Coordination of autonomous devices over noisy channels: capacity results and coding techniques

Giulia Cervia

Université / / Paris Seine

École Doctorale EM2P
Economie, Management, Mathématiques et Physique

THÈSE

pour obtenir le titre de

Docteur en Sciences

Spécialité Sciences et Technologies
de l'Information et de la Communication

par

Giulia Cervia

Coordination of autonomous devices over noisy channels: capacity results and coding techniques

30 Novembre 2018

Devant le jury composé de:

Rapporteure	Michèle Wigger	Professeure à Télécom ParisTech
Rapporteur	Albert Guillén i Fàbregas	Professeur à l'Universitat Pompeu Fabra
Président du jury	Pierre Duhamel	Directeur de recherches CNRS à L2S
Examinateur	Jean-Claude Belfiore	Professeur à Télécom ParisTech
Directrice de thèse	Inbar Fijalkow	Professeure des universités à l'ENSEA
Co-encadrante	Laura Luzzi	Maître de conférences à l'ENSEA
Co-encadrant	Maël Le Treust	Chargé de recherche CNRS à ETIS

ACKNOWLEDGEMENTS

I would like to start by thanking my advisor Inbar Fijalkow, for being an invaluable point of reference during these three years at ETIS.
This thesis has been possible because of the unfailing support of my supervisors Laura Luzzi and Maël Le Treust. I am very grateful to them for their help and guidance, as well as for the capacity of sharing a laugh every time we were stuck. A separate thanks to Laura for choosing me among other candidates and giving me this opportunity.
Thanks to Michèle Wigger and Albert Guillén i Fàbregas for being the referees for my thesis. They provided many helpful comments and their meticulous work has been of great value for me.
I also would like to thank Jean-Claude Belfiore and Michel Kieffer for accepting to participate to my Ph.D. committee, and Pierre Duhamel for accepting a very last minute invitation to join (and be president of) the committee, litterally saving the day.
Last but not the least, I would like to thank Matthieu Bloch for the essential contributions he gave to my research.

CONTENTS

1 | Introduction

1.1 Context and motivation

The fifth generation of wireless networks (5G), whose standardization is planned for 2020, envisions machine to machine communication and the Internet of Things: a unified network of connected objects including embedded sensors, medical devices, smart meters, and autonomous vehicles. While 4G revolutionized the smartphone experience, 5G will have an even bigger impact on smart consumer items. The anticipated explosion of device-to-device communications, with the perspective of 30 billion connected devices by 2020, creates new challenges. The new networking standard will not just be about faster communications, but it will also support the next wave of technological innovation, from connected cars to factory automation, smart cities, robot-assisted surgery, virtual reality and edge computing.

Although in the past communication networks have traditionally been designed with the purpose of reliably conveying information, it is crucial for the next generation of networks to ensure the *cooperation* and *coordination* of the constituent devices, viewed as autonomous decision makers. In this context, we consider coordination as a way to enforce a prescribed behavior.

We can think of a variety of applications that would benefit from coordination in this sense. The nodes might be agents playing in a cooperative game, having either the same objective or different purposes [50]. Alternatively, they might be computers which are part of the same network and cooperate in order to distribute their tasks appropriately, dealing with a work load that might be varying over time. They could be smart vehicles that receive live traffic information and improve the calculated route based on the current situation, or household appliances that not only manage their operation times to reduce peak power consumption, but also take into account external conditions (such as time of the day, temperature, humidity) in order to assign priority. One other interesting application might involve medical sensors that monitor health parameters and coordinate with the infusion pump to deliver the medication.

As a motivating example, consider the particular case of autonomous agents in charge of airport security. Suppose the agents have to check the airport surface, but the insufficiency of security resources prevents complete coverage and allows adversaries to exploit patterns in patrolling. Thus, the agents' walk should mimic a random behavior, otherwise its purpose will be defeated rather easily by observing the agents for a certain amount of time. Moreover, the agents want to coordinate their behavior in order to ensure the highest level of security, and over time all the terminals surface should be checked with the same probability. In this example, communication takes place between the agents, but they may also be assisted by a common source of information. In particular, this permits them to communicate as little as possible, so that if someone could intercept the messages exchanged by the agents, it would not be enough to guess their next move.

We now provide a more formal description of the above example. We consider the case of n agents. For simplicity, we model their actions as movements on a finite grid as in Figure 1.1, in which the agents perform random walks by taking actions left, right, up and down, represented respectively by 0, 1, 2, 3. Observe that when the agents reach the border of the graph not all the directions are still possible, and their walk is constrained by the nature of the perimeter. Not only the agents communicate to each other, we also assume that they have access to a source of common randomness such as a satellite time-stamp. Here, coordination is meant as a way to impose a known and fixed joint distribution of actions at all nodes in the network. Then, the goal is to ensure that the distribution of the actions taken by the agents after communication is close to a prescribed distribution, which in this example is the uniform distribution.

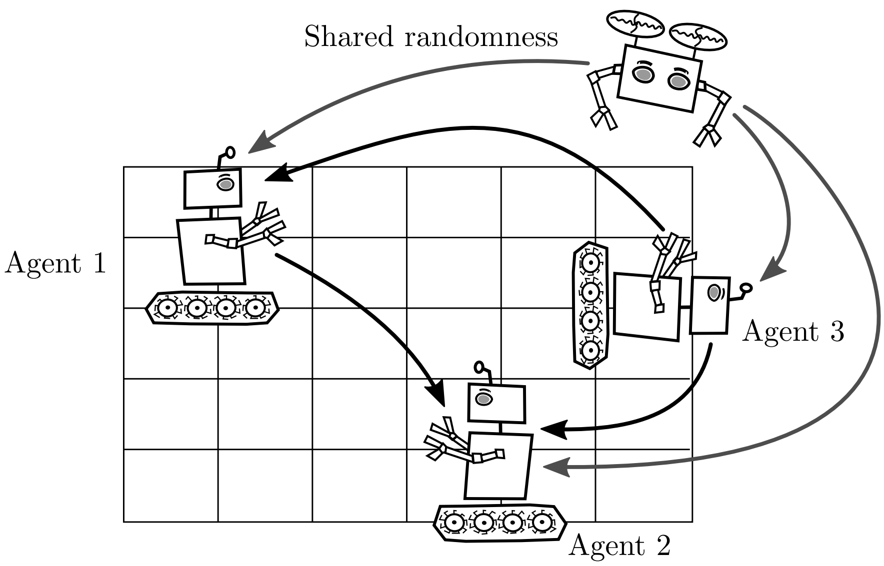

Figure 1.1: Autonomous agents perform random walks on a grid. They communicate with each other and share a source of common information

1.2 State of the art

A large variety of research addresses the topic of coordination: the information-theoretic formulation of the coordination problem that we consider in this thesis was put forward in [25], related to earlier work on quantum information theory [4, 65, 36, 68]. This framework also relates to the game-theoretic perspective on coordination [29] with applications, for instance, to power control [39]. Here, we outline the state of the art which is more significant for this work.

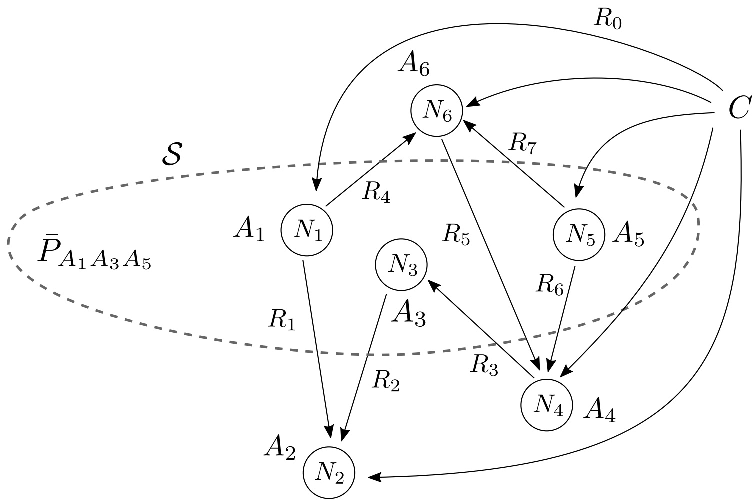

Figure 1.2: Example of a general network, where the actions A_1, A_3 and A_5 are chosen by nature and the actions A_2, A_4 and A_6 are generated through communication and common randomness C. The purpose of coordination is to establish which joint distributions $\bar{P}_{A_1A_2A_3A_4A_5A_6}$ can be achieved.

Information theoretic framework A general information-theoretic framework for coordination in networks was presented in [22, 25]. In [25], the authors consider a network represented by a graph with a certain number of nodes, as in Figure 1.2. Some of the nodes N_i, $i = 1, \dots, 6$, are connected by unidirectional communication links with rates R_j, $j = 1, \dots, 7$. A uniform source of common randomness C of rate R_0 is available to all nodes N_i, $i = 1, \dots, 6$. For every $i = 1, \dots, 6$, the node N_i performs an action A_i which belongs to a set $\mathcal{A}_i$ of possible actions. Some of these actions, A_1, A_3 and A_5, which belong to the set $\mathcal{S} \subseteq [\![1, 6]\!]$, are assigned by nature and behave according to the fixed distribution $\bar{P}_{A_1A_3A_5}$. On the other hand, the actions A_2, A_4 and A_6 are generated through communication between the nodes and through common randomness C that is available to all nodes. An interesting problem is to characterize which target joint distributions $\bar{P}_{A_1A_2A_3A_4A_5A_6}$ are achievable, in the sense that it is possible to induce, through the communication among the nodes, a joint distribution $P_{A_1A_2A_3A_4A_5A_6}$ that approximates $\bar{P}_{A_1A_2A_3A_4A_5A_6}$.

Observe that, if the set $\mathcal{S}$ is empty and common randomness is unexpensive, the problem becomes trivial. In fact, the nodes can agree on how they will behave and use common randomness to generate any distribution $\bar{P}_{A_2A_4A_5}$ [25].

Two notions of coordination have been proposed to measure how the induced joint distribution approximates the target distribution [25, 23]: *empirical coordination*, which requires the joint histogram of the devices' actions to approach a target distribution, and *strong coordination*, which requires the joint distribution of sequences of actions to converge to an i.i.d. target distribution, e.g., in variational distance.

A variety of works has focused on solving the coordination problem in different network settings based on both empirical and strong coordination. Coordination via a relay has been considered in [30, 10], while in [9, 66, 67] the authors study coordination over a line network. While most works on coordination involve an error-free communication link, in [24] the authors consider joint empirical coordination of signals and actions with a noisy communication channel. This setting was extended in [44] to include two-sided state information, and in [43] to include channel feedback available at the encoder.

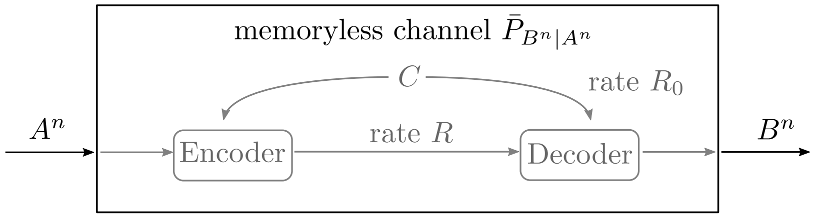

Figure 1.3: Synthesis of a memoryless channel by using communication and common randomness: it is possible to characterize the rate R and R_0 such that the conditional distribution induced by the code $P_{B^n|A^n}$ is close in total variational distance to the target channel $\bar{P}_{B^n|A^n}$.

Channel simulation Particularly relevant for this work, coordination through interactive communication can also be studied from the point of view of channel simulation [23]. In fact, since some of the actions are imposed by nature, achieving a certain i.i.d. joint distribution is equivalent to inducing the corresponding conditional distribution, which can be viewed as a discrete memoryless channel (DMC). The notion of strong coordination can be revisited as mimicking a DMC, and [23, 71] characterize the minimum rate of communication and of common randomness needed to achieve a joint distribution that is statistically indistinguishable (as measured by total variation) from the distribution induced by a memoryless channel. The channel simulation problem was further studied in [31], in which the authors simulate a noisy channel using an other noisy channel.

Quantum information theory These works on channel simulation are also related to earlier work on “Shannon’s reverse coding theorem” [4]. The idea of [4] is to to generalize Shannon’s coding theorem [63] to channels with quantum effects, and to prove its optimality by showing that a noisy channel can be simulated through a binary error-free link sending information at channel capacity. The construction based on channel simulation proposed in [4] has inspired the work on efficient compression of stochastic sources of probability distributions and mixed quantum states [68]. Independently, in [65] the authors consider channel synthesis with unlimited common randomness for the quantum data compression problem. The problem of compressing not only quantum states, but also of compressing the outcomes of quantum measurements, is studied in [36]. In [35, 36], the authors solve a rate distortion problem by proving the existence of a code with minimal amount of information exchanged, such that the joint empirical probability distribution of the signals is close to some desired distribution.

Game theory The topic of coordination has also been of interest in game theory, using the notion of *implementable* probability distribution [29], related to empirical coordination. In [29], the authors investigate a cooperative sender-receiver game by formulating a coding problem and by using information theoretic tools. The game theory setting differs from the information theory one because it considers several rounds of interactive communication: a repeated game takes place between the agents and at each stage the chosen actions may take into account past actions and the state of nature. The authors characterize the set of target joint probability distributions that are implementable by a choice of strategy of the agents; these distributions determine the expected utility of the game. Note that the concept of utility function is general and captures the different objectives of the coding process: it can be a distortion function for the source coding problem, a cost function for the channel coding problem or a payoff function for the players in a repeated game as in [29]. In [38, 39], the authors extend [29] by considering a noisy channel with an application to power control.

Strategical coordination The problem of the strategical coordination, at the intersection between game theory and information theory, considers encoder and decoder as players that have distinct objectives and choose their encoding and decoding strategies accordingly. This represents a significant difference with classical information theory, where transmitters are either allies that have a common goal, or they act as opponents. In contrast, in strategic communication the nodes disclose their information according to their own objectives. Repeated games involving distinct utility functions are studied in [3, 54, 28], without explicitly mentioning coordination. Related to these works, in [48–51] the authors investigate strategic coordination with a noisy channel.

State dependent networks The analysis in [29] emphasizes the need for updating information, and studies the communication dynamics in a model where the state of nature evolves over time. In [38, 39] power control is used to encode embedded data about the channel state information, and this method is further considered in [40, 41] that extends [39] to the symmetric scenario where each agent has access to state information. Joint empirical coordination in state dependent networks has been studied in [42, 44].

Coding theory All the previously cited works focused on establishing the fundamental limits of coordination, and their aim is to characterize the set of achievable behaviors and minimal communication rates to obtain coordination. Another class of works considers the design of practical schemes for coordination. In the field of coding theory, polar codes [1, 2] prove themselves to be particularly well suited to translate information theoretic properties such as coordination. Schemes for coordination based on polar codes have been proposed for empirical coordination in cascade networks [5], and for strong coordination with error free links [11, 15]. Polar codes for strong coordination of actions with noisy links have been designed in [56, 57].

1.3 Contributions and main results of this thesis

In this work, we address the problem of strong coordination in the two-node network of Figure 1.4 comprised of an i.i.d information source U^n with distribution $\bar{P}_U$, and a noisy channel parametrized by the conditional distribution $\bar{P}_{Y|X}$. Both nodes have access to a common source of randomness C of rate R_0. The encoder selects a signal $X^n = f_n(U^n, C)$, $f_n : \mathcal{U}^n \times [\![1, 2^{nR_0}]\!] \to \mathcal{X}^n$ as a stochastic function of the source and the common randomness.

The signal X^n is transmitted over the channel and the decoder selects an action $V^n = g_n(Y^n, C)$ as a stochastic function of the output of the channel Y^n and the common randomness.

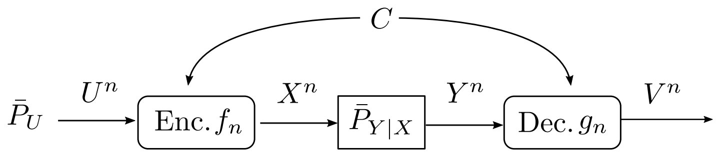

Figure 1.4: Coordination of signals and actions for a two-node network with a noisy channel.

For this setting, if the communication link is error-free, in [25, 23] the authors completely characterize both the empirical and the strong coordination region. Contrarily to the majority of works on coordination that have dealt with noise-free channels, we consider the more realistic case of coordination over noisy channels. This problem has already been considered for empirical coordination by [24, 46], and for strong coordination of actions in [31, 56, 57]. However, in this setting, the signals exchanged between the users can also be viewed as part of their observable behavior. Therefore we investigate the joint strong coordination of signals and actions. In this case, not only the signals sent from the encoder to the decoder allow to coordinate the actions, but additionally they have to be coordinated. In the setting of Figure 1.4 we derive an inner and an outer bound for the coordination region both with a non-causal encoder and a non-causal decoder and with a strictly causal encoder and a non-causal decoder.

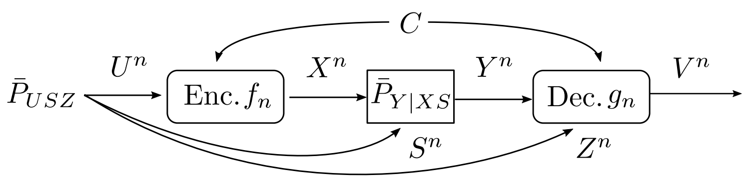

Figure 1.5: Coordination of signals and actions for a two-node network with a noisy channel with state and side information at the decoder.

This setting can be generalized to include situations where the environment changes over time such as state-dependent channels and and side/state information at the decoder, as in Figure 1.5. As in [42, 44], we introduce a random variable representing the effect of the environment and the side information available at the decoder, and we prove and inner and an outer bound for the strong coordination region when the encoder and decoder are both non-causal. In particular, the inner bound is proved by proposing a random binning scheme and a random coding scheme that have similar statistics.

As is the case for the best currently known bounds for empirical coordination [24, 44], our inner and outer bounds do not match, keeping the generality of our achievability scheme an open problem. However, we show that we successfully characterize the strong coordination region exactly in some special cases: i) when the channel is noiseless; ii) when the decoder is lossless; and iii) when the random variables of the channel are independent from the random variables of the source. In particular, the inner bound in all these cases is derived by specializing the random binning and random coding schemes to the specifics of every setting, suggesting that our achievability scheme is general enough.

Then, inspired by the binning technique using polar codes in [12], we propose an explicit polar coding scheme that achieves the inner bound for the coordination capacity region.

1.4 Organization of the thesis

The remainder of the document is organized as follows.

Chapter 2 introduces the notation and the information-theoretic framework of the coordination problem. We define *empirical* and *strong* coordination: empirical coordination captures an "average behavior" over multiple repeated actions of the devices; in contrast, strong coordination captures the behavior of time sequences. We state the main properties of the metrics of choice of strong coordination, K-L divergence and total variational distance. Furthermore, we introduce the random binning tools that we use to prove the achievability for strong coordination.

Chapter 3 considers the case of a non-causal encoder/decoder and derives an inner and an outer bound for the strong coordination region, first in a simple model in which there is no state and no side information, and then in the general case of a noisy channel with state and side information at the decoder.

Chapter 4 characterizes the strong coordination region for three special cases: i) perfect channel; ii) lossless decoder; and iii) independence between source and channel. The complete characterization of the strong coordination region, even if only in specific cases, allows us to derive some conclusions on the nature of coordination. In particular, we show that the separation principle does not hold for strong coordination. A byproduct of strong coordination is that it enforces some level of "security", in the sense of guaranteeing that the sequences of actions will be unpredictable to an outside observer beyond what is known about the target joint distribution of sequences. Consequently, at the end of the chapter we investigate the secrecy implications of strong coordination.

Chapter 5 introduces a general technique to turn achievability proofs based on random binning into explicit polar coding schemes, and it presents a polar coding scheme for the simpler setting where there is no state and no side information.

Chapter 6 considers strong coordination with a strictly causal encoder. We derive an inner and an outer bound for the coordination region, and we propose a polar coding scheme for this setting.

Conclusions and perspectives examines the implications of our results and proposes some ideas for further developments.

Some complementary material is provided in the Appendices, including definitions, classical information theory results, and additional proofs of technical lemmas and theorems. More precisely:

Appendix A consists in a summary of useful information theoretic definitions and properties, and proofs of preliminary results.

Appendix B details all the achievability and converse proofs omitted in Chapters 3, 4 and 6. In particular, it details the complete proofs of the inner bound for i) the general case of channel with state and side information; ii) the case of the strictly causal encoder.

Appendix C includes the missing proofs of Chapter 5. Moreover, it presents the polar coding achievability schemes for empirical coordination with both a non-causal encoder and a strictly causal encoder.

1.5 Publications

This work was partly presented in the following publications:

Journal paper (submitted)

- G. Cervia, L. Luzzi, M. Le Treust and M. R. Bloch, "Strong coordination of signals and actions over noisy channels with two-sided state information", submitted to *IEEE Transactions on Information Theory* (https://arxiv.org/abs/1801.10543v2)

International conferences

- G. Cervia, L. Luzzi, M. R. Bloch and M. Le Treust, " Polar coding for empirical coordination of signals and actions over noisy channels", in *IEEE Information Theory Workshop*, Cambridge (U.K.), September 2016
- G. Cervia, L. Luzzi, M. Le Treust and M. R. Bloch, " Strong coordination of signals and actions over noisy channels", in *IEEE International Symposium on Information Theory*, Aachen (Germany), June 2017
- G. Cervia, L. Luzzi, M. Le Treust and M. R. Bloch, "Strong coordination over noisy channels with strictly causal encoding", in *Proc. of Allerton Conference on Communication, Control and Computing*, Monticello (Illinois, U.S.A.), October 2018

National conferences

- G. Cervia, L. Luzzi, M. Le Treust and M. R. Bloch, "Polar codes for empirical coordination over noisy channels with strictly causal encoding", in *Colloque GRETSI*, Juan-les-Pins (France), September 2017

2 | Preliminaries

In this chapter, we give a general introduction to the coordination problem. First, Section 2.1 introduces the notation that we use throughout this document. The definitions of empirical and strong coordination are presented in Section 2.2. Then, since both the achievability and the converse proof rely on properties of the total variational distance and Kullback-Leibler divergence, these are presented in Section 2.3. Finally, a general achievability technique based on random binning is explained in Section 2.4. The proofs of the original results are in Appendix A.3.

2.1 Notation

We define the integer interval $[\![a, b]\!]$ as the set of integers between a and b. Given a random vector $X^n := (X_1, \dots, X_n)$, we note X^i the first i components of X^n, $X_{\sim i}$ the vector $(X_j)_{j \neq i}$, $j \in [\![1, n]\!]$, where the component X_i has been removed and $X[A]$ the vector $(X_j)_{j \in A}$, $A \subseteq [\![1, n]\!]$. Given two random vectors A and B, $A \perp B$ indicates that A and B are independent. We denote with Q_A the uniform distribution over $\mathcal{A}$, and with $Ber(p)$ the Bernoulli distribution of parameter $p \in [0, 1]$ which takes the value 1 with probability p, and the value 0 with probability $1 - p$. We define the following conditional distribution

$$\mathbb{1}_{\hat{A}|A}(a|a') := \begin{cases} 1 & \text{if } a = a', \\ 0 & \text{if } a \neq a'. \end{cases}$$

We use the notation $f(\varepsilon)$ to denote a function which tends to zero as ε does, and the notation $\delta(n)$ to denote a function which tends to zero exponentially as n goes to infinity. The *total variational distance* between two probability mass functions P and P' taking value in $\mathcal{A}$ is given by

$$\mathbb{V}(P, P') := \frac{1}{2} \|P - P'\|_{L^1} = \frac{1}{2} \sum_{a \in \mathcal{A}} |P(a) - P'(a)|.$$

The *Kullback-Leibler divergence* (or K-L divergence) between two probability mass functions P and P' is

$$\mathbb{D}(P \| P') := \sum_{a \in \mathcal{A}} P(a) \log \frac{P(a)}{P'(a)},$$

where the logarithm function $\log$ is assumed to be base 2 unless specified otherwise.

2.2 Empirical and strong coordination of actions

In a decentralized network of connected objects, devices communicate with each other while sensing or acting on their environment. It is essential that these devices, considered as autonomous decision-makers, cooperate and coordinate their actions. From an information theory perspective, rather than sending information from one point to another with a fidelity constraint, we are interested in quantifying the amount of communication needed to control the joint probability distribution of behavior among the nodes in the network.

Before presenting our setting, we try to clarify the purpose of coordination in the simplified scenario of Figure 2.1 studied in [25].

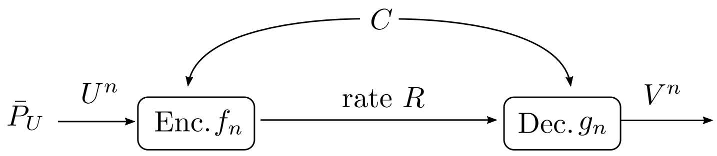

Figure 2.1: Coordination of the actions U^n and V^n for a two-node network with an error-free link of rate R.

In [25], the authors consider two nodes connected by a one-directional error-free link of rate R and sharing a common source of uniform randomness C of rate R_0. At time $i = 1, \dots, n$, the nodes perform the actions U_i and V_i respectively. The source sequence U^n is assigned by nature and behaves according to the fixed distribution $\bar{P}_{U^n}$. Then, the encoder generates a message M as a stochastic function of U^n and the common randomness C of rate R_0, that is available to all nodes. The message is sent through the error-free link of rate R and the sequence of actions V^n is generated as a function of the message M and of the common randomness C.

In [25] two different notions of coordination are proposed, empirical and strong coordination, both associated with a desired joint distribution of actions. *Empirical coordination* is achieved if the joint histogram of the actions in the network is close with high probability to the desired distribution $\bar{P}_{UV}$, and does not require common randomness.

Definition 2.1 - ***Achievability for empirical coordination [25, 22]*** *A distribution $\bar{P}_{UV}$ and a rate of communication R are achievable for empirical coordination if there exists a sequence (f_n, g_n) of encoders-decoders such that for all $\varepsilon > 0$, there exists $\bar{n}$ such that for all $n \geq \bar{n}$:*

$$\mathbb{P}\left\{\mathbb{V}\left(T_{U^nV^n}, \bar{P}_{UV}\right) > \varepsilon\right\} < \varepsilon$$

where

$$T_{U^nV^n}(u,v) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{(U_i,V_i) = (u,v)\}$$

is the joint histogram of the actions induced by the code (f_n, g_n).

Definition 2.2 - ***Empirical coordination region [25, 22]*** *The empirical coordination region $\mathcal{R}_e$ is the closure of the set of achievable pairs $(\bar{P}_{UV}, R)$.*

While empirical coordination is only interested in controlling the joint histogram of the actions, *strong coordination* deals instead with the joint probability distribution of the actions.

Definition 2.3 - ***Achievability for strong coordination [25, 22]*** *A sequence $(\bar{P}_{UV}, R, R_0)$ is achievable for strong coordination if there exists a sequence (f_n, g_n) of encoders-decoders with rate of common randomness R_0, such that*

$$\lim_{n\to\infty} \mathbb{V}\left(P_{U^nV^n}, \bar{P}_{UV}^{\otimes n}\right) = 0$$

where $P_{U^nV^n}$ is the joint distribution induced by the code (f_n, g_n).

Definition 2.4 - ***Strong coordination region [25, 22]*** *The strong coordination region $\mathcal{R}$ is the closure of the set of achievable $(\bar{P}_{UV}, R, R_0)$.*

Remark 2.5 - ***Convexity.*** While the empirical coordination region $\mathcal{R}_e$ is convex, the strong coordination capacity $\mathcal{R}$ region is not convex in general. Suppose for example that we consider a network with no communication and no common randomness. An arbitrary joint distribution is not achievable for strong coordination without communication or common randomness, but in [25] the authors prove that any extreme point in the probability simplex corresponds to a trivial distribution that is achievable. Hence, the convex combination of these distributions is not necessarily strongly achievable, in contrast with empirical coordination.

Remark 2.6 - ***Boundary issues.*** As in [25], we define the achievable region as the closure of the set of achievable rates and distributions. This definition allows to avoid boundary complications. For a thorough discussion on the boundaries of the achievable region when $\mathcal{R}$ is defined as the closure of the set of rates for a given distribution, see Appendix A.2.

Remark 2.7 - ***Empirical versus strong coordination.*** Whenever average behavior over time is the concern, the control of the empirical joint distribution is enough. However, strong coordination is to be preferred form a security standpoint. For example, in situations where an adversary is involved, it might be useful to make the sequence of actions appear impenetrable to an outside observer. Suppose an opponent observes the actions of the nodes and tries to anticipate and exploit patterns; in this case empirical coordination is not a constraint stringent enough to prevent the adversary to guess some information. On the other hand, suppose the adversary performs a statistical test to decide if the distribution detected, the distribution P induced by the code, is indistinguishable in total variational distance from the i.i.d. distribution $\bar{P}$ (hypothesis H_0). We denote α the probability of Type I error (rejecting H_0 when true) and β the probability of Type II error (accepting H_0 when wrong). The analysis of α and β has the Kullback-Leibler divergence and total variational distance as metrics of choice [6], linking hypothesis testing to strong coordination. In [52] it is proved that it is possible for the adversary to design blind tests ignoring his channel observations that achieve any pair (α, β) such that $\alpha + \beta = 1$, and that the adversary's optimal test satisfies

$$\alpha + \beta \geq 1 - \mathbb{V}(P, \bar{P}).$$

Therefore, by minimizing the total variational distance between the two distributions, we ensure that the adversary's best statistical test produces a trade-off between α and β that is not much better than that of a blind test.

Remark 2.8 - ***Actions assigned by nature.*** Observe that without the assumption that some actions are assigned by nature, the problem becomes trivial. Suppose that the two nodes can choose their actions and that common randomness is available at both nodes. In this case, no communication is required between the nodes and, if the nodes can agree ahead of time on how they will behave in the presence of common randomness (for example, a time stamp used as a seed for a random number generator), any conditional distribution $\bar{P}_{V^n|U^n}$ can be generated [22]. However, the problem becomes interesting when the actions of certain nodes are imposed by nature. Hence, since $\bar{P}_{U^n}$ is fixed, the joint distribution $P_{U^nV^n}$ depends only on the conditional distribution $P_{V^n|U^n}$. Then, we want to characterize the conditional distributions $P_{V^n|U^n}$ which are compatible with the network constraints.

Remark 2.9 - ***Communication is suboptimal.*** Of course, in the setting of Figure 2.1, a trivial solution to the coordination problem would be to have the first node communicate its randomized actions to the second node using the error-free link, which would require a rate of at least $H(U)$ bits per action. Then, the second node would just have to simulate a discrete memoryless channel $P_{V|U}$ using local randomness. However, it turns out that this strategy is an excessive use of communication. The empirical and strong coordination regions are characterized in [25]:

$$\mathcal{R}_{\text{Cuff,e}} := \left\{ (\bar{P}_{UV}, R) \;\middle|\; \begin{array}{l} \bar{P}_{UV} = \bar{P}_U \bar{P}_{V|U} \\ R \geq I(U;V) \end{array} \right\}, \tag{2.1}$$

$$\mathcal{R}_{\text{Cuff}} := \left\{ (\bar{P}_{UV}, R, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{UV} = \bar{P}_U \bar{P}_{V|U} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UWV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{V|W} \\ R \geq I(U;W) \\ R + R_0 \geq I(UV;W) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{V}| + 1 \end{array} \right\}. \tag{2.2}$$

Remark 2.10 - ***Relation to Wyner common information.*** Note that

$$R + R_0 \geq I(UV;W) \geq C(U;V),$$

where

$$C(U;V) := \min_{U-W-V} I(UV;W)$$

is Wyner common information. In absence of common randomness, the minimum communication rate R required to achieve strong coordination is Wyner's common information. This result is coherent with Wyner's intuition [69] that, in order to generate U^n and V^n separately as an i.i.d. source pair, they must share bits at a rate of at least the common information $C(U;V)$ of the joint distribution.

Example 2.11 The trade-off between R_0 and R characterized in $\mathcal{R}_{\text{Cuff}}$ (2.2) is better shown in an example. We consider the case of a source U generated according to $Ber(1/2)$, and a target distribution $\bar{P}_{UV}$ where V is an erasure with probability p_e and is equal to U otherwise. In [23, Appendix 1] the authors show that the optimal choice for the joint distribution $\bar{P}_{UWV}$ in (2.2) is the concatenation of two erasure channels $\bar{P}_{W|U}$ and $\bar{P}_{V|W}$.

To prove it, observe that for any $w \in \mathcal{W}$, by the Markov property $U - W - V$, we have

$$\bar{P}_{UV|W=w} = \bar{P}_{U|W=w}\bar{P}_{V|W=w}.$$

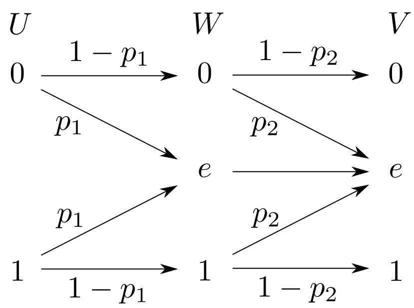

Figure 2.2: Optimal choice for W.

Then, since the events $\{(U, V) = (0, 1)\}$ and $\{(U, V) = (1, 0)\}$ have probability zero according to $\bar{P}_{UWV}$, we are left with the following possibilities:

- *Option A*: $V = 0$, then $U = 0$ with probability 1,
- *Option B*: $V = 1$, then $U = 1$ with probability 1,
- *Option C*: V is the erasure symbol and U is either 0 or 1.

These options lead to the following alternatives:

- *Option A*: if $\bar{P}_{UV|W=w}(u, 0) = \bar{P}_{U|W=w}(u)\bar{P}_{V|W=w}(0) > 0$, then $u = 0$ since:

$$\begin{aligned}\bar{P}_{V|W=w}(0) &= \bar{P}_{UV|W=w}(0, 0) + \bar{P}_{UV|W=w}(1, 0)\\ &= \bar{P}_{UV|W=w}(0, 0) = \bar{P}_{U|W=w}(0)\bar{P}_{V|W=w}(0)\end{aligned}$$

  So either $\bar{P}_{V|W=w}(0) = 0$, or $\bar{P}_{V|W=w}(0) > 0$ and $\bar{P}_{U|W=w}(0) = 1$.

- *Option B*: similarly, if $\bar{P}_{UV|W=w}(u, 1) = \bar{P}_{U|W=w}(u)\bar{P}_{V|W=w}(1) > 0$, then $u = 1$ since:

$$\begin{aligned}\bar{P}_{V|W=w}(1) &= \bar{P}_{UV|W=w}(0, 1) + \bar{P}_{UV|W=w}(1, 1)\\ &= \bar{P}_{UV|W=w}(1, 1) = \bar{P}_{U|W=w}(1)\bar{P}_{V|W=w}(1)\end{aligned}$$

  So either $\bar{P}_{V|W=w}(1) = 0$, or $\bar{P}_{V|W=w}(1) > 0$ and $\bar{P}_{U|W=w}(1) = 1$.

- *Option C*: alternatively, in this case the event $V \in \{0, 1\}$ has zero probability.

Based on the previous options, we define the sets:

$$\begin{aligned}\mathcal{W}_A &:= \{w \in \mathcal{W} \mid P_{U|W=w}, P_{V|W=w} \text{ verifies option A}\}\\ \mathcal{W}_B &:= \{w \in \mathcal{W} \mid P_{U|W=w}, P_{V|W=w} \text{ verifies option B}\}\\ \mathcal{W}_C &:= \{w \in \mathcal{W} \mid P_{U|W=w}, P_{V|W=w} \text{ verifies option C}\}\end{aligned} \tag{2.3}$$

Now, notice that it is sufficient to consider only distributions which have at most one value of $w \in \mathcal{W}$ for each of one of the three sets in (2.3). In fact, suppose that there exist two values of W that belong to the same set. Then, we define a function $f : \mathcal{W} \to f(\mathcal{W})$ that associates the same label to all the values $w \in W$ that belong to the same set. We denote $f(W) = W'$ and by the data processing inequality

$$I(U; W') \leq I(U; W) \quad \text{and} \quad I(UV; W') \leq I(UV; W).$$

Moreover $U - W' - V$ form a Markov chain because for every one of the three options, either U or V is deterministic. Then, we can consider the cardinality bound $|\mathcal{W}| = 3$, smaller than $|\mathcal{W}| \leq |\mathcal{U} \times \mathcal{V}| + 1 = 7$ and we take $\mathcal{W} = \{0, 1, e\}$.

Then, note that the distribution $\bar{P}_{UWV}$ is symmetric. To prove it, for every distribution $\bar{P}_{UWV}$ in the region, denote $\bar{P}_{\tilde{U}\tilde{W}\tilde{V}}$ the distribution for $\tilde{U} = 1 - U$, $\tilde{V} = 1 - V$, and $\tilde{W}$ the auxiliary random variable that belongs to $\mathcal{W}_A$ if W belongs to $\mathcal{W}_B$, belongs to $\mathcal{W}_B$ if W belongs to $\mathcal{W}_A$ and belongs to $\mathcal{W}_C$ if W does. Observe that $\bar{P}_{\tilde{U}\tilde{W}\tilde{V}}$ is still achievable and we define the symmetric distribution $\bar{P}'_{UWV}$ to be the average of $\bar{P}_{\tilde{U}\tilde{W}\tilde{V}}$ and P_{UWV}. Then by the convexity of the mutual information

$$I_{\bar{P}'}(U; W) \leq I_{\bar{P}}(U; W) \quad \text{and} \quad I_{\bar{P}'}(UV; W) \leq I_{\bar{P}}(UV; W)$$

and $\bar{P}'_{UWV}$ belongs to the achievable region.

