we provide new structural results that link it with other problems, such as the zero-error capacity of a channel C_0 (see Figure 3.2).

Our main study concerns the "linearization" problem. In the zero-error Slepian-Wolf setting, if the encoder is doing several independent compression tasks with their respective

independent side information, then the optimal rate is given by $\overline{H}(\wedge\,\cdot)$, i.e. $\overline{H}$ of a graph with an AND product structure. As shown by Tuncel et al. in [71], "separating" the independent tasks yields an achievability scheme, hence $\overline{H}(\wedge\,\cdot) \leq \sum \overline{H}(\cdot)$. Another interesting case of zero-error Slepian-Wolf problem is when the encoder has a partial information $g(Y)$ on the decoder's side information. The optimal rate is given by $\overline{H}(\sqcup\,\cdot)$, i.e. $\overline{H}$ of a graph with a disjoint union structure. "Separating" the source realizations w.r.t. the encoder's side information also yields an achievability scheme, hence $\overline{H}(\sqcup^{P_{g(Y)}}\,\cdot) \leq \sum_z P_{g(Y)}(z)\overline{H}(\cdot)$. When equality holds in either case we say that "linearization" of $\overline{H}$ holds.

Our contribution is to prove the equivalences of linearizations between $C_0(\wedge\,\cdot)$, $C_0(\sqcup\,\cdot)$, $C(\wedge\,\cdot, P_V)$, $C(\sqcup\,\cdot, P_V)$, $\overline{H}(\wedge\,\cdot)$, and $\overline{H}(\sqcup\,\cdot)$; where $C(\cdot, P_V)$ is the optimal rate in the zero-error channel coding problem where the encoder has to use codewords that are typical w.r.t. P_V. Therefore, the "separated" coding scheme is optimal in the zero-error channel coding problem, if and only if it is optimal in the zero-error Slepian-Wolf setting with the same characteristic graph. Furthermore, in each of these settings, the optimalities of the respective "separated" coding strategies for the product $\wedge$ and the disjoint union $\sqcup$ are equivalent. In order to prove that, we define the zero-error capacity achieving distributions of a channel, and derive several results on them; which enables us to link C_0 with $C(\cdot, P_V)$ and $\overline{H}$. Furthermore, we give examples and counterexamples of linearization for all these equivalences. Firstly, this yields a single-letter formula for $\overline{H}$ in cases where it was unknown: product of perfect graphs, which are not necessarily perfect. Secondly, the counterexamples developed illustrate that the "separated" coding schemes are not always optimal: in the zero-error Slepian-Wolf setting, the rate may strictly decrease when compressing together independent sources.

2.3 Coding for computing zero-error problems

The *coding for computing* problems are a generalization of graph-based problems, and also make use of characteristic graphs. However, the function to be retrieved also impacts the adequate graph-theoretic tools to be used. In the setting depicted in Figure 6.2, the decoder has to retrieve a function f of its side information and of the source; the encoder does not know the realizations of the decoder's side information but observes a deterministically degraded version of them, represented by the function g.

Our first contribution is to give an asymptotic formula for the optimal rate in this setting. Our second contribution is to formulate an hypothesis that we call "pairwise

shared side information" that allows us to derive a single-letter characterization of the optimal rate. This hypothesis is satisfied if every pair of source symbols "share" at least one side information symbol for all output of g. It has graph-theoretic interpretations, as the single-letter formula stems from the particular structure of the characteristic graph: a disjoint union of OR products. Moreover, this result is of practical interest as it covers the cases where the source distribution is full-support, without any assumption on f, g. Finally, we give an interpretation of this condition, in terms of the worst optimal rate in an auxiliary zero-error Slepian-Wolf problem.

2.4 Organization of the manuscript

This manuscript is organized as follows. In Chapter 3 we provide a detailed presentation of all the problems studied and the definitions used in this manuscript, and we describe the state of the art for each one of them. In Chapter 4, Chapter 5, and Chapter 6, we present our results for the *correction-based*, *graph-based*, and *coding for computing* zero-error problems, respectively. The proofs of these results can be found in Appendix A, Appendix B, and Appendix C, respectively.

Chapter 3

State of the art

In this Chapter we give a detailed presentation of the different problems studied in this manuscript, in both zero-error and vanishing error regime. As illustrated in the following, the zero-error and the vanishing error regimes may lead to different optimal rates in several settings; in particular in the channel coding problem, and in the Slepian-Wolf problem.

3.1 Source coding

The source coding problem is one of the fundamental settings that appear in data compression; it is introduced by Shannon in [59].

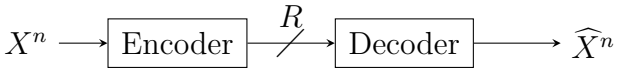


Figure 3.1 – The source coding problem.

Definition 3.1.1 (Source coding problem) *The source coding problem of Figure 3.1 is described by:*

- *A finite set $\mathcal{X}$, and a source distribution $P_X \in \Delta(\mathcal{X})$;*
- *For all $n \in \mathbb{N}^\star$, X^n is the random sequence of n copies of X, drawn in an i.i.d. fashion using P_X.*
- *An encoder that knows X^n sends binary strings over a noiseless channel to a decoder that wants to retrieve X^n without error.*

A coding scheme in this setting is described by:

- *A time horizon $n \in \mathbb{N}^\star$, and an encoding function $\phi_e : \mathcal{X}^n \to \{0,1\}^*$ such that $\operatorname{Im} \phi_e$ is prefix-free;*
- *A decoding function $\phi_d : \{0,1\}^* \to \mathcal{X}^n$;*

- *The rate is the average length of the codeword per source symbol, i.e.* $R \doteq \frac{1}{n}\mathbb{E}[\ell \circ \phi_e(X^n)]$.

Depending whether a perfect source reconstruction is required, two different regimes can be considered.

Definition 3.1.2 (Zero-error regime, vanishing error regime) *In the* ***zero-error regime****, the coding schemes* (n, ϕ_e, ϕ_d) *must satisfy the* ***zero-error property****:*

$$\mathbb{P}\Big(\widehat{X}^n \neq X^n\Big) = 0, \tag{3.1}$$

where $\widehat{X}^n = \phi_d(\phi_e(X^n))$*. The objective is to find the minimal rate among all coding schemes under the zero-error constraint:*

$$R^*_{SC0} \doteq \inf_{\substack{n,\phi_e,\phi_d \\ \text{zero-error}}} \frac{1}{n}\mathbb{E}[\ell \circ \phi_e(X^n)]. \tag{3.2}$$

In the ***vanishing error regime****, the coding schemes* (n, ϕ_e, ϕ_d) *must satisfy the* ϵ***-error property****:*

$$\mathbb{P}\Big(\widehat{X}^n \neq X^n\Big) \leq \epsilon, \tag{3.3}$$

where $\widehat{X}^n = \phi_d(\phi_e(X^n))$*. The objective is to find the minimal rate among all coding schemes under the* ϵ*-error constraint, with* $\epsilon \to 0$*:*

$$R^*_{SC} \doteq \lim_{\epsilon \to 0} \inf_{\substack{n,\phi_e,\phi_d \\ \epsilon\text{-error}}} \frac{1}{n}\mathbb{E}[\ell \circ \phi_e(X^n)]. \tag{3.4}$$

Determining the optimal rate means determining theoretic limit of all possible coding schemes in the source coding problem. The zero-error regime induces stronger constraints on the coding schemes than the vanishing error regime, thus leading to a higher optimal rate.

Theorem 3.1.3 (from [59])

$$R^*_{SC} = H(X). \tag{3.5}$$

Theorem 3.1.4 (from [34])

$$R^*_{SC0} = H(X). \tag{3.6}$$

The optimal rate is characterized by the entropy of the information source: Theorem 3.1.3 and Theorem 3.1.4 show that the optimal rate in this setting equals $H(X)$, for both vanishing error and zero-error regimes. This stems from the fact that the optimal rate in the vanishing error regime can be achieved with a Huffman algorithm, which also satisfies the zero-error property.

Alternatively by Remark 3.1.5, a possible zero-error coding strategy consists in indexing the typical set w.r.t. P_X following the approach from [59], and using the correcting device.

Remark 3.1.5 (Correcting device) *In this setting, since the encoder knows the realizations of every random variable in the problem, it can simulate every decoding. Therefore, every coding scheme in the vanishing error regime can be turned into zero-error ones with negligible impact on the rate in the following way: the encoder uses the coding scheme in the vanishing error regime, along with a bit of error. The latter equals* 0 *in case no decoding error occurs and* 1 *if any decoding error happens, and all source sequences are sent to the decoder in that case. The correcting device has a negligible additional cost on the rate, due to the negligible cost of the flag bit and the low probability of error of a coding scheme in the vanishing error regime. As a consequence, both optimal rates of this problem in the vanishing error and in the zero-error regime are equal.*

3.2 Channel coding

The channel coding problem is introduced in [59] in the vanishing error regime; and in [58] in the zero-error regime. It is a well-known example where the respective optimal rates in these regimes are different. In particular, a full-support distribution $P_{Y|X}$ for the channel may yield a positive channel capacity (i.e. optimal rate) in the vanishing error regime, and a zero-error capacity equal to 0.

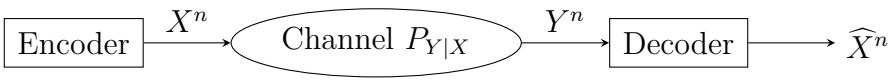


Figure 3.2 – The channel coding problem.

Definition 3.2.1 (Channel coding problem) *The channel coding problem of Figure 3.2 is described by:*

- *Two finite sets $\mathcal{X}$, $\mathcal{Y}$;*
- *A Discrete Memoryless Channel (DMC) with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and transition probability $P_{Y|X} \in \Delta(\mathcal{Y})^{|\mathcal{X}|}$;*
- *An encoder that sends inputs over the DMC, and a decoder that receives the DMC's outputs.*

A coding scheme in this setting is described by:

- *A time horizon $n \in \mathbb{N}^\star$, and a codebook $\mathcal{C}_n \subseteq \mathcal{X}^n$;*
- *A decoding function $\phi_d : \mathcal{Y}^n \to \mathcal{X}^n$;*
- *The rate is the average number of messages transmitted per channel use, i.e. $\frac{1}{n}\log|\mathcal{C}_n|$.*

The channel coding problem is formally defined for two different regimes, depending whether a positive probability of error is allowed when communicating through the DMC.

Definition 3.2.2 (Zero-error regime, vanishing error regime) *In the **zero-error regime**, the coding scheme $(n, \mathcal{C}_n, \phi_d)$ must satisfy the **zero-error property**:*

$$\mathbb{P}\Big(\widehat{X}^n \neq X^n\Big) = 0; \tag{3.7}$$

where $X^n \sim \mathrm{Unif}(\mathcal{C}_n)$, $\widehat{X}^n = \phi_d(Y^n)$, and Y^n is drawn conditionally w.r.t. X^n using $P_{Y|X}^n$. The objective is to find the maximal rate among all coding schemes that satisfy the zero-error property:

$$R_{C0}^* \doteq \sup_{\substack{n,\mathcal{C}_n,\phi_d \\ \text{zero-error}}} \frac{1}{n}\log|\mathcal{C}_n|. \tag{3.8}$$

*In the **vanishing error regime**, the coding schemes $(n, \mathcal{C}_n, \phi_d)$ must satisfy the ϵ-**error property**:*

$$\mathbb{P}\Big(\widehat{X}^n \neq X^n\Big) \leq \epsilon, \tag{3.9}$$

where $X^n \sim \mathrm{Unif}(\mathcal{C}_n)$, $\widehat{X}^n = \phi_d(Y^n)$, and Y^n is drawn conditionally w.r.t. X^n using $P_{Y|X}^n$. The objective is to find the maximal rate among all coding schemes that satisfy the ϵ-error

property, with $\epsilon \to 0$*:*

$$R_C^* \doteq \lim_{\epsilon \to 0} \sup_{\substack{n, \mathcal{C}_n, \phi_d \\ \epsilon\text{-error}}} \frac{1}{n} \log |\mathcal{C}_n|. \tag{3.10}$$

3.2.1 Channel coding in the vanishing error regime

The channel capacity in the vanishing error regime is defined by $\max_{P_X \in \Delta(\mathcal{X})} I(X;Y)$; and as shown by Shannon in Theorem 3.2.3, this quantity characterizes the optimal rate of communication over a DMC in the vanishing error regime. This comes from the fact that for all $P_X \in \Delta(\mathcal{X})$, the encoder may choose $2^{nI(X;Y)}$ random codewords from $\mathcal{X}^n$ by using P_X, and expect a correct decoding with high probability by typicality arguments. Taking the maximum over P_X yields the highest achievable rate.

Theorem 3.2.3 (from [59])

$$R_C^* = \max_{P_X \in \Delta(\mathcal{X})} I(X;Y). \tag{3.11}$$

3.2.2 Zero-error channel coding

In the zero-error regime, an adequate graph G_C can be associated to a given instance of channel coding problem in Figure 3.2. This graph is called "characteristic graph" of the problem, as it encompasses the problem data in its structure: the vertices are the source alphabet, and two channel input symbols x, x' are adjacent if they are "confusable", i.e. $P_{Y|X}(y|x)P_{Y|X}(y|x') > 0$ for some channel output symbol y. In other words, when x and x' are adjacent, knowing the realization y does not allow to distinguish between the realizations x and x' with probability of error 0.

Definition 3.2.4 (Characteristic graph) *Let* $\mathcal{X}, \mathcal{Y}$ *be two finite sets and* $P_{Y|X}$ *be a conditional distribution from* $\Delta(\mathcal{Y})^{|\mathcal{X}|}$*. The characteristic graph associated to* $P_{Y|X}$ *is defined by:*

- $\mathcal{X}$ *as set of vertices,*
- $x, x' \in \mathcal{X}$ *are adjacent if* $P_{Y|X}(y|x)P_{Y|X}(y|x') > 0$ *for some* $y \in \mathcal{Y}$*.*

The AND product $\wedge$ is a binary operator on graphs, and is used to build the characteristic graph for more than one channel use. More precisely, two sequences of channel

inputs x^n, x'^n are adjacent in the n-th AND power $G_C^{\wedge n}$ (i.e. iterated AND product of G_C) if $P_{Y|X}^n(y^n|x^n)P_{Y|X}^n(y^n|x'^n) > 0$ for some sequence of channel outputs y^n.

In Definition 3.2.6, the AND product of graphs with an underlying distribution is defined, as it will be useful for the zero-error source coding problem. In the case of zero-error channel coding, graphs without distribution are considered, but the structure of the AND product is the same regardless of the distribution. The AND product is also called "strong product" or "normal product" in the literature (for example in [44, 47]).

Definition 3.2.5 (Probabilistic graph) *A probabilistic graph G is a tuple $(\mathcal{V}, \mathcal{E}, P_V)$, where $\mathcal{V}$ and $\mathcal{E}$ are respectively the sets of vertices and edges; and $P_V \in \Delta(\mathcal{V})$ is a probability distribution on the vertices.*

Definition 3.2.6 (AND product $\wedge$) *Let $G_1 = (\mathcal{V}_1, \mathcal{E}_1, P_{V_1})$, $G_2 = (\mathcal{V}_2, \mathcal{E}_2, P_{V_2})$ be two probabilistic graphs, their AND product $G_1 \wedge G_2$ is a probabilistic graph defined by:*

- *$\mathcal{V}_1 \times \mathcal{V}_2$ as set of vertices,*
- *$P_{V_1} \otimes P_{V_2}$ as probability distribution on the vertices,*
- *$(v_1 v_2), (v_1' v_2')$ are adjacent if*

$$v_1 v_1' \in \mathcal{E}_1 \text{ AND } v_2 v_2' \in \mathcal{E}_2, \tag{3.12}$$

with the convention of self-adjacency for all vertices.

We denote by $G_1^{\wedge n}$ the n-th AND power:

$$G_1^{\wedge n} = G_1 \wedge ... \wedge G_1 \qquad (n \text{ times}). \tag{3.13}$$

The AND product of graphs without probability distribution has the vertex set and edges defined above, without underlying probability distribution.

If there is a pair of codewords x^n, x'^n adjacent in $G_C^{\wedge n}$ in the codebook used by the encoder, then upon receiving the y^n such that $P_{X|Y}^n(x^n|y^n)P_{X|Y}^n(x'^n|y^n) > 0$, the decoder is unable to determine whether the encoder sent x^n or x'^n, which prevents zero-error decoding. Therefore, a zero-error decoding at the end of the time horizon n is possible if and only if the encoder uses a codebook formed of pairwise non-adjacent symbols in $G_C^{\wedge n}$, i.e. an independent set, which is formally defined below.

Definition 3.2.7 (Independent subset, independence number α) *Let $G = (\mathcal{V}, \mathcal{E})$ be a graph. A subset $\mathcal{S} \subseteq \mathcal{V}$ is independent in G if $xx' \notin \mathcal{E}$ for all $x, x' \in \mathcal{S}$. The independence number is the maximal size of an independent set in G, and is denoted by $\alpha(G)$.*

Definition 3.2.8 (Zero-error capacity of a graph C_0) *Let G be a graph, its zero-error capacity is defined by*

$$C_0(G) \doteq \lim_{n\to\infty} \frac{1}{n} \log \alpha(G^{\wedge n}). \tag{3.14}$$

By construction, the optimal rate for a coding scheme with the zero-error property and time horizon n is $\frac{1}{n}\log\alpha(G_C^{\wedge n})$. As shown in Proposition 3.2.9, the asymptotic optimal rate is $\lim_{n\to\infty}\frac{1}{n}\log\alpha(G_C^{\wedge n})$; this quantity is called the zero-error capacity of a channel (or the zero-error capacity of its characteristic graph), and represents the best zero-error communication rate through a DMC.

Note that, by convention, we define the zero-error capacity with the logarithm. Another existing convention (for example in [44]) for the zero-error capacity is $\Theta(G) \doteq \lim_{n\to\infty} \sqrt[n]{\alpha(G^{\wedge n})}$; which is equivalent in the sense that $C_0 = \log\Theta$.

Theorem 3.2.9 (from [58]) *The optimal rate in the zero-error channel coding setting writes*

$$R_{C0}^* = C_0(G_C), \tag{3.15}$$

where G_C is the characteristic graph associated to the distribution $P_{Y|X}$.

Determining a single-letter expression for C_0 is a wide open problem. We present in Section 5.5 some examples from the literature where C_0 is known, in particular perfect graphs. The Lovász θ function, introduced in [44], is an upper bound on the zero-error capacity. This function is used to show that $C_0(C_5) = \frac{\log 5}{2}$, which makes C_5 the minimally non-perfect graph for which C_0 is known. Further observations on the θ function are derived by Sason in [54]. The zero-error capacity of C_7 is still unknown. Several existing lower bounds on $C_0(C_7)$ result from an independent set found by a computer program; in particular by Vesel and Žerovnik in [72], by Mathew and Östergård in [48], and by Polak and Schrijver in [51].

Figure 3.3 – The cycle graphs C_5, C_7.

3.2.3 Related works

The computability of C_0 is investigated in [9] by Boche and Deppe, and they prove that C_0 is not computable. An asymptotic expression for C_0 using semiring homomorphisms is given by Zuiddam et al. in [78]. In [31], Gu and Shayevitz study the two-way channel case. An extension of C_0 for secure communication is developed in [74] by Wiese et al.

Another related setting is zero-error transmission over a discrete channel with memory. The case of binary channels with one memory is studied by Ahlswede et al. in [1], followed by Cohen et al. in [17], and the remaining unsolved cases are solved by Cao et al. in [12]. The case of binary channels with two memories is studied by Zhang et al. in [77].

Finally, the Sperner capacity is an extension of the zero-error capacity to digraphs introduced in [28] by Gargano et al. Upper bounds on the Sperner capacity are developed by Alon in [3] based on the maximum outdegree of the digraph; and by Körner et al. in [40] based on an adaptation to digraphs of the local chromatic number.

The interested reader may refer to Körner and Orlitsky's survey in [38], and Simonyi's survey in [63].

3.3 Slepian-Wolf problem

The Slepian-Wolf problem is introduced in [64] by Slepian and Wolf, in the vanishing error regime; and the zero-error variant of this problem is presented by Alon and Orlitsky, in [5]. This problem corresponds to a situation in data compression where the decoder has a side-information Y about the source X that has to be retrieved. Its optimal rate in the zero-error regime is given by the complementary graph entropy $\overline{H}$ of the characteristic graph, and finding a single-letter formula for $\overline{H}$ is an open question. As presented in Theorem 3.3.7 and Theorem 3.3.3, the respective optimal rates in the zero-error regime and in the vanishing error regime are different, in particular when the source distribution is full-support. Note that the correcting device described in Remark 3.1.5 cannot be used here for the zero-error regime, as the encoder does not know the realizations of Y.

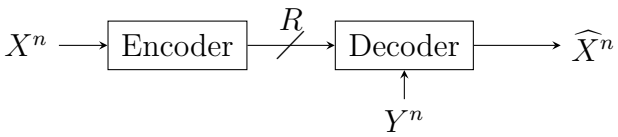


Figure 3.4 – The zero-error Slepian-Wolf problem.

Definition 3.3.1 (Zero-error Slepian-Wolf problem) *The zero-error source coding problem of Figure 3.4 is described by:*

- *Two finite sets $\mathcal{X}$, $\mathcal{Y}$, and a source distribution $P_{X,Y} \in \Delta(\mathcal{X} \times \mathcal{Y})$;*
- *For all $n \in \mathbb{N}^\star$, (X^n, Y^n) is the random sequence of n copies of (X, Y), drawn in an i.i.d. fashion using $P_{X,Y}$.*
- *An encoder that knows X^n sends binary strings over a noiseless channel to a decoder that knows Y^n, and that wants to retrieve X^n without error.*

A coding scheme in this setting is described by:

- *A time horizon $n \in \mathbb{N}^\star$, and an encoding function $\phi_e : \mathcal{X}^n \to \{0,1\}^*$ such that $\operatorname{Im} \phi_e$ is prefix-free;*
- *A decoding function $\phi_d : \mathcal{Y}^n \times \{0,1\}^* \to \mathcal{X}^n$;*
- *The rate is the average length of the codeword per source symbol, i.e. $R \doteq \frac{1}{n}\mathbb{E}[\ell \circ \phi_e(X^n)]$.*

Definition 3.3.2 (Zero-error regime, vanishing error regime) *In the **zero-error regime**, the coding schemes (n, ϕ_e, ϕ_d) must satisfy the **zero-error property**:*

$$\mathbb{P}\Big(\widehat{X}^n \neq X^n\Big) = 0, \tag{3.16}$$

where $\widehat{X}^n = \phi_d(Y^n, \phi_e(X^n))$. The objective is to find the minimal rate among all coding schemes under the zero-error constraint:

$$R^*_{SW0} \doteq \inf_{\substack{n,\phi_e,\phi_d \\ \text{zero-error}}} \frac{1}{n}\mathbb{E}[\ell \circ \phi_e(X^n)]. \tag{3.17}$$

*In the **vanishing error regime**, the coding schemes (n, ϕ_e, ϕ_d) must satisfy the **ϵ-error property**:*

$$\mathbb{P}\Big(\widehat{X}^n \neq X^n\Big) \leq \epsilon, \tag{3.18}$$

where $\widehat{X}^n = \phi_d(Y^n, \phi_e(X^n))$*. The objective is to find the minimal rate among all coding schemes under the* ϵ*-error constraint, with* $\epsilon \to 0$*:*

$$R^*_{SW} \doteq \lim_{\epsilon \to 0} \inf_{\substack{n, \phi_e, \phi_d \\ \epsilon\text{-error}}} \frac{1}{n} \mathbb{E}[\ell \circ \phi_e(X^n)]. \tag{3.19}$$

Theorem 3.3.3 (from [64]) *The optimal rate in the Slepian-Wolf setting in the vanishing error regime writes*

$$R^*_{SW} = H(X|Y). \tag{3.20}$$

In the vanishing error regime, the optimal rate $H(X|Y)$ in Theorem 3.3.3 remains the same if the encoder also observes the sequence Y^n.

3.3.1 Zero-error Slepian-Wolf problem

In a similar way to zero-error channel coding, the characteristic graph G_{SW} of an instance of zero-error Slepian-Wolf problem in Figure 3.4 is defined in Definition 3.2.4, with the conditional source distribution $P_{Y|X}$. This graph was first used by Witsenhausen in [75]. However in the zero-error Slepian-Wolf setting it is a probabilistic graph, as it has the underlying distribution P_X on its vertices.

Assume that $\phi_e(x^n) = \phi_e(x'^n)$ for some x^n, x'^n adjacent in $G^{\wedge n}_{SW}$, then upon receiving the y^n such that $P^{\otimes n}_{X|Y}(x^n|y^n)P^{\otimes n}_{X|Y}(x'^n|y^n) > 0$, the decoder is unable to determine whether the encoder sent x^n or x'^n, which prevents zero-error decoding. Therefore, the source sequences that are adjacent in the characteristic graph must be mapped to different codewords; hence the use of graph colorings, which map adjacent vertices to different colors.

Definition 3.3.4 (Coloring, chromatic number χ) *Let* $G = (\mathcal{V}, \mathcal{E})$ *be a graph. A mapping* $c : \mathcal{V} \to \mathcal{C}$ *is a coloring if* $c^{-1}(i)$ *is independent for all* $i \in \mathcal{C}$*. The chromatic number* $\chi(G)$ *is the smallest* $|\mathcal{C}|$ *such that there exists a coloring* $c : \mathcal{V} \to \mathcal{C}$ *of* G*.*

Definition 3.3.5 (Induced subgraph $G[\mathcal{S}]$) *Let* $G = (\mathcal{V}, \mathcal{E})$ *be a graph, and let* $\mathcal{S} \subseteq \mathcal{V}$ *be a subset of vertices. The subgraph of* G *induced by* $\mathcal{S}$ *is the graph denoted by* $G[\mathcal{S}]$ *with* $\mathcal{S}$ *as set of vertices, and* vv' *are adjacent in* $G[\mathcal{S}]$ *if and only if they are adjacent in* G*.*

In the case where G *is a probabilistic graph with underlying distribution* P_V*, we give* $G[\mathcal{S}]$ *the normalized underlying distribution* $P_V / P_V(\mathcal{S})$*.*

With high probability, the source sequence X^n will be typical w.r.t. P_X, therefore one can consider only the subgraph of $G_{SW}^{\wedge n}$ induced by the set of typical sequences $\mathcal{T}_\epsilon^n(P_X)$. Then the encoder colors this induced subgraph $G_{SW}^{\wedge n}[\mathcal{T}_\epsilon^n(P_X)]$ with a coloring that has the minimum number of colors, and sends the color index to the decoder if X^n is typical, or the index of X^n in $\mathcal{X}^n$ otherwise. This coding strategy has a rate upper-bounded by

$$\frac{1}{n} + \mathbb{P}\Big(X^n \notin \mathcal{T}_\epsilon^n(P_X)\Big) \log|\mathcal{X}| + \frac{1}{n} \log \chi\Big(G_{SW}^{\wedge n}[\mathcal{T}_\epsilon^n(P_X)]\Big); \tag{3.21}$$

where χ is the chromatic number and the decoder is able to retrieve X^n thanks to the color and Y^n. Koulgi et al. have shown in [41, Theorem 1] that taking the limit when n goes to infinity and ϵ goes to 0 yields the best achievable rate in the zero-error Slepian Wolf problem. This quantity is called complementary graph entropy, and is defined by Körner and Longo in [37].

Definition 3.3.6 (Complementary graph entropy $\overline{H}$) *For all probabilistic graph $G = (\mathcal{V}, \mathcal{E}, P_V)$, the quantity $\overline{H}(G)$ is defined by:*

$$\overline{H}(G) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \chi\Big(G^{\wedge n}[\mathcal{T}_\epsilon^n(P_V)]\Big). \tag{3.22}$$

Theorem 3.3.7 (from [41, Theorem 1]) *The optimal rate in the zero-error Slepian-Wolf setting writes*

$$R_{SW0}^* = \overline{H}(G_{SW}), \tag{3.23}$$

where G_{SW} is the probabilistic graph formed of the characteristic graph associated to the distribution $P_{Y|X}$, with the underlying distribution P_X on its vertices.

The zero-error Slepian Wolf setting that we study is called "restricted inputs" by Alon and Orlitsky in [5] where they give another characterization of the optimal rate with chromatic entropies. The chromatic entropy $H_\chi(G_{SW}^{\wedge n})$ corresponds to the minimal entropy of a coloring of $G_{SW}^{\wedge n}$. Therefore, after normalization, it characterizes the best rate at a fixed number n of source uses with a perfect compression of the color. As stated in Theorem 3.3.9, by taking the limit when n goes to infinity, one obtains the optimal rate; and also, another expression for $\overline{H}$.

A stronger notion of zero-error coding, called "unrestricted inputs", is also introduced in [5]. It requires the zero-error property to be satisfied even for the sequences of symbols

(X^n, Y^n) that take values out of the support of $P_{X,Y}^n$. Alon and Orlitsky provide in [5], a single-letter formula for the optimal rate.

Definition 3.3.8 (Chromatic entropy H_χ) *Let $G = (\mathcal{V}, \mathcal{E}, P_V)$ be a probabilistic graph, its chromatic entropy is defined by*

$$H_\chi(G) = \inf\{H(c(V)) \mid c \text{ is a coloring of } G\}. \tag{3.24}$$

Theorem 3.3.9 (from [5, Lemma 6]) *For all probabilistic graph G,*

$$R_{SW0}^* = \lim_{n\to\infty} \frac{1}{n} H_\chi(G_{SW}^{\wedge n}). \tag{3.25}$$

The two distinct asymptotic expressions are equal to the optimal rate.

$$R_{SW0}^* = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \chi\Big(G^{\wedge n}[\mathcal{T}_\epsilon^n(P_V)]\Big) = \lim_{n\to\infty} \frac{1}{n} H_\chi(G_{SW}^{\wedge n}). \tag{3.26}$$

Proposition 3.3.10 (Full support) *If the distribution $P_{X,Y}$ has full support, then $R_{SW0}^* = H(X)$.*

The case of $P_{X,Y}$ with full support is a worst case for the zero-error regime, for which the optimal rate is $H(X)$, instead of $H(X|Y)$ in the vanishing error regime.

There is no known single letter expression for the optimal rate R_{SW0}^* except for a few special cases; in particular for perfect graphs and $\Big(C_5, \mathrm{Unif}(\{1, ..., 5\})\Big)$.

Chapter 4

Contributions for correction-based zero-error problems

In Section 4.1, we present the source coding problems considered in this chapter. We call these problems "correction-based" as they can be tackled with a correcting device which gives the zero-error property to coding schemes designed for the vanishing error regime. As a result, for this class of problems, the optimal rate in the zero-error regime coincides with the optimal rate in the vanishing error regime.

We determine in Section 4.2 the feasible rate region of the zero-error source coding problem when side information may be present at the decoder, presented in Section 4.1.1. In Section 4.3 we give an inner and an outer bound on the feasible rate region of the zero-error source coding problem for a Gray-Wyner network, presented in Section 4.1.3. We show that these bounds coincide in several cases of interest.

4.1 Correction-based zero-error problems

4.1.1 Zero-error source coding when side information may be present at the decoder

This scenario arises in interactive compression, where the user can randomly access part of the data directly in the compressed domain. A source sequence X^n models the smallest entity that can be requested, for instance a file of a database, a frame of a video, or a block of an omnidirectional image [8, 49, 53].

In this setting, the encoder has access to a side information but does not know whether the decoder has access to it; and the encoding is divided into two steps. In the first step, the encoder sends information without knowing whether side information is available at

the decoder. The message sent in the first step must be exploitable independently from the presence of the side information; furthermore, a zero-error decoding must be done if the decoder has access to the side information. In the second step, an information complement is sent by the encoder in the case where the decoder has no side information, and the decoder must be able to retrieve the source with zero-error. The two steps are modeled by two decoders and two noiseless channels.

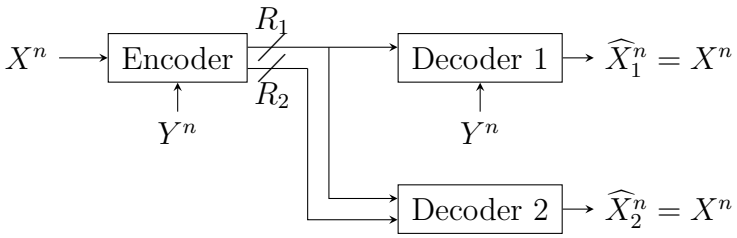


Figure 4.1 – Source coding when side information may be present.

Definition 4.1.1 *The setting of Figure 4.1 is described by:*

- *Two finite sets $\mathcal{X}$, $\mathcal{Y}$, and a source distribution $P_{X,Y} \in \Delta(\mathcal{X} \times \mathcal{Y})$;*
- *For all $n \in \mathbb{N}^\star$, (X^n, Y^n) is the random sequence of n copies of (X, Y), drawn in an i.i.d. fashion using $P_{X,Y}$.*
- *An encoder that knows (X^n, Y^n) sends binary strings over two noiseless channels to two decoders that want to retrieve X^n without error. Decoder 1 has Y^n as side information.*

A coding scheme in this setting is described by:

- *A time horizon $n \in \mathbb{N}^\star$,*
- *Two encoding functions and two decoding functions*

$$\phi_e^{(1)} : \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}^*, \quad \phi_e^{(2)} : \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}^*, \tag{4.1}$$

$$\phi_d^{(1)} : \{0,1\}^* \times \mathcal{Y}^n \to \mathcal{X}^n, \quad \phi_d^{(2)} : \left(\{0,1\}^*\right)^2 \to \mathcal{X}^n, \tag{4.2}$$

  such that $\operatorname{Im} \phi_e^{(1)}$ and $\operatorname{Im} \phi_e^{(2)}$ are prefix-free;

- *The rates over each channel are the average length of the codeword per source symbol:*

$$R_1 \doteq \frac{1}{n}\mathbb{E}[\ell \circ \phi_e^{(1)}(X^n, Y^n)] \qquad R_2 \doteq \frac{1}{n}\mathbb{E}[\ell \circ \phi_e^{(2)}(X^n, Y^n)]; \tag{4.3}$$

- n, $\phi_e^{(1)}$, $\phi_e^{(2)}$, $\phi_d^{(1)}$, $\phi_d^{(2)}$ *must satisfy the zero-error property:*

$$\mathbb{P}\Big(\widehat{X}_1^n \neq X^n\Big) = \mathbb{P}\Big(\widehat{X}_2^n \neq X^n\Big) = 0; \tag{4.4}$$

where

$$\widehat{X}_1^n = \phi_d^{(1)}\Big(\phi_e^{(1)}(X^n, Y^n), Y^n\Big), \tag{4.5}$$

$$\widehat{X}_2^n = \phi_d^{(2)}\Big(\phi_e^{(1)}(X^n, Y^n), \phi_e^{(2)}(X^n, Y^n)\Big). \tag{4.6}$$

The objective is to find the feasible rate region $\mathcal{R}_{SI}$, which is the closure of the set

$$\Big\{(R_1, R_2) \in \mathbb{R}^2 \;\Big|\; \exists\Big(n, \phi_e^{(1)}, \phi_e^{(2)}, \phi_d^{(1)}, \phi_d^{(2)}\Big) \text{ zero-error, with rates } (R_1, R_2)\Big\}. \tag{4.7}$$

A way to achieve zero-error coding in the setting of Figure 4.1 is to use conditional Huffman coding of X knowing Y; and send the source X to decoder 1 at rate $R_1 = H(X|Y)$. Then, to recover the source X, decoder 2 needs to obtain the side information Y, which requires a rate of $R_2 = H(Y) \geq I(X;Y)$.

In order to be exploitable by both decoders, part of the information sent through the common channel must be independent from Y. For this reason the setting of Figure 4.1 is closely related to the Slepian and Wolf (SW) problem in [64], seen as lossless source coding with side information at the decoder only. In this work, it is shown that the corner point $(R_1, R_2) = (H(X|Y), I(X;Y))$ is achievable in the vanishing probability of error regime. In [19], Csiszar proved in that linear codes achieve the optimal SW rate region. Several works in [13, 14, 73] investigate the duality between SW setting and channel coding using linear codes, as the side information Y can be seen as the output of a virtual channel with input X. However these tools cannot be straightforwardly adapted to the zero-error setting, as the linear codes proposed also present a vanishing probability of error.

The setting of Figure 4.1 can be seen as a zero-error variant with side informations known at the encoder of the successive refinement problem proposed by Kaspi in [35]; later generalized by Timo et al. in [67] for more than two decoders. Even if the lossy reconstruction of the source makes it fundamentally different from the zero-error setting, there are notable examples that present the same tools as in SW. The side information scalable source coding (i.e. the decoder 2 has a side information Y' s.t. $X \to Y \to Y'$) in [66] for instance uses nested random binning. This random binning approach was further developed in [2] to give a unified coding scheme that works for both scalable source coding

and Wyner-Ziv successive refinement in [65] (i.e. the decoder 2 has a side information Y' s.t. $X \to Y' \to Y$).

In [45], Ma and Cheng use linear codes in a zero-error SW restriction, under symmetry assumptions on the source. However, a zero-error SW coding scheme in our setting does not use at all the side information knowledge at the encoder. A more in-depth review of the literature for the zero-error SW setting can be found in Section 3.3.1.

In Section 4.2, we characterize the feasible rate region. More precisely, we show that the pair of rates $(R_1, R_2) = (H(X|Y), I(X;Y))$ is achievable in the zero-error regime and moreover, it is the corner-point of the feasible rate region.

4.1.2 Gray-Wyner problem

The problem built by Gray and Wyner in [30], aims at capturing the common information between two random variables X and Y.

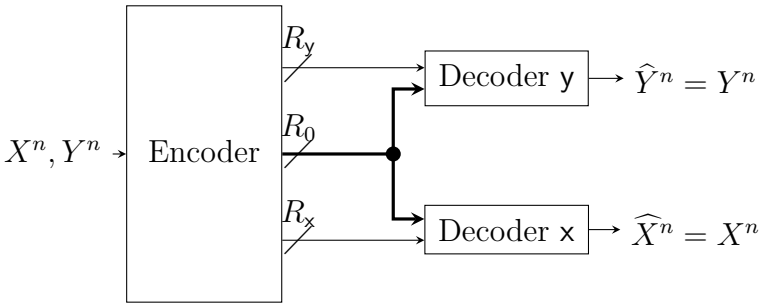


Figure 4.2 – The Gray-Wyner source coding problem.

Definition 4.1.2 *The setting of Figure 4.2 is described by:*

- *Two finite sets $\mathcal{X}$, $\mathcal{Y}$, and a source distribution $P_{X,Y} \in \Delta(\mathcal{X} \times \mathcal{Y})$;*
- *For all $n \in \mathbb{N}^\star$, (X^n, Y^n) is the random sequence of n copies of (X, Y), drawn in an i.i.d. fashion using $P_{X,Y}$.*
- *An encoder that knows (X^n, Y^n) sends binary strings over three noiseless channels to the decoder x (resp. y) that wants to retrieve X^n (resp. Y^n) without error.*

A coding scheme in this setting is described by:

- *A time horizon $n \in \mathbb{N}^\star$,*

- *Three encoding functions and two decoding functions*

$$\phi_e^{(0)}, \phi_e^{(\mathsf{x})}, \phi_e^{(\mathsf{y})} : \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}^*, \tag{4.8}$$

$$\phi_d^{(\mathsf{x})} : \left(\{0,1\}^*\right)^2 \to \mathcal{X}^n, \tag{4.9}$$

$$\phi_d^{(\mathsf{y})} : \left(\{0,1\}^*\right)^2 \to \mathcal{Y}^n, \tag{4.10}$$

such that $\operatorname{Im} \phi_e^{(i)}$ *is prefix-free for all* $i \in \{0, \mathsf{x}, \mathsf{y}\}$*;*

- *The rates over each channel are the average length of the codeword per source symbol:*

$$R_i \doteq \frac{1}{n}\mathbb{E}[\ell \circ \phi_e^{(i)}(X^n, Y^n)] \tag{4.11}$$

for all $i \in \{0, \mathsf{x}, \mathsf{y}\}$*;*

- n, $\left(\phi_e^{(i)}\right)_{i\in\{0,\mathsf{x},\mathsf{y}\}}$, $\left(\phi_d^{(j)}\right)_{j\in\{\mathsf{x},\mathsf{y}\}}$ *must satisfy the* ϵ*-error property:*

$$\mathbb{P}\left(\widehat{X}^n \neq X^n\right) \leq \epsilon \tag{4.12}$$

$$\mathbb{P}\left(\widehat{Y}^n \neq Y^n\right) \leq \epsilon; \tag{4.13}$$

where

$$\widehat{X}^n = \phi_d^{(\mathsf{x})}\left(\phi_e^{(0)}(X^n, Y^n), \phi_e^{(\mathsf{x})}(X^n, Y^n)\right), \tag{4.14}$$

$$\widehat{Y}^n = \phi_d^{(\mathsf{y})}\left(\phi_e^{(0)}(X^n, Y^n), \phi_e^{(\mathsf{y})}(X^n, Y^n)\right). \tag{4.15}$$

The objective is to find the feasible rate region $\mathcal{R}$*, which is the closure of the set*

$$\bigcap_{\epsilon>0} \left\{(R_0, R_\mathsf{x}, R_\mathsf{y}) \in \mathbb{R}^3 \,\middle|\, \exists n, \left(\phi_e^{(i)}\right)_{i\in\{0,\mathsf{x},\mathsf{y}\}}, \left(\phi_d^{(j)}\right)_{j\in\{\mathsf{x},\mathsf{y}\}} \epsilon\text{-error, with rates } (R_0, R_\mathsf{x}, R_\mathsf{y})\right\}. \tag{4.16}$$

In [30], Gray and Wyner have determined the feasible rate region in the vanishing error regime, as illustrated in Theorem 4.1.3.

Theorem 4.1.3 *The feasible rate region is the closure of the following set:*

$$\bigcup_{\substack{\mathcal{W} \text{ finite set,} \\ P_{W|X,Y} \in \Delta(\mathcal{W})^{|\mathcal{X}||\mathcal{Y}|}}} \left\{ (R_0, R_{\mathsf{x}}, R_{\mathsf{y}}) \in \mathbb{R}^3 \,\middle|\, R_0 \geq I(W;X,Y), R_{\mathsf{x}} \geq H(X|W), R_{\mathsf{y}} \geq H(Y|W) \right\}. \tag{4.17}$$

Following the same idea as in Remark 3.1.5, this region is also the feasible rate region in the zero-error regime (i.e. $\mathbb{P}(\widehat{X}^n \neq X^n) = \mathbb{P}(\widehat{Y}^n \neq Y^n) = 0$).

4.1.3 Zero-error source coding for a Gray-Wyner network when side information may be present at the decoder

The following setting is an extension of both problems presented in Section 4.1.2 and Section 4.1.1, in which there are two uncertainties: firstly, the encoder has access to two sources and their respective side information but does not know which one will be requested by the decoder; and secondly, the side information may be present at the decoder. The encoding is divided into three steps. In the first step, the encoder sends information without knowing which source is requested; we aim at sending the common information between the two sources, hence the "Gray-Wyner network" name for this problem. In the second step, the encoder knows which source is requested but does not know if the decoder has access to the side information; if the latter is present at the decoder, then a zero-error decoding must be done. The third step is an information complement sent by the encoder in the case where the decoder does not have the side information; and the decoder must be able to retrieve the source with zero-error.