Therefore, by symmetry and the cardinality bound, the optimal construction of $\bar{P}_{UWV}$ for synthesizing the symmetric binary erasure channel for symmetric inputs is a concatenation of two symmetric binary erasure channels as in Figure 2.2 with erasure probability p_1 and p_2 respectively. The probabilities p_1 and p_2 have to verify

$$\begin{aligned}(1 - p_1)(1 - p_2) &= 1 - p_e,\\ p_1 + (1 - p_1)p_2 &= p_e.\end{aligned}$$

This implies $p_2 \leq p_e$ since

$$p_1 = 1 - \frac{1 - p_e}{1 - p_2} \geq 0 \quad \Longrightarrow \quad \frac{1 - p_e}{1 - p_2} \leq 1 \quad \Longrightarrow \quad p_2 \leq p_e.$$

Moreover, $p_2 \leq 1/2$: suppose $p_e \geq 1/2$, we have

$$p_1 + (1 - p_1)p_2 = p_2 + (1 - p_2)p_1 \geq 1/2 \quad \Longrightarrow \quad p_1 \geq \frac{1/2 - p_2}{1 - p_2}$$

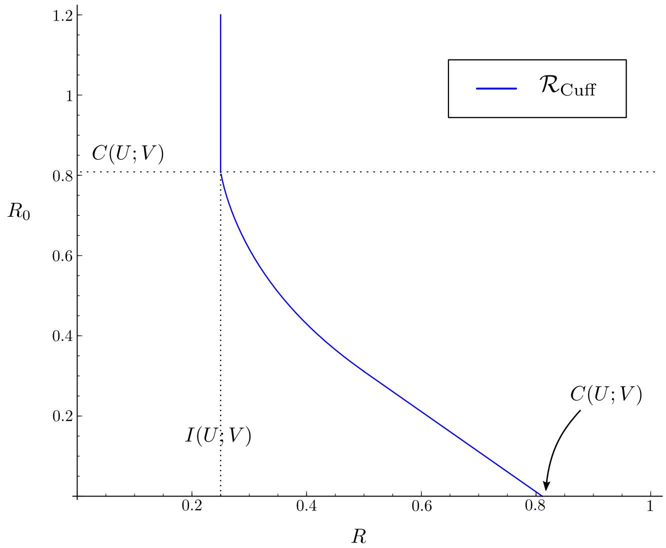

Figure 2.3: Boundary of the region $\mathcal{R}_{\text{Cuff}}$ for a binary erasure channel with erasure probability $p_e = 0.75$ and a Bernoulli-half input [23, 21].

where in particular since p_1 is greater or equal than 0, the right-hand side implies

$$\begin{cases} p_2 \leq 1/2 \text{ and } p_2 \geq 1 \\ 0 \leq p_2 \leq 1 \end{cases} \implies 0 \leq p_2 \leq 1/2.$$

Therefore we have

$$p_2 \in [0, \min\{1/2; p_e\}], \quad p_1 = 1 - \frac{1 - p_e}{1 - p_2}$$

and we obtain

$$I(U;W) = 1 - p_1, \qquad I(UV;W) = h(p_e) + (1 - p_1)(1 - h(p_2))$$

where h is the binary entropy function.

Figure 2.3 shows the boundary of $\mathcal{R}_{\text{Cuff}}$ for $p_e = 0.75$ and a Bernoulli-half input. The dotted bound $R \geq I(U;V)$ comes directly from combining $R \geq I(U;W)$ with the Markov chain $U - W - V$. At the other extreme, if $R_0 = 0$ in (2.2), $R \geq I(UV;W) \geq C(U;V)$ where $C(U;V)$ is Wyner common information [21]. Then, $p_e = 0.75$ and a Bernoulli-half input, the rate of communication R varies from $I(U;V) = 0.25$ bits and $C(U;V) = 0.811$ bits. Then, the rate of common randomness R_0 varies from 0 when R is maximal, $R = C(U;V) = 0.811$ bits, and it is at least $C(U;V) = 0.811$ bits, when R is minimal, $R = I(U;V) = 0.25$ bits.

Note that, if instead of strong coordination we choose a straightforward communication strategy in which the first agents communicates, this requires at least $H(U) = 1$ bit/action. Since best rate pair minimizing R is $(R, R_0) = (0.25, 0.811)$, if common randomness is inexpensive to obtain, strong coordination provides a significant reduction in communication rate to 0.25 bits/action.

2.3 Properties of the total variational distance and K-L divergence

In this section, we introduce some properties of the total variation and Kullback-Leibler metrics that will be used in our proofs.

The following result is a consequence of the triangle inequality.

Lemma 2.12 - ***Total variation: marginal [22, Lemma 16]*** *For any two joint distributions P_{AB} and $\hat{P}_{AB}$, the total variation distance between them can only be reduced when attention is restricted to P_A and $\hat{P}_A$. That is,*

$$\mathbb{V}(P_A, \hat{P}_A) \leq \mathbb{V}(P_{AB}, \hat{P}_{AB}).$$

More interestingly, this result shows in terms of the total variational distance what happens if the same conditional distribution is applied to two distributions defined on the same set:

Lemma 2.13 - ***Total variation: same channel [22, Lemma 17]*** *For any two random variables A and $\hat{A}$ on the same set $\mathcal{A}$ generated with distributions P_A and $\hat{P}_A$ respectively, the total variation distance between them remains the same when they are passed through the same channel:*

$$\mathbb{V}(P_A, \hat{P}_A) = \mathbb{V}(P_A P_{B|A}, \hat{P}_A P_{B|A}).$$

A similar result holds for Kullback-Leibler divergence and it comes directly from the chain rule for the divergence.

Lemma 2.14 - ***Kullback-Leibler divergence: same channel*** *For any two random variables A and $\hat{A}$ on the same set $\mathcal{A}$ generated with distributions P_A and $\hat{P}_A$ respectively, the Kullback-Leibler divergence between them remains the same when they are passed through the same channel:*

$$\mathbb{D}(P_A \| \hat{P}_A) = \mathbb{D}(P_A P_{B|A} \| \hat{P}_A P_{B|A}).$$

Finally, if two joint distributions of (A, B) are close in variational distance, then there exists a fixed symbol a such that the corresponding conditional distributions on $\mathcal{B}$ are close.

Lemma 2.15 - ***Total variation: fixed symbol [70, Lemma 4]*** *If $\mathbb{V}(P_A P_{B|A}, P'_A P'_{B|A}) = \varepsilon$, then there exists $a \in \mathcal{A}$ such that*

$$\mathbb{V}(P_{B|A=a}, P'_{B|A=a}) \leq 2\varepsilon.$$

Now, we state some results on almost i.i.d. sequences that we need in order to prove the outer bounds for the coordination region.

Lemma 2.16 - ***Entropy and timing information of nearly i.i.d. sequences [23, Lemma VI.3]*** *Let P_{A^n} be such that $\mathbb{V}(P_{A^n}, \bar{P}_A^{\otimes n}) \leq \varepsilon < 1/4$. Then we have*

$$\sum_{t=1}^{n} I(A_t; A^{t-1}) \leq 4\varepsilon \left(\log |\mathcal{A}| + \log \frac{1}{\varepsilon} \right)$$

and for any uniform random variable $T \in [\![1,n]\!]$ independent of A^n serving as time index

$$I(A_T;T) \leq 4\varepsilon\left(\log|\mathcal{A}| + \log\frac{1}{\varepsilon}\right).$$

The proofs of the following new results are in Appendix A.3. The following lemma is in the same spirit as Lemma 2.16. We state a slightly different version which is more convenient for our proofs.

Lemma 2.17 *Let P_{A^n} be such that $\mathbb{V}(P_{A^n}, \bar{P}_A^{\otimes n}) \leq \varepsilon$. Then we have*

$$\sum_{t=1}^{n} I(A_t; A_{\sim t}) \leq nf(\varepsilon).$$

In particular, if P_{AB} is such that $\mathbb{V}(P_{AB}, \bar{P}_A\bar{P}_B) \leq \varepsilon$, then $I(A;B) \leq f(\varepsilon)$.

The following is a consequence of Lemma 2.17.

Lemma 2.18 *Let $P_{A^nB^n}$ be such that $\mathbb{V}(P_{A^nB^n}, \bar{P}_{AB}^{\otimes n}) \leq \varepsilon$. Then we have*

$$\sum_{t=1}^{n} I(A_t; A^{t-1}B_{\sim t}|B_t) \leq nf(\varepsilon). \tag{2.4}$$

Let the variable T serve as a random time index, for any random variable C we have

$$H(C|B^n) \geq nI(A_T; CB_{\sim T}T|B_T) - nI(A_TB_T;T) - nf(\varepsilon). \tag{2.5}$$

2.4 Random binning

In strong coordination problems, the goal is to generate random variables whose joint distribution is close to a desired i.i.d. distribution in total variation distance. In traditional random coding, following the approach introduced by Shannon, a codebook

$$\mathcal{C} = \{X^n(m) \mid m = 1, \ldots, 2^{nR}\}$$

of rate R is constructed by generating the components of each codeword $X_i^n(m)$, $i = 1, \ldots, n$, $m = 1, \ldots, 2^{nR}$, independently at random according to the distribution P_X. Random binning is an alternative approach in which the set of all possible codewords $\mathcal{X}^n$ is randomly partitioned into 2^{nR} bins.

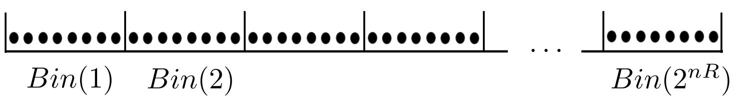

Figure 2.4: Random binning for $\mathcal{X}^n$.

Given $\mathbf{x} \in \mathcal{X}^n$, $\mathbf{x}$ is associated to an index $m \in [\![1, 2^{nR}]\!]$ drawn independently according to a uniform distribution on $[\![1, 2^{nR}]\!]$, via the encoder

$$\varphi_n : \mathcal{X}^n \to [\![1, 2^{nR}]\!].$$

Just to cite some examples, the achievability proofs of strong secrecy of the wiretap channel [19], the Slepian-Wolf Theorem [26, Theorem 10.1], and lossless source coding with a helper [26, Theorem 10.2] use random binning arguments.

While empirical coordination proofs are based on typicality arguments, strong coordination requires instead to prove results about approximation of probability distributions in the sense of vanishing total variation distance and random binning is particularly well suited for this purpose.

To prove the existence of coding schemes which induce a certain target joint distribution, we use a method introduced by [70] that, despite being rather general, has a fairly simple structure. We proceed in two steps. First, we define a random binning scheme for the n-letter target i.i.d. distribution. Then, we define a random coding scheme. We show that the joint distributions induced by the two schemes, the random binning and the random coding scheme, are close in total variational distance. To prove it, we need to impose rate conditions that satisfy two different objectives, here presented in two lemmas.

Properties of random binning The following lemma is a direct consequence of the Slepian-Wolf Theorem [64]. For each sequence $\mathbf{a} \in \mathcal{A}^n$, let $\varphi_n(\mathbf{a})$ be a bin index drawn independently according to a uniform distribution on $[\![1, 2^{nR}]\!]$. Then, the following result allows a decoder to recover the value of a random variable from its binning index with high probability, provided that the rate R of the binning is large enough.

Lemma 2.19 - ***Source coding with side information at the decoder [64]*** *Given a discrete memoryless source (A^n, B^n), where B^n is side information available at the decoder, let $\varphi_n : \mathcal{A}^n \to [\![1, 2^{nR}]\!]$ be a uniform random binning of A^n, and let $C := \varphi_n(A^n)$. Then if $R > H(A|B)$, the decoder can recover A^n from C and B^n with:*

$$\mathbb{E}_{\varphi_n}[\mathbb{P}\{\hat{A}^n \neq A^n\}] \leq \delta(n).$$

The second objective consists of ensuring that the binning is almost uniform and almost independent from the source, so that the random binning scheme and the random coding scheme generate joint distributions that have the same statistics. Given two correlated random variables A^n and B^n, the following result allows to obtain a binning for B^n which is almost independent of A^n, provided that the rate is small enough. The lemma is inspired by the discussion in [58, Section III.A], and proved in Appendix A.3.

Lemma 2.20 - ***Channel randomness extraction for discrete memoryless sources and channels [7]*** *Let A^n with distribution P_{A^n} be a discrete memoryless source and $P_{B^n|A^n}$ a discrete memoryless channel. Let $\varphi_n : \mathcal{B}^n \to [\![1, 2^{nR}]\!]$ be a uniform random binning of B^n, and let $K := \varphi_n(B^n)$.*

Then if $R \leq H(B|A)$*, there exists a constant* $\alpha > 0$ *such that*

$$\mathbb{E}_{\varphi_n}[\mathbb{D}(P_{A^n K} \| P_{A^n} Q_K)] \leq 2^{-\alpha n}. \tag{2.6}$$

Channel coding example How to relate a random coding scheme with a random binning one is better shown with the channel coding example [70], as an alternative to Shannon's classical random coding proof (see [63] and [18, Theorem 7.7.1]).

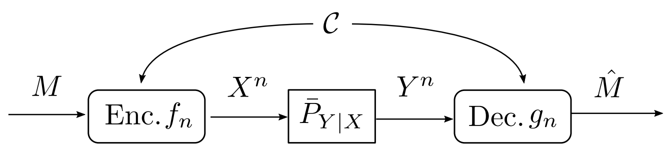

Figure 2.5: Point to point communication system: a transmitter wants to reliably send a uniform message M of rate R over a discrete memoryless channel $\bar{P}_{X|Y}$.

Suppose we are in the setting of Figure 2.5. The encoder wants to send a uniform message M of rate R over a discrete memoryless channel $\bar{P}_{X|Y}$, and encoder and decoder share a source of uniform randomness of rate R_0. In Shannon's classical proof, the codewords are generated independently according to an i.i.d. distribution $\bar{P}_{X^n}$ and the codebook

$$\mathcal{C} = \{X^n(m) \mid m = 1, \ldots, 2^{nR}\}$$

is known at both the encoder and the decoder. The encoder sends $X^n(M, \mathcal{C})$ over the channel, and at the output of the channel, the decoder estimates the message sent through its knowledge of the codebook and of Y^n. Note that the codebook $\mathcal{C}$ here plays the part of the common randomness and that the codebook and the message have to be independent of each other. Finally, the decoder has to be able to estimate correctly the message M:

$$p_e := \mathbb{P}\{\hat{M} \neq M\} \to 0.$$

Shannon's classical proof shows that the error probability, averaged over the random codebooks, is small if $R < I(X;Y)$. Here, we focus on the joint distribution induced by the code

$$P_{M\mathcal{C}X^nY^n} = Q_M Q_{\mathcal{C}} P^{\text{RC}}_{X^n|M\mathcal{C}} \bar{P}_{Y^n|X^n}$$

where we denote with Q_A the uniform distribution over $\mathcal{A}$, and we show that we can use random binning to define a stochastic encoder $P^{\text{RB}}_{X^n|M\mathcal{C}}$ which also guarantees vanishing error probability.

We consider X^n generated i.i.d. according to $\bar{P}_{X^n} := \bar{P}_X^{\otimes n}$ and we define the random binning scheme as follows. All symbols $\mathbf{x} \in \mathcal{X}^n$ are assigned indices through the following two binning functions as in Figure 2.6, hence defining random variables $\mathcal{C}$ and M:

- first binning $\mathcal{C} = \varphi_1(X^n)$, representing the codebook, where $\varphi_1 : \mathcal{X}^n \to [\![1, 2^{nR_0}]\!]$ is an encoder which maps each sequence of $\mathcal{X}^n$ uniformly and independently to the set $[\![1, 2^{nR_0}]\!]$;

- second binning $M = \varphi_2(X^n)$, representing the message, where $\varphi_2 : \mathcal{X}^n \to [\![1, 2^{nR}]\!]$ is an encoder which maps each sequence of $\mathcal{X}^n$ uniformly and independently to the set $[\![1, 2^{nR}]\!]$.

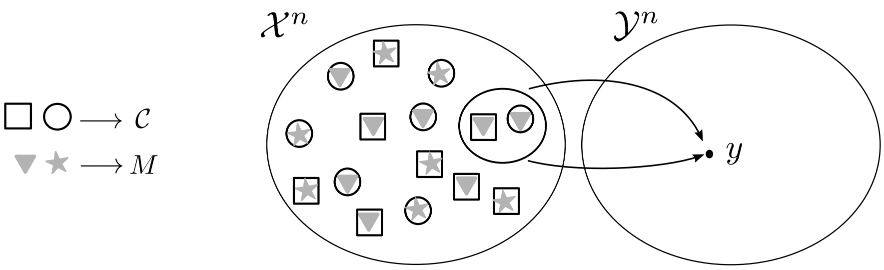

Figure 2.6: The square and the circle represent the possible outputs $\mathcal{C}$ of the first binning and the star and the triangle the outputs M of the second binning. Given $\mathbf{y}$ and the realization of $\mathcal{C}$, it is possible to recover $\hat{X}^n$ and thus $\hat{M}$.

We impose the following rate conditions:

- $R_0 > H(X|Y)$, so that the decoder can reconstruct X^n reliably from the output of the channel Y^n and the codebook $\mathcal{C}$ using a Slepian-Wolf decoder via the conditional distribution $P^{\text{SW}}_{\hat{X}^n|\mathcal{C}Y^n}$ by applying Lemma 2.19 to $A^n = X^n$, $B^n = Y^n$, $C = \mathcal{C}$. Therefore the message $\hat{M} = \varphi_2(\hat{X}^n)$ is reconstructed reliably;
- $R + R_0 < H(X)$ so that the message M is almost independent of the codebook $\mathcal{C}$ and they are almost uniform by applying Lemma 2.20 to $B^n = X^n$, $A^n = \emptyset$, $K = (M, \mathcal{C})$.

Then, the random binning induces a joint distribution

$$P^{\text{RB}} := \bar{P}_{M\mathcal{C}X^nY^n\hat{X}^n\hat{M}} = \bar{P}_{X^n}\bar{P}_{Y^n|X^n}P_{\mathcal{C}|X^n}P_{M|X^n}P^{\text{SW}}_{\hat{X}^n|\mathcal{C}Y^n}P_{\hat{M}|\hat{X}^n}.$$

Now, we *invert* the binning and consider the well-defined distribution $P^{\text{RB}}_{X^n|M\mathcal{C}}$. Note that for every $c \in [\![1, 2^{nR_0}]\!]$, we define a codebook $\mathcal{C}_c = \{\mathbf{x} \mid \varphi_1(\mathbf{x}) = c\}$. We associate to the codebook $\mathcal{C}_c$ a stochastic encoder, which maps any message $m \in [\![1, 2^{nR}]\!]$ into a codeword of $\mathcal{C}_c$ as follows:

$$f_n^c : [\![1, 2^{nR}]\!] \to \mathcal{C}_c, \quad m \mapsto \mathbf{x}$$

where $\mathbf{x} \in \mathcal{X}^n$ is generated according to $P^{\text{RB}}_{X^n|M=m,\mathcal{C}=c}$. As in Shannon's traditional approach, the decoder g_n estimates the message sent through its knowledge of the codebook and of Y^n: it reconstructs $\hat{X}^n$ via the conditional distribution $P^{\text{SW}}_{\hat{X}^n|\mathcal{C}Y^n}$ and $\hat{M} = \varphi_2(\hat{X}^n)$. This induces a joint distribution:

$$P^{\text{RC}} := P_{M\mathcal{C}X^nY^n\hat{X}^n\hat{M}} = Q_M Q_{\mathcal{C}} P^{\text{RB}}_{X^n|M\mathcal{C}}\bar{P}_{Y^n|X^n}P^{\text{SW}}_{\hat{X}^n|\mathcal{C}Y^n}P_{\hat{M}|\hat{X}^n}.$$

Note that we have imposed rate conditions such that there exists at least a pair of binnings $(\varphi_1^*, \varphi_2^*)$ that verifies:

$$\mathbb{P}_{P^{\text{RB},(\varphi_1^*,\varphi_2^*)}}\{\hat{X}^n \neq X^n\} \leq \delta(n), \tag{2.7}$$

$$\mathbb{D}(P_{MC}^{(\varphi_1^*,\varphi_2^*)} \| Q_M Q_{\mathcal{C}}) \leq 2^{-\alpha n}. \tag{2.8}$$

Then, by (2.7) the probability of error for the random binning scheme when estimating X^n tends to zero when n tends to infinity, and therefore

$$p_e' := \mathbb{P}_{P^{\text{RB},(\varphi_1^*,\varphi_2^*)}}\{M^n \neq M\} \to 0.$$

The total variational distance between P^{RC} and $P^{\text{RB},(\varphi_1^*,\varphi_2^*)}$ is

$$\begin{aligned} \mathbb{V}(P^{\text{RC}}, P^{\text{RB},(\varphi_1^*,\varphi_2^*)}) &= \mathbb{V}(P^{\text{RC}}_{M\mathcal{C}X^nY^n\hat{X}^n\hat{M}}, P^{\text{RB},(\varphi_1^*,\varphi_2^*)}_{M\mathcal{C}X^nY^n\hat{X}^n\hat{M}}) \\ &\overset{(a)}{=} \mathbb{V}(P^{\text{RC}}_{M\mathcal{C}X^n}, P^{\text{RB},(\varphi_1^*,\varphi_2^*)}_{M\mathcal{C}X^n}) \\ &\overset{(b)}{=} \mathbb{V}(P^{\text{RC}}_{M\mathcal{C}}, P^{\text{RB},(\varphi_1^*,\varphi_2^*)}_{M\mathcal{C}}) \\ &\overset{(c)}{=} \mathbb{V}(Q_M Q_{\mathcal{C}}, P^{\text{RB},(\varphi_1^*,\varphi_2^*)}_{M\mathcal{C}}) \leq \delta(n). \end{aligned} \tag{2.9}$$

where (a) and (b) follow from Lemma 2.13 because $\bar{P}_{Y^n|X^n}$, $P^{\text{SW}}_{\hat{X}^n|\mathcal{C}Y^n}$, and $P_{\hat{M}|\hat{X}^n}$ are the same in both distributions and by definition of the encoder, and (c) comes from (2.8) and Pinsker's inequality.

Note that, by imposing the rate conditions

$$R > H(X) - H(X|Y) = I(X;Y),$$

we have proved that the random binning distribution has vanishing probability of error p_e' and it is indistinguishable in total variational distance from the random coding distribution. We prove that these two facts imply that the probability of error $p_e := \mathbb{P}_{P^{\text{RC}}}\{\hat{M} \neq M\}$ vanishes. In fact, we have

$$\begin{aligned} \mathbb{P}_{P^{\text{RC}}}\{\hat{X}^n \neq X^n\} &= \sum_{m,c,\mathbf{x},\mathbf{y}} P^{\text{RC}}_{M\mathcal{C}X^n}(m,c,\mathbf{x},\mathbf{y}) \mathbf{1}\{\hat{\mathbf{x}} \neq \mathbf{x}\} \\ &= \sum_{m,c,\mathbf{x},\mathbf{y}} \left(P^{\text{RC}}_{M\mathcal{C}X^n}(m,c,\mathbf{x},\mathbf{y}) - P^{\text{RB},(\varphi_1^*,\varphi_2^*)}_{M\mathcal{C}X^n}(m,c,\mathbf{x},\mathbf{y}) \right) \mathbf{1}\{\hat{\mathbf{x}} \neq \mathbf{x}\} \\ &\quad + \sum_{m,c,\mathbf{x},\mathbf{y}} P^{\text{RB},(\varphi_1^*,\varphi_2^*)}_{M\mathcal{C}X^n}(m,c,\mathbf{x},\mathbf{y}) \mathbf{1}\{\hat{\mathbf{x}} \neq \mathbf{x}\} \\ &\leq \mathbb{V}(P^{\text{RC}}, P^{\text{RB},(\varphi_1^*,\varphi_2^*)}) + \mathbb{P}_{P^{\text{RB},(\varphi_1^*,\varphi_2^*)}}\{\hat{X}^n \neq X^n\} \\ &\overset{(d)}{\leq} 2\delta(n) \end{aligned}$$

where (d) follows from (2.7) and (2.9). Hence, p_e vanishes when n tends to infinity.

To complete the proof, observe that this is slightly different from using a channel code, because for every channel use we are choosing a codebook uniformly at random in the set of codebooks $\{\mathcal{C}_c \mid c \in [\![1, 2^{nR_0}]\!]\}$, instead of using the same codebook at all times.

3 | Strong coordination of signals and actions over noisy channels

In Section 2.2 we have introduced the coordination problem in a setting that involves an error-free line of communication between the agents. However, seeing that real-life communication is usually noisy, this assumption is quite optimistic and it is more reasonable to consider a noisy channel between the agents instead. Therefore, from now on we focus the more realistic scenario of a two node network comprised of an information source and a noisy channel given by nature. Consequently, the study of coordination has to involve the statistics of the channel.

In [31, 56, 57] the authors consider the strong coordination of the actions of the nodes when there is a noisy channel between the encoder and the decoder. In this thesis, we adopt a more general point of view, noting that the input and output signals exchanged over the noisy link are part of what can be observed about the behavior of the nodes. Therefore, we want to coordinate these signals in addition to their actions. This point of view has already been considered for empirical coordination in [24, 44, 46].

The sequence of both signals and actions should follow a known and fixed joint distribution, and this scenario presents two conflicting goals: the encoder needs to convey a message to the decoder to coordinate the actions, while simultaneously coordinating the signals coding the message.

Example: insider information We can clarify the interest in this joint coordination with an example. We can imagine that the agents are two stockbrokers who work for two competing companies, and we call them Agent 1 and Agent 2. Agent 1 has access to a source of information on future market developments: he sees the symbol 0 if it is a good moment to buy, and the symbol 1 if it is a good moment to sell. Agent 2 can choose between two possible actions in $\{0, 1\}$, either to buy, which corresponds to 0, or to sell, which corresponds to 1. Although they can profit from helping each other, they want to keep their cooperation secret. Both agents have access to a source of common randomness. The informed agent, Agent 1, wants to share part of this knowledge by sending a signal to Agent 2, but, since the line of communication that the agents can use is noisy, Agent 2 has to use both the source of common randomness and his observation of the output of the channel to reconstruct the information on the action to take. The two agents want the signal distributions to be statistically indistinguishable from i.i.d., so that an outside observer working for a competing company, and without access to the common randomness, would not be able to prove that an exchange of information had taken place.

As the example above suggests, the joint strong coordination of signals and actions with a noisy link is particularly interesting if security is required: if for example we require the actions of the agents to appear independent of the communication, a malicious eavesdropper who observes the output of the channel, cannot infer anything about the source and the reconstruction without having access to the source of common randomness [62].

Throughout this chapter, we consider the case in which both the encoder and the decoder are non-causal. We start with the simple setting of a two-node network comprised of an information source and a noisy channel in Section 3.1. Then, as in [42, 44, 40], we introduce a random state capturing the effect of the environment, to model actions and channels that change with external factors, and in Section 3.2 we consider a general setting in which state information and side information about the source may or may not be available at the decoder.

For these settings, we derive an inner and an outer bound for the strong coordination region.

We design an achievability proof by developing a joint source-channel scheme in which an auxiliary codebook allows us to simultaneously coordinate signals and actions. The random binning method used to prove the inner bound is quite general and can be generalized from the simpler setting of Section 3.1 to the one involving state and side information of Section 3.2. On the other hand, when state and side information are considered, more caution is required for the outer bound.

As for the best-known bounds for empirical coordination [24, 44], our bounds do not match and the optimality of our general achievability scheme remains an open question. Despite that, we will show that the joint source-channel coding scheme proposed in this chapter is optimal in some special cases, detailed in Chapter 4. Of course, one other possible approach would be to treat the coordination of the signals separately from the coordination of the actions. However, it is not clear a priori whether the concatenation of channel coordination and source coordination is optimal. In fact, this would mean that the separation principle holds, which later we prove being false in Section 4.4.

3.1 Two-node network

Now, we provide the mathematical description of this setting, depicted in Figure 4.5. Two agents, the encoder and the decoder, wish to coordinate their behaviors: the stochastic actions and the signals of the agents should follow a known and fixed joint distribution.

We suppose that the encoder and the decoder have access to a shared source of uniform randomness $C \in [\![1, 2^{nR_0}]\!]$. Let $U^n \in \mathcal{U}^n$ be an i.i.d. source with distribution $\bar{P}_U$. The encoder observes the sequence $U^n \in \mathcal{U}^n$ and selects a signal $X^n = f_n(U^n, C)$, $f_n : \mathcal{U}^n \times [\![1, 2^{nR_0}]\!] \to \mathcal{X}^n$. The signal X^n is transmitted over a discrete memoryless channel parametrized by the conditional distribution $\bar{P}_{Y|X}$. Upon observing Y^n and common randomness C, the decoder selects an action $V^n = g_n(Y^n, C)$, where $g_n : \mathcal{Y}^n \times [\![1, 2^{nR_0}]\!] \to \mathcal{V}^n$ is a stochastic map. For block length n, the pair (f_n, g_n) constitutes a code.

Here, we extend Definition 2.3 and Definition 2.4 to the noisy link setting where the signals should be coordinated together with the actions.

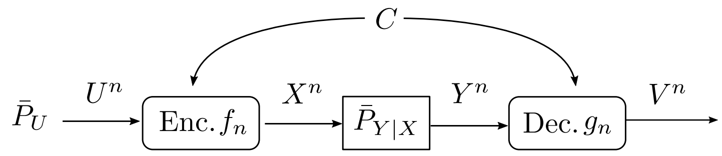

Figure 3.1: Coordination of signals and actions for a two-node network with a noisy channel with non-causal encoder and decoder.

Definition 3.1 - ***Achievability for strong coordination [25, 22]*** *A pair $(\bar{P}_{UXYV}, R_0)$ is achievable for strong coordination if there exists a sequence (f_n, g_n) of encoders-decoders with rate of common randomness R_0, such that*

$$\lim_{n\to\infty} \mathbb{V}\left(P_{U^nX^nY^nV^n}, \bar{P}_{UXYV}^{\otimes n}\right) = 0,$$

where $P_{U^nX^nY^nV^n}$ is the joint distribution induced by the code.

Definition 3.2 - ***Strong coordination region [25, 22]*** *The strong coordination region $\mathcal{R}$ is the closure of the set of achievable pairs $(\bar{P}_{UXYV}, R_0)$.*

Our first result is an inner and outer bound for the strong coordination region $\mathcal{R}$.

Theorem 3.3 *Let $\bar{P}_U$ and $\bar{P}_{Y|X}$ be the given source and channel parameters, then $\mathcal{R}_{in} \subseteq \mathcal{R} \subseteq \mathcal{R}_{out}$ where:*

$$\mathcal{R}_{\text{in}} := \left\{ (\bar{P}_{UXYV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{UXYV} = \bar{P}_U \bar{P}_{X|U} \bar{P}_{Y|X} \bar{P}_{V|UXY} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UXYWV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{X|UW} \bar{P}_{Y|X} \bar{P}_{V|WY} \\ I(W;U) \leq I(W;Y) \\ R_0 \geq I(W;UXV|Y) \end{array} \right\}, \tag{3.1}$$

$$\mathcal{R}_{\text{out}} := \left\{ (\bar{P}_{UXYV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{UXYV} = \bar{P}_U \bar{P}_{X|U} \bar{P}_{Y|X} \bar{P}_{V|UXY} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UXYWV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{X|UW} \bar{P}_{Y|X} \bar{P}_{V|WY} \\ I(W;U) \leq I(X;Y) \\ R_0 \geq I(W;UXV|Y) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| + 4 \end{array} \right\}. \tag{3.2}$$

Remark 3.4 - ***Markov chain decomposition.*** Observe that the decomposition of the joint distributions $\bar{P}_{UXYV}$ and $\bar{P}_{UWXYV}$ is equivalently characterized in terms of Markov chains:

$$Y - X - U, \qquad \begin{cases} Y - X - (U, W), \\ V - (Y, W) - (X, U). \end{cases} \tag{3.3}$$

Remark 3.5 - ***Relation to conditional common information.*** Observe that bound on the rate of common randomness is

$$R_0 \geq I(W;UXV|Y) \geq C(UX;V|Y),$$

where $C(UX;V|Y) := \min_{V-(W,Y)-(U,X)} I(UXV;W|Y)$ is the conditional common information. In [37] the authors prove that, given two terminals that, with the same side information Y^n, generate random vector A^n and B^n respectively, the minimum rate of common randomness required to coordinate the triple (A^n, B^n, Y^n) according to a target distribution is the conditional common information $C(A;B|Y)$. Here, we coordinate $(A^n, B^n, Y^n) = (U^n, X^n, V^n, Y^n)$, and we obtain the same lower bound on R_0 with $A^n = (U^n, X^n)$, $B^n = V^n$ and Y^n, but in our setting Y^n is side information only for the decoder.

Comparison with empirical coordination The empirical coordination region for the setting of Figure 3.1 was investigated in [24, Theorem 1], in which the authors derived the following inner bound:

$$\mathcal{R}_{\text{e,in}} := \left\{ (\bar{P}_{UXYV}) \;\middle|\; \begin{array}{l} \bar{P}_{UXYV} = \bar{P}_U \bar{P}_{X|U} \bar{P}_{Y|X} \bar{P}_{V|UXY} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UXYWV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{X|UW} \bar{P}_{Y|X} \bar{P}_{V|WY} \\ I(W;U) \leq I(W;Y) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| + 4 \end{array} \right\}. \tag{3.4}$$

Note that the information constraint $I(W;U) \leq I(W;Y)$ and the decomposition of the joint probability distribution $\bar{P}_U \bar{P}_{W|U} \bar{P}_{X|UW} \bar{P}_{Y|X} \bar{P}_{V|WY}$ are the same for $\mathcal{R}_{\text{e,in}}$ and $\mathcal{R}_{\text{in}}$. The main difference is that strong coordination requires a positive rate of common randomness $R_0 > I(W;UXV|Y)$. This is consistent with the conjecture, stated in [25], that with enough common randomness the strong coordination capacity region is the same as the empirical coordination capacity region for any specific network setting.

3.1.1 Inner bound

The achievability proof of Theorem 3.3 uses the same techniques as in [31] inspired by [70]. As anticipated in Section 2.4, the key idea of the proof is to define two coding schemes. First, we define a random binning scheme for the target i.i.d. distribution. Then, we define a random coding scheme and we prove that it has almost the same statistics as the random binning scheme.

Random binning scheme Assume that the sequences U^n, X^n, W^n, Y^n and V^n are jointly i.i.d. with distribution

$$\bar{P}_{U^n} \bar{P}_{W^n|U^n} \bar{P}_{X^n|W^nU^n} \bar{P}_{Y^n|X^n} \bar{P}_{V^n|W^nY^n}, \tag{3.5}$$

where we use the general notation $\bar{P}_{A^n} := \bar{P}_A^{\otimes n}$ for n-letter target distributions. We consider two uniform random binnings for W^n:

- first binning $C = \varphi_1(W^n)$, where $\varphi_1 : \mathcal{W}^n \to [\![1, 2^{nR_0}]\!]$ maps each sequence of $\mathcal{W}^n$ uniformly and independently to the set $[\![1, 2^{nR_0}]\!]$;

- second binning $F = \varphi_2(W^n)$, where $\varphi_2 : \mathcal{W}^n \to [\![1, 2^{nR}]\!]$ maps each sequence of $\mathcal{W}^n$ uniformly and independently to the set $[\![1, 2^{nR}]\!]$.

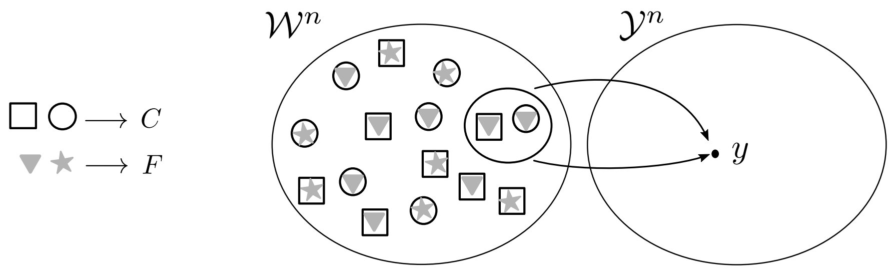

Figure 3.2: The square and the circle represent the possible outputs C of the first binning and the star and the triangle the outputs F of the second binning. Given $\mathbf{y}$ and the realizations of C and F, it is possible to recover $\mathbf{w}$.

Note that if $R + R_0 > H(W|Y)$, by Lemma 2.19, it is possible to recover W^n from Y^n and (C, F) with high probability using a Slepian-Wolf decoder via the conditional distribution $P^{\text{SW}}_{\hat{W}^n|CFY^n}$. This defines a joint distribution:

$$P^{\text{RB}} := \bar{P}_{U^n} \bar{P}_{W^n|U^n} \bar{P}_{X^n|W^nU^n} \bar{P}_{C|W^n} \bar{P}_{F|W^n} \bar{P}_{Y^n|X^n} \bar{P}_{V^n|W^nY^n} P^{\text{SW}}_{\hat{W}^n|CFY^n}.$$

In particular, $P^{\text{RB}}_{W^n|CFU^n}$ is well defined.