Definition 4.1.4 *The setting of Figure 4.3 is described by:*

- *Four finite sets $\mathcal{U}$, $\mathcal{V}$, $\mathcal{X}$, $\mathcal{Y}$, and a source distribution $P_{U,V,X,Y} \in \Delta(\mathcal{U} \times \mathcal{V} \times \mathcal{X} \times \mathcal{Y})$;*
- *For all $n \in \mathbb{N}^\star$, (U^n, V^n, X^n, Y^n) is the random sequence of n copies of (U, V, X, Y), drawn in an i.i.d. fashion using $P_{U,V,X,Y}$.*
- *An encoder that knows (U^n, V^n, X^n, Y^n) sends binary strings over four noiseless channels to three decoders that want to retrieve X^n or Y^n without error. Decoder x_1 (resp. y) has U^n (resp. V^n) as side information.*

A coding scheme in this setting is described by:

- *A time horizon $n \in \mathbb{N}^\star$,*

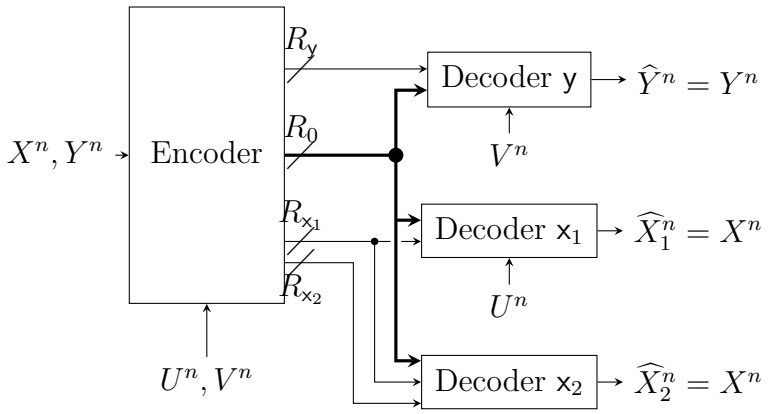


Figure 4.3 – Zero-error source coding for a Gray-Wyner network when side information may be present.

- *Four encoding functions and three decoding functions*

$$\phi_e^{(0)}, \phi_e^{(\mathsf{x}_1)}, \phi_e^{(\mathsf{x}_2)}, \phi_e^{(\mathsf{y})} : \mathcal{U}^n \times \mathcal{V}^n \times \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}^*, \tag{4.18}$$

$$\phi_d^{(\mathsf{x}_1)} : \left(\{0,1\}^*\right)^2 \times \mathcal{U}^n \to \mathcal{X}^n, \quad \phi_d^{(\mathsf{x}_2)} : \left(\{0,1\}^*\right)^3 \to \mathcal{X}^n, \tag{4.19}$$

$$\phi_d^{(\mathsf{y})} : \left(\{0,1\}^*\right)^2 \times \mathcal{V}^n \to \mathcal{Y}^n, \tag{4.20}$$

such that $\operatorname{Im} \phi_e^{(i)}$ *is prefix-free for all* $i \in \{0, \mathsf{x}_1, \mathsf{x}_2, \mathsf{y}\}$*;*

- *The rates over each channel are the average length of the codeword per source symbol:*

$$R_i \doteq \frac{1}{n}\mathbb{E}[\ell \circ \phi_e^{(i)}(U^n, V^n, X^n, Y^n)] \tag{4.21}$$

for all $i \in \{0, \mathsf{x}_1, \mathsf{x}_2, \mathsf{y}\}$*;*

- n, $\left(\phi_e^{(i)}\right)_{i\in\{0,\mathsf{x}_1,\mathsf{x}_2,\mathsf{y}\}}$, $\left(\phi_d^{(j)}\right)_{j\in\{\mathsf{x}_1,\mathsf{x}_2,\mathsf{y}\}}$ *must satisfy the zero-error property:*

$$\mathbb{P}\left(\widehat{X}_1^n \neq X^n\right) = \mathbb{P}\left(\widehat{X}_2^n \neq X^n\right) = \mathbb{P}\left(\widehat{Y}^n \neq Y^n\right) = 0; \tag{4.22}$$

where

$$\widehat{X}_1^n = \phi_d^{(1)}\left(\phi_e^{(0)}(S^n), \phi_e^{(\mathsf{x}_1)}(S^n), U^n\right), \tag{4.23}$$

$$\widehat{X}_2^n = \phi_d^{(2)}\left(\phi_e^{(0)}(S^n), \phi_e^{(\mathsf{x}_1)}(S^n), \phi_e^{(\mathsf{x}_2)}(S^n)\right), \tag{4.24}$$

$$\widehat{Y}^n = \phi_d^{(\mathsf{y})}\left(\phi_e^{(0)}(S^n), \phi_e^{(\mathsf{y})}(S^n), V^n\right), \tag{4.25}$$

and $S^n = (U^n, V^n, X^n, Y^n)$.

The objective is to find the feasible rate region $\mathcal{R}_{GW}$*, which is the closure of the set*

$$\Big\{(R_0, R_{\mathsf{x}_1}, R_{\mathsf{x}_2}, R_{\mathsf{y}}) \in \mathbb{R}^4 \,\Big| \tag{4.26}$$

$$\exists n, \left(\phi_e^{(i)}\right)_{i\in\{0,\mathsf{x}_1,\mathsf{x}_2,\mathsf{y}\}}, \left(\phi_d^{(j)}\right)_{j\in\{\mathsf{x}_1,\mathsf{x}_2,\mathsf{y}\}} \textit{ zero-error, with rates } (R_0, R_{\mathsf{x}_1}, R_{\mathsf{x}_2}, R_{\mathsf{y}})\Big\}. \tag{4.27}$$

This scenario can be likened to zero-error source coding problems with side information at the decoders, but not at the encoder. In such problems, the encoder has partial or no knowledge of the side information, and zero-error and vanishing error constraints do not lead to the same rate, in general. The zero-error constraint leads to the use of graph theoretical concepts for the characterization of the optimal rate. Witsenhausen first studied in [75] whether Slepian and Wolf's results can be extended to the zero-error case, using the chromatic number of the characteristic graph of the source distribution. In [62], Simonyi studied the generalization to more than one decoder with different side information, and proved that the optimal rate is the one associated to the worst side information (in the sense of Witsenhausen). Tuncel and Rose later extended this result to variable-length codes in [70]. A more in-depth review of the literature for the zero-error Slepian-Wolf setting can be found in Section 3.3.1.

A second class is the zero-error source coding problems with an encoder that has access to all side information, and decoders have access to the different side information. In these problems, the characterization of the optimal rates can be done using a packing lemma in type classes and the following correcting argument: the source sequence can be fully transmitted to each decoder whenever a decoding fails. This correcting argument allows for the extension of lossless results into zero-error results with same rates. A particular case of Sgarro's problem in [57] when side information are fully transmitted by the helpers can be extended to the zero-error case when the side information is also available at the encoder. In Section 4.2, we characterize the feasible rate region of a problem in this class presented in Section 4.1.1, where conditional coding does not achieve optimal rates.

Finally, the setting of Figure 4.3 is closely related to several problems in the literature, with different side information at the decoders and different desired source random variable, which have been studied under a vanishing error hypothesis. A covering lemma in type classes can be used to characterize the optimal rates in such problems, for exemple the lossless Gray and Wyner's (GW) problem [30]. In [68], Timo et al. study the GW setting with side information at the decoders only and give an inner bound on the

feasible rate region. In [7], Benammar and Zaidi consider a lossy GW setting with side information at the decoders only and with an hypothesis on the variables requested at the decoders. Laich and Wigger study in [43] the influence of the side information at the encoder in the lossless Kaspi/Heegard-Berger problem. They show that, for some source distributions, adding side information at the encoder lowers the minimal required rate to decode without loss.

In Section 4.3 we give an inner bound and an outer bound on the feasible rate region of the setting in Figure 4.3.

4.2 Main results for the zero-error source coding problem when side information may be present at the decoder

We now determine the feasible rate region of the zero-error source coding problem when side information may be present at the decoder, presented in Section 4.1.1

Theorem 4.2.1

$$\mathcal{R}_{SI} = \Big\{(R_1, R_2) \in \mathbb{R}^2 \;\Big|\; R_1 \geq H(X|Y),\ R_1 + R_2 \geq H(X)\Big\}. \tag{4.28}$$

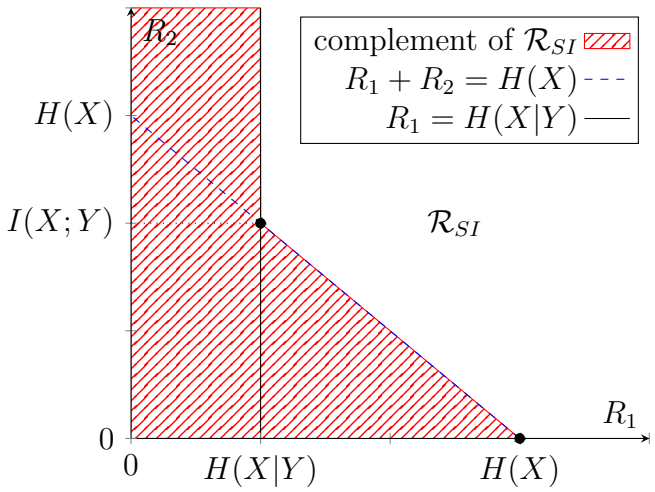


Figure 4.4 – Zero-error achievable rate region $\mathcal{R}_{SI}$.

Proof. [Converse of Theorem 4.2.1] In this setting, each decoder must retrieve X with zero-error. Using Shannon lossless source coding result [18, Theorem 5.3.1], we have $R_1 + R_2 \geq H(X)$.

Now let us prove $R_1 \geq H(X|Y)$

Let $M_0 \doteq \phi_e^{(1)}(S^n)$, where $S^n = (S_1, ..., S_n)$ is sequence of n iid copies of the source random variables (X, Y). Let $R_0 = \frac{1}{n}\mathbb{E}[\ell \circ \phi_e^{(1)}(S^n)]$.

Firstly, we can use Kraft inequality as the set $\operatorname{Im}\phi_e^{(1)}$ is prefix-free, we have

$$\sum_{w \in \operatorname{Im}\phi_e^{(1)}} 2^{-\ell(w)} \leq 1. \tag{4.29}$$

Let $\kappa \doteq \sum_{w \in \operatorname{Im}\phi_e^{(1)}} 2^{-\ell(w)}$, we have $\kappa \leq 1$ and $\left(\frac{2^{-\ell(w)}}{\kappa}\right)_{w \in \operatorname{Im}\phi_e^{(1)}}$ is a probability distribution.

Secondly,

$$\begin{aligned}
nR_0 &= \mathbb{E}[\ell \circ \phi_e^{(1)}(S^n)] && (4.30)\\
&= \sum_{w \in \operatorname{Im}\phi_e} \mathbb{P}(\phi_e^{(1)}(S^n) = w)\ell(w) && (4.31)\\
&= -\sum_{w \in \operatorname{Im}\phi_e} \mathbb{P}(\phi_e^{(1)}(S^n) = w) \log 2^{-\ell(w)} && (4.32)\\
&= -\log\kappa \; - \sum_{w \in \operatorname{Im}\phi_e^{(1)}} \mathbb{P}(\phi_e^{(1)}(S^n) = w) \log \frac{2^{-\ell(w)}}{\kappa} && (4.33)\\
&\geq -\sum_{w \in \operatorname{Im}\phi_e} \mathbb{P}(\phi_e^{(1)}(S^n) = w) \log \frac{2^{-\ell(w)}}{\kappa} && (4.34)\\
&\geq -\sum_{w \in \operatorname{Im}\phi_e} \mathbb{P}(\phi_e^{(1)}(S^n) = w) \log \mathbb{P}(\phi_e^{(1)}(S^n) = w) && (4.35)\\
&= H(\phi_e^{(1)}(S^n)) = H(M_0) && (4.36)
\end{aligned}$$

where (4.34) and (4.35) respectively come from $\kappa \leq 1$ and Gibbs inequality.

Therefore,

$$\begin{aligned} nR_0 &\geq H(M_0) && (4.37)\\ &\geq H(M_0|Y^n) && (4.38)\\ &= H(M_0, Y^n) - H(Y^n) && (4.39)\\ &= H(X^n, Y^n) - H(Y^n) && (4.40)\\ &= nH(X|Y), && (4.41) \end{aligned}$$

where (4.40) comes from the zero-error property. □

Our achievability result relies on a random coding argument. We use Csiszar and Körner's method of types [20, Chapter 2] in order to calibrate a linear code which is used to partition the set of source sequences. The encoder sends the coset of the source sequence to all decoders and the index of the source sequence in its coset to decoder 2. We show that the zero-error property is satisfied and the corresponding rates converge to the pair of target rates $(H(X|Y), I(X;Y))$.

In order to prove Theorem 4.2.1, we show that

$$\Big(H(X|Y), I(X;Y)\Big) \in \mathcal{R}_{SI}. \tag{4.42}$$

In order to complete the achievability result we use a time sharing with the point $\Big(H(X), 0\Big)$, which is known to be achievable by compressing X using a Huffman code and sending the resulting binary sequence via $\phi_e^{(1)}$.

4.2.1 Definitions for the achievability proof

Definition 4.2.2 (Type class, V-shell) *Let $Q_{X,Y} \in \Delta_n(\mathcal{X} \times \mathcal{Y})$, its type class is the set $\mathcal{T}_0^n(Q_{X,Y}) = \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n \mid T_{x^n,y^n} = Q_{X,Y}\}$. The $Q_{X|Y}$-shell of a sequence $y^n \in \mathcal{Y}^n$ is the set $\mathcal{T}_{Q_{X|Y}}(y^n) = \{x^n \in \mathcal{X}^n \mid T_{x^n,y^n} = Q_{X,Y}\}$.*

Definition 4.2.3 (Generator/parity matrix, syndrome, coset) *Let $\mathcal{A}$ be a finite set such that $|\mathcal{A}|$ is prime, so we can give $\mathcal{A} \simeq \mathbb{Z}/|\mathcal{A}|\mathbb{Z}$ a field structure. For all $n, k \in \mathbb{N}^\star$, we denote by $\mathcal{M}_{n,k}(\mathcal{A})$ the set of $n \times k$ matrices over the finite field $\mathcal{A}$.*

Let $k \in \mathbb{N}^\star$, a generator matrix is a matrix $\boldsymbol{G} \in \mathcal{M}_{n,k}(\mathcal{A})$. An associated parity matrix is a matrix $\boldsymbol{H} \in \mathcal{M}_{n-k,n}(\mathcal{A})$ such that Im $\boldsymbol{G}$ = Ker $\boldsymbol{H}$*, where* Im *and* Ker *denote the image and the kernel, respectively.*

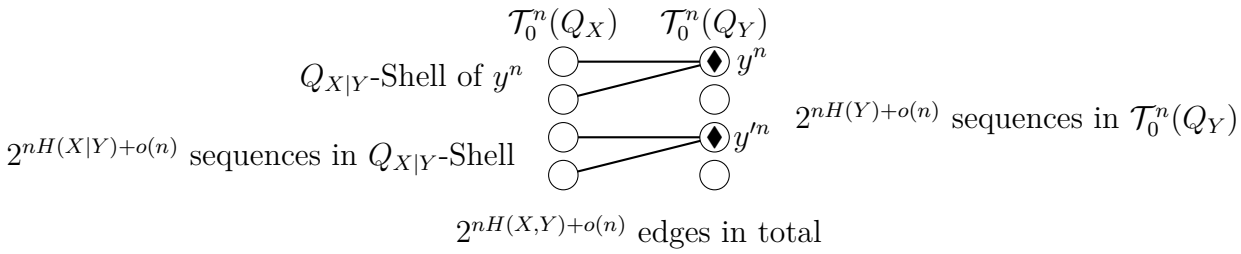


Figure 4.5 – An illustration of type classes and $Q_{X|Y}$-shells. Edges correspond to jointly typical sequences with $\epsilon = 0$, i.e. $T_{x^n,y^n} = Q_{X,Y}$. At most $2^{nI(X;Y)+o(n)}$ disjoint $Q_{X,Y}$-shells can be packed in $\mathcal{T}_0^n(Q_X)$, as $\frac{|\mathcal{T}_0^n(Q_Y)|}{|\mathcal{T}_{Q_{X|Y}}(y^n)|} = \frac{2^{nH(X)+o(n)}}{2^{nH(X|Y)+o(n)}} = 2^{nI(X;Y)+o(n)}$.

The syndrome of a sequence $a^n \in \mathcal{A}^n$ is $\boldsymbol{H}a^n$. The coset associated to the syndrome $\boldsymbol{H}a^n$ is the set $\operatorname{Im} \boldsymbol{G} + a^n = \{a'^n \in \mathcal{A}^n \mid \boldsymbol{H}a'^n = \boldsymbol{H}a^n\}$.

4.2.2 Coding scheme

For all $n \in \mathbb{N}^\star$, we show the existence of a sequence of $(n, R_1^{(n)}, R_2^{(n)})$-zero-error source codes that achieves the corner-point $\Big(H(X|Y), I(X;Y)\Big)$ of the zero-error rate region $\mathcal{R}_{SI}$. Our proof is based on a linear code adjusted depending on the random type T_{X^n,Y^n}, and coset partitioning of the Hamming space.

We assume w.l.o.g. that $P_{X,Y} \neq P_X \otimes P_Y$. We also assume w.l.o.g. that $|\mathcal{X}|$ is prime number by padding (i.e. extending with zeros) $P_{X,Y}$ if necessary. We fix the block-length n and a constant parameter $\delta \in (0; \log|\mathcal{X}| - H(X|Y))$ that will represent a rate penalty.

- Random code generation: For each pair of sequences (x^n, y^n), we define the parameter
$$k \doteq \left\lceil n - n\frac{H(T_{x^n,y^n}) - H(T_{y^n}) + \delta}{\log|\mathcal{X}|} \right\rceil^+. \tag{4.43}$$
where $\lceil\cdot\rceil$ denotes the ceiling function and $(\cdot)^+$ denotes $\max(\cdot, 0)$. We denote by K the random variable induced by k defined in (4.43), for the random sequences (X^n, Y^n). A generator matrix $\mathbf{G} \in \mathcal{M}_{n,n}(\mathcal{X})$ is randomly drawn, with i.i.d. entries drawn according to the uniform distribution on $\mathcal{X}$. If $k \neq 0$, let $\mathbf{G}_k$ be the matrix obtained by extracting the k first lines of $\mathbf{G}$, and $\mathbf{H}_k$ a parity matrix associated to $\mathbf{G}_k$.
The random code $\mathcal{C}$ consists of the set of random matrices $\mathcal{C} = (\mathbf{G}_k, \mathbf{H}_k)_{k \leq n}$. Before the transmission starts, a code realization is chosen and revealed to the encoder and both decoders.

- Encoding function $\phi_e^{(1)}$: Let $E \in \{0,1\}$ be such that $E = 0$ if $K \neq 0$ and $\Big(\operatorname{Im}\mathbf{G}_K + X^n\Big) \cap \mathcal{T}_{T_{X^n|Y^n}}(Y^n) = \{X^n\}$, $E = 1$ otherwise; where $T_{X^n|Y^n}$ is the conditional distribution obtained from T_{X^n,Y^n}. Then we define

$$\phi_e^{(1)}(X^n, Y^n) = \begin{cases} b(T_{X^n,Y^n}, E, \mathbf{H}_K X^n) & \text{if } E = 0, \\ b(T_{X^n,Y^n}, E, X^n) & \text{if } E = 1, \end{cases} \tag{4.44}$$

  where $b(\cdot)$ denotes the binary expansion.

- Encoding function $\phi_e^{(2)}$: If $E = 0$, the index of X^n in its coset $\operatorname{Im}\mathbf{G}_K + X^n$ is compressed using a Huffman code with the distribution P_{X^n}. Let $\iota(\mathbf{G}_K, X^n, Y^n)$ be the resulting binary sequence, then we set

$$\phi_e^{(2)}(X^n, Y^n) = \iota(\mathbf{G}_K, X^n, Y^n). \tag{4.45}$$

  Otherwise, $\phi_e^{(2)}(X^n, Y^n) = 0$.

- Decoding function $\phi_d^{(1)}$: It observes $\phi_e^{(1)}(X^n, Y^n)$ and extracts E and T_{X^n,Y^n}. If $E = 1$,

$$\phi_d^{(1)}(\phi_e^{(1)}(X^n, Y^n), Y^n) = X^n. \tag{4.46}$$

  Otherwise $E = 0$, it extracts $\mathbf{H}_K X^n$ and determines the coset $\operatorname{Im}\mathbf{G}_K + X^n$. Moreover, by using T_{X^n,Y^n} and Y^n it determines the $T_{X^n|Y^n}$-shell $\mathcal{T}_{T_{X^n|Y^n}}(Y^n)$, and therefore returns an element

$$\phi_d^{(1)}(\phi_e^{(1)}(X^n, Y^n), Y^n) \in \Big(\operatorname{Im}\mathbf{G}_K + X^n\Big) \cap \mathcal{T}_{T_{X^n|Y^n}}(Y^n).$$

- Decoding function $\phi_d^{(2)}$: It observes $\phi_e^{(1)}(X^n, Y^n)$ and extracts E and T_{X^n,Y^n}. If $E = 0$, it extracts $\mathbf{H}_K X^n$ and determines the coset $\operatorname{Im}\mathbf{G}_K + X^n$, and it returns $\phi_d^{(2)}(\phi_e^{(1)}(X^n, Y^n), \phi_e^{(2)}(X^n, Y^n)\Big)$, the element of $\operatorname{Im}\mathbf{G}_K + X^n$ with index $\phi_e^{(2)}(X^n, Y^n)$. If $E = 1$, it returns

$$\phi_d^{(2)}\Big(\phi_e^{(1)}(X^n, Y^n), \phi_e^{(2)}(X^n, Y^n)\Big) = X^n.$$

Remark 4.2.4 *The parameter K is selected so that when $K > 0$, the number of parity bits of the linear code asymptotically matches the conditional entropy.*

4.2.3 Zero-error property and rate analysis

We now prove that the code built in Section 4.2.2 satisfies the zero-error property. It is clear that both decoders retrieve X^n with zero-error when $E = 1$.

If $E = 0$, then by definition of E we have $(\operatorname{Im} \mathbf{G}_K + X^n) \cap \mathcal{T}_{T_{X^n|Y^n}}(Y^n) = \{X^n\}$, hence $\phi_d^{(1)}\left(\phi_e^{(1)}(X^n, Y^n), Y^n\right) = X^n$ with probability 1. On the other hand, $\phi_e^{(2)}(X^n, Y^n) = \iota(\mathbf{G}_K, X^n, Y^n)$, so the element of $\operatorname{Im} \mathbf{G}_K + X^n$ with index $\phi_e^{(2)}(X^n, Y^n)$ is X^n. Thus, $\phi_d^{(2)}(\phi_e^{(1)}(X^n, Y^n), \phi_e^{(2)}(X^n, Y^n)\big) = X^n$ with probability 1.

Lemma 4.2.5 (Rate analysis) *For all parameter $\delta > 0$, the sequence of rates of the codes built in Section 4.2.2 satisfy*

$$R_1^{(n)} \underset{n\to\infty}{\to} H(X|Y) + \delta, \qquad R_2^{(n)} \underset{n\to\infty}{\to} I(X;Y). \tag{4.47}$$

The proof can be found in Appendix A.1.

4.3 Main results for the zero-error source coding problem for a Gray-Wyner network when side information may be present at the decoder

We first give an inner bound on the feasible rate region in Theorem 4.3.1, based on the following coding strategy. For all realization (x^n, y^n) of the source, an auxiliary sequence w^n is used to capture the common information between the source sequences, and the sets of possible sequences x^n and y^n are partitioned into cosets. As the side information is available at some decoders, random binning is done so that less information is transmitted on w^n through the common channel. Decoding algorithms are based on V-shells of side information sequences, and the zero-error property is ensured by an error bit that is accompanied by all source sequences if set to 1. We also give an outer bound on the feasible rate region in Theorem 4.3.2, and we show in Section 4.3.1 that the inner and outer bound coincide in several cases of interest.

Theorem 4.3.1 (Inner bound) *The rate tuples* $(R_0, R_{\mathsf{x}_1}, R_{\mathsf{x}_2}, R_{\mathsf{y}})$ *that satisfy*

$$R_0 \geq I(U,V,X,Y;W) - \min\Big(I(U;W), I(V;W)\Big), \tag{4.48}$$
$$R_{\mathsf{x}_1} \geq H(X|U,W), \tag{4.49}$$
$$R_{\mathsf{x}_2} \geq I(X;U|W) + \min\Big(I(U;W), I(V;W)\Big), \tag{4.50}$$
$$R_{\mathsf{y}} \geq H(Y|V,W), \tag{4.51}$$

for some finite set $\mathcal{W}$ *and distribution* $P_{U,V,W,X,Y} = P_{U,V,X,Y} P_{W|U,V,X,Y}$ *are achievable.*

Theorem 4.3.2 (Outer bound) *The rate tuples* $(R_0, R_{\mathsf{x}_1}, R_{\mathsf{x}_2}, R_{\mathsf{y}})$ *that are achievable must satisfy*

$$R_0 \geq I(X,Y;W|U,V), \tag{4.52}$$
$$R_{\mathsf{x}_1} \geq H(X|U,W), \tag{4.53}$$
$$R_{\mathsf{y}} \geq H(Y|V,W), \tag{4.54}$$
$$R_0 + R_{\mathsf{x}_1} + R_{\mathsf{x}_2} \geq H(X), \tag{4.55}$$

for some finite set $\mathcal{W}$ *and distribution* $P_{U,V,W,X,Y} = P_{U,V,X,Y} P_{W|U,V,X,Y}$.

The proofs of Theorem 4.3.1 and Theorem 4.3.2 are respectively given in Appendix A.2 and in Appendix A.3. These proofs are based on the following Lemma 4.3.3 and Lemma 4.3.5. Lemma 4.3.3 is a covering/packing lemma for type classes. It is different from the packing lemma in [20, Lemma 10.1], as the latter states the existence of a family of codewords that satisfy the packing property, instead of drawing them at random following a given distribution.

Lemma 4.3.3 (Covering/Packing lemma in type classes) *Let* $k \in \mathbb{N}^\star$, $P_{A,B} \in \Delta_k(\mathcal{A} \times \mathcal{B})$, $P_{\bar{A}} \in \Delta(\mathcal{A})$, *and*

$$R_c \doteq I(A;B) + D(P_A \| P_{\bar{A}}). \tag{4.56}$$

Let $R > 0$ *such that* $R \neq R_c$. *For all* $n \in k\mathbb{N}^\star$, *let* $\mathcal{C}^{(n)} \doteq \left(\bar{A}^n_{[1]}, ..., \bar{A}^n_{[2^{nR}]}\right)$ *be a codebook of random sequences, drawn with a joint distribution that satisfies the marginal condition* $\bar{A}^n_{[i]} \sim P_{\bar{A}}^{\otimes n}$ *for all* $i \in \{1, \ldots, 2^{nR}\}$.

- *If $R < R_c$, then we have for all $b^n \in \mathcal{T}_0^n(P_B)$:*

$$\mathbb{P}\left(\exists a^n \in \mathcal{C}^{(n)},\ T_{a^n,b^n} = P_{A,B}\right) \leq 2^{-n(R_c-R)+o(n)}. \tag{4.57}$$

- *If $R > R_c$, assuming that the sequences in $\mathcal{C}^{(n)}$ are iid, we have for all $b^n \in \mathcal{T}_0^n(P_B)$:*

$$\begin{aligned}&\mathbb{P}\left(\exists a^n \in \mathcal{C}^{(n)},\ s.t.\ T_{a^n,b^n} = P_{A,B}\right) \\ &= 1 - \exp\left[-2^{n(R-R_c)+o(n)}\right].\end{aligned} \tag{4.58}$$

The proof of Lemma 4.3.3 can be found in Appendix A.4.

Remark 4.3.4 *For $P_{\bar{A}} = P_A$, we have $R_c = I(A;B)$. For $P_{\bar{A}} = \mathrm{Unif}(\mathcal{X})$, we have $R_c = \log|\mathcal{A}| - H(A|B)$.*

Lemma 4.3.5 (Coset partition) *Assume that $|\mathcal{A}|$ is prime. Let $k \in \mathbb{N}$, $P_{A,B} \in \Delta_k(\mathcal{A}\times\mathcal{B})$, $R_c = \log|\mathcal{A}| - H(A|B)$ and $R < R_c$ such that $kR/\log|\mathcal{A}| \in \mathbb{N}$. For all $n \in k\mathbb{N}$, let $\mathbf{G}^{(n)} \in \mathcal{M}_{n,nR/\log|\mathcal{X}|}(\mathcal{A})$ be a generator matrix whose entries are iid random variables drawn with the distribution $\mathrm{Unif}(\mathcal{A})$. Then we have for all $(a^n,b^n) \in \mathcal{T}_0^n(P_{A,B})$:*

$$\mathbb{P}\left(\exists \tilde{a}^n \in \mathrm{Im}\,\mathbf{G}^{(n)} + a^n \setminus \{a^n\},\ T_{\tilde{a}^n,b^n} = P_{A,B}\right) \leq 2^{-n(R_c-R)+o(n)}. \tag{4.59}$$

Proof. For all a^n, we have by construction that the coset $\mathrm{Ker}\,\mathbf{H}^{(n)} + a^n \setminus \{a^n\}$ is formed of $|\mathcal{A}|^{nR/\log|\mathcal{X}|} - 1 = 2^{nR} - 1$ random codewords, which are pairwise independent and identically distributed: they follow the distribution $\mathrm{Unif}(\mathcal{A})^{\otimes n}$. The result follows from Lemma 4.3.3. □

4.3.1 Comparison with previous results

In this section, we derive the zero-error achievable rate region for several special cases, in which our inner bound is optimal.

First, consider the zero-error variant of Gray-Wyner problem [30] by setting U and V constant, and removing the decoder x_2. Our inner bound allows to derive the zero-error rate achievable region for this problem:

$$R_0 \geq I(X,Y;W), \qquad R_{\mathsf{x}_1} \geq H(X|W), \tag{4.60}$$

$$R_{\mathsf{y}} \geq H(Y|W), \qquad R_{\mathsf{x}_2} \geq 0, \tag{4.61}$$

for some distribution $P_{U,V,W,X,Y} = P_{U,V,X,Y}P_{W|U,V,X,Y}$. Due to the converse from [30, Theorem 4], this bound is optimal.

Another scheme of interest is when Y and V are constant, which gives the problem presented in Section 4.1.1. We obtain the following

Inner bound:	Outer bound:	
$R_0 \geq I(X;W\|U),$	$R_0 \geq I(X;W\|U),$	(4.62)
$R_{\mathsf{x}_1} \geq H(X\|U,W),$	$R_{\mathsf{x}_1} \geq H(X\|U,W),$	(4.63)
$R_{\mathsf{x}_2} \geq I(U;X,W),$	$R_0 + R_{\mathsf{x}_1} + R_{\mathsf{x}_2} \geq H(X),$	(4.64)
$R_{\mathsf{y}} \geq 0,$	$R_{\mathsf{y}} \geq 0,$	(4.65)

for some distribution $P_{U,V,W,X,Y} = P_{U,V,X,Y}P_{W|U,V,X,Y}$. By choosing W constant (resp. $W = X$) we retrieve the achievability of the tuple $(R_0, R_{\mathsf{x}_1}, R_{\mathsf{x}_2}, R_{\mathsf{y}}) = (0, H(X|U), I(U;X), 0)$ (resp. $(H(X|U), 0, I(U;X), 0)$). It proves that the bound is optimal as the outer bound gives $R_0 + R_{\mathsf{x}_1} \geq H(X|U)$ and $R_0 + R_{\mathsf{x}_1} + R_{\mathsf{x}_2} \geq H(X)$.

We obtain the problem of Timo et al. [68] by removing the side information at the encoder, and removing the decoder x_2. Then the possible distributions $P_{U,V,W,X,Y}$ must satisfy the Markov chain $(U,V) \to (X,Y) \to W$, and we obtain the same inner bound as them:

$$R_0 \geq \max\Big(I(X,Y;W|U), I(X,Y;W|V)\Big), \tag{4.66}$$
$$R_{\mathsf{x}_1} \geq H(X|U,W), \tag{4.67}$$
$$R_{\mathsf{x}_2} \geq 0, \tag{4.68}$$
$$R_{\mathsf{y}} \geq H(Y|V,W), \tag{4.69}$$

for some distribution $P_{U,V,W,X,Y} = P_{U,V,X,Y}P_{W|X,Y}$.

Chapter 5

Contributions for graph-based zero-error problems

In the zero-error problems that we call "graph-based", zero-error constraints are captured by a characteristic graph, and determining the optimal rate requires to solve hard graph-theoretic questions. For instance, the zero-error Slepian-Wolf problem presented in Section 3.3.1 is one of them, its optimal rate is given by the complementary graph entropy $\overline{H}$ of the characteristic graph, and finding a single-letter formula for $\overline{H}$ is an open question. However, in order to understand better the nature of the zero-error Slepian-Wolf problem, we provide new structural results that link it with other problems, such as the zero-error capacity of a channel C_0 (see Figure 3.2).

Our approach for these structural results is the following. If the encoder is doing several independent compression tasks with their respective independent side information, then the optimal rate is given by $\overline{H}(\wedge\,\cdot)$, i.e. $\overline{H}$ of a graph with an AND product structure. As shown by Tuncel et al. in [71], "separating" the independent tasks yields an achievability scheme, hence $\overline{H}(\wedge\,\cdot) \leq \sum \overline{H}(\cdot)$. An important question is to determine whether this equation is satisfied with equality. When equality holds we say that "linearization" of $\overline{H}(\wedge\cdot)$ holds. Similarly, we define the linearization of optimal zero-error capacities $C(\cdot, P_V)$ and C_0. Although, linearization always holds in the vanishing error regime, it does not hold anymore in the zero-error regime.

Another interesting case of zero-error Slepian-Wolf problem is when the encoder has a partial information $g(Y)$ on the decoder's side information. The optimal rate is given by $\overline{H}(\sqcup\,\cdot)$, i.e. $\overline{H}$ of a graph with a disjoint union structure. "Separating" the source realizations w.r.t. the encoder's side information also yields an achievability scheme, hence $\overline{H}(\sqcup^{P_{g(Y)}}\cdot) \leq \sum_z P_{g(Y)}(z)\overline{H}(\cdot)$. Similarly, we question the "linearization" of $\overline{H}(\sqcup^{P_{g(Y)}}\cdot)$.

In this Chapter, we show the equivalences of the linearizations of the optimal source and channel zero-error rates $\overline{H}$, $C(\cdot, P_V)$ and C_0, when considering the AND product and the disjoint union of graphs, as depicted in Figure 5.1. More details are given about

existing results in Section 5.1, where we also present formally the linearization problems. The equivalence of linearizations between $\overline{H}$ and $C(\cdot, P_V)$ is proved in Section 5.2. In order to link the linearizations of C_0 and $C(\cdot, P_V)$, it is necessary to study the capacity-achieving distributions of a graph; the results needed are given in Section 5.3. In Section 5.4 we show the equivalence of linearizations between C_0 and $C(\cdot, P_V)$. In Section 5.5 we develop examples and counterexamples of linearization for C_0, $C(\cdot, P_V)$ and $\overline{H}$. The linearization problems and the zero-error channel coding problem with typical input constraint are formally presented in Section 5.1. The proofs of all results in this Chapter can be found in Appendix B.

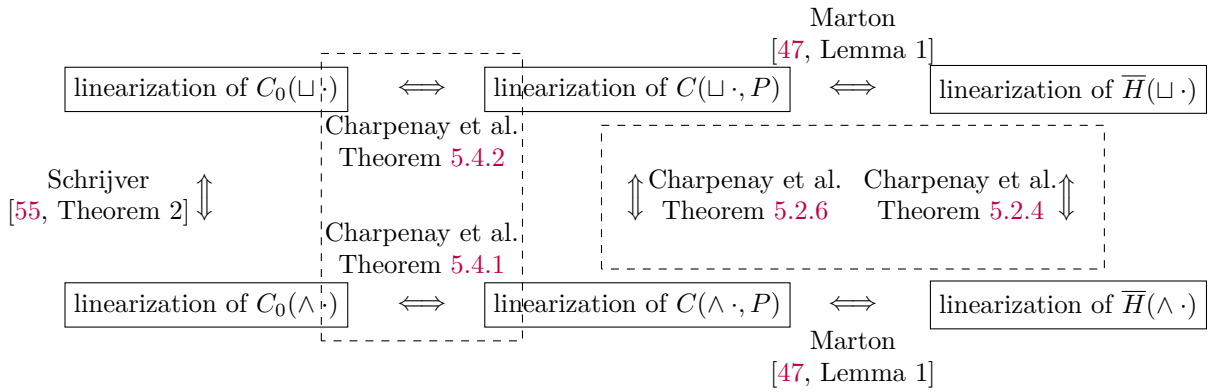


Figure 5.1 – Equivalences of linearizations between the zero-error capacity $C_0(\cdot)$, the zero-error capacity relative to a distribution $C(\cdot, P_V)$, and the complementary graph entropy $\overline{H}(\cdot)$. Our results are represented in the dashed rectangles.

5.1 Graph-based zero-error problems

5.1.1 Zero-error channel coding with typical input constraint

In the channel coding problem in the vanishing error regime, $I(X;Y)$ is the best rate one can achieve with codebooks formed of typical codewords w.r.t. P_X; and the channel capacity is the maximum of $I(X;Y)$ taken over P_X.

Similarly, we can think of a more constrained version of the zero-error channel coding problem, where the channel inputs have to be typical w.r.t. some distribution. The zero-error capacity of a graph relative to a distribution $C(\cdot, P_V)$ was introduced by Csiszar and Körner in [21].

Definition 5.1.1 (Zero-error capacity of a graph relative to P_V) *The zero-error capacity of a graph $G = (\mathcal{V}, \mathcal{E})$ relative to $P_V \in \Delta(\mathcal{V})$ is defined by*

$$C(G, P_V) \doteq \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \alpha\Big(G^{\wedge n}[\mathcal{T}_\epsilon^n(P_V)]\Big). \tag{5.1}$$

Remark 5.1.2 *We show in Lemma B.3.1 that the superior limit when $n \to \infty$ in the definition of $C(\cdot, P_V)$ can be replaced with a regular limit, thanks to the superadditivity of the sequence $\left(\frac{1}{n} \log \alpha\Big(G^{\wedge n}[\mathcal{T}_\epsilon^n(P_V)]\Big)\right)_{n \in \mathbb{N}^\star}$.*

We show in Lemma 5.1.3 that $C(\cdot, P_V)$ characterizes the optimal rate in the zero-error channel coding problem with typical inputs. We use this alternative characterization of $C(\cdot, P_V)$ to prove several other results, in particular for capacity achieving distributions. We define capacity achieving distributions in Section 5.3, and derive several results on them.

Lemma 5.1.3 *Let $G = (\mathcal{V}, \mathcal{E})$ be a graph and $P_V \in \Delta(\mathcal{V})$. Then there exists a sequence $(\mathcal{C}_n)_{n \in \mathbb{N}^\star}$ such that*

$$\forall n \in \mathbb{N}^\star,\ \mathcal{C}_n \subseteq \mathcal{V}^n \text{ is an independent set in } G^{\wedge n}, \tag{5.2}$$

$$\max_{v^n \in \mathcal{C}_n} \|T_{v^n} - P_V\|_\infty \underset{n \to \infty}{\to} 0, \tag{5.3}$$

$$\frac{\log |\mathcal{C}_n|}{n} \underset{n \to \infty}{\to} C(G, P_V). \tag{5.4}$$

Furthermore, any sequence $(\mathcal{C}_n)_{n \in \mathbb{N}^\star}$ that satisfies (5.2) and (5.3) also satisfies

$$\limsup_{n \to \infty} \frac{\log |\mathcal{C}_n|}{n} \leq C(G, P_V). \tag{5.5}$$

The proof of Lemma 5.1.3 is developed in Appendix B.3.1. Theorem 5.1.4 from Simonyi's survey [63, Theorem 20] states that C_0 is equal to the maximum of $C(\cdot, P_V)$, taken over the channel input distributions P_V, paralleling the channel capacity in the vanishing error regime.

Theorem 5.1.4 (from [63, Theorem 20]) *For all graph $G = (\mathcal{V}, \mathcal{E})$,*

$$C_0(G) = \max_{P_V \in \Delta(\mathcal{V})} C(G, P_V). \tag{5.6}$$

Theorem 5.1.4 is obtained from [63, Theorem 20] when the family of graph has only

one element. The result of [63, Theorem 20] is proved in [29, Sec. 2] for a more general setting that involves the Sperner capacity of directed graphs. For the sake of completeness, we provide a proof for Theorem 5.1.4 in Appendix B.3.2 that does not rely on directed graphs.

The complementary graph entropy is closely related to the zero-error capacity relative to a distribution. The link between these two quantities is provided by Marton in [47, Lemma 1], see Theorem 5.1.5 below. This formula is the cornerstone of our results that extend the properties of $\overline{H}$ to $C(G,\cdot)$. As stated in Corollary 5.1.6 which makes the link between $\overline{H}$ and C_0, a single letter formula for $\overline{H}$ would also yield a single-letter formula for C_0. It is worth noting that, similarly to C_0, $\overline{H}$ has a known single-letter expression for perfect graphs (see Section 5.5).

Theorem 5.1.5 (from [47, Lemma 1]) *For all graph $G=(\mathcal{V},\mathcal{E})$ and $P_V \in \Delta(\mathcal{V})$,*

$$C(G,P_V) + \overline{H}(G,P_V) = H(V). \tag{5.7}$$

Corollary 5.1.6 *For all graph $G=(\mathcal{V},\mathcal{E})$,*

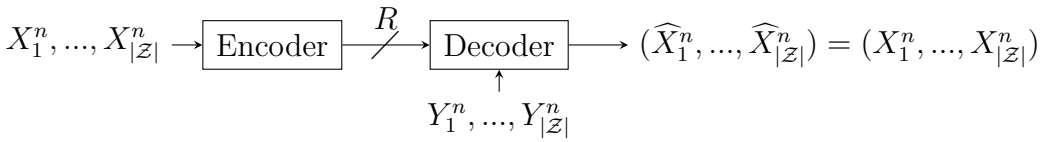
$$C_0(G) = \max_{P_V \in \Delta(\mathcal{V})} \Big(H(V) - \overline{H}(G,P_V) \Big). \tag{5.8}$$

We can interpret the formula in Theorem 5.1.5 the following way. The quantities $\overline{H}(G,P_V)$ and $C(G,P_V)$ are respectively the minimum number of colors, and the maximum size of an independent set. A color class (i.e. vertices of the same color) is an independent subset of vertices: in the case with same-sized color classes we would need $\log \alpha(G)$ bits to describe the source sequence in its color. Therefore, $C(G,P_V)$ can be seen as the information needed to describe the index of the source sequence in its color class. These two quantities sum up to $H(V)$, which is the information needed to describe the source sequence without loss; thus 5.1.5 can be seen as an analog for zero-error regime of the formula $I(X;Y)+H(X|Y)=H(X)$.

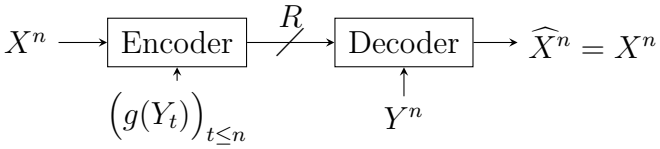
5.1.2 Linearization in the zero-error Slepian-Wolf problem

Let us consider the particular instances of zero-error Slepian-Wolf source coding depicted in Figure 5.2; where $g:\mathcal{Y}\to\mathcal{Z}$ is a deterministic function, $\mathcal{Z}$ is a finite set, and the pairs $((X_z,Y_z))_{z\in\mathcal{Z}}$ are mutually independent. They present a practical interest, as Figure 5.2a models the case where the encoder carries several independent compression

tasks with different respective distributions; and Figure 5.2b models the case where the encoder has partial information on the realizations of the decoder's side information.



(a) Zero-error Slepian-Wolf problem with several independent sources.



(b) Zero-error Slepian-Wolf problem with side information at the encoder.

Figure 5.2 – Two particular instances of zero-error Slepian-Wolf source coding problem.

As shown in Proposition 5.1.7 and Proposition 5.1.9, in these cases the optimal rate is the complementary graph entropy of a graph with a particular structure: disjoint union $\sqcup$, and AND product $\wedge$ respectively. Note that the setting in Figure 5.2b is a particular case of Figure 3.4, as it is equivalent to a setting with source $(X, g(Y))$ that must be retrieved by the decoder, and side information Y at the decoder; and also, Figure 3.4 is a particular case of Figure 5.2b by removing the encoder's side information.

Proposition 5.1.7 *The optimal rate in Figure 5.2a writes*

$$\overline{H}\left(\bigwedge_{z\in\mathcal{Z}} G_z\right); \tag{5.9}$$

where for all $z \in \mathcal{Z}$, G_z is the characteristic graph associated to the conditional distribution $P_{X_z|Y_z}$, with the underlying probability distribution P_{X_z} on its vertices.

Definition 5.1.8 (Disjoint union of probabilistic graphs $\sqcup$) *Let $\mathcal{A}$ be a finite set, and let $P_A \in \Delta(\mathcal{A})$. For all $a \in \mathcal{A}$, let $G_a = (\mathcal{V}_a, \mathcal{E}_a, P_{V_a})$ be probabilistic graphs, their disjoint union w.r.t. P_A is a probabilistic graph $(\mathcal{V}, \mathcal{E}, P_V)$ denoted by $\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a$ and defined by:*

- *$\mathcal{V} = \bigsqcup_{a\in\mathcal{A}} \mathcal{V}_a$ is the disjoint union of the sets $(\mathcal{V}_a)_{a\in\mathcal{A}}$;*
- *For all $v, v' \in \mathcal{V}$, $vv' \in \mathcal{E}$ iff they both belong to the same $\mathcal{V}_a$ and $vv' \in \mathcal{E}_a$;*
- *$P_V = \sum_{a\in\mathcal{A}} P_A(a) P_{V_a}$, note that the $(P_{V_a})_{a\in\mathcal{A}}$ have disjoint support in $\mathcal{V}$.*

The disjoint union of graphs without probability distribution has the vertex set and edges defined above, without underlying probability distribution. Therefore we do not specify a distribution P_A when considering a disjoint union of such graphs.

The disjoint union is also called "sum of graphs" in [71]. An example of disjoint union and AND product of probabilistic graphs can be found in Figure 5.3.

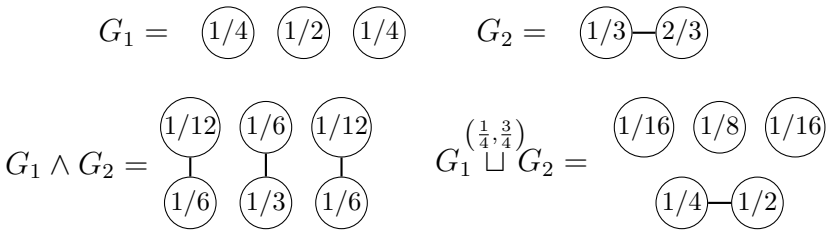


Figure 5.3 – An empty graph $G_1 = (N_3, (\frac{1}{4}, \frac{1}{2}, \frac{1}{4}))$ and a complete graph $G_2 = (K_2, (\frac{1}{3}, \frac{2}{3}))$, along with their AND product $G_1 \wedge G_2$ and their disjoint union $G_1 \sqcup G_2$ w.r.t. $(\frac{1}{4}, \frac{3}{4})$. The underlying distributions are represented by the numbers on each vertex.

Proposition 5.1.9 *The optimal rate in Figure 5.2b writes*

$$\overline{H}\left(\bigsqcup_{z\in\mathcal{Z}}^{P_{g(Y)}} G_z\right); \tag{5.10}$$

where for all $z \in \mathcal{Z}$, G_z is the characteristic graph associated to the conditional distribution $(P_{X|Y}(x,y))_{(x,y)\in\mathcal{X}\times g^{-1}(z)}$, with the underlying probability distribution $P_{X|g(Y)=z}$ on its vertices.

A natural coding strategy for the setting in Figure 5.2a consists in separating the problem into $|\mathcal{Z}|$ source coding sub-problems, where the random variable X_z has to be transmitted to the decoder that knows Y_z. Concatenating the optimal coding strategies in each sub-problem yields a zero-error coding scheme for the problem in Figure 5.2a, with rate $\sum_{z\in\mathcal{Z}} \overline{H}(G_z)$.

Definition 5.1.10 *The* linearization *with respect to the AND product $\wedge$ of the optimal rate $\overline{H}$ is satisfied when*

$$\overline{H}\left(\bigwedge_{z\in\mathcal{Z}} G_z\right) = \sum_{z\in\mathcal{Z}} \overline{H}(G_z). \tag{5.11}$$

The linearization *with respect to the disjoint union of graphs* $\bigsqcup$ *of the optimal rate* $\overline{H}$ *is satisfied when*

$$\overline{H}\left(\bigsqcup_{z\in\mathcal{Z}}^{P_{g(Y)}} G_z\right) = \sum_{z\in\mathcal{Z}} P_{g(Y)}(z)\overline{H}(G_z). \tag{5.12}$$

Similarly, a natural coding strategy for the setting in Figure 5.2b consists in separating the source realizations depending on the value of $g(Y)$, and use the optimal coding scheme designed for the distribution $P_{X,Y|g(Y)=z}$ for all z. The associated rate is $\sum_{z\in\mathcal{Z}} P_{g(Y)}(z)\overline{H}(G_z)$. When this coding strategy is optimal, the following holds:

$$\overline{H}\left(\bigsqcup_{z\in\mathcal{Z}}^{P_{g(Y)}} G_z\right) = \sum_{z\in\mathcal{Z}} P_{g(Y)}(z)\overline{H}(G_z), \tag{5.13}$$

i.e. $\overline{H}$ can be *linearized* w.r.t. $\sqcup$.

These natural coding schemes are both zero-error. For this reason, inequality always holds in (5.11) and (5.13), as captured in Tuncel et al.'s formulae in Theorem 5.1.11.

Theorem 5.1.11 (from [71, Theorem 2]) *For all probabilistic graphs* G, G' *and* $s \in [0,1]$,

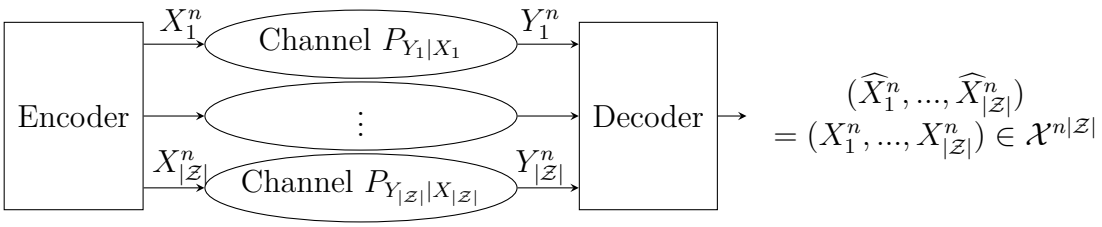
$$\overline{H}(G \wedge G') \leq \overline{H}(G) + \overline{H}(G'); \tag{5.14}$$

$$\overline{H}(G \overset{(s,1-s)}{\sqcup} G') \leq s\overline{H}(G) + (1-s)\overline{H}(G'). \tag{5.15}$$

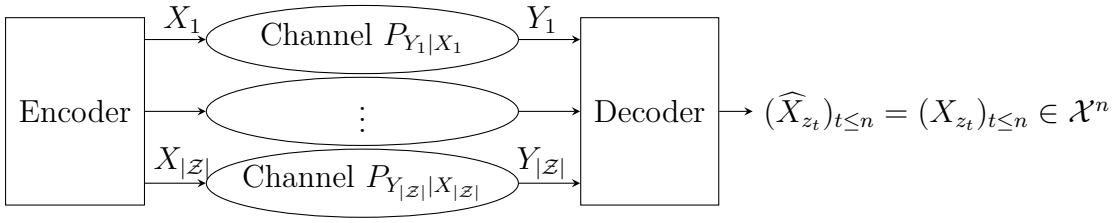
In the vanishing error regime, these natural approaches are always optimal. However in the zero-error regime they are not: we develop a counterexample in Theorem 5.5.13. We also show in Section 5.5 that for some classes of graphs (5.11) and (5.13) hold for all underlying distributions on the vertices. This enables us to derive a single-letter characterization of $\overline{H}$ for a new class of graphs. We will see in the following that the linearization of (5.14) is equivalent to the linearisation of (5.15).

5.1.3 Linearization in zero-error channel coding

In the zero-error channel coding setting, a case of interest is when encoder can use multiple independent channels to communicate with the decoder, depicted in Figure 5.4. We can think of two transmission regimes in this setting:



(a) Product of channels.



(b) Sum of channels.

Figure 5.4 – Two particular instances of zero-error channel coding problem.

- **Product of channels**: all the channels are used by the encoder at each time step, and the decoder observes all channel outputs.
- **Sum of channels**: at each time step $t \leq n$, the encoder has to use exactly one channel z_t among the $|\mathcal{Z}|$ channels. The decoder observes the chosen channel index and its output, and has to retrieve the input of the chosen channel.

In the vanishing error regime, the respective optimal rates for these settings depend on the channel capacities $C_z \doteq \max_{P_{X_z} \in \Delta(\mathcal{X})} I(X_z; Y_z)$: $\sum_{z\in\mathcal{Z}} C_z$ for the product of channels, and $\log\left(\sum_{z\in\mathcal{Z}} 2^{C_z}\right)$ for the sum of channels. However, in the zero-error regime, the optimal rates for the sum and product of channels are respectively given by the zero-error capacities of a product graph and a disjoint union graph; as shown in Proposition 5.1.12 and Proposition 5.1.13.

Proposition 5.1.12 (from [58]) *The optimal rate in Figure 5.4a writes*

$$C_0\left(\bigwedge_{z\in\mathcal{Z}} G_z\right); \tag{5.16}$$

where for all $z \in \mathcal{Z}$, G_z is the characteristic graph associated to the conditional distribution $P_{X_z|Y_z}$.

Proposition 5.1.13 (from [58]) *The optimal rate in Figure 5.4b writes*

$$C_0\left(\bigsqcup_{z\in\mathcal{Z}} G_z\right); \tag{5.17}$$

where for all $z \in \mathcal{Z}$, G_z is the characteristic graph associated to the conditional distribution $P_{X_z|Y_z}$.

A possible zero-error coding scheme in Figure 5.4 in the product of channels consists in separately using optimal codebooks, designed respectively for each channel. The associated rate is $\sum_{z\in\mathcal{Z}} C_0(G_z)$ and satisfies

$$C_0\left(\bigwedge_{z\in\mathcal{Z}} G_z\right) \geq \sum_{z\in\mathcal{Z}} C_0(G_z). \tag{5.18}$$

There are cases where equality holds in (5.18), i.e.

$$C_0\left(\bigwedge_{z\in\mathcal{Z}} G_z\right) = \sum_{z\in\mathcal{Z}} C_0(G_z), \tag{5.19}$$

which we call "linearization" of C_0 w.r.t. $\wedge$; and cases where inequality in (5.18) is strict, for instance the Schläfli graph and its complement (see Section 5.5.2).

For the sum of channels, we can use the optimal codebooks designed for each channel, with a time-sharing w.r.t. the following distribution:

$$P_Z^* \doteq \left(\frac{2^{C_0(G_z)}}{\sum_{z'\in\mathcal{Z}} 2^{C_0(G_{z'})}}\right)_{z\in\mathcal{Z}}. \tag{5.20}$$

The distribution P_Z^* represents the optimal time-sharing between the different codebooks, as we prove in Lemma B.4.5 that P_Z^* defined in (5.20) is the maximizer of the function $P_Z \mapsto H(P_Z) + \sum_{z\in\mathcal{Z}} P_Z(z)C_0(G_z)$.

The rate writes

$$H(P_Z^*) + \sum_{z\in\mathcal{Z}} P_Z^*(z)C_0(G_z) = \sum_{z\in\mathcal{Z}} P_Z^*(z)\left(\log\left(\frac{\sum_{z'\in\mathcal{Z}} 2^{C_0(G_{z'})}}{2^{C_0(G_z)}}\right) + C_0(G_z)\right) \tag{5.21}$$

$$= \log\left(\sum_{z'\in\mathcal{Z}} 2^{C_0(G_{z'})}\right). \tag{5.22}$$

and if this strategy is optimal, then the following holds:

$$C_0\left(\bigsqcup_{z\in\mathcal{Z}} G_z\right) = \log\left(\sum_{z\in\mathcal{Z}} 2^{C_0(G_z)}\right), \tag{5.23}$$

i.e. C_0 can be *linearized* w.r.t. $\sqcup$.

Definition 5.1.14 *The* linearization *with respect to the AND product $\wedge$ of the optimal rate C_0 is satisfied when*

$$C_0\left(\bigwedge_{z\in\mathcal{Z}} G_z\right) = \sum_{z\in\mathcal{Z}} C_0(G_z). \tag{5.24}$$

The linearization *with respect to the disjoint union of graphs of the optimal rate C_0 is satisfied when*

$$C_0\left(\bigsqcup_{z\in\mathcal{Z}} G_z\right) = \log\left(\sum_{z\in\mathcal{Z}} 2^{C_0(G_z)}\right). \tag{5.25}$$

Remark 5.1.15 *Linarization always holds in the vanishing error regime, in the sense that the optimal coding strategy for communicating over several independent channels consists in using respective optimal codebooks for each channel. In other words,*

$$\max_{P_{X_1,\dots,X_{|\mathcal{Z}|}}\in\Delta(\mathcal{X}^{|\mathcal{Z}|})} I(X_1,\dots,X_{|\mathcal{Z}|};Y_1,\dots,Y_{|\mathcal{Z}|}) = \sum_{z\in\mathcal{Z}} \max_{P_{X_z}\in\Delta(\mathcal{X})} I(X_z;Y_z). \tag{5.26}$$

Remark 5.1.16 *Note that P_Z^* is full-support: it can be observed $P_Z \mapsto H(P_Z)$ has an infinite slope at the frontier of $\Delta(\mathcal{Z})$, consequently the maximizer of $P_Z \mapsto H(P_Z) + \sum_{z\in\mathcal{Z}} P_Z(z)C_0(G_z)$ is always an interior point. In other words, the information carried by the channel index $H(P_Z)$ offsets the loss in rate, if the channels with smaller capacities are not chosen too often. Therefore, in the sum of channels setting, always choosing the channel with highest capacity is suboptimal; and never choosing a channel is also suboptimal, even if this channel has zero-error capacity* 0.

These natural coding schemes are both zero-error. For this reason, inequality always holds in (5.19) and (5.23), as captured in Shannon's formulae in Theorem 5.1.17.

Theorem 5.1.17 (from [58, Theorem 4]) *For all graphs G, G',*

$$C_0(G \sqcup G') \leq \log\left(2^{C_0(G)} + 2^{C_0(G')}\right); \tag{5.27}$$

$$C_0(G \wedge G') \leq C_0(G) + C_0(G'). \tag{5.28}$$

The Schläfli graph S is a counterexample of linearization of C_0 used by Haemers in [33] (see [15, Section 6.1] for an explicit construction), as stated in Theorem 5.1.18. To prove this result, Haemers uses a bound on the zero-error capacity based on the rank of the adjacency matrix of the graph. Refinements of this bound are developed by Bukh and Cox in [10], and by Gao et al. in [26]. Corollary 5.1.19 follows from Schrijver's result in Theorem 5.1.20.

Theorem 5.1.18 (from [33]) *Let S be the Schläfli graph and $\overline{S}$ its complementary graph, then $C_0(S \wedge \overline{S}) > C_0(S) + C_0(\overline{S})$.*

Corollary 5.1.19 $C_0(S \sqcup \overline{S}) > \log\left(2^{C_0(S)} + 2^{C_0(\overline{S})}\right)$.

In [36], Keevash and Long study the maximal value of $C_0(G \sqcup G')$ under the assumption $C_0(G), C_0(G') < \epsilon$.

As stated in Theorem 5.1.20, Schrijver has shown that the cases of equality in Theorem 5.1.17 coincide. In Section 5.2, we show the equivalence of the linearization for $\overline{H}$ and $C(\cdot, P_V)$, this requires new proof techniques. Furthermore, as we point out in Section 5.3 and Section 5.4, studying the capacity achieving distributions is necessary in order to link C_0 and $C(\cdot, P_V)$ and their linearizations.

Theorem 5.1.20 (from [55, Theorem 2]) *For all graphs G, G',*

$$C_0(G \sqcup G') = \log\left(2^{C_0(G)} + 2^{C_0(G')}\right) \tag{5.29}$$

$$\iff C_0(G \wedge G') = C_0(G) + C_0(G'). \tag{5.30}$$

In Section 5.4 and Section 5.2, we show that the graphs that allow for a linearization of C_0, $C(\cdot, P_V)$, and $\overline{H}$ are the same, when P_V is a capacity-achieving distribution.

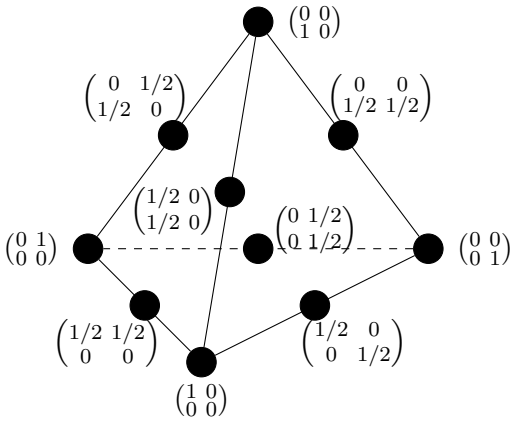


Figure 5.5 – The 2-discretized probability simplex $\Delta_2(\mathcal{X} \times \mathcal{Y})$ for $\mathcal{X} = \mathcal{Y} = \{0, 1\}$.

5.2 Main results on the linearization of $\overline{H}$ and $C(\cdot, P_V)$

In this Section we show the equivalence between the linearizations of $\overline{H}(\sqcup\, \cdot\,)$, $\overline{H}(\wedge\, \cdot\,)$, $C(\sqcup \cdot, P_V)$ and $\overline{H}(\wedge \cdot, P_V)$. In the following $\mathcal{A}$ is a finite set and $(G_a)_{a\in\mathcal{A}} = (\mathcal{V}_a, \mathcal{E}_a, P_{V_a})_{a\in\mathcal{A}}$ is an arbitrary finite family of probabilistic graphs.

A crucial result for several linearization proofs is Lemma 5.2.1, which give an expression for the complementary graph entropy of a disjoint union w.r.t. a type; the proof is given in Appendix B.2.1. The main reasons why $\wedge$ appears in (5.31) in Lemma 5.2.1 are the AND powers used in $\overline{H}$, and the distributivity of $\wedge$ w.r.t. $\sqcup$ (see Lemma B.2.4).

Lemma 5.2.1 *If $P_A \in \Delta_k(\mathcal{A})$ for some $k \in \mathbb{N}^\star$ then*

$$\overline{H}\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right) = \frac{1}{k}\overline{H}\left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge kP_A(a)}\right). \tag{5.31}$$

When $P_A = \mathrm{Unif}(\mathcal{A})$, we show in Corollary 5.2.2 that $\overline{H}(\sqcup\, \cdot\,)$ and $\overline{H}(\wedge\, \cdot\,)$ are equal up to a multiplicative constant.

Corollary 5.2.2

$$\overline{H}\left(\bigsqcup_{a\in\mathcal{A}}^{\mathrm{Unif}(\mathcal{A})} G_a\right) = \frac{1}{|\mathcal{A}|}\overline{H}\left(\bigwedge_{a\in\mathcal{A}} G_a\right). \tag{5.32}$$

Remark 5.2.3 *In Lemma 5.2.1, if $P_A \in \Delta_k(\mathcal{A})$, then $P_A \in \Delta_{ik}(\mathcal{A})$ for all $i \in \mathbb{N}^\star$.*

However, this does not change the expression (5.31) *as*

$$\frac{1}{ik}\overline{H}\left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge ikP_A(a)}\right) = \frac{1}{ik}\overline{H}\left(\left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge kP_A(a)}\right)^i\right) = \frac{1}{k}\overline{H}\left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge kP_A(a)}\right). \tag{5.33}$$

In Theorem 5.2.4, we show that the cases of equality coincide in Tuncel et al.'s inequalities in Theorem 5.1.11.

Theorem 5.2.4 *Let $P_A \in \Delta(\mathcal{A})$ with full-support, then the following equivalence holds:*

$$\overline{H}\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right) = \sum_{a\in\mathcal{A}} P_A(a)\overline{H}(G_a) \tag{5.34}$$

$$\iff \overline{H}\left(\bigwedge_{a\in\mathcal{A}} G_a\right) = \sum_{a\in\mathcal{A}} \overline{H}(G_a). \tag{5.35}$$

The proof of Theorem 5.2.4 is given in Appendix B.2.2. The key results of this proof are Lemma 5.2.1 which gives the equivalence between linearizations of $\overline{H}(\wedge\,\cdot)$ and $\overline{H}(\sqcup^{P_A}\cdot)$; and Lemma B.2.2, Lemma B.2.3 that extend this equivalence to all distributions P_A with full-support.

As a consequence of Marton's formula $\overline{H}(G) + C(G, P_V) = H(P_V)$, and the fact that the entropy $H(\cdot)$ always behaves additively w.r.t. independent random variables, we show with Proposition 5.2.5 the equivalence of linearizations between $\overline{H}$ and $C(\cdot, P_V)$. We also show superadditivity inequalities on $C(\wedge\,\cdot, P_V)$ and $C(\sqcup\,\cdot, P_V)$ based on Tuncel et al's subadditivity inequalities on $\overline{H}$ in Theorem 5.1.11. The proof of Proposition 5.2.5 is given in Appendix B.2.4.

The symmetric behaviors of $\overline{H}$ and $C(\cdot, P_V)$ lead to the equivalence between the linearizations of $C(\wedge\,\cdot, P_V)$ and $C(\sqcup\,\cdot, P_V)$ stated in Theorem 5.2.6, as a direct consequence of Theorem 5.2.4 and Proposition 5.2.5.

Proposition 5.2.5 *Let $P_A \in \Delta(\mathcal{A})$, then*

$$C\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a,\ \sum_{a\in\mathcal{A}} P_A(a)P_{V_a}\right) \geq H(P_A) + \sum_{a\in\mathcal{A}} P_A(a)C(G_a, P_{V_a}), \tag{5.36}$$

$$C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P_{V_a}\right) \geq \sum_{a\in\mathcal{A}} C(G_a, P_{V_a}); \tag{5.37}$$

and the following equivalences hold:

$$C\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a,\ \sum_{a\in\mathcal{A}} P_A(a)P_{V_a}\right) = H(P_A) + \sum_{a\in\mathcal{A}} P_A(a)C(G_a, P_{V_a}) \tag{5.38}$$

$$\iff \overline{H}\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right) = \sum_{a\in\mathcal{A}} P_A(a)\overline{H}(G_a); \tag{5.39}$$

$$C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P_{V_a}\right) = \sum_{a\in\mathcal{A}} C(G_a, P_{V_a}) \tag{5.40}$$

$$\iff \overline{H}\left(\bigwedge_{a\in\mathcal{A}} G_a\right) = \sum_{a\in\mathcal{A}} \overline{H}(G_a). \tag{5.41}$$

A natural question is about the equivalence between (5.38) and (5.40). In the next Theorem (which follows from Theorem 5.2.4), we show that this equivalence holds.

Theorem 5.2.6 *Let $P_A \in \Delta(\mathcal{A})$ with full-support, then the following equivalence holds:*

$$C\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a,\ \sum_{a\in\mathcal{A}} P_A(a)P_{V_a}\right) = H(P_A) + \sum_{a\in\mathcal{A}} P_A(a)C(G_a, P_{V_a}) \tag{5.42}$$

$$\iff C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P_{V_a}\right) = \sum_{a\in\mathcal{A}} C(G_a, P_{V_a}). \tag{5.43}$$

In this Section we proved that the equivalence of the linearisations between $\overline{H}$ and $C(\cdot, P_V)$ holds. In the next Sections, we investigate the linearisation of the zero-error capacity C_0.

5.3 Main results on capacity-achieving distributions

In this Section, we define the set of capacity-achieving distributions of a graph, which are the distributions P_V such that $C(\cdot, P_V)$ is equal to C_0. We give results on the capacity-achieving distributions when the graph is a product, and when the graph is vertex-transitive. Furthermore, understanding the capacity-achieving distributions is a necessary step towards the equivalence of linearization between C_0 and $C(\cdot, P_V)$, which is presented in Section 5.4. In the following, $\mathcal{A}$ is a finite set and $(G_a)_{a\in\mathcal{A}} = (\mathcal{V}_a, \mathcal{E}_a)_{a\in\mathcal{A}}$ is an arbitrary finite family of graphs.

Definition 5.3.1 (Set of capacity-achieving distributions $\mathcal{P}^*$) *Let $G = (\mathcal{V}, \mathcal{E})$ be a graph. The set of capacity-achieving distributions of G is the subset of $\Delta(\mathcal{V})$ denoted by $\mathcal{P}^*(G)$ and is defined by*

$$\mathcal{P}^*(G) \doteq \operatorname*{argmax}_{P_V \in \Delta(\mathcal{V})} C(G, P_V). \tag{5.44}$$

Proposition 5.3.2 *For all graph G, the mapping $P_V \mapsto C(G, P_V)$ is concave. The set of capacity-achieving distributions $\mathcal{P}^*(G)$ defined in Definition 5.3.1 is convex, nonempty, and satisfies*

$$\forall P_V \in \mathcal{P}^*(G),\ C_0(G) = C(G, P_V). \tag{5.45}$$

The proof of Proposition 5.3.2 is developed in Appendix B.3.3; the key result used in the proof is the formula $\max_{P_V} C(G, P_V) = C_0(G)$ from Theorem 5.1.4.

In Theorem 5.3.3, we show that if a joint distribution is capacity-achieving for a product of graphs, then the product of its marginals is also capacity-achieving. The proof of Theorem 5.3.3 is given in Appendix B.3.4, and is based on a codebook shifting argument: with a given zero-error codebook for $G \wedge G'$ with average type $P_{V,V'}$, one can build a shifted zero-error codebook (for an increased number of channel uses) with the same rate and average type $P_V \otimes P_{V'}$.

Theorem 5.3.3 *If $P_{V_1,\dots,V_{|\mathcal{A}|}} \in \mathcal{P}^*(\bigwedge_{a\in\mathcal{A}} G_a)$, then $\bigotimes_{a\in\mathcal{A}} P_{V_a} \in \mathcal{P}^*(\bigwedge_{a\in\mathcal{A}} G_a)$.*

Corollary 5.3.4

$$C_0\left(\bigwedge_{a\in\mathcal{A}} G_a\right) = \max_{\substack{(P_{V_a})_{a\in\mathcal{A}} \\ \in\prod_{a\in\mathcal{A}}\Delta(\mathcal{V}_a)}} C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P_{V_a}\right). \tag{5.46}$$

We show in Lemma 5.3.7 that the uniform distribution achieves the zero-error capacity for vertex-transitive graphs, i.e. a graph in which all the vertices play the same "role". The proof of Lemma 5.3.7 is given in Appendix B.3.5.

Definition 5.3.5 (Group of automorphisms Aut) *An automorphism of a graph $G = (\mathcal{V}, \mathcal{E})$ is a bijection $\psi : \mathcal{V} \to \mathcal{V}$ such that for all $v, v' \in \mathcal{V}$, $vv' \in \mathcal{E}$ if and only if $\psi(v)\psi(v') \in \mathcal{E}$. The group of automorphisms of G is denoted by $\mathrm{Aut}(G)$.*

Definition 5.3.6 (Vertex-transitive graph) *A graph $G = (\mathcal{V}, \mathcal{E})$ is vertex-transitive if* $\mathrm{Aut}(G)$ *acts transitively on its vertices, i.e. for all $v, v' \in \mathcal{V}$, there exists $\psi \in \mathrm{Aut}(G)$ such that $\psi(v) = v'$.*

Lemma 5.3.7 *If $G = (\mathcal{V}, \mathcal{E})$ is vertex-transitive, then*

$$\mathrm{Unif}(\mathcal{V}) \in \mathcal{P}^*(G). \tag{5.47}$$

Corollary 5.3.8 *Let $(G_a)_{a\in\mathcal{A}} = (\mathcal{V}_a, \mathcal{E}_a)_{a\in\mathcal{A}}$ be vertex-transitive graphs, their product is also vertex-transitive and*

$$\mathrm{Unif}\left(\prod_{a\in\mathcal{A}} \mathcal{V}_a\right) = \bigotimes_{a\in\mathcal{A}} \mathrm{Unif}(\mathcal{V}_a) \in \mathcal{P}^*\left(\bigwedge_{a\in\mathcal{A}} G_a\right). \tag{5.48}$$

5.4 Linking the linearizations of C_0 and $C(\cdot, P_V)$

In this Section, we show the equivalences of linearizations between C_0 and $C(\cdot, P_V)$. In the following, $\mathcal{A}$ is a finite set and $(G_a)_{a\in\mathcal{A}} = (\mathcal{V}_a, \mathcal{E}_a)_{a\in\mathcal{A}}$ is an arbitrary finite family of graphs.

Theorem 5.4.1 states that $C_0(\wedge\,\cdot\,)$ can be linearized if and only if $C(\wedge\,\cdot\,, P_V)$ can be linearized for some distribution P_V which is capacity-achieving for the product. When the linearisation of $C(\wedge\,\cdot\,, P_V)$ holds, we also show that the marginals of such a distribution are capacity-achieving for the respective graphs in the product. A similar result is derived for the disjoint union in Theorem 5.4.2; this result makes use of the optimal distribution $P_A(a) = \frac{2^{C_0(G_a)}}{\sum_{a'\in\mathcal{A}} 2^{C_0(G_{a'})}}$, which is the maximizer of $P_A \mapsto H(P_A) + \sum_{a\in\mathcal{A}} P_A(a) C_0(G_a)$ as stated in Lemma B.4.5. The proof of Theorem 5.4.1 and Theorem 5.4.2 are respectively given in Appendix B.4.1 and Appendix B.4.4.

Theorem 5.4.1 *The following holds*

$$C_0\left(\bigwedge_{a\in\mathcal{A}} G_a\right) = \sum_{a\in\mathcal{A}} C_0(G_a) \tag{5.49}$$

$$\iff \exists P_{V_1,\dots,V_{|\mathcal{A}|}} \in \mathcal{P}^*\left(\bigwedge_{a\in\mathcal{A}} G_a\right),\ C\left(\bigwedge_{a\in\mathcal{A}} G_a, \bigotimes_{a\in\mathcal{A}} P_{V_a}\right) = \sum_{a\in\mathcal{A}} C(G_a, P_{V_a}). \tag{5.50}$$

Furthermore, any distribution $\bigotimes_{a\in\mathcal{A}} P_{V_a} \in \mathcal{P}^*\left(\bigwedge_{a\in\mathcal{A}} G_a\right)$ that satisfies (5.50) also satisfies the following: $\forall a \in \mathcal{A},\ P_{V_a} \in \mathcal{P}^*(G_a)$.

Theorem 5.4.2 *The following equivalence holds*

$$C_0\left(\bigsqcup_{a\in\mathcal{A}} G_a\right) = \log\left(\sum_{a\in\mathcal{A}} 2^{C_0(G_a)}\right) \tag{5.51}$$

$$\iff \exists P_V \in \mathcal{P}^*\left(\bigsqcup_{a\in\mathcal{A}} G_a\right), \tag{5.52}$$

$$C\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a,\ \sum_{a\in\mathcal{A}} P_A(a)P_{V_a}\right) = H(P_A) + \sum_{a\in\mathcal{A}} P_A(a)C(G_a, P_{V_a}), \tag{5.53}$$

where $P_{V_a} = P_{V|V\in\mathcal{V}_a}$ *and* $P_A(a) = P_V(\mathcal{V}_a)$ *for all* $a \in \mathcal{A}$*. Furthermore, any* $\sum_{a\in\mathcal{A}} P_A(a)P_{V_a}$ *that satisfies* (5.53) *also satisfies the following for all* $a \in \mathcal{A}$*:*

$$P_A(a) = \frac{2^{C_0(G_a)}}{\sum_{a'\in\mathcal{A}} 2^{C_0(G_{a'})}},\ \text{and } P_{V_a} \in \mathcal{P}^*(G_a). \tag{5.54}$$

Remark 5.4.3 *One could think of a possible proof strategy for Theorem 5.4.2, which is successively using the equivalences in Theorem 5.1.20, Theorem 5.4.1, and Theorem 5.2.6. However, doing so yields the following statement*

$$C_0\left(\bigsqcup_{a\in\mathcal{A}} G_a\right) = \log\left(\sum_{a\in\mathcal{A}} 2^{C_0(G_a)}\right) \tag{5.55}$$

$$\iff \exists P_A \in \Delta(\mathcal{A})\ \textit{full-support},\ \exists P_{V_1,\dots,V_{|\mathcal{A}|}} \in \mathcal{P}^*\left(\bigwedge_{a\in\mathcal{A}} G_a\right), \tag{5.56}$$

$$C\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a,\ \sum_{a\in\mathcal{A}} P_A(a)P_{V_a}\right) = H(P_A) + \sum_{a\in\mathcal{A}} P_A(a)C(G_a, P_{V_a});$$

a missing step consists in linking $\mathcal{P}^*\left(\bigsqcup_{a\in\mathcal{A}} G_a\right)$ *with* $\mathcal{P}^*\left(\bigwedge_{a\in\mathcal{A}} G_a\right)$*.*

5.5 Main examples and counterexamples of linearization

5.5.1 Perfect graphs

In this Section, we show that perfect graphs allow for linearization of C_0, $C(\cdot, P_V)$ and $\overline{H}$ w.r.t. both $\sqcup$ and $\wedge$ with any underlying distribution. Perfect graphs are one of the only known examples of graphs with a single-letter formula for $\overline{H}$ and C_0, as stated in Theorem 5.5.4 and Theorem 5.5.5. We also give single-letter formulae for C_0, $C(\cdot, P_V)$ and $\overline{H}$ for products of perfect graphs, which are not perfect in general. Therefore, our

results yield new examples of single-letter formulae C_0, $C(\cdot, P_V)$ and $\overline{H}$ for cases where such formulae were previously unknown.

Definition 5.5.1 (Graph complement, clique number ω) *For all $G = (\mathcal{V}, \mathcal{E})$, the complementary graph of G is defined by $\overline{G} \doteq (\mathcal{V}, \mathcal{E}^c)$. The clique number of G is defined by $\omega(G) \doteq \alpha(\overline{G})$.*

Definition 5.5.2 (Perfect graph) *A graph $G = (\mathcal{V}, \mathcal{E})$ is perfect if $\forall \mathcal{S} \subseteq \mathcal{V}$, $\chi(G[\mathcal{S}]) = \omega(G[\mathcal{S}])$. A probabilistic graph $(\mathcal{V}, \mathcal{E}, P_V)$ is perfect if $(\mathcal{V}, \mathcal{E})$ is perfect.*

Definition 5.5.3 (Körner graph entropy H_κ) *For all $G = (\mathcal{V}, \mathcal{E}, P_V)$, let $\Gamma(G)$ be the collection of independent sets of vertices in G. The Körner graph entropy of G is defined by*

$$H_\kappa(G) = \min_{V \in W \in \Gamma(G)} I(W;V), \tag{5.57}$$

where the minimum is taken over all distributions $P_{W|V} \in \Delta(\mathcal{W})^{\mathcal{V}}$, with $\mathcal{W} = \Gamma(G)$ and with the constraint that the random vertex V belongs to the random independent set W with probability one, i.e. $V \in W \in \Gamma(G)$ in (5.57).

Theorem 5.5.4 (from [22, Corollary 12]) *Let G be a perfect probabilistic graph, then*

$$\overline{H}(G) = H_\kappa(G). \tag{5.58}$$

Shannon proved in [58, Theorem 3] that a graph G whose vertex set can be partitioned into $\alpha(G)$ cliques (i.e. complete induced subgraphs) satisfies $C_0(G) = \alpha(G)$. Perfect graphs are an example of that, as their complementary is also perfect, and satisfy $\chi(\overline{G}) = \omega(\overline{G}) = \alpha(G)$, where $\chi(\overline{G})$ is the clique cover number.

Theorem 5.5.5 (from [58, Theorem 3]) *If G is a perfect graph, then $C_0(G) = \log \alpha(G)$.*

As stated in Proposition 5.5.6, perfect graphs are an example of linearization for C_0: since $C_0(G \sqcup G') = \log \alpha(G \sqcup G') = \log(\alpha(G) + \alpha(G')) = \log(2^{C_0(G)} + 2^{C_0(G')})$ holds for all perfect G, G', we obtain $C_0(G \wedge G') = C_0(G) + C_0(G')$ by Schrijver's result in Theorem 5.1.20.

Proposition 5.5.6 *Let G and G' be perfect graphs, then*

$$C_0(G \sqcup G') = \log\left(2^{C_0(G)} + 2^{C_0(G')}\right) = \log(\alpha(G) + \alpha(G')); \tag{5.59}$$

$$C_0(G \wedge G') = C_0(G) + C_0(G') = \log \alpha(G) + \log \alpha(G'). \tag{5.60}$$

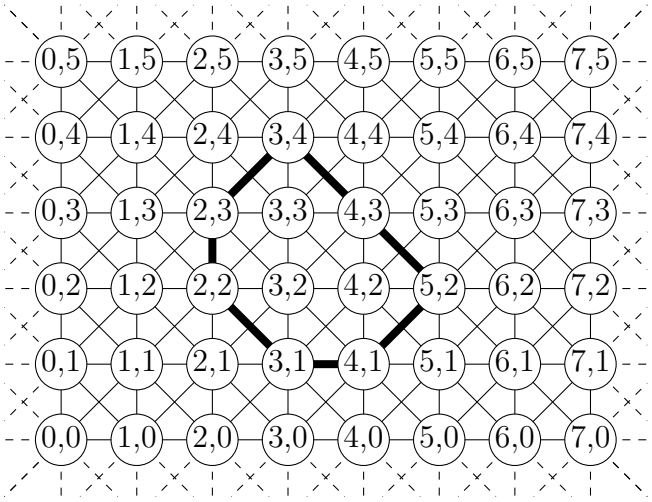


Figure 5.6 – A non-perfect AND product of perfect graphs: $C_6 \wedge C_8$ with an induced C_7.

AND products of perfect graphs are not perfect in general; see for example Figure 5.6 where a product of two perfect graphs contains an induced C_7, which makes it non-perfect by Theorem 5.5.7. However, a disjoint union of perfect graphs is always perfect, as stated in Lemma B.5.2; therefore the results on equivalence of linearizations are useful to extend the linearization properties and single-letter expressions to the AND product of perfect graphs. In Theorem 5.5.8 we show that perfect graphs allow for the linearization of $C(\cdot, P_V)$ and $\overline{H}$, for all underlying probability distributions; and we also give a single-letter expression for $\overline{H}$ and $C(\cdot, P_V)$ in that case. The proof of Theorem 5.5.8 is given in Appendix B.5.1.

Theorem 5.5.7 (Strong perfect graph theorem, from [16, Theorem 1.2]) *A graph G is perfect if and only if neither G nor $\overline{G}$ have an induced odd cycle of length at least 5.*

Theorem 5.5.8 *When $(G_a)_{a\in\mathcal{A}} = (\mathcal{V}_a, \mathcal{E}_a, P_{V_a})_{a\in\mathcal{A}}$ is a family of perfect probabilistic*

graphs, the following single-letter characterizations hold:

$$\overline{H}\left(\bigwedge_{a\in\mathcal{A}} G_a\right) = \sum_{a\in\mathcal{A}} \overline{H}(G_a) = \sum_{a\in\mathcal{A}} H_\kappa(G_a), \tag{5.61}$$

$$\overline{H}\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right) = \sum_{a\in\mathcal{A}} P_A(a)\overline{H}(G_a) = \sum_{a\in\mathcal{A}} P_A(a) H_\kappa(G_a), \tag{5.62}$$

$$C\left(\bigwedge_{a\in\mathcal{A}} G_a, \bigotimes_{a\in\mathcal{A}} P_{V_a}\right) = \sum_{a\in\mathcal{A}} C(G_a, P_{V_a}) = \sum_{a\in\mathcal{A}} \Big(H(V_a) - H_\kappa(G_a)\Big), \tag{5.63}$$

$$\begin{aligned} C\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a, \sum_{a\in\mathcal{A}} P_A(a) P_{V_a}\right) &= H(P_A) + \sum_{a\in\mathcal{A}} P_A(a) C(G_a, P_{V_a}) \\ &= H(P_A) + \sum_{a\in\mathcal{A}} P_A(a)\Big(H(V_a) - H_\kappa(G_a)\Big). \end{aligned} \tag{5.64}$$

Another interesting example is the graph $C_5 \sqcup G$ where G is perfect, which is an example of linearization of $\overline{H}$ with single-letter formula developed by Tuncel et al. in [71]. The pentagon graph C_5 is not perfect, and makes non-perfect any disjoint union or AND product that is made with it. However we can use Theorem 5.2.4, and derive in Corollary 5.5.10 another non-perfect example of linearization with single-letter formula for $\overline{H}$: $C_5 \wedge G$ with G perfect.