Random coding scheme In this section we follow the approach in [70, Section IV.E]. Suppose that in the setting of Figure 3.3, encoder and decoder have access not only to common randomness C but also to extra randomness F, where C is generated uniformly at random in $[\![1, 2^{nR_0}]\!]$ with distribution Q_C and F is generated uniformly at random in $[\![1, 2^{nR}]\!]$ with distribution Q_F independently of C. Then, the encoder generates W^n according to $P^{\text{RB}}_{W^n|CFU^n}$ defined above and X^n according to $\bar{P}_{X^n|U^nW^n}$. The encoder sends X^n through the channel. The decoder obtains Y^n and (C, F) and reconstructs W^n via the conditional distribution $P^{\text{SW}}_{\hat{W}^n|CFY^n}$. The decoder then generates V^n letter by letter according to the distribution

$$P^{\text{RC}}_{V^n|\hat{W}^nY^n}(\hat{\mathbf{v}}|\hat{\mathbf{w}}, \mathbf{y}) = \bar{P}_{V^n|W^nY^n}(\hat{\mathbf{v}}|\hat{\mathbf{w}}, \mathbf{y}), \tag{3.6}$$

where $\hat{\mathbf{w}}$ is the output of the Slepian-Wolf decoder. This defines a joint distribution:

$$P^{\text{RC}} := Q_C Q_F P^{\text{RC}}_{U^n} \bar{P}_{W^n|CFU^n} \bar{P}_{X^n|W^nU^n} \bar{P}_{Y^n|X^n} P^{\text{SW}}_{\hat{W}^n|CFY^n} P^{\text{RC}}_{V^n|\hat{W}^nY^n}.$$

Strong coordination of $(U^n, X^n, W^n, Y^n, V^n)$ We want to show that the distribution P^{RB} is achievable for strong coordination:

$$\lim_{n\to\infty} \mathbb{V}(P^{\text{RB}}_{U^nX^nW^n\hat{W}^nY^nV^n}, P^{\text{RC}}_{U^nX^nW^n\hat{W}^nY^nV^n}) = 0. \tag{3.7}$$

Note that

$$\begin{aligned}\mathbb{D}(P^{\text{RB}} \| P^{\text{RC}}) &\overset{(a)}{=} \mathbb{D}(\bar{P}_{U^n} \bar{P}_{W^n|U^n} \bar{P}_{C|W^n} \bar{P}_{F|W^n} \| Q_C Q_F P^{\text{RC}}_{U^n} P^{\text{RB}}_{W^n|CFU^n}) \\ &\overset{(b)}{=} \mathbb{D}(\bar{P}_{U^n CF} \| P^{\text{RC}}_{U^n} Q_C Q_F)\end{aligned} \tag{3.8}$$

where (a) comes from Lemma 2.14 because $\bar{P}_{Y^n|X^n}$ and $P^{\text{SW}}_{\hat{W}^n|CFY^n}$ are the same in both distributions. Note that (b) follows from Lemma 2.14 as well, since W^n is generated according to $P^{\text{RB}}_{W^n|CFU^n}$. Then if $R_0 + R < H(W|U)$, we can apply Lemma 2.20 to $B^n = W^n$, $K = (C, F)$, and $A^n = U^n$, and claim that there exists a fixed pair of binnings $\varphi' := (\varphi'_1, \varphi'_2)$ such that, if we denote with $P^{\text{RB},\varphi'}$ and $P^{\text{RC},\varphi'}$ the distributions P^{RB} and P^{RC} with respect to the choice of a binning φ', we have

$$\mathbb{D}(P^{\text{RB},\varphi'}_{U^n CF} \| P^{\text{RC},\varphi'}_{U^n} Q_C Q_F) = \delta(n), \tag{3.9}$$

which by (3.8) implies

$$\mathbb{D}(P^{\text{RB},\varphi'}_{U^n W^n \hat{W}^n X^n Y^n CF} \| P^{\text{RC},\varphi'}_{U^n W^n \hat{W}^n X^n Y^n CF}) = \delta(n).$$

Then, by Pinsker's inequality we have

$$\mathbb{V}(P^{\text{RB},\varphi'}_{U^n W^n \hat{W}^n X^n Y^n CF}, P^{\text{RC},\varphi'}_{U^n W^n \hat{W}^n X^n Y^n CF}) = \delta(n). \tag{3.10}$$

From now on, we omit φ' to simplify the notation.

The next step is to show that we have strong coordination for V^n as well. The main difficulty is that in the second coding scheme V^n is generated using the output of the Slepian-Wolf decoder $\hat{W}^n$ and not W^n as in the first scheme. Because of Lemma 2.19, the inequality $\tilde{R} + R_0 > H(W|Y)$ implies that $\hat{W}^n$ is equal to W^n with high probability and we will use this fact to show that the probability distributions are close in total variational distance.

Now, we need to establish a technical lemma, whose proof can be found in Appendix B.1.

Lemma 3.6 *Let A^n and $\hat{A}^n$ be such that $\mathbb{P}\{\hat{A}^n \neq A^n\} \to 0$ when $n \to \infty$. Then for any random variable B^n and for any joint distribution $P_{B^n A^n \hat{A}^n}$ we have:*

$$\lim_{n\to\infty} \mathbb{V}(P_{B^n A^n \hat{A}^n}, P_{B^n A^n} \mathbb{1}_{\hat{A}^n|A^n}) = 0,$$

$$\textit{where } \mathbb{1}_{\hat{A}^n|A^n}(\mathbf{a}|\mathbf{a}') = \begin{cases} 1 & \textit{if } \mathbf{a} = \mathbf{a}', \\ 0 & \textit{if } \mathbf{a} \neq \mathbf{a}'. \end{cases}$$

Then, Lemma 3.6 implies that:

$$\mathbb{V}(P^{\text{RB}}_{U^n \hat{W}^n W^n X^n Y^n CF}, P^{\text{RB}}_{U^n W^n X^n Y^n CF} \mathbb{1}_{\hat{W}^n|W^n}) = \delta(n). \tag{3.11}$$

Similarly, we apply the same reasoning to the random coding scheme, and we have

$$\mathbb{V}(P^{\text{RC}}_{U^nW^n\hat{W}^nX^nY^nCF}, P^{\text{RC}}_{U^nW^nX^nY^nCF}\mathbb{1}_{\hat{W}^n|W^n}) = \delta(n). \tag{3.12}$$

Then using the triangle inequality, we find that

$$\begin{aligned}
\mathbb{V}(P^{\text{RB}}, P^{\text{RC}}) &= \mathbb{V}(P^{\text{RB}}_{U^nW^n\hat{W}^nX^nY^nCF}P^{\text{RB}}_{V^n|W^nY^n}, P^{\text{RC}}_{U^nW^n\hat{W}^nX^nY^nCF}P^{\text{RC}}_{V^n|\hat{W}^nY^n}) \\
&\leq \mathbb{V}(P^{\text{RB}}_{U^nW^n\hat{W}^nX^nY^nCF}\bar{P}_{V^n|W^nY^n}, P^{\text{RB}}_{U^nW^nX^nY^nCF}\mathbb{1}_{\hat{W}^n|W^n}\bar{P}_{V^n|W^nY^n}) \\
&\quad + \mathbb{V}(P^{\text{RB}}_{U^nW^nX^nY^nCF}\mathbb{1}_{\hat{W}^n|W^n}\bar{P}_{V^n|W^nY^n}, P^{\text{RC}}_{U^nW^nX^nY^nCF}\mathbb{1}_{\hat{W}^n|W^n}P^{\text{RC}}_{V^n|\hat{W}^nY^n}) \\
&\quad + \mathbb{V}(P^{\text{RC}}_{U^nW^nX^nY^nCF}\mathbb{1}_{\hat{W}^n|W^n}P^{\text{RC}}_{V^n|W^nY^n}, P^{\text{RC}}_{U^nW^n\hat{W}^nX^nY^nCF}P^{\text{RC}}_{V^n|\hat{W}^nY^n}).
\end{aligned} \tag{3.13}$$

The first and the third term go to zero exponentially by applying Lemma 2.13 to (3.11) and (3.12) respectively. Now observe that

$$\mathbb{1}_{\hat{W}^n|W^n}\bar{P}_{V^n|W^nY^n} = \mathbb{1}_{\hat{W}^n|W^n}P^{\text{RC}}_{V^n|\hat{W}^nY^n}$$

by definition of $P^{\text{RC}}_{V^n|\hat{W}^nY^n}$ (3.6). Then by using Lemma 2.13 again the second term is equal to

$$\mathbb{V}(P^{\text{RB}}_{U^nW^nX^nY^nCF}, P^{\text{RC}}_{U^nW^nX^nY^nCF}),$$

and goes to zero by (3.10) and Lemma 2.12. Hence, we have

$$\mathbb{V}(P^{\text{RB}}, P^{\text{RC}}) = \delta(n). \tag{3.14}$$

Using Lemma 2.12, we conclude that

$$\mathbb{V}(P^{\text{RB}}_{U^nX^nW^n\hat{W}^nY^nV^n}, P^{\text{RC}}_{U^nX^nW^n\hat{W}^nY^nV^n}) = \delta(n).$$

Remove the extra randomness F Even though the extra common randomness F is required to coordinate $(U^n, X^n, Y^n, V^n, W^n)$ we will show that we do not need it in order to coordinate only (U^n, X^n, Y^n, V^n). Observe that by Lemma 2.12, equation (3.14) implies that

$$\mathbb{V}(P^{\text{RB}}_{U^nX^nY^nV^nF}, P^{\text{RC}}_{U^nX^nY^nV^nF}) = \delta(n). \tag{3.15}$$

As in [70], we would like to reduce the amount of common randomness by having the two nodes agree on an instance $F = f$. To do so, we apply Lemma 2.20 to $B^n = W^n$, $K = F$, and $A^n = U^nX^nY^nV^n$. If $R < H(W|UXYV)$, Lemma 2.20 implies that there exists a fixed binning φ''_2 such that

$$\mathbb{V}(P^{\text{RB},\varphi''_2}_{U^nX^nY^nV^nF}, Q_F P^{\text{RB},\varphi''_2}_{U^nX^nY^nV^n}) = \delta(n). \tag{3.16}$$

Remark 3.7 - ***One binning for two conditions*.** Note that we had already chosen a specific binning φ_2', we can prove that there exists a binning which works for both conditions (3.9) and (3.16), i.e. we can take $\varphi_2' = \varphi_2''$.

Proof. If we denote with $\mathbb{E}_{\varphi_1 \varphi_2}$ the expected value with respect to the random binnings, for all ε, there exists $\bar{n}$ such that $\forall n \geq \bar{n}$

$$\mathbb{E}_{\varphi_1 \varphi_2}\left[\mathbb{V}(P^{\text{RB},(\varphi_1,\varphi_2)}_{U^n F C}, Q_F Q_C \bar{P}_{U^n})\right] < \frac{\varepsilon}{2}$$
$$\mathbb{E}_{\varphi_1 \varphi_2}[\mathbb{V}(P^{\text{RB},(\varphi_1,\varphi_2)}_{U^n X^n Y^n V^n F}, Q_F P^{\text{RB}}_{U^n X^n Y^n V^n})] < \frac{\varepsilon}{2}$$

which implies by Markov's inequality

$$\mathbb{P}_{\varphi_1 \varphi_2}\left\{\mathbb{V}(P^{\text{RB},(\varphi_1,\varphi_2)}_{U^n F C}, Q_F Q_C \bar{P}_{U^n}) < \varepsilon\right\} > \frac{1}{2}$$
$$\mathbb{P}_{\varphi_1 \varphi_2}\{\mathbb{V}(P^{\text{RB},(\varphi_1,\varphi_2)}_{U^n X^n Y^n V^n F}, Q_F P^{\text{RB}}_{U^n X^n Y^n V^n}) < \varepsilon\} > \frac{1}{2}. \tag{3.17}$$

Note that in (3.16) we have not imposed any restrictions on the choice of φ_1'', so we can take $\varphi_1'' = \varphi_1'$. Thus, we have chosen the binnings (φ_1', φ_2') and $(\varphi_1'', \varphi_2'')$ respectively such that

$$\lim_{n\to\infty} \mathbb{V}(P^{\text{RB},(\varphi_1',\varphi_2')}_{U^n F C}, Q_F Q_C \bar{P}_{U^n}) = 0$$
$$\lim_{n\to\infty} \mathbb{V}(P^{\text{RB},(\varphi_1'',\varphi_2'')}_{U^n X^n Y^n V^n F}, Q_F P^{\text{RB}}_{U^n X^n Y^n V^n}) = 0.$$

It follows from (3.17) that the intersection of the following two sets is non-empty

$$\left\{ (\varphi_1, \varphi_2) \;\middle|\; \mathbb{V}(P^{\text{RB},(\varphi_1,\varphi_2)}_{U^n F C}, Q_F Q_C \bar{P}_{U^n}) < \varepsilon\right\},$$
$$\left\{ (\varphi_1, \varphi_2) \;\middle|\; \mathbb{V}(P^{\text{RB},(\varphi_1,\varphi_2)}_{U^n X^n Y^n V^n F}, Q_F P^{\text{RB}}_{U^n X^n Y^n V^n}) < \varepsilon\right\},$$

therefore there exists a binning $(\varphi_1^*, \varphi_2^*)$ that satisfies both conditions. □

Because of (3.15), (3.16) implies

$$\mathbb{V}(P^{\text{RC}}_{U^n X^n Y^n V^n F}, Q_F P^{\text{RB}}_{U^n X^n Y^n V^n}) = \delta(n). \tag{3.18}$$

By Lemma 2.15, there exists an instance $f \in [\![1, 2^{nR}]\!]$ such that

$$\mathbb{V}(P^{\text{RB}}_{U^n X^n Y^n V^n | F=f}, P^{\text{RC}}_{U^n X^n Y^n V^n | F=f}) = \delta(n).$$

Then, by fixing $F = f$ and using common randomness C, we have coordination for (U^n, X^n, Y^n, V^n).

Rate constraints We have imposed the following rate constraints:

$$H(W|Y) < R + R_0 < H(W|U),$$

$$R < H(W|UXYV).$$

Therefore we obtain:

$$R_0 > H(W|Y) - H(W|UXYV) = I(W;UXV|Y),$$
$$I(W;U) < I(W;Y). \qquad \square$$

Comparison with strong coordination of actions With the same random binning techniques, [31] characterizes an inner bound for the strong coordination region in the slightly different scenario in which only U^n and V^n need to be coordinated. Given the source and channel parameters $\bar{P}_U$ and $\bar{P}_{Y|X}$ respectively, the inner bound in [31] is:

$$\mathcal{R}_{\text{Hadd,in}} := \left\{ (\bar{P}_{UV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{UXYV} = \bar{P}_U \bar{P}_{X|U} \bar{P}_{Y|X} \bar{P}_{V|UXY} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UXYWV} = \bar{P}_U \bar{P}_{WX|U} \bar{P}_{Y|X} \bar{P}_{V|WY} \\ I(W;U) \leq I(W;Y) \\ R_0 \geq I(W;UV) - I(W;Y) \end{array} \right\}. \tag{3.19}$$

Note that the decomposition of the joint distribution and the information constraints are the same as in (3.1). The only difference is that the rate of common randomness in (3.1) is larger since

$$I(W;UXV|Y) = I(W;UXYV) - I(W;Y) \geq I(W;UV) - I(W;Y).$$

The difference in common randomness rate $I(W;XY|UV)$ stems from the requirement in [31], which coordinates U^n and V^n only instead of coordinating (U^n, X^n, Y^n, V^n).

3.1.2 Outer bound

Consider a code (f_n, g_n) that induces a distribution $P_{U^n X^n Y^n V^n}$ that is ε-close in total variational distance to the i.i.d. distribution $\bar{P}_{UXYV}^{\otimes n}$. Let the random variable T be uniformly distributed over the set $[\![1,n]\!]$ and independent of sequence (U^n, X^n, Y^n, V^n, C). The variable T will serve as a random time index. The variable U_T is independent of T because U^n is an i.i.d. source sequence [25, Section VII.B].

Bound on R_0 We apply Lemma 2.18 to $A^n := U^n X^n V^n$, $B^n := Y^n$ and C and, using (2.5), we have

$$\begin{aligned} nR_0 &= H(C) \geq H(C|B^n) \\ &\overset{(a)}{\geq} nI(A_T; CB_{\sim T}T|B_T) - nI(A_T B_T; T) - nf(\varepsilon) \\ &\overset{(b)}{\geq} nI(A_T; CB_{\sim T}T|B_T) - 2nf(\varepsilon) = nI(U_T X_T V_T; CY_{\sim T}T|Y_T) - 2nf(\varepsilon) \end{aligned} \tag{3.20}$$

where (a) follows from (2.5) in Lemma 2.18 and (b) comes from Lemma 2.16.

Information constraint We have

$$
\begin{aligned}
0 &\overset{(a)}{\leq} I(X^n;Y^n) - I(C,U^n;Y^n)\\
&\leq I(X^n;Y^n) - I(U^n;Y^n|C)\\
&= H(Y^n) - H(Y^n|X^n) + H(U^n|Y^nC) - H(U^n|C)\\
&\overset{(b)}{\leq} \sum_{t=1}^{n} H(Y_t) - \sum_{t=1}^{n} H(Y_t|X_t) + \sum_{t=1}^{n} H(U_t|U^{t-1}Y_tY_{\sim t}C) - \sum_{t=1}^{n} H(U_t)\\
&\overset{(c)}{\leq} \sum_{t=1}^{n} (H(Y_t) - H(Y_t|X_t) + H(U_t|Y_{\sim t}C) - H(U_t))\\
&\overset{(d)}{\leq} nH(Y_T) - nH(Y_T|X_TT) + nH(U_T|Y_{\sim T}CT) - nH(U_T|T)\\
&\overset{(e)}{=} nH(Y_T) - nH(Y_T|X_T) + nH(U_T|Y_{\sim T}CT) - nH(U_T)\\
&= nI(X_T;Y_T) - nI(U_T;Y_{\sim T},C,T)
\end{aligned}
$$

where (a) comes from the Markov chain $Y^n - X^n - (C,U^n)$ and (b) comes from the chain rule for the conditional entropy and the fact that U^n is an i.i.d. source independent of C. The inequalities (c) and (d) come from the fact that conditioning does not increase entropy and (e) from the memoryless nature of the channel $\bar{P}_{Y|X}$ and the i.i.d. nature of the source $\bar{P}_U$.

Identification of the auxiliary random variable We identify the auxiliary random variables W_t with $(C,Y_{\sim t})$ for each $t \in [\![1,n]\!]$ and W with $(W_T,T) = (C,Y_{\sim T},T)$. For each $t \in [\![1,n]\!]$ the following two Markov chains hold:

$$Y_t - X_t - (C,Y_{\sim t},U_t) \iff Y_t - X_t - (W_t,U_t) \tag{3.21}$$
$$V_t - (C,Y_{\sim t},Y_t) - (U_t,X_t) \iff V_t - (W_t,Y_t) - (U_t,X_t) \tag{3.22}$$

where (3.21) comes from the fact that the channel is memoryless and (3.22) from the fact that the decoder is non-causal and for each $t \in [\![1,n]\!]$ the decoder generates V_t from Y^n and common randomness C. Then, we have

$$Y_T - X_T - (C,Y_{\sim T},U_T,T) \iff Y_T - X_T - (W_T,U_T,T) \tag{3.23}$$
$$V_T - (C,Y_{\sim T},Y_T,T) - (U_T,X_T) \iff V_T - (W_T,Y_T,T) - (U_T,X_T) \tag{3.24}$$

where (3.23) holds because

$$\mathbb{P}\{Y_T = y \mid X_T = x, Y_{\sim T} = \tilde{\mathbf{y}}, U_T = u, T = t, C\} = \mathbb{P}\{Y_T = y \mid X_T = x\}$$

since the channel is memoryless. Then by (3.22), (3.24) holds because

$$I(V_T;U_TX_T|CY^nT) = \sum_{i=1}^{n}\frac{1}{n}I(V_t;U_tX_t|CY^nT=t) = 0.$$

Since $W = W_t$ when $T = t$, we also have $(U,X) - (W,Y) - V$ and $Y - X - (U,W)$.

Proof of cardinality bound First, we state a direct consequence of the Fenchel-Eggleston-Carathéodory theorem [26, Appendix A].

Lemma 3.8 - ***Support Lemma [26, Appendix C]*** *Let $\mathcal{A}$ a finite set and $\mathcal{W}$ be an arbitrary set. Let $\mathcal{P}$ be a connected compact subset of probability mass functions on $\mathcal{A}$ and $P_{A|W}$ be a collection of conditional probability mass functions on $\mathcal{A}$. Suppose that $h_i(\pi)$, $i=1,\dots,d$, are real-valued continuous functions of $\pi\in\mathcal{P}$. Then for every W defined on $\mathcal{W}$ there exists a random variable W' with $|\mathcal{W}'|\leq d$ and a collection of conditional probability mass functions $P_{A|W'}\in\mathcal{P}$ such that*

$$\sum_{w\in\mathcal{W}} P_W(w)h_i(P_{A|W}(a|w)) = \sum_{w\in\mathcal{W}'} P_{W'}(w)h_i(P_{A|W'}(a|w)) \quad i=1,\dots,d.$$

Now, we consider the probability distribution $\bar{P}_U\bar{P}_{W|U}\bar{P}_{X|UW}\bar{P}_{Y|X}\bar{P}_{V|WY}$ that is ε-close in total variational distance to the i.i.d. distribution. We identify $\mathcal{A}$ with $\{1,\dots,|\mathcal{A}|\}$ and we consider $\mathcal{P}$ a connected compact subset of probability mass functions on $\mathcal{A}=\mathcal{U}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{V}$. Similarly to [46], suppose that (U,X,Y,V) has distribution π and $h_i(\pi)$, $i=1,\dots,|\mathcal{A}|+4$, are real-valued continuous functions of $\pi\in\mathcal{P}$ such that:

$$h_i(\pi) = \begin{cases} \pi(i) & \text{for } i=1,\dots,|\mathcal{A}|-1 \\ H(U) & \text{for } i=|\mathcal{A}| \\ H(UXV|Y) & \text{for } i=|\mathcal{A}|+1 \\ H(Y|UX) & \text{for } i=|\mathcal{A}|+2 \\ H(V|Y) & \text{for } i=|\mathcal{A}|+3 \\ H(V|UXY) & \text{for } i=|\mathcal{A}|+4 \end{cases}.$$

Then by Lemma 3.8 there exists an auxiliary random variable W' taking at most $|\mathcal{U}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{V}|+4$ values such that:

$$H(U|W) = \sum_{w\in\mathcal{W}} P_W(w)H(U|W=w) = \sum_{w\in\mathcal{W}'} P_{W'}(w)H(U|W'=w) = H(U|W'),$$
$$H(UXV|YW) = \sum_{w\in\mathcal{W}} P_W(w)H(UXV|YW=w) = \sum_{w\in\mathcal{W}'} P_{W'}(w)H(UXV|YW'=w) = H(UXV|YW'),$$
$$H(Y|UXW) = \sum_{w\in\mathcal{W}} P_W(w)H(Y|UXW=w) = \sum_{w\in\mathcal{W}'} P_{W'}(w)H(Y|UXW'=w) = H(Y|UXW'),$$
$$H(V|YW) = \sum_{w\in\mathcal{W}} P_W(w)H(V|YW=w) = \sum_{w\in\mathcal{W}'} P_{W'}(w)H(V|YW'=w) = H(V|YW'),$$

$$H(V|UXYW) = \sum_{w\in\mathcal{W}} P_W(w) H(V|UXYW=w) = \sum_{w\in\mathcal{W}'} P_{W'}(w) H(V|UXYW'=w) = H(V|UXYW').$$

Hence, the constraints on the conditional distributions, the information constraints and the Markov chains are still verified since we can rewrite the inequalities in (3.2) and the Markov chains in (3.3) as

$$\begin{aligned}
&H(U) - H(U|W) \leq I(X;Y),\\
&R_0 \geq H(UXV|Y) - H(UXV|WY),\\
&I(Y;UW|X) = H(Y|X) - H(Y|UXW) = 0,\\
&I(V;UX|YW) = H(V|YW) - H(V|UXYW) = 0.
\end{aligned}$$

Note that we are not forgetting any constraints: to preserve $H(U) - H(U|W) \leq I(X;Y)$ we only need to fix $H(U|W)$ because the other quantities depend only on the joint distribution P_{UXYV} (which is preserved). Similarly, once the distribution $\bar{P}_{UXYV}$ is preserved, the difference $H(UXV|Y) - H(UXV|WY)$ only depends on the conditional entropy $H(UXV|WY)$ and the difference $H(Y|X) - H(Y|UXW)$ only depends on $H(Y|UXW)$. □

3.2 Two-node network with two-sided state information

The coordination problem introduced up to now assumes that the source and the channel follow distributions which are fixed ahead of time and known by the agents. However, this constraint prevents us from modeling situations in which the agent reacts to an external stimulus, and in which the channel statistics depend on the environment. For instance, consider a situation where the actions of an agent might be constrained by obstacles that prevent it from making certain choices. In this case the probability distributions given by nature could change with time and would be partially if not completely unknown to some of the agents. To include such situations in the coordination framework, we extend the model to take into account the uncertainty about the source and channel distribution, as in [40–42, 44] for empirical coordination.

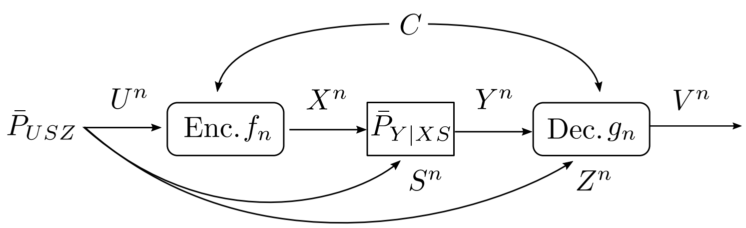

Figure 3.3: Coordination of signals and actions for a two-node network with a noisy channel with state and side information at the decoder.

We consider the model depicted in Figure 3.3. It is a generalization of the simpler setting of Figure 3.1, where we introduce a state in the description of the behavior. Here, we introduce a state-dependent i.i.d. source (U, S, Z) generated according to $\bar{P}_{USZ}$ and a state-dependent noisy channel $\bar{P}_{Y|XS}$. The encoder selects a signal $X^n = f_n(U^n, C)$, with $f_n : \mathcal{U}^n \times [\![1, 2^{nR_0}]\!] \to \mathcal{X}^n$, and transmits it over the discrete memoryless channel $\bar{P}_{Y|XS}$ where S represents the state. The decoder then selects an action $V^n = g_n(Y^n, Z^n, C)$, where $g_n : \mathcal{Y}^n \times \mathcal{Z}^n \times [\![1, 2^{nR_0}]\!] \to \mathcal{V}^n$ is a stochastic map and Z^n represents the side information available at the decoder.

Remark 3.9 - ***Generality of the setting*****.** Note that the model is quite general and includes scenarios where partial or perfect channel state information is available at the encoder as well, since the variables U^n and S^n are possibly correlated. Moreover the side information Z^n can be about the source and/or the state.

In the case of non-causal encoder and decoder, the problem of characterizing the strong coordination region $\mathcal{R}_{\text{state}}$ for the system model in Figure 3.3 is still open, but we establish the following inner and outer bounds.

Theorem 3.10 *Let $\bar{P}_{USZ}$ and $\bar{P}_{Y|XS}$ be the given source and channel parameters, then $\mathcal{R}_{\text{state,in}} \subseteq \mathcal{R}_{\text{state}} \subseteq \mathcal{R}_{\text{state,out}}$ where:*

$$\mathcal{R}_{\text{state,in}} := \left\{ (\bar{P}_{USZXYV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{USZXYV} = \bar{P}_{USZ}\bar{P}_{X|U}\bar{P}_{Y|XS}\bar{P}_{V|UXYSZ} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{USZWXYV} = \bar{P}_{USZ}\bar{P}_{W|U}\bar{P}_{X|UW}\bar{P}_{Y|XS}\bar{P}_{V|WYZ} \\ I(W;U) \leq I(W;YZ) \\ R_0 \geq I(W;USXV|YZ) \end{array} \right\}, \quad (3.25)$$

$$\mathcal{R}_{\text{state,out}} := \left\{ (\bar{P}_{USZXYV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{USZXYV} = \bar{P}_{USZ}\bar{P}_{X|U}\bar{P}_{Y|XS}\bar{P}_{V|UXYSZ} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{USZWXYV} = \bar{P}_{USZ}\bar{P}_{W|U}\bar{P}_{X|UW}\bar{P}_{Y|XS}\bar{P}_{V|WYZ} \\ I(W;U) \leq \min\{I(XUS;YZ), I(XS;Y) + I(U;Z)\} \\ R_0 \geq I(W;USXV|YZ) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| + 5 \end{array} \right\}. \quad (3.26)$$

Remark 3.11 - ***Two outer bounds*****.** As in Theorem 3.3, even if inner and outer bounds do not match, they only differ on the information constraint involving $I(W;U)$. Note that we cannot compare $I(XUS;YZ)$ and $I(XS;Y) + I(U;Z)$:

$$\begin{aligned} I(XUS;YZ) &= I(XS;YZ) + I(U;YZ|XS) \\ &\overset{(a)}{=} I(XS;Y) + I(U;Z) + I(SXY;Z) \\ &\quad - I(Y;Z) + I(UZ;SXY) - I(SXY;Z) - I(U;XS) \\ &\overset{(b)}{=} I(XS;Y) + I(U;Z) - I(Y;Z) + I(UZ;SX) - I(U;XS) \\ &= I(XS;Y) + I(U;Z) - I(Y;Z) + I(Z;SX|U) + I(U;XS) - I(U;XS) \\ &\overset{(c)}{=} I(XS;Y) + I(U;Z) - I(Y;Z) + I(Z;S|U) \end{aligned}$$

where (a) follows from basic properties of the mutual information, (b) and (c) from the Markov chains $Y - XS - UZ$ and $X - US - Z$ respectively. If we note $\Delta := I(Z;S|U) - I(Y;Z)$, then $I(XUS;YZ) = I(XS;Y) + I(U;Z) + \Delta$ where Δ may be either positive or negative, for instance:

- in the special case where $S - U - Z$ holds and $Y = Z$, $\Delta = -H(Y) \leq 0$;
- if we suppose Y independent of Z, $\Delta = I(Z;S|U) \geq 0$.

Hence, in $\mathcal{R}_{\text{state,out}}$ the upper bound on the mutual information $I(W;U)$ is the minimum of the two.

Remark 3.12 - *Markov chain decomposition*. Observe that the decomposition of the joint distributions $\bar{P}_{USZXYV}$ and $\bar{P}_{USZWXYV}$ is equivalently characterized in terms of Markov chains:

$$\begin{cases} Z - (U,S) - (X,Y), \\ Y - (X,S) - U, \end{cases} \qquad \begin{cases} Z - (U,S) - (X,Y,W), \\ Y - (X,S) - (U,W), \\ V - (Y,Z,W) - (X,S,U). \end{cases} \tag{3.27}$$

3.2.1 Inner bound

The achievability proof is a generalization of the inner bound in Theorem 3.3 proved in Section 3.1.1 and can be found in Appendix B.1.2.

3.2.2 Outer bound

Consider a code (f_n, g_n) that induces a distribution $P_{U^nS^nZ^nX^nY^nV^n}$ that is ε-close in total variational distance to the i.i.d. distribution $\bar{P}^{\otimes n}_{USZXYV}$. Let the random variable T be uniformly distributed over the set $[\![1,n]\!]$ and independent of the sequence $(U^n, S^n, Z^n, X^n, Y^n, V^n, C)$. The variable (U_T, S_T, Z_T) is independent of T because (U^n, S^n, Z^n) is an i.i.d. source sequence [25, Section VII.B].

Bound on R_0 The proof is the same as in Section 3.1.2 (based on Lemma 2.18): in (3.20), we identify $A^n := U^nS^nX^nV^n$ and $B^n := Y^nZ^n$. Then, we obtain $R_0 \geq I(W;USXV|YZ)$.

Information constraint As shown in Remark 3.11, in the general case we are not able to compare $I(XUS;YZ)$ and $I(XS;Y) + I(U;Z)$. Then, we show separately that:

$$I(W;U) \leq I(XUS;YZ), \tag{3.28}$$
$$I(W;U) \leq I(XS;Y) + I(U;Z). \tag{3.29}$$

Proof of (3.28) We have

$$0 \overset{(a)}{\leq} I(X^nS^n;Y^n) - I(CU^n;Y^n) = H(Y^n|CU^n) - H(Y^n|X^nS^n)$$
$$\overset{(b)}{=} H(Y^n|CU^n) - H(Y^n|CU^nX^nS^n) = I(Y^n;X^nS^n|CU^n) \leq I(Y^nZ^n;X^nS^n|CU^n) \tag{3.30}$$

$$= I(Y^nZ^n; X^nS^nU^n|C) - I(Y^nZ^n; U^n|C) \overset{(c)}{\leq} nI(Y_TZ_T; X_TS_TU_T|T) - nI(U_T; Y_{\sim T}Z_{\sim T}CT)$$

where (a) and (b) come from the Markov chain $Y^n - (X^n, S^n) - (C, U^n)$. To prove (c), we show separately that:

(i) $I(Y^nZ^n; U^n|C) \geq nI(U_T; Y_{\sim T}Z_{\sim T}CT)$,

(ii) $I(Y^nZ^n; X^nS^nU^n|C) \leq nI(Y_TZ_T; X_TS_TU_T|T)$.

Proof of (i) Observe that

$$\begin{aligned}
I(Y^nZ^n; U^n|C) &= H(U^n|C) - H(U^n|Y^nZ^nC) \overset{(d)}{=} H(U^n) - H(U^n|Y^nZ^nC)\\
&\overset{(e)}{=} \sum_{t=1}^{n}(H(U_t) - H(U_t|U^{t-1}Y_tZ_tY_{\sim t}Z_{\sim t}C)) \geq \sum_{t=1}^{n}(H(U_t) - H(U_t|Y_{\sim t}Z_{\sim t}C))\\
&= nH(U_T|T) - nH(U_T|Y_{\sim T}Z_{\sim T}CT) \overset{(f)}{=} nH(U_T) - nH(U_T|Y_{\sim T}Z_{\sim T}CT)\\
&= nI(U_T; Y_{\sim T}Z_{\sim T}CT)
\end{aligned}$$

where (d) comes from the independence between U^n and C and (e) and (f) follow from the i.i.d. nature of U^n.

Proof of (ii) First, we need the following result (proved in Appendix B.2).

Lemma 3.13 *For every $t \in [\![1, n]\!]$ the following Markov chain holds:*

$$(Y_t, Z_t) - (X_t, U_t, S_t) - (C, X_{\sim t}, U_{\sim t}, S_{\sim t}, Y_{\sim t}, Z_{\sim t}). \tag{3.31}$$

Then, observe that

$$\begin{aligned}
I(Y^nZ^n; X^nS^nU^n|C) &\leq I(Y^nZ^n; X^nS^nU^nC)\\
&= \sum_{t=1}^{n} I(Y_tZ_t; X^nS^nU^nC|Y^{t-1}Z^{t-1}) \leq \sum_{t=1}^{n} I(Y_tZ_t; X^nS^nU^nCY^{t-1}Z^{t-1})\\
&= \sum_{t=1}^{n} I(Y_tZ_t; X_tS_tU_t) + \sum_{t=1}^{n} I(Y_tZ_t; X_{\sim t}S_{\sim t}U_{\sim t}CY^{t-1}Z^{t-1}|X_tS_tU_t)\\
&\overset{(g)}{=} \sum_{t=1}^{n} I(Y_tZ_t; X_tS_tU_t) = nI(Y_TZ_T; X_TS_TU_T|T)
\end{aligned}$$

where (g) follows from Lemma 3.13. Moreover, since the distributions are ε-close to i.i.d. by hypothesis, the last term is close to $nI(YZ; XSU)$. In fact, we have

$$\begin{aligned}
I(Y_TZ_T; X_TS_TU_T|T) &= H(Y_TZ_T|T) + H(X_TS_TU_T|T) - H(Y_TZ_TX_TS_TU_T|T)\\
&= \sum_{t=1}^{n}\frac{1}{n}H(Y_tZ_t) + \sum_{t=1}^{n}\frac{1}{n}H(X_tS_tU_t) - \sum_{t=1}^{n}\frac{1}{n}H(Y_tZ_tX_tS_tU_t).
\end{aligned}$$

Then, as in the proof of Lemma 2.16,

$$\begin{aligned}
|H(Y_t Z_t) - H(YZ)| &\leq 2\varepsilon \log\left(\frac{|\mathcal{Y}\times\mathcal{Z}|}{\varepsilon}\right) := \varepsilon_1,\\
|H(X_t S_t U_t) - H(XSU)| &\leq 2\varepsilon \log\left(\frac{|\mathcal{X}\times\mathcal{S}\times\mathcal{U}|}{\varepsilon}\right) := \varepsilon_2,\\
|H(Y_t Z_t X_t S_t U_t) - H(YZXSU)| &\leq 2\varepsilon \log\left(\frac{|\mathcal{Y}\times\mathcal{Z}\times\mathcal{X}\times\mathcal{S}\times\mathcal{U}|}{\varepsilon}\right) := \varepsilon_3.
\end{aligned}$$

This implies that

$$|I(Y_T Z_T; X_T S_T U_T|T) - I(YZ;XSU)| \leq g(\varepsilon), \tag{3.32}$$

where $g(\varepsilon) := (\varepsilon_1 + \varepsilon_2 + \varepsilon_3)$. Then, (3.30) becomes

$$0 \leq nI(YZ;XSU) - nI(U_T; Y_{\sim T} Z_{\sim T} CT) + g(\varepsilon).$$

Proof of (3.29) In this case, for the second part of the converse, we have

$$\begin{aligned}
0 &\overset{(a)}{\leq} I(X^n S^n; Y^n) - I(CZ^n U^n; Y^n) \overset{(b)}{\leq} I(X^n S^n; Y^n) - I(U^n; Y^n C|Z^n)\\
&= H(Y^n) - H(Y^n|X^n S^n) - H(U^n) + I(U^n;Z^n) + H(U^n|Y^n Z^n C)\\
&\overset{(c)}{\leq} \sum_{t=1}^n H(Y_t) - \sum_{t=1}^n H(Y_t|X_t S_t) - \sum_{t=1}^n H(U_t) + \sum_{t=1}^n I(U_t;Z_t) + \sum_{t=1}^n H(U_t|U^{t-1} Y_t Z_t Y_{\sim t} Z_{\sim t} C)\\
&\overset{(d)}{\leq} nH(Y_T) - nH(Y_T|X_T S_T T) - nH(U_T|T) + nI(U_T;Z_T|T) + nH(U_T|Y_{\sim T} Z_{\sim T} CT)\\
&\overset{(e)}{=} nH(Y_T) - nH(Y_T|X_T S_T) - nH(U_T) + nI(U_T;Z_T) + nH(U_T|Y_{\sim T} Z_{\sim T} CT)\\
&= nI(X_T, S_T; Y_T) - nI(U_T; Y_{\sim T} Z_{\sim T} CT) + nI(U_T; Z_T)
\end{aligned}$$

where (a) comes from the Markov chain $Y^n - (X^n, S^n) - (C, Z^n, U^n)$, (b) from the fact that

$$I(CZ^n U^n; Y^n) \geq I(Z^n U^n; Y^n|C) = I(Z^n U^n; Y^n C) \geq I(U^n; Y^n C|Z^n)$$

by the chain rule and the fact that U^n and Z^n are independent of C. Then (c) comes from the chain rule for the conditional entropy. The inequalities (d) comes from the fact that conditioning does not increase entropy (in particular $H(Y_T|T) \leq H(Y_T)$) and (e) from the memoryless channel $\bar{P}_{Y|XS}$ and the i.i.d. source $\bar{P}_{UZ}$. Finally, since the source is i.i.d. the last term is $nI(U;Z)$.

Remark 3.14 Note that if U is independent of Z the upper bound for $I(U;W)$ is $I(XS;Y)$.

Identification of the auxiliary random variable For each $t \in [\![1,n]\!]$ we identify the auxiliary random variables W_t with $(C, Y_{\sim t}, Z_{\sim t})$ and W with $(W_T, T) = (C, Y_{\sim T}, Z_{\sim T}, T)$.

The following Markov chains hold for each $t \in [\![1, n]\!]$:

$$Z_t - (U_t, S_t) - (C, X_t, Y_t, Y_{\sim t}, Z_{\sim t}) \iff Z_t - (U_t, S_t) - (X_t, Y_t, W_t), \quad (3.33)$$
$$Y_t - (X_t, S_t) - (C, Y_{\sim t}, Z_{\sim t}, U_t) \iff Y_t - (X_t, S_t) - (W_t, U_t), \quad (3.34)$$
$$V_t - (C, Y_{\sim t}, Z_{\sim t}, Y_t, Z_t) - (U_t, S_t, X_t) \iff V_t - (W_t, Y_t, Z_t) - (U_t, S_t, X_t). \quad (3.35)$$

Then we have

$$Z_T - (U_T, S_T) - (C, X_T, Y_T, Y_{\sim T}, Z_{\sim T}, T) \Leftrightarrow Z_T - (U_T, S_T) - (X_T, Y_T, W_T, T), \quad (3.36)$$
$$Y_T - (X_T, S_T) - (C, Y_{\sim T}, Z_{\sim T}, U_T, T) \Leftrightarrow Y_T - (X_T, S_T) - (W_T, U_T, T), \quad (3.37)$$
$$V_T - (C, Y_{\sim T}, Z_{\sim T}, Y_T, Z_T, T) - (U_T, S_T, X_T) \Leftrightarrow V_T - (W_T, Y_T, Z_T, T) - (U_T, S_T, X_T). \quad (3.38)$$

where (3.36) and (3.37) come from the fact that

$$\begin{aligned}
&\mathbb{P}\{Z_T = z | S_T = s, U_T = u, X_T = x, Y_T = y, Y_{\sim T} = \tilde{\mathbf{y}}, Z_{\sim T} = \tilde{\mathbf{z}}, T = t, C\} \\
&\qquad = \mathbb{P}\{Z_T = z | S_T = s, U_T = u\}, \\
&\mathbb{P}\{Y_T = y | X_T = x, S_T = s, Y_{\sim T} = \tilde{\mathbf{y}}, Z_{\sim T} = \tilde{\mathbf{z}}, U_T = u, T = t, C\} \\
&\qquad = \mathbb{P}\{Y_T = y | X_T = x, S_T = s\}
\end{aligned}$$

since the source is i.i.d. and the channel is memoryless. Then by (3.35), (3.38) holds because

$$I(V_T; U_T S_T X_T | C Y^n Z^n T) = \sum_{i=1}^{n} \frac{1}{n} I(V_t; U_t S_t X_t | C Y^n Z^n T = t) = 0.$$

Since $W = W_t$ when $T = t$, we also have $Z - (U, S) - (X, Y, W)$, $(U, S, X) - (W, Y, Z) - V$ and $Y - (X, S) - (W, U)$. The cardinality bound is proved in Appendix B.5. □

4 | Capacity region for special cases

For non-causal encoding and decoding, although the inner and outer bounds do not match in general, we characterize the strong coordination region in three special cases: when the channel is perfect; when the decoding is lossless; and when the random variables of the channel are independent from the random variables of the source. In all these cases, the achievability proof is merely a consequence of the general achievability proof of Theorem 3.10. The converse proofs, on the other hand, rely of the specifics of each setting, and are therefore different from each other.

The study of these particular cases allows us to derive some interesting considerations on the information-theoretic nature of coordination. First, observe that the empirical coordination region for these three settings was derived in [44]. In this section we recover the same decomposition of the joint distributions and the same information constraints as in [44], but we show that for strong coordination a positive rate of common randomness is also necessary. This reinforces the conjecture, stated in [25, Conjecture 1], that with enough common randomness the strong coordination capacity region is the same as the empirical coordination capacity region for any specific network setting.

Moreover, through the study of these special cases, in Section 4.4 we discuss whether the concatenation of channel codes and coordination codes is optimal, i.e. whether separation holds. Surprisingly, the intuition that empirical and strong coordination have the same properties fails on this matter and we prove that the separation principle does not hold for joint source-channel strong coordination.

Finally, perhaps one of the most interesting byproduct of strong coordination is the fact that in some particular cases, it offers security "for free". In Section 4.5 we explore this aspect. We observe that if the random variables of the channel are independent from the random variables of the source, we are imposing constraints such that an eavesdropper, even with perfect knowledge of the channel output, could not infer anything about the source and the reconstruction, provided that he has no access to common randomness [62]. These considerations lead to the notion of *secure strong coordination*, which combines strong coordination and strong secrecy.

4.1 Perfect channel

Instead of having a noisy link between the encoder and the decoder, suppose that there is a perfect channel between the two agents, so that they observe the same signal, as in Figure 4.1.

In this case $X^n = Y^n$, and the variable Z^n plays the role of side information at the decoder.

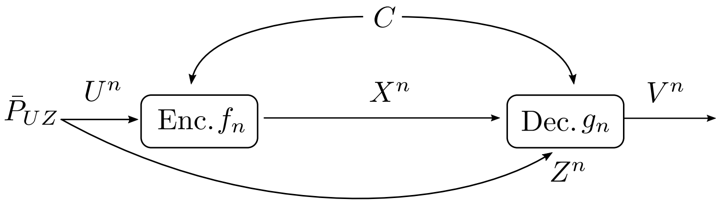

Figure 4.1: Coordination of signals and actions for a two-node network with a perfect channel.

The strong coordination region $\mathcal{R}_{\text{PC}}$ is characterized in the following result.

Theorem 4.1 *Suppose that $\bar{P}_{Y|XS}(\mathbf{y}|\mathbf{x},\mathbf{s}) = \mathbb{1}_{Y|X}(\mathbf{y}|\mathbf{x})$. Then the strong coordination region is*

$$\mathcal{R}_{\text{PC}} := \left\{ (\bar{P}_{UZXV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{UZXV} = \bar{P}_{UZ}\bar{P}_{X|U}\bar{P}_{V|UXZ} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UZWXV} = \bar{P}_{UZ}\bar{P}_{W|U}\bar{P}_{X|UW}\bar{P}_{V|WXZ} \\ I(WX;U) \leq H(X) + I(W;Z|X) \\ R_0 \geq I(W;UV|XZ) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{V}| + 4 \end{array} \right\}. \tag{4.1}$$

Remark 4.2 - *Markov chain decomposition*. Observe that the decomposition of the joint distributions $\bar{P}_{UZXV}$ and $\bar{P}_{UZWXV}$ is equivalently characterized in terms of Markov chains:

$$Z - U - X, \qquad \begin{cases} Z - U - (X,W), \\ V - (X,Z,W) - U. \end{cases} \tag{4.2}$$

4.1.1 Achievability

We show that $\mathcal{R}_{\text{PC}}$ is contained in the region $\mathcal{R}_{\text{state,in}}$ defined in (3.25) and thus it is achievable. We note $\mathcal{R}_{\text{state,in}}(W)$ the subset of $\mathcal{R}_{\text{state,in}}$, here specialized to the case without state and with a perfect channel, for a fixed $W \in \mathcal{W}$ that satisfies:

$$\begin{aligned} &\bar{P}_{UZWXV} = \bar{P}_{USZ}\bar{P}_{W|U}\bar{P}_{X|UW}\bar{P}_{V|WXZ}, \\ &I(W;U) \leq I(W;XZ), \\ &R_0 \geq I(W;UV|XZ). \end{aligned} \tag{4.3}$$

Then the set $\mathcal{R}_{\text{state,in}}$ is the union over all the possible choices for W that satisfy (4.3). Similarly, $\mathcal{R}_{\text{PC}}$ is the union of all $\mathcal{R}_{\text{PC}}(W)$ with W that satisfies

$$\begin{aligned} &\bar{P}_{UZWXV} = \bar{P}_{UZ}\bar{P}_{W|U}\bar{P}_{X|UW}\bar{P}_{V|WXZ}, \\ &I(W,X;U) \leq H(X) + I(W;Z|X), \\ &R_0 \geq I(W;UV|XZ). \end{aligned} \tag{4.4}$$

Let $(\bar{P}_{UZXV}, R_0) \in \mathcal{R}_{\text{PC}}(W)$ for some $W \in \mathcal{W}$. Then W verifies the Markov chains $Z - U - (X, W)$ and $V - WXZ - U$ and the information constraints for $\mathcal{R}_{\text{PC}}$. Note that $(\bar{P}_{UZXV}, R_0) \in \mathcal{R}_{\text{state,in}}(W')$, where $W' = (W, X)$. The Markov chains are still valid for W' and the information constraints in (4.4) imply the information constraints for $\mathcal{R}_{\text{state,in}}(W')$ since:

$$\begin{aligned} I(W'; U) &= I(W, X; U) \leq H(X) + I(W; Z|X) \\ &= I(W, X; X) + I(W, X; Z|X) = I(W, X; XZ) = I(W'; XZ), \\ R_0 &\geq I(W'; UV|XZ) = I(WX; UV|XZ). \end{aligned} \tag{4.5}$$

Then $(\bar{P}_{UZXV}, R_0) \in \mathcal{R}_{\text{state,in}}(W')$ and if we consider the union over all suitable W, we have

$$\bigcup_W \mathcal{R}_{\text{PC}}(W) \subseteq \bigcup_{(W,X)} \mathcal{R}_{\text{state,in}}(W, X) \subseteq \bigcup_W \mathcal{R}_{\text{state,in}}(W).$$

Finally, $\mathcal{R}_{\text{PC}} \subseteq \mathcal{R}_{\text{state,in}}$. □

4.1.2 Converse

Consider a code (f_n, g_n) that induces a distribution $P_{U^n Z^n X^n V^n}$ that is ε-close in total variational distance to the i.i.d. distribution $\bar{P}_{UZXV}^{\otimes n}$. Let T be the uniform random variable defined in Section 3.1.2.

We would like to prove that

$$0 \leq H(X) + I(W; Z|X) - I(W, X; U) = I(W, X; XZ) - I(W, X; U).$$

The following proof is inspired by [44]. We have

$$\begin{aligned} 0 &= H(X^n, Z^n) - I(X^n Z^n; U^n C) - H(X^n Z^n | U^n C) \\ &\overset{(a)}{\leq} \sum_{t=1}^n H(X_t, Z_t) - \sum_{t=1}^n I(X^n Z^n; U_t | U_{t+1}^n C) - H(X^n Z^n | U^n C) \\ &\overset{(b)}{=} \sum_{t=1}^n I(X^n Z^n C; X_t Z_t) - \sum_{t=1}^n I(X^n Z^n U_{t+1}^n C; U_t) + \sum_{t=1}^n I(U_{t+1}^n C; U_t) - H(X^n Z^n | U^n C) \\ &\overset{(c)}{=} \sum_{t=1}^n I(X^n Z^n C; X_t Z_t) - \sum_{t=1}^n I(X^n Z^n U_{t+1}^n C; U_t) - H(X^n Z^n | U^n C) \\ &\leq \sum_{t=1}^n I(X^n Z^n C; X_t Z_t) - \sum_{t=1}^n I(X^n Z^n C; U_t) - H(X^n Z^n | U^n C) \\ &\overset{(d)}{=} \sum_{t=1}^n I(X^n Z_{\sim t} C; X_t Z_t) + \sum_{t=1}^n I(Z_t; X_t Z_t | X^n Z_{\sim t} C) - \sum_{t=1}^n I(X^n Z_{\sim t} C; U_t) \\ &\quad - \sum_{t=1}^n I(Z_t; U_t | X^n Z_{\sim t} C) - H(X^n Z^n | U^n C) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{t=1}^{n} I(X^n Z_{\sim t} C; X_t Z_t) - \sum_{t=1}^{n} I(X^n Z_{\sim t} C; U_t) - H(X^n Z^n | U^n C) + \sum_{t=1}^{n} H(Z_t | X^n Z_{\sim t} C) \\
&\quad - \sum_{t=1}^{n} H(Z_t | X^n Z^n C) - \sum_{t=1}^{n} H(Z_t | X^n Z_{\sim t} C) + \sum_{t=1}^{n} H(Z_t | U_t X^n Z_{\sim t} C) \\
&= \sum_{t=1}^{n} I(X^n Z_{\sim t} C; X_t Z_t) - \sum_{t=1}^{n} I(X^n Z_{\sim t} C; U_t) + \sum_{t=1}^{n} H(Z_t | U_t X^n Z_{\sim t} C) - H(X^n Z^n | U^n C) \\
&\overset{(e)}{\leq} \sum_{t=1}^{n} I(X^n Z_{\sim t} C; X_t Z_t) - \sum_{t=1}^{n} I(X^n Z_{\sim t} C; U_t) + \sum_{t=1}^{n} H(Z_t | U_t C) - H(Z^n | U^n C) \\
&\overset{(f)}{=} \sum_{t=1}^{n} I(X^n Z_{\sim t} C; X_t Z_t) - \sum_{t=1}^{n} I(X^n Z_{\sim t} C; U_t) \\
&= n I(X^n Z_{\sim T} C; X_T Z_T | T) - n I(X^n Z_{\sim T} C; U_T | T) \\
&\leq n I(X^n Z_{\sim T} C T; X_T Z_T) - n I(X^n Z_{\sim T} C T; U_T) + n I(T; U_T) \\
&\overset{(g)}{=} n I(X_T X_{\sim T} Z_{\sim T} C T; X_T Z_T) - n I(X_T X_{\sim T} Z_{\sim T} C T; U_T)
\end{aligned}
$$

where (a) and (b) follow from the properties of the mutual information and (c) comes from the independence between U^n and C and the i.i.d. nature of the source. Then (d) comes from the chain rule, (e) from the properties of conditional entropy, (f) from the independence between (U^n, Z^n) and C and the i.i.d. nature of the source. Finally, (g) comes from the fact that $I(T; U_T)$ is zero due to the i.i.d. nature of the source.

We identify the auxiliary random variable W_t with $(C, X_{\sim t}, Z_{\sim t})$ for each $t \in [\![1, n]\!]$ and W with $(W_T, T) = (C, X_{\sim T}, Z_{\sim T}, T)$. Observe that with this identification of W the bound for R_0 follows from Section 3.2.2 with the substitution $Y = X$. Moreover, the following Markov chains are verified for each $t \in [\![1, n]\!]$:

$$
\begin{aligned}
&Z_t - U_t - (W_t, X_t), \\
&V_t - (W_t, X_t, Z_t) - U_t.
\end{aligned}
$$

The first one holds because the source is i.i.d. and Z_t does not belong to W_t. The second Markov chain follows from the fact that V is generated using C, X^n and Z^n that are included in $(W_t, X_t, Z_t) = (C, X_{\sim t}, Z_{\sim t}, X_t, Z_t)$. With a similar approach as in Section 3.1.2 and Section 3.2.2, the Markov chains with T hold. Then since $W = W_t$ when $T = t$, we also have $Z - U - (W, X)$ and $V - (W, X, Z) - U$. The cardinality bound is proved in Appendix B.5. □

Comparison with empirical coordination In [44, Section IV.A], the empirical coordination region is characterized with the same decomposition of the joint distributions and the information constraint $I(W, X; U) \leq H(X) + I(W; Z|X)$, the same as in (4.1).

4.2 Lossless decoder

Up to now, we refer to the random variable V as the *reconstruction* of the source U. This might be misleading, since V is actually generated according to the conditional distribution $\bar{P}_{V|UXYSZ}$ and it is consequently a stochastic function of the source, the signals, the state and side information. Here, we investigate a special case by considering when the decoder wants to reconstruct the source losslessly, i.e., $V = U$ as in Figure 4.2. The strong coordination region $\mathcal{R}_{\text{LD}}$ is characterized in the following result.

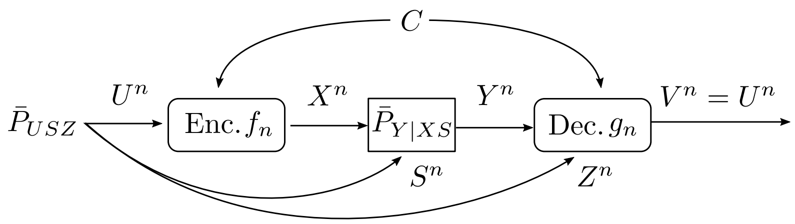

Figure 4.2: Coordination of signals and actions for a two-node network with a noisy channel and a lossless decoder.

Theorem 4.3 *Suppose that* $\bar{P}_{V|USXYZ}(\mathbf{v}|\mathbf{u},\mathbf{s},\mathbf{x},\mathbf{y},\mathbf{z}) = \mathbb{1}_{V|U}(\mathbf{v}|\mathbf{u})$. *Then the strong coordination region is*

$$\mathcal{R}_{\text{LD}} := \left\{ (\bar{P}_{USZXY}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{USZXYV} = \bar{P}_{USZ}\bar{P}_{X|U}\bar{P}_{Y|XS}\mathbb{1}_{V|U} \\ \exists\, W \textit{ taking values in } \mathcal{W} \\ \bar{P}_{USZWXYV} = \bar{P}_{USZ}\bar{P}_{W|U}\bar{P}_{X|UW}\bar{P}_{Y|XS}\mathbb{1}_{V|U} \\ I(W;U) \leq I(W;YZ) \\ R_0 \geq I(W;USX|YZ) \\ |\mathcal{W}| \leq |\mathcal{U}\times\mathcal{S}\times\mathcal{Z}\times\mathcal{X}\times\mathcal{Y}| + 3 \end{array} \right\}. \tag{4.6}$$

Remark 4.4 - *Markov chain decomposition*. Observe that the decomposition of the joint distributions $\bar{P}_{USZXYV}$ and $\bar{P}_{USZWXYV}$ is equivalently characterized in terms of Markov chains:

$$\begin{cases} Z - (U,S) - (X,Y), \\ Y - (X,S) - U, \end{cases} \qquad \begin{cases} Z - (U,S) - (X,Y,W), \\ Y - (X,S) - (U,W). \end{cases} \tag{4.7}$$

4.2.1 Achievability

We show that $\mathcal{R}_{\text{LD}} \subseteq \mathcal{R}_{\text{state,in}}$ and thus it is achievable. Similarly to the achievability proof in Theorem 4.1, let $(\bar{P}_{USZXYV}, R_0) \in \mathcal{R}_{\text{LD}}(W)$ for some $W \in \mathcal{W}$. Then, W verifies the Markov chains $Z - (U,S) - (X,Y,W)$ and $Y - (X,S) - (U,Z,W)$ and the information constraints for $\mathcal{R}_{\text{LD}}$. We want to show that $(\bar{P}_{USZXYV}, R_0) \in \mathcal{R}_{\text{state,in}}(W)$. Observe that the Markov chains are still valid. Hence, the only difference is the bound on R_0, but $I(W;USXV|YZ) = I(W;USX|YZ)$ when $U = V$. Then, $(\bar{P}_{USZXYV}, R_0) \in \mathcal{R}_{\text{state,in}}(W)$ and if we consider the union over all suitable W, we have

$$\bigcup_W \mathcal{R}_{\text{LD}}(W) \subseteq \bigcup_W \mathcal{R}_{\text{state,in}}(W).$$

Finally, $\mathcal{R}_{\text{LD}} \subseteq \mathcal{R}_{\text{state,in}}$. □

4.2.2 Converse

Consider a code (f_n, g_n) that induces a distribution $P_{U^n S^n Z^n X^n Y^n V^n}$ that is ε-close in total variational distance to the i.i.d. distribution $\bar{P}_{USZXY}^{\otimes n} \mathbb{1}_{V|U}^{\otimes n}$. Let T be the uniform random variable defined in Section 3.1.2.

We have

$$\begin{aligned}
nR_0 &= H(C) \geq H(C|Y^nZ^n) = H(CU^n|Y^nZ^n) - H(U^n|CY^nZ^n)\\
&\overset{(a)}{\geq} H(CU^n|Y^nZ^n) - nf(\varepsilon) \geq I(U^nS^nX^n;CU^n|Y^nZ^n) - nf(\varepsilon)\\
&= \sum_{t=1}^{n} I(U_tS_tX_t;CU^n|U^{t-1}S^{t-1}X^{t-1}Y_{\sim t}Z_{\sim t}Y_tZ_t) - nf(\varepsilon)\\
&\overset{(b)}{\geq} \sum_{t=1}^{n} I(U_tS_tX_t;CU^nY_{\sim t}Z_{\sim t}U^{t-1}S^{t-1}X^{t-1}|Y_tZ_t) - 2nf(\varepsilon)\\
&\geq \sum_{t=1}^{n} I(U_tS_tX_t;CU^nY_{\sim t}Z_{\sim t}|Y_tZ_t) - 2nf(\varepsilon)\\
&= nI(U_TS_TX_T;CU^nY_{\sim T}Z_{\sim T}|Y_TZ_TT) - 2nf(\varepsilon)\\
&= nI(U_TS_TX_T;CU^nY_{\sim T}Z_{\sim T}T|Y_TZ_T) - nI(U_TS_TX_T;T|Y_TZ_T) - 2nf(\varepsilon)\\
&= nI(U_TS_TX_T;CU^nY_{\sim T}Z_{\sim T}T|Y_TZ_T) - nI(U_TS_TX_TY_TZ_T;T) + nI(Y_TZ_T;T) - 2nf(\varepsilon)\\
&\overset{(c)}{\geq} nI(U_TS_TX_T;CU^nY_{\sim T}Z_{\sim T}T|Y_TZ_T) - 3nf(\varepsilon)
\end{aligned}$$

where (a) follows from Fano's Inequality which implies that

$$H(U^n|CY^nZ^n) \leq nf(\varepsilon) \tag{4.8}$$

as proved in Appendix B.2. To prove (b), observe that

$$\begin{aligned}
I(U_tS_tX_t;CU^n|U^{t-1}S^{t-1}X^{t-1}Y_{\sim t}Z_{\sim t}Y_tZ_t) =& I(U_tS_tX_t;CU^nY_{\sim t}Z_{\sim t}U^{t-1}S^{t-1}X^{t-1}|Y_tZ_t)\\
&- I(U_tS_tX_t;Y_{\sim t}Z_{\sim t}U^{t-1}S^{t-1}X^{t-1}|Y_tZ_t)
\end{aligned}$$

and $I(U_tS_tX_t;Y_{\sim t}Z_{\sim t}U^{t-1}S^{t-1}X^{t-1}|Y_tZ_t) \leq f(\varepsilon)$ by Lemma 2.18. Finally, (c) comes from the fact, proved in Lemma 2.16, that $I(U_TS_TX_TY_TZ_T;T)$ vanishes since the distribution is ε-close to i.i.d. by hypothesis. With the identifications $W_t = (C, U^n, Y_{\sim t}, Z_{\sim t})$ for each $t \in [\![1,n]\!]$ and $W = (W_T, T) = (C, U^n, Y_{\sim T}, Z_{\sim T}, T)$, we have $R_0 \geq I(W;USX|YZ)$.

For the second part of the converse, we have

$$
\begin{aligned}
nI(U;W) &\le nH(U) = H(U^n) = H(U^n|C) = I(U^n;Y^nZ^nC) + H(U^n|Y^nZ^nC)\\
&\overset{(d)}{\le} \sum_{t=1}^{n} I(U^n;Y_tZ_t|Y^{t-1}Z^{t-1}C) + nf(\varepsilon) \le \sum_{t=1}^{n} I(U^nY^{t-1}Z^{t-1}C;Y_tZ_t) + nf(\varepsilon)\\
&\le \sum_{t=1}^{n} I(U^nY_{\sim t}Z_{\sim t}C;Y_tZ_t) + nf(\varepsilon) = nI(U^nY_{\sim T}Z_{\sim T}C;Y_TZ_T|T) + nf(\varepsilon)\\
&\le nI(U^nY_{\sim T}Z_{\sim T}CT;Y_TZ_T) + nf(\varepsilon) \overset{(e)}{=} nI(W;YZ) + nf(\varepsilon)
\end{aligned}
$$

where (d) comes from Fano's Inequality and (e) comes from the identification

$$W = (C, U^n, Y_{\sim T}, Z_{\sim T}, T).$$

In order to complete the converse, we show that the following Markov chains hold for each $t \in [\![1,n]\!]$:

$$
\begin{aligned}
&Y_t - (X_t, S_t) - (U_t, Z_t, W_t),\\
&Z_t - (U_t, S_t) - (X_t, Y_t, W_t).
\end{aligned}
$$

The first one is verified because the channel is memoryless and Y_t does not belong to W_t and the second one holds because of the i.i.d. nature of the source and because Z_t does not belong to W_t. With a similar approach as in Section 3.1.2 and Section 3.2.2, the Markov chains with T hold. Then, since $W = W_t$ when $T = t$, we also have $Y - (X,S) - (U,Z,W)$ and $Z - (U,S) - (X,Y,W)$. The cardinality bound is proved in Appendix B.5. □

Comparison with empirical coordination An equivalent characterization of the region is:

$$
\mathcal{R}_{\text{LD}} := \left\{ (\bar{P}_{USZXY}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{USZXY} = \bar{P}_{USZ}\bar{P}_{X|U}\bar{P}_{Y|XS}\\ \exists\, W \text{ taking values in } \mathcal{W}\\ \bar{P}_{USZWXY} = \bar{P}_{USZ}\bar{P}_{W|U}\bar{P}_{X|UW}\bar{P}_{Y|XS}\\ H(U) \le I(WU;YZ)\\ R_0 \ge I(W;USX|YZ) + H(U|WYZ)\\ |\mathcal{W}| \le |\mathcal{U}\times\mathcal{S}\times\mathcal{Z}\times\mathcal{X}\times\mathcal{Y}| + 1 \end{array} \right\}. \tag{4.9}
$$

The region in (4.9) is achievable since with the choice of the auxiliary random variable $W'' = (W,U)$, the constraints in (4.6) become

$$
I(WU;U) = H(U) \le I(WU;YZ) \tag{4.10}
$$

$$
\begin{aligned}
R_0 &\ge I(WU;USX|YZ) = I(W;USX|YZ) + I(U;USX|WYZ)\\
&= I(W;USX|YZ) + H(U|WYZ) - H(U|USXWYZ) \qquad (4.11)\\
&= I(W;USX|YZ) + H(U|WYZ).
\end{aligned}
$$

Moreover, the converse in the proof of Theorem 4.3 is still valid with the identification

$$W = (C, U_{\sim T}, Y_{\sim T}, Z_{\sim T}, T).$$

Note that [44, Section IV.B] gives a characterization of the empirical coordination region, and the constraint for the mutual information is

$$0 \leq I(WU;YZ) - H(U) = I(WU;YZ) - H(U) - I(W;S|U)$$

which is the same as in (4.10) because of the Markov chain $(S, Z) - U - W$.

4.3 Independence between source and channel

In this section, we investigate the special case of separation between the random variables of the source and the random variables of the channel: suppose that the channel state $\bar{P}_{S^n}$ is independent of the source and of the side information $\bar{P}_{U^n Z^n}$, and that the target joint distribution is of the form $\bar{P}^{\otimes n}_{UZV} \bar{P}^{\otimes n}_{SXY}$. For simplicity, we will suppose that the encoder has perfect state information as in Figure 4.3.

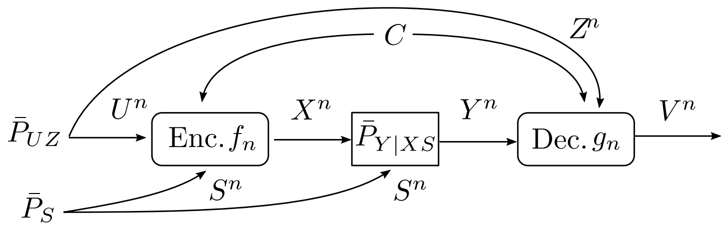

Figure 4.3: Coordination of signals and actions for a two-node network with a noisy channel where the source is separated from the channel and the encoder has perfect state information.

This scenario, for which we fully characterize the strong coordination region, is particularly interesting because it sheds some light on the nature of strong coordination. In fact, we use the results of this section to prove that the separation principle does not hold for strong coordination in Section 4.4, and to show the connection between strong coordination and secrecy in Section 4.5.

Note that in this case the coordination requirements are three-fold: the random variables (U^n, Z^n, V^n) should be coordinated, the random variables (S^n, X^n, Y^n) should be coordinated and finally (U^n, Z^n, V^n) should be independent of (S^n, X^n, Y^n). We introduce two auxiliary random variables W_1 and W_2, where W_2 is used to accomplish the coordination of (U^n, Z^n, V^n), while W_1 has the double role of ensuring the independence of source and state as well as coordinating (S^n, X^n, Y^n).

The strong coordination region $\mathcal{R}_{\text{SEP}}$ is characterized in the following result.

Theorem 4.5 *Suppose that $\bar{P}_{USXYZV} = \bar{P}_{UZV}\bar{P}_{SXY}$. Then, the strong coordination region is*

$$\mathcal{R}_{\text{SEP}} := \left\{ (\bar{P}_{USZXY}, R_0) \left| \begin{array}{l} \bar{P}_{USZXYV} = \bar{P}_{UZ}\bar{P}_{V|UZ}\bar{P}_S\bar{P}_{X|S}\bar{P}_{Y|XS} \\ \exists\, (W_1, W_2) \text{ taking values in } \mathcal{W}_1 \times \mathcal{W}_2 \\ \bar{P}_{USZW_1W_2XYV} = \bar{P}_{UZ}\bar{P}_{W_2|U}\bar{P}_{V|ZW_2}\bar{P}_S\bar{P}_{X|S}\bar{P}_{W_1|SX}\bar{P}_{Y|XS} \\ I(W_1;S) + I(W_2;U) \leq I(W_1;Y) + I(W_2;Z) \\ R_0 \geq I(W_1;SX|Y) + I(W_2;UV|Z) \\ (|\mathcal{W}_1|, |\mathcal{W}_2|) \leq |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| + 3. \end{array} \right. \right\}. \tag{4.12}$$

Remark 4.6 - *Markov chain decomposition*. Observe that the decomposition of the joint distribution $\bar{P}_{USZW_1W_2XYV}$ is equivalently characterized in terms of Markov chains:

$$\begin{cases} Z - U - W_2, \\ Y - (X,S) - W_1, \\ V - (Z, W_2) - U. \end{cases} \tag{4.13}$$

4.3.1 Achievability

We show that $\mathcal{R}_{\text{SEP}}$ is contained in the achievable region $\mathcal{R}_{\text{state,in}}$ in (3.25) specialized to this specific setting. In this case we are also supposing that the encoder has perfect state information, i.e. the input of the encoder is the pair (U^n, S^n) as in Figure 4.3 as well as common randomness C. Then, the joint distribution $\bar{P}_{USZXYV}$ in $R_{\text{state,in}}$ becomes $\bar{P}_{UZ}\bar{P}_{V|UZ}\bar{P}_S\bar{P}_{X|S}\bar{P}_{Y|XS}$ since (U, Z, V) is independent of (S, X, Y).

Observe that the set $\mathcal{R}_{\text{state,in}}$ is the union over all the possible choices for W that satisfy the joint distribution, rate and information constraints in (3.25). Similarly, $\mathcal{R}_{\text{SEP}}$ is the union of all $\mathcal{R}_{\text{SEP}}(W_1, W_2)$ with (W_1, W_2) that satisfies the joint distribution, rate and information constraints in (4.12). Let $(\bar{P}_{USZXY}, R_0) \in \mathcal{R}_{\text{SEP}}(W_1, W_2)$ for some (W_1, W_2) taking values in $\mathcal{W}_1 \times \mathcal{W}_2$. Then, (W_1, W_2) verifies the Markov chains $Z - U - W_2$, $V - (W_2, Z) - U$ and $Y - (S, X) - W_1$, and the information constraints for $\mathcal{R}_{\text{SEP}}$. We will show that $(\bar{P}_{USZXY}, R_0)$ belongs to $\mathcal{R}_{\text{state,in}}(W')$, where $W' = (W_1, W_2)$. The information constraints in (4.12) imply the information constraints for $\mathcal{R}_{\text{state,in}}(W')$ since:

$$\begin{aligned}
&I(W_1W_2;YZ) - I(W_1W_2;US) \\
&= I(W_1;YZ) + I(W_2;YZ|W_1) - I(W_1;US) - I(W_2;US|W_1) \\
&= I(W_1;Y) + I(W_2;YZW_1) - I(W_1;S) - I(W_2;USW_1) \\
&= I(W_1;Y) + I(W_2;Z) - I(W_1;S) - I(W_2;U) \geq 0, \\
&I(W_1W_2;USXV|YZ), \\
&= I(W_1;USXV|YZ) + I(W_2;USXV|YZW_1) \\
&= I(W_1;USXVZ|Y) + I(W_2;USXVY|ZW_1) \\
&= I(W_1;SX|Y) + I(W_2;USXVY|Z) \\
&= I(W_1;SX|Y) + I(W_2;UV|Z) \leq R_0,
\end{aligned}$$

because by construction W_1 and W_2 are independent of each other and W_1 is independent of (U, Z, V) and W_2 is independent of (S, X, Y). Then $(\bar{P}_{USZXY}, R_0) \in \mathcal{R}_{\text{state,in}}(W_1, W_2)$ and if we consider the union over all suitable (W_1, W_2), we have

$$\bigcup_{(W_1,W_2)} \mathcal{R}_{\text{SEP}}(W_1, W_2) \subseteq \bigcup_{(W_1,W_2)} \mathcal{R}_{\text{state,in}}(W_1, W_2) \subseteq \bigcup_{W} \mathcal{R}_{\text{state,in}}(W).$$

Finally, $\mathcal{R}_{\text{SEP}} \subseteq \mathcal{R}_{\text{state,in}}$. □

4.3.2 Converse

Let T be the uniform random variable defined in Section 3.1.2. Consider a code (f_n, g_n) that induces a distribution $P_{U^n S^n Z^n X^n Y^n V^n}$ that is ε-close in total variational distance to the i.i.d. distribution $\bar{P}_{UZV}^{\otimes n} \bar{P}_{SXY}^{\otimes n}$. Then, we have

$$\mathbb{V}(P_{S^n X^n Y^n Z^n U^n V^n}, \bar{P}_{UZV}^{\otimes n} \bar{P}_{SXY}^{\otimes n}) < \varepsilon.$$

If we apply Lemma 2.17 to $A = S^n X^n Y^n$ and $B = Z^n U^n V^n$, we have

$$I(S^n X^n Y^n; Z^n U^n V^n) < f(\varepsilon). \tag{4.14}$$

Then, we have

$$\begin{aligned}
nR_0 &= H(C) \overset{(a)}{\geq} I(U^n S^n X^n V^n; C|Y^n Z^n) = I(S^n X^n; C|Y^n Z^n U^n V^n) + I(U^n V^n; C|Y^n Z^n) \\
&= I(S^n X^n; CZ^n U^n V^n|Y^n) - I(S^n X^n; Z^n U^n V^n|Y^n) + I(U^n V^n; CY^n|Z^n) - I(U^n V^n; Y^n|Z^n) \\
&\overset{(b)}{\geq} I(S^n X^n; CZ^n U^n V^n|Y^n) + I(U^n V^n; CY^n|Z^n) - 2f(\varepsilon) \\
&\overset{(c)}{=} \sum_{t=1}^{n} I(S_t X_t; CZ^n U^n V^n|S_{t+1}^n X_{t+1}^n Y_t Y_{\sim t}) + \sum_{t=1}^{n} I(U_t V_t; CY^n|U^{t-1} V^{t-1} Z_t Z_{\sim t}) - 2f(\varepsilon) \\
&= \sum_{t=1}^{n} I(S_t X_t; CZ^n U^n V^n S_{t+1}^n X_{t+1}^n Y_{\sim t}|Y_t) - \sum_{t=1}^{n} I(S_t X_t; S_{t+1}^n X_{t+1}^n Y_{\sim t}|Y_t) \\
&\quad + \sum_{t=1}^{n} I(U_t V_t; CY^n U^{t-1} V^{t-1} Z_{\sim t}|Z_t) - \sum_{t=1}^{n} I(U_t V_t; U^{t-1} V^{t-1} Z_{\sim t}|Z_t) - 2nf(\varepsilon) \\
&\overset{(d)}{\geq} \sum_{t=1}^{n} I(S_t X_t; CZ^n U^n V^n S_{t+1}^n X_{t+1}^n Y_{\sim t}|Y_t) + \sum_{t=1}^{n} I(U_t V_t; CY^n U^{t-1} V^{t-1} Z_{\sim t}|Z_t) - 2f(\varepsilon) - 2nf(\varepsilon) \\
&= nI(S_T X_T; CU^n S_{T+1}^n Y_{\sim T}|Y_T T) + nI(U_T V_T; CY^n U^{T-1} V^{T-1} Z_{\sim T}|Z_T T) - 2(n+1)f(\varepsilon) \\
&\geq nI(S_T X_T; CU^n S_{T+1}^n Y^{T-1}|Y_T T) + nI(U_T V_T; CY^n U^{T-1} Z_{\sim T}|Z_T T) - 2(n+1)f(\varepsilon) \\
&= nI(S_T X_T; CU^n S_{T+1}^n Y^{T-1} T|Y_T) - nI(S_T X_T; T|Y_T) \\
&\quad + nI(U_T V_T; CY^n U^{T-1} Z_{\sim T} T|Z_T) - nI(U_T V_T; T|Z_T) - 2(n+1)f(\varepsilon) \\
&= nI(S_T X_T; CU^n S_{T+1}^n Y^{T-1} T|Y_T) - nI(S_T X_T Y_T; T) + nI(Y_T; T) \\
&\quad + nI(U_T V_T; CY^n U^{T-1} Z_{\sim T} T|Z_T) - nI(U_T V_T Z_T; T) + nI(Z_T; T) - 2(n+1)f(\varepsilon)
\end{aligned}$$

$$\overset{(e)}{\geq} nI(S_T X_T; CU^n S_{T+1}^n Y^{T-1} T | Y_T) + nI(U_T V_T; CY^n U^{T-1} Z_{\sim T} T | Z_T) - 2(2n+1) f(\varepsilon)$$

where (a) follows from basic properties of entropy and mutual information. To prove (b), note that

$$\begin{aligned} I(S^n X^n; Z^n U^n V^n | Y^n) &\leq I(S^n X^n Y^n; Z^n U^n V^n), \\ I(U^n V^n; Y^n | Z^n) &\leq I(S^n X^n Y^n; Z^n U^n V^n), \end{aligned}$$

and $I(S^n X^n Y^n; Z^n U^n V^n) < f(\varepsilon)$ by (4.14). Then (c) comes from the chain rule for mutual information, (d) follows from Lemma 2.18 and (e) from Lemma 2.16 since the distributions are close to i.i.d. by hypothesis. The lower bound on R_0 follows from the identifications

$$\begin{aligned} &W_{1,t} = (C, U^n, S_{t+1}^n, Y^{t-1}) && t \in [\![1, n]\!], \\ &W_{2,t} = (C, Y^n, U^{t-1}, Z_{\sim t}) && t \in [\![1, n]\!], \\ &W_1 = (W_{1,T}, T) = (C, U^n, S_{T+1}^n, Y^{T-1}, T), \\ &W_2 = (W_{2,T}, T) = (C, Y^n, U^{T-1}, Z_{\sim T}, T). \end{aligned}$$

Following the same approach as [44, 45], we divide the second part of the converse in two steps. First, we have the following upper bound:

$$\begin{aligned} I(CU^n; Y^n) &= \sum_{t=1}^n I(CU^n; Y_t | Y^{t-1}) \leq \sum_{t=1}^n I(CU^n Y^{t-1}; Y_t) \\ &= \sum_{t=1}^n I(CU^n Y^{t-1} S_{t+1}^n; Y_t) - \sum_{t=1}^n I(S_{t+1}^n; Y_t | CU^n Y^{t-1}) \\ &\overset{(f)}{=} \sum_{t=1}^n I(CU^n Y^{t-1} S_{t+1}^n; Y_t) - \sum_{t=1}^n I(S_t; Y^{t-1} | CU^n S_{t+1}^n) \\ &= \sum_{t=1}^n I(CU^n Y^{t-1} S_{t+1}^n; Y_t) - \sum_{t=1}^n I(S_t; CY^{t-1} U^n S_{t+1}^n) + \sum_{t=1}^n I(S_t; CU^n S_{t+1}^n) \\ &\overset{(g)}{=} \sum_{t=1}^n I(CU^n Y^{t-1} S_{t+1}^n; Y_t) - \sum_{t=1}^n I(S_t; CY^{t-1} U^n S_{t+1}^n) \\ &\overset{(h)}{=} \sum_{t=1}^n I(Y_t; W_{1,t}) - \sum_{t=1}^n I(S_t; W_{1,t}) \end{aligned} \tag{4.15}$$

where (f) comes from the Csiszár sum identity, (g) from the fact that $I(S_t; CU^n S_{t+1}^n)$ is zero because the source and the common randomness are independent of the state, which is i.i.d. by hypothesis. Finally, (h) comes from the identification of the auxiliary random variable $W_{1,t}$ for $t \in [\![1, n]\!]$.