Theorem 5.5.9 (from [71, Lemma 3]) *Let $s \in [0,1]$, let G be a perfect probabilistic graph, and let $G_5 \doteq (C_5, \mathrm{Unif}(\{0,...,4\}))$, we have*

$$\overline{H}(G_5 \overset{(s,1-s)}{\sqcup} G) = s\overline{H}(G_5) + (1-s)\overline{H}(G) \tag{5.65}$$

$$= \frac{s}{2}\log 5 + (1-s)H_\kappa(G). \tag{5.66}$$

Corollary 5.5.10 *For all perfect probabilistic graph G,*

$$\overline{H}(G \wedge G_5) = \overline{H}(G) + \overline{H}(G_5) = H_\kappa(G) + \tfrac{1}{2}\log 5. \tag{5.67}$$

5.5.2 The Schläfli graph

In order to use the Schläfli graph S as a counterexample for $C(\cdot, P_V)$ and $\overline{H}$, we need a capacity-achieving distribution of $S \wedge \overline{S}$. As stated in Lemma 5.5.11, the Schläfli graph

is vertex transitive. By Lemma 5.3.7 and Corollary 5.5.12, the uniform distribution is capacity-achieving for S, $\overline{S}$, and $S \wedge \overline{S}$.

Lemma 5.5.11 (from [11, Lemma 3.7]) *The Schläfli graph is vertex-transitive.*

Corollary 5.5.12 *The sets $\mathcal{P}^*(S)$, $\mathcal{P}^*(\overline{S})$, and $\mathcal{P}^*(S \wedge \overline{S})$ all contain the uniform distribution.*

In Theorem 5.5.13 we adapt the Schläfli graph counterexample to $C(\cdot, P_V)$ and $\overline{H}$ by using our results from the previous sections.

Theorem 5.5.13 *Let $s \in (0,1)$, let $S = (\mathcal{V}_S, \mathcal{E}_S, \mathrm{Unif}(\mathcal{V}_S))$ be the Schläfli graph and let $\overline{S} = (\mathcal{V}_S, \mathcal{E}_{\overline{S}}, \mathrm{Unif}(\mathcal{V}_S))$ be the complementary of the the Schläfli graph with uniform distribution on their vertices. Then*

$$C(S \wedge \overline{S}, \mathrm{Unif}(\mathcal{V}_S) \otimes \mathrm{Unif}(\mathcal{V}_{\overline{S}})) > C(S, \mathrm{Unif}(\mathcal{V}_S)) + C(\overline{S}, \mathrm{Unif}(\mathcal{V}_{\overline{S}})), \tag{5.68}$$

$$\begin{aligned} C(S \sqcup \overline{S}, s\,\mathrm{Unif}(\mathcal{V}_S) + (1-s)\,\mathrm{Unif}(\mathcal{V}_{\overline{S}})) &> h_b(s) + sC(S, \mathrm{Unif}(\mathcal{V}_S)) \\ &\quad + (1-s)C(\overline{S}, \mathrm{Unif}(\mathcal{V}_{\overline{S}})), \end{aligned} \tag{5.69}$$

$$\overline{H}(S \wedge \overline{S}) < \overline{H}(S) + \overline{H}(\overline{S}), \tag{5.70}$$

$$\overline{H}(S \overset{(s,1-s)}{\sqcup} \overline{S}) < s\overline{H}(S) + (1-s)\overline{H}(\overline{S}); \tag{5.71}$$

where h_b is the binary entropy.

We obtain the first inequality from Theorem 5.4.1 and Corollary 5.5.12; the second one from Theorem 5.2.6; the third one comes from Proposition 5.2.5; and the last one from Theorem 5.2.4.

Remark 5.5.14 *Alon has built in [4] infinite families of graphs that satisfy $C_0(G \sqcup G') > \log(2^{C_0(G)} + 2^{C_0(G')})$. Similar results as in Theorem 5.5.13 can be derived for these graphs, by using their respective capacity-achieving distributions.*

Chapter 6

Contributions for coding for computing zero-error problems

In this Chapter, we study the zero-error coding for computing problem with side information at the encoder, presented in Section 6.1. We fist give an asymptotic expression for the optimal rate in the general case, then we formulate an hypothesis on $P_{X,Y}$ and g that we call "pairwise shared side information" that allows us to derive a single-letter characterization of the optimal rate. In particular it covers the cases with $P_{X,Y}$ full-support, without any assumption on f, g.

This hypothesis is satisfied if every pair of source symbols "share" at least one side information symbol for all output of g. It has graph-theoretic interpretations, as the single-letter formula stems from the particular structure of the characteristic graph of the problem: a disjoint union of OR products. We also prove that this condition is equivalent to the worst optimal rate in an auxiliary Slepian-Wolf problem.

6.1 Coding for computing problems

6.1.1 Coding for computing in the vanishing error regime

The coding for computing problem is defined in [50] by Orlitsky and Roche. In this setting the decoder wants to retrieve a function of both encoder's and decoder's data.

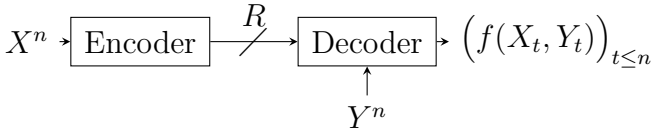


Figure 6.1 – Coding for computing.

Definition 6.1.1 *The source coding problem of Figure 6.2 is described by:*

- *Finite sets* $\mathcal{U}$, $\mathcal{X}$, $\mathcal{Y}$ *and a source distribution* $P_{X,Y} \in \Delta(\mathcal{X} \times \mathcal{Y})$;
- *For all* $n \in \mathbb{N}^\star$, (X^n, Y^n) *is the random sequence of* n *copies of* (X, Y), *drawn in an i.i.d. fashion using* $P_{X,Y}$.
- *Two deterministic functions*

$$f : \mathcal{X} \times \mathcal{Y} \to \mathcal{U}. \tag{6.1}$$

- *An encoder that knows* X^n *sends binary strings over a noiseless channel to a decoder that knows* Y^n, *and that wants to retrieve* $\Big(f(X_t, Y_t)\Big)_{t \leq n}$ *without error.*

A coding scheme in this setting is decribed by:

- *A time horizon* $n \in \mathbb{N}^\star$, *and an encoding function* $\phi_e : \mathcal{X}^n \to \{0,1\}^*$ *such that* $\operatorname{Im} \phi_e$ *is prefix-free;*
- *A decoding function* $\phi_d : \mathcal{Y}^n \times \{0,1\}^* \to \mathcal{U}^n$;
- *The rate is the average length of the codeword per source symbol, i.e.* $R \doteq \frac{1}{n}\mathbb{E}\Big[\ell \circ \phi_e\Big(X^n\Big)\Big]$;
- n, ϕ_e, ϕ_d *must satisfy the* ϵ*-error property:*

$$\mathbb{P}\left(\phi_d\Big(Y^n, \phi_e\Big(X^n\Big)\Big) \neq \Big(f(X_t, Y_t)\Big)_{t \leq n}\right) \leq \epsilon. \tag{6.2}$$

The objective is to find the minimal rate among all coding schemes under the vanishing error constraint:

$$R^*_{CFC} \doteq \lim_{\epsilon \to 0} \inf_{\substack{n, \phi_e, \phi_d \\ \epsilon\text{-error}}} \frac{1}{n}\mathbb{E}\Big[\ell \circ \phi_e\Big(X^n, (g(Y_t))_{t \leq n}\Big)\Big]. \tag{6.3}$$

As illustrated in Theorem 6.1.4, Orlitsky and Roche give in [50] a single-letter expression of the optimal rate. This optimal rate is characterized with a characteristic graph G_{CFC} defined below, and a conditional version of Körner's graph entropy (the latter is defined in Definition 5.5.3).

Definition 6.1.2 (Conditional Körner graph entropy) *For all* $G = (\mathcal{V}, \mathcal{E}, P_V)$, *let* $\Gamma(G)$ *be the collection of independent sets of vertices in* G. *Let* $P_{Y|V} \in \Delta(\mathcal{Y})^{|\mathcal{V}|}$, *the*

conditional Körner graph entropy of G is defined by

$$H_\kappa(G|Y) = \min_{\substack{W \to V \to Y \\ V \in W \in \Gamma(G)}} I(W;V|Y), \tag{6.4}$$

where the minimum is taken over all distributions $P_{W|V,Y} \in \Delta(\mathcal{W})^{\mathcal{V}}$, with $W \to V \to Y$ (i.e. $P_{W|V,Y} = P_{W|V}$), $\mathcal{W} = \Gamma(G)$ and with the constraint that the random vertex V belongs to the random independent set W with probability one, i.e. $V \in W \in \Gamma(G)$ in (6.4).

Definition 6.1.3 (Characteristic graph G_{CFC}) *The auxiliary graph G_{CFC} is defined by*

- *$\mathcal{X}$ as set of vertices with distribution P_X,*
- *xx' are adjacent if $f(x,y) \neq f(x',y)$ for some $y \in \operatorname{supp} P_{Y|X=x} \cap \operatorname{supp} P_{Y|X=x'}$.*

Theorem 6.1.4 (from [50])

$$R^*_{CFC} = H_\kappa(G_{CFC}|Y).$$

6.1.2 Zero-error coding for computing with side information at the encoder

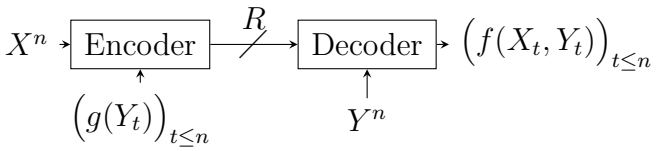


Figure 6.2 – Zero-error coding for computing with side information at the encoder.

The problem of Figure 6.2 is a zero-error setting that relates to Orlitsky and Roche's coding for computing problem from [50]. This coding problem appears in video compression [23, 27], where X^n models a set of images known at the encoder. The decoder does not always want to retrieve each image, but has instead a sequence Y^n of particular requests for each image, e.g. detection: cat, dog, car, bike; or scene recognition: street/city/mountain, etc... The encoder does not know the decoder's exact request but has prior information about it (e.g. type of request), which is modeled by $(g(Y_t))_{t\leq n}$. This problem also relates to the zero-error Slepian-Wolf open problem presented in Section 3.3.1, as it is obtained as a special case by taking g constant and $f(X,Y) = X$.

Definition 6.1.5 *The zero-error source coding problem of Figure 6.2 is described by:*

- *Four finite sets $\mathcal{U}$, $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ and a source distribution $P_{X,Y} \in \Delta(\mathcal{X} \times \mathcal{Y})$;*
- *For all $n \in \mathbb{N}^\star$, (X^n, Y^n) is the random sequence of n copies of (X, Y), drawn in an i.i.d. fashion using $P_{X,Y}$.*
- *Two deterministic functions*

$$f : \mathcal{X} \times \mathcal{Y} \to \mathcal{U}, \tag{6.5}$$
$$g : \mathcal{Y} \to \mathcal{Z}. \tag{6.6}$$

- *An encoder that knows X^n and $\Big(g(Y_t)\Big)_{t\leq n}$ sends binary strings over a noiseless channel to a decoder that knows Y^n, and that wants to retrieve $\Big(f(X_t, Y_t)\Big)_{t\leq n}$ without error.*

A coding scheme in this setting is decribed by:

- *A time horizon $n \in \mathbb{N}^\star$, and an encoding function $\phi_e : \mathcal{X}^n \times \mathcal{Z}^n \to \{0,1\}^*$ such that $\operatorname{Im} \phi_e$ is prefix-free;*
- *A decoding function $\phi_d : \mathcal{Y}^n \times \{0,1\}^* \to \mathcal{U}^n$;*
- *The rate is the average length of the codeword per source symbol, i.e. $R \doteq \frac{1}{n}\mathbb{E}\Big[\ell \circ \phi_e\Big(X^n, (g(Y_t))_{t\leq n}\Big)\Big]$;*
- *n, ϕ_e, ϕ_d must satisfy the zero-error property:*

$$\mathbb{P}\left(\phi_d\Big(Y^n, \phi_e\Big(X^n, (g(Y_t))_{t\leq n}\Big)\Big) \neq \Big(f(X_t, Y_t)\Big)_{t\leq n}\right) = 0. \tag{6.7}$$

The objective is to find the minimal rate among all coding schemes under the zero-error constraint:

$$R^*_{CFC0} \doteq \inf_{\substack{n,\phi_e,\phi_d \\ \text{zero-error}}} \frac{1}{n}\mathbb{E}\Big[\ell \circ \phi_e\Big(X^n, (g(Y_t))_{t\leq n}\Big)\Big]. \tag{6.8}$$

Now the scheme of Figure 6.2 has been studied with different coding constraints than zero-error, and the optimal rate has been characterized in each case: the lossless case by Orlitsky and Roche in [50], the lossy case by Yamamoto in [76], and the zero-error "unrestricted inputs" case by Shayevitz in [60]. These results can only be used as bounds here: the zero-error problem depicted in Figure 6.2 does not have a characterization of the optimal rate.

Numerous extensions of the problem depicted in Figure 6.2 have been studied recently. The distributed context, for instance, has an additional encoder which encodes Y before transmitting it to the decoder. Achievability schemes have been proposed for this setting by Krithivasan and Pradhan in [42] using abelian groups; by Basu et al. in [6] using hypergraphs for the case with maximum distortion criterion; and by Malak and Médard in [46] using hyperplane separations for the continuous lossless case.

Another related context is the network setting, where the function of source random variables from source nodes has to be retrieved at the sink node of a given network. For tree networks, the feasible rate region is characterized by Feizi and Médard in [24] for networks of depth one; and by Sefidgaran and Tchamkerten in [56] under a Markov source distribution hypothesis. In [52], Ravi and Dey consider a bidirectional relay with zero-error "unrestricted inputs" and characterize the rate region for a specific class of functions. In [32], Guang et al. study zero-error function computation on acyclic networks with limited capacities, and give an inner bound based on network cut-sets. For both distributed and network settings, the zero-error coding for computing problem with encoder side information remains open.

In Chapter 6, we formulate an hypothesis on $P_{X,Y}$ and g that we call "pairwise shared side information" that allows us to derive a single-letter characterization of the optimal rate. In particular it covers the cases with $P_{X,Y}$ full-support, without any assumption on f, g.

6.2 General case

We first build the characteristic graph $G_{[n]}$, which is a probabilistic graph that captures the zero-error encoding constraints on a given number n of source uses. It differs from the graphs used in [60], as we do not need a cartesian representation of these graphs to study the optimal rates. Furthermore, it has a vertex for each possible realization of $\left(X^n, \left(g(Y_t)\right)_{t\leq n}\right)$ known at the encoder, instead of $\mathcal{X}^n$ as in the zero-error Slepian-Wolf problem presented in Section 3.3.1.

Definition 6.2.1 (Characteristic graph $G_{[n]}$) *The characteristic graph $G_{[n]}$ is defined by:*

- *$\mathcal{X}^n \times \mathcal{Z}^n$ as set of vertices with distribution $P^n_{X,g(Y)}$,*

- $(x^n, z^n)(x'^n, z'^n)$ *are adjacent if* $z^n = z'^n$ *and there exists* $y^n \in g^{-1}(z^n)$ *such that:*

$$\forall t \leq n,\ P_{X,Y}(x_t, y_t)P_{X,Y}(x'_t, y_t) > 0, \tag{6.9}$$

$$\text{and } \exists t \leq n,\ f(x_t, y_t) \neq f(x'_t, y_t); \tag{6.10}$$

where $g^{-1}(z^n) = \left\{ y^n \in \mathcal{Y}^n \,\middle|\, \Big(g(y_t)\Big)_{t\leq n} = z^n \right\}$.

The characteristic graph $G_{[n]}$ is designed with the same core idea as in [75]: (x^n, z^n) and (x'^n, z'^n) are adjacent if there exists a side information symbol y^n compatible with the observation of the encoder (i.e. $z^n = z'^n$ and $y^n \in g^{-1}(z^n)$), such that $f(x^n, y^n) \neq f(x'^n, y^n)$. In order to prevent erroneous decodings, the encoder must map adjacent pairs of sequences to different codewords; hence the use of graph colorings.

Theorem 6.2.2 (Optimal rate) *The optimal rate writes:*

$$R^*_{CFC0} = \lim_{n\to\infty} \frac{1}{n} H_\chi(G_{[n]}). \tag{6.11}$$

Proof. By construction the following holds: for all encoding function ϕ_e, ϕ_e is a coloring of $G_{[n]}$ if and only if there exists a decoding function ϕ_d such that (n, ϕ_e, ϕ_d) satisfies the zero-error property. Thus the best achievable rate writes

$$R^*_{CFC} = \inf_n \inf_{\phi_e \text{ coloring of } G_{[n]}} H\Big(\phi_e\Big(X^n, \Big(g(Y_t)\Big)_{t\leq n}\Big)\Big) \tag{6.12}$$

$$= \lim_{n\to\infty} \frac{1}{n} H_\chi(G_{[n]}). \tag{6.13}$$

where (6.13) comes from Fekete's lemma and from the definition of H_χ. □

A general single-letter expression for R^*_{CFC} is missing, due to the lack of intrinsic structure of $G_{[n]}$. In Section 6.3, we introduce a hypothesis that gives structure to $G_{[n]}$ and allows us to derive a single-letter expression for R^*_{CFC}.

6.3 Pairwise shared side information

Definition 6.3.1 *The distribution* $P_{X,Y}$ *and the function* g *satisfy the "pairwise shared side information" condition if*

$$\forall z \in \mathcal{Z}, \forall x, x' \in \mathcal{X}, \exists y \in g^{-1}(z), P_{XY}(x, y)P_{XY}(x', y) > 0. \tag{6.14}$$

This means that for all z output of g, every pair (x, x') "shares" at least one side information symbol $y \in g^{-1}(z)$.

Note that any full-support distribution $P_{X,Y}$ satisfies the "pairwise shared side information" hypothesis. In Theorem 6.3.2 we give an interpretation of the "pairwise shared side information" condition in terms of the optimal rate in an auxiliary zero-error Slepian-Wolf problem. The proof of Theorem 6.3.2 is given in Appendix C.1

Theorem 6.3.2 *The tuple $(P_{X,Y}, g)$ satisfies the condition "pairwise shared side information" (6.14)*

$\iff R^* = H(X|g(Y))$ in the case $f(X,Y) = X$, and for all $z \in \mathcal{Z}$, $P_{X|g(Y)=z}$ is full-support.

In the zero-error Slepian-Wolf problem presented in Section 4.1.1, the optimal rate $\lim_{n\to\infty} \frac{1}{n} H_\chi(G^{\wedge n})$ does not have a single-letter expression. However, as shown in Proposition 6.3.4, there exists such a formula for $\lim_{n\to\infty} \frac{1}{n} H_\chi(G^{\vee n})$: the Körner graph entropy introduced in [39]. By using a convex combination of Körner graph entropies, we provide a single-letter expression in Theorem 6.3.6 for the optimal rate R^*_{CFC}.

Definition 6.3.3 (OR product) *Let $G_1 = (\mathcal{V}_1, \mathcal{E}_1, P_{V_1})$, $G_2 = (\mathcal{V}_2, \mathcal{E}_2, P_{V_2})$ be two probabilistic graphs; their OR product denoted $G_1 \vee G_2$ is defined by: $\mathcal{V}_1 \times \mathcal{V}_2$ as set of vertices, $P_{V_1} P_{V_2}$ as probability distribution on the vertices, and $(v_1 v_2), (v_1' v_2')$ are adjacent if*

$$(v_1 v_1' \in \mathcal{E}_1 \text{ and } v_1 \neq v_1') \text{ OR } (v_2 v_2' \in \mathcal{E}_2 \text{ and } v_2 \neq v_2');$$

with the convention that all vertices are self-adjacent. We denote by $G_1^{\vee n}$ the n-th OR power.

Proposition 6.3.4 (Properties of H_κ) *[5, Theorem 5] For all probabilistic graphs G and G',*

$$H_\kappa(G) = \lim_{n\to\infty} \frac{1}{n} H_\chi(G^{\vee n}), \tag{6.15}$$

$$H_\kappa(G \vee G') = H_\kappa(G) + H_\kappa(G'). \tag{6.16}$$

Definition 6.3.5 (Auxiliary graph G_z^f) *For all $z \in \mathcal{Z}$, we define the auxiliary graph G_z^f by*

- $\mathcal{X}$ as set of vertices with distribution $P_{X|g(Y)=z}$,

- xx' *are adjacent if* $f(x,y) \neq f(x',y)$ *for some* $y \in g^{-1}(z) \cap \operatorname{supp} P_{Y|X=x} \cap \operatorname{supp} P_{Y|X=x'}$.

Theorem 6.3.6 (Pairwise shared side information) *If* $P_{X,Y}$ *and* g *satisfy* (6.14)*, the optimal rate writes:*

$$R^*_{CFC0} = \sum_{z \in \mathcal{Z}} P_{g(Y)}(z) H_\kappa(G^f_z). \tag{6.17}$$

The proof is in Section C.2, the keypoint is the particular structure of $G_{[n]}$: a disjoint union of OR products.

Remark 6.3.7 *The "pairwise shared side information" assumption* (6.14) *implies that the adjacency condition* (6.9) *is satisfied, which makes* $G_{[n]}$ *a disjoint union of OR products. Moreover, Körner graph entropies appear in the final expression for* R^*_{CFC}*, even if* $G_{[n]}$ *is not an* n*-th OR power.*

Now consider the case where $P_{X,Y}$ is full-support. This is a sufficient condition to have (6.14). The optimal rate in this setting is derived from Theorem 6.3.6, which leads to the analytic expression in Theorem 6.3.8.

Theorem 6.3.8 (Optimal rate when $P_{X,Y}$ is full-support) *When* $P_{X,Y}$ *is full-support, the optimal rate writes:*

$$R^*_{CFC0} = H\Big(j(X, g(Y)) \Big| g(Y)\Big), \tag{6.18}$$

where the function j *returns a word in* $\mathcal{U}^*$*, defined by*

$$\begin{aligned} j : \mathcal{X} \times \mathcal{Z} &\to \mathcal{U}^* \\ (x,z) &\mapsto \Big(f(x,y')\Big)_{y' \in g^{-1}(z)}. \end{aligned} \tag{6.19}$$

Proof. By Theorem 6.3.6, $R^*_{CFC} = \sum_{z \in \mathcal{Z}} P_{g(Y)}(z) H_\kappa(G^f_z)$. It can be shown that G^f_z is complete multipartite for all z as $P_{X,Y}$ is full support; and it satisfies $H_\kappa(G^f_z) = H\Big(j(X, g(Y)) \Big| g(Y) = z\Big)$. □

6.4 Example

In this example, the "pairwise shared side information" assumption is satisfied and R^*_{CFC} is strictly less than a conditional Huffman coding of X knowing $g(Y)$; and also strictly less than the optimal rate without exploiting $g(Y)$ at the encoder.

$P_{X,Y}$

X \ Y	0	1	2	3	4	5	6	7
	$g(Y)=0$				$g(Y)=1$			
0	0.1	0.05	$*$	$*$	0.05	0.05	$*$	$*$
1	0.1	$*$	0.05	$*$	0.05	0.05	0.05	$*$
2	0.1	$*$	$*$	0.05	$*$	0.05	$*$	$*$
3	$*$	0.05	0.05	0.05	$*$	0.05	$*$	0.05

$f(\cdot,\cdot)$

X \ Y	0	1	2	3	4	5	6	7
	$g(Y)=0$				$g(Y)=1$			
0	a	b	$*$	$*$	b	a	$*$	$*$
1	a	$*$	b	$*$	a	a	b	$*$
2	b	$*$	$*$	c	$*$	b	$*$	$*$
3	$*$	c	c	c	$*$	c	$*$	c

Figure 6.3 – An example of $P_{X,Y}$ and g that satisfy (6.14); along with the outcomes $f(X,Y)$. The elements outside $\operatorname{supp} P_{X,Y}$ are denoted by $*$.

Consider the probability distribution and function outcomes depicted in Figure 6.3, with $\mathcal{U}=\{a,b,c\}$, $\mathcal{X}=\{0,...,3\}$, $\mathcal{Y}=\{0,...,7\}$, and $\mathcal{Z}=\{0,1\}$. Let us show that the "pairwise shared side information" assumption is satisfied. The source symbols $0,1,2\in\mathcal{X}$ share the side information symbol 0 (resp. 5) when $g(Y)=0$ (resp. $g(Y)=1$). The source symbol $3\in\mathcal{X}$ shares the side information symbols $1,2,3$ with the source symbols $0,1,2$, respectively, when $g(Y)=0$; and the source symbol 3 shares the side information symbol 5 with all other source symbols when $g(Y)=1$.

Since the "pairwise shared side information" assumption is satisfied, we can use Theorem 6.3.6; the optimal rate writes

$$R^*_{CFC0}=P_{g(Y)}(0)H_\kappa(G_0^f)+P_{g(Y)}(1)H_\kappa(G_1^f). \tag{6.20}$$

First we need to determine the probabilistic graphs G_0^f and G_1^f. In G_0^f, the vertex 0 is adjacent to 2 and 3, as $f(0,0)\neq f(2,0)$ and $f(0,1)\neq f(3,1)$. The vertex 1 is also adjacent to 2 and 3 as $f(1,0)\neq f(2,0)$ and $f(1,2)\neq f(3,2)$. Furthermore $P_{X|g(Y)=0}$ is uniform, hence $G_0^f=(C_4,\mathrm{Unif}(\mathcal{X}))$ where C_4 is the cycle graph with 4 vertices.

In G_1^f, the vertices 1, 2, 3 are pairwise adjacent as $f(1,5)$, $f(2,5)$ and $f(3,5)$ are pairwise different; and 0 is adjacent to 1, 2 and 3 because of the different function outputs generated by $Y=4$ and $Y=5$. Thus, $G_1^f=(K_4,P_{X|g(Y)=1})$ with $P_{X|g(Y)=1}=(\frac{1}{4},\frac{3}{8},\frac{1}{8},\frac{1}{4})$

and K_4 is the complete graph with 4 vertices.

Now let us determine $H_\kappa(G_0^f)$ and $H_\kappa(G_1^f)$. On one hand,

$$H_\kappa(G_0^f) = H(V_0) - \max_{V_0 \in W \in \Gamma(G_0^f)} H(V_0|W) \tag{6.21}$$

$$= 2 - 1 = 1, \tag{6.22}$$

with $V_0 \sim P_{X|g(Y)=0} = \text{Unif}(\mathcal{X})$; and where $H(V_0|W)$ in (6.21) is maximized by taking $W = \{0, 1\}$ when $V \in \{0, 1\}$, and $W = \{2, 3\}$ otherwise.

On the other hand,

$$H_\kappa(G_1^f) = \min_{V_1 \in W \in \Gamma(G_1^f)} I(W; V_1) \tag{6.23}$$

$$= H(V_1) \approx 1.906, \tag{6.24}$$

with $V_1 \sim P_{X|g(Y)=1}$; where (6.24) follows from $\Gamma(G_1^f) = \{\{0\}, ..., \{3\}\}$, as G_1^f is complete. Hence $R^*_{CFC} \approx 1.362$.

The rate that we would obtain by transmitting X knowing $g(Y)$ at both encoder and decoder with a conditional Huffman algorithm writes: $R_{\text{Huff}} = H(X|g(Y)) \approx 1.962$.

The rate that we would obtain without exploiting $g(Y)$ at the encoder is $R_{\text{No } g} = H(X) \approx 1.985$, because of the different function outputs generated by $Y = 4$ and $Y = 5$.

Finally, $H(f(X, Y)|Y) \approx 0.875$.

In this example we have

$$H(X) = R_{\text{No } g} > R_{\text{Huff}} > R^*_{CFC0} > H(f(X, Y)|Y). \tag{6.25}$$

This illustrates the impact of the side information at the encoder in this setting, as we can observe a large gap between the optimal rate R^*_{CFC} and $R_{\text{No } g}$.

Conclusion

We have shown the equivalences of linearization between C_0, $C(\cdot, P_V)$, and $\overline{H}$. Therefore, we proved the equivalence between the suboptimality of separated zero-error coding on independent channels; and the suboptimality of separated compression of independent sources in the zero-error Slepian-Wolf setting, with same characteristic graph and capacity-achieving distribution.

We also state the following open questions:

- As pointed out in Lemma 5.3.3, for all capacity-achieving distribution of a product graph, the product of its marginals is also capacity-achieving. Are these marginals capacity-achieving for the respective graphs in the product; and conversely, if we consider the product of capacity-achieving distributions of graphs, is this distribution capacity-achieving for the product of graphs? In other words,

$$\mathcal{P}^*\left(\bigwedge_{a\in\mathcal{A}} G_a\right) \cap \bigotimes_{a\in\mathcal{A}} \Delta(\mathcal{V}_a) \stackrel{?}{=} \bigotimes_{a\in\mathcal{A}} \mathcal{P}^*(G_a). \tag{6.26}$$

  We gave a partial answer in Theorem 5.4.1, in the sense that inclusion holds when the linearization of the product holds.

- We have shown in Theorem 5.4.1 and Theorem 5.4.2 that the linearization of C_0 holds if and only if the linearization of $C(\cdot, P_V)$ holds, where P_V is any capacity-achieving distribution. Can we find graphs such that the linearization of $C(\cdot, P_V)$ holds when P_V is capacity-achieving, but does not hold for some P_V that is not capacity-achieving? A negative answer would imply that the linearization of C_0 is equivalent to the linearization of $C(\cdot, P_V)$ and $\overline{H}$ for *all* P_V, similarly to perfect graphs.

- Finally, we have seen in Corollary 5.5.10 that $\overline{H}\Big(G \wedge (C_5, \mathrm{Unif}(\{1, ..., 5\}))\Big)$ with G perfect is an example of linearization. Is the non-linearization of $\overline{H}(\wedge\, \cdot)$ tied to specific non-perfect induced subgraphs in each graph in the product? And if so, can we find a minimal family of these graphs?

Appendix A

Proofs for correction-based zero-error results

A.1 Proof of Lemma 4.2.5

Let us prove that

$$R_1^{(n)} \underset{n\to\infty}{\to} H(X|Y) + \delta, \qquad R_2^{(n)} \underset{n\to\infty}{\to} I(X;Y). \tag{A.1}$$

A.1.1 Preliminaries

Lemma A.1.1 *Let X' be a random variable such that $P_{X'} = \mathrm{Unif}(\mathcal{X})$. Then for all $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$, we have:*

$$\mathbb{P}\left(T_{X'^n,y^n} = T_{x^n,y^n}\right) = 2^{nH(T_{x^n,y^n}) - nH(T_{y^n}) - n\log|\mathcal{X}| + o(n)} \tag{A.2}$$

Proof. Since $P_{X'}$ is uniform:

$$\begin{aligned}\mathbb{P}\left(T_{X'^n,y^n} = T_{x^n,y^n}\right) &= |\mathcal{X}|^{-n} \left|\mathcal{T}_{Q_{x^n|y^n}}(y^n)\right| \\ &= 2^{-n\log|\mathcal{X}|} 2^{nH(T_{x^n,y^n}) - nH(T_{y^n}) + o(n)},\end{aligned} \tag{A.3}$$

as [20, Lemma 2.5] gives the asymptotic size of the $T_{x^n|y^n}$-shell $\mathcal{T}_{T_{x^n|y^n}}(y^n)$.
□

A.1.2 Probability of decoding ambiguity

We need to estimate $\mathbb{P}(E = 1)$. We have $E = 1$ *iff* $K = 0$ or there exists $(a_1, ..., a_K) \in \mathcal{X}^K \setminus \{0, ..., 0\}$ such that $T_{\left(X^n + \sum_{i\leq K} a_i \mathbf{G}_K^{(i)}\right), Y^n} = T_{X^n,Y^n}$, where $\mathbf{G}_K^{(i)}$ denotes the i-th

column of the matrix $\mathbf{G}_K$. Thus

$$\mathbb{P}(E=1) \leq \mathbb{P}(K=0) + \mathbb{P}\left(\bigcup_{\substack{a^k \in \mathcal{X}^k \\ a \neq 0}} \left[T_{\left(X^n + \sum_{i \leq K} a_i \mathbf{G}_K^{(i)}\right), Y^n} = T_{X^n, Y^n} \right] \middle| K \neq 0 \right). \tag{A.4}$$

We provide an upper bound on the second term in (A.4). For all (x^n, y^n) such that $k \neq 0$, we have:

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{\substack{a^k \in \mathcal{X}^k \\ a \neq 0}} \left[T_{\left(x^n + \sum_{i \leq k} a_i \mathbf{G}_k^{(i)}\right), y^n} = T_{x^n, y^n} \right] \right) \\
&\leq \sum_{\substack{a^k \in \mathcal{X}^k \\ a \neq 0}} \mathbb{P}\left(T_{\left(x^n + \sum_{i \leq k} a_i \mathbf{G}_k^{(i)}\right), y^n} = T_{x^n, y^n} \right) && \text{(A.5)} \\
&\leq |\mathcal{X}|^k 2^{nH(T_{x^n,y^n}) - nH(T_{y^n}) - n \log |\mathcal{X}| + o(n)} && \text{(A.6)} \\
&\leq 2^{n \log |\mathcal{X}| - nH(T_{x^n,y^n}) + nH(T_{y^n}) - \delta n + o(n)} \\
&\quad \cdot 2^{nH(T_{x^n,y^n}) - nH(T_{y^n}) - n \log |\mathcal{X}| + o(n)} && \text{(A.7)} \\
&\leq 2^{-\delta n + o(n)}, && \text{(A.8)}
\end{aligned}$$

where (A.6) comes from Lemma A.1.1 and (A.8) comes from (4.43). Therefore,

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{\substack{a^K \in \mathcal{X}^K \\ a \neq 0}} \left[T_{\left(X^n + \sum_{i \leq K} \alpha_i \mathbf{G}_K^{(i)}\right), Y^n} = T_{X^n, Y^n} \right] \middle| K \neq 0 \right) \\
&= \sum_{x^n, y^n} \mathbb{P}\Big((X^n, Y^n) = (x^n, y^n) \Big| K \neq 0 \Big) && \text{(A.9)} \\
&\quad \cdot \mathbb{P}\left(\bigcup_{\substack{a^K \in \mathcal{X}^K \\ a \neq 0}} \left[T_{\left(X^n + \sum_{i \leq K} \alpha_i \mathbf{G}_K^{(i)}\right), Y^n} = T_{X^n, Y^n} \right] \middle| K \neq 0, (X^n, Y^n) = (x^n, y^n) \right) \\
&\leq \sum_{x^n, y^n} \mathbb{P}\Big((X^n, Y^n) = (x^n, y^n) \Big| K \neq 0 \Big) 2^{-\delta n + o(n)} && \text{(A.10)} \\
&\leq 2^{-\delta n + o(n)}, && \text{(A.11)}
\end{aligned}$$

where (A.10) comes from (A.8) and the fact that $\mathbf{G}$ is independent of (X, Y).

We now provide an upper bound on the first term in (A.4).

$$\mathcal{S} \doteq \left\{ P_{X',Y'} \in \Delta(\mathcal{X}\times\mathcal{Y}) \;\middle|\; 1 - \frac{H(X'|Y')+\delta}{\log|\mathcal{X}|} \leq 0 \right\}. \tag{A.12}$$

Then we have:

$$\begin{aligned}
&\mathbb{P}(K=0) && \text{(A.13)}\\
&= \sum_{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n} \mathbb{P}\left(n - n\frac{H(T_{x^n,y^n}) - H(T_{y^n}) + \delta}{\log|\mathcal{X}|} \leq 0 \right) P_{X,Y}^{\otimes n}(x^n,y^n) && \text{(A.14)}\\
&= \mathbb{P}(T_{X^n,Y^n} \in \mathcal{S}) && \text{(A.15)}\\
&= \sum_{P_{X',Y'}\in\mathcal{S}\cap\Delta_n(\mathcal{X}\times\mathcal{Y})} \mathbb{P}(T_{X^n,Y^n} = P_{X',Y'}) && \text{(A.16)}\\
&\leq |\mathcal{S}\cap\Delta_n(\mathcal{X}\times\mathcal{Y})| \sup_{P_{X',Y'}\in\mathcal{S}\cap\Delta_n(\mathcal{X}\times\mathcal{Y})} \mathbb{P}(T_{X^n,Y^n} = P_{X',Y'}) && \text{(A.17)}\\
&\leq |\mathcal{S}\cap\Delta_n(\mathcal{X}\times\mathcal{Y})| \sup_{P_{X',Y'}\in\mathcal{S}\cap\Delta_n(\mathcal{X}\times\mathcal{Y})} 2^{-nD(P_{X',Y'}\|P_{X,Y})} && \text{(A.18)}\\
&\leq |\mathcal{S}\cap\Delta_n(\mathcal{X}\times\mathcal{Y})| \sup_{P_{X',Y'}\in\mathcal{S}} 2^{-nD(P_{X',Y'}\|P_{X,Y})} && \text{(A.19)}\\
&\leq 2^{-n\inf_{P_{X',Y'}\in\mathcal{S}} D(P_{X',Y'}\|P_{X,Y})+o(n)}, && \text{(A.20)}
\end{aligned}$$

where (A.18) comes from [20, Lemma 2.6]. Since $P_{X,Y} \notin \mathcal{S}$ by definition of δ, we have $\inf_{P_{X',Y'}\in\mathcal{S}} D(P_{X,Y}\|P_{X',Y'}) > 0$. Thus there exists a positive constant $\beta > 0$ such that

$$\mathbb{P}(K=0) \leq 2^{-\beta n+o(n)}. \tag{A.21}$$

Thus by combining (A.4), (A.11), (A.21), we have:

$$\mathbb{P}(E=1) \leq 2^{-\delta n+o(n)} + 2^{-\beta n+o(n)}. \tag{A.22}$$

A.1.3 Rate on the common channel

The encoding function $\phi_e^{(1)}$ defined in (4.44) returns T_{X^n,Y^n} and E. When $E=0$, it sends the syndrome $\mathbf{H}_K X^n$ at rate $\frac{n-K}{n}\log|\mathcal{X}|$, otherwise, it sends X^n. Therefore,

$$\begin{aligned}
&nR_1^{(n)} && \text{(A.23)}\\
&= 1 + |\mathcal{X}||\mathcal{Y}|\log(n+1) + \mathbb{P}(E=1)n\log|\mathcal{X}| && \text{(A.24)}
\end{aligned}$$

$$+ \mathbb{P}(E=0) \sum_{x^n,y^n} \mathbb{P}\Big((X^n,Y^n)=(x^n,y^n)\Big|E=0\Big)\cdot(n-k)\log|\mathcal{X}|$$

$$\leq 1+|\mathcal{X}||\mathcal{Y}|\log(n+1)+\mathbb{P}(E=1)n\log|\mathcal{X}|+(n-\mathbb{E}[K])\log|\mathcal{X}| \tag{A.25}$$

$$\leq 1+|\mathcal{X}||\mathcal{Y}|\log(n+1)+\mathbb{P}(E=1)n\log|\mathcal{X}|+n\mathbb{E}\Big[H_{T_{X^n,Y^n}}(X|Y)\Big]+n\delta+1, \tag{A.26}$$

where $H_{T_{X^n,Y^n}}(X|Y)$ denotes a random variable, which is the conditional entropy computed with the distribution T_{X^n,Y^n}; (A.25) comes from $n-k\geq 0$ for all (x^n,y^n), and (A.26) comes from (4.43).

By the law of large numbers [18, Theorem 11.2.1] $\mathbb{E}\Big[H_{T_{X^n,Y^n}}(X|Y)\Big] \underset{n\to\infty}{\to} H(X|Y)$, and by using (A.22), we obtain

$$\lim_{n\to\infty} R_1^{(n)} \leq H(X|Y)+\delta. \tag{A.27}$$

A.1.4 Rate on the secondary channel

The encoding function $\phi_e^{(2)}$ is defined in (4.45). If $E=0$, then $K\neq 0$ and the encoder transmits the index of X^n in its coset. The Huffman algorithm has an average output length $R_2^{(n)}$ that satisfies

$$R_2^{(n)} \leq \frac{1}{n}\Big(1+\sum_{k\neq 0}\mathbb{P}(K=k|E=0)\cdot H(X^n|\mathbf{H}_kX^n,K=k,\mathcal{C},E=0)\Big) \tag{A.28}$$

$$= \frac{1}{n}+\frac{1}{n}H(X^n|K,\mathcal{C},E=0)-\frac{1}{n}H(\mathbf{H}_KX^n|K,\mathcal{C},E=0), \tag{A.29}$$

where (A.29) follows from the fact that $\mathbf{H}_KX^n$ is a deterministic function of X^n, given a random code $\mathcal{C}$.

We now provide an upper bound to the last term $-\frac{1}{n}H(\mathbf{H}_KX^n|K,\mathcal{C},E=0)$ in (A.29). To do so, we introduce a new encoding scheme that first encodes the sequences X^n and Y^n with the encoding function $\phi_e^{(1)}$, and then encode the output by using an entropy coder. The rate of this code r is upperbounded by $H(\phi_e^{(1)}(X^n,Y^n)|\mathcal{C})+1$.

Moreover, the decoder 1 retrieves X^n with zero error (see Section 4.2.3), and the entropy coder is also lossless. Thus r is greater than the rate achieved by a conditional entropy coder that compresses X^n knowing the side information Y^n, whose rate is lower bounded by $nH(X|Y)$.

Therefore, we have

$$
\begin{aligned}
nH(X|Y) &\leq r && \text{(A.30)}\\
&< H(\phi_e^{(1)}(X^n, Y^n)|\mathcal{C}) + 1 && \text{(A.31)}\\
&= 1 + H(T_{X^n,Y^n}, E|\mathcal{C}) + \mathbb{P}(E=0)H(\mathbf{H}_K X^n|T_{X^n,Y^n}, \mathcal{C}, E=0)\\
&\quad + \mathbb{P}(E=1)H(X^n|T_{X^n,Y^n}, \mathcal{C}, E=1) && \text{(A.32)}\\
&\leq H(\mathbf{H}_K X^n|T_{X^n,Y^n}, \mathcal{C}, E=0) + o(n) && \text{(A.33)}\\
&= H(\mathbf{H}_K X^n|T_{X^n,Y^n}, K, \mathcal{C}, E=0) + o(n) && \text{(A.34)}\\
&\leq H(\mathbf{H}_K X^n|K, \mathcal{C}, E=0) + o(n) && \text{(A.35)}
\end{aligned}
$$

where $o(n)$ in (A.33) corresponds to the term $1+H(T_{X^n,Y^n}, E|\mathcal{C})+\mathbb{P}(E=1)H(X^n|T_{X^n,Y^n}, \mathcal{C}, E=1)$, and (A.34) follows from the fact that K is a deterministic function of T_{X^n,Y^n}.

We now provide an upper bound on the second term of (A.29).

$$
\begin{aligned}
&\frac{1}{n}H(X^n|K, \mathcal{C}, E=0) \leq \frac{1}{n\mathbb{P}(E=0)}\Big(H(X^n|K, \mathcal{C}, E)\\
&- \mathbb{P}(E=1)H(X^n|K, \mathcal{C}, E=1)\Big)\\
&\leq \frac{1}{n}H(X^n|K, \mathcal{C}, E) + o(1) && \text{(A.36)}\\
&\leq H(X) + o(1). && \text{(A.37)}
\end{aligned}
$$

By combining (A.29), (A.35) and (A.37), we obtain

$$\lim_{n\to\infty} R_2^{(n)} \leq I(X;Y). \tag{A.38}$$

Conclusion. The rates in (A.27) and (A.38) are evaluated on average over the random code $\mathcal{C}$ with a parameter $\delta > 0$ arbitrarily small. This shows that there exists a sequence of $\left(n, R_1^{(n)}, R_2^{(n)}\right)$-zero-error source codes, such that

$$\left(R_1^{(n)}, R_2^{(n)}\right) \underset{n\to\infty}{\to} \Big(H(X|Y), I(X;Y)\Big). \tag{A.39}$$

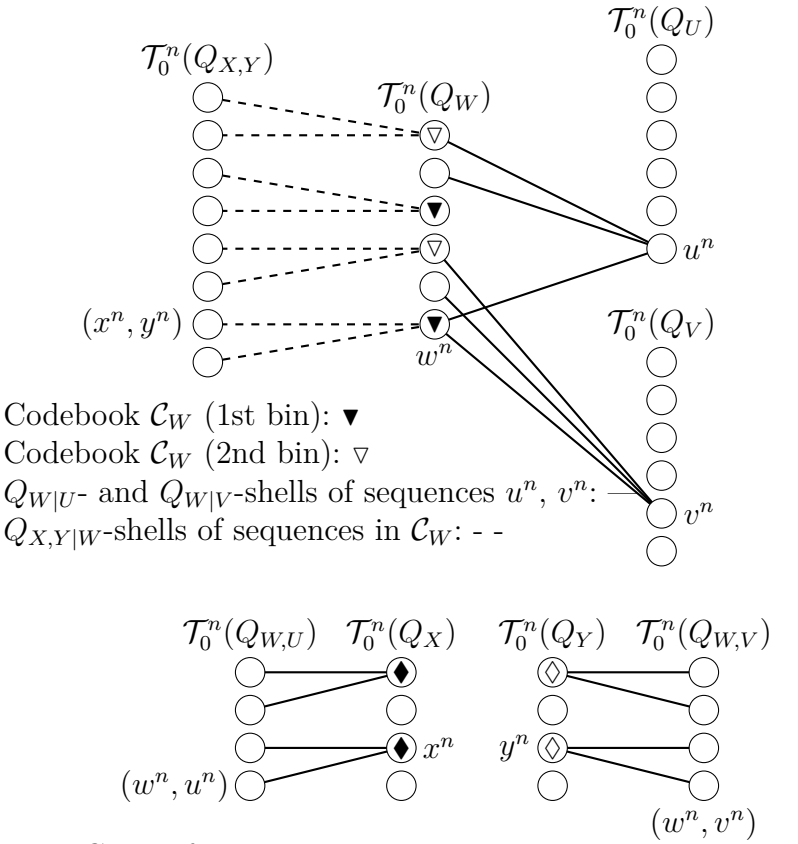


Figure A.1 – An illustration of the encoding algorithm.

A.2 Proof of Theorem 4.3.1

A.2.1 Outline

The encoding algorithm is depicted in Figure A.1. First, for all realization (x^n, y^n) of the source, a codeword w^n is selected from a random codebook that captures the common information between X and Y. Then the bin index of w^n is sent on the channel with encoding function $\phi_e^{(0)}$; the number of bins is adjusted to the worst side information. Using the V-shells of their respective side information u^n and v^n, both decoders retrieve w^n. Finally, the coset of x^n sent on the channel with encoding function $\phi_e^{(\mathsf{x}_1)}$, and the pair (w^n, u^n) (resp. (w^n, v^n) and the coset of y^n) enables the decoder x_1 (resp. Y) to recover x^n (resp. y^n). An information complement is sent to the decoder x_2 so that it can recover

x^n as well.

A.2.2 Encoding algorithm

Let $P_{W|U,V,X,Y} \in \Delta(\mathcal{W})^{|\mathcal{U}\times\mathcal{V}\times\mathcal{X}\times\mathcal{Y}|}$, let $\delta > 0$, and let $n \in \mathbb{N}^\star$. For all type $Q_{U,V,X,Y} \in \Delta_n(\mathcal{U}\times\mathcal{V}\times\mathcal{X}\times\mathcal{Y})$, let $Q_{U,V,W,X,Y}$ be the type from $\Delta_n(\mathcal{U}\times\mathcal{V}\times\mathcal{W}\times\mathcal{X}\times\mathcal{Y})$ that has $Q_{U,V,X,Y}$ as marginal distribution and minimizes $D(P_{U,V,W,X,Y}\|Q_{U,V,W,X,Y})$, so we have

$$Q_{U,V,W,X,Y} \underset{Q_{U,V,X,Y}\to P_{U,V,X,Y}}{\to} P_{U,V,W,X,Y}. \tag{A.40}$$

In the following, we denote by $H_Q(\cdot)$ and $I_Q(\cdot\,;\cdot)$ the entropy and mutual information that are computed w.r.t. Q, instead of the true distribution of $P_{U,V,W,X,Y}$.

Now, for a fixed type $Q_{U,V,W,X,Y}$, partition $\mathcal{T}_0^n(Q_X)$ (resp. $\mathcal{T}_0^n(Q_Y)$) into $2^{nH_Q(X|U,W)+n\delta}$ (resp. $2^{nH_Q(Y|V,W)+n\delta}$) cosets, using adequate generator and parity matrices $\mathbf{G}_X$ and $\mathbf{H}_X$ (resp. $\mathbf{G}_Y$ and $\mathbf{H}_Y$), cf. Lemma 4.3.5. This is possible as $|\mathcal{X}|$ and $|\mathcal{Y}|$ can be assumed prime w.l.o.g. by padding (i.e. completing with zeros) Q if necessary.

Let $\mathcal{C}_W$ be a set of codewords formed with $2^{nI_Q(U,V,X,Y;W)+n\delta}$ random sequences drawn from $\mathcal{W}^n$ independently and following the distribution Q_W. Let $\mathcal{Z}$ be a finite set such that

$$|\mathcal{Z}| = \frac{|\mathcal{C}_W|}{2^{n\min(I_Q(U;W),I_Q(V;W))-n\delta}} \tag{A.41}$$

$$= 2^{nI_Q(U,V,X,Y;W)-n\min(I_Q(U;W),I_Q(V;W))+2n\delta}. \tag{A.42}$$

A bin label $b(\overline{w}^n)$ drawn uniformly in $\mathcal{Z}$ is assigned to each sequence $\overline{w}^n \in \mathcal{C}_W$, so we have $2^{n\min(I_Q(U;W),I_Q(V;W))-n\delta+o(n)}$ sequences in each bin. In the following we will denote by $\mathcal{C}_W^{[b(w^n)]}$ the subset of sequences from $\mathcal{C}_W$ in the bin labelled $b(w^n)$.

Now let (u^n, v^n, x^n, y^n) be a realization of the source, and $Q_{U,V,X,Y}$ the corresponding type. Let $Q_{U,V,W,X,Y}$ be the completed type as described above. The encoder determines a sequence $w^n \in \mathcal{T}_{Q_{W|U,V,X,Y}}(u^n, v^n, x^n, y^n)$ and its bin label $b(w^n)$; if no such sequence exists then the decoding ambiguity $\mathcal{E}_1$ is declared.

The encoder sends:

$$\phi_e^{(\mathsf{x}_1)}(u^n, v^n, x^n, y^n) = B(\mathbf{H}_X x^n), \tag{A.43}$$

$$\phi_e^{(\mathsf{x}_2)}(u^n, v^n, x^n, y^n) = \iota, \tag{A.44}$$

$$\phi_e^{(\mathsf{y})}(u^n, v^n, x^n, y^n) = B(\mathbf{H}_Y y^n), \tag{A.45}$$

$$\phi_e^{(0)}(u^n, v^n, x^n, y^n) \tag{A.46}$$
$$= B\left(Q_{U,V,W,X,Y}, b(w^n), \begin{cases} E, x^n, y^n & \text{if } E = 1 \\ E & \text{if } E = 0 \end{cases}\right),$$

where $B(\cdot)$ denotes the binary expansion, and ι denotes the index of x^n in $(\operatorname{Ker}\mathbf{H}_X + x^n) \cap \bigcup_{\overline{w}^n \in \mathcal{C}_W^{[b(w^n)]}} \mathcal{T}_{Q_{X|W}}(\overline{w}^n)$.

The error bit E is set to 1 iff there is a decoding ambiguity, i.e. one of the following events occur:

$$\mathcal{E}_1 = \left\{\nexists w^n \in \mathcal{C}_W, \text{ s.t. } T_{u^n,v^n,w^n,x^n,y^n} = Q_{U,V,W,X,Y}\right\},$$
$$\mathcal{E}_2 = \left\{\exists \tilde{w}^n \neq w^n, \ \tilde{w}^n \in \mathcal{C}_W^{[b(w^n)]} \cap \mathcal{T}_{Q_{W|U}}(u^n)\right\},$$
$$\mathcal{E}_3 = \left\{\exists \tilde{w}^n \neq w^n, \ \tilde{w}^n \in \mathcal{C}_W^{[b(w^n)]} \cap \mathcal{T}_{Q_{W|V}}(v^n)\right\},$$
$$\mathcal{E}_4 = \left\{\exists \tilde{x}^n \neq x^n, \ \tilde{x}^n \in \mathcal{T}_{Q_{X|U,W}}(u^n, w^n) \cap (\operatorname{Ker}\mathbf{H}_X + x^n)\right\},$$
$$\mathcal{E}_5 = \left\{\exists \tilde{y}^n \neq y^n, \ \tilde{y}^n \in \mathcal{T}_{Q_{Y|V,W}}(v^n, w^n) \cap (\operatorname{Ker}\mathbf{H}_Y + y^n)\right\}.$$

A.2.3 Decoding algorithm and zero-error property

All decoders determine whether there is a decoding ambiguity. If so, they recover their respective source sequence from the channel with encoding function $\phi_e^{(0)}$. If not, then the following procedure is followed.

The decoder x_1 recovers $b(w^n)$ from the channel with encoding function $\phi_e^{(0)}$ ($\mathcal{E}_1^c \Rightarrow w^n$ is defined), and using the side information u^n, it searches the V-shell $\mathcal{T}_{Q_{W|U}}(u^n)$ to find the sequences with label $b(w^n)$, and retrieves w^n ($\mathcal{E}_2^c \Rightarrow w^n$ is retrieved at decoder x_1). The coset $(\operatorname{Ker}\mathbf{H}_X + x^n)$ is then extracted from the channel with encoding function $\phi_e^{(\mathsf{x}_1)}$, the decoder x_1 then searches the set $\mathcal{T}_{Q_{X|U,W}}(u^n, w^n) \cap (\operatorname{Ker}\mathbf{H}_Y + y^n)$ and retrieves x^n with zero-error ($\mathcal{E}_4^c \Rightarrow x^n$ is retrieved).

The decoder y proceeds symmetrically and recovers y^n with the help of v^n as $\mathcal{E}_1, \mathcal{E}_3, \mathcal{E}_5$ are not realised.

The decoder x_2 determines $(\operatorname{Ker}\mathbf{H}_X + x^n) \cap \bigcup_{\overline{w}^n \in \mathcal{C}_W^{[b(w^n)]}} \mathcal{T}_{Q_{X|W}}(\overline{w}^n)$ using the syndrome $\mathbf{H}_X x^n$ transmitted in the channel with encoding function $\phi_e^{(\mathsf{x}_1)}$, and $Q_{U,V,W,X,Y}$, $b(w^n)$ sent through the channel with encoding function $\phi_e^{(0)}$. Then by using the index sent on the

channel with encoding function $\phi_e^{(\mathsf{x}_2)}$, the decoder x_2 recovers x^n with zero-error.

A.2.4 Probability of a decoding ambiguity

Let us bound $\mathbb{P}(\mathcal{E}_1)$. We have $\mathbb{P}(\mathcal{E}_1) \leq 2^{-n\delta+o(n)}$ by using Lemma 4.3.3, with $\mathcal{A} = \mathcal{W}$, $\mathcal{B} = \mathcal{U} \times \mathcal{V} \times \mathcal{X} \times \mathcal{Y}$, $P_{A,B} = Q_{W,(U,V,X,Y)}$ and $P_{\bar{A}} = Q_W$.

Now let us bound $\mathbb{P}(\mathcal{E}_2)$. The set $\mathcal{C}_W^{[b(w^n)]} \setminus \{w^n\}$ is composed of $2^{n\min(I_Q(U;W),I_Q(V;W))-n\delta+o(n)}$ sequences drawn with the same distribution Q_W. We can use Lemma 4.3.3 with $\mathcal{A} = \mathcal{W}$, $\mathcal{B} = \mathcal{U}$, $P_{A,B} = Q_{W,U}$ and $P_{\bar{A}} = Q_W$:

$$\begin{aligned}\mathbb{P}(\mathcal{E}_2) &\leq 2^{n\min(I_Q(U;W),I_Q(V;W))-n\delta-nI_Q(U;W)+o(n)} &\text{(A.47)}\\ &\leq 2^{-n\delta+o(n)}. &\text{(A.48)}\end{aligned}$$

Symmetrically, we have $\mathbb{P}(\mathcal{E}_3) \leq 2^{-n\delta+o(n)}$.

In order to bound the remaining probabilities we use Lemma 4.3.5 with $\mathcal{A} = \mathcal{X}$, $\mathcal{B} = \mathcal{U} \times \mathcal{W}$, $P_{A,B} = Q_{X,(U,W)}$:

$$\begin{aligned}\mathbb{P}(\mathcal{E}_4) &\leq 2^{n\left(\log|\mathcal{X}|-H_Q(X|U,W)-\delta-\log|\mathcal{X}|+H_Q(X|U,W)\right)+o(n)}\\ &\leq 2^{-n\delta+o(n)}, &\text{(A.49)}\end{aligned}$$

and symmetrically, we have $\mathbb{P}(\mathcal{E}_5) \leq 2^{-n\delta+o(n)}$. Thus, $\mathbb{P}(E=1) \leq 5 \cdot 2^{-\delta n+o(n)}$.

A.2.5 Rate analysis

Let $\mathcal{S} = \mathcal{U} \times \mathcal{V} \times \mathcal{W} \times \mathcal{X} \times \mathcal{Y}$, the rate on the channel with encoding function $\phi_e^{(0)}$ writes:

$$\begin{aligned}R_0^{(n)} &= \frac{1}{n}\log|\mathcal{Z}| + \frac{1}{n}\Big(1 + \mathbb{P}(E=1)n\log\big(|\mathcal{X}||\mathcal{Y}|\big)\Big)\\ &+ \frac{1}{n}\log|\Delta_n(\mathcal{S})| &\text{(A.50)}\\ &\leq I_Q(U,V,X,Y;W) - \min\Big(I_Q(U;W), I_Q(V;W)\Big) + 2\delta\\ &+ \frac{1}{n} + 5\log\big(|\mathcal{X}||\mathcal{Y}|\big)2^{-n\delta+o(n)} + \frac{1}{n}\log|\Delta_n(\mathcal{S})| &\text{(A.51)}\\ &\underset{n\to\infty}{\to} I(U,V,X,Y;W) - \min\Big(I(U;W), I(V;W)\Big) + 2\delta.\end{aligned}$$

The following rates are the exponent of the number of cosets:

$R_{\mathsf{x}_1}^{(n)} = \frac{1}{n}\log\left(2^{nH_Q(X|U,W)+n\delta}\right) \underset{n\to\infty}{\to} H(X|U,W)+\delta$, and symmetrically, $R_{\mathsf{y}}^{(n)} \underset{n\to\infty}{\to} H(Y|V,W)+\delta$.

The rate $R_{\mathsf{x}_2}^{(n)}$ on the channel with encoding function $\phi_e^{(\mathsf{x}_2)}$ writes:

$$\frac{1}{n}\log\left|(\operatorname{Ker}\mathbf{H}_X + x^n)\cap \bigcup_{\overline{w}^n\in\mathcal{C}_W^{[b(w^n)]}} \mathcal{T}_{Q_{X|W}}(\overline{w}^n)\right| \tag{A.52}$$

$$= \frac{1}{n}\log\left|\bigcup_{\overline{w}^n\in\mathcal{C}_W^{[b(w^n)]}} \mathcal{T}_{Q_{X|W}}(\overline{w}^n)\right|$$

$$+\frac{1}{n}\log|\operatorname{Ker}\mathbf{H}_X + x^n| - \log|\mathcal{X}| + o(1) \tag{A.53}$$

$$\leq \frac{1}{n}\log|\mathcal{C}_W^{[b(w^n)]}| + \frac{1}{n}\log|\mathcal{T}_{Q_{X|W}}(w^n)|$$

$$+(\log|\mathcal{X}| - H_Q(X|U,W)-\delta) - \log|\mathcal{X}| + o(1) \tag{A.54}$$

$$= \min(I_Q(U;W), I_Q(V;W)) - \delta + H_Q(X|W)$$

$$- H_Q(X|U,W) - \delta + o(1) \tag{A.55}$$

$$= \min(I_Q(U;W), I_Q(V;W)) + I_Q(X;U|W) - 2\delta + o(1) \tag{A.56}$$

$$\underset{n\to\infty}{\to} \min(I(U;W), I(V;W)) + I(X;U|W) - 2\delta \tag{A.57}$$

where (A.53) follows from the independence of the sequences in $\mathcal{C}_W$ and the entries of $\mathbf{G}_X$.

These equations hold for all $\delta > 0$ small enough, thus the rate tuple

$$\begin{pmatrix} I(X,Y,U,V;W) - \min\big(I(U;W), I(V;W)\big), \\ H(X|U,W), \\ I(X;U|W) + \min\big(I(U;W), I(V;W)\big), \\ H(Y|U,W) \end{pmatrix}, \tag{A.58}$$

is achievable. By taking the union over all distributions $P_{U,V,W,X,Y}$ we obtain the desired result.

A.3 Proof of Theorem 4.3.2

Let $S^t \doteq (U^t, V^t, X^t, Y^t)$, for all $t \leq n$. For all $i \in \{0, \mathsf{x}_1, \mathsf{y}\}$ $M_i \doteq \phi_e^{(i)}(S^n)$.

Firstly, we can use Kraft inequality as the set $\operatorname{Im}\phi_e^{(0)}$ is prefix-free, we have

$$\sum_{w^n\in\operatorname{Im}\phi_e^{(0)}} 2^{-\ell(w^n)} \leq 1. \tag{A.59}$$

Let $\kappa \doteq \sum_{w^n\in\operatorname{Im}\phi_e^{(0)}} 2^{-\ell(w^n)}$, we have $\kappa \leq 1$ and $\left(\frac{2^{-\ell(w^n)}}{\kappa}\right)_{w^n\in\operatorname{Im}\phi_e^{(0)}}$ is a probability distribution.

Secondly,

$$\begin{aligned}
nR_0 &= \mathbb{E}[\ell\circ\phi_e^{(0)}(S^n)] && \text{(A.60)}\\
&= \sum_{w^n\in\operatorname{Im}\phi_e^{(0)}} \mathbb{P}(\phi_e^{(0)}(S^n)=w^n)\ell(w^n) && \text{(A.61)}\\
&= -\sum_{w^n\in\operatorname{Im}\phi_e^{(0)}} \mathbb{P}(\phi_e^{(0)}(S^n)=w^n)\log 2^{-\ell(w^n)} && \text{(A.62)}\\
&= -\log\kappa - \sum_{w^n\in\operatorname{Im}\phi_e^{(0)}} \mathbb{P}(\phi_e^{(0)}(S^n)=w^n)\log\frac{2^{-\ell(w^n)}}{\kappa} && \text{(A.63)}\\
&\geq -\sum_{w^n\in\operatorname{Im}\phi_e^{(0)}} \mathbb{P}(\phi_e^{(0)}(S^n)=w^n)\log\frac{2^{-\ell(w^n)}}{\kappa} && \text{(A.64)}\\
&\geq -\sum_{w^n\in\operatorname{Im}\phi_e^{(0)}} \mathbb{P}(\phi_e^{(0)}(S^n)=w^n)\log\mathbb{P}(\phi_e^{(0)}(S^n)=w^n) && \text{(A.65)}\\
&= H(\phi_e^{(0)}(S^n)) = H(M_0) && \text{(A.66)}
\end{aligned}$$

where (A.64) and (A.65) respectively come from $\kappa\leq 1$ and Gibbs inequality.

In the following, $T\sim\operatorname{Unif}(\{1,...,n\})$ is a random variable independent from the other random variables of the model, $W_t \doteq (M_0, S^{t-1}, U_{t+1}^n, V_{t+1}^n)$ for all $t\leq n$. We identify $W\doteq(W_T,T)$, $(U,V,X,Y)\doteq(U_T,V_T,X_T,Y_T)$. We have

$$\begin{aligned}
nR_0 &\geq H(M_0) && \text{(A.67)}\\
&\geq I(M_0; X^n, Y^n | U^n, V^n) && \text{(A.68)}\\
&= \sum_{t\leq n} I(M_0; X_t, Y_t | U^n, V^n, X^{t-1}, Y^{t-1}) && \text{(A.69)}\\
&= \sum_{t\leq n} I(M_0, S^{t-1}, U_{t+1}^n, V_{t+1}^n; X_t, Y_t | U_t, V_t) && \text{(A.70)}\\
&\quad - I(S^{t-1}, U_{t+1}^n, V_{t+1}^n; X_t, Y_t | U_t, V_t)
\end{aligned}$$

$$= \sum_{t\leq n} I(W_t; X_t, Y_t | U_t, V_t) \tag{A.71}$$
$$= nI(W_T; X_T, Y_T | U_T, V_T, T) \tag{A.72}$$
$$= nI(W; X, Y | U, V), \tag{A.73}$$

where (A.67) comes from (A.60) to (A.66); (A.69) comes from the chain rule, (A.71) comes from the independence of $S_1, ..., S_n$, and (A.73) come from the independence of T and S^n.

Let us bound R_{x_1}:

$$nR_{\mathsf{x}_1} \geq H(M_{\mathsf{x}_1}) \tag{A.74}$$
$$\geq I(M_{\mathsf{x}_1}; X^n | M_0, U^n) \tag{A.75}$$
$$= H(X^n | M_0, U^n) - H(X^n | M_{\mathsf{x}_1}, M_0, U^n) \tag{A.76}$$
$$= \sum_{t\leq n} H(X_t | M_0, U^n, X^{t-1}) \tag{A.77}$$
$$\geq \sum_{t\leq n} H(X_t | W_t, U_t) \tag{A.78}$$
$$= nH(X_T | W_T, U_T, T) \tag{A.79}$$
$$= nH(X | W, U), \tag{A.80}$$

where (A.77) comes from the zero-error constraint. Symmetrically we obtain $R_{\mathsf{y}} \geq H(Y|W, V)$.

Now, the decoder x_2 must retrieve X with zero-error. Using Shannon lossless source coding result [18, Theorem 5.3.1], we have $R_0 + R_{\mathsf{x}_1} + R_{\mathsf{x}_2} \geq H(X)$.

A.4 Proof of Lemma 4.3.3

We use the following result, which is a consequence of [20, Lemma 2.6].

Lemma A.4.1 (Large deviations) *Let $n \in \mathbb{N}$, and $P_{A,B} \in \Delta_n(\mathcal{A}\times\mathcal{B})$. Let $P_{\bar{A}} \in \Delta(\mathcal{A})$ and $\bar{A}^n \sim (P_{\bar{A}})^{\otimes n}$, then we have for all $b^n \in \mathcal{T}_0^n(P_B)$: $\mathbb{P}\left(T_{\bar{A}^n, b^n} = P_{A,B}\right) = 2^{-n\Big(I(A;B)+D(P_A\|P_{\bar{A}})\Big)+o(n)}$.*

$$\text{If } R > R_c: \quad \mathbb{P}\left(\exists \tilde{a}^n \in \mathcal{C}^{(n)}, \text{ s.t. } T_{\tilde{a}^n, \tilde{b}^n} = P_{A,B}\right) \tag{A.81}$$
$$= 1 - \mathbb{P}\left(T_{A^n_{[1]}, \tilde{b}^n} \neq P_{A,B}\right)^{2^{nR}} \tag{A.82}$$

$$= 1 - \exp\left[2^{nR} \ln\left(1 - 2^{-nR_c + o(n)}\right)\right] \tag{A.83}$$

$$= 1 - \exp\left[-2^{n(R-R_c)+o(n)}\right], \tag{A.84}$$

where (A.82) comes from the fact that the random variables $(A^n_{[i]})$ are iid, and (A.83) is a consequence of Lemma A.4.1.

$$\text{If } R < R_c: \quad \mathbb{P}\left(\exists \tilde{a}^n \in \mathcal{C}^{(n)}, \text{ s.t. } T_{\tilde{a}^n, \tilde{b}^n} = P_{A,B}\right) \tag{A.85}$$

$$\leq \sum_{i \leq 2^{nR}} \mathbb{P}\left(T_{A^n_{[1]}, \tilde{b}^n} \neq P_{A,B}\right) \tag{A.86}$$

$$= 2^{nR} 2^{-nR_c + o(n)}. \tag{A.87}$$

The last equation comes from Lemma A.4.1, and from the fact that the variables $(A^n_{[i]})$ are identically distributed.

Appendix B

Proofs for graph-based zero-error results

B.1 Proof dependencies

An illustration of the dependencies between the results can be found in Figure B.1. Note that Theorem 5.1.4 already exists in the literature, but we provide a proof based on Lemma 5.1.3 for the sake of completeness.

B.2 Main proofs for the linearization of $\overline{H}$ and $C(\cdot, P_V)$

B.2.1 Proof of Lemma 5.2.1

In order to prove Lemma 5.2.1, we prove an asymptotic version stated in Lemma B.2.1. The proof of Lemma B.2.1 is developed in Appendix B.2.3.

Lemma B.2.1 *Let $(\overline{a}_n)_{n\in\mathbb{N}^\star} \in \mathcal{A}^{\mathbb{N}^\star}$ be any sequence such that $T_{\overline{a}^n} \to P_A$ when $n \to \infty$. Then we have*

$$\overline{H}\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right) = \lim_{n\to\infty} \frac{1}{n} H_\chi\left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge nT_{\overline{a}^n}(a)}\right). \tag{B.1}$$

Now let us prove Lemma 5.2.1. Let $(\overline{a}_n)_{n\in\mathbb{N}^\star}$ be a k-periodic sequence such that $T_{\overline{a}^k} = P_A$, then $T_{\overline{a}^{nk}} = T_{\overline{a}^k}$ for all $n \in \mathbb{N}^\star$, and $T_{\overline{a}^n} \underset{n\to\infty}{\to} P_A$. We can use Lemma B.2.1 and consider every k-th term in the limit:

$$\overline{H}\Big(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\Big) = \lim_{n\to\infty} \frac{1}{kn} H_\chi\Big(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge knT_{\overline{a}^{kn}}(a)}\Big) \tag{B.2}$$

$$= \lim_{n\to\infty} \frac{1}{kn} H_\chi\Big(\Big(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge kT_{\overline{a}^k}(a)}\Big)^{\wedge n}\Big) \tag{B.3}$$

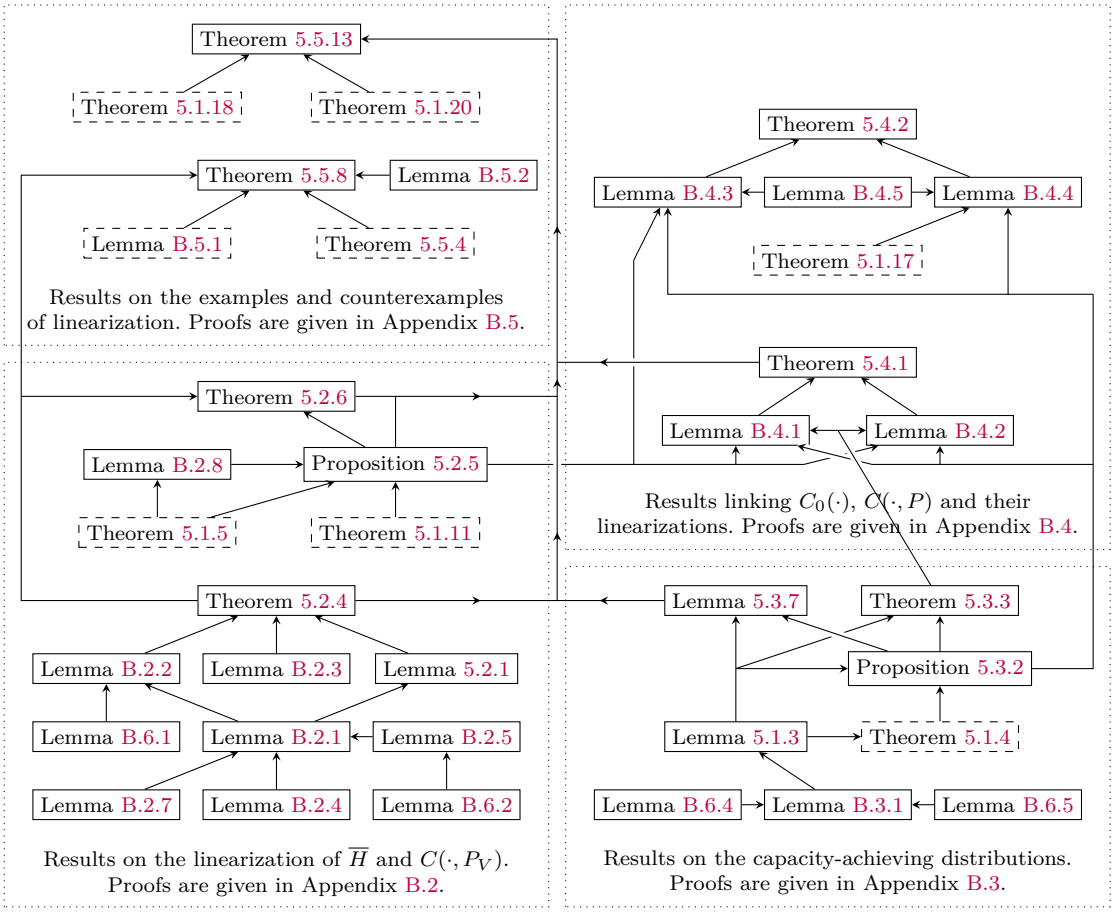


Figure B.1 – An arrow from A to B means that A is used in the proof of B. Results from the literature are represented with a dashed outline.

$$= \frac{1}{k}\overline{H}\Big(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge k P_A(a)} \Big). \tag{B.4}$$

B.2.2 Proof of Theorem 5.2.4

In order to prove Theorem 5.2.4, we will need Corollary 5.2.2, Lemma B.2.2 and Lemma B.2.3. In Lemma B.2.2 we give regularity properties of $P_A \mapsto \overline{H}\Big(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\Big)$. The proof of Lemma B.2.2 is developed in Appendix B.6.1. Lemma B.2.3 states that if a convex function γ of $\Delta(\mathcal{A})$ meets the linear interpolation of $(\gamma(\mathbb{1}_a))_{a\in\mathcal{A}}$, where $(\mathbb{1}_a)_{a\in\mathcal{A}}$ are the extreme points of $\Delta(\mathcal{A})$, then γ is linear. We use it for proving the equivalence in Theorem 5.2.4, by considering $\gamma = P_A \mapsto \overline{H}\Big(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\Big)$. The proof of Lemma B.2.3 is given in Appendix B.6.2.

Lemma B.2.2 *The function* $P_A \mapsto \overline{H}\Big(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\Big)$ *is convex and* $(\log \max_a |\mathcal{V}_a|)$*-Lipschitz.*

Lemma B.2.3 *Let $\mathcal{A}$ be a finite set, and $\gamma : \Delta(\mathcal{A}) \to \mathbb{R}$ be a convex function, and for all $a \in \mathcal{A}$, let $\mathbb{1}_a$ be the distribution that assigns 1 to the symbol a and 0 to the others. Then the following holds:*

$$\exists P_A \in \mathrm{int}(\Delta(\mathcal{A})),\ \gamma(P_A) = \sum_{a\in\mathcal{A}} P_A(a)\gamma(\mathbb{1}_a) \tag{B.5}$$

$$\iff \forall P_A \in \Delta(\mathcal{A}),\ \gamma(P_A) = \sum_{a\in\mathcal{A}} P_A(a)\gamma(\mathbb{1}_a) \tag{B.6}$$

where $\mathrm{int}(\Delta(\mathcal{A}))$ *is the interior of* $\Delta(\mathcal{A})$ *(i.e. the full-support distributions on* $\mathcal{A}$*).*

Now let us prove Theorem 5.2.4:

($\Longrightarrow$) Assume that $\overline{H}\left(\bigwedge_{a\in\mathcal{A}} G_a\right) = \sum_{a\in\mathcal{A}} \overline{H}(G_a)$.

We can use Corollary 5.2.2, which states that $\overline{H}\Big(\bigsqcup_{a\in\mathcal{A}}^{\mathrm{Unif}(\mathcal{A})} G_a\Big) = \frac{1}{|\mathcal{A}|}\overline{H}\left(\bigwedge_{a\in\mathcal{A}} G_a\right)$; hence $\overline{H}\Big(\bigsqcup_{a\in\mathcal{A}}^{\mathrm{Unif}(\mathcal{A})} G_a\Big) = \sum_{a\in\mathcal{A}} \frac{1}{|\mathcal{A}|}\overline{H}(G_a)$. Thus, the function $P_A \mapsto \overline{H}\Big(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\Big)$ satisfies (B.5) with the interior point $P_A = \mathrm{Unif}(\mathcal{A})$, and is convex by Lemma B.2.2: by Lemma B.2.3 we have

$$\forall P_A \in \Delta(\mathcal{A}),\ \overline{H}\Big(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\Big) = \sum_{a\in\mathcal{A}} P_A(a)\overline{H}(G_a). \tag{B.7}$$

($\Longleftarrow$) Conversely, assume (B.7), then $P_A \mapsto \overline{H}\Big(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\Big)$ is linear. We can use Corollary 5.2.2, and we have $\overline{H}\Big(\bigwedge_{a\in\mathcal{A}} G_a\Big) = |\mathcal{A}|\overline{H}\Big(\bigsqcup_{a\in\mathcal{A}}^{\mathrm{Unif}(\mathcal{A})} G_a\Big) = \sum_{a\in\mathcal{A}} \overline{H}(G_a)$.

B.2.3 Proof of Lemma B.2.1

We need several lemmae for this result. Lemma B.2.4 establishes the distributivity of $\wedge$ w.r.t. $\sqcup$ for probabilistic graphs, similarly as in [78] for graphs without underlying distribution. Lemma B.2.5 states that $\overline{H}$ can be computed with subgraphs induced by sets that have an asymptotic probability one, in particular we will use it with typical sets of vertices. Lemma B.2.7 gives the chromatic entropy of a disjoint union of isomorphic probabilistic graphs. The proofs of Lemma B.2.4, Lemma B.2.5 and Lemma B.2.7 are respectively given in Appendix B.6.3, Appendix B.6.4, and Appendix B.6.5.

Lemma B.2.4 *Let $\mathcal{A}, \mathcal{B}$ be finite sets, let $P_A \in \Delta(\mathcal{A})$ and $P_B \in \Delta(\mathcal{B})$. For all $a \in \mathcal{A}$ and*

$b \in \mathcal{B}$, let $G_a = (\mathcal{V}_a, \mathcal{E}_a, P_{V_a})$ and $G_b = (\mathcal{V}_b, \mathcal{E}_b, P_{V_b})$ be probabilistic graphs. Then

$$\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right) \wedge \left(\bigsqcup_{b\in\mathcal{B}}^{P_B} G_b\right) = \bigsqcup_{(a,b)\in\mathcal{A}\times\mathcal{B}}^{P_A P_B} G_a \wedge G_b. \tag{B.8}$$

Lemma B.2.5 *Let $G = (\mathcal{V}, \mathcal{E}, P_V)$, and $(\mathcal{S}_n)_{n\in\mathbb{N}^\star}$ be a sequence of sets such that for all $n \in \mathbb{N}^\star$, $\mathcal{S}_n \subseteq \mathcal{V}^n$, and $P_V^n(\mathcal{S}_n) \to 1$ when $n \to \infty$. Then $\overline{H}(G) = \lim_{n\to\infty} \frac{1}{n} H_\chi\Big(G^{\wedge n}[\mathcal{S}_n]\Big)$.*

Definition B.2.6 (Isomorphic probabilistic graphs) *Let $G_1 = (\mathcal{V}_1, \mathcal{E}_1, P_{V_1})$ and $G_2 = (\mathcal{V}_2, \mathcal{E}_2, P_{V_2})$ be two probabilistic graphs. We say that G_1 is isomorphic to G_2 (denoted by $G_1 \simeq G_2$) if there exists an isomorphism between them, i.e. a bijection $\psi : \mathcal{V}_1 \to \mathcal{V}_2$ such that:*

- *For all $v_1, v_1' \in \mathcal{V}_1$, $v_1 v_1' \in \mathcal{E}_1 \iff \psi(v_1)\psi(v_1') \in \mathcal{E}_2$,*
- *For all $v_1 \in \mathcal{V}_1$, $P_{V_1}(v_1) = P_{V_2}\Big(\psi(v_1)\Big)$.*

Lemma B.2.7 *Let $\mathcal{B}$ be a finite set, let $P_B \in \Delta(\mathcal{B})$ and let $(G_b)_{b\in\mathcal{B}}$ be a family of isomorphic probabilistic graphs, then $H_\chi\Big(\bigsqcup_{b'\in\mathcal{B}}^{P_B} G_{b'}\Big) = H_\chi(G_b)$ for all $b \in \mathcal{B}$.*

Now let us prove Lemma B.2.1. Let $P_A \in \Delta(\mathcal{A})$, and let $G = \bigsqcup_{a\in\mathcal{A}}^{P_A} G_a$. Let $(\overline{a}_n)_{n\in\mathbb{N}^\star} \in \mathcal{A}^{\mathbb{N}^\star}$ be a sequence such that $T_{\overline{a}^n} \to P_A$ when $n \to \infty$.

Let $\epsilon > 0$, and for all $n \in \mathbb{N}^\star$ let

$$\begin{aligned} &\mathcal{T}_\epsilon^n(P_A) \doteq \Big\{a^n \in \mathcal{A}^n \;\Big|\; \|T_{a^n} - P_A\|_\infty \leq \epsilon\Big\}, \\ &P'^n \doteq \frac{P_A^n}{P_A^n(\mathcal{T}_\epsilon^n(P_A))}, \qquad \mathcal{S}_{n,\epsilon} \doteq \bigsqcup_{a^n\in\mathcal{T}_\epsilon^n(P_A)} \prod_{t\leq n} \mathcal{V}_{a_t}. \end{aligned} \tag{B.9}$$

Since $P_V^n(\mathcal{S}_{n,\epsilon}) \to 1$ when $n \to \infty$, we have by Lemma B.2.5

$$\overline{H}(G) = \lim_{n\to\infty} \frac{1}{n} H_\chi\Big(G^{\wedge n}[\mathcal{S}_{n,\epsilon}]\Big), \tag{B.10}$$

Let us study the limit in (B.10). For all n large enough, $\overline{a}^n \in \mathcal{T}_\epsilon^n(P_A)$ as $T_{\overline{a}^n} \to P_A$. Therefore, for all $a^n \in \mathcal{T}_\epsilon^n(P_A)$, $a' \in \mathcal{A}$, and n large enough, we have

$$\Big|T_{\overline{a}^n}(a') - T_{a^n}(a')\Big| \leq 2\epsilon. \tag{B.11}$$

We have on one hand

$$\begin{aligned}
&H_\chi\left(\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right)^{\wedge n}[\mathcal{S}_{n,\epsilon}]\right)\\
&= H_\chi\left(\left(\bigsqcup_{a^n\in\mathcal{A}^n}^{P_A^n} \bigwedge_{t\le n} G_{a_t}\right)[\mathcal{S}_{n,\epsilon}]\right) && \text{(B.12)}\\
&= H_\chi\left(\bigsqcup_{a^n\in\mathcal{T}_\epsilon^n(P_A)}^{P'^n} \bigwedge_{t\le n} G_{a_t}\right) && \text{(B.13)}\\
&= H_\chi\left(\bigsqcup_{a^n\in\mathcal{T}_\epsilon^n(P_A)}^{P'^n} \bigwedge_{a'\in\mathcal{A}} G_{a'}^{\wedge nT_{a^n}(a')}\right) && \text{(B.14)}\\
&\le H_\chi\left(\bigsqcup_{a^n\in\mathcal{T}_\epsilon^n(P_A)}^{P'^n} \bigwedge_{a'\in\mathcal{A}} G_{a'}^{\wedge nT_{\overline{a}^n}(a')+\lceil 2n\epsilon\rceil}\right) && \text{(B.15)}\\
&= H_\chi\left(\bigwedge_{a'\in\mathcal{A}} G_{a'}^{\wedge nT_{\overline{a}^n}(a')+\lceil 2n\epsilon\rceil}\right) && \text{(B.16)}\\
&\le H_\chi\left(\bigwedge_{a'\in\mathcal{A}} G_{a'}^{\wedge nT_{\overline{a}^n}(a')}\right) + H_\chi\left(\bigwedge_{a'\in\mathcal{A}} G_{a'}^{\wedge\lceil 2n\epsilon\rceil}\right) && \text{(B.17)}\\
&\le H_\chi\left(\bigwedge_{a'\in\mathcal{A}} G_{a'}^{\wedge nT_{\overline{a}^n}(a')}\right) + \lceil 2n\epsilon\rceil|\mathcal{A}|\log|\mathcal{V}|; && \text{(B.18)}
\end{aligned}$$

where (B.12) comes from Lemma B.2.4; (B.13) comes from the definition of $\mathcal{S}_{n,\epsilon}$ and P'^n in (B.9); (B.14) is a rearrangement of the terms inside the product; (B.15) comes from (B.11); (B.16) follows from Lemma B.2.7, the graphs $\left(\bigwedge_{a'\in\mathcal{A}} G_{a'}^{\wedge nT_{\overline{a}^n}(a')+\lceil 2n\epsilon\rceil}\right)_{a^n\in\mathcal{T}_\epsilon^n(P_A)}$ are isomorphic as they do not depend on a^n; (B.17) follows from the subadditivity of H_χ; and (B.18) is the upper bound on H_χ given by the highest entropy of a coloring.

On the other hand, we obtain with similar arguments

$$\begin{aligned}
&H_\chi\left(\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right)^{\wedge n}[\mathcal{S}_{n,\epsilon}]\right)\\
&\ge H_\chi\left(\bigwedge_{a'\in\mathcal{A}} G_{a'}^{\wedge nT_{\overline{a}^n}(a')-\lceil 2n\epsilon\rceil}\right) && \text{(B.19)}\\
&\ge H_\chi\left(\bigwedge_{a'\in\mathcal{A}} G_{a'}^{\wedge nT_{\overline{a}^n}(a')}\right) - H_\chi\left(\bigwedge_{a'\in\mathcal{A}} G_{a'}^{\wedge\lceil 2n\epsilon\rceil}\right), && \text{(B.20)}\\
&\ge H_\chi\left(\bigwedge_{a'\in\mathcal{A}} G_{a'}^{\wedge nT_{\overline{a}^n}(a')}\right) - \lceil 2n\epsilon\rceil|\mathcal{A}|\log|\mathcal{V}|. && \text{(B.21)}
\end{aligned}$$

Note that (B.20) also comes from the subadditivity of H_χ, as $H_\chi(G_2) \ge H_\chi(G_1 \wedge G_2) - H_\chi(G_1)$ for all G_1, G_2.