Then, we show a lower bound:

$$I(CU^n; Y^n) \geq I(U^n; Y^n | C) \overset{(i)}{=} I(U^n; CY^n) \overset{(j)}{=} I(U^n Z^n; CY^n)$$

$$
\begin{aligned}
&\geq I(U^n; CY^n|Z^n) = \sum_{t=1}^{n} I(U_t; CY^n|Z^nU^{t-1}) \\
&= \sum_{t=1}^{n} I(U_t; CY^nZ_{\sim t}U^{t-1}|Z_t) - \sum_{t=1}^{n} I(U_t; Z_{\sim t}U^{t-1}|Z_t) \\
&\overset{(k)}{=} \sum_{t=1}^{n} I(U_t; CY^nZ_{\sim t}U^{t-1}|Z_t) \\
&= \sum_{t=1}^{n} I(U_tZ_t; CY^nZ_{\sim t}U^{t-1}) - \sum_{t=1}^{n} I(Z_t; CY^nZ_{\sim t}U^{t-1}) \\
&\overset{(l)}{=} \sum_{t=1}^{n} I(U_t; CY^nZ_{\sim t}U^{t-1}) - \sum_{t=1}^{n} I(Z_t; CY^nZ_{\sim t}U^{t-1}) \\
&\overset{(m)}{=} \sum_{t=1}^{n} I(U_t; W_{2,t}) - \sum_{t=1}^{n} I(Z_t; W_{2,t})
\end{aligned}
\tag{4.16}
$$

where (i) comes from the fact that $I(U^n; C)$ is zero because U^n and C are independent, (j) from the Markov chain $Z^n - U^n - Y^nC$, (k) from the fact that U^n and Z^n are i.i.d. by hypothesis, (l) follows from the the Markov chain $Z_t - U_t - (Y^n, Z_{\sim t}, U^{t-1}, C)$ for $t \in [\![1, n]\!]$ and finally (m) comes from the identification of the auxiliary random variable $W_{2,t}$ for $t \in [\![1, n]\!]$.

By combining upper and lower bound, we have

$$
\begin{aligned}
0 &\overset{(n)}{\leq} \sum_{t=1}^{n} I(Y_t, W_{1,t}) - \sum_{t=1}^{n} I(S_t; W_{1,t}) + \sum_{t=1}^{n} I(Z_t; W_{2,t}) - \sum_{t=1}^{n} I(U_t; W_{2,t}) \\
&= nI(Y_T, W_{1,T}|T) - nI(S_T; W_{1,T}|T) + nI(Z_T; W_{2,T}|T) - nI(U_T; W_{2,T}|T) \\
&\leq nI(Y_T, W_{1,T}T) - nI(S_T; W_{1,T}T) + nI(S_T; T) + nI(Z_T; W_{2,T}T) - nI(U_T; W_{2,T}T) + nI(U_T; T) \\
&\overset{(o)}{=} nI(Y_T, W_{1,T}T) - nI(S_T; W_{1,T}T) + nI(Z_T; W_{2,T}T) - nI(U_T; W_{2,T}T) \\
&\overset{(p)}{=} nI(Y; W_1) - nI(S; W_1) + nI(Z; W_2) - nI(U; W_2)
\end{aligned}
$$

where (n) comes from (4.15) and (4.16) and (o) follows from the i.i.d. nature of the source and state. Finally (p) follows from the identifications for W_1 and W_2.

With the chosen identification, the Markov chains are verified for each $t \in [\![1, n]\!]$:

$$
\begin{aligned}
&Y_t - (X_t, S_t) - W_{1,t} \\
&Z_t - U_t - W_{2,t} \\
&V_t - (W_{2,t}, Z_t) - U_t.
\end{aligned}
$$

The first Markov chain holds because the channel is memoryless and Y_t does not belong to $W_{1,t}$. The second one holds because Z^n is i.i.d. and Z_t does not belong to $W_{2,t}$. Finally, the third one is verified because the decoder is non-causal and V_t is a function of (Y^n, Z^n) that is included in $(W_{2,t}, Z_t) = (Y^n, U^{t-1}, Z_{\sim t}, Z_t)$. With a similar approach as in Section 3.1.2 and Section 3.2.2, the Markov chains with T hold. Then since $W_1 = W_{1,t}$ and $W_2 = W_{2,t}$ when $T = t$, we also have $Y - (X,S) - W_1$, $Z - U - W_2$ and $V - (W_2, Z) - U$. The cardinality bound is proved in Appendix B.5. □

Remark 4.7 - ***Correlation of the auxiliary random variables.*** Note that even if in the converse proof W_1 and W_2 are correlated, from them we can define two new variables W_1' and W_2' independent of each other, with the same marginal distributions $P_{W_1'SXY} = P_{W_1SXY}$ and $P_{W_2'UVZ} = P_{W_2UVZ}$, such that the joint distribution $P_{W_1'W_2'SXYUVZ}$ splits as $P_{W_1'SXY}P_{W_2'UVZ}$. Since we are supposing (U,V,Z) and (S,X,Y) independent of each other and the constraints only depend on the marginal distributions P_{W_1SXY} and P_{W_2UVZ}, the converse is still satisfied with the new auxiliary random variables W_1' and W_2'. Moreover the new variables still verify the cardinality bounds, since they also depend only on the marginal distributions, as shown in Appendix B.5.

Comparison with empirical coordination In the case of separation between the random variables of the source and the random variables of the channel, the empirical coordination region is characterized in [44, Section IV.C] with the same decomposition of the joint distributions and the same information constraint as in (4.12).

Special case $X = U$ Observe that for the three special cases considered up to now the general achievability scheme of Chapter 3 was optimal, while each converse relies on the specifics of the setting and has to be rederived for every region. Now, we consider the case where $X = U$. For simplicity, we do not take into account state or side information. In this case, the inner and outer bound of Theorem 3.3 become

$$\mathcal{R}_{X=U,in} := \left\{ (\bar{P}_{XYV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{XYV} = \bar{P}_X \bar{P}_{Y|X} \bar{P}_{V|XY} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{XYWV} = \bar{P}_X \bar{P}_{W|X} \bar{P}_{Y|X} \bar{P}_{V|WY} \\ I(W;X) = I(W;Y) \\ R_0 \geq I(W;XV|Y) \end{array} \right\},$$

$$\mathcal{R}_{X=U,out} := \left\{ (\bar{P}_{XYV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{XYV} = \bar{P}_X \bar{P}_{Y|X} \bar{P}_{V|UXY} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{XYWV} = \bar{P}_X \bar{P}_{W|X} \bar{P}_{Y|X} \bar{P}_{V|WY} \\ I(W;X) \leq I(X;Y) \\ R_0 \geq I(W;XV|Y) \\ |\mathcal{W}| \leq |\mathcal{X} \times \mathcal{Y} \times \mathcal{V}| + 4 \end{array} \right\}.$$

Observe that in $\mathcal{R}_{X=U,out}$ the mutual information $I(W;X)$ is upper bounded by the capacity of the channel. Moreover, the equality in the information constraint $I(W;X) = I(W;Y)$ in $\mathcal{R}_{in,X=U}$ comes from the intersection of the condition $I(W;X) \geq I(W;Y)$, implied by the the Markov chain $W - X - Y$, and the information constraint $I(W;X) \leq I(W;Y)$ of the achievability proof. Then, we have

$$I(W;XY) = I(W;Y) + I(W;X|Y) \overset{(a)}{=} I(W;X) + I(W;X|Y)$$

$$= I(W;X) + I(W;Y|X) \overset{(b)}{=} I(W;X)$$

where (a) comes from the equality $I(W;X) = I(W;Y)$ and (b) from $I(W;Y|X) = 0$, which is equivalent to the Markov chain $W - X - Y$. Thus, $I(W;X|Y) = 0$ and the Markov chain $W - Y - X$ holds. Then, if $\bar{P}_{XY}$ has full support, by [47, Lemma 1] the auxiliary random variable W is independent of X and Y and the region $\mathcal{R}_{X=U,in}$ becomes

$$\left\{(\bar{P}_{XYV}, R_0) \;\middle|\; \bar{P}_{XYV} = \bar{P}_X \bar{P}_{Y|X} \bar{P}_{V|Y}, R_0 \geq 0\right\}.$$

We observe that even in this rather simple example, it is still an open problem whether the inner and outer bound (or both of them) can be improved [1].

4.4 Is separation optimal?

Strong coordination over error-free channels was investigated in [25, 21]. When extending this analysis to noisy channels, it natural to ask whether some form of separation theorem holds between source and channel coordination. In fact, if the separation principle were still valid for strong coordination, by concatenating the strong coordination of the source and the strong coordination of the input and output of the channel we should retrieve the same mutual information and rate constraints. If that were the case, it would mean that a joint source-channel coordination version of Shannon's source-channel separation theorem [63] holds. Following the intuition provided by empirical coordination [46], it would be natural to think that separation holds. However, this is not the case for strong coordination, and we prove it in this section.

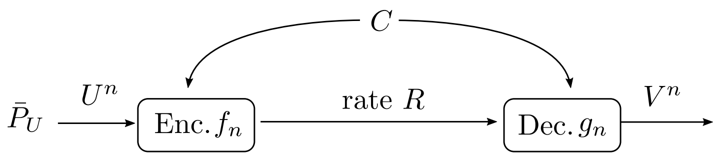

Figure 4.4: Coordination of the actions U^n and V^n for a two-node network with an error-free link of rate R.

In order to prove that separation does not hold, we first consider the optimal result for coordination of actions with error-free links in [25, 21] and then we compare it with our result on joint coordination of signals and actions. In particular, since we want to compare the result in [25, 21] with an exact region, we consider the case in which the channel is perfect and the target joint distribution is of the form $\bar{P}_{UV}^{\otimes n}\bar{P}_X^{\otimes n}$. The choice of a perfect channel might appear counterintuitive but it is motivated by the fact that we are trying to find a counterexample. As a matter of fact, if the separation principle holds for any noisy link, it should in particular hold for a perfect one.

We start by considering the two-node network with fixed source $\bar{P}_U$ and an error-free link of rate R (Figure 4.4). As anticipated in (2.2), for this setting [25, 21] characterize the strong

[1] The authors thank Michèle Wigger and Albert Guillén i Fàbregas for suggesting this example.

coordination region as

$$\mathcal{R}_{\text{Cuff}} := \left\{ (\bar{P}_{UV}, R, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{UV} = \bar{P}_U \bar{P}_{V|U} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UWV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{V|W} \\ R \geq I(U;W) \\ R + R_0 \geq I(UV;W) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{V}| + 1 \end{array} \right\}. \tag{4.17}$$

The result in [25, 21] characterizes the trade-off between the rate R_0 of available common randomness and the required description rate R for simulating a discrete memoryless channel for a fixed input distribution.

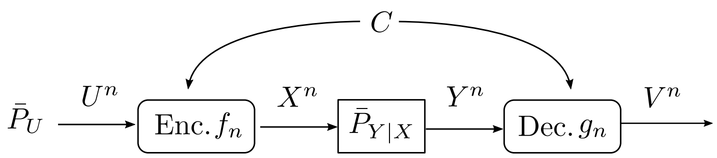

Figure 4.5: Two-node network with a noisy channel with non-causal encoder and decoder.

We consider, in the simpler scenario with no state and no side information of Figure 4.5, the intersection $\mathcal{R}_{UV\otimes X} := \mathcal{R}_{\text{PC}} \cap \mathcal{R}_{\text{SEP}}$. The following result, proved in Appendix B.3, characterizes the strong coordination region.

Proposition 4.8 *In a two-node network with comprised of an i.i.d. source and a noisy channel with non-causal encoder and decoder, suppose that $\bar{P}_{Y|X}(\mathbf{y}|\mathbf{x}) = \mathbb{1}_{Y|X}(\mathbf{y}|\mathbf{x})$ and $\bar{P}_{UXV} = \bar{P}_{UV}\bar{P}_X$. Then, the strong coordination region is*

$$\mathcal{R}_{UV\otimes X} := \left\{ (\bar{P}_{UXV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{UXV} = \bar{P}_U \bar{P}_{V|U} \bar{P}_X \\ \exists\, W \textit{ taking values in } \mathcal{W} \\ \bar{P}_{UWXV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{V|W} \bar{P}_X \\ I(W;U) \leq H(X) \\ R_0 \geq I(UV;W) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{V}| + 1 \end{array} \right\}. \tag{4.18}$$

To compare $\mathcal{R}_{\text{Cuff}}$ and $\mathcal{R}_{UV\otimes X}$, suppose that in the setting of Figure 4.4 we use a codebook to send a message to coordinate U^n and V^n. In order to do so we introduce an i.i.d. source X^n with distribution P_X over $\mathcal{X}$ in the model and we use the typical sequences of X^n as a codebook $\mathcal{C}$. Note that the codebook $\mathcal{C}$ can be seen as an optimal channel code for the perfect channel. Hence, asymptotically $R = H(X)$ and we rewrite the information constraints in (4.17) as

$$\begin{aligned} &H(X) \geq I(U;W), \\ &R_0 \geq I(UV;W) - H(X). \end{aligned}$$

This defines a new region:

$$\mathcal{R}_{\text{Cuff},H(X)} := \left\{ (\bar{P}_{UXV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{UV} = \bar{P}_U \bar{P}_{V|U} \\ X \text{ generated according to } \bar{P}_X \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UWV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{V|W} \bar{P}_X \\ I(W;U) \leq H(X) \\ R_0 \geq I(UV;W) - H(X) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{V}| + 1 \end{array} \right\}. \tag{4.19}$$

The distributions $\bar{P}_{UXV}$ in $\mathcal{R}_{\text{Cuff},H(X)}$ coordinate separately X^n and (U^n, V^n): in [21] the request is to induce a joint distribution $P_{U^nV^n}$ that is ε-close in total variational distance to the i.i.d. distribution $\bar{P}_{UV}^{\otimes n}$, and we have imposed that X^n is generated according to the uniform distribution.

Observe that, while the information constraint is the same in the two regions (4.19) and (4.18), the rate of common randomness R_0 required for strong coordination region in (4.18) is larger than the rate of common randomness in (4.17). In fact, in the setting of Figure 4.4 both X^n and the pair (U^n, V^n) achieve coordination separately (i.e. P_X^n is close to $\bar{P}_X^{\otimes n}$ and $P_{U^nV^n}$ is close to $\bar{P}_{UV}^{\otimes n}$ in total variational distance), but there is no extra constraint on the joint distribution $P_{U^nX^nV^n}$. On the other hand, the structure of our setting in (4.18) is different and requires the control of the joint distribution $P_{U^nX^nV^n}$ which has to be ε-close in total variational distance to the i.i.d. distribution $\bar{P}_{UV}^{\otimes n}\bar{P}_X^{\otimes n}$. Since we are imposing a more stringent constraint, it requires more common randomness.

Remark 4.9 - ***Interpretation of*** $\mathcal{R}_{UV\otimes X}$**.** We found $\mathcal{R}_{UV\otimes X}$ as the intersection of two regions, but we can give it the following interpretation starting from $\mathcal{R}_{\text{Cuff}}$. By identifying $R = H(X)$ in $\mathcal{R}_{\text{Cuff}}$, we find that the rate of common randomness has to be greater than $I(UV;W) - H(X)$. But this is not enough to ensure that X^n is independent of (U^n, V^n). In order to guarantee that, we apply a one-time pad on X^n (which requires an amount of fresh randomness equal to $H(X)$) and we have

$$R_0 \geq I(UV;W) - H(X) + H(X) = I(UV;W)$$

which is the condition on the rate of common randomness in (4.18).

Remark 4.10 - ***Comparison with previous results*****.** Note that separation holds for empirical coordination, as shown in [46]. For strong coordination of actions over noisy links, in [31] the authors derive an inner and an outer bound for the region and prove that separation in the sense of concatenating strong coordination of the actions and a good channel code is suboptimal.

Example 4.11 The difference in terms of rate of common randomness R_0 is better shown in an example: when separately coordinating the two blocks X^n and (U^n, V^n) without imposing a joint behavior $P_{U^nV^nX^n}$, the same bits of common randomness can be reused for both purposes, and the required rate R_0 is lower. We consider the case, already analyzed in Example

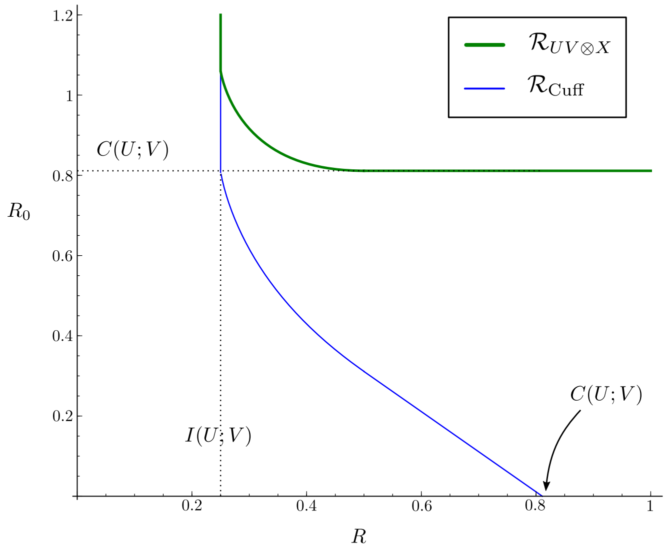

Figure 4.6: Comparison of the joint coordination region $\mathcal{R}_{UV\otimes X}$ with $\mathcal{R}_{\text{Cuff}}$ when $R = H(X)$ [23, 21]: boundaries of the regions for a binary erasure channel with erasure probability $p_e = 0.75$ and a Bernoulli-half input.

2.11, of a Bernoulli-half source U, and V which is an erasure with probability p_e and is equal to U otherwise. As we already proved, the optimal choice for the joint distributed P_{UWV} is the concatenation of two erasure channels $\bar{P}_{W|U}$ and $\bar{P}_{V|W}$ with erasure probability p_1 and p_2 respectively. Then, we recall that

$$p_2 \in [0, \min\{1/2; p_e\}], \quad p_1 = 1 - \tfrac{1-p_e}{1-p_2},$$
$$I(U;W) = 1 - p_1, \qquad I(UV;W) = h(p_e) + (1-p_1)(1-h(p_2)),$$

where h is the binary entropy function. Figure 4.6 shows the boundaries of the regions (4.17) (blue) and (4.18) (green) for $p_e = 0.75$ and a Bernoulli-half input. As anticipated in Example 2.11, the dotted bound $R \geq I(U;V)$ comes directly from combining $R \geq I(U;W)$ with the Markov chain $U - W - V$. At the other extreme, if $R_0 = 0$ in (4.17), $R + R_0 \geq I(UV;W) \geq C(U;V)$, where $C(U;V)$ is Wyner common information [21]. On the other hand, in the region defined in (4.18), $R_0 \geq I(UV;W) \geq C(U;V)$ for any value of $R = H(X)$. Moreover, note that as $R = H(X)$ tends to infinity, there is no constraint on the auxiliary random variable W (aside from the Markov chain $U - W - V$) and similarly to [37] the minimum rate of common randomness R_0 needed for strong coordination is Wyner common information $C(U;V)$. In particular to achieve joint strong coordination of (U, X, V) a positive rate of common randomness is required. The boundaries of the rate regions only coincide for the minimum rate $R = H(X)$, and $\mathcal{R}_{UV\otimes X}$ is strictly contained in $\mathcal{R}_{\text{Cuff}}$.

4.5 Coordination under secrecy constraints

Strong coordination requirements lead to synthesize joint distributions that are close to a desired joint distribution in total variational distance. Considering that the desired joint distribution is i.i.d. and that hypothesis tests will produce similar outcomes for distributions that are extremely close in total variation, the synthesized sequences are immune to statistical tests designed to detect i.i.d. correlated sequences, as mentioned in Remark 2.7. The above observations lead to applications for physical layer security. In this section we briefly discuss how in the separation setting of Section 4.3, strong coordination offers additional security guarantees “for free”. In many control settings, one would like the actions at various nodes to be independent of the communication so the actions cannot be anticipated by malicious eavesdroppers. In this context, common randomness is not only useful to coordinate signals and actions of the nodes but plays the role of a secret key shared between the two legitimate users.

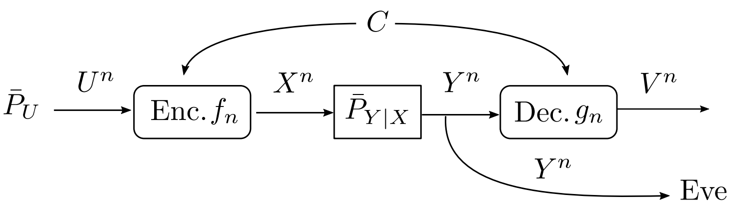

Figure 4.7: Wiretap channel: strong coordination implies secrecy.

For simplicity, we do not consider channel state and side information at the decoder. Suppose there is an eavesdropper who observes the output signals of the noisy channel. We show that not knowing the common randomness, Eve, the eavesdropper, cannot infer any information about the actions. More precisely, we want the joint distribution induced by the code $P_{U^nV^nY^n}$ to satisfy the *strong secrecy condition* [8]:

$$\lim_{n\to\infty} \mathbb{D}(P_{U^nV^nY^n} \| P_{U^nV^n}P_{Y^n}) = \lim_{n\to\infty} I(U^nV^n; Y^n) = 0 \tag{4.20}$$

while strongly coordinating (U^n, V^n):

$$\lim_{n\to\infty} \mathbb{V}(P_{U^nV^n}, \bar{P}_{UV}^{\otimes n}) = 0.$$

Corollary 4.12 *Suppose that there is an eavesdropper that receives the same sequence Y^n as the decoder but has no knowledge of the common randomness. Then, there exists a sequence (f_n, g_n) of strong coordination codes achieving the pair $(\bar{P}_{UV}\bar{P}_{XY}, R_0) \in \mathcal{R}_{SEP}$ such that the induced joint distribution $P_{U^nV^nX^nY^n}$ satisfies the strong secrecy condition:*

$$\lim_{n\to\infty} \mathbb{D}(P_{U^nV^nY^n} \| P_{U^nV^n}P_{Y^n}) = \lim_{n\to\infty} I(U^nV^n; Y^n) = 0.$$

Proof. Observe that in this setting the target joint distribution is of the form $\bar{P}_{UV}^{\otimes n}\bar{P}_{XY}^{\otimes n}$. Therefore achieving strong coordination means that $\mathbb{V}(P_{U^nV^nY^n}, \bar{P}_{UV}^{\otimes n}\bar{P}_{Y}^{\otimes n})$ vanishes. By the upper bound on the mutual information in Lemma A.15, we have secrecy if $\mathbb{V}(P_{U^nV^nY^n}, \bar{P}_{UV}^{\otimes n}\bar{P}_{Y}^{\otimes n})$ goes to zero exponentially. But we have proved in Section 3.1.1 that there exists a sequence of codes such that $\mathbb{V}(\bar{P}_{U^nX^nY^nV^n}, P_{U^nX^nY^nV^n})$ goes to zero exponentially. Hence, so does $\mathbb{V}(P_{U^nV^nY^n}, \bar{P}_{UV}^{\otimes n}\bar{P}_{Y}^{\otimes n})$. □

4.5.1 Secure strong coordination region

Inspired by the result of Corollary 4.12, we want to understand the interplay between the strong coordination of the actions U^n and V^n and secrecy. We define secure strong coordination as follows.

Definition 4.13 *A pair $(\bar{P}_{UV}, R_0)$ is achievable for secure strong coordination if there exists a sequence (f_n, g_n) of encoders-decoders with rate of common randomness R_0, such that*

$$\lim_{n\to\infty} \mathbb{V}\left(P_{U^nV^n}, \bar{P}_{UV}^{\otimes n}\right) = 0, \tag{4.21}$$

$$\lim_{n\to\infty} \mathbb{D}(P_{U^nV^nY^n} \| P_{U^nV^n} P_{Y^n}) = \lim_{n\to\infty} I(U^nV^n; Y^n) = 0, \tag{4.22}$$

where $P_{U^nX^nY^nV^n}$ is the joint distribution induced by the code. The secure strong coordination region $\mathcal{S}$ is the closure of the set of achievable pairs $(\bar{P}_{UV}, R_0)$.

Remark 4.14 - ***No coordination for the channel input.*** Note that here we are not requiring the coordination of the signals X^n.

In the setting of Figure 4.7, the following result fully characterizes the secure strong coordination region.

Theorem 4.15 *The secure strong coordination region is*

$$\mathcal{S} := \left\{ (\bar{P}_{UV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{UV} = \bar{P}_U \bar{P}_{V|U} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UWV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{V|W} \\ I(W;U) \leq \max_{\bar{P}_X} I(X;Y) \\ R_0 \geq I(UV;W) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{V}| + 1 \end{array} \right\}. \tag{4.23}$$

Comparison with cascade channel synthesis Although we derived the region $\mathcal{S}$ of (4.23) autonomously, it is possible to deduce it as a special case of strong coordination in a cascade network with secrecy constraints [62, Theorem 1]. In [62], the authors consider the three-nodes cascade network of Figure 4.8: the i.i.d. sequence U^n is given by nature, messages M_1 and M_2 are sent along noiseless links at rates R_1 and R_2 respectively. The three nodes share a source of common randomness C of rate R_0.

The coordination region for the cascade network of Figure 4.8 is characterized in [62, Theorem 1]: the sequence (U^n, V^n, Z^n) has to be i.i.d. correlated and independent of the messages (M_1, M_2):

$$\mathbb{V}(P_{U^nV^nZ^nM_1M_2}, P_{M_1M_2}\bar{P}_{UVZ}^{\otimes n}) \to 0. \tag{4.24}$$

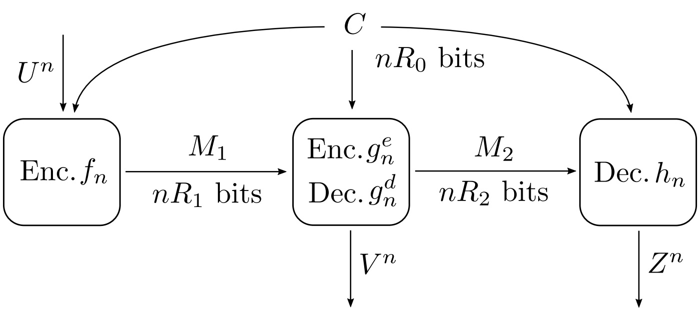

Figure 4.8: Three-nodes cascade network

The strong coordination region under security constraints in this setting is

$$\mathcal{R}_{\text{cascade}} := \left\{ (\bar{P}_{UVZ}, R_0, R_1, R_2) \left| \begin{array}{l} \bar{P}_{UVZ} = \bar{P}_U \bar{P}_{VZ|U} \\ \exists\, (W_1, W_2) \text{ taking values in } \mathcal{W}_1 \times \mathcal{W}_2 \\ U - (W_1, W_2) - V \\ (U, V, W_1) - W_2 - Z \\ R_0 \geq I(UVZ; W_1 W_2) \\ R_1 \geq I(U; W_1 W_2) \\ R_2 \geq I(U; W_2) \\ |\mathcal{W}_1| \leq |\mathcal{U} \times \mathcal{V} \times \mathcal{Z}| + 3 \\ |\mathcal{W}_2| \leq |\mathcal{U} \times \mathcal{V} \times \mathcal{Z} \times \mathcal{W}_1| + 3 \end{array} \right. \right\}. \tag{4.25}$$

Remark 4.16 The problem of finding the strong coordination region for the cascade setting is still open, but under the secrecy constraints, the region is easier to derive.

Notice that, if we merge the second and third node by identifying

$$\begin{aligned} &V = Z, \\ &W_1 = W_2 = W, \\ &R_1 = R_2 = R, \end{aligned}$$

the region in (4.23) can be derived as a special case of [62, Theorem 1]. First, with the above identifications, (4.25) becomes:

$$\mathcal{R}'_{\text{cascade}} := \left\{ (\bar{P}_{UV}, R_0, R) \left| \begin{array}{l} \bar{P}_{UV} = \bar{P}_U \bar{P}_{VZ|U} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ U - W - V \\ R_0 \geq I(UV; W) \\ R \geq I(U; W) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{V}| + 1 \end{array} \right. \right\}. \tag{4.26}$$

Second, we concatenate this result with a good channel code. We suppose that the encoder generates X^n and sends a message (M_1 or M_2) through a channel of rate R, which has output Y^n. If we send the message using a good channel code, we can take $R = \max_{P_X} I(X;Y)$ in (4.26), and we recover exactly the region $\mathcal{S}$ defined in (4.23). However, the achievability proof of [62, Theorem 1] involves soft covering instead of random binning techniques and is slightly more complicated.

Proof of Theorem 4.15: achievability The achievability of Theorem 4.15 uses similar random binning techniques as the ones of Theorem 3.3 and is presented in detail in Appendix B.4.1.

Proof of Theorem 4.15: converse Consider a code (f_n, g_n) that induces a distribution $P^{\text{RC}}_{U^n V^n}$ that is ε-close in total variational distance to the i.i.d. distribution $\bar{P}^{\otimes n}_{UV}$ and such that $I(U^n V^n; Y^n) \leq \varepsilon$. Let the random variable T be uniformly distributed over the set $[\![1, n]\!]$ and independent of the sequence (U^n, X^n, Y^n, V^n, C).

Then, we have

$$\begin{aligned}
nR_0 = H(C) &\overset{(a)}{\geq} I(U^n V^n; C|Y^n) = I(U^n V^n; CY^n) - I(U^n V^n; Y^n) \\
&\overset{(b)}{\geq} I(U^n V^n; CY^n) - f(\varepsilon) = \sum_{t=1}^{n} I(U_t V_t; CY^n | U^{t-1} V^{t-1}) - f(\varepsilon) \\
&= \sum_{t=1}^{n} I(U_t V_t; CY^n U^{t-1} V^{t-1}) - \sum_{t=1}^{n} I(U_t V_t; U^{t-1} V^{t-1}) - f(\varepsilon) \\
&\overset{(c)}{\geq} \sum_{t=1}^{n} I(U_t V_t; CY^n U^{t-1} V^{t-1}) - (n+1) f(\varepsilon) \geq \sum_{t=1}^{n} I(U_t V_t; CY^n) - (n+1) f(\varepsilon) \\
&= nI(U_T V_T; CY^n | T) - (n+1) f(\varepsilon) = nI(U_T V_T; CY^n T) - nI(U_T V_T; T) - (n+1) f(\varepsilon) \\
&\overset{(d)}{\geq} nI(U_T V_T; CY^n T) - (2n+1) f(\varepsilon)
\end{aligned} \tag{4.27}$$

where (a) follows from basic properties of entropy and mutual information and (b) from the fact that $I(U^n V^n; Y^n) \leq \varepsilon$ because of the secrecy conditions (4.22) and from Pinsker's inequality. Finally, since the distribution $P_{U^n V^n}$ is close to i.i.d. by hypothesis, (c) and (d) come from Lemma 2.17 and [23, Lemma VI.3] respectively.

For the second part of the converse, observe that

$$\begin{aligned}
0 &\overset{(e)}{\leq} I(X^n; Y^n) - I(Y^n; U^n C) \leq I(X^n; Y^n) - I(Y^n; U^n | C) \\
&\overset{(f)}{\leq} n \max_{\bar{P}_X} I(X; Y) - \sum_{t=1}^{n} I(Y^n; U_t | U^{t-1} C) \\
&= n \max_{\bar{P}_X} I(X; Y) - \sum_{t=1}^{n} I(Y^n U^{t-1} C; U_t) + \sum_{t=1}^{n} I(U^{t-1} C; U_t) \\
&\overset{(g)}{=} n \max_{\bar{P}_X} I(X; Y) - \sum_{t=1}^{n} I(Y^n U^{t-1} C; U_t) \leq n \max_{\bar{P}_X} I(X; Y) - \sum_{t=1}^{n} I(Y^n C; U_t) \\
&= n \max_{\bar{P}_X} I(X; Y) - nI(Y^n C; U_T | T) = n \max_{\bar{P}_X} I(X; Y) - nI(Y^n CT; U_T) + nI(T; U_T)
\end{aligned} \tag{4.28}$$

$$\overset{(h)}{=} n \max_{\bar{P}_X} I(X;Y) - nI(Y^n CT;U).$$

where (e) follows from the Markov chain $(U^n, C) - X^n - Y^n$, (f) from Lemma A.11 which proves that the capacity per transmission is not increased if we use a discrete memoryless channel many times. Finally, (g) and (h) come from the i.i.d. nature of the source $\bar{P}_U$ and the independence of the source from the common randomness.

Then, we conclude by identifying the auxiliary random variable W_t with (C, Y^n) for each $t \in [\![1, n]\!]$ and W with $(W_T, T) = (C, Y^n, T)$. □

A more general model Now suppose that Eve, the eavesdropper, observes the signal Z^n sent over the channel $\bar{P}_{Z|X}$ as in Figure 4.9. We want the induced joint distribution $P_{U^nV^nZ^n}$ to satisfy the strong secrecy condition [8] while strongly coordinating (U^n, V^n):

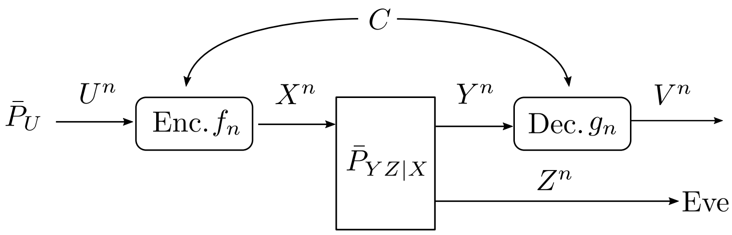

Figure 4.9: Wiretap channel: the eavesdropper observes Z^n.