By combining (B.18) and (B.21) we obtain

$$\begin{aligned}
&\left|\lim_{n\to\infty}\frac{1}{n}H_\chi(G^{\wedge n}[\mathcal{S}_{n,\epsilon}]) - \lim_{n\to\infty}\frac{1}{n}H_\chi\left(\bigwedge_{a'\in\mathcal{A}} G_{a'}^{\wedge nT_{\overline{a}^n}(a')}\right)\right|\\
&\le 2\epsilon|\mathcal{A}|\log|\mathcal{V}|. && \text{(B.22)}
\end{aligned}$$

As this holds for all $\epsilon > 0$, combining (B.10) and (B.22) yields the desired result.

B.2.4 Proof of Proposition 5.2.5

In order to prove Proposition 5.2.5 we need Lemma B.2.8, which is a consequence of Marton's formula in Theorem 5.1.5 applied to a disjoint union. The proof of Lemma B.2.8 can be found in Appendix B.6.6.

Lemma B.2.8 *Let $P_A \in \Delta(\mathcal{A})$, then*

$$\overline{H}\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right) + C\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a,\ \sum_{a\in\mathcal{A}} P_A(a)P_{V_a}\right) = H(P_A) + \sum_{a\in\mathcal{A}} P_A(a)H(P_{V_a}). \tag{B.23}$$

Let us prove Proposition 5.2.5. We have on one hand:

$$\begin{aligned}
&H(P_A) + \textstyle\sum_{a\in\mathcal{A}} P_A(a)C(G_a, P_{V_a}) && \text{(B.24)}\\
&= H(P_A) + \textstyle\sum_{a\in\mathcal{A}} -P_A(a)\overline{H}(G_a) + P_A(a)H(P_{V_a}) && \text{(B.25)}\\
&\leq H(P_A) - \overline{H}\left(\textstyle\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right) + \textstyle\sum_{a\in\mathcal{A}} P_A(a)H(P_{V_a}) && \text{(B.26)}\\
&= C\left(\textstyle\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a,\ \textstyle\sum_{a\in\mathcal{A}} P_A(a)P_{V_a}\right); && \text{(B.27)}
\end{aligned}$$

where (B.25) comes from Theorem 5.1.5; (B.26) follows from Theorem 5.1.11; and (B.27) follows from Lemma B.2.8. Therefore,

$$\begin{aligned}
&C\left(\textstyle\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a,\ \textstyle\sum_{a\in\mathcal{A}} P_A(a)P_{V_a}\right) = H(P_A) + \textstyle\sum_{a\in\mathcal{A}} P_A(a)C(G_a, P_{V_a}) && \text{(B.28)}\\
\iff &\overline{H}\left(\textstyle\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right) = \textstyle\sum_{a\in\mathcal{A}} P_A(a)\overline{H}(G_a). && \text{(B.29)}
\end{aligned}$$

On the other hand:

$$\begin{aligned}
\textstyle\sum_{a\in\mathcal{A}} C(G_a, P_{V_a}) &= \textstyle\sum_{a\in\mathcal{A}} -\overline{H}(G_a) + H(P_{V_a}) && \text{(B.30)}\\
&\leq -\overline{H}\left(\textstyle\bigwedge_{a\in\mathcal{A}} G_a\right) + \textstyle\sum_{a\in\mathcal{A}} H(P_{V_a}) && \text{(B.31)}\\
&= -\overline{H}\left(\textstyle\bigwedge_{a\in\mathcal{A}} G_a\right) + H\left(\textstyle\bigotimes_{a\in\mathcal{A}} P_{V_a}\right) && \text{(B.32)}\\
&= C\left(\textstyle\bigwedge_{a\in\mathcal{A}} G_a,\ \textstyle\bigotimes_{a\in\mathcal{A}} P_{V_a}\right); && \text{(B.33)}
\end{aligned}$$

where (B.30) comes from Theorem 5.1.5; (B.31) follows from Theorem 5.1.11; and (B.33) also follows from Theorem 5.1.5. Therefore,

$$\textstyle\sum_{a\in\mathcal{A}} C(G_a, P_{V_a}) = C\left(\textstyle\bigwedge_{a\in\mathcal{A}} G_a,\ \textstyle\bigotimes_{a\in\mathcal{A}} P_{V_a}\right) \tag{B.34}$$

$$\iff \overline{H}\left(\bigwedge_{a\in\mathcal{A}} G_a\right) = \sum_{a\in\mathcal{A}} \overline{H}(G_a). \tag{B.35}$$

B.3 Main proofs for the capacity-achieving distributions

B.3.1 Proof of Lemma 5.1.3

Lemma B.3.1 enables us to replace the limit superior in the definition of $C(G, P_V)$ by a regular limit. The proof is given in Appendix B.6.7 and uses superadditivity arguments.

Lemma B.3.1 *For all $\epsilon > 0$, we have*

$$\limsup_{n\to\infty} \frac{1}{n}\alpha\Big(G^{\wedge n}[\mathcal{T}_\epsilon^n(P_V)]\Big) = \sup_{n\in\mathbb{N}^\star} \frac{1}{n}\alpha\Big(G^{\wedge n}[\mathcal{T}_\epsilon^n(P_V)]\Big) = \lim_{n\to\infty} \frac{1}{n}\alpha\Big(G^{\wedge n}[\mathcal{T}_\epsilon^n(P_V)]\Big). \tag{B.36}$$

Now let us prove Lemma 5.1.3.

By definition of $C(G, P_V)$ and Lemma B.3.1,

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha(G^{\wedge n}[\mathcal{T}_{\epsilon'}^n(P_V)]) \underset{\epsilon'\to 0}{\to} C(G, P_V). \tag{B.37}$$

Thus, there exists a mapping $\epsilon : \mathbb{N}^\star \to \mathbb{R}_+^\star$ such that $\epsilon(k) \underset{k\to\infty}{\to} 0$ and for all $k \in \mathbb{N}^\star$,

$$\left|\lim_{n'\to\infty} \frac{1}{n'} \log \alpha(G^{\wedge n'}[\mathcal{T}_{\epsilon(k)}^{n'}(P_V)]) - C(G, P_V)\right| \leq \frac{1}{k}. \tag{B.38}$$

Now, let $m(k) \in \mathbb{N}^\star$ such that for all $n \geq m(k)$,

$$\left|\frac{1}{n} \log \alpha(G^{\wedge n}[\mathcal{T}_{\epsilon(k)}^{n}(P_V)]) - \lim_{n'\to\infty} \frac{1}{n'} \log \alpha(G^{\wedge n'}[\mathcal{T}_{\epsilon(k)}^{n'}(P_V)])\right| \leq \frac{1}{k}. \tag{B.39}$$

It can be easily observed that (B.39) is also satisfied if $n \geq \max(k, m(k))$. Therefore, for all $k \in \mathbb{N}^\star$, and $n \geq \max(k, m(k))$, we have

$$\left|\frac{1}{n} \log \alpha(G^{\wedge n}[\mathcal{T}_{\epsilon(k)}^{n}(P_V)]) - C(G, P_V)\right| \leq \frac{1}{2k}. \tag{B.40}$$

Now we can build the desired sequence of codebooks. For all $n \in \mathbb{N}^\star$, let $\mathcal{C}_n \subseteq \mathcal{V}^n$ be a maximal independent set in $G^{\wedge n}[\mathcal{T}_{\epsilon(k(n))}^n(P_V)]$; where $k(n)$ is the biggest k such that

$n \geq \max(k, m(k))$. Note that $k(n)$ exists as $k(n) \leq n$, and

$$k(n) \underset{n\to\infty}{\to} \infty. \tag{B.41}$$

We have

$$\left|\frac{\log|\mathcal{C}_n|}{n} - C(G, P_V)\right| = \left|\frac{1}{n}\log\alpha(G^{\wedge n}[\mathcal{T}^n_{\epsilon(k(n))}(P_V)]) - C(G, P_V)\right| \tag{B.42}$$

$$\leq \frac{1}{2k(n)} \tag{B.43}$$

$$\underset{n\to\infty}{\to} 0; \tag{B.44}$$

where (B.42) follows from the construction of $\mathcal{C}_n$; (B.43) is a consequence of (B.40); and (B.44) follows from (B.41).

Since for all $k \in \mathbb{N}^\star$, $\mathcal{C}_n \subseteq \mathcal{T}^n_{\epsilon(k(n))}(P_V)$, we also have

$$\max_{v^n\in\mathcal{C}_n} \|T_{v^n} - P_V\|_\infty \leq \epsilon(k(n)) \underset{n\to\infty}{\to} 0. \tag{B.45}$$

Conversely, assume that for all $n \in \mathbb{N}^\star$, $\mathcal{C}_n \subseteq \mathcal{V}^n$ is an independent set in $G^{\wedge n}$, such that

$$\max_{v^n\in\mathcal{C}_n} \|T_{v^n} - P_V\|_\infty \underset{n\to\infty}{\to} 0. \tag{B.46}$$

Then for all $\epsilon > 0$ and n large enough, $\mathcal{C}_n \in \mathcal{T}^n_\epsilon(P_V)$ and

$$\frac{1}{n}\log|\mathcal{C}_n| \leq \frac{1}{n}\log\alpha\Big(G^{\wedge n}[\mathcal{T}^n_\epsilon(P_V)]\Big). \tag{B.47}$$

By taking the limit superior when $n \to \infty$ and $\epsilon \to 0$ we obtain

$$\limsup_{n\to\infty} \frac{1}{n}\log|\mathcal{C}_n| \leq C(G, P_V). \tag{B.48}$$

B.3.2 Proof of Theorem 5.1.4

($\leq$) By definition of C_0 and $C(G, P_V)$ we have

$$\sup_{P_V\in\Delta(\mathcal{V})} C(G, P_V) = \sup_{P_V\in\Delta(\mathcal{V})} \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n}\log\alpha\Big(G^{\wedge n}[\mathcal{T}^n_\epsilon(P_V)]\Big) \tag{B.49}$$

$$\leq \sup_{P_V \in \Delta(\mathcal{V})} \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \alpha\Big(G^{\wedge n}\Big) \tag{B.50}$$

$$= C_0(G). \tag{B.51}$$

($\geq$) Let $(\mathcal{C}_n)_{n \in \mathbb{N}^\star}$ be a sequence such that for all $n \in \mathbb{N}^\star$, $\mathcal{C}_n$ is an independent set in $G^{\wedge n}$, and

$$\lim_{n \to \infty} \frac{1}{n} \log |\mathcal{C}_n| = C_0(G); \tag{B.52}$$

the existence of the sequence $(\mathcal{C}_n)_{n \in \mathbb{N}^\star}$ follows from the definition of C_0.

Let $(\tau_n)_{n \in \mathbb{N}^\star}$ be the sequence defined by: for all $n \in \mathbb{N}^\star$,

$$\tau_n \doteq \frac{1}{|\mathcal{C}_n|} \sum_{v^n \in \mathcal{C}_n} T_{v^n}. \tag{B.53}$$

The terms of the sequence $(\tau_n)_{n \in \mathbb{N}^\star}$ are in $\Delta(\mathcal{V})$, which is a compact set. Therefore, by Bolzano-Weierstrass theorem, $(\tau_n)_{n \in \mathbb{N}^\star}$ has a convergent subsequence $(\tau_{\phi(n)})_{n \in \mathbb{N}^\star}$, where $\phi : \mathbb{N}^\star \to \mathbb{N}^\star$ is strictly increasing. We denote by $(\mathcal{C}_{\phi(n)})_{n \in \mathbb{N}^\star}$ the corresponding subsequence of independent sets, and

$$P_V^* \doteq \lim_{n \to \infty} \tau_{\phi(n)} \in \Delta(\mathcal{V}). \tag{B.54}$$

By construction, we also have

$$\lim_{n \to \infty} \frac{\log |\mathcal{C}_{\phi(n)}|}{\phi(n)} = C_0(G). \tag{B.55}$$

Let us build an adequate sequence of codebooks with type converging uniformly to P_V^*, and with asymptotic rate $C_0(G)$. For all $n \in \mathbb{N}^\star$, let

$$\mathcal{C}^*_{n\phi(n)} \doteq (\mathcal{C}_{\phi(n)})^n \cap \mathcal{T}^{n\phi(n)}_{\epsilon_n}(P_V^*), \tag{B.56}$$

where $\epsilon_n \doteq \|P_V^* - \tau_{\phi(n)}\|_\infty + \frac{1}{\sqrt[4]{n}}$.

It can be easily observed that $\epsilon_n \underset{n \to \infty}{\to} 0$ and $\mathcal{C}^*_{n\phi(n)} \subseteq \mathcal{T}^{n\phi(n)}_{\epsilon_n}(P_V^*)$: by construction we have

$$\max_{v^{n\phi(n)} \in \mathcal{C}^*_{n\phi(n)}} \|T_{v^{n\phi(n)}} - P_V^*\|_\infty \underset{n \to \infty}{\to} 0. \tag{B.57}$$

Furthermore, for all $n \in \mathbb{N}^\star$, $\mathcal{C}^*_{n\phi(n)}$ is independent in $G^{\wedge n\phi(n)}$, as it is contained in the independent set $(\mathcal{C}_{\phi(n)})^n$.

Now let us prove that $\frac{\log|\mathcal{C}^*_{n\phi(n)}|}{n\phi(n)} \to C_0(G)$ when $n \to \infty$. Let us draw a codeword

$$C^{n\phi(n)} = (C_1^{\phi(n)}, ..., C_n^{\phi(n)}) \tag{B.58}$$

uniformly from $(\mathcal{C}_{\phi_n})^n$, and show that it is in $\mathcal{T}^{n\phi(n)}_{\epsilon_n}(P_V^*)$ with high probability. On one hand, for all $t \leq n$, the average type of $C_t^{\phi(n)}$ writes

$$\mathbb{E}\left[T_{C_t^{\phi(n)}}\right] = \frac{1}{|\mathcal{C}_{\phi(n)}|} \sum_{c^{\phi(n)} \in \mathcal{C}_{\phi(n)}} T_{c^{\phi(n)}} = \tau_{\phi(n)}. \tag{B.59}$$

On the other hand,

$$\begin{aligned}
\frac{|\mathcal{C}^*_{n\phi(n)}|}{|(\mathcal{C}_{\phi(n)})^n|} &= \frac{|(\mathcal{C}_{\phi(n)})^n \cap \mathcal{T}^{n\phi(n)}_{\epsilon_n}(P_V^*)|}{|(\mathcal{C}_{\phi(n)})^n|} && \text{(B.60)}\\
&= \mathbb{P}\Big(C^{n\phi(n)} \in \mathcal{T}^{n\phi(n)}_{\epsilon_n}(P_V^*)\Big) && \text{(B.61)}\\
&= \mathbb{P}\left(\|T_{C^{n\phi(n)}} - P_V^*\|_\infty \leq \epsilon_n\right) && \text{(B.62)}\\
&\geq \mathbb{P}\left(\|T_{C^{n\phi(n)}} - \tau_{\phi(n)}\|_\infty + \|\tau_{\phi(n)} - P_V^*\|_\infty \leq \epsilon_n\right) && \text{(B.63)}\\
&= \mathbb{P}\left(\|T_{C^{n\phi(n)}} - \tau_{\phi(n)}\|_\infty \leq n^{-1/4}\right) && \text{(B.64)}\\
&= \mathbb{P}\left(\left\|\tfrac{1}{n}\textstyle\sum_{t\leq n} T_{C_t^{\phi(n)}} - \tau_{\phi(n)}\right\|_\infty \leq n^{-1/4}\right) && \text{(B.65)}\\
&= 1 - \mathbb{P}\left(\left\|\left(\textstyle\sum_{t\leq n} T_{C_t^{\phi(n)}}\right) - n\tau_{\phi(n)}\right\|_\infty > n^{3/4}\right) && \text{(B.66)}\\
&\geq 1 - \textstyle\sum_{v\in\mathcal{V}} \mathbb{P}\left(\left|\sum_{t\leq n} T_{C_t^{\phi(n)}}(v) - n\tau_{\phi(n)}(v)\right| > n^{3/4}\right) && \text{(B.67)}\\
&\geq 1 - \textstyle\sum_{v\in\mathcal{V}} \frac{1}{n^{3/2}} \mathbb{V}\left[\sum_{t\leq n} T_{C_t^{\phi(n)}}(v)\right] && \text{(B.68)}\\
&\geq 1 - \tfrac{|\mathcal{V}|}{n^{1/2}} \underset{n\to\infty}{\to} 1; && \text{(B.69)}
\end{aligned}$$

where (B.61) and (B.65) come from the construction of $C^{n\phi(n)}$; (B.64) comes from the construction of ϵ_n; (B.67) follows from the union bound; (B.68) comes from Chebyshev's inequality and (B.59); (B.69) follows from $\mathbb{V}\left[\sum_{t\leq n} T_{C_t^{\phi(n)}}(v)\right] = \sum_{t\leq n} \mathbb{V}\left[T_{C_t^{\phi(n)}}(v)\right] \leq n$, as the random variables $T_{C_t^{\phi(n)}}(v)$ are iid and takes values in $[0,1]$. Hence

$$\lim_{n\to\infty} \frac{\log|\mathcal{C}^*_{n\phi(n)}|}{n\phi(n)} = \lim_{n\to\infty} \frac{\log|(\mathcal{C}_{\phi(n)})^n|}{n\phi(n)} = C_0(G). \tag{B.70}$$

By combining (B.57), (B.70) and Lemma 5.1.3, it follows that

$$C_0(G) = \lim_{n\to\infty} \frac{\log |\mathcal{C}^*_{n\phi(n)}|}{n\phi(n)} \leq C(G, P_V). \tag{B.71}$$

B.3.3 Proof of Proposition 5.3.2

Let us show that for all graph $G = (\mathcal{V}, \mathcal{E})$, the function $P_V \mapsto C(G, P_V)$ is concave. Let $P_V, P'_V \in \Delta(\mathcal{V})$ and $\beta \in [0, 1]$. Let $(b_n)_{n\in\mathbb{N}}$ be a sequence of integers such that $\frac{b_n}{n} \underset{n\to\infty}{\to} \beta$.

By Lemma 5.1.3, there exists two sequences $(\mathcal{C}_n)_{n\in\mathbb{N}}$ and $(\mathcal{C}'_n)_{n\in\mathbb{N}}$ that satisfy the following:

$$\forall n \in \mathbb{N}^\star,\ \mathcal{C}_n \subseteq \mathcal{V}^n \text{ and } \mathcal{C}'_n \subseteq \mathcal{V}^n \text{ are independent in } G^{\wedge n}; \tag{B.72}$$

and

$$\frac{\log |\mathcal{C}_n|}{n} \underset{n\to\infty}{\to} C(G, P_V), \qquad \frac{\log |\mathcal{C}'_n|}{n} \underset{n\to\infty}{\to} C(G, P'_V), \tag{B.73}$$

$$\max_{v^n\in\mathcal{C}_n} \|T_{v^n} - P_V\|_\infty \underset{n\to\infty}{\to} 0, \qquad \max_{v^n\in\mathcal{C}'_n} \|T_{v^n} - P_V\|_\infty \underset{n\to\infty}{\to} 0. \tag{B.74}$$

Let us build a sequence of codebooks $(\mathcal{C}''_n)_{n\in\mathbb{N}^\star}$ adapted to the distribution $\beta P_V + (1-\beta)P'_V$ by using a time-sharing between $(\mathcal{C}_n)_{n\in\mathbb{N}^\star}$ and $(\mathcal{C}'_n)_{n\in\mathbb{N}^\star}$. For all $n \in \mathbb{N}^\star$, let

$$\mathcal{C}''_n \doteq \mathcal{C}_n^{b_n} \times \mathcal{C}'^{n-b_n}_n. \tag{B.75}$$

For all $n \in \mathbb{N}^\star$, $\mathcal{C}''_n \subseteq \mathcal{V}^{n^2}$ is independent in $G^{\wedge n^2}$ as a product of independent sets.

The rate associated to $\mathcal{C}''_n$ writes

$$\frac{\log |\mathcal{C}''_n|}{n^2} = \frac{b_n \log |\mathcal{C}_n| + (n - b_n) \log |\mathcal{C}'_n|}{n^2} \tag{B.76}$$

$$= \frac{b_n}{n}\frac{\log |\mathcal{C}_n|}{n} + \frac{n - b_n}{n}\frac{\log |\mathcal{C}'_n|}{n} \tag{B.77}$$

$$\underset{n\to\infty}{\to} \beta C(G, P_V) + (1-\beta)C(G, P'_V); \tag{B.78}$$

and the types of the codewords in $\mathcal{C}''_n$ satisfy

$$\max_{v^{n^2}\in\mathcal{C}''_n} \left\| T_{v^{n^2}} - \beta P_V - (1-\beta)P'_V \right\|_\infty \tag{B.79}$$

$$= \max_{v^{nb_n}\in\mathcal{C}_n} \max_{v'^{n(n-b_n)}\in\mathcal{C}'_n} \left\| \frac{nb_n}{n^2} T_{v^{nb_n}} + \frac{n(n-b_n)}{n^2} T_{v'^{n(n-b_n)}} - \beta P_V - (1-\beta)P'_V \right\|_\infty \tag{B.80}$$

$$\leq \max_{v^{nb_n}\in\mathcal{C}_n} \left\| \frac{b_n}{n} T_{v^{nb_n}} - \beta P_V \right\|_\infty + \max_{v'^{n(n-b_n)}\in\mathcal{C}'_n} \left\| \frac{n-b_n}{n} T_{v'^{n(n-b_n)}} - (1-\beta)P'_V \right\|_\infty \tag{B.81}$$

$$= \beta \max_{v^{nb_n}\in\mathcal{C}_n} \left\| T_{v^{nb_n}} - P_V + o(1) T_{v^{nb_n}} \right\|_\infty$$

$$+ (1-\beta) \max_{v'^{n(n-b_n)}\in\mathcal{C}'_n} \left\| T_{v'^{n(n-b_n)}} - P'_V + o(1) T_{v'^{n(n-b_n)}} \right\|_\infty \tag{B.82}$$

$$\leq \beta \max_{v^{nb_n}\in\mathcal{C}_n} \left\| T_{v^{nb_n}} - P_V \right\|_\infty + o(1) \left\| T_{v^{nb_n}} \right\|_\infty$$

$$+ (1-\beta) \max_{v'^{n(n-b_n)}\in\mathcal{C}'_n} \left\| T_{v'^{n(n-b_n)}} - P'_V \right\|_\infty + o(1) \left\| T_{v'^{n(n-b_n)}} \right\|_\infty \tag{B.83}$$

$$\underset{n\to\infty}{\to} 0. \tag{B.84}$$

By Lemma 5.1.3, $\lim_{n\to\infty} \frac{\log|\mathcal{C}''_n|}{n^2} \leq C(G, \beta P_V + (1-\beta)P'_V)$, thus

$$\beta C(G, P_V) + (1-\beta) C(G, P'_V) \leq C(G, \beta P_V + (1-\beta)P'_V). \tag{B.85}$$

The function $P_V \mapsto C(G, P_V)$ is concave on the convex compact set $\Delta(\mathcal{V})$, therefore its set of maximizers $\mathcal{P}^*(G) = \operatorname{argmax}_{P_V\in\Delta(\mathcal{V})} C(G, P_V)$ is convex. Furthermore, by Theorem 5.1.4, the set $\mathcal{P}^*(G)$ is nonempty and satisfies

$$\forall P_V \in \mathcal{P}^*(G),\ C(G, P_V) = C_0(G). \tag{B.86}$$

B.3.4 Proof of Theorem 5.3.3

The proof techniques used here are similar as in the proof of Theorem 5.1.4 in Appendix B.3.2.

Let us start by showing that Theorem 5.3.3 is true when $\mathcal{A}$ has two elements. Let $G = (\mathcal{V}, \mathcal{E})$, and $G' = (\mathcal{V}', \mathcal{E}')$ be two graphs, and let $P_{V,V'} \in \mathcal{P}^*(G \wedge G')$. We will prove that $P_V \otimes P_{V'}$ is also capacity-achieving by building an adequate sequence of codebooks.

For all $n \in \mathbb{N}^\star$, let $\mathcal{C}_n \subseteq (\mathcal{V}\times\mathcal{V}')^n$ such that $\mathcal{C}_n$ is an independent set in $(G \wedge G')^{\wedge n}$, and

$$\frac{1}{n}\log|\mathcal{C}_n| \underset{n\to\infty}{\to} C_0(G \wedge G'), \tag{B.87}$$

$$\max_{(v^n, v'^n)\in\mathcal{C}_n} \left\| T_{v^n, v'^n} - P_{V,V'} \right\|_\infty \underset{n\to\infty}{\to} 0. \tag{B.88}$$

The existence of such a sequence is given by Lemma 5.1.3, and Proposition 5.3.2. Let

$$Q_{V,V'}^{(n)} \doteq \frac{1}{|\mathcal{C}_n|} \sum_{(v^n,v'^n)\in\mathcal{C}_n} T_{v^n,v'^n}. \tag{B.89}$$

An immediate observation is that

$$Q_{V,V'}^{(n)} \underset{n\to\infty}{\to} P_{V,V'} \tag{B.90}$$

as a consequence of (B.88).

Let us build a sequence of codebooks with asymptotic rate $C_0(G \wedge G')$, such that the type of their codewords converge uniformly to $P_V \otimes P_{V'}$:

$$\mathcal{C}_{n^3}^* \doteq \mathcal{T}_{\epsilon_n}^{n^3}(P_V \otimes P_{V'}) \cap \left(\prod_{t\leq n} \mathcal{C}_n^{(t)}\right)^n; \tag{B.91}$$

where

$$\epsilon_n \doteq \|Q_V^{(n)} \otimes Q_{V'}^{(n)} - P_V \otimes P_{V'}\|_\infty + \tfrac{1}{\sqrt[4]{n}}; \tag{B.92}$$

and where for all $t \leq n$, the shifted codebook $\mathcal{C}_n^{(t)}$ is defined by

$$\mathcal{C}_n^{(t)} \doteq \left\{ \left((v_t, v_{t+1}, ..., v_n, v_1, ..., v_{t-1}), v'^n \right) \;\middle|\; (v^n, v'^n) \in \mathcal{C}_n \right\}. \tag{B.93}$$

By construction, $\mathcal{C}_{n^3}^* \subseteq \mathcal{T}_{\epsilon_n}^{n^3}(P_V \otimes P_{V'})$ thanks to (B.91), and $\epsilon_n \underset{n\to\infty}{\to} 0$ thanks to (B.92) and (B.90); therefore we have

$$\max_{v^{n^3}\in\mathcal{C}_n^*} \|T_{v^{n^3}} - P_V \otimes P_{V'}\|_\infty \underset{n\to\infty}{\to} 0. \tag{B.94}$$

Furthermore, $\mathcal{C}_{n^3}^*$ is an independent set in $(G \wedge G')^{\wedge n^3}$, as it is contained in the product independent set $\left(\prod_{t\leq n} \mathcal{C}_n^{(t)}\right)^n$; note that this holds because the shifted codebook $\mathcal{C}_n^{(t)}$ is an independent set in $(G \wedge G')^{\wedge n}$ for all $t \leq n$.

Now let us prove that $\frac{\log|\mathcal{C}_{n^3}^*|}{n^3} \underset{n\to\infty}{\to} C_0(G \wedge G')$. Let us draw a codeword uniformly from $\left(\prod_{t\leq n} \mathcal{C}_n^{(t)}\right)^n$:

$$C^{n^3} \doteq (C_1^{n^2}, ..., C_n^{n^2}), \tag{B.95}$$

where for all $t \leq n$, $C_t^{n^2}$ is a random $n \times n$-sequence drawn uniformly from $\prod_{t\leq n} \mathcal{C}_n^{(t)}$. We want to prove that $C^{n^3} \in \mathcal{T}_{\epsilon_n}^{n^3}(P_V \otimes P_{V'})$ with high probability.

On one hand we have to determine the average type of the random variables $(C_t^{n^2})_{t\leq n}$ which are iid copies of $C^{n^2} = (C_1^n, ..., C_n^n)$; where each C_t^n is drawn uniformly from $\mathcal{C}_n^{(t)}$, and the $(C_t^n)_{t\leq n}$ are mutually independent.

$$\mathbb{E}\left[T_{C_t^{n^2}}\right] = \frac{1}{n}\sum_{t\leq n}\mathbb{E}\left[T_{C_t^n}\right] \tag{B.96}$$

$$= \frac{1}{n}\sum_{t\leq n}\frac{1}{|\mathcal{C}_n^{(t)}|}\sum_{(v^n,v'^n)\in\mathcal{C}_n^{(t)}} T_{v^n,v'^n} \tag{B.97}$$

$$= \frac{1}{n}\sum_{t\leq n}\frac{1}{|\mathcal{C}_n|}\sum_{(v^n,v'^n)\in\mathcal{C}_n} T_{\sigma_t(v^n),v'^n} \tag{B.98}$$

$$= \frac{1}{|\mathcal{C}_n|}\sum_{(v^n,v'^n)\in\mathcal{C}_n}\frac{1}{n}\sum_{t\leq n} T_{\sigma_t(v^n),v'^n} \tag{B.99}$$

$$= \frac{1}{|\mathcal{C}_n|}\sum_{(v^n,v'^n)\in\mathcal{C}_n} T_{v^n}\otimes T_{v'^n} \tag{B.100}$$

$$= Q_V^{(n)}\otimes Q_{V'}^{(n)}, \tag{B.101}$$

where $\sigma_t(v^n) = (v_t, v_{t+1}, ..., v_n, v_1, ..., v_{n-1})$; (B.98) comes from the construction of $\mathcal{C}_n^{(t)}$ in (B.93); and (B.100) comes from the following observation:

$$\sum_{t\leq n} T_{\sigma_t(v^n),v'^n} = \sum_{t\leq n}\sum_{s\leq n} T_{v_{s+t},v'_s} = \sum_{s\leq n}\sum_{t\leq n} T_{v_{s+t},v'_s} \tag{B.102}$$

$$= \sum_{s\leq n} T_{v^n,(v'_s,...,v'_s)} = \sum_{s\leq n} T_{v^n}\otimes T_{v'_s} = T_{v^n}\otimes T_{v'^n}, \tag{B.103}$$

where the index $s+t$ is taken modulo n.

On the other hand we have

$$\frac{|\mathcal{C}_{n^3}^*|}{\left|\left(\prod_{t\leq n}\mathcal{C}_n^{(t)}\right)^n\right|} \tag{B.104}$$

$$= \frac{\left|\mathcal{T}_{\epsilon_n}^{n^3}(P_V\otimes P_{V'})\cap\left(\prod_{t\leq n}\mathcal{C}_n^{(t)}\right)^n\right|}{\left|\left(\prod_{t\leq n}\mathcal{C}_n^{(t)}\right)^n\right|} \tag{B.105}$$

$$= \mathbb{P}\left(C^{n^3}\in\mathcal{T}_{\epsilon_n}^{n^3}(P_V\otimes P_{V'})\right) \tag{B.106}$$

$$= \mathbb{P}\left(\left\|\tfrac{1}{n}\textstyle\sum_{t\leq n} T_{C_t^{n^2}} - P_V\otimes P_{V'}\right\|_\infty \leq \epsilon_n\right) \tag{B.107}$$

$$\geq \mathbb{P}\left(\left\|\tfrac{1}{n}\sum_{t\leq n} T_{C_t^{n^2}} - Q_V^{(n)}\otimes Q_{V'}^{(n)}\right\|_\infty + \left\|Q_V^{(n)}\otimes Q_{V'}^{(n)} - P_V\otimes P_{V'}\right\|_\infty \leq \epsilon_n\right) \tag{B.108}$$

$$= \mathbb{P}\left(\left\|\sum_{t\leq n} T_{C_t^{n^2}} - nQ_V^{(n)}\otimes Q_{V'}^{(n)}\right\|_\infty \leq n^{3/4}\right) \tag{B.109}$$

$$\geq 1 - \sum_{(v,v')\in\mathcal{V}\times\mathcal{V}'} \mathbb{P}\left(\left|\sum_{t\leq n} T_{C_t^{n^2}}(v,v') - nQ_V^{(n)}\otimes Q_{V'}^{(n)}(v,v')\right| > n^{3/4}\right) \tag{B.110}$$

$$\geq 1 - \sum_{(v,v')\in\mathcal{V}\times\mathcal{V}'} \tfrac{1}{n^{3/2}}\mathbb{V}\left[\sum_{t\leq n} T_{C_t^{n^2}}(v,v')\right] \tag{B.111}$$

$$\geq 1 - \tfrac{|\mathcal{V}||\mathcal{V}'|}{n^{1/2}} \underset{n\to\infty}{\to} 1; \tag{B.112}$$

where (B.106) and (B.107) come from the construction of C^{n^3}; (B.109) comes from the construction of ϵ_n; (B.110) follows from the union bound; (B.111) comes from Chebyshev's inequality and (B.101); and (B.112) comes from the fact that $\mathbb{V}\left[\sum_{t\leq n} T_{C_t^{n^2}}(v,v')\right] = \sum_{t\leq n}\mathbb{V}\left[T_{C_t^{n^2}}(v,v')\right] \leq n$, as the random variables $T_{C_t^{n^2}}(v,v')$ are iid and takes values in $[0,1]$. Hence

$$\lim_{n\to\infty}\frac{\log|\mathcal{C}^*_{n^3}|}{n^3} = \lim_{n\to\infty}\frac{\log\left|\left(\prod_{t\leq n}\mathcal{C}_n^{(t)}\right)^n\right|}{n^3} = \lim_{n\to\infty}\frac{\log|\mathcal{C}_n|}{n} = C_0(G\wedge G'); \tag{B.113}$$

where the second equality holds as the shifted codebooks $(\mathcal{C}_n^{(t)})_{t\leq n}$ all have cardinality $|\mathcal{C}_n|$.

Thus, by combining (B.113), Lemma 5.1.3, and Proposition 5.3.2 we obtain

$$C_0(G\wedge G') = \lim_{n\to\infty}\frac{\log|\mathcal{C}^*_{n^3}|}{n^3} \leq C(G\wedge G', P_V\otimes P_{V'}) \leq C_0(G\wedge G'), \tag{B.114}$$

hence $P_V\otimes P_{V'}\in\mathcal{P}^*(G\wedge G')$.

Therefore, Theorem 5.3.3 is proved when $\mathcal{A}$ has two elements:

$$P_{V,V'}\in\mathcal{P}^*(G\wedge G') \Longrightarrow P_V\otimes P_{V'}\in\mathcal{P}^*(G\wedge G'). \tag{B.115}$$

Now let us consider the case where $\mathcal{A}$ has a cardinality greater than 2. Let $P_{V_1,\dots,V_{\mathcal{A}}}\in\mathcal{P}^*(\bigwedge_{a\in\mathcal{A}}G_a)$. By considering the product graphs

$$\bigwedge_{a\in\mathcal{A}}G_a = \left(\bigwedge_{1\leq i<i^*}G_i\right)\wedge\left(\bigwedge_{i^*\leq i\leq|\mathcal{A}|}G_i\right); \tag{B.116}$$

for all $i^*\leq|\mathcal{A}|$, and applying (B.115) successively, we obtain

$$P_{V_1,\dots,V_{\mathcal{A}}}\in\mathcal{P}^*\left(\bigwedge_{a\in\mathcal{A}}G_a\right) \Longrightarrow P_{V_1}\otimes P_{V_2,\dots,V_{|\mathcal{A}|}}\in\mathcal{P}^*\left(\bigwedge_{a\in\mathcal{A}}G_a\right) \tag{B.117}$$

$$\implies (P_{V_1} \otimes P_{V_2}) \otimes P_{V_3,\dots,V_{|\mathcal{A}|}} \in \mathcal{P}^* \left(\bigwedge_{a\in\mathcal{A}} G_a\right) \tag{B.118}$$

$$\implies \dots \tag{B.119}$$

$$\implies \bigotimes_{a\in\mathcal{A}} P_{V_a} \in \mathcal{P}^* \left(\bigwedge_{a\in\mathcal{A}} G_a\right). \tag{B.120}$$

B.3.5 Proof of Lemma 5.3.7

Let G be a vertex-transitive graph, and let $P_V \in \mathcal{P}^*(G)$. Let $\psi \in \mathrm{Aut}(G)$, we first prove that $P_{\psi(V)} \in \mathcal{P}^*(G)$, then we will conclude by using the convexity of $\mathcal{P}^*(G)$.

Let $(\mathcal{C}_n)_{n\in\mathbb{N}^\star}$ be a sequence such that

$$\forall n \in \mathbb{N}^\star,\ \mathcal{C}_n \subseteq \mathcal{V}^n \text{ is an independent set in } G^{\wedge n}, \tag{B.121}$$

$$\max_{v^n\in\mathcal{C}_n} \|T_{v^n} - P_V\|_\infty \underset{n\to\infty}{\to} 0, \tag{B.122}$$

$$\frac{\log|\mathcal{C}_n|}{n} \underset{n\to\infty}{\to} C(G, P_V) = C_0(G). \tag{B.123}$$

The existence of such a sequence is given by Lemma 5.1.3. Note that the last equality in (B.123) comes from the assumption $P_V \in \mathcal{P}^*(G)$.

Now, for all $n \in \mathbb{N}^\star$ the codebook

$$\psi(\mathcal{C}_n) \doteq \{(\psi(v_1), \dots, \psi(v_n)) \mid v^n \in \mathcal{C}_n\} \tag{B.124}$$

is also independent in $G^{\wedge n}$, as ψ is a graph automorphism and therefore preserves adjacencies. We have by construction

$$\max_{v^n\in\psi(\mathcal{C}_n)} \|T_{v^n} - P_{\psi(V)}\|_\infty \underset{n\to\infty}{\to} 0. \tag{B.125}$$

Furthermore, since ψ is a bijection we have $|\psi(\mathcal{C}_n)| = |\mathcal{C}_n|$ for all $n \in \mathbb{N}^\star$, thus

$$\frac{\log|\psi(\mathcal{C}_n)|}{n} = \frac{\log|\mathcal{C}_n|}{n} \underset{n\to\infty}{\to} C_0(G). \tag{B.126}$$

Hence

$$P_{\psi(V)} \in \mathcal{P}^*(G). \tag{B.127}$$

Now, for all $v, v' \in \mathcal{V}$, denote by $\mathcal{S}_{v'\to v} \subseteq \mathrm{Aut}(G)$ the set of automorphisms that map v' to v; note that this set is nonempty thanks to the vertex-transitivity of G. We have for

all $v \in \mathcal{V}$

$$\mathrm{Aut}(G) = \bigsqcup_{v' \in \mathcal{V}} \mathcal{S}_{v' \to v}. \tag{B.128}$$

Furthermore, for all $v \in \mathcal{V}$, all the sets $(\mathcal{S}_{v' \to v})_{v' \to v}$ have the same cardinality: for all $v', v'' \in \mathcal{V}$,

$$\mathcal{S}_{v'' \to v} \circ \psi_1 \subseteq \mathcal{S}_{v' \to v}, \tag{B.129}$$

where $\psi_1 \in \mathcal{S}_{v' \to v''}$. It follows that for all $v, v' \in \mathcal{V}$,

$$|\mathcal{S}_{v' \to v}| = \frac{|\mathrm{Aut}(G)|}{|\mathcal{V}|}. \tag{B.130}$$

Therefore, for all $v \in \mathcal{V}$ we have

$$\frac{1}{|\mathrm{Aut}(G)|} \sum_{\psi \in \mathrm{Aut}(G)} P_{\psi(V)} \qquad \in \mathcal{P}^*(G) \tag{B.131}$$

$$= \left(\frac{1}{|\mathrm{Aut}(G)|} \sum_{\psi \in \mathrm{Aut}(G)} P_V(\psi^{-1}(v)) \right)_{v \in \mathcal{V}} \tag{B.132}$$

$$= \left(\frac{1}{|\mathrm{Aut}(G)|} \sum_{v' \in \mathcal{V}} |\mathcal{S}_{v' \to v}| P_V(v') \right)_{v \in \mathcal{V}} \tag{B.133}$$

$$= \left(\frac{1}{|\mathrm{Aut}(G)|} \sum_{v' \in \mathcal{V}} \frac{|\mathrm{Aut}(G)|}{|\mathcal{V}|} P_V(v') \right)_{v \in \mathcal{V}} \tag{B.134}$$

$$= \mathrm{Unif}(\mathcal{V}); \tag{B.135}$$

where (B.131) comes from the convexity of $\mathcal{P}^*(G)$ given by Proposition 5.3.2 and (B.127); (B.133) comes from (B.128); and (B.134) comes from (B.130).

B.4 Main proofs for the link between linearizations of $C(\cdot, P_V)$ and C_0

B.4.1 Proof of Theorem 5.4.1

We prove Theorem 5.4.1 in two steps, which are Lemma B.4.1 and Lemma B.4.2. The proofs are respectively given in Appendix B.4.2 and B.4.3.

Lemma B.4.1

$$C_0\left(\bigwedge_{a\in\mathcal{A}} G_a\right) = \sum_{a\in\mathcal{A}} C_0(G_a) \tag{B.136}$$

$$\implies \forall (P^*_{V_a})_{a\in\mathcal{A}} \in \prod_{a\in\mathcal{A}} \mathcal{P}^*(G_a), \begin{cases} \bigotimes_{a\in\mathcal{A}} P^*_{V_a} \in \mathcal{P}^*\left(\bigwedge_{a\in\mathcal{A}} G_a\right) \text{ and} \\ C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P^*_{V_a}\right) = \sum_{a\in\mathcal{A}} C(G_a, P^*_{V_a}). \end{cases} \tag{B.137}$$

Lemma B.4.2 *For all $P_{V_1,\dots,V_{|\mathcal{A}|}} \in \mathcal{P}^*\left(\bigwedge_{a\in\mathcal{A}} G_a\right)$, the following holds*

$$C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P_{V_a}\right) = \sum_{a\in\mathcal{A}} C(G_a, P_{V_a}) \tag{B.138}$$

$$\implies C_0\left(\bigwedge_{a\in\mathcal{A}} G_a\right) = \sum_{a\in\mathcal{A}} C_0(G_a) \text{ and } \forall a\in\mathcal{A},\ P_{V_a} \in \mathcal{P}^*(G_a). \tag{B.139}$$

Let us prove Theorem 5.4.1. Let $(P^*_{V_a})_{a\in\mathcal{A}} \in \prod_{a\in\mathcal{A}} \mathcal{P}^*(G_a)$, we have by Lemma B.4.1

$$C_0\left(\bigwedge_{a\in\mathcal{A}} G_a\right) = \sum_{a\in\mathcal{A}} C_0(G_a) \tag{B.140}$$

$$\implies \bigotimes_{a\in\mathcal{A}} P^*_{V_a} \in \mathcal{P}^*\left(\bigwedge_{a\in\mathcal{A}} G_a\right) \text{ and } C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P^*_{V_a}\right) = \sum_{a\in\mathcal{A}} C(G_a, P^*_{V_a}) \tag{B.141}$$

$$\implies \exists P_{V_1,\dots,V_{|\mathcal{A}|}} \in \mathcal{P}^*\left(\bigwedge_{a\in\mathcal{A}} G_a\right),\ C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P_{V_a}\right) = \sum_{a\in\mathcal{A}} C(G_a, P_{V_a}). \tag{B.142}$$

Conversely, by Lemma B.4.2 we have

$$\exists P_{V_1,\dots,V_{|\mathcal{A}|}} \in \mathcal{P}^*\left(\bigwedge_{a\in\mathcal{A}} G_a\right),\ C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P_{V_a}\right) = \sum_{a\in\mathcal{A}} C(G_a, P_{V_a}) \tag{B.143}$$

$$\implies C_0\left(\bigwedge_{a\in\mathcal{A}} G_a\right) = \sum_{a\in\mathcal{A}} C_0(G_a), \tag{B.144}$$

and any $P_{V_1,\dots,V_{|\mathcal{A}|}}$ that satisfies (B.143) also satisfies $(P_{V_a})_{a\in\mathcal{A}} \in \prod_{a\in\mathcal{A}} \mathcal{P}^*(G_a)$.