Definition 4.17 *A pair $(\bar{P}_{UV}, R_0)$ is achievable for secure strong coordination if there exists a sequence (f_n, g_n) of encoders-decoders with rate of common randomness R_0, such that*

$$\lim_{n\to\infty} \mathbb{V}\left(P_{U^nV^n}, \bar{P}_{UV}^{\otimes n}\right) = 0, \tag{4.29}$$

$$\lim_{n\to\infty} \mathbb{D}(P_{U^nV^nZ^n} \| P_{U^nV^n}P_{Z^n}) = \lim_{n\to\infty} I(U^nV^n; Z^n) = 0, \tag{4.30}$$

where $P_{U^nX^nY^nZ^nV^n}$ is the joint distribution induced by the code. The secure strong coordination region $\mathcal{S}_{\mathcal{Z}}$ is the closure of the set of achievable pairs $(\bar{P}_{UV}, R_0)$.

We have the following inner bound, proved in Appendix B.4.2.

Proposition 4.18 *An inner bound for the secure strong coordination region is*

$$\mathcal{S}_{Z,in} := \left\{ (\bar{P}_{UV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{UV} = \bar{P}_U \bar{P}_{V|U} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UWV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{V|W} \\ \exists\, \bar{P}_X \quad I(W;U) \leq I(X;Y) \\ R_0 \geq I(UV;W) + (I(X;Z) - I(X;Y)) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{V}| + 1 \end{array} \right\}. \tag{4.31}$$

Remark 4.19 - *Degraded channel.* Suppose the eavesdropper observes a degraded version of the signal obtained by the legitimate receiver as in Figure 4.10: $\bar{P}_{YZ|X} = \bar{P}_{Y|X}\bar{P}_{Z|Y}$.

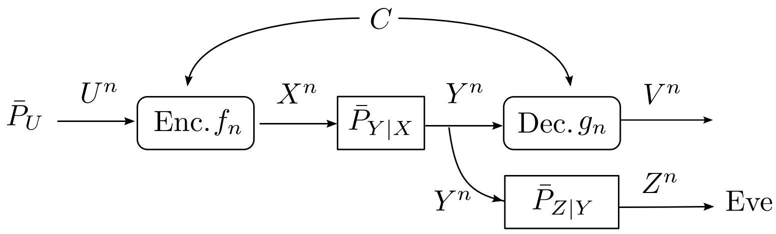

Figure 4.10: Degraded wiretap channel.

In this case $C_{\text{S}} := \max_{\bar{P}_X} (I(X;Y) - I(X;Z))$ is the secrecy capacity of a degraded wiretap channel [8, Corollary 3.1] . If we also suppose all the channels to be symmetric, the uniform distribution $\bar{P}_X(x) = 1/|\mathcal{X}|$, $\forall x \in \mathcal{X}$, maximizes the mutual information $I(X;Y)$ as well as the difference $I(X;Y) - I(X;Z)$. In this case, the inner bound in (4.31) becomes:

$$\mathcal{S}_{\text{in, deg}} := \left\{ (\bar{P}_{UV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{UV} = \bar{P}_U \bar{P}_{V|U} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UWV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{V|W} \\ I(W;U) \le \max_{\bar{P}_X} I(X;Y) \\ R_0 \ge I(UV;W) - C_{\text{S}} \\ |\mathcal{W}| \le |\mathcal{U} \times \mathcal{V}| + 1 \end{array} \right\}. \tag{4.32}$$

Note that we can give the following interpretation of $\mathcal{S}_{\text{in, deg}}$: it is possible to save on common randomness by generating random bits and sending them securely over the wiretap channel.

5 | Explicit schemes for coordination: polar codes

Although our achievability results in Chapter 3 and 4 shed some light on the fundamental limits of coordination over noisy channels, the problem of designing practical codes for strong coordination in this setting is still open. Then, the objective of this chapter is to provide a constructive counterpart to the theoretical research on coordination. Specifically, we investigate the design of explicit codes for strong coordination, and, because of their strong theoretical properties and of the analogy with random binning, we choose polar codes for our construction.

Note that polar codes have already been proposed for coordination in other settings: [5] proposes polar coding schemes for point-to-point empirical coordination with error-free links and uniform actions, while [15, 13] generalize the polar coding scheme to the case of non-uniform actions. Polar coding for strong point-to-point coordination has been presented in [11, 15, 13]. In [56] the authors construct a joint coordination-channel polar coding scheme for strong coordination of actions.

In this chapter, we focus on designing codes that allow joint strong coordination of signals and actions over noisy channels, and we prove in particular that the inner bound of Theorem 3.3 is achievable with polar codes.

We dedicate Section 5.1 to summarize the results and properties of source polarization [1, 2], and in Section 5.2 we introduce the polar coding properties which represent the equivalents of the random binning properties of Section 2.4. Then, in Section 5.3 we present an explicit scheme for strong coordination in the simplified scenario of no state and no side information. A polar coding scheme for empirical coordination for the same setting is detailed in Appendix C.3.1.

5.1 Source polarization

Polar codes, introduced by Arıkan in [1], provide the first deterministic construction of capacity-achieving codes for any binary symmetric discrete memoryless channel (B-DMC). In [1], Arıkan applies a simple linear transform to the channel inputs before transmission and a successive cancellation decoder at the output. The idea of polar codes is based on the recursive repetition of the same linear transform in order to obtain, from n independent copies of a given B-DMC W, a second set of n binary-input channels such that, as n becomes large, these effective channels tend towards either a completely noisy channel or an error-free channel with the fraction of error-free channels approaching the capacity of W.

In this thesis we are interested in the notion of *source polarization*, introduced by Arıkan in [2], which complements *channel polarization*. Let (X, Y) generated according to P_{XY} be an arbitrary pair of random variables over $\mathcal{X} \times \mathcal{Y}$ with $\mathcal{X} = \{0, 1\}$ and $\mathcal{Y}$ an arbitrary finite set. The pair (X, Y) constitutes a discrete memoryless source (DMS), with X as the part to be compressed and Y in the role of side-information about X. Shannon's lossless source coding theorem states that an encoder can compress (X^n, Y^n) into a codeword of length roughly $nH(X|Y)$ bits so that a decoder observing the codeword and Y^n can recover X^n reliably, provided n is sufficiently large. In [2], Arıkan describes a method based on polarization that achieves this compression bound.

Consider n independent copies of the source $(X^n, Y^n) = \{(X_i, Y_i)\}_{i=1}^n$, where $n = 2^m$, $m \geq 1$. We note $Z^n := X^n G_n$ the *polarization* of X^n, where the polarization transform G_n is defined as

$$G_n := \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes m} B_n.$$

The symbol $\otimes$ denotes the Kronecker product, and B_n is the "bit-reversal" permutation defined in [1] as

$$B_n := R_n(I_2 \otimes R_{n/2})(I_4 \otimes R_{n/4}) \dots (I_{n/2} \otimes R_2)$$

where $R_n : \mathcal{A}^n \to \mathcal{A}^n$ is the permutation operator defined as follows

$$R_n(a_i) := \begin{cases} a_{2i-1} & i \in [\![1, n/2]\!], \\ a_{2i-n} & i \in [\![n/2+1, n]\!]. \end{cases}$$

Note that G_n is invertible and $G_n = G_n^{-1}$.

The main result on source polarization for binary alphabets $\mathcal{X} = \{0, 1\}$ is the following.

Theorem 5.1 - ***Source polarization [2, Theorem 1]*** *Let (X, Y) be a DMS. For any $n = 2^m$, $m \geq 1$, let $Z^n = X^n G_n$. Then, for any $\delta \in (0, 1)$, as n tends to infinity*

$$\frac{|\{i \in [\![1, n]\!] \mid H(Z_i|Y^n Z^{i-1}) \in (1-\delta, 1]\}|}{n} \to H(X|Y),$$
$$\frac{|\{i \in [\![1, n]\!] \mid H(Z_i|Y^n Z^{i-1}) \in [0, \delta)\}|}{n} \to 1 - H(X|Y).$$

Now, we introduce the following notation.

Definition 5.2 - ***Very high and high entropy sets*** *For some $0 < \beta < 1/2$, let $\delta_n = 2^{-n^\beta}$ and define the very high and high entropy sets:*

$$\begin{aligned} \mathcal{V}_{X|Y} &:= \{j \in [\![1, n]\!] \mid H(Z_j|Z^{j-1} Y^n) > 1 - \delta_n\} && \textit{very high entropy set,} \\ \mathcal{H}_{X|Y} &:= \{j \in [\![1, n]\!] \mid H(Z_j|Z^{j-1} Y^n) > \delta_n\} && \textit{high entropy set.} \end{aligned} \tag{5.1}$$

The cardinality of the very high and high entropy set bits is characterized in the following result.

Lemma 5.3 - ***Cardinality of the very high and high entropy sets [2, 14]*** *The very high entropy and high entropy sets have the following properties:*

$$\lim_{n\to\infty} \frac{|\mathcal{V}_{X|Y}|}{n} = H(X|Y), \qquad \text{[14, Lemma 1]}$$
$$\lim_{n\to\infty} \frac{|\mathcal{H}_{X|Y}|}{n} = H(X|Y), \qquad \text{[2, Theorem 1]}$$
$$\lim_{n\to\infty} \frac{|\mathcal{H}^c_{X|Y}|}{n} = 1 - H(X|Y). \quad \text{[2, Theorem 1]}$$

One consequence of source polarization is the fact that it is possible to compress the source X^n using Y^n as side information by computing $Z^n = X^n G_n$ and selecting its high entropy bits $Z^n[\mathcal{H}_{X|Y}]$. We recall the definition of successive cancellation decoder.

Definition 5.4 - ***Successive cancellation decoding [2, 61]*** *Let (X, Y) be a DMS and let $Z^n = X^n G_n$. Having received $Z[\mathcal{H}_{X|Y}]$ and knowing Y^n the decoder sequentially builds an estimate $\hat{Z}^n$ as follows*

$$\hat{Z}_i = \begin{cases} Z_i & \text{if } i \in \mathcal{H}_{X|Y}, \\ 0 & \text{if } i \in \mathcal{H}^c_{X|Y} \text{ and } L_n(Y^n, Z^{i-1}) \geq 1, \\ 1 & \text{else}, \end{cases}$$

where L_n is the likelihood ratio defined as

$$L_n(Y^n, Z^{i-1}) := \frac{\bar{P}_{Z_i|Z^{i-1}Y^n}(0|\hat{Z}^{i-1}Y^n)}{\bar{P}_{Z_i|Z^{i-1}Y^n}(1|\hat{Z}^{i-1}Y^n)}.$$

Then, the decoder computes $\hat{X}^n = \hat{Z}^n G_n$.

Theorem 5.5 - ***Error probability for source polar coding [2, 61]*** *For any fixed rate $R > H(X|Y)$ and $\beta < 1/2$, the probability of error for the successive cancellation decoding method for polar source coding is bounded as $p_e = O(2^{-n^\beta})$.*

Moreover, a number of works employs the successive cancellation operation at the encoder side to generate a realization of X^n given Y^n and $Z[\mathcal{H}_{X|Y}]$ using a stochastic encoder rather than a hard decision based on the likelihood ratio [34, 15, 13]. Here, we consider the definition of [15, 13].

Definition 5.6 - ***Successive cancellation encoding [15, 13]*** *Let (X, Y) be a DMS and let $Z^n = X^n G_n$. Given Y^n and $Z[\mathcal{H}_{X|Y}]$, the encoder sequentially builds an estimate $\hat{Z}^n$ of Z^n generated* *according to the conditional distribution*

$$P_{\hat{Z}_i|\hat{Z}^{i-1}Y^n} = \begin{cases} \mathbb{1}_{\hat{Z}_i|Z_i} & \text{if } i \in \mathcal{V}_{X|Y}, \\ P_{Z_i|Z^{i-1}Y^n} & \text{if } i \in \mathcal{V}_{X|Y}^c. \end{cases}$$

5.2 From random binning to polar codes

Our random binning achievability proofs are based on proving that the random binning scheme we defined is close in total variational distance to a random coding scheme, and this relies on two results: Lemma 2.19 and Lemma 2.20. Inspired by [12, 14], the key idea to translate a random binning achievability proof into a polar coding achievability proof is to use the following results.

The following result, adapted from [12, Lemma 3], is a consequence of Theorem 5.1 and Theorem 5.5.

Lemma 5.7 - *Source coding with side information using polar codes* *Consider a discrete memoryless source (A^n, B^n) generated i.i.d. according to P_{AB}, where A^n is the part to be compressed and B^n is side information. Suppose $|\mathcal{A}| = \{0, 1\}$. We denote $Z^n := A^n G_n$ the polarization of A^n. We consider the high entropy set $\mathcal{H}_{A|B}$ and the restriction of Z^n to the very high entropy bits $Z[\mathcal{H}_{A|B}]$. For every $j \in [\![1, n]\!]$, we generate $\hat{Z}^n$ as in Definition 5.4 and compute $\hat{A}^n := \hat{Z}^n G_n$. Then,*

$$\mathbb{P}\{\hat{A}^n \neq A^n\} \leq \delta_n.$$

Note that the high entropy bits in positions $\mathcal{H}_{A|B}$ play the same role as the random binning index in Lemma 2.19.

Moreover, the lemma is still valid if, instead of the decoder of Definition 5.4, we use the a stochastic decoder, as proved in Appendix C.1.

Remark 5.8 - *Source coding with side information using a stochastic decoder.* Consider a DMS (A^n, B^n) as above, and $Z^n := A^n G_n$ the polarization of A^n. We consider the very high entropy set $\mathcal{H}_{A|B}$ and $Z[\mathcal{H}_{A|B}]$. For every $j \in [\![1, n]\!]$, we generate $\hat{Z}^n$ according to the conditional distribution

$$P_{\hat{Z}_i|\hat{Z}^{i-1}B^n} = \begin{cases} \mathbb{1}_{\hat{Z}_i|Z_i} & \text{if } i \in \mathcal{H}_{A|B}, \\ P_{Z_i|Z^{i-1}B^n} & \text{if } i \in \mathcal{H}_{A|B}^c. \end{cases}$$

The decoder computes $\hat{A}^n := \hat{Z}^n G_n$. Then, $\mathbb{P}\{\hat{A}^n \neq A^n\} \leq \delta_n$.

Now, we state the counterpart of Lemma 2.20. This result, adapted from [12, Lemma 4], is proved in Appendix C.2.

Lemma 5.9 - *Channel randomness extraction using polar codes* *Consider a discrete memoryless source (A^n, B^n) generated according to P_{AB}, and suppose $|\mathcal{A}| = \{0, 1\}$. Let $Z^n := A^n G_n$ be the polarization of A^n. We consider the restriction of Z^n to the very high entropy bits $Z[\mathcal{V}_{A|B}]$.*

Then,

$$\mathbb{D}(P_{Z[\mathcal{V}_{A|B}]B^n} \| Q_{Z[\mathcal{V}_{A|B}]} P_{B^n}) \leq n\delta_n,$$

and therefore by Pinsker's inequality we have

$$\mathbb{V}(P_{Z[\mathcal{V}_{A|B}]B^n}, Q_{Z[\mathcal{V}_{A|B}]} P_{B^n}) \leq \sqrt{2\log 2}\sqrt{n\delta_n}.$$

Note that the very high entropy bits in positions $\mathcal{V}_{A|B}$ play the same role as the random binning index in Lemma 2.20.

The previous lemmas suggest that achievability results proved via random binning arguments can be turned into explicit coding schemes using polar codes. Intuitively, every time that the random binning proof involves Lemma 2.19, this can be substituted with Lemma 5.7 using the high entropy set $\mathcal{H}_{A|B}$. On the other hand, whenever Lemma 2.20 is needed, we can use the very high entropy set $\mathcal{V}_{A|B}$ and Lemma 5.9.

However, as we will see in the next section, if in the proof the side information in Lemma 5.7 is different from the side information in Lemma 5.9, the high entropy and very high entropy sets may not necessarily be aligned. Thus, the coding scheme requires to deal with this issue carefully, and in order to realign the indices we use a chaining construction over k blocks.

5.3 Polar coding for strong coordination of signals and actions over noisy channels

We present a joint source-channel polar coding scheme for strong coordination that achieves joint coordination of signals and actions over a noisy channel. We focus on channels without state and side information for simplicity, and we show that the inner bound $\mathcal{R}_{\text{in}}$ for the coordination region of Theorem 3.3 is achievable using polar codes, if we assume that between the encoder and decoder there is an error-free channel of negligible rate.

For brevity, we only focus on the set of achievable distributions in $\mathcal{R}_{\text{in}}$ for which the auxiliary variable W is binary. The scheme can be extended to the case of a non-binary random variable W using non-binary polar codes when $|\mathcal{W}|$ is a prime number [60].

Theorem 5.10 *The subset of the region $\mathcal{R}_{\text{in}}$ defined in (3.1) for which the auxiliary random variable W is binary is achievable using polar codes, provided there exists an error-free channel of negligible rate between the encoder and decoder.*

5.3.1 Polar coding scheme

To convert the information-theoretic achievability proof of Theorem 3.3 into a polar coding proof, we use source polarization [2] to induce the desired joint distribution. Inspired by [12], we want to translate the random binning scheme into a polar coding scheme.
The key step for coordination is to generate the same auxiliary sequence W^n at both decoder and encoder. Once this is accomplished, the task is essentially done because the sequences X^n and Y^n with the correct distribution can be generated via the conditional distributions $\bar{P}_{X|U}$ and the channel $\bar{P}_{Y|X}$; hence, the appropriate V^n can be drawn at the decoder.
As we will show, the information constraints and rate conditions found in the random binning

proof directly convert into the definition of the polarization sets. In the random binning scheme we reduced the amount of common randomness F by having the nodes to agree on an instance of F, here we recycle some common randomness using a chaining construction as in [32, 55].

Consider random vectors U^n, W^n, X^n, Y^n and V^n generated i.i.d. according to (3.5)

$$\bar{P}_{U^n}\bar{P}_{W^n|U^n}\bar{P}_{X^n|W^nU^n}\bar{P}_{Y^n|X^n}\bar{P}_{V^n|W^nY^n}$$

that satisfies the inner bound of (3.1):

$$\mathcal{R}_{\text{in}} = \left\{ (\bar{P}_{UXYV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{UXYV} = \bar{P}_U \bar{P}_{X|U} \bar{P}_{Y|X} \bar{P}_{V|UXY} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UXYWV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{X|UW} \bar{P}_{Y|X} \bar{P}_{V|WY} \\ I(W;U) \leq I(W;Y) \\ R_0 \geq I(W;UXV|Y) \end{array} \right\}.$$

Let $Z^n := W^n G_n$ be the polarization of W^n. For some $0 < \beta < 1/2$, let $\delta_n = 2^{-n^\beta}$ and define the very high entropy and high entropy sets:

$$\begin{aligned} \mathcal{V}_W &:= \left\{ j \in [\![1,n]\!] \mid H(Z_j|Z^{j-1}) > 1-\delta_n \right\}, \\ \mathcal{V}_{W|U} &:= \left\{ j \in [\![1,n]\!] \mid H(Z_j|Z^{j-1}U^n) > 1-\delta_n \right\}, \\ \mathcal{V}_{W|Y} &:= \left\{ j \in [\![1,n]\!] \mid H(Z_j|Z^{j-1}Y^n) > 1-\delta_n \right\}, \\ \mathcal{H}_{W|Y} &:= \left\{ j \in [\![1,n]\!] \mid H(Z_j|Z^{j-1}Y^n) > \delta_n \right\}. \end{aligned} \tag{5.2}$$

We define the following disjoint sets:

$$\begin{aligned} A_1 &:= \mathcal{V}_{W|U} \cap \mathcal{H}_{W|Y}, \quad & A_2 &:= \mathcal{V}_{W|U} \cap \mathcal{H}_{W|Y}^c, \\ A_3 &:= \mathcal{V}_{W|U}^c \cap \mathcal{H}_{W|Y}, \quad & A_4 &:= \mathcal{V}_{W|U}^c \cap \mathcal{H}_{W|Y}^c. \end{aligned} \tag{5.3}$$

The cardinality of these sets is characterized as follows.

Remark 5.11 - *Cardinality*. It follows from Corollary 5.3 that:

- $\mathcal{V}_{W|Y} \subset \mathcal{H}_{W|Y}$ and $\lim_{n\to\infty} \frac{|\mathcal{H}_{W|Y} \setminus \mathcal{V}_{W|Y}|}{n} = 0$,
- $\lim_{n\to\infty} \frac{|\mathcal{V}_{W|U}|}{n} = H(W|U)$,
- $\lim_{n\to\infty} \frac{|\mathcal{H}_{W|Y}|}{n} = H(W|Y)$.

Since $H(W|U) - H(W|Y) = I(W;Y) - I(W;U)$, for sufficiently large n the assumption $I(W;Y) \geq I(W;U)$ directly implies that $|A_2| \geq |A_3|$.

Remark 5.12 - ***Side information asymmetry.*** Note that the encoder can use its observation of the source to generate the bits of Z^n (and therefore W^n). The decoder, on the other hand, reconstructs $\hat{Z}^n$ (and therefore $\hat{W}^n$) using the output of the channel as side information. More precisely, at the encoder we have two sets of bits, A_1 and A_2, which are almost uniformly random given the source since $A_1 \cup A_2 = \mathcal{V}_{W|U}$, and two sets of bits, A_3 and A_4, which are almost deterministic given the source since $A_3 \cup A_4 = \mathcal{V}^c_{W|U}$. Then, the bits A_1 and A_2 can be generated uniformly at random, and the bits A_3 and A_4 can be generated using successive cancellation encoding according to $\bar{P}_{Z_i|Z^{i-1}U^n}$ as in Definition 5.6.
At the decoder the situation is slightly different, because the output of the channel Y^n is playing the role of the side information. Then, the sets A_1 and A_3 are almost uniformly random given the output of the channel, and A_2 and A_4 almost deterministic and can be generated using successive cancellation decoding as in Definition 5.4 according to $\bar{P}_{Z_i|Z^{i-1}Y^n}$.
We individuate a "problematic" set A_3, which is a non-empty set of bits that are almost deterministic for the encoder but can not be recovered reliably at the decoder. To solve this asymmetry, we use a chaining construction over k blocks as in [32, 55] to ensure proper alignment of the polarized sets.

Encoding The encoder observes k blocks of the source $U^n_{(1:k)} := (U^n_{(1)}, \dots, U^n_{(k)})$ and generates for each block $i \in [\![1, k]\!]$ a random variable $Z^n_{(i)}$ following the procedure described in Algorithm 1. The chaining construction proceeds as follows:

- Let $A'_1 := \mathcal{V}_{W|UXYV}$, observe that A'_1 is a subset of A_1 since $\mathcal{V}_{W|UXYV} \subset \mathcal{V}_{W|U}$ and $\mathcal{V}_{W|UXYV} \subset \mathcal{V}_{W|Y} \subset \mathcal{H}_{W|Y}$. The bits in $A'_1 \subset \mathcal{V}_{W|U}$ in block $i \in [\![1, k]\!]$ are chosen with uniform probability using a uniform randomness source $\bar{C}'$ shared with the decoder, and their value is reused over all blocks;
- The bits in $A_1 \setminus A'_1 \subset \mathcal{V}_{W|U}$ in block $i \in [\![1, k]\!]$ are chosen with uniform probability using a uniform randomness source $\bar{C}_i$ shared with the decoder;
- In the first block the bits in $A_2 \subset \mathcal{V}_{W|U}$ are chosen with uniform probability using a local randomness source M;
- For the following blocks, let A'_3 be a subset of A_2 such that $|A'_3| = |A_3|$. The bits of A_3 in block i are sent to A'_3 in the block $i+1$ using a one-time pad with key C_i. Thanks to the Crypto Lemma, if we choose C_i of size $|A_3|$ to be a uniform random key, the bits in A'_3 in the block $i+1$ are uniform. The bits in $A_2 \setminus A'_3$ are chosen with uniform probability using the local randomness source M;
- The bits in A_3 and in A_4 are generated according to the previous bits using successive cancellation encoding as in Definition 5.6. Note that it is possible to sample efficiently from $\bar{P}_{Z_i|Z^{i-1}U^n}$ given U^n.

As in [12], to deal with unaligned indices, chaining also requires in the last encoding block to transmit $Z_{(k)}[A_3]$ to the decoder. Hence the coding scheme requires an error-free channel between the encoder and decoder which has negligible rate since $|Z_{(k)}[A_3]| \leq |\mathcal{H}_{W|Y}|$ and

$$\lim_{\substack{n\to\infty\\k\to\infty}} \frac{|\mathcal{H}_{W|Y}|}{kn} = \lim_{k\to\infty} \frac{H(W|Y)}{k} = 0.$$

The encoder then computes $W_{(i)}^n = Z_{(i)}^n G_n$ for $i = 1, \dots, k$ and generates $X_{(i)}^n$ symbol by symbol from $W_{(i)}^n$ and $U_{(i)}^n$ using the conditional distribution

$$\bar{P}_{X_{j,(i)}|W_{j,(i)}U_{j,(i)}}(x|\tilde{w}_{j,(i)}, u_{j,(i)}) = \bar{P}_{X|WU}(x|w_{j,(i)}, u_{j,(i)})$$

and sends $X_{(i)}^n$ over the channel.

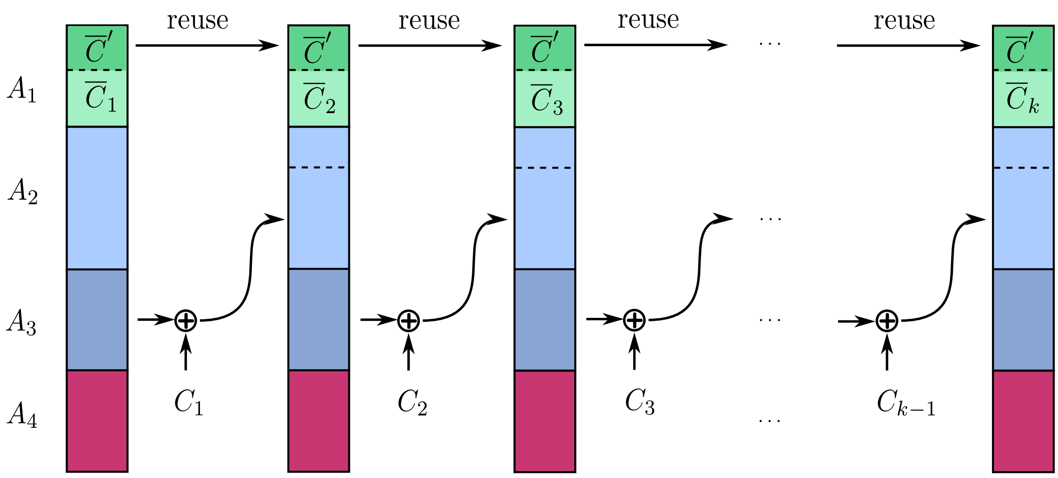

Figure 5.1: Chaining construction for block Markov encoding

Decoding The decoding procedure described in Algorithm 2 proceeds as follows. The decoder observes $(Y_{(1)}^n, \dots, Y_{(k)}^n)$ and $Z_{(k)}[A_3]$ which allows it to decode in reverse order. For $i \in [\![1, k]\!]$, we note $\hat{Z}_{(i)}^n$ the estimate of $Z_{(i)}^n$ at the decoder generated as in Definition 5.4. By Lemma 5.7, Z^n is equal to $\hat{Z}^n$ with high probability. In block $i \in [\![1, k]\!]$, the decoder has access to $\hat{Z}_{(i)}[A_1 \cup A_3] = \hat{Z}_{(i)}[\mathcal{H}_{W|Y}]$:

- the bits in A_1 in block i correspond to shared randomness $\bar{C}'$ and $\bar{C}_i$ for A_1' and $A_1 \setminus A_1'$ respectively;
- in block $i \in [1, k-1]$ the bits in A_3 are obtained by successfully recovering A_2 in block $i+1$.

Rate of common randomness The rate of common randomness is $I(W; UXV|Y)$ since:

$$\lim_{n\to\infty} \frac{k|A_1| - (k-1)|A_1'| + (k-1)|A_3|}{kn} = \lim_{n\to\infty} \frac{|A_1| + |A_3| - |A_1'|}{n}$$
$$= H(W|Y) - H(W|UXYV) = I(W; UXV|Y).$$

Algorithm 1: Encoding

Input : $(U^n_{(1)}, \ldots, U^n_{(k)})$, M local randomness (uniform random bits), common randomness $(\bar{C}', \bar{C}_{1:k}, C_{1:k-1})$ shared with the decoder, where $\bar{C}'$ has size $|A'_1|$, $\bar{C}_{1:k}$ has size $k|A_1 \setminus A'_1|$, $C_{1:k-1}$ has size $(k-1)|A_3|$.

Output: $(Z^n_{(1)}, \ldots, Z^n_{(k)})$

if $i = 1$ **then**

$$Z_{(1)}[A'_1] \longleftarrow \bar{C}'$$
$$Z_{(1)}[A_1 \setminus A'_1] \longleftarrow \bar{C}_1$$
$$Z_{(1)}[A_2] \longleftarrow M$$

for $j \in A_3 \cup A_4$ **do**

Given $U^n_{(1)}$, successively choose the bits $Z_{j,(1)}$ according to

$$\bar{P}_{Z_j|Z^{j-1}U^n}(Z_{j,(1)}|Z^{j-1}_{(1)}U^n_{(1)}) \tag{5.4}$$

end

end

for $i = 2, \ldots, k$ **do**

$$Z_{(i)}[A'_1] \longleftarrow \bar{C}'$$
$$Z_{(i)}[A_1 \setminus A'_1] \longleftarrow \bar{C}_i$$
$$Z_{(i)}[A'_3] \longleftarrow Z_{(i-1)}[A_3] \oplus C_{i-1}$$
$$Z_{(i)}[A_2 \setminus A'_3] \longleftarrow M$$

for $j \in A_3 \cup A_4$ **do**

Given $U^n_{(i)}$, successively choose the bits $Z_{j,(i)}$ according to

$$\bar{P}_{Z_j|Z^{j-1}U^n}(Z_{j,(i)}|Z^j_{(i-1)}U^n_{(i)}) \tag{5.5}$$

end

end

Remark 5.13 - ***Polar coding for empirical coordination.*** The proposed scheme can easily be adapted to prove empirical coordination with polar codes. Since empirical coordination requires less common randomness, the polar coding schemes changes in the amount of recycled common randomness. Moreover, empirical coordination is not compromised if the last block is not coordinated, as long as the number of blocks k is large enough, and it is not necessary to require an error-free channel of negligible rate. The complete scheme and the achievability proof for empirical coordination are in Appendix C.3.

Algorithm 2: Decoding

Input : $(Y^n_{(1)}, \dots, Y^n_{(k)})$, $Z_{(k)}[A_3]$ shared with the encoder, common randomness $(\bar{C}', \bar{C}_{1:k-1}, C_{1:k-1})$ shared with the encoder, where $\bar{C}'$ has size $|A_1'|$, $\bar{C}_{1:k}$ has size $k|A_1 \setminus A_1'|$ and $C_{1:k-1}$ has size $(k-1)|A_3|$.

Output: $(\hat{Z}^n_{(1)}, \dots, \hat{Z}^n_{(k)})$

for $i = k, \dots, 1$ **do**
    $\hat{Z}_{(i)}[A_1'] \longleftarrow \bar{C}'$ $\hat{Z}_{(i)}[A_1 \setminus A_1'] \longleftarrow \bar{C}_i$
    if $i = k$ **then**
        $\hat{Z}_{(i)}[A_3]$ shared with the decoder
    end
    else
        $\hat{Z}_{(i)}[A_3] \longleftarrow \hat{Z}_{(i+1)}[A_3']$
    end
    for $j \in A_2 \cup A_4$ **do**
        Successively choose the bits according to $\hat{Z}_{j,(i)} = \begin{cases} 0 & \text{if } L_n(Y^n_{(i)}, Z^{j-1}_{(i-1)}) \geq 1 \\ 1 & \text{else} \end{cases}$
        where
$$L_n(Y^n_{(i)}, Z^{j-1}_{(i-1)}) = \frac{\bar{P}_{Z_{j,(i)}|Z^{j-1}_{(i-1)}Y^n_{(i)}}(0|\hat{Z}^{j-1}_{(i-1)}Y^n_{(i)})}{\bar{P}_{Z_{j,(i)}|Z^{j-1}_{(i-1)}Y^n_{(i)}}(1|\hat{Z}^{j-1}_{(i-1)}Y^n_{(i)})}$$
    end
end

5.3.2 Achievability

We note with P the joint distribution induced by the encoding and decoding algorithm of the previous sections. The proof of Theorem 5.10 requires a few steps, here presented as different lemmas.

Coordination in one block First, we want to show that we have strong coordination in each block.

Lemma 5.14 *In each block $i \in [\![1, k]\!]$, we have*

$$\mathbb{V}(P_{U^n_{(i)}W^n_{(i)}X^n_{(i)}Y^n_{(i)}V^n_{(i)}}, \bar{P}^{\otimes n}_{UWXYV}) \leq \delta^{(1)}_n \tag{5.6}$$

where $\delta^{(1)}_n := 2\mathbb{P}\left\{\hat{W}^n_{(i)} \neq W^n_{(i)}\right\} + \sqrt{2\log 2}\sqrt{n\delta_n}$.

Proof. First, we need the following result, which is a direct consequence of Lemma 5.9.

Lemma 5.15 *For $i \in [\![1, k]\!]$, we have*

$$\mathbb{V}(P_{U^n_{(i)}W^n_{(i)}}, \bar{P}^{\otimes n}_{UW}) \leq \delta^{(2)}_n$$

where $\delta_n^{(2)} := \sqrt{2\log 2}\sqrt{n\delta_n}$.

Proof. Recall that $W_{(i)}^n = Z_{(i)}^n G_n$, $W^n = Z^n G_n$. Then, in each block $i \in [\![1,k]\!]$, we have

$$\mathbb{D}(\bar{P}_{UW}^{\otimes n} \| P_{U_{(i)}^n W_{(i)}^n}) \overset{(a)}{=} \mathbb{D}(\bar{P}_{UZ}^{\otimes n} \| P_{U_{(i)}^n Z_{(i)}^n}) \overset{(b)}{=} \mathbb{D}(\bar{P}_{UZ[A_1\cup A_2]}^{\otimes n} \| P_{U_{(i)}^n Z_{(i)}[A_1\cup A_2]}) \overset{(c)}{\leq} n\delta_n, \tag{5.7}$$

where (a) comes from the invertibility of G_n, (b) from the fact that we can consider only the very high entropy bits since

$$\begin{aligned}
\mathbb{D}(\bar{P}_{UZ}^{\otimes n} \| P_{U_{(i)}^n Z_{(i)}^n}) &\overset{(d)}{=} \mathbb{D}(\bar{P}_{Z^n|U^n} \| P_{Z_{(i)}^n|U_{(i)}^n} | \bar{P}_{U^n}) \\
&\overset{(e)}{=} \sum_{j=1}^{n} \mathbb{D}(\bar{P}_{Z_j|Z^{j-1}U^n} \| P_{Z_{(i),j}|Z_{(i)}^{j-1}U_{(i)}^n} | \bar{P}_{Z^{j-1}U^n}) \\
&\overset{(f)}{=} \sum_{j\in A_1\cup A_2} \mathbb{D}(P_{Z_j|Z^{j-1}U^n} \| P_{Z_{(i),j}|Z_{(i)}^{j-1}U_{(i)}^n} | \bar{P}_{Z^{j-1}U^n}) \\
&\overset{(g)}{=} \mathbb{D}(\bar{P}_{UZ[A_1\cup A_2]}^{\otimes n} \| P_{U_{(i)}^n Z_{(i)}^n[A_1\cup A_2]}),
\end{aligned} \tag{5.8}$$

where (d), (e) and (g) come by the chain rule for the divergence (Lemma A.5), and (f) is true because $Z_{(i)}^n$ is generated according to Definition 5.6.

Finally (c) comes from Lemma 5.9.

Therefore, applying Pinsker's inequality to (5.7) we have

$$\mathbb{V}(P_{U_{(i)}^n W_{(i)}^n}, \bar{P}_{UW}^{\otimes n}) \leq \sqrt{2\log 2}\sqrt{n\delta_n} := \delta_n^{(2)} \to 0. \qquad \square$$

Now, note that $X_{(i)}^n$ is generated symbol by symbol from $U_{(i)}^n$ and $W_{(i)}^n$ via the conditional distribution $\bar{P}_{X|UW}$ and $Y_{(i)}^n$ is generated symbol by symbol via the channel $\bar{P}_{Y|X}$. By Lemma 2.13, we add first $X_{(i)}^n$ and then $Y_{(i)}^n$ and we obtain that for each $i \in [\![1,k]\!]$,

$$\mathbb{V}(P_{U_{(i)}^n W_{(i)}^n X_{(i)}^n Y_{(i)}^n}, \bar{P}_{UWXY}^{\otimes n}) = \mathbb{V}(P_{U_{(i)}^n W_{(i)}^n}, \bar{P}_{UW}^{\otimes n}) \leq \delta_n^{(2)} \tag{5.9}$$

and therefore the left-hand side of (5.9) vanishes.