B.4.2 Proof of Lemma B.4.1

For all family of graphs $(G_a)_{a\in\mathcal{A}}$, and distributions $(P^*_{V_a})_{a\in\mathcal{A}} \in \prod_{a\in\mathcal{A}} \mathcal{P}^*(G_a)$ the following holds:

$$C_0\left(\bigwedge_{a\in\mathcal{A}} G_a\right) = \max_{(P_{V_a})_{a\in\mathcal{A}} \in \prod_{a\in\mathcal{A}} \Delta(\mathcal{V}_a)} C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P_{V_a}\right) \tag{B.145}$$

$$\geq C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P^*_{V_a}\right) \tag{B.146}$$
$$\geq \sum_{a\in\mathcal{A}} C(G_a, P^*_{V_a}) \tag{B.147}$$
$$= \sum_{a\in\mathcal{A}} C_0(G_a); \tag{B.148}$$

where (B.145) and (B.146) follow from Corollary 5.3.4; (B.147) comes from Proposition 5.2.5; and (B.148) follows from Proposition 5.3.2.

Now assume that $\sum_{a\in\mathcal{A}} C_0(G_a) = C_0\left(\bigwedge_{a\in\mathcal{A}} G_a\right)$, then equality holds in (B.145) to (B.148). In particular, we have

$$C_0\left(\bigwedge_{a\in\mathcal{A}} G_a\right) = C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P^*_{V_a}\right),\ \text{hence}\ \bigotimes_{a\in\mathcal{A}} P^*_{V_a} \in \mathcal{P}^*\left(\bigwedge_{a\in\mathcal{A}} G_a\right); \tag{B.149}$$
$$C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P^*_{V_a}\right) = \sum_{a\in\mathcal{A}} C(G_a, P^*_{V_a}). \tag{B.150}$$

B.4.3 Proof of Lemma B.4.2

Let $P_{V_1,\dots,V_{|\mathcal{A}|}} \in \mathcal{P}^*(\bigwedge_{a\in\mathcal{A}} G_a)$, by Theorem 5.3.3, we have

$$\bigotimes_{a\in\mathcal{A}} P_{V_a} \in \mathcal{P}^*\left(\bigwedge_{a\in\mathcal{A}} G_a\right). \tag{B.151}$$

Additionally, for all $(P^*_{V_a})_{a\in\mathcal{A}} \in \prod_{a\in\mathcal{A}} \mathcal{P}^*(G_a)$ we have for all $a \in \mathcal{A}$

$$C(G_a, P_{V_a}) \leq C(G_a, P^*_{V_a}), \tag{B.152}$$

and the following holds

$$C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P_{V_a}\right) \tag{B.153}$$
$$\geq C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P^*_{V_a}\right) \tag{B.154}$$
$$\geq \sum_{a\in\mathcal{A}} C(G_a, P^*_{V_a}) \tag{B.155}$$
$$\geq \sum_{a\in\mathcal{A}} C(G_a, P_{V_a}); \tag{B.156}$$

where (B.154) comes from (B.151); (B.155) comes from Proposition 5.2.5; and (B.156) comes from (B.152).

Now assume that

$$C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P_{V_a}\right) = \sum_{a\in\mathcal{A}} C(G_a, P_{V_a}), \tag{B.157}$$

then equality holds in (B.153) to (B.156). In particular, we have for all $a \in \mathcal{A}$:

$$C(G_a, P_{V_a}) = C(G_a, P^*_{V_a}) \tag{B.158}$$

as a consequence of $\sum_{a\in\mathcal{A}} C(G_a, P^*_{V_a}) = \sum_{a\in\mathcal{A}} C(G_a, P_{V_a})$ and (B.152). Hence P_{V_a} also maximizes $C(G_a, \cdot)$ for all $a \in \mathcal{A}$:

$$\forall a \in \mathcal{A},\ P_{V_a} \in \mathcal{P}^*(G_a). \tag{B.159}$$

Furthermore,

$$\begin{aligned} C_0\left(\bigwedge_{a\in\mathcal{A}} G_a\right) &= C\left(\bigwedge_{a\in\mathcal{A}} G_a,\ \bigotimes_{a\in\mathcal{A}} P_{V_a}\right) && \text{(B.160)}\\ &= \sum_{a\in\mathcal{A}} C(G_a, P_{V_a}) && \text{(B.161)}\\ &= \sum_{a\in\mathcal{A}} C_0(G_a); && \text{(B.162)} \end{aligned}$$

where (B.160) comes from Corollary 5.3.4; (B.161) follows from (B.157); and (B.162) comes from (B.159) and Proposition 5.3.2.

B.4.4 Proof of Theorem 5.4.2

The techniques used in this proof are the same as in the proof of Theorem 5.4.1. We prove Theorem 5.4.2 in two steps, which are Lemma B.4.3 and Lemma B.4.4; their proofs are respectively given in Appendix B.4.5 and B.4.6.

Lemma B.4.3 *Let*

$$P^*_A \doteq \left(\frac{2^{C_0(G_a)}}{\sum_{a'\in\mathcal{A}} 2^{C_0(G_{a'})}}\right)_{a\in\mathcal{A}}, \tag{B.163}$$

we have

$$\begin{aligned} & C_0\left(\bigsqcup_{a\in\mathcal{A}} G_a\right) = \log\left(\sum_{a\in\mathcal{A}} 2^{C_0(G_a)}\right)\\ \Longrightarrow\ & \forall (P^*_{V_a})_{a\in\mathcal{A}} \in \prod_{a\in\mathcal{A}} \mathcal{P}^*(G_a),\ \sum_{a\in\mathcal{A}} P^*_A(a) P^*_{V_a} \in \mathcal{P}^*\left(\bigsqcup_{a\in\mathcal{A}} G_a\right) \ \textit{and}\\ & C\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a,\ \sum_{a\in\mathcal{A}} P^*_A(a) P^*_{V_a}\right) = H(P^*_A) + \sum_{a\in\mathcal{A}} P^*_A(a) C(G_a, P^*_{V_a}), \end{aligned} \tag{B.164}$$

Lemma B.4.4 *Let*

$$P_A^* \doteq \left(\frac{2^{C_0(G_a)}}{\sum_{a' \in \mathcal{A}} 2^{C_0(G_{a'})}} \right)_{a \in \mathcal{A}}, \tag{B.165}$$

for all $\sum_{a \in \mathcal{A}} P_A(a) P_{V_a} \in \mathcal{P}^* \left(\bigsqcup_{a \in \mathcal{A}} G_a \right)$ *the following holds*

$$\begin{aligned} & C \left(\bigsqcup_{a \in \mathcal{A}}^{P_A} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{V_a} \right) = H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C(G_a, P_{V_a}) \\ \implies & C_0 \left(\bigsqcup_{a \in \mathcal{A}} G_a \right) = \log \left(\sum_{a \in \mathcal{A}} 2^{C_0(G_a)} \right), \ (P_{V_a})_{a \in \mathcal{A}} \in \prod_{a \in \mathcal{A}} \mathcal{P}^*(G_a), \ \textit{and} \ P_A = P_A^*. \end{aligned} \tag{B.166}$$

Now let us prove Theorem 5.4.2. Let $(P_{V_a}^*)_{a \in \mathcal{A}} \in \prod_{a \in \mathcal{A}} \mathcal{P}^*(G_a)$, we have by Lemma B.4.3

$$C_0 \left(\bigsqcup_{a \in \mathcal{A}} G_a \right) = \log \left(\sum_{a \in \mathcal{A}} 2^{C_0(G_a)} \right) \tag{B.167}$$

$$\implies \sum_{a \in \mathcal{A}} P_A^*(a) P_{V_a}^* \in \mathcal{P}^* \left(\bigsqcup_{a \in \mathcal{A}} G_a \right) \text{ and}$$

$$C \left(\bigsqcup_{a \in \mathcal{A}}^{P_A} G_a, \sum_{a \in \mathcal{A}} P_A^*(a) P_{V_a}^* \right) = H(P_A^*) + \sum_{a \in \mathcal{A}} P_A^*(a) C(G_a, P_{V_a}^*), \tag{B.168}$$

$$\implies \exists P_V \in \mathcal{P}^* \left(\bigsqcup_{a \in \mathcal{A}} G_a \right),$$

$$C \left(\bigsqcup_{a \in \mathcal{A}}^{P_A} G_a, \sum_{a \in \mathcal{A}} P_A^*(a) P_{V_a}^* \right) = H(P_A^*) + \sum_{a \in \mathcal{A}} P_A^*(a) C(G_a, P_{V_a}^*),$$

where $P_{V_a}^* = P_{V | V \in \mathcal{V}_a}$ and $P_A^*(a) = P_V(\mathcal{V}_a)$ for all $a \in \mathcal{A}$.

Conversely, by Lemma B.4.4 we have

$$\exists P_V \in \mathcal{P}^* \left(\bigsqcup_{a \in \mathcal{A}} G_a \right), \tag{B.169}$$

$$C \left(\bigsqcup_{a \in \mathcal{A}}^{P_A} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{V_a} \right) = H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C(G_a, P_{V_a})$$

$$\implies C_0 \left(\bigsqcup_{a \in \mathcal{A}} G_a \right) = \log \left(\sum_{a \in \mathcal{A}} 2^{C_0(G_a)} \right), \tag{B.170}$$

and any $P_V = \sum_{a \in \mathcal{A}} P_A(a) P_{V_a}$ that satisfies (B.169) also satisfies

$$(P_{V_a})_{a \in \mathcal{A}} \in \prod_{a \in \mathcal{A}} \mathcal{P}^*(G_a), \text{ and } P_A = \left(\frac{2^{C_0(G_a)}}{\sum_{a' \in \mathcal{A}} 2^{C_0(G_{a'})}} \right)_{a \in \mathcal{A}}. \tag{B.171}$$

B.4.5 Proof of Lemma B.4.3

Lemma B.4.5 states that the function $(w_a)_{a\in\mathcal{A}} \mapsto \log\left(\sum_{a\in\mathcal{A}} 2^{w_a}\right)$ is the Legendre-Fenchel conjugate [69] of the entropy function $P_A \mapsto H(P_A)$. The proof of Lemma B.4.5 is given in Appendix B.6.8.

Lemma B.4.5 *Let $(w_a)_{a\in\mathcal{A}} \in \mathbb{R}^{|\mathcal{A}|}$, then the mapping*

$$\zeta : P_A \mapsto H(P_A) + \sum_{a\in\mathcal{A}} P_A(a) w_a \tag{B.172}$$

has a unique maximum

$$P_A^* = \left(\frac{2^{w_a}}{\sum_{a'\in\mathcal{A}} 2^{w_{a'}}}\right)_{a\in\mathcal{A}}; \tag{B.173}$$

and

$$\zeta(P_A^*) = \log\left(\sum_{a\in\mathcal{A}} 2^{w_a}\right). \tag{B.174}$$

Now let us prove Lemma B.4.3. Let

$$\textstyle\sum_{a\in\mathcal{A}} P_A(a) P_{V_a} \in \mathcal{P}^*\left(\bigsqcup_{a\in\mathcal{A}} G_a\right), \tag{B.175}$$

$$(P_{V_a}^*)_{a\in\mathcal{A}} \in \textstyle\prod_{a\in\mathcal{A}} \mathcal{P}^*(G_a), \tag{B.176}$$

$$P_A^* \doteq \left(\frac{2^{C_0(G_a)}}{\sum_{a'\in\mathcal{A}} 2^{C_0(G_{a'})}}\right)_{a\in\mathcal{A}}. \tag{B.177}$$

We have

$$\begin{aligned}
C_0\left(\textstyle\bigsqcup_{a\in\mathcal{A}} G_a\right) &= C\left(\textstyle\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a,\ \sum_{a\in\mathcal{A}} P_A(a) P_{V_a}\right) && \text{(B.178)}\\
&\geq C\left(\textstyle\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a,\ \sum_{a\in\mathcal{A}} P_A^*(a) P_{V_a}^*\right) && \text{(B.179)}\\
&\geq H(P_A^*) + \textstyle\sum_{a\in\mathcal{A}} P_A^*(a) C(G_a, P_{V_a}^*) && \text{(B.180)}\\
&= H(P_A^*) + \textstyle\sum_{a\in\mathcal{A}} P_A^*(a) C_0(G_a) && \text{(B.181)}\\
&= \log\left(\textstyle\sum_{a\in\mathcal{A}} 2^{C_0(G_a)}\right); && \text{(B.182)}
\end{aligned}$$

where (B.178) and (B.179) come from (B.175) and Proposition 5.3.2; (B.180) comes from

Proposition 5.2.5; (B.181) comes from (B.176) and Proposition 5.3.2; and (B.182) comes from (B.177) and Lemma B.4.5.

Assume that $C_0\left(\bigsqcup_{a\in\mathcal{A}} G_a\right) = \log\left(\sum_{a\in\mathcal{A}} 2^{C_0(G_a)}\right)$, then equality holds in (B.178) to (B.182), therefore the following holds:

$$\begin{aligned}
&C_0\left(\bigsqcup_{a\in\mathcal{A}} G_a\right) = \log\left(\sum_{a\in\mathcal{A}} 2^{C_0(G_a)}\right)\\
\implies &\forall (P^*_{V_a})_{a\in\mathcal{A}} \in \prod_{a\in\mathcal{A}} \mathcal{P}^*(G_a),\ \sum_{a\in\mathcal{A}} P^*_A(a) P^*_{V_a} \in \mathcal{P}^*\left(\bigsqcup_{a\in\mathcal{A}} G_a\right) \text{ and}\\
&C\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a,\ \sum_{a\in\mathcal{A}} P^*_A(a) P^*_{V_a}\right) = H(P^*_A) + \sum_{a\in\mathcal{A}} P^*_A(a) C(G_a, P^*_{V_a}).
\end{aligned} \tag{B.183}$$

B.4.6 Proof of Lemma B.4.4

Let

$$\sum_{a\in\mathcal{A}} P_A(a) P_{V_a} \in \mathcal{P}^*\left(\bigsqcup_{a\in\mathcal{A}} G_a\right), \tag{B.184}$$

$$(P^*_{V_a})_{a\in\mathcal{A}} \in \prod_{a\in\mathcal{A}} \mathcal{P}^*(G_a), \tag{B.185}$$

$$P^*_A \doteq \left(\frac{2^{C_0(G_a)}}{\sum_{a'\in\mathcal{A}} 2^{C_0(G_{a'})}}\right)_{a\in\mathcal{A}}. \tag{B.186}$$

We have

$$\begin{aligned}
C\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a,\ \sum_{a\in\mathcal{A}} P_A(a) P_{V_a}\right) &= C_0\left(\bigsqcup_{a\in\mathcal{A}} G_a\right) && \text{(B.187)}\\
&\geq \log\left(\sum_{a\in\mathcal{A}} 2^{C_0(G_a)}\right) && \text{(B.188)}\\
&= H(P^*_A) + \sum_{a\in\mathcal{A}} P^*_A(a) C_0(G_a) && \text{(B.189)}\\
&\geq H(P_A) + \sum_{a\in\mathcal{A}} P_A(a) C_0(G_a) && \text{(B.190)}\\
&= H(P_A) + \sum_{a\in\mathcal{A}} P_A(a) C(G_a, P^*_{V_a}) && \text{(B.191)}\\
&\geq H(P_A) + \sum_{a\in\mathcal{A}} P_A(a) C(G_a, P_{V_a}); && \text{(B.192)}
\end{aligned}$$

where (B.187) comes from (B.184) and Proposition 5.3.2; (B.188) comes from Theorem 5.1.17; (B.189) and (B.190) come from (B.186) and Lemma B.4.5, which can be found in Appendix B.4.5; (B.191) and (B.192) come from (B.185) and Proposition 5.3.2.

Assume that $C\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a,\ \sum_{a\in\mathcal{A}} P_A(a) P_{V_a}\right) = H(P_A) + \sum_{a\in\mathcal{A}} P_A(a) C(G_a, P_{V_a})$, then equality holds in (B.187) to (B.192). In particular $P_A = P^*_A$ as a consequence of the equality between (B.189) and (B.190); and $(P_{V_a})_{a\in\mathcal{A}} \in \prod_{a\in\mathcal{A}} \mathcal{P}^*(G_a)$ as a consequence of the equality between (B.191) and (B.192). Thus, for all $\sum_{a\in\mathcal{A}} P_A(a) P_{V_a} \in \mathcal{P}^*\left(\bigsqcup_{a\in\mathcal{A}} G_a\right)$

the following holds:

$$\begin{aligned} & C\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a,\ \sum_{a\in\mathcal{A}} P_A(a)P_{V_a}\right) = H(P_A) + \sum_{a\in\mathcal{A}} P_A(a)C(G_a, P_{V_a}) \\ \implies & C_0\left(\bigsqcup_{a\in\mathcal{A}} G_a\right) = \log\left(\sum_{a\in\mathcal{A}} 2^{C_0(G_a)}\right),\ (P_{V_a})_{a\in\mathcal{A}} \in \prod_{a\in\mathcal{A}} \mathcal{P}^*(G_a),\ \text{and } P_A = P_A^*. \end{aligned} \tag{B.193}$$

B.5 Main proofs for the examples and counterexamples

B.5.1 Proof of Theorem 5.5.8

Lemma B.5.1 comes from [63, Corollary 1], and states that the function $P_A \mapsto H_\kappa\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right)$, defined analogously to $P_A \mapsto \overline{H}\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right)$, is always linear. The proof of Lemma B.5.2 is given in Appendix B.6.9.

Lemma B.5.1 (from [61, Corollary 3.4]) *For all probabilistic graphs $(G_a)_{a\in\mathcal{A}}$ and $P_A \in \Delta(\mathcal{A})$, we have $H_\kappa\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right) = \sum_{a\in\mathcal{A}} P_A(a)H_\kappa(G_a)$.*

Lemma B.5.2 *The probabilistic graph $\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a$ is perfect if and only if G_a is perfect for all $a \in \mathcal{A}$.*

Now let us prove Theorem 5.5.8.

For all $a \in \mathcal{A}$, let $G_a = (\mathcal{V}_a, \mathcal{E}_a, P_{V_a})$ be a perfect probabilistic graph. By Lemma B.5.2, $\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a$ is also perfect; and we have $\overline{H}\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right) = H_\kappa\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right)$ by Theorem 5.5.4. We also have $H_\kappa\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right) = \sum_{a\in\mathcal{A}} P_A(a)H_\kappa(G_a) = \sum_{a\in\mathcal{A}} P_A(a)\overline{H}(G_a)$ by Lemma B.5.1 and Theorem 5.5.4 used on the perfect graphs $(G_a)_{a\in\mathcal{A}}$. Thus

$$\overline{H}\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right) = \sum_{a\in\mathcal{A}} P_A(a)\overline{H}(G_a). \tag{B.194}$$

By Theorem 5.2.4, it follows that $\overline{H}\left(\bigwedge_{a\in\mathcal{A}} G_a\right) = \sum_{a\in\mathcal{A}} \overline{H}(G_a) = \sum_{a\in\mathcal{A}} H_\kappa(G_a)$, where the last equality comes from Theorem 5.5.4.

B.6 Secondary proofs

B.6.1 Proof of Lemma B.2.2

In order to prove Lemma B.2.2 we need Lemma B.2.1, which can be found in Appendix B.2.1; and Lemma B.6.1, which is a generalization for infinite sequences of the following observation: if $T_{\overline{a}^n} = P_A \in \Delta_n(\mathcal{A})$ satisfies $P_A = \frac{i}{n}P'_A + \frac{n-i}{n}P''_A$ with $P'_A \in \Delta_i(\mathcal{A})$ and $P''_A \in \Delta_{n-i}(\mathcal{A})$, then $\overline{a}^n$ can be separated into two subsequences a'^i and a''^{n-i} such that $T_{a'^i} = P'_A$ and $T_{a''^{n-i}} = P''_A$. The proof is given in Appendix B.6.10.

Lemma B.6.1 (Type-splitting lemma) *Let $(\overline{a}_n)_{n\in\mathbb{N}^\star} \in \mathcal{A}^{\mathbb{N}^\star}$ be a sequence such that $T_{\overline{a}^n} \to P_A \in \Delta(\mathcal{A})$ when $n \to \infty$, let $\beta \in (0,1)$ and $P'_A, P''_A \in \Delta(\mathcal{A})$ such that*

$$P_A = \beta P'_A + (1-\beta)P''_A. \tag{B.195}$$

Then there exists a sequence $(b_n)_{n\in\mathbb{N}^\star} \in \{0,1\}^{\mathbb{N}^\star}$ such that the two extracted sequences $a' \doteq (\overline{a}_n)_{\substack{n\in\mathbb{N}^\star,\\ b_n=0}}$ and $a'' \doteq (\overline{a}_n)_{\substack{n\in\mathbb{N}^\star,\\ b_n=1}}$ satisfy

$$T_{b^n} \underset{n\to\infty}{\to} (\beta, 1-\beta), \tag{B.196}$$

$$T_{a'^n} \underset{n\to\infty}{\to} P'_A, \qquad T_{a''^n} \underset{n\to\infty}{\to} P''_A. \tag{B.197}$$

Now let us prove Lemma B.2.2. Let $\eta : P_A \mapsto \overline{H}\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right)$.

(η Lipschitz) Let us first prove that η is Lipschitz. For all $P_A, P'_A \in \Delta(\mathcal{A})$ we need to bound the quantity $|\eta(P_A) - \eta(P'_A)|$; by Lemma B.2.1 this is equivalent to bounding

$$\lim_{n\to\infty} \frac{1}{n} \left| H_\chi\left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge nT_{\overline{a}^n}(a)}\right) - H_\chi\left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge nT_{\overline{a}'^n}(a)}\right)\right| \tag{B.198}$$

where $(T_{\overline{a}^n}, T_{\overline{a}'^n}) \to (P_A, P'_A)$ when $n \to \infty$.

Fix $n \in \mathbb{N}^\star$, we assume that the quantity inside $|\cdot|$ in (B.198) is positive; the other case can be treated with the same arguments by symmetry of the roles. We have

$$H_\chi\left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge nT_{\overline{a}^n}(a)}\right) - H_\chi\left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge nT_{\overline{a}'^n}(a)}\right) \tag{B.199}$$

$$\leq H_\chi\left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge nT_{\overline{a}^n}(a)}\right) - H_\chi\left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge n\min(T_{\overline{a}^n}(a), T_{\overline{a}'^n}(a))}\right) \tag{B.200}$$

$$= H_\chi\left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge n\min(T_{\overline{a}^n}(a), T_{\overline{a}'^n}(a))} \bigwedge_{a\in\mathcal{A}} G_a^{\wedge n|T_{\overline{a}^n}(a) - T_{\overline{a}'^n}(a)|_+}\right)$$
$$- H_\chi\left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge n\min(T_{\overline{a}^n}(a), T_{\overline{a}'^n}(a))}\right) \tag{B.201}$$

$$\leq H_\chi\left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge n|T_{\overline{a}^n}(a)-T_{\overline{a}'^n}(a)|_+}\right) \tag{B.202}$$

$$\leq \log\left(\max_a |\mathcal{V}_a|\right)\sum_{a\in\mathcal{A}} n|T_{\overline{a}^n}(a)-T_{\overline{a}'^n}(a)|_+ \tag{B.203}$$

$$\leq n\log\left(\max_a |\mathcal{V}_a|\right)\|T_{\overline{a}^n}-T_{\overline{a}'^n}\|_1, \tag{B.204}$$

where $|\cdot|_+ = \max(\cdot,0)$ and $\|T_{\overline{a}^n}-T_{\overline{a}'^n}\|_1 = \sum_{a\in\mathcal{A}}|T_{\overline{a}^n}(a)-T_{\overline{a}'^n}(a)|$; (B.200) follows from the removal of terms in the second product, as $H_\chi(G\wedge G')\geq H_\chi(G)$ for all probabilistic graphs G, G'; (B.201) is an arrangement of the terms in the first product, as $\min(s,t)+\max(s-t,0)=s$ for all real numbers s,t; (B.202) comes from the subadditivity of H_χ; (B.203) follows from $H_\chi(G_a)\leq\log\max_{a'}|\mathcal{V}_{a'}|$ for all $a\in\mathcal{A}$; (B.204) results from $|T_{\overline{a}^n}(a)-T_{\overline{a}'^n}(a)|_+\leq|T_{\overline{a}^n}(a)-T_{\overline{a}'^n}(a)|$ for all $a\in\mathcal{A}$.

By normalization and limit, it follows that

$$|\eta(P_A)-\eta(P'_A)| \leq \lim_{n\to\infty}\log\left(\max_a|\mathcal{V}_a|\right)\cdot\|T_{\overline{a}^n}-T_{\overline{a}'^n}\|_1 \tag{B.205}$$

$$= \log\left(\max_a|\mathcal{V}_a|\right)\cdot\|P_A-P'_A\|_1. \tag{B.206}$$

Hence η is $(\log\max_a|\mathcal{V}_a|)$-Lipschitz.

(η convex) Let us now prove that η is convex. Let $P'_A, P''_A\in\Delta(\mathcal{A})$, and $\beta\in(0,1)$, we have by Lemma B.2.1

$$\eta\Big(\beta P'_A+(1-\beta)P''_A\Big) = \lim_{n\to\infty}\frac{1}{n}H_\chi\left(\bigwedge_{a\in\mathcal{A}}G_a^{\wedge nT_{\overline{a}^n}(a)}\right), \tag{B.207}$$

where $T_{\overline{a}^n}\to\beta P'_A+(1-\beta)P''_A$ when $n\to\infty$. By Lemma B.6.1, there exists $(b_n)_{n\in\mathbb{N}^\star}\in\{0,1\}^{\mathbb{N}^\star}$ such that the decomposition of $(\overline{a}_n)_{n\in\mathbb{N}^\star}$ into two subsequences $a'\doteq(\overline{a}_n)_{\substack{n\in\mathbb{N}^\star,\\ b_n=0}}$ and $a''\doteq(\overline{a}_n)_{\substack{n\in\mathbb{N}^\star,\\ b_n=1}}$ satisfies

$$T_{b^n}\underset{n\to\infty}{\to}(\beta,1-\beta), \tag{B.208}$$

$$T_{a'^n}\underset{n\to\infty}{\to}P'_A, \qquad T_{a''^n}\underset{n\to\infty}{\to}P''_A. \tag{B.209}$$

For all $n\in\mathbb{N}^\star$, let $\Xi(n)\doteq nT_{b^n}(0)$, we have

$$\eta\Big(\beta P'_A+(1-\beta)P''_A\Big) \tag{B.210}$$

$$= \lim_{n\to\infty} \frac{1}{n} H_\chi \left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge \Xi(n) T_{a'\Xi(n)}(a) + (n-\Xi(n)) T_{a''n-\Xi(n)}(a)} \right) \tag{B.211}$$

$$\leq \lim_{n\to\infty} \frac{\Xi(n)}{n} \frac{1}{\Xi(n)} H_\chi \left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge \Xi(n) T_{a'\Xi(n)}(a)} \right) \tag{B.212}$$

$$+ \frac{n-\Xi(n)}{n} \frac{1}{n-\Xi(n)} H_\chi \left(\bigwedge_{a\in\mathcal{A}} G_a^{\wedge (n-\Xi(n)) T_{a''n-\Xi(n)}(a)} \right) \tag{B.213}$$

$$= \beta\eta(P'_A) + (1-\beta)\eta(P''_A); \tag{B.214}$$

where (B.211) comes from (B.207); (B.213) follows from the subadditivity of H_χ; (B.214) comes from (B.208), (B.209) and Lemma B.2.1. Since (B.214) holds for all $P'_A, P''_A \in \Delta(\mathcal{A})$ and $\beta \in (0,1)$, we have that η is convex.

B.6.2 Proof of Lemma B.2.3

It can be easily observed that

$$\exists P_A \in \operatorname{int}(\Delta(\mathcal{A})),\ \gamma(P_A) = \textstyle\sum_{a\in\mathcal{A}} P_A(a)\gamma(\mathbb{1}_a) \tag{B.215}$$

$$\Longleftarrow\ \forall P_A \in \Delta(\mathcal{A}),\ \gamma(P_A) = \textstyle\sum_{a\in\mathcal{A}} P_A(a)\gamma(\mathbb{1}_a). \tag{B.216}$$

Now let us prove (B.215) $\Rightarrow$ (B.216). Let $P^*_A \in \operatorname{int}\Delta(\mathcal{A})$ such that $\gamma(P^*_A) = \sum_{a\in\mathcal{A}} P^*_A(a)\gamma(\mathbb{1}_a)$. Let $m : \Delta(\mathcal{A}) \to \mathbb{R}$ linear such that $m(P^*_A) = \gamma(P^*_A)$ and $\forall P_A \in \Delta(\mathcal{A})$, $m(P_A) \leq \gamma(P_A)$. We have

$$0 = \gamma(P^*_A) - m(P^*_A) = \textstyle\sum_{a\in\mathcal{A}} P^*_A(a)\big(\gamma(\mathbb{1}_a) - m(\mathbb{1}_a)\big); \tag{B.217}$$

and therefore $\gamma(\mathbb{1}_a) = m(\mathbb{1}_a)$ for all $a \in \mathcal{A}$, as $\gamma - m \geq 0$ and $P^*_A(a) > 0$ for all $a \in \mathcal{A}$. For all $P_A \in \Delta(\mathcal{A})$, we have

$$f(P_A) \leq \textstyle\sum_{a\in\mathcal{A}} P_A(a)\gamma(\mathbb{1}_a) \tag{B.218}$$

$$= \textstyle\sum_{a\in\mathcal{A}} P_A(a)m(\mathbb{1}_a) = m(P_A), \tag{B.219}$$

hence $\gamma = m$ and γ is linear.

B.6.3 Proof of Lemma B.2.4

The probabilistic graphs in both sides of (B.8) have

$$(\bigsqcup_{a\in\mathcal{A}} \mathcal{V}_a) \times (\bigsqcup_{b\in\mathcal{B}} \mathcal{V}_b) = \bigsqcup_{(a,b)\in\mathcal{A}\times\mathcal{B}} \mathcal{V}_a \times \mathcal{V}_b \tag{B.220}$$

as set of vertices, with underlying distribution

$$\begin{aligned}&(\textstyle\sum_{a\in\mathcal{A}} P_A(a)P_{V_a})\,(\sum_{b\in\mathcal{B}} P_B(b)P_{V_b})\\ &= \textstyle\sum_{(a,b)\in\mathcal{A}\times\mathcal{B}} P_A(a)P_B(b)P_{V_a}P_{V_b}.\end{aligned} \tag{B.221}$$

Now let us show that these two graphs have the same edges. Let $(v_\mathcal{A}, v_\mathcal{B}), (v'_\mathcal{A}, v'_\mathcal{B}) \in (\bigsqcup_{a\in\mathcal{A}} \mathcal{V}_a) \times (\bigsqcup_{b\in\mathcal{B}} \mathcal{V}_b)$; let $a_*, a'_* \in \mathcal{A}$ and $b_*, b'_* \in \mathcal{B}$ be the unique indexes such that

$$(v_\mathcal{A}, v_\mathcal{B}) \in \mathcal{V}_{a_*} \times \mathcal{V}_{b_*} \quad \text{and} \quad (v'_\mathcal{A}, v'_\mathcal{B}) \in \mathcal{V}_{a'_*} \times \mathcal{V}_{b'_*}. \tag{B.222}$$

We have:

$$(v_\mathcal{A}, v_\mathcal{B}), (v'_\mathcal{A}, v'_\mathcal{B}) \text{ are adjacent in } \left(\bigsqcup\nolimits_{a\in\mathcal{A}}^{P_A} G_a\right) \wedge \left(\bigsqcup\nolimits_{b\in\mathcal{B}}^{P_B} G_b\right) \tag{B.223}$$

$$\iff v_\mathcal{A}, v'_\mathcal{A} \text{ adjacent in } \bigsqcup\nolimits_{a\in\mathcal{A}}^{P_A} G_a \text{ and } v_\mathcal{B}, v'_\mathcal{B} \text{ adjacent in } \bigsqcup\nolimits_{b\in\mathcal{B}}^{P_B} G_b \tag{B.224}$$

$$\iff a_* = a'_* \text{ and } v_\mathcal{A}v'_\mathcal{A} \in \mathcal{E}_{a_*} \text{ and } b_* = b'_* \text{ and } v_\mathcal{B}v'_\mathcal{B} \in \mathcal{E}_{b_*} \tag{B.225}$$

$$\iff (a_*, b_*) = (a'_*, b'_*) \text{ and } (v_\mathcal{A}, v_\mathcal{B}), (v'_\mathcal{A}, v'_\mathcal{B}) \text{ are adjacent in } G_{a_*} \wedge G_{b_*} \tag{B.226}$$

$$\iff (v_\mathcal{A}, v_\mathcal{B}), (v'_\mathcal{A}, v'_\mathcal{B}) \text{ are adjacent in } \bigsqcup\nolimits_{(a,b)\in\mathcal{A}\times\mathcal{B}}^{P_AP_B} G_a \wedge G_b. \tag{B.227}$$

B.6.4 Proof of Lemma B.2.5

In order to prove Lemma B.2.5, we need Lemma B.6.2. In Lemma B.6.2 we give upper and lower bounds on the chromatic entropy of an induced subgraph $G[\mathcal{S}]$, using the chromatic entropy of the whole graph G and the probability $P_V(\mathcal{S})$. The core idea is that if $P_V(\mathcal{S})$ is close to 1 and $H_\chi(G)$ is big, then $H_\chi(G[\mathcal{S}])$ is close to $H_\chi(G)$. The proof of Lemma B.6.2 is given in Appendix B.6.11

Lemma B.6.2 *Let $G = (\mathcal{V}, \mathcal{E}, P_V)$ and $\mathcal{S} \subseteq \mathcal{V}$, then*

$$H_\chi(G) - 1 - (1 - P_V(\mathcal{S})) \log |\mathcal{V}| \leq H_\chi(G[\mathcal{S}]) \leq \frac{H_\chi(G)}{P_V(\mathcal{S})}. \tag{B.228}$$

Remark B.6.3 *$H_\chi(G[\mathcal{S}])$ can be greater than $H_\chi(G)$, even if $G[\mathcal{S}]$ has less vertices and inherits the structure of G. This stems from the normalized distribution $P_V/P_V(\mathcal{S})$ on the vertices of $G[\mathcal{S}]$ which gives more weight to the vertices in $\mathcal{S}$. For example, consider*

$$G = \Big(N_5, \mathrm{Unif}(\{1, ..., 5\})\Big) \overset{(1-\epsilon,\epsilon)}{\sqcup} \Big(K_5, \mathrm{Unif}(\{1, ..., 5\})\Big);$$

where K_n (resp. N_n) is the complete (resp. empty) graph with n vertices, i.e. there is an edge (resp. no edge) between any pair of distinct vertices; and with $\mathcal{S}$ being the vertices in the connected component K_5 in G. Then $H_\chi(G) = \epsilon \log 5$ and $H_\chi(G[\mathcal{S}]) = \log 5$.

Now let us prove Lemma B.2.5. By Lemma B.6.2, we have for all $n \in \mathbb{N}^\star$:

$$\begin{aligned} &H_\chi(G^{\wedge n}) - 1 - (1 - P_V^n(\mathcal{S}_n)) \log |\mathcal{V}| \\ &\leq H_\chi(G^{\wedge n}[\mathcal{S}^n]) \leq \frac{H_\chi(G^{\wedge n})}{P_V^n(\mathcal{S}_n)}. \end{aligned} \tag{B.229}$$

Since $P_V^n(\mathcal{S}_n) \to 1$, and $H_\chi(G^{\wedge n}) = n\overline{H}(G) + o(n)$ when $n \to \infty$, the desired results follows immediately by normalization and limit.

B.6.5 Proof of Lemma B.2.7

Let $(\tilde{G}_i)_{i \leq N}$ be isomorphic probabilistic graphs and G such that $G = \bigsqcup_i \tilde{G}_i$. Let $c_1^* : \mathcal{V}_1 \to \mathcal{C}$ be the coloring of $\tilde{G}_1$ with minimal entropy, and let c^* be the coloring of G defined by

$$c^* : \mathcal{V} \to \mathcal{C} \tag{B.230}$$

$$v \mapsto c_1^* \circ \psi_{i_v \to 1}(v), \tag{B.231}$$

where i_v is the unique integer such that $v \in \mathcal{V}_{i_v}$, and $\psi_{i_v \to 1} : \mathcal{V}_{i_v} \to \mathcal{V}_1$ is an isomorphism between $\tilde{G}_{i_v}$ and $\tilde{G}_1$. In other words c^* applies the same coloring pattern c_1^* on each connected component of G. We have

$$H_\chi(G) \leq H(c^*(V)) \tag{B.232}$$

$$= h\Big(\textstyle\sum_{j \leq N} P_{i_V}(j) P_{c^*(V_j)}\Big) \tag{B.233}$$

$$= h\Big(\textstyle\sum_{j \leq N} P_{i_V}(j) P_{c_1^*(V_1)}\Big) \tag{B.234}$$

$$= H(c_1^*(V_1)) \tag{B.235}$$
$$= H_\chi(\tilde{G}_1), \tag{B.236}$$

where h denotes the entropy of a distribution; (B.234) comes from the definition of c^*; and (B.236) comes from the definition of c_1^*.

Now let us prove the upper bound on $H_\chi(\tilde{G}_1)$. Let c be a coloring of G, and let $i^* \doteq \operatorname{argmin}_i H(c(V_i))$ (i.e. i^* is the index of the connected component for which the entropy of the coloring induced by c is minimal). We have

$$H(c(V)) = h\Big(\textstyle\sum_{j\leq N} P_{i_V}(j) P_{c(V_j)}\Big) \tag{B.237}$$
$$\geq \textstyle\sum_{j\leq N} P_{i_V}(j) h(P_{c(V_j)}) \tag{B.238}$$
$$\geq \textstyle\sum_{j\leq N} P_{i_V}(j) H(c(V_{i^*})) \tag{B.239}$$
$$\geq H_\chi(\tilde{G}_{i^*}), \tag{B.240}$$
$$= H_\chi(\tilde{G}_1), \tag{B.241}$$

where (B.238) follows from the concavity of h; (B.239) follows from the definition of i^*; (B.240) comes from the fact that c induces a coloring of $\tilde{G}_{i^*}$; (B.241) comes from the fact that $\tilde{G}_1$ and $\tilde{G}_{i^*}$ are isomorphic. Now, we can combine the bounds (B.236) and (B.241): for all coloring c of G we have

$$H_\chi(G) \leq H_\chi(\tilde{G}_1) \leq H(c(V)), \tag{B.242}$$

which yields the desired equality when taking the infimum over c.

B.6.6 Proof of Lemma B.2.8

The probabilistic graph $\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a$ has $\sum_{a\in\mathcal{A}} P_A(a) P_{V_a}$ as underlying distribution. Let A, V be two random variables such that A is drawn with P_A, and V is drawn with $P_{V|A}(\cdot|a) \doteq P_{V_a}$, so that

$$P_V = \textstyle\sum_{a\in\mathcal{A}} P_A(a) P_{V_a}. \tag{B.243}$$

We have

$$\overline{H}\left(\textstyle\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right) + C\left(\textstyle\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a, \sum_{a\in\mathcal{A}} P_A(a) P_{V_a}\right) \tag{B.244}$$

$$= H(V) \tag{B.245}$$
$$= H(A, V) \tag{B.246}$$
$$= H(A) + H(V|A) \tag{B.247}$$
$$= H(P_A) + \textstyle\sum_{a\in\mathcal{A}} P_A(a) H(P_{V_a}); \tag{B.248}$$

where (B.245) comes from Theorem 5.1.5 and (B.243); and (B.246) comes from the fact that A can be written as a function of V: by definition, the vertex set of $\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a$ writes $\mathcal{V} = \bigsqcup_{a\in\mathcal{A}} \mathcal{V}_a$ and $\operatorname{supp} P_{V_a} \subseteq \mathcal{V}_a$, therefore A is the unique index such that $V \in \mathcal{V}_A$.

B.6.7 Proof of Lemma B.3.1

In this proof, we need Lemma B.6.4 and Lemma B.6.5; their proofs are respectively given in Appendix and Appendix B.6.12 and Appendix B.6.13.

Lemma B.6.4 *For all distributions $P_V, P'_V \in \Delta(\mathcal{V})$, for all $\epsilon \geq 0$, and $m, n \in \mathbb{N}^\star$,*

$$\mathcal{T}_\epsilon^m(P_V) \times \mathcal{T}_\epsilon^n(P'_V) \subseteq \mathcal{T}_\epsilon^{m+n}\left(\frac{m}{m+n}P_V + \frac{n}{m+n}P'_V\right). \tag{B.249}$$

Lemma B.6.5 *For all graphs $G = (\mathcal{V}, \mathcal{E})$, $G' = (\mathcal{V}', \mathcal{E}')$ and sets $\mathcal{S} \subseteq \mathcal{V}$, $\mathcal{S} \subseteq \mathcal{V}'$ we have*

$$(G \wedge G')[\mathcal{S} \times \mathcal{S}'] \simeq G[\mathcal{S}] \wedge G'[\mathcal{S}'], \tag{B.250}$$

where $\simeq$ denotes isomorphic graphs, defined in Definition B.2.6.

Let us prove that for all $\epsilon > 0$ and $P_V \in \Delta(\mathcal{V})$, the sequence

$$\left(\log \alpha\Big(G^{\wedge n}[\mathcal{T}_\epsilon^n(P_V)]\Big)\right)_{n\in\mathbb{N}^\star} \tag{B.251}$$

is superadditive. For all integers m, n we have

$$\log \alpha\Big(G^{\wedge n+m}[\mathcal{T}_\epsilon^{n+m}(P_V)]\Big) \geq \log \alpha\Big(G^{\wedge m+n}[\mathcal{T}_\epsilon^m(P_V) \times \mathcal{T}_\epsilon^n(P_V)]\Big) \tag{B.252}$$
$$= \log \alpha\Big(G^{\wedge m}[\mathcal{T}_\epsilon^m(P_V)] \wedge G^{\wedge n}[\mathcal{T}_\epsilon^n(P_V)]\Big) \tag{B.253}$$
$$\geq \log \alpha\Big(G^{\wedge m}[\mathcal{T}_\epsilon^m(P_V)]\Big) + \log \alpha\Big(G^{\wedge n}[\mathcal{T}_\epsilon^n(P_V)]\Big), \tag{B.254}$$

where (B.252) and (B.253) respectively come from Lemma B.6.4 and Lemma B.6.5.