Observe that we cannot use Lemma 2.13 again because $V_{(i)}^n$ is generated using $\hat{W}_{(i)}^n$ (i.e. the estimate of $W_{(i)}^n$ at the decoder) and not $W_{(i)}^n$. By the triangle inequality for all $i \in [\![1,k]\!]$

$$\mathbb{V}(P_{U_{(i)}^n \hat{W}_{(i)}^n X_{(i)}^n Y_{(i)}^n}, \bar{P}_{UWXY}^{\otimes n}) \leq \mathbb{V}(P_{U_{(i)}^n \hat{W}_{(i)}^n X_{(i)}^n Y_{(i)}^n}, P_{U_{(i)}^n W_{(i)}^n X_{(i)}^n Y_{(i)}^n}) + \mathbb{V}(P_{U_{(i)}^n W_{(i)}^n X_{(i)}^n Y_{(i)}^n}, \bar{P}_{UWXY}^{\otimes n}). \tag{5.10}$$

We have proved in (5.9) that the second term of the right-hand side in (5.10) goes to zero, we show that the first term tends to zero as well.

We recall the definition of coupling and the basic coupling inequality for two random variables [53].

Definition 5.16 *A coupling of two probability distributions P_A and $P_{A'}$ on the same measurable* *space $\mathcal{A}$ is any probability distribution $\hat{P}_{AA'}$ on the product measurable space $\mathcal{A} \times \mathcal{A}$ whose marginals are P_A and $P_{A'}$.*

Proposition 5.17 - ***Coupling property [53, I.2.6]*** *Given two random variables A, A' with probability distributions P_A, $P_{A'}$, any coupling $\hat{P}_{AA'}$ of P_A, $P_{A'}$ satisfies*

$$\mathbb{V}(P_A, P_{A'}) \leq 2\mathbb{P}_{\hat{P}_{AA'}}\{A \neq A'\}.$$

Now, we apply the coupling property to

$$\begin{array}{ll} A = U^n_{(i)} \hat{W}^n_{(i)} X^n_{(i)} Y^n_{(i)}, & A' = U^n_{(i)} W^n_{(i)} X^n_{(i)} Y^n_{(i)}, \\ P = P_{U^n_{(i)} \hat{W}^n_{(i)} X^n_{(i)} Y^n_{(i)}}, & P' = P_{U^n_{(i)} W^n_{(i)} X^n_{(i)} Y^n_{(i)}}, \end{array}$$

on $\mathcal{A} = \mathcal{U} \times \mathcal{W} \times \mathcal{X} \times \mathcal{Y}$. Since by Lemma 5.7 we have

$$p_e := \mathbb{P}\left\{\hat{W}^n_{(i)} \neq W^n_{(i)}\right\} \leq \delta_n,$$

we find that $\mathbb{V}(P_{U^n_{(i)} \hat{W}^n_{(i)} X^n_{(i)} Y^n_{(i)}}, P_{U^n_{(i)} W^n_{(i)} X^n_{(i)} Y^n_{(i)}}) \leq 2p_e$ and therefore

$$\mathbb{V}(P_{U^n_{(i)} \hat{W}^n_{(i)} X^n_{(i)} Y^n_{(i)}}, \bar{P}^{\otimes n}_{UWXY}) \leq 2p_e + \delta_n^{(2)} = \delta_n^{(1)} \to 0.$$

Since V_i^n is generated symbol by symbol from $\hat{W}_i^n$ and Y_i^n, we apply Lemma 2.13 and find

$$\mathbb{V}(P_{U^n_{(i)} \hat{W}^n_{(i)} X^n_{(i)} Y^n_{(i)} V^n_{(i)}}, \bar{P}^{\otimes n}_{UWXYV}) \leq \delta_n^{(1)} \to 0.$$

□

Coordination of two consecutive blocks Now, we want to show that two consecutive blocks are almost independent. To simplify the notation, we set

$$\begin{array}{ll} L := U^n X^n Y^n V^n & \\ L_i := U^n_{(i)} X^n_{(i)} Y^n_{(i)} V^n_{(i)} & i \in [\![1, k]\!] \\ L_{a:b} := U^n_{(a:b)} X^n_{(a:b)} Y^n_{(a:b)} V^n_{(a:b)} & [\![a, b]\!] \subset [\![1, k]\!] \end{array}$$

Lemma 5.18 *For $i \in [\![2, k]\!]$, we have*

$$\mathbb{V}(P_{L_{i-1:i}\bar{C}'}, P_{L_{i-1}\bar{C}'} P_{L_i}) \leq \delta_n^{(3)}$$

where $\delta_n^{(3)} := \sqrt{2\log 2}\sqrt{n\delta_n + 2\delta_n^{(1)}(\log|\mathcal{U} \times \mathcal{X} \times \mathcal{W} \times \mathcal{Y} \times \mathcal{V}| - \log \delta_n^{(1)})}$ and $\delta_n^{(1)}$ is defined in Lemma 5.14.

Proof. For $i \in [\![2, k]\!]$, we have

$$\mathbb{D}(P_{L_{i-1:i}\bar{C}'} \| P_{L_{i-1}\bar{C}'} P_{L_i}) = I(L_{i-1}\bar{C}'; L_i) \overset{(a)}{=} I(L_i; \bar{C}') + I(L_{i-1}; L_i|\bar{C}')$$

$$\overset{(b)}{=} I(L_i; \bar{C}') = I(L_i; Z_{(i)}[A'_1]) \overset{(c)}{=} |A'_1| - H(Z_{(i)}[A'_1]|L_i) \tag{5.11}$$
$$\overset{(d)}{=} |A'_1| - H(Z[A'_1]|L) + \delta_n^{(4)} \overset{(e)}{\leq} |A'_1| - \sum_{j \in A'_1} H(Z_j|Z^{j-1}L) + \delta_n^{(4)}$$
$$\overset{(f)}{\leq} |A'_1| - |A'_1|(1-\delta_n) + \delta_n^{(4)} \leq n\delta_n + \delta_n^{(4)}$$

where (a) comes from the chain rule, (b) from the Markov chain $L_{i-1} - \bar{C}' - L_i$, (c) from the fact that the bits in A'_1 are uniform. To prove (d) observe that

$$H(Z_{(i)}[A'_1]|L_i) - H(Z[A'_1]|L) = H(Z_{(i)}[A'_1]L_i) - H(Z[A'_1]L) - H(L_i) + H(L)$$
$$\overset{(g)}{\leq} \delta_n^{(1)} \log \frac{|\mathcal{U} \times \mathcal{X} \times \mathcal{W} \times \mathcal{Y} \times \mathcal{V}|}{\delta_n^{(1)}} + \delta_n^{(1)} \log \frac{|\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}|}{\delta_n^{(1)}}$$
$$\leq 2\delta_n^{(1)} (\log |\mathcal{U} \times \mathcal{X} \times \mathcal{W} \times \mathcal{Y} \times \mathcal{V}| - \log \delta_n^{(1)}) := \delta_n^{(4)}$$

where (g) comes from Lemma A.16 since

$$\mathbb{V}(P_{L_i}, \bar{P}^{\otimes n}_{UXYV}) \leq \mathbb{V}(P_{L_i W^n_{(i)}}, \bar{P}^{\otimes n}_{UWXYV}) \leq \delta_n^{(1)}$$

that vanishes as n goes to infinity. Finally (e) is true because conditioning does not increase entropy and (f) comes by definition of the set A'_1. Then from Pinsker's inequality

$$\mathbb{V}(P_{L_{i-1:i}\bar{C}'}, P_{L_{i-1}\bar{C}'} P_{L_i}) \leq \sqrt{2\log 2} \sqrt{n\delta_n + \delta_n^{(4)}} = \delta_n^{(3)} \to 0. \qquad \square$$

Coordination of all blocks Now that we have proven the asymptotical independence of two consecutive blocks, we use Lemma 5.18 to prove the asymptotical independence of all blocks. First we need an intermediate step.

Lemma 5.19 *We have*

$$\mathbb{V}(P_{L_{1:k}}, \prod_{i=1}^{k} P_{L_i}) \leq \sqrt{k-1}\,\delta_n^{(3)}$$

where $\delta_n^{(3)}$ is defined in Lemma 5.18.

Proof. We have

$$\mathbb{D}(P_{L_{1:k}} \| \prod_{i=1}^{k} P_{L_i}) \overset{(a)}{=} \sum_{i=2}^{k} I(L_i; L_{1:i-1}) \leq \sum_{i=2}^{k} I(L_i; L_{1:i-1}\bar{C}')$$
$$= \sum_{i=2}^{k} \left(I(L_i; L_{i-1}\bar{C}') + \sum_{j=1}^{i-2} I(L_i; L_{i-j-1}|L_{i-j:i-1}\bar{C}') \right)$$
$$\leq \sum_{i=2}^{k} \left(I(L_i; L_{i-1}\bar{C}') + \sum_{j=1}^{i-2} I(L_i; L_{i-j-1:i-2}|L_{i-1}\bar{C}') \right)$$

$$\overset{(b)}{=}\sum_{i=2}^{k} I(L_i; L_{i-1}\bar{C}')\overset{(c)}{\leq}(k-1)(n\delta_n+\delta_n^{(4)})$$

where (a) comes from [13, Lemma 15], (b) is true because the dependence structure of the blocks gives the Markov chain $L_{i-j-1:i-2} - L_{i-1}\bar{C}' - L_i$ and (c) follows from (5.11). We conclude with Pinsker's inequality. $\square$

Finally, we prove the asymptotical independence of all blocks.

Lemma 5.20 *We have*

$$\mathbb{V}(P_{L_{1:k}}, \bar{P}_{UXYV}^{\otimes nk}) \leq \delta_n^{(5)}$$

where $\delta_n^{(5)} := \sqrt{k}(\delta_n^{(3)} + \delta_n^{(2)})$ *and* $\delta_n^{(2)}$ *and* $\delta_n^{(3)}$ *are defined in Lemma 5.15 and Lemma 5.18 respectively.*

Proof. By the triangle inequality

$$\mathbb{V}\left(P_{L_{1:k}}, \bar{P}_{UXYV}^{\otimes nk}\right) \leq \mathbb{V}\left(P_{L_{1:k}}, \prod_{i=1}^{k} P_{L_i}\right) + \mathbb{V}\left(\prod_{i=1}^{k} P_{L_i}, \bar{P}_{UXYV}^{\otimes nk}\right) \tag{5.12}$$

where the first term is smaller than $\sqrt{k-1}\delta_n^{(3)}$ by Lemma 5.19. To bound the second term, observe that

$$\mathbb{D}\left(\prod_{i=1}^{k} P_{L_i} \middle\| \bar{P}_{UXYV}^{\otimes nk}\right) = \mathbb{D}\left(\prod_{i=1}^{k} P_{L_i} \middle\| \prod_{i=1}^{k} \bar{P}_{UXYV}^{\otimes n}\right) = \sum_{i=1}^{k} \mathbb{D}\left(P_{L_i} \| \bar{P}_{UXYV}^{\otimes n}\right). \tag{5.13}$$

By the chain rule we have that $\mathbb{D}(P_{L_i} \| \bar{P}_{UXYV}^{\otimes n}) \leq \mathbb{D}(P_{L_i W_{(i)}^n} \| \bar{P}_{UWXYV}^{\otimes n})$. Since $X_{(i)}^n$, $Y_{(i)}^n$ and $V_{(i)}^n$ are generated symbol by symbol via the conditional distributions $\bar{P}_{X|UW}$, $\bar{P}_{Y|X}$ and $\bar{P}_{V|WY}$ respectively, by Lemma 2.14 we have that

$$\mathbb{D}(P_{L_i W_{(i)}^n} \| \bar{P}_{UWXYV}^{\otimes n}) = \mathbb{D}(P_{U_{(i)}^n W_{(i)}^n} \| \bar{P}_{UW}). \tag{5.14}$$

Hence, we have

$$\mathbb{D}(\prod_{i=1}^{k} P_{L_i} \| \bar{P}_{UXYV}^{\otimes kn}) = \sum_{i=1}^{k} \mathbb{D}(P_{L_i} \| \bar{P}_{UXYV}^{\otimes n})\overset{(a)}{\leq}\sum_{i=1}^{k} \mathbb{D}(P_{U_{(i)}^n W_{(i)}^n} \| \bar{P}_{UW}^{\otimes n})\overset{(b)}{=}kn\delta_n$$

where (a) follows from the chain rule and (5.14) and (b) comes from (5.7). Then, by Pinsker's inequality, (5.12) becomes:

$$\mathbb{V}(P_{L_{1:k}}, \bar{P}_{UXYV}^{\otimes nk}) \leq \sqrt{k-1}\delta_n^{(3)} + \sqrt{k}\delta_n^{(2)} \leq \sqrt{k}(\delta_n^{(3)} + \delta_n^{(2)}) = \delta_n^{(5)} \to 0. \qquad \square$$

6 | Coordination of signals and actions with strictly causal encoder

Until now we have considered joint source-channel coordination in the presence of a non-causal encoder and non-causal decoder. In this chapter, we examine the case in which the encoder is strictly causal, which has the benefit of shortening the transmission delay. For empirical coordination, [24] provides a complete characterization of the region using similar methods to those used in the analysis of causal state amplification in [16, 17]. Although the strong coordination region is still unknown, in Section 6.1 we provide an inner and an outer bound that differ only in the amount of common randomness needed to strongly coordinate signals and actions. The achievability proof relies on a random binning argument as in Section 3.1.1, but the nature of this setting presents some extra difficulties. In fact, the information about the source at time i is needed for the reconstruction, but is observed by the encoder only at time $i+1$. So this information must be recovered by the decoder at a later time. In order to ensure coordination, we use a block-Markov scheme and a one-time pad.

Finally, similarly to what we have done in Chapter 5, in Section 6.2 we prove that polar codes provide a constructive alternative to random binning proofs and we describe the explicit scheme for strong coordination.

6.1 Strong coordination with a strictly causal encoder

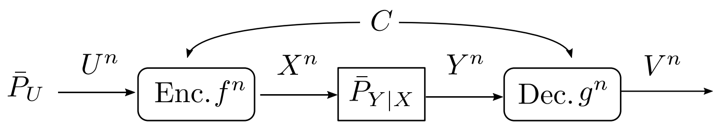

Figure 6.1: Coordination of signals and actions for a two-node network with strictly causal encoder and non-causal decoder.

We focus here on the setting, depicted in Figure 6.1 in which the encoder is strictly causal. For simplicity, we consider the setting without state and side information, and we suppose that the encoder and the decoder have access to a shared source of uniform randomness $C \in [\![1, 2^{nR_0}]\!]$. Let $U^n \in \mathcal{U}^n$ be an i.i.d. source with distribution $\bar{P}_U$. At time $i \in [\![1, n]\!]$, the strictly causal encoder observes the sequence $U^{i-1} \in \mathcal{U}^n$, common randomness C and selects a signal $X_i = f_i(U^{i-1}, C)$, where $f_i : \mathcal{U}^{i-1} \times [\![1, 2^{nR_0}]\!] \to \mathcal{X}$ is a stochastic function. The signal

$X^n = (X_1, \dots, X_n)$ is transmitted over a discrete memoryless channel $\bar{P}_{Y|X}$. Upon observing Y^n and the common randomness C, the decoder selects an action $V^n = g^n(Y^n, C)$, where $g^n : \mathcal{Y}^n \times [\![1, 2^{nR_0}]\!] \to \mathcal{V}^n$ is a stochastic map. Let $f^n := \{f_i\}_{i=1}^n$ for block length n. The pair (f^n, g^n) constitutes a code. We slightly modify the definitions of achievability and of the strong coordination in this setting.

Definition 6.1 *A pair $(\bar{P}_{UXYV}, R_0)$ is achievable for strong coordination if there exists a sequence (f^n, g^n) of strictly causal encoders and non-causal decoders with rate of common randomness R_0, such that for every $\varepsilon > 0$ there exists $n \in \mathbb{N}$ and $(U^{\tilde{n}}, X^{\tilde{n}}, Y^{\tilde{n}}, V^{\tilde{n}})$ a sufficiently long sub-sequence of (U^n, X^n, Y^n, V^n) with $\tilde{n} > (1-\varepsilon)n$ that satisfies*

$$\lim_{n\to\infty} \mathbb{V}\left(P_{U^{\tilde{n}}X^{\tilde{n}}Y^{\tilde{n}}V^{\tilde{n}}}, \bar{P}_{UXYV}^{\otimes \tilde{n}}\right) = 0$$

where P is the joint distribution induced by the code. The strong coordination region $\mathcal{R}_{SC}$ is the closure of the set of achievable pairs $(\bar{P}_{UXYV}, R_0)$.

Remark 6.2 - ***The classical definition does not apply.*** The definition for strong coordination in this setting is slightly different from Definition 3.1, which for the strictly causal encoder would be satisfied only by trivial distributions since the last block of the source will never be observed by the encoder. Here, we avoid this issue by losing coordination in a negligible fraction of time slots.

This precaution is not necessary when we consider empirical coordination, since losing coordination in only a few time slots does not affect the average behavior over time.

The problem of characterizing the strong coordination region is still open, but we establish the following inner and outer bounds.

Theorem 6.3 *Let $\bar{P}_U$ and $\bar{P}_{Y|X}$ be the given source and channel parameters, then $\mathcal{R}_{\text{SC,in}} \subseteq \mathcal{R}_{SC} \subseteq \mathcal{R}_{SC,out}$*

$$\mathcal{R}_{SC,in} := \left\{ (\bar{P}_{UXYV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{UXYV} = \bar{P}_U \bar{P}_X \bar{P}_{Y|X} \bar{P}_{V|UXY} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UXYVW} = \bar{P}_U \bar{P}_X \bar{P}_{W|UX} \bar{P}_{Y|X} \bar{P}_{V|WY} \\ I(WX;U) \leq I(WX;Y) \\ R_0 \geq I(W;UXV|Y) + H(X|WY) \end{array} \right\}, \tag{6.1}$$

$$\mathcal{R}_{\text{SC,out}} := \left\{ (\bar{P}_{UXYV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{UXYV} = \bar{P}_U \bar{P}_X \bar{P}_{Y|X} \bar{P}_{V|UXY} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UXYVW} = \bar{P}_U \bar{P}_X \bar{P}_{W|UX} \bar{P}_{Y|X} \bar{P}_{V|WY} \\ I(WX;U) \leq I(WX;Y) \\ R_0 \geq I(W;UXV|Y) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| + 4 \end{array} \right\}. \tag{6.2}$$

Remark 6.4 - ***Equivalent characterization.*** By the chain rule, we have

- $I(XW;U) = I(W;U|X) + I(X;U) = I(W;U|X)$ since U and X are independent;

- $I(XW;Y) = I(W;Y|X) + I(X;Y) = I(X;Y)$ because of the Markov chain $W - X - Y$.

Hence the condition $I(WX;U) \leq I(WX;Y)$ in (6.1) and (6.2) is equivalent to $I(W;U|X) \leq I(X;Y)$.

Remark 6.5 - ***Markov chain decomposition.*** Observe that the decomposition of the joint distributions $\bar{P}_{UXYV}$ and $\bar{P}_{UWXYV}$ is equivalently characterized in terms of Markov chains:

$$U \perp X, \qquad \begin{cases} U \perp X, \\ Y - X - W, \\ V - (Y,W) - (X,U). \end{cases} \tag{6.3}$$

Comparison with empirical coordination For empirical coordination, [24, Theorem 3] gives the following characterization of the region with strictly causal encoding:

$$\mathcal{R}_{\text{SC,e}} := \left\{ \bar{P}_{UXYV} \;\middle|\; \begin{array}{l} \bar{P}_{UXYV} = \bar{P}_U \bar{P}_X \bar{P}_{Y|X} \bar{P}_{V|UXY} \\ \exists\, W \text{ taking values in } \mathcal{U} \\ P_{UWXYV} = \bar{P}_U \bar{P}_X \bar{P}_{W|XU} \bar{P}_{Y|X} \bar{P}_{V|WY} \\ I(X,W;U) \leq I(X,W;Y) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| + 1 \end{array} \right\}. \tag{6.4}$$

Similarly to the case of non-causal encoder and decoder of Section 3.1, in $\mathcal{R}_{\text{SC,in}}$ and $\mathcal{R}_{\text{SC,out}}$ the decomposition of the joint distribution and the information constraints are the same as in $\mathcal{R}_{\text{SC,e}}$, but for strong coordination a positive rate of common randomness is also necessary.

6.1.1 Inner bound

As in Section 3.1.1, the crucial point to prove the achievability is to define a random binning for the target joint distribution, and a random coding scheme, each of which induces a joint distribution, and to prove that the two schemes have almost the same statistics. However, the strictly causal nature of the encoder requires a more subtle random coding scheme with a block-Markov structure. Here, we present the coding schemes, while the proof that the two schemes have almost the same statistics is in Appendix B.1.3.

Random binning scheme Assume that the sequences U^n, X^n, W^n, Y^n and V^n are jointly i.i.d. with distribution

$$\bar{P}_{U^n} \bar{P}_{X^n} \bar{P}_{W^n|U^n X^n} \bar{P}_{Y^n|X^n} \bar{P}_{V^n|W^n Y^n}. \tag{6.5}$$

While in the achievability proof for non-causal encoder and decoder in Theorem 3.3 the random binning was only defined for the auxiliary variable W^n, to deal with the strictly causal encoder we bin the variable X^n as well. We consider the two following uniform random binnings for X^n:

- $M_1 = \varphi_1(X^n)$, where $\varphi_1 : \mathcal{X}^n \to [\![1, 2^{nR_1}]\!]$,

- $M_2 = \varphi_2(X^n)$, $\varphi_2 : \mathcal{X}^n \to [\![1, 2^{nR_2}]\!]$.

The rates R_1 and R_2 are chosen as follows:

- $R_1 + R_2 < H(X)$, so that by Lemma 2.20 there exists one binning (φ'_1, φ'_2) of X such that M_1 and M_2 are almost uniform and almost independent of each other;
- $R_1 > H(X|Y)$, so that by Lemma 2.19 there exists one binning φ'_1 of X such that it is possible to reconstruct X from Y and M_1 with high probability using a Slepian-Wolf decoder via the conditional distribution $P^{\text{SW}}_{\hat{X}^n|M_1Y^n}$

where we can use the same binning φ'_1 for both conditions, as proved in Remark 3.7.

Then, we consider three uniform random binnings for W^n:

- $M_3 = \varphi_3(W^n)$, $\varphi_3 : \mathcal{W}^n \to [\![1, 2^{nR_3}]\!]$,
- $M_4 = \varphi_4(W^n)$, $\varphi_4 : \mathcal{W}^n \to [\![1, 2^{nR_4}]\!]$,
- $F = \psi(W^n)$, $\psi : \mathcal{W}^n \to [\![1, 2^{nR}]\!]$,

where the rates R_3, R_4 and R are chosen as follows:

- $R_3 + R < H(W|XU)$, so that by Lemma 2.20 there exists one binning (φ'_3, ψ') of W such that M_3 and F are almost uniform and almost independent of X and U;
- $R_3 + R_4 + R > H(W|X)$, so that by Lemma 2.19 there exists one binning $(\varphi'_3, \varphi'_4, \psi')$ of W such that it is possible to reconstruct W from X and (M_3, M_4, F) with high probability using a Slepian-Wolf decoder via the conditional distribution $P^{\text{SW}}_{\hat{W}^n|M_3M_4FX^n}$,

and we can use the same binning (φ'_3, ψ') for both conditions, as proved in Remark 3.7.

This defines a joint distribution:

$$P^{\text{RB}} := \bar{P}_{U^n} \bar{P}_{X^n} \bar{P}_{W^n|U^nX^n} \bar{P}_{M_1|X^n} \bar{P}_{M_2|X^n} \bar{P}_{M_3|W^n} \bar{P}_{M_4|W^n} \bar{P}_{F|W^n} \bar{P}_{Y^n|X^n} \bar{P}_{V^n|W^nY^n}. \tag{6.6}$$

In particular, the conditional distributions $P^{\text{RB}}_{M_4|M_3X^nU^n}$, $P^{\text{RB}}_{W^n|M_3M_4FX^n}$ and $P^{\text{RB}}_{X^n|M_1M_2M_3F}$ are well-defined.

Random coding scheme In this section we follow the approach in [70, Section IV.E] and [31]. Suppose that encoder and decoder have access to extra randomness F, where F is generated uniformly at random in $[\![1, 2^{nR}]\!]$ with distribution Q_F independently of the rest of the common randomness.

Encoder We use a chaining construction over k blocks of length n in which the encoder observes $U^n_{(1:k)} := (U^n_{(1)}, \dots, U^n_{(k)})$, where $U^n_{(i)}$ for $i \in [\![1, k]\!]$ are k blocks of the source. The encoder has access to common randomness $(M_{1,(1:k)}, M_{3,(1:k)}, F_{(1:k)}, K_{(2:k)})$ and the block-Markov scheme depicted in Figure 6.2 proceeds as follows:

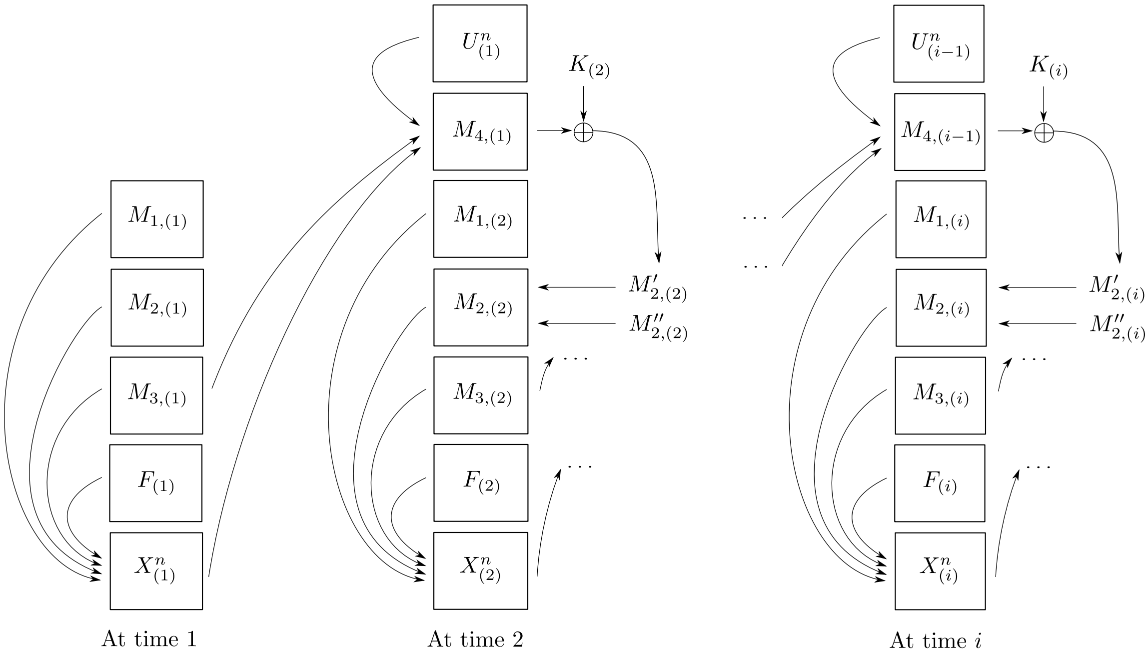

Figure 6.2: Block-Markov encoding scheme.

- For $i \in [\![1, k]\!]$, $M_{3,(i)}$ and $F_{(i)}$ are generated independently and uniformly over $[\![1, 2^{nR_3}]\!]$ and $[\![1, 2^{nR}]\!]$ using common randomness with distributions Q_{M_3} and Q_F respectively;
- $M_{1,(i)}$ is generated independently and uniformly over $[\![1, 2^{nR_1}]\!]$ using common randomness with distribution Q_{M_1};
- In the first block, $M_{2,(1)}$ is generated uniformly at random using some independent local randomness;
- For $i \in [\![2, k]\!]$, $M_{4,(i-1)}$ is generated according to the distribution defined earlier

$$P^{\text{RB}}_{M_4|M_3X^nU^n}(\mathbf{m}_{4,(i-1)}|\mathbf{m}_{3,(i-1)}, \mathbf{x}_{(i-1)}, \mathbf{u}_{(i-1)});$$

  where $(\mathbf{m}_{3,(i-1)}, \mathbf{x}_{(i-1)}, \mathbf{u}_{(i-1)})$ are generated at time $i-1$;
- For $i \in [\![2, k]\!]$, $M_{2,(i)} = (M'_{2,(i)}, M''_{2,(i)})$, where

$$M'_{2,(i)} = M_{4,(i-1)} \oplus K_i \tag{6.7}$$

  and K_i is generated uniformly over $[\![1, 2^{nR_4}]\!]$ using common randomness, while $M''_{2,(i)}$ is generated uniformly at random using some independent local randomness. Thanks to the Crypto Lemma, the distribution on $M_{2,(i)}$ is uniform and we denote it with Q_{M_2};
- The encoder generates $X^n_{(i)}$ according to the distribution defined earlier

$$P^{\text{RB}}_{X^n|M_1M_2M_3F}(\mathbf{x}_{(i)}|\mathbf{m}_{1,(i)}, \mathbf{m}_{2,(i)}, \mathbf{m}_{3,(i)}, \mathbf{f}_{(i)});$$

  Note that this distribution satisfies the strictly causal constraint, since $X^n_{(i)}$ is generated knowing the common randomness and $M'_{2,(i)} = M_{4,(i-1)} \oplus K_i$, where $M_{4,(i-1)}$ depends on the source at time $i-1$;

Then, the sequence $X^n_{(i)}$ is sent through the channel.

Remark 6.6 - ***Rate condition implies the information constraint.*** Observe that we have imposed the condition $|M_{4,(i-1)}| = |M'_{2,(i)}|$, which holds as long as $R_4 \leq R_2$. We have

$$R_2 < H(X) - R_1 < H(X) - H(X|Y) = I(X;Y),$$
$$R_4 > I(W;U|X).$$

Then, $R_4 \leq R_2$ implies $I(W;U|X) < I(X;Y)$.

Decoder Since the decoder is non-causal, it observes $Y^n_{(1:k)}$ and common randomness $(M_{1,(1:k)}, M_{3,(1:k)}, F_{(1:k)}, K_{(2:k)})$ and the decoding algorithm proceeds as follows:

- The decoder reconstructs $\hat{X}^n_{(1:k)}$, where, for all $i \in [\![1,k]\!]$, $\hat{X}^n_{(i)}$ is generated via the conditional distributions

$$P^{\text{SW}}_{\hat{X}^n|M_1Y^n}(\mathbf{x}_{(i)}|\mathbf{m}_{1,(i)}, \mathbf{y}_{(i)});$$

- The decoder recovers $\hat{M}_{2,(1:k)}$, where, for all $i \in [\![1,k]\!]$, $\hat{M}_{2,(i)}$ is generated via

$$\varphi_2(\hat{\mathbf{x}}_{(i)}) = \mathbf{m}_{2,(i)};$$

  where $\hat{\mathbf{x}}_{(i)}$ is the output of the Slepian-Wolf decoder;

- For all $i \in [\![2,k]\!]$, with the key of the one-time pad $K_{(i)}$ and $\hat{M}'_{2,(i)}$, the decoder recovers

$$\hat{M}_{4,(i-1)} = \hat{M}'_{2,(i)} \oplus K_{(i)};$$

- Observe that at time i, the decoder knows an estimate of $\hat{M}_{4,(i)}$ because the non-causal nature of the decoder allows us to decode in reverse order and we note its distribution $P^{\text{RC}}_{\hat{M}_4}(\hat{\mathbf{m}}_{4,(i)})$. Therefore, once the decoder has $\hat{M}_{4,(i)}$, it reconstructs $W^n_{(i)}$, $i \in [\![1,k-1]\!]$, via

$$P^{\text{SW}}_{W^n|M_3\hat{M}_4F\hat{X}^n}(\mathbf{w}_{(i)}|\mathbf{m}_{3,(i)}, \hat{\mathbf{m}}_{4,(i)}, \mathbf{f}_{(i)}, \hat{\mathbf{x}}_{(i)});$$

- Finally, the decoder generates $V^n_{(i)}$, $i \in [\![1,k-1]\!]$, letter by letter according to the distribution

$$P^{\text{RC}}_{V^n|W^nY^n}(\mathbf{v}_{(i)}|\mathbf{w}_{(i)}, \mathbf{y}_{(i)}).$$

For all $i \in [\![1, k-1]\!]$, the block-Markov coding scheme defines the joint distribution $P^{\text{RC}}_{(i)}$:

$$\begin{aligned}
P^{\text{RC}}_{(i)} :=& P^{\text{RC}}_{(U^n X^n \hat{X}^n Y^n V^n W^n M_1 M_2 M_3 M_4 \hat{M}_4 F)_{(i)}} \\
=& \bar{P}_{U^n}(\mathbf{u}_{(i)}) Q_{M_1}(\mathbf{m}_{1,(i)}) Q_{M_2}(\mathbf{m}_{2,(i)}) Q_{M_3}(\mathbf{m}_{3,(i)}) Q_F(\mathbf{f}_{(i)}) \\
& P^{\text{RB}}_{X^n|M_1 M_2 M_3 F}(\mathbf{x}_{(i)}|\mathbf{m}_{1,(i)}, \mathbf{m}_{2,(i)}, \mathbf{m}_{3,(i)}, \mathbf{f}_{(i)}) P^{\text{RB}}_{M_4|M_3 X^n U^n}(\mathbf{m}_{4,(i)}|\mathbf{m}_{3,(i)}, \mathbf{x}_{(i)}, \mathbf{u}_{(i)}) \\
& \bar{P}_{Y^n|X^n}(\mathbf{y}_{(i)}|\mathbf{x}_{(i)}) P^{\text{SW}}_{\hat{X}^n|M_1 Y^n}(\hat{\mathbf{x}}_{(i)}|\mathbf{m}_{1,(i)}, \mathbf{y}_{(i)}) P^{\text{RC}}_{\hat{M}_4}(\hat{\mathbf{m}}_{4,(i)}) \\
& P^{\text{SW}}_{W^n|M_3 \hat{M}_4 F \hat{X}^n}(\mathbf{w}_{(i)}|\mathbf{m}_{3,(i)}, \hat{\mathbf{m}}_{4,(i)}, \mathbf{f}_{(i)}, \hat{\mathbf{x}}_{(i)}) P^{\text{RC}}_{V^n|W^n Y^n}(\mathbf{v}_{(i)}|\mathbf{w}_{(i)}, \mathbf{y}_{(i)}).
\end{aligned} \tag{6.8}$$

Remark 6.7 - ***Last block is not coordinated.*** Observe that, even though the block-Markov algorithm is over k blocks, the last block is only used to convey information on the source at time $k-1$ through $M_{4,(k-1)}$ which is generated at time k. In fact, if k is large enough, Definition 6.1 allows us to coordinate only the first $k-1$ blocks.

Notice that we impose rate conditions $R_1 > H(X|Y)$ such that $\mathbb{P}\{\hat{X}^n_{(i)} \neq X^n_{(i)}\} \leq \delta(n)$ which in turn implies $\mathbb{P}\{\hat{M}_{2,(i)} \neq M_{2,(i)}\} \leq \delta(n)$, $\mathbb{P}\{\hat{M}_{4,(i)} \neq M_{4,(i)}\} \leq \delta(n)$. Moreover,

$$\begin{aligned}
&\mathbb{P}\{\hat{X}^n_{(1:k)} \neq X^n_{(1:k)}\} \leq \sum_{i=1}^{k} \mathbb{P}\{\hat{X}^n_{(i)} \neq X^n_{(i)}\} \leq k\delta(n), \\
&\mathbb{P}\{\hat{M}_{2,(1:k)} \neq M_{2,(1:k)}\} \leq \sum_{i=1}^{k} \mathbb{P}\{\hat{M}_{2,(i)} \neq M_{2,(i)}\} \leq k\delta(n), \\
&\mathbb{P}\{\hat{M}_{4,(1:k-1)} \neq M_{4,(1:k-1)}\} \leq \sum_{i=1}^{k-1} \mathbb{P}\{\hat{M}_{4,(i)} \neq M_{4,(i)}\} \leq (k-1)\delta(n),
\end{aligned} \tag{6.9}$$

where $k\delta(n)$ and $(k-1)\delta(n)$ vanish since $\delta(n)$ goes to zero exponentially fast.

Then, we apply the coupling property to

$$\begin{aligned}
&A = (U^n X^n Y^n V^n W^n M_1 M_2 M_3 M_4 F)_{(i)}, \quad A' = (U^n \hat{X}^n Y^n V^n W^n M_1 M_2 M_3 \hat{M}_4 F)_{(i)}, \\
&P_A = P^{\text{RC}}_{A}, \quad P_{A'} = P^{\text{RC}}_{A'}, \\
&\mathcal{A} = \mathcal{U} \times \mathcal{X} \times \mathcal{W} \times \mathcal{Y} \times [\![1, 2^{nR_1}]\!] \times [\![1, 2^{nR_2}]\!] \times [\![1, 2^{nR_3}]\!] \times [\![1, 2^{nR_4}]\!] \times [\![1, 2^{nR}]\!],
\end{aligned}$$

and, because of (6.9), the distribution $\hat{P}^{\text{RC}}_{(i)}$, defined as

$$\begin{aligned}
\hat{P}^{\text{RC}}_{(i)} :=& \hat{P}^{\text{RC}}_{(U^n X^n Y^n V^n W^n M_1 M_2 M_3 M_4 F)_{(i)}} \\
=& \bar{P}_{U^n}(\mathbf{u}_{(i)}) Q_{M_1}(\mathbf{m}_{1,(i)}) Q_{M_2}(\mathbf{m}_{2,(i)}) Q_{M_3}(\mathbf{m}_{3,(i)}) Q_F(\mathbf{f}_{(i)}) \\
& P^{\text{RB}}_{X^n|M_1 M_2 M_3 F}(\mathbf{x}_{(i)}|\mathbf{m}_{1,(i)}, \mathbf{m}_{2,(i)}, \mathbf{m}_{3,(i)}, \mathbf{f}_{(i)}) P^{\text{RB}}_{M_4|M_3 X^n U^n}(\mathbf{m}_{4,(i)}|\mathbf{m}_{3,(i)}, \mathbf{x}_{(i)}, \mathbf{u}_{(i)}) \\
& \bar{P}_{Y^n|X^n}(\mathbf{y}_{(i)}|\mathbf{x}_{(i)}) P^{\text{SW}}_{W^n|M_3 M_4 F X^n}(\mathbf{w}_{(i)}|\mathbf{m}_{3,(i)}, \mathbf{m}_{4,(i)}, \mathbf{f}_{(i)}, \mathbf{x}_{(i)}) P^{\text{RC}}_{V^n|W^n Y^n}(\mathbf{v}_{(i)}|\mathbf{w}_{(i)}, \mathbf{y}_{(i)}),
\end{aligned} \tag{6.10}$$

has almost the same statistics of $P^{\text{RC}}_{(i)}$:

$$\mathbb{V}(P^{\text{RC}}_{(i)}, \hat{P}^{\text{RC}}_{(i)}) \leq \delta(n).$$

To conclude, we want prove that

$$\lim_{n\to\infty} \mathbb{V}\left(P^{\text{RB}}, \hat{P}^{\text{RC}}_{(i)}\right) = 0.$$

The proof is similar to the achievability of Section 3.1.1 and it is detailed in Appendix B.1.3.