We can use Fekete's lemma [25]: for all $\epsilon > 0$, the following limit exists

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha(G^{\wedge n}[\mathcal{T}_\epsilon^n(P_V)]) = \limsup_{n\to\infty} \frac{1}{n} \log \alpha(G^{\wedge n}[\mathcal{T}_\epsilon^n(P_V)]) \tag{B.255}$$

$$= \sup_{n\in\mathbb{N}^\star} \frac{1}{n} \log \alpha(G^{\wedge n}[\mathcal{T}_\epsilon^n(P_V)]). \tag{B.256}$$

B.6.8 Proof of Lemma B.4.5

Let us maximize

$$\zeta : P_A \mapsto H(P_A) + \sum_{a\in\mathcal{A}} P_A(a) w_a. \tag{B.257}$$

It can be easily observed that ζ is strictly concave, hence the existence and uniqueness of the maximum. We have

$$\nabla\zeta(P_A) = \left(-\log P_A(a) - \frac{1}{\ln 2} + w_a \right)_{a\in\mathcal{A}}, \tag{B.258}$$

hence

$$\nabla\zeta(P_A) \perp \Delta(\mathcal{A}) \iff \exists C \in \mathbb{R},\ \nabla\zeta(P_A) = (C, ..., C) \tag{B.259}$$

$$\iff \exists C' \in \mathbb{R},\ (-\log P_A(a) + w_a)_{a\in\mathcal{A}} = (C', ..., C') \tag{B.260}$$

$$\iff \exists C' \in \mathbb{R},\ P_A = 2^{-C'} \left(2^{w_a}\right)_{a\in\mathcal{A}} \tag{B.261}$$

The value of C' can be deduced from the fact that P_A is a probability distribution: $2^{C'}$ is the normalization constant $\sum_{a'\in\mathcal{A}} 2^{w_{a'}}$. Hence the maximum of ζ writes

$$P_A^* = \left(\frac{2^{w_a}}{\sum_{a'\in\mathcal{A}} 2^{w_{a'}}} \right)_{a\in\mathcal{A}}; \tag{B.262}$$

and we have

$$\zeta(P_A^*) = \sum_{a\in\mathcal{A}} P_A^*(a) \left(\log \left(\frac{\sum_{a'\in\mathcal{A}} 2^{w_{a'}}}{2^{w_a}} \right) + w_a \right) \tag{B.263}$$

$$= \log \left(\sum_{a'\in\mathcal{Z}} 2^{w_{a'}} \right). \tag{B.264}$$

B.6.9 Proof of Lemma B.5.2

($\Longrightarrow$) Let $G = \bigsqcup_{a\in\mathcal{A}}^{P_A} G_a$ be a perfect probabilistic graph. Let $a' \in \mathcal{A}$ and $\mathcal{S}_{a'} \subset \mathcal{V}_{a'}$. We have $\chi\Big(\Big(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\Big)[\mathcal{S}_{a'}]\Big) = \omega\Big(\Big(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\Big)[\mathcal{S}_{a'}]\Big)$ since G is perfect, and therefore $\chi(G_{a'}[\mathcal{S}_{a'}]) = \omega(G_{a'}[\mathcal{S}_{a'}])$, as $\Big(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\Big)[\mathcal{S}_{a'}] = G_{a'}[\mathcal{S}_{a'}]$. Thus all the graphs $(G_a)_{a\in\mathcal{A}}$ are perfect.

($\Longleftarrow$) Conversely, assume that for all $a \in \mathcal{A}$, $G_a = (\mathcal{V}_a, \mathcal{E}_a, P_{V_a})$ is perfect. Then for all $\mathcal{S} \subset \bigsqcup_{a\in\mathcal{A}} \mathcal{V}_a$, $\mathcal{S}$ can be written as $\bigsqcup_{a\in\mathcal{A}} \mathcal{S}_a$ where $\mathcal{S}_a \subset \mathcal{V}_a$ for all $a \in \mathcal{A}$, and we have for all $P_A \in \Delta(\mathcal{A})$:

$$\chi\left(\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\right)[\mathcal{S}]\right) = \chi\left(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a[\mathcal{S}_a]\right) \tag{B.265}$$

$$= \max_{a\in\mathcal{A}} \chi\left(G_a[\mathcal{S}_a]\right) \tag{B.266}$$

$$= \max_{a\in\mathcal{A}} \omega\left(G_a[\mathcal{S}_a]\right), \tag{B.267}$$

and similarly, $\omega\Big(\Big(\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a\Big)[\mathcal{S}]\Big) = \max_{a\in\mathcal{A}} \omega\left(G_a[\mathcal{S}_a]\right)$. Hence $\bigsqcup_{a\in\mathcal{A}}^{P_A} G_a$ is also perfect.

B.6.10 Proof of Lemma B.6.1

Let $(\overline{a}_n)_{n\in\mathbb{N}^\star} \in \mathcal{A}^{\mathbb{N}^\star}$ be a sequence such that $T_{\overline{a}^n} \to P_A = \beta P'_A + (1-\beta)P''_A$ when $n \to \infty$.

Consider a sequence $(B_n)_{n\in\mathbb{N}^\star}$ of independent Bernoulli random variables such that for all $n \in \mathbb{N}^\star$,

$$\mathbb{P}(B_n = 0) = \frac{\beta P'_A(\overline{a}_n)}{P_A(\overline{a}_n)}. \tag{B.268}$$

By the strong law of large numbers,

$$\mathbb{P}\left(T_{B^n,\overline{a}^n} \underset{n\to\infty}{\to} (\beta P'_A, (1-\beta)P''_A)\right) = 1. \tag{B.269}$$

Therefore, there exists at least one realization $(b_n)_{n\in\mathbb{N}^\star}$ of $(B_n)_{n\in\mathbb{N}^\star}$ such that $T_{b^n,\overline{a}^n}$ converges to $\left(\beta P'_A, (1-\beta)P''_A\right)$. The convergences of marginal and conditional types yield

$$T_{b^n} \underset{n\to\infty}{\to} (\beta, 1-\beta), \tag{B.270}$$

$$T_{a'^n} \underset{n\to\infty}{\to} P'_A, \qquad T_{a''^n} \underset{n\to\infty}{\to} P''_A, \tag{B.271}$$

where $a' \doteq (\overline{a}_n)_{\substack{n\in\mathbb{N}^\star,\\ b_n=0}}$, and $a'' \doteq (\overline{a}_n)_{\substack{n\in\mathbb{N}^\star,\\ b_n=1}}$, are the extracted sequences.

B.6.11 Proof of Lemma B.6.2

Let $c^* : \mathcal{V} \to \mathcal{C}$ and $c^*_{\mathcal{S}} : \mathcal{S} \to \mathcal{C}$ be the optimal colorings of G and $G[\mathcal{S}]$, respectively. Consider the coloring $c : \mathcal{V} \to \mathcal{C} \sqcup \mathcal{V}$ of G defined by $c(v) = c^*_{\mathcal{S}}$ if $v \in \mathcal{S}$, $c(v) = v$ otherwise.

(Lower bound) On one hand, we have

$$\begin{aligned}
H_\chi(G) &\leq H(c(V), \mathbb{1}_{V\in\mathcal{S}}) && \text{(B.272)}\\
&= H(\mathbb{1}_{V\in\mathcal{S}}) + P_V(\mathcal{S})H(c(V)|V \in \mathcal{S}) \\
&\quad + (1 - P_V(\mathcal{S}))H(c(V)|V \notin \mathcal{S}) && \text{(B.273)}\\
&\leq 1 + H(c^*_{\mathcal{S}}(V)|V \in \mathcal{S}) + (1 - P_V(\mathcal{S})) \log |\mathcal{V}| && \text{(B.274)}\\
&= H_\chi(G[\mathcal{S}]) + 1 + (1 - P_V(\mathcal{S})) \log |\mathcal{V}|; && \text{(B.275)}
\end{aligned}$$

where (B.272) comes from the fact that c is a coloring of G; (B.273) is a decomposition using conditional entropies; (B.274) comes from the construction of c: $c|_{\mathcal{S}} = c^*_{\mathcal{S}}$; (B.275) follows from the optimality of $c^*_{\mathcal{S}}$ as a coloring of $G[\mathcal{S}]$.

(Upper bound) On the other hand,

$$\begin{aligned}
&H_\chi(G[\mathcal{S}]) \\
&\leq H(c^*(V)|V \in \mathcal{S}) && \text{(B.276)}\\
&= \frac{1}{P_V(\mathcal{S})}\Big(H(c^*(V)|\mathbb{1}_{V\in\mathcal{S}}) - (1 - P_V(\mathcal{S}))H(c^*(V)|V \notin \mathcal{S})\Big) && \text{(B.277)}\\
&\leq \frac{H(c^*(V))}{P_V(\mathcal{S})} = \frac{H_\chi(G)}{P_V(\mathcal{S})} && \text{(B.278)}
\end{aligned}$$

where (B.276) comes from the fact that c^* induces a coloring of $G[\mathcal{S}]$; (B.277) is a decomposition using conditional entropies; (B.278) results from the elimination of negative terms and the optimality of c^*.

B.6.12 Proof of Lemma B.6.4

Let $(v^m, v^n) \in \mathcal{T}_\epsilon^m(P_V) \times \mathcal{T}_\epsilon^n(P'_V)$. We have

$$\left\| T_{v^{m+n}} - \left(\frac{m}{n+m} P_V + \frac{n}{n+m} P'_V \right) \right\|_\infty \tag{B.279}$$

$$= \left\| \frac{m}{n+m} T_{v^m} + \frac{n}{n+m} T_{v^n} - \frac{m}{n+m} P_V - \frac{n}{n+m} P'_V \right\|_\infty \tag{B.280}$$

$$\leq \frac{m}{n+m} \left\| T_{v^m} - P_V \right\|_\infty + \frac{n}{n+m} \left\| T_{v^n} - P'_V \right\|_\infty \tag{B.281}$$

$$\leq \frac{m}{n+m} \epsilon + \frac{n}{n+m} \epsilon = \epsilon, \tag{B.282}$$

hence the desired result.

B.6.13 Proof of Lemma B.6.5

The graphs $G[\mathcal{S}] \wedge G'[\mathcal{S}']$ and $(G \wedge G')[\mathcal{S} \times \mathcal{S}']$ both have $\mathcal{S} \times \mathcal{S}'$ as set of vertices. For all $(v_1, v'_1), (v_2, v'_2) \in \mathcal{S} \times \mathcal{S}'$, we have:

$$(v_1, v'_1), (v_2, v'_2) \text{ are adjacent in } G[\mathcal{S}] \wedge G'[\mathcal{S}'] \tag{B.283}$$

$$\iff v_1, v_2 \text{ are adjacent in } G \text{ and } v'_1, v'_2 \text{ are adjacent in } G' \tag{B.284}$$

$$\iff (v_1, v'_1), (v_2, v'_2) \text{ are adjacent in } (G \wedge G')[\mathcal{S} \times \mathcal{S}']. \tag{B.285}$$

Appendix C

Proofs for coding for computing zero-error results

C.1 Proof of Theorem 6.3.2

Consider the particular case $f(X,Y) = X$ of Figure 6.2. The optimal rate in this particular case equals the optimal rate R^* in the following auxiliary problem, depicted in Figure C.1: $(X, g(Y))$ as source available at the encoder and to be retrieved by the decoder which knows Y (thus expecting it to retrieve $g(Y)$ in addition to X does not change the optimal rate).

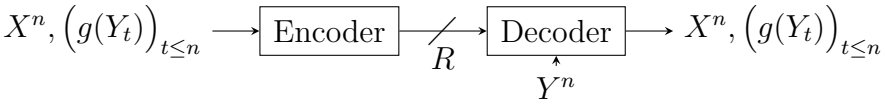


Figure C.1 – An auxiliary zero-error Slepian-Wolf problem.

This auxiliary problem is a particular instance of the zero-error Slepian-Wolf problem; its optimal rate writes $\overline{H}(G)$, where G is the characteristic graph defined in Definition 3.2.4 for the pair $\Big((X, g(Y)), Y\Big)$. The graph G has $\mathcal{X} \times \mathcal{Z}$ as set of vertices, and (x,z) is adjacent to (x',z') if there exists a side information symbol $y \in \mathcal{Y}$ such that $P_{X,Y,g(Y)}(x,y,z)P_{X,Y,g(Y)}(x',y,z') > 0$. It can be observed that the vertices (x,z) and (x',z') such that $z \neq z'$ are not adjacent in G. The graph G is therefore a disjoint union indexed by $\mathcal{Z}$:

$$G = \bigsqcup_{z\in\mathcal{Z}}^{P_{g(Y)}} G_z; \tag{C.1}$$

$$R^* = \overline{H}(G) = \overline{H}\left(\bigsqcup_{z\in\mathcal{Z}}^{P_{g(Y)}} G_z\right); \tag{C.2}$$

where for all $z \in \mathcal{Z}$, G_z is the characteristic graph defined in Definition 3.2.4 for the pair $(X'_z, Y'_z) \sim P_{X,Y|g(Y)=z}$.

($\Rightarrow$) Assume that g and $P_{X,Y}$ satisfy the "pairwise shared side information" condition. It directly follows that $P_{X|g(Y)=z}$ is full-support for all $z \in \mathcal{Z}$. Let $z \in \mathcal{Z}$, and let $(x,z),(x',z)$ be any two vertices of G_z. By construction, there exists $y \in g^{-1}(z)$ such that $P_{X,Y}(x,y)P_{X,Y}(x',y) > 0$; hence $P_{X,Y,g(Y)}(x,y,z)P_{X,Y,g(Y)}(x',y,z) > 0$, and $(x,z),(x',z)$ are adjacent in G_z. Each graph G_z is therefore complete, and perfect; the graph $G = \bigsqcup_{z\in\mathcal{Z}}^{P_{g(Y)}} G_z$ is a disjoint union of perfect graphs and is also perfect by Lemma B.5.2. We have:

$$\begin{aligned} R^* &= \overline{H}\left(\bigsqcup_{z\in\mathcal{Z}}^{P_{g(Y)}} G_z\right) && \text{(C.3)}\\ &= H_\kappa\left(\bigsqcup_{z\in\mathcal{Z}}^{P_{g(Y)}} G_z\right) && \text{(C.4)}\\ &= \sum_{z\in\mathcal{Z}} P_{g(Y)}(z) H_\kappa(G_z) && \text{(C.5)}\\ &= \sum_{z\in\mathcal{Z}} P_{g(Y)}(z) H(P_{X|g(Y)=z}) && \text{(C.6)}\\ &= H(X|g(Y)); && \text{(C.7)} \end{aligned}$$

where (C.3) comes from (C.2); (C.4) and (C.5) follow from Theorem 5.5.8 used on the perfect graph $\bigsqcup_{z\in\mathcal{Z}}^{P_{g(Y)}} G_z$; and (C.6) holds as the independent subsets of the complete graph G_z are singletons containing one of its vertices.

($\Leftarrow$) Conversely, assume that $P_{X|g(Y)=z}$ is full-support for all $z \in \mathcal{Z}$, and $R^* = H(X|g(Y))$.

Assume, *ad absurdum*, that at least one of the $(G_z)_{z\in\mathcal{Z}}$ is not complete; then there exists a coloring of that graph that maps two different vertices to the same color. Thus, there exists $z \in \mathcal{Z}$ such that

$$\overline{H}(G_z) < H(P_{X|g(Y)=z}), \tag{C.8}$$

as $P_{X|g(Y)=z}$ is full-support. We have

$$\begin{aligned} H(X|g(Y)) &= R^* && \text{(C.9)}\\ &= \overline{H}\left(\bigsqcup_{z\in\mathcal{Z}}^{P_{g(Y)}} G_z\right) && \text{(C.10)}\\ &\leq \sum_{z\in\mathcal{Z}} P_{g(Y)}(z)\overline{H}(G_z) && \text{(C.11)}\\ &< H(X|g(Y)); && \text{(C.12)} \end{aligned}$$

where (C.10) comes from (C.2), (C.11) results from [71, Theorem 2], and (C.12) follows from (C.8). We arrive at a contradiction, hence all the graphs $(G_z)_{z\in\mathcal{Z}}$ are complete: for all $z \in \mathcal{Z}$ and $x, x' \in \mathcal{X}$, there exists a side information symbol $y \in \mathcal{Y}$ such that $P_{X,Y,g(Y)}(x, y, z)P_{X,Y,g(Y)}(x', y, z) > 0$; hence $y \in g^{-1}(z)$, and satisfies $P_{X,Y}(x, y)P_{X,Y}(x', y) > 0$. The condition "pairwise shared side information" is satisfied by $P_{X,Y}, g$.

C.2 Proof of Theorem 6.3.6

Let us specify the adjacency condition in $G_{[n]}$ under the assumption (6.14). Two vertices are adjacent if they satisfy (6.9) and (6.10); however (6.9) is always satisfied under (6.14). Thus $(x^n, z^n)(x'^n, z^n)$ are adjacent if $z^n = z'^n$ and

$$\exists y^n \in g^{-1}(z^n), \exists t \leq n, f(x_t, y_t) \neq f(x'_t, y_t). \tag{C.13}$$

It can be observed that the condition (C.13) is the adjacency condition of an OR product of adequate graphs; more precisely,

$$G_{[n]} = \bigsqcup_{z^n\in\mathcal{Z}^n} \bigvee_{t\leq n} G^f_{z_t}. \tag{C.14}$$

Although $G_{[n]}$ cannot be expressed as an n-th OR power, we will show that its chromatic entropy asymptotically coincide with that of an appropriate OR power: we now search for an asymptotic equivalent of $H_\chi(G_{[n]})$.

Definition C.2.1 *$\mathcal{S}_n$ is the set of colorings of $G_{[n]}$ that can be written as $(x^n, z^n) \mapsto (T_{z^n}, \tilde{c}(x^n, z^n))$ for some mapping $\tilde{c} : \mathcal{X}^n \times \mathcal{Z}^n \to \tilde{\mathcal{C}}$; where T_{z^n} denotes the type of z^n.*

In the following, we define $Z^n \doteq \big(g(Y_t)\big)_{t\leq n}$. Now we need several Lemmas. Lemma C.2.2 states that the optimal coloring $c(x^n, z^n)$ of $G_{[n]}$ has the type of z^n as a prefix at a negligible rate cost. Lemma C.2.3 gives an asymptotic formula for the minimal entropy of the colorings from $\mathcal{S}_n$.

Lemma C.2.2 *The following asymptotic comparison holds:*

$$H_\chi(G_{[n]}) = \inf_{\substack{c \text{ coloring of } G_{[n]} \\ \text{s.t. } c\in\mathcal{S}_n}} H(c(X^n, Z^n)) + O(\log n). \tag{C.15}$$

Lemma C.2.3 *The following asymptotic comparison holds:*

$$\inf_{\substack{c \text{ coloring of } G_{[n]} \\ \text{s.t. } c \in \mathcal{S}_n}} H(c(X^n, Z^n)) = n \sum_{z \in \mathcal{Z}} P_{g(Y)}(z) H_\kappa(G_z^f) + o(n). \tag{C.16}$$

The proof of Lemma C.2.2 is given in Appendix C.3, its keypoint is the asymptotically negligible entropy of the prefix T_{Z^n} of the colorings of $\mathcal{S}_n$.

The proof of Lemma C.2.3 is given in Appendix C.4, and relies on the decomposition $G_{[n]} = \bigsqcup_{Q_n \in \Delta_n(\mathcal{Z})} G_{[n]}^{Q_n}$, where $G_{[n]}^{Q_n}$ is the subgraph induced by the vertices (x^n, z^n) such that the type of z^n is Q_n. We show that $G_{[n]}^{Q_n}$ is a disjoint union of isomorphic graphs whose chromatic entropy is given by Lemma B.2.7 and (6.16): $\left| H_\chi(G_{[n]}^{Q_n}) - n \sum_{z \in \mathcal{Z}} Q_n(z) H_\kappa(G_z^f) \right| \leq n\epsilon_n$. Finally, uniform convergence arguments enable us to conclude.

Now let us combine these results together:

$$\begin{aligned} R^* &= \frac{1}{n} H_\chi(G_{[n]}) + o(1) && \text{(C.17)} \\ &= \frac{1}{n} \inf_{\substack{c \text{ coloring of } G_{[n]} \\ \text{s.t. } c \in \mathcal{S}_n}} H(c(X^n, Z^n)) + o(1) && \text{(C.18)} \\ &= \sum_{z \in \mathcal{Z}} P_{g(Y)}(z) H_\kappa(G_z^f) + o(1), && \text{(C.19)} \end{aligned}$$

where (C.17) comes from Theorem 6.2.2, (C.18) comes from Lemma C.2.2, and (C.19) comes from Lemma C.2.3. The proof of Theorem 6.3.6 is complete.

C.3 Proof of Lemma C.2.2

Let c_n^* be the coloring of $G_{[n]}$ with minimal entropy. Then we have:

$$\begin{aligned} H_\chi(G_{[n]}) &= \inf_{c \text{ coloring of } G_{[n]}} H(c(X^n, Z^n)) && \text{(C.20)} \\ &\leq \inf_{\substack{c \text{ coloring of } G_{[n]} \\ \text{s.t. } c \in \mathcal{S}_n}} H(c(X^n, Z^n)) && \text{(C.21)} \\ &= \inf_{\substack{c : (x^n, z^n) \\ \mapsto (T_{z^n}, \tilde{c}(x^n, z^n))}} H(T_{Z^n}, \tilde{c}(X^n, Z^n)) && \text{(C.22)} \\ &\leq H(T_{Z^n}) + H(c_n^*(X^n, Z^n)) && \text{(C.23)} \\ &= H_\chi(G_{[n]}) + O(\log n), && \text{(C.24)} \end{aligned}$$

where (C.22) comes from Definition C.2.1; (C.23) comes from the subadditivity of the entropy, and the fact that $(x^n, z^n) \mapsto (T_{z^n}, c_n^*(x^n, z^n))$ is a coloring of $G_{[n]}$ that belongs to $\mathcal{S}_n$; and (C.24) comes from $H(T_{Z^n}) = O(\log n)$, as $\log |\Delta_n(\mathcal{Z})| = O(\log n)$. The desired equality comes from the bounds $H_\chi(G_{[n]})$ and $H_\chi(G_{[n]}) + O(\log n)$ on (C.21).

C.4 Proof of Lemma C.2.3

For all $Q_n \in \Delta_n(\mathcal{Z})$, let

$$G_{[n]}^{Q_n} = \bigsqcup_{\substack{z^n \in \mathcal{Z}^n \\ T_{z^n} = Q_n}} \bigvee_{t \le n} G_{z_t}^f, \tag{C.25}$$

with the probability distribution induced by $P_{X,Z}^n$. This graph is formed of the connected components of $G_{[n]}$ whose corresponding z^n has type Q_n. We need to find an equivalent for $H_\chi(G_{[n]}^{Q_n})$. Since $G_{[n]}^{Q_n}$ is a disjoint union of isomorphic graphs, we can use Lemma B.2.7:

$$H_\chi(G_{[n]}^{Q_n}) = H_\chi\left(\bigvee_{z \in \mathcal{Z}} (G_z^f)^{\vee n Q_n(z)} \right). \tag{C.26}$$

On one hand,

$$\begin{aligned} H_\chi\left(\bigvee_{z \in \mathcal{Z}} (G_z^f)^{\vee n Q_n(z)} \right) &\ge H_\kappa\left(\bigvee_{z \in \mathcal{Z}} (G_z^f)^{\vee n Q_n(z)} \right) && \text{(C.27)} \\ &= n \sum_{z \in \mathcal{Z}} Q_n(z) H_\kappa(G_z^f), && \text{(C.28)} \end{aligned}$$

where (C.27) comes from $H_\kappa \le H_\chi$ [5, Lemma 14], (C.28) comes from (6.16). On the other hand,

$$\begin{aligned} H_\chi\left(\bigvee_{z \in \mathcal{Z}} (G_z^f)^{\vee n Q_n(z)} \right) &\le \sum_{z \in \mathcal{Z}} Q_n(z) H_\chi((G_z^f)^{\vee n}) && \text{(C.29)} \\ &= n \sum_{z \in \mathcal{Z}} Q_n(z) H_\kappa(G_z^f) + n\epsilon_n, && \text{(C.30)} \end{aligned}$$

where $\epsilon_n \doteq \max_z \frac{1}{n} H_\chi((G_z^f)^{\vee n}) - H_\kappa(G_z^f)$ is a quantity that does not depend on Q_n and satisfies $\lim_{n\to\infty} \epsilon_n = 0$; (C.29) comes from the subadditivity of H_χ. Combining equations

(C.26), (C.28) and (C.30) yields

$$\left| H_\chi(G_{[n]}^{Q_n}) - n \sum_{z \in \mathcal{Z}} Q_n(z) H_\kappa(G_z^f) \right| \leq n\epsilon_n. \tag{C.31}$$

Now, we have an equivalent for $H_\chi(G_{[n]}^{Q_n})$.

$$\begin{aligned}
&\inf_{\substack{c \text{ coloring of } G_{[n]} \\ \text{s.t. } c \in \mathcal{S}_n}} H(c(X^n, Z^n)) && \text{(C.32)}\\
&= \inf_{\substack{c:(x^n,z^n) \\ \mapsto (T_{z^n}, \tilde{c}(x^n,z^n))}} H(\tilde{c}(X^n, Z^n)|T_{Z^n}) + H(T_{Z^n}) && \text{(C.33)}\\
&= \inf_{\substack{c:(x^n,z^n) \\ \mapsto (T_{z^n}, \tilde{c}(x^n,z^n))}} \sum_{Q_n \in \Delta_n(\mathcal{Z})} P_{T_{Z^n}}(Q_n) H(\tilde{c}(X^n, Z^n)|T_{Z^n} = Q_n) + O(\log n) && \text{(C.34)}\\
&= \sum_{Q_n \in \Delta_n(\mathcal{Z})} P_{T_{Z^n}}(Q_n) \inf_{c_{Q_n} \text{ coloring of } G_{[n]}^{Q_n}} H(c_{Q_n}(X^n, Z^n)|T_{Z^n} = Q_n) + O(\log n) && \text{(C.35)}\\
&= \sum_{Q_n \in \Delta_n(\mathcal{Z})} P_{T_{Z^n}}(Q_n) H_\chi(G_{[n]}^{Q_n}) + O(\log n) && \text{(C.36)}\\
&= \sum_{Q_n \in \Delta_n(\mathcal{Z})} P_{T_{Z^n}}(Q_n) \left(n \sum_{z \in \mathcal{Z}} Q_n(z) H_\kappa(G_z^f) \pm n\epsilon_n \right) + O(\log n) && \text{(C.37)}\\
&= n \sum_{Q_n \in \Delta_n(\mathcal{Z})} 2^{-nD(Q_n \| P_{g(Y)}) + o(n)} \left(\sum_{z \in \mathcal{Z}} Q_n(z) H_\kappa(G_z^f) \right) \pm n\epsilon_n + O(\log n) && \text{(C.38)}\\
&= n \sum_{z \in \mathcal{Z}} P_{g(Y)}(z) H_\kappa(G_z^f) + o(n), && \text{(C.39)}
\end{aligned}$$

where (C.34) comes from $H(T_{Z^n}) = O(\log n)$, as $\log |\Delta_n(\mathcal{Z})| = O(\log n)$; (C.35) follows from the fact that the entropy of $\tilde{c}$ can be minimized independently on each $G_{[n]}^{Q_n}$; (C.36) follows from the definition of $G_{[n]}^{Q_n}$; (C.37) comes from (C.31); (C.38) comes from [20, Lemma 2.6] and the fact that ϵ_n does not depend on Q_n.

BIBLIOGRAPHY

[1] Rudolf Ahlswede, Ning Cai, and Zhen Zhang, « Zero-error capacity for models with memory and the enlightened dictator channel », *in*: *IEEE Transactions on Information Theory* 44.*3* (1998), pp. 1250–1252.

[2] Emrah Akyol et al., « On scalable coding in the presence of decoder side information », *in*: *2014 IEEE International Symposium on Information Theory (ISIT)*.

[3] Noga Alon, « On the capacity of digraphs », *in*: *European Journal of Combinatorics* 19.*1* (1998), pp. 1–5.

[4] Noga Alon, « The Shannon capacity of a union », *in*: *Combinatorica* 18.*3* (1998), pp. 301–310.

[5] Noga Alon and Alon Orlitsky, « Source coding and graph entropies », *in*: *IEEE Transactions on Information Theory* 42.*5* (1996), pp. 1329–1339.

[6] Sourya Basu, Daewon Seo, and Lav R Varshney, « Hypergraph-based Coding Schemes for Two Source Coding Problems under Maximal Distortion », *in*: *IEEE International Symposium on Information Theory (ISIT)*, 2020.

[7] Meryem Benammar and Abdellatif Zaidi, « Rate-Distortion Region of a Gray–Wyner Model with Side Information », *in*: *Entropy* 20.*1* (2018), p. 2.

[8] Navid Mahmoudian Bidgoli, Thomas Maugey, and Aline Roumy, « Fine granularity access in interactive compression of 360-degree images based on rate-adaptive channel codes », *in*: *IEEE Transactions on Multimedia* 23 (2020), pp. 2868–2882.

[9] Holger Boche and Christian Deppe, « Computability of the zero-error capacity of noisy channels », *in*: *2021 IEEE Information Theory Workshop (ITW)*, IEEE, 2021, pp. 1–6.

[10] Boris Bukh and Christopher Cox, « On a fractional version of Haemers' bound », *in*: *IEEE Transactions on Information Theory* 65.*6* (2018), pp. 3340–3348.

[11] Peter J Cameron, « 6-transitive graphs », *in*: *Journal of Combinatorial Theory, Series B* 28.*2* (1980), pp. 168–179.

[12] Qi Cao et al., « On zero-error capacity of binary channels with one memory », *in*: *IEEE Transactions on Information Theory* 64.*10* (2018), pp. 6771–6778.

[13] Jun Chen, Da-Ke He, and Ashish Jagmohan, « The equivalence between Slepian-Wolf coding and channel coding under density evolution », *in*: *IEEE Transactions on Communications* 57.*9* (2009), pp. 2534–2540.

[14] Jun Chen et al., « On the linear codebook-level duality between Slepian–Wolf coding and channel coding », *in*: *IEEE Transactions on Information Theory* 55.*12* (2009), pp. 5575–5590.

[15] Maria Chudnovsky and Paul D Seymour, « The structure of claw-free graphs. », *in*: *BCC*, 2005, pp. 153–171.

[16] Maria Chudnovsky et al., « The strong perfect graph theorem », *in*: *Annals of mathematics* (2006), pp. 51–229.

[17] Gérard Cohen, Emanuela Fachini, and János Körner, « Zero-error capacity of binary channels with memory », *in*: *arXiv preprint arXiv:1402.0349* (2014).

[18] Thomas M Cover, *Elements of information theory*, John Wiley & Sons, 1999.

[19] Imre Csiszar, « Linear codes for sources and source networks: Error exponents, universal coding », *in*: *IEEE Transactions on Information Theory* 28.*4* (1982), pp. 585–592.

[20] Imre Csiszár and János Körner, *Information theory: coding theorems for discrete memoryless systems*, Cambridge University Press, 2011.

[21] Imre Csiszár and János Körner, « On the capacity of the arbitrarily varying channel for maximum probability of error », *in*: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 57.*1* (1981), pp. 87–101.

[22] Imre Csiszár et al., « Entropy splitting for antiblocking corners and perfect graphs », *in*: *Combinatorica* 10.*1* (1990), pp. 27–40.

[23] Lingyu Duan et al., « Video coding for machines: A paradigm of collaborative compression and intelligent analytics », *in*: *IEEE Transactions on Image Processing* 29 (2020), pp. 8680–8695.

[24] Soheil Feizi and Muriel Médard, « On network functional compression », *in*: *IEEE transactions on information theory* 60.*9* (2014), pp. 5387–5401.

[25] Michael Fekete, « Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten », *in*: *Mathematische Zeitschrift* 17.*1* (1923), pp. 228–249.

[26] Li Gao, Sander Gribling, and Yinan Li, « On a tracial version of Haemers bound », *in*: *IEEE Transactions on Information Theory* (2022).

[27] Wen Gao et al., « Recent standard development activities on video coding for machines », *in*: *arXiv preprint arXiv:2105.12653* (2021).

[28] L Gargano, J Körner, and U Vaccaro, « Sperner theorems on directed graphs and qualitative independence », *in*: *J. Combin. Theory Ser. A* 61 (1992), pp. 173–192.

[29] Luisa Gargano, János Körner, and Ugo Vaccaro, « Capacities: from information theory to extremal set theory », *in*: *Journal of Combinatorial Theory, Series A* 68.*2* (1994), pp. 296–316.

[30] RM Gray and AD Wyner, « Source coding for a simple network », *in*: *Bell System Technical Journal* 53.*9* (1974), pp. 1681–1721.

[31] Yujie Gu and Ofer Shayevitz, « On the non-adaptive zero-error capacity of the discrete memoryless two-way channel », *in*: *Entropy* 23.*11* (2021), p. 1518.

[32] Xuan Guang et al., « Improved upper bound on the network function computing capacity », *in*: *IEEE Transactions on Information Theory* 65.*6* (2019), pp. 3790–3811.

[33] Willem Haemers et al., « On some problems of Lovász concerning the Shannon capacity of a graph », *in*: *IEEE Transactions on Information Theory* 25.*2* (1979), pp. 231–232.

[34] David A Huffman, « A method for the construction of minimum-redundancy codes », *in*: *Proceedings of the IRE* 40.*9* (1952), pp. 1098–1101.

[35] Amiram H. Kaspi, « Rate-distortion function when side-information may be present at the decoder », *in*: *IEEE Transactions on Information Theory* 40.*6* (1994), pp. 2031–2034.

[36] Peter Keevash and Eoin Long, « On the normalized Shannon capacity of a union », *in*: *Combinatorics, Probability and Computing* 25.*5* (2016), pp. 766–767.

[37] J Korner and Giuseppe Longo, « Two-step encoding for finite sources », *in*: *IEEE Transactions on Information Theory* 19.*6* (1973), pp. 778–782.

[38] Janos Korner and Alon Orlitsky, « Zero-error information theory », *in*: *IEEE Transactions on Information Theory* 44.*6* (1998), pp. 2207–2229.

[39] János Körner, « Coding of an information source having ambiguous alphabet and the entropy of graphs », *in*: *6th Prague conference on information theory*, 1973, pp. 411–425.

[40] János Körner, Concetta Pilotto, and Gábor Simonyi, « Local chromatic number and Sperner capacity », *in*: *Journal of Combinatorial Theory, Series B* 95.*1* (2005), pp. 101–117.

[41] Prashant Koulgi et al., « On zero-error source coding with decoder side information », *in*: *IEEE Transactions on Information Theory* 49.*1* (2003), pp. 99–111.

[42] Dinesh Krithivasan and S Sandeep Pradhan, « Distributed source coding using abelian group codes: A new achievable rate-distortion region », *in*: *IEEE Transactions on Information Theory* 57.*3* (2011).

[43] Thomas Laich and Michèle Wigger, « Utility of encoder side information for the lossless kaspi/heegard-b erger problem », *in*: *2013 IEEE International Symposium on Information Theory*, IEEE, 2013, pp. 3065–3069.

[44] László Lovász, « On the Shannon capacity of a graph », *in*: *IEEE Transactions on Information theory* 25.*1* (1979), pp. 1–7.

[45] Rick Ma and Samuel Cheng, « Zero-error Slepian–Wolf coding of confined-correlated sources with deviation symmetry », *in*: *IEEE Transactions on Information Theory* 59.*12* (2013), pp. 8195–8209.

[46] Derya Malak and Muriel Médard, « Hyper Binning for Distributed Function Coding », *in*: *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, IEEE, 2020, pp. 1–5.

[47] Katalin Marton, « On the Shannon capacity of probabilistic graphs », *in*: *Journal of Combinatorial Theory, Series B* 57.*2* (1993), pp. 183–195.

[48] K Ashik Mathew and Patric RJ Östergård, « New lower bounds for the Shannon capacity of odd cycles », *in*: *Designs, Codes and Cryptography* 84 (2017), pp. 13–22.

[49] Thomas Maugey et al., « Incremental coding for extractable compression in the context of massive random access », *in*: *IEEE transactions on Signal and Information Processing over Networks* 6 (2020), pp. 251–260.

[50] Alon Orlitsky and James R Roche, « Coding for computing », *in*: *Proceedings of IEEE 36th Annual Foundations of Computer Science*, IEEE, 1995, pp. 502–511.

[51] Sven C Polak and Alexander Schrijver, « New lower bound on the Shannon capacity of C7 from circular graphs », *in*: *Information Processing Letters* 143 (2019), pp. 37–40.

[52] Jithin Ravi and Bikash Kumar Dey, « Function Computation Through a Bidirectional Relay », *in*: *IEEE Transactions on Information Theory* 65.*2* (2018), pp. 902–916.

[53] Aline Roumy and Thomas Maugey, « Universal lossless coding with random user access: the cost of interactivity », *in*: *2015 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2015, pp. 1870–1874.

[54] Igal Sason, « Observations on the Lovász θ-Function, Graph Capacity, Eigenvalues, and Strong Products », *in*: *Entropy* 25.*1* (2023), p. 104.

[55] Alexander Schrijver, « On the Shannon capacity of sums and products of graphs », *in*: *Indagationes Mathematicae* 34.*1* (2023), pp. 37–41.

[56] Milad Sefidgaran and Aslan Tchamkerten, « Distributed function computation over a rooted directed tree », *in*: *IEEE Transactions on Information Theory* 62.*12* (2016), pp. 7135–7152.

[57] Andrea Sgarro, « Source coding with side information at several decoders », *in*: *IEEE Transactions on Information Theory* 23.*2* (1977), pp. 179–182.

[58] Claude Shannon, « The zero error capacity of a noisy channel », *in*: *IRE Transactions on Information Theory* 2.*3* (1956), pp. 8–19.

[59] Claude Elwood Shannon, « A mathematical theory of communication », *in*: *The Bell system technical journal* 27.*3* (1948), pp. 379–423.

[60] Ofer Shayevitz, « Distributed computing and the graph entropy region », *in*: *IEEE transactions on information theory* 60.*6* (2014), pp. 3435–3449.

[61] Gábor Simonyi, « Graph entropy: a survey », *in*: *Combinatorial Optimization* 20 (1995), pp. 399–441.

[62] Gábor Simonyi, « On Witsenhausen's zero-error rate for multiple sources », *in*: *IEEE Transactions on Information Theory* 49.*12* (2003), pp. 3258–3260.

[63] Gábor Simonyi, « Perfect graphs and graph entropy. An updated survey », *in*: *Perfect graphs* (2001), pp. 293–328.

[64] David Slepian and Jack Wolf, « Noiseless coding of correlated information sources », *in*: *IEEE Transactions on information Theory* 19.*4* (1973), pp. 471–480.

[65] Yossef Steinberg and Neri Merhav, « On successive refinement for the Wyner–Ziv problem », *in*: *IEEE Transactions on Information Theory* 50.*8* (2004), pp. 1636–1654.

[66] Chao Tian and Suhas N Diggavi, « Side-information scalable source coding », *in*: *IEEE Transactions on Information Theory* 54.*12* (2008), pp. 5591–5608.

[67] Roy Timo, Terence Chan, and Alex Grant, « Rate distortion with side-information at many decoders », *in*: *IEEE Transactions on Information Theory* 57.*8* (2011), pp. 5240–5257.

[68] Roy Timo et al., « Source coding for a simple network with receiver side information », *in*: *2008 IEEE International Symposium on Information Theory*, IEEE, 2008, pp. 2307–2311.

[69] Hugo Touchette, « Legendre-Fenchel transforms in a nutshell », *in*: *URL http://www. maths. qmul. ac. uk/˜ ht/archive/lfth2. pdf* (2005).

[70] Ertem Tuncel and Kenneth Rose, « On variable-length coding of sources with side information at multiple decoders », *in*: *IEEE International Symposium on Information Theory, 2003. Proceedings.* IEEE, 2003, p. 140.

[71] Ertem Tuncel et al., « On complementary graph entropy », *in*: *IEEE transactions on information theory* 55.*6* (2009), pp. 2537–2546.

[72] Aleksander Vesel and Janez Žerovnik, « Improved lower bound on the Shannon capacity of C7 », *in*: *Information processing letters* 81.*5* (2002), pp. 277–282.

[73] Lele Wang and Young-Han Kim, « Linear code duality between channel coding and Slepian-Wolf coding », *in*: *2015 53rd IEEE Annual Allerton Conference on Communication, Control, and Computing (Allerton).*

[74] Moritz Wiese et al., « Secure estimation and zero-error secrecy capacity », *in*: *IEEE Transactions on Automatic Control* 64.*3* (2018), pp. 1047–1062.

[75] Hans Witsenhausen, « The zero-error side information problem and chromatic numbers (corresp.) », *in*: *IEEE Transactions on Information Theory* 22.*5* (1976), pp. 592–593.

[76] Hirosuke Yamamoto, « Wyner-ziv theory for a general function of the correlated sources (corresp.) », *in*: *IEEE Transactions on Information Theory* 28.*5* (1982), pp. 803–807.

[77] Guofen Zhang et al., « The zero-error capacity of binary channels with 2-memories », *in*: *Advances in Mathematics of Communications* (2022).

[78] Jeroen Zuiddam et al., *Algebraic complexity, asymptotic spectra and entanglement polytopes*, Institute for Logic, Language and Computation, 2018.

COLLEGE MATHS, TELECOMS
DOCTORAL INFORMATIQUE, SIGNAL
BRETAGNE SYSTEMES, ELECTRONIQUE

Université de Rennes

Titre : Théorie de l'information zéro-erreur dans les réseaux : graphes, codage pour le calcul de fonctions et dualité source-canal

Mot clés : Théorie de l'information zéro-erreur, codage de source, dualité source-canal

Résumé : Cette thèse de doctorat porte sur la théorie de l'information zéro-erreur, notamment sur le codage de source avec information adjacente, le codage de canal et la dualité source-canal. Ces travaux s'articulent autour de trois axes : 1. les problèmes de codage zéro-erreur basés sur la correction, 2. les problèmes de codage zéro-erreur basés sur les graphes de confusion, 3. les problèmes de codage zéro-erreur pour le calcul d'une fonction. Une contribution importante de ce travail concerne le problème de la linéarisation du débit optimal, lorsque l'encodeur traite plusieurs tâches conjointement. Nous démontrons l'équivalence de la linéarisation pour les produits de graphes et pour leurs unions disjointes. Cette observation permet de caractériser les débits optimaux pour une classe de problèmes irrésolus depuis les années 1950. Les contributions de cette thèse ont donné lieu à des publications dans les actes des meilleurs conférences internationales de théorie de l'information et deux articles de revues sont en cours de préparation.

Title: Zero-error network information theory: graphs, coding for computing and source-channel duality

Keywords: Zero-error Information Theory, Source coding, Source-channel duality

Abstract: This doctoral thesis focuses on zero-error information theory, particularly on source coding with side information, channel coding and source-channel duality. These works revolve around three axes: 1. zero-error coding problems based on correction, 2. zero-error coding problems based on confusion graphs, 3. zero-error coding problems for calculating a function. An important contribution of this work concerns the problem of linearizing the optimal bitrate, when the encoder processes several tasks jointly. We demonstrate the equivalence of linearization for graph products and for their disjoint unions. This observation makes it possible to characterize the optimal rates for a class of unsolved problems since the 1950s. The contributions of this thesis have given rise to publications in the proceedings of the best international Information Theory conferences and two journal articles are in preparation.