6.1.2 Outer bound

Consider a code (f^n, g^n) that induces a distribution $P_{U^nX^nY^nV^n}$ that is ε-close in total variational distance to the i.i.d. distribution $\bar{P}^{\otimes n}_{UXYV}$. Let the random variable T be uniformly distributed over the set $[\![1, n]\!]$ and independent of the sequence (U^n, X^n, Y^n, V^n, C). The variable T will serve as a random time index. The variable U_T is independent of T because U^n is an i.i.d. source [25, Section VII.B].

Bound on R_0 We have

$$\begin{aligned}
&nR_0 = H(C) \geq H(C|Y^n) \geq I(C; U^nV^nX^n|Y^n)\\
&= \sum_{t=1}^{n} I(U_tV_tX_t; C|U^{t-1}V^{t-1}X^{t-1}Y_{\sim t}Y_t)\\
&= \sum_{t=1}^{n} I(U_tV_tX_t; CY_{\sim t}U^{t-1}V^{t-1}X^{t-1}|Y_t) - \sum_{t=1}^{n} I(U_tV_tX_t; Y_{\sim t}U^{t-1}V^{t-1}X^{t-1}|Y_t)\\
&\overset{(a)}{\geq} \sum_{t=1}^{n} I(U_tV_tX_t; CY_{\sim t}U^{t-1}V^{t-1}X^{t-1}|Y_t) - nf(\varepsilon)\\
&\geq \sum_{t=1}^{n} I(U_tV_tX_t; CY^n_{t+1}U^{t-1}|Y_t) - nf(\varepsilon)\\
&= nI(U_TV_TX_T; CY^n_{T+1}U^{T-1}|Y_TT) - nf(\varepsilon)\\
&= nI(U_TV_TX_T; CY^n_{T+1}U^{T-1}T|Y_T) - nI(U_TV_TX_T; T|Y_T) - nf(\varepsilon)\\
&\overset{(b)}{\geq} nI(U_TV_TX_T; CY^n_{T+1}U^{T-1}T|Y_T) - 2nf(\varepsilon)
\end{aligned}$$

where (a) comes from Lemma 2.17 and (b) comes from [23, Lemma VI.3].

Information constraint We have

$$\begin{aligned}
&nI(U_T; CY^n_{T+1}U^{T-1}X_TT)\\
&\overset{(a)}{=} nI(U_T; CY^n_{T+1}U^{T-1}X_T|T) = \sum_{t=1}^{n} I(U_t; CY^n_{t+1}U^{t-1}X_t)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{t=1}^{n} I(U_t; CY_{t+1}^n U^{t-1}) + \sum_{t=1}^{n} I(U_t; X_t | CY_{t+1}^n U^{t-1}) \\
&\leq \sum_{t=1}^{n} I(U_t; CY_{t+1}^n U^{t-1}) + \sum_{t=1}^{n} I(U_t Y_{t+1}^n; X_t | CU^{t-1}) \\
&\overset{(b)}{=} \sum_{t=1}^{n} I(U_t; CY_{t+1}^n U^{t-1}) \overset{(c)}{=} \sum_{t=1}^{n} I(U_t; Y_{t+1}^n | U^{t-1} C) \\
&\overset{(d)}{=} \sum_{t=1}^{n} I(Y_t; U^{t-1} | Y_{t+1}^n C) \leq \sum_{t=1}^{n} I(Y_t; U^{t-1} Y_{t+1}^n C) \\
&\leq \sum_{t=1}^{n} I(Y_t; U^{t-1} Y_{t+1}^n C X_t) = nI(Y_T; U^{T-1} Y_{T+1}^n C X_T | T) \\
&\leq nI(Y_T; U^{T-1} Y_{T+1}^n C X_T T)
\end{aligned}
$$

where (a) follows from the i.i.d. nature of the source, (b) from the following Markov chain

$$X_t - (C, U^{t-1}) - (U_t, Y_{t+1}^n)$$

that holds because of the strictly causal nature of the encoder. Then, (c) comes from the fact that the source is generated i.i.d. and independent of C and (d) from the Csiszár sum identity.

We identify the auxiliary random variables W_t with (U^{t-1}, Y_{t+1}^n, C) for each $t \in [\![1, n]\!]$ and W with $(W_T, T) = (U^{T-1}, Y_{T+1}^n, C, T)$.

Identification of the auxiliary random variable For each $t \in [\![1, n]\!]$, W_t satisfies the following conditions:

$$
\begin{aligned}
& U_t \perp X_t, \\
& Y_t - X_t - (U_t, W_t), \\
& V_t - (Y_t, W_t) - (U_t, X_t).
\end{aligned}
\tag{6.11}
$$

Then, we have

$$
\begin{aligned}
& U_T \perp X_T \\
& Y_T - X_T - (U_T, W_T), \\
& V_T - (Y_T, W_T) - (U_T, X_T),
\end{aligned}
\tag{6.12}
$$

and, since $W = W_t$ when $T = t$, it implies

$$
\begin{aligned}
& U \perp X, \\
& Y - X - (U, W), \\
& V - (Y, W) - (U, X).
\end{aligned}
\tag{6.13}
$$

To complete the proof, the cardinality bound is proved in Appendix B.5.

6.2 Explicit coding schemes for coordination: polar codes

As for the case of non-causal encoder, polar codes serve as an explicit coding scheme for coordination. For simplicity, we only focus on the set of achievable distributions in $\mathcal{R}_{\text{SC,in}}$ for which the auxiliary variable W is binary. The scheme can be extended to the case of a non-binary random variable W using non-binary polar codes when $|\mathcal{W}|$ is a prime number [60].

Theorem 6.8 *The subset of the region $\mathcal{R}_{SC,in}$ defined in (6.1) for which the auxiliary random variable W is binary is achievable using polar codes, provided there exists an error-free channel of negligible rate between the encoder and decoder.*

Following the intuition provided in Section 5.2, to design the polar coding schemes we have to translate the rate conditions that descend from the random binning properties into very high and high entropy bits sets. In this section we detail the block-Markov polar coding schemes for strong coordination.

We assume that the sequences U^n, X^n, W^n, Y^n and V^n are jointly i.i.d. with distribution defined in (6.5):

$$\bar{P}_{U^n}\bar{P}_{X^n}\bar{P}_{W^n|U^nX^n}\bar{P}_{Y^n|X^n}\bar{P}_{V^n|W^nY^n}.$$

We propose an explicit coding scheme that induces a joint distribution close to (6.5) in total variational distance.

6.2.1 Polar coding scheme

Recall that for the strictly causal encoder, the strong coordination achievability proof in Section 6.1.1 requires to bin both the input of the channel X^n and the auxiliary random variable W^n, which translates into polarizing both X^n and W^n.

Polarize X Let $S^n = X^n G_n$ be the polarization of X^n, where G_n is the source polarization transform. For some $0 < \beta < 1/2$, let $\delta_n := 2^{-n^\beta}$ and define the very high and high entropy sets:

$$\begin{aligned}
\mathcal{V}_X &:= \left\{ j \in [\![1,n]\!] \mid H(S_j|S^{j-1}) > 1-\delta_n \right\}, \\
\mathcal{H}_X &:= \left\{ j \in [\![1,n]\!] \mid H(S_j|S^{j-1}) > \delta_n \right\}, \\
\mathcal{H}_{X|Y} &:= \left\{ j \in [\![1,n]\!] \mid H(S_j|S^{j-1}Y^n) > \delta_n \right\}.
\end{aligned} \tag{6.14}$$

Partition the set $[\![1,n]\!]$ into four disjoint sets:

$$\begin{aligned}
A_1 &:= \mathcal{V}_X \cap \mathcal{H}_{X|Y}, & A_2 &:= \mathcal{V}_X \cap \mathcal{H}^c_{X|Y}, \\
A_3 &:= \mathcal{V}^c_X \cap \mathcal{H}_{X|Y}, & A_4 &:= \mathcal{V}^c_X \cap \mathcal{H}^c_{X|Y}.
\end{aligned} \tag{6.15}$$

Remark 6.9 - *Cardinality of $\mathcal{V}_X$ and $\mathcal{H}_{X|Y}$*. It follows from Corollary 5.3 that:

- $\mathcal{V}_X \subset \mathcal{H}_X$ and $\lim_{n\to\infty} \frac{|\mathcal{H}_X \setminus \mathcal{V}_X|}{n} = 0$,
- $A_1 \cup A_2 = \mathcal{V}_X$ and $\lim_{n\to\infty} \frac{|\mathcal{V}_X|}{n} = H(X)$,
- $A_1 \cup A_3 = \mathcal{H}_{X|Y}$ and $\lim_{n\to\infty} \frac{|\mathcal{H}_{X|Y}|}{n} = H(X|Y)$.

Since $\lim_{n\to\infty} \frac{|A_2| - |A_3|}{n} = H(X) - H(X|Y) = I(X;Y) \geq 0$ this implies directly that for n large enough $|A_2| \geq |A_3|$.

Polarize W Let $Z^n = W^n G_n$ be the polarization of W^n and define:

$$\begin{aligned}
\mathcal{V}_{W|XU} &:= \left\{ j \in [\![1,n]\!] \mid H(Z_j|Z^{j-1}X^nU^n) > 1 - \delta_n \right\}, \\
\mathcal{H}_{W|XU} &:= \left\{ j \in [\![1,n]\!] \mid H(Z_j|Z^{j-1}X^nU^n) > \delta_n \right\}, \\
\mathcal{H}_{W|X} &:= \left\{ j \in [\![1,n]\!] \mid H(Z_j|Z^{j-1}X^n) > \delta_n \right\}.
\end{aligned} \tag{6.16}$$

Partition the set $[\![1,n]\!]$ into four disjoint sets:

$$\begin{aligned}
B_1 &:= \mathcal{V}_{W|XU} \cap \mathcal{H}_{W|X} = \mathcal{V}_{W|XU}, \\
B_2 &:= \mathcal{V}_{W|XU} \cap \mathcal{H}^c_{W|X} = \emptyset, \\
B_3 &:= \mathcal{V}^c_{W|XU} \cap \mathcal{H}_{W|X}, \\
B_4 &:= \mathcal{V}^c_{W|XU} \cap \mathcal{H}^c_{W|X} = \mathcal{H}^c_{W|X}.
\end{aligned} \tag{6.17}$$

Remark 6.10 - *Cardinality of $\mathcal{V}_{W|XU}$ and $\mathcal{V}_{W|X}$*. It follows from Corollary 5.3 that:

- $\mathcal{V}_{W|XU} \subset \mathcal{H}_{W|XU}$ and $\lim_{n\to\infty} \frac{|\mathcal{H}_{W|XU} \setminus \mathcal{V}_{W|XU}|}{n} = 0$,
- $B_1 = \mathcal{V}_{W|XU}$ and $\lim_{n\to\infty} \frac{|\mathcal{V}_{W|XU}|}{n} = H(W|XU)$,
- $B_4 = \mathcal{H}^c_{W|X}$ and $\lim_{n\to\infty} \frac{|\mathcal{V}^c_{W|X}|}{n} = 1 - H(W|X)$,
- $B_1 \cup B_3 = \mathcal{V}_{W|X}$ and $\lim_{n\to\infty} \frac{|\mathcal{V}_{W|X}|}{n} = H(W|X)$,
- $B_3 \cup B_4 = \mathcal{V}^c_{W|XU}$ and $\lim_{n\to\infty} \frac{|\mathcal{V}^c_{W|XU}|}{n} = 1 - H(W|XU)$.

Note that

$$H(W|X) - H(W|XU) = I(W;U|X) = I(WX;U) \geq 0$$

and $|B_3|/n$ tends to $I(WX;U)$. Since $I(WX;Y) = I(X;Y)$, the inequality $I(WX;U) \leq I(WX;Y)$ implies directly that for n large enough $|B_3| \leq |A_2| - |A_3|$.

Algorithm 3: Encoding algorithm at Node 1

Input : $(U^n_{(0)}, \dots, U^n_{(k)})$, local randomness (uniform random bits) M and common randomness $C = (\{C_i\}_{i=1,\dots,k}, \{C'_i\}_{i=1,\dots,k}, \bar{C}', \{K_i\}_{i=1,\dots,k}, \{K'_i\}_{i=1,\dots,k})$ shared with Node 2:

- C_i of size $|A_1|$ and K_i of size $|A_3|$
- $\bar{C}'$ of size $|B'_1|$, C'_i of size $|B_1 \setminus B'_1|$ and K'_i of size $|B_3|$

Output: $(S^n_{(1)}, \dots, S^n_{(k)})$, $(Z^n_{(1)}, \dots, Z^n_{(k)})$

if $i = 1$ **then**

$S_{(1)}[A_1] \longleftarrow C_i,\ S_{(1)}[A_2] \longleftarrow M$

for $j \in A_3 \cup A_4$ **do**

Successively draw the bits $S_{j,(1)}$ according to

$$\bar{P}_{S_j|S^{j-1}}(S_{(i),j}|S^{j-1}_{(i)}) \tag{6.18}$$

$Z_{(1)}[B'_1] \longleftarrow \bar{C}' \quad Z_{(1)}[B_1 \setminus B'_1] \longleftarrow C'_i$

for $j \in B_3 \cup B_4$ **do**

Given $U^n_{(1)}$, successively draw the bits $Z^j_{(1)}$ according to

$$\bar{P}_{Z_j|Z^{j-1}X^nU^n}(Z_{(i),j}|Z^{j-1}_{(i)}X^n_{(i)}U^n_{(i-1)}) \tag{6.19}$$

for $i = 2, \dots, k$ **do**

$$S_{(i)}[A_1] \longleftarrow C_i, \quad S_{(i)}[A'_2] \longleftarrow M,$$
$$S_{(i)}[B'_3] \longleftarrow Z_{(i-1)}[B_3] \oplus K_{i-1}, \quad S_{(i)}[A'_3] \longleftarrow S_{(i-1)}[A_3] \oplus K_{i-1}$$

for $j \in A_3 \cup A_4$ **do**

Successively draw the bits $S_{(i),j}$ according to (6.18)

$Z_{(i)}[B'_1] \longleftarrow \bar{C}', \quad Z_{(i)}[B_1 \setminus B'_1] \longleftarrow C'_i$

for $j \in B_3 \cup B_4$ **do**

Succ. draw the bits $Z_{(i),j}$ according to (6.19)

Encoding The encoder observes $U^n_{(0:k)} := (U^n_{(0)}, U^n_{(1)}, \dots, U^n_{(k)})$, where $U^n_{(0)}$ is a uniform random sequence and $U^n_{(i)}$ for $i \in [\![1, k]\!]$ are k blocks of the source. It then generates for each block $i \in [\![1, k]\!]$ random variables $S^n_{(i)}$ and $Z^n_{(i)}$ following the procedure described in Algorithm 3, The chaining construction proceeds as follows:

- The bits in $A_1 \subset \mathcal{V}_X$ in block $i \in [\![1, k]\!]$ are chosen with uniform probability using a uniform randomness source C_i shared with the decoder;
- In the first block the bits in $A_2 \subset \mathcal{V}_X$ are chosen with uniform probability using a local randomness source M;
- Let $B'_1 := \mathcal{V}_{W|UXYV}$, observe that B'_1 is a subset of B_1 since $\mathcal{V}_{W|UXYV} \subset \mathcal{V}_{W|XU}$. The bits in $B'_1 \subset \mathcal{V}_{W|XU}$ in block $i \in [\![1, k]\!]$ are chosen with uniform probability using a

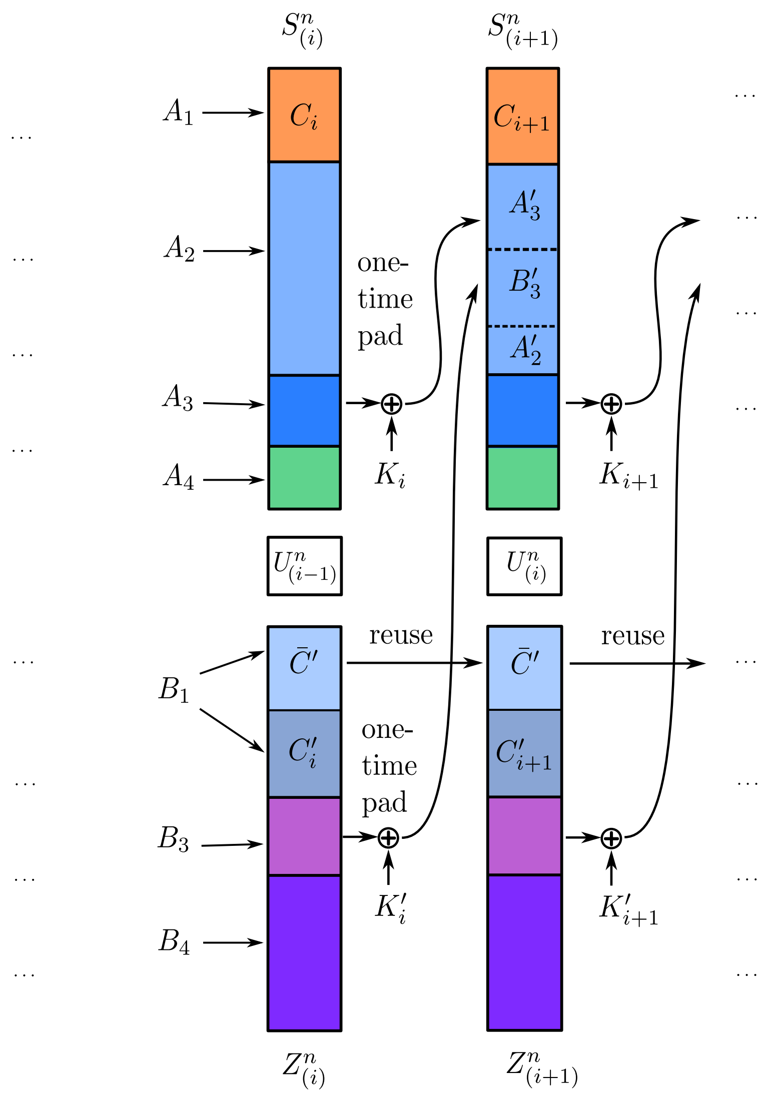

Figure 6.3: Chaining construction for block Markov encoding

uniform randomness source $\bar{C}'$ shared with the decoder, and their value is reused over all blocks;

- The bits in $B_1 \setminus B'_1 \subset \mathcal{V}_{W|XU}$ in block $i \in [\![1, k]\!]$ are chosen with uniform probability using a uniform randomness source C'_i shared with the decoder;

- The bits in $A_3 \cup A_4$ and $B_3 \cup B_4$ are generated according to the previous bits using successive cancellation encoding as in Definition 5.6. Note that it is possible to sample efficiently from $\bar{P}_{S_j|S^{j-1}}$ and $\bar{P}_{Z_j|Z^{j-1}X^nU^n}$ (given U^n and X^n) respectively;

- From the second block, the encoder generates the bits of A_2 in the following way. Let A'_3 and B'_3 be two disjoint subsets of A_2 such that $|A'_3| = |A_3|$ and $|B'_3| = |B_3|$. The existence of those disjoint subsets is guaranteed by Remark 6.9 and Remark 6.10. The bits of A_3 and B_3 in block i are used as A'_3 and B'_3 in block $i+1$ using one-time pads with keys K_i and K'_i respectively:

$$S_{(i+1)}[A'_3] = S_{(i)}[A'_3] \oplus K_i \quad i = 1, \dots, k-1,$$
$$S_{(i+1)}[B'_3] = Z_{(i)}[A'_3] \oplus K'_i \quad i = 1, \dots, k-1.$$

Thanks to the Crypto Lemma, if we choose K_i of size $|A_3|$ and K'_i of size $|B_3|$ to be uniform random keys, the bits in A'_3 and B'_3 in the block $i+1$ are uniform. The bits in $A'_2 := A_2 \setminus (A'_3 \cup B'_3)$ are chosen with uniform probability using the local randomness source M.

The encoder then computes $X_i^n = S_i^n G_n$ for $i = 1, \dots, k$ and sends it over the channel. As in [12], to deal with unaligned indices, chaining also requires in the last encoding block to transmit $S_{(k)}[A_3] \cup Z_{(k)}[B_3]$ to the decoder. Hence the coding scheme requires an error-free channel between the encoder and decoder which has negligible rate since $|S_{(k)}[A_3] \cup Z_{(k)}[B_3]| \leq |\mathcal{H}_X|$ and

$$\lim_{\substack{n\to\infty \\ k\to\infty}} \frac{|\mathcal{H}_X|}{kn} = \lim_{k\to\infty} \frac{H(X)}{k} = 0.$$

Decoding The decoder observes $(Y_{(1)}^n, \dots, Y_{(k)}^n)$ and $S_{(k)}[A_3] \cup Z_{(k)}[B_3]$ allows it to decode in reverse order. The decoding algorithm, detailed in Algorithm 4, proceeds as follows:

- In every block $i \in [\![1, k]\!]$, the decoder has access to $\hat{S}_{(i)}[A_1] \subseteq \hat{S}_{(i)}[\mathcal{H}_{X|Y}]$ and $\hat{Z}_{(i)}[B_1] \subseteq \hat{Z}_{(i)}[\mathcal{H}_{W|X}]$ because the bits in A_1 and B_1 correspond to shared randomness $(\{C_i\}_{i=1,\dots,k}$, $\{C'_i\}_{i=1,\dots,k}, \bar{C}')$,
- In block $i \in [\![1, k-1]\!]$ the bits in A_3 and B_3 are obtained by successfully recovering A_2 in block $i+1$, which is possible because the keys of the one-time pad are part of the common randomness;
- From $Y_{(i)}^n$ and $\hat{S}_{(i)}[A_1 \cup A_3]$ the successive cancellation decoder defined in Definition 5.4 can retrieve $\hat{S}_{(i)}[A_2 \cup A_4]$ and $\hat{Z}_{(i)}[B_4]$. Note that, by Lemma 5.7, Z^n is equal to $\hat{Z}^n$ and S^n is equal to $\hat{S}^n$ with high probability.
- The decoder computes $\hat{W}_{(i)}^n = \hat{Z}_{(i)}^n G_n$
- Finally, the decoder generates $V_{(i)}^n$ symbol by symbol using

$$P_{V_{(i),j}|\hat{W}(i),jY_{(i),j}}(v|w,y) = \bar{P}_{V|WY}(v|w,y).$$

Rate of common randomness The rate of common randomness is $I(W;UXV|Y)+H(X|WY)$ since:

$$\lim_{\substack{n\to\infty \\ k\to\infty}} \frac{k|A_1|+(k-1)|A_3|+k|B_1|+(k-1)|B_3|-(k-1)|B'_1|}{kn}$$
$$= \lim_{n\to\infty} \frac{|A_1|+|A_3|+|B_1|+|B_3|-|B'_1|}{n}$$

Algorithm 4: Decoding algorithm at Node 2

Input : $(Y_{(1)}^n, \ldots, Y_{(k)}^n)$, $S_{(k)}[A_3] \cup Z_{(k)}[B_3]$ and C common randomness shared with Node 1

Output: $(\hat{S}_{(1)}^n, \ldots, \hat{S}_{(k)}^n)$, $(\hat{Z}_{(1)}^n, \ldots, \hat{Z}_{(k)}^n)$

for $i = k, \ldots, 1$ **do**

$$\hat{S}_{(i)}[A_1] \longleftarrow C_i$$
$$\hat{Z}_{(i)}[B_1'] \longleftarrow \bar{C}' \quad \hat{Z}_{(i)}[B_1 \backslash B_1'] \longleftarrow C_i'$$

if $i \neq k$ **then**

$$\hat{S}_{(i)}[A_3] \leftarrow \hat{S}_{(i+1)}[A_3'] \oplus K_i$$
$$\hat{Z}_{(i)}[B_3] \leftarrow \hat{S}_{(i+1)}[B_3'] \oplus K_i'$$

for $j \in A_2 \cup A_4$ **do**

Successively draw the bits according to

$$\hat{S}_{(i),j} = \begin{cases} 0 & \text{if } L_n(Y_{(i)}^n, \hat{S}_{(i)}^{j-1}) \geq 1 \\ 1 & \text{else} \end{cases}$$

$$\text{where} \quad L_n(Y_{(i)}^n, \hat{S}_{(i)}^{j-1}) = \frac{\bar{P}_{S_j|S^{j-1}Y^n}\left(0|\hat{S}_{(i)}^{j-1} Y_{(i)}^n\right)}{\bar{P}_{S_j|S^{j-1}Y^n}\left(1|\hat{S}_{(i)}^{j-1} Y_{(i)}^n\right)}$$

for $j \in B_4$ **do**

Successively draw the bits according to

$$\hat{Z}_{(i),j} = \begin{cases} 0 & \text{if } L_n(X_{(i+1)}^n, \hat{Z}_{(i)}^{j-1}) \geq 1 \\ 1 & \text{else} \end{cases}$$

$$= H(X|Y) + H(W|X) - H(W|UXYV)$$
$$\stackrel{(a)}{=} I(W; UXV|Y) + H(X|YW)$$

where (a) has been proved in (B.19).

Remark 6.11 - ***Polar coding for empirical coordination.*** With a similar coding scheme, we can prove that polar codes are coordination codes for empirical coordination. The two schemes only differ on the amount of common randomness that it is possible to recycle, and on the fact that empirical coordination always allows to loose coordination on one block, making the request of an error-free channel of negligible rate unnecessary. The complete scheme and the achievability proof for empirical coordination are in Appendix C.3.

6.2.2 Achievability

We note with P the joint distribution induced by the encoding and decoding algorithm of Section 6.2.1. and we prove that this is an explicit scheme for strong coordination.

Coordination in one block The proof requires a few steps. Similarly to the achievability proof of Theorem 5.10, we first prove the following lemma, which is a direct consequence of Lemma 5.9.

Lemma 6.12 *For every $i \in [\![1, k]\!]$, we have*

$$\mathbb{V}(P_{U^n_{(i)} X^n_{(i)} W^n_{(i)}}, \bar{P}^{\otimes n}_{UXW}) \leq \delta_n^{(6)}$$

where $\delta_n^{(6)} := 2\sqrt{\log 2}\sqrt{n\delta_n}$

Proof. We prove that in each block $i \in [\![1, k]\!]$

$$\mathbb{D}\left(\bar{P}^{\otimes n}_{UXW} \middle\| P_{U^n_{(i)} X^n_{(i)} W^n_{(i)}}\right) = 2n\delta_n. \tag{6.20}$$

In fact, we have

$$\mathbb{D}(\bar{P}^{\otimes n}_{UXW} \| P_{U^n_{(i)} X^n_{(i)} W^n_{(i)}}) = \mathbb{D}(\bar{P}_{X^n|U^n} \| P_{X^n_{(i)}|U^n_{(i)}} | \bar{P}_{U^n}) + \mathbb{D}(\bar{P}_{W^n|X^n U^n} \| P_{W^n_{(i)}|X^n_{(i)} U^n_{(i)}} | \bar{P}_{X^n U^n})$$

We call D_1 and D_2 the first and the second term. Then:

$$D_1 \overset{(a)}{=} \mathbb{D}(\bar{P}_{X^n} \| P_{X^n_{(i)}}) \overset{(b)}{=} \mathbb{D}(\bar{P}_{S^n} \| P_{S^n_{(i)}}) \overset{(c)}{=} \mathbb{D}(\bar{P}_{S[A_1 \cup A_2]} \| P_{S_{(i)}[A_1 \cup A_2]}) \overset{(d)}{\leq} n\delta_n$$

where (a) follows from the fact that X is independent of U, (b) from the invertibility of G_n, (c) from the fact that we can consider only the very high entropy bits as in (5.8), and (d) from Lemma 5.9. Similarly,

$$D_2 \overset{(a)}{=} \mathbb{D}(\bar{P}_{Z^n|X^n U^n} \| P_{Z^n_{(i)}|X^n_{(i)} U^n_{(i)}} | \bar{P}_{X^n U^n}) \overset{(b)}{=} \mathbb{D}(\bar{P}_{Z[B_1]|X^n U^n} \| P_{Z_{(i)}[B_1]|X^n_{(i)} U^n_{(i)}} | \bar{P}_{X^n U^n}) \overset{(c)}{\leq} n\delta_n,$$

where (a) comes from the invertibility of G_n, (b) from the fact that we can consider only the very high entropy bits as in (5.8), and (c) follows from Lemma 5.9.Therefore, applying Pinsker's inequality to (5.7) we have

$$\mathbb{V}(P_{U^n_{(i)} X^n_{(i)} W^n_{(i)}}, \bar{P}^{\otimes n}_{UXW}) \leq 2\sqrt{\log 2}\sqrt{n\delta_n} := \delta_n^{(6)} \to 0. \qquad \square$$

Note that $Y^n_{(i)}$ is generated symbol by symbol via the channel $\bar{P}_{Y|X}$. By Lemma 2.13, for each $i \in [\![1, k]\!]$,

$$\mathbb{V}(P_{U^n_{(i)} W^n_{(i)} X^n_{(i)} Y^n_{(i)}}, \bar{P}^{\otimes n}_{UWXY}) = \mathbb{V}(P_{U^n_{(i)} W^n_{(i)}}, \bar{P}^{\otimes n}_{UW}) \leq \delta_n^{(6)} \tag{6.21}$$

and therefore the left-hand side of (6.21) vanishes.

Observe that $V_{(i)}^n$ is generated using $\hat{W}_{(i)}^n$ (i.e. the estimate of $W_{(i)}^n$ at the decoder) and not $W_{(i)}^n$. By the triangle inequality for all $i \in [\![1,k]\!]$

$$\mathbb{V}(P_{U_{(i)}^n \hat{W}_{(i)}^n X_{(i)}^n Y_{(i)}^n}, \bar{P}_{UWXY}^{\otimes n}) \leq \mathbb{V}(P_{U_{(i)}^n \hat{W}_{(i)}^n X_{(i)}^n Y_{(i)}^n}, P_{U_{(i)}^n W_{(i)}^n X_{(i)}^n Y_{(i)}^n}) + \mathbb{V}(P_{U_{(i)}^n W_{(i)}^n X_{(i)}^n Y_{(i)}^n}, \bar{P}_{UWXY}^{\otimes n}). \tag{6.22}$$

We have proved in (6.21) that the second term of the right-hand side in (6.22) goes to zero, we show that the first term tends to zero as well. To do so, we apply the coupling property to

$$\begin{aligned} A &= U_{(i)}^n \hat{W}_{(i)}^n X_{(i)}^n Y_{(i)}^n, & A' &= U_{(i)}^n W_{(i)}^n X_{(i)}^n Y_{(i)}^n,\\ P &= P_{U_{(i)}^n \hat{W}_{(i)}^n X_{(i)}^n Y_{(i)}^n}, & P' &= P_{U_{(i)}^n W_{(i)}^n X_{(i)}^n Y_{(i)}^n}, \end{aligned}$$

on $\mathcal{A} = \mathcal{U} \times \mathcal{W} \times \mathcal{X} \times \mathcal{Y}$. Since by Lemma 5.7 we have

$$p_e := \mathbb{P}\left\{\hat{W}_{(i)}^n \neq W_{(i)}^n\right\} \leq \delta_n,$$

we find that $\mathbb{V}\left(P_{U_{(i)}^n \hat{W}_{(i)}^n X_{(i)}^n Y_{(i)}^n}, P_{U_{(i)}^n W_{(i)}^n X_{(i)}^n Y_{(i)}^n}\right) \leq 2p_e$, and therefore

$$\mathbb{V}\left(P_{U_{(i)}^n \hat{W}_{(i)}^n X_{(i)}^n Y_{(i)}^n}, \bar{P}_{UWXY}^{\otimes n}\right) \leq 2p_e + \delta_n^{(6)} := \delta_n^{(7)} \to 0.$$

Since V_i^n is generated symbol by symbol from $\hat{W}_i^n$ and Y_i^n, we apply Lemma 2.13 again and find

$$\mathbb{V}\left(P_{U_{(i)}^n \hat{W}_{(i)}^n X_{(i)}^n Y_{(i)}^n V_{(i)}^n}, \bar{P}_{UWXYV}^{\otimes n}\right) \leq \delta_n^{(7)} \to 0. \tag{6.23}$$

□

Coordination of all blocks First, we want to show that two consecutive blocks are almost independent.

Lemma 6.13 *For $i \in [\![2,k]\!]$, we have*

$$\mathbb{V}\left(P_{L_{i-1:i}\bar{C}'}, P_{L_{i-1}\bar{C}'} P_{L_i}\right) \leq \delta(n).$$

Proof. For $i \in [\![2,k]\!]$, we have

$$\begin{aligned} &\mathbb{D}\left(P_{L_{i-1:i}\bar{C}'} \| P_{L_{i-1}\bar{C}'} P_{L_i}\right)\\ &= I(L_{i-1}\bar{C}'; L_i) = I(L_i; \bar{C}') + I(L_{i-1}; L_i|\bar{C}')\\ &\overset{(a)}{=} I(L_i; \bar{C}') = I(L_i; Z_{(i)}[B_1'])\overset{(b)}{=}|B'| - H(Z_{(i)}[B_1']|L_i)\\ &\overset{(c)}{=}|B_1'| - H(Z[B_1']|L) + \delta_n^{(8)} \overset{(d)}{\leq} |B_1'| - \sum_{j \in B_1'} H(Z_j|Z^{j-1}L) + \delta_n^{(8)} \end{aligned}$$

$$\overset{(e)}{\leq} |B_1'| - |B_1'|(1-\delta_n) + \delta_n^{(8)} \leq n\delta_n + \delta_n^{(8)}.$$

To prove (a), observe that, because of the one-time pads on A_3 and B_3, $(U^n_{(i-1)}, X^n_{(i-1)}, Y^n_{(i-1)}, V^n_{(i-1)})$ and $(U^n_{(i)}, X^n_{(i)}, Y^n_{(i)}, V^n_{(i)})$ are dependent only through the recycled common randomness $\bar{C}'$. Therefore, the Markov chain $L_{i-1} - \bar{C}' - L_i$ holds. Then, (b) comes from from the fact that the bits in B_1' are uniform. To prove (c), note that

$$\begin{aligned} H(Z_{(i)}[B_1']|L_i) - H(Z[B_1']|L) &= H(Z_{(i)}[B_1']L_i) - H(Z[B_1']L) - H(L_i) + H(L) \\ &\overset{(f)}{\leq} \delta_n^{(7)} \log \frac{|\mathcal{U} \times \mathcal{X} \times \mathcal{W} \times \mathcal{Y} \times \mathcal{V}|}{\delta_n^{(7)}} + \delta_n^{(7)} \log \frac{|\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}|}{\delta_n^{(7)}} \\ &\leq 2\delta_n^{(7)}(\log |\mathcal{U} \times \mathcal{X} \times \mathcal{W} \times \mathcal{Y} \times \mathcal{V}| - \log \delta_n^{(7)}) := \delta_n^{(8)} \end{aligned}$$

where (f) comes from Lemma A.16 since by (6.23) we have

$$\mathbb{V}\left(P_{L_i}, \bar{P}^{\otimes n}_{UXYV}\right) \leq \mathbb{V}\left(P_{L_i W^n_{(i)}}, \bar{P}^{\otimes n}_{UWXYV}\right) \leq \delta_n^{(7)}$$

that vanishes as n goes to infinity.

Finally, (d) is true because conditioning does not increase entropy and (e) comes by definition of the set B_1'. Then we conclude with Pinsker's inequality. □

Now that we have the asymptotical independence of two consecutive blocks, we conclude using Lemma 5.19 and Lemma 5.20. □

Conclusions and perspectives

In this thesis, we have addressed the problem of strong coordination in a two-node network comprised of an information source and a noisy channel, in which both nodes have access to a common source of randomness. We have considered a general setting in which state information and side information about the source may or may not be available at the decoder. By exploiting the properties of random binning, we have outlined a general joint source-channel achievability scheme, which has the benefit of being rather simple and flexible enough to be adapted to different requests.
Despite the fact that in the general case the inner bound does not match the outer bound, the achievability scheme can be applied directly to three special cases for which we can exactly characterize the strong coordination region. This suggests that our random binning scheme is quite general and that we should look for a refinement of the outer bound to characterize the region. However, this is a difficult problem and has not been solved even for empirical coordination [44].

In this work, we have only focused on the case of two-node networks. For more general networks, the current understanding of the fundamental limits of coordination remains rather limited, and characterizing the strong coordination region (or even finding bounds for it) is still an open problem. For the simpler case of error-free links, several works have gone in this direction, see for example [25, 23, 10, 9, 66, 67]. We observe that, even with this less stringent constraint, characterizing the region for more complicated network topologies becomes quickly very difficult.

On a different note, the characterization of the region in three special cases has brought light on one of the most interesting consequences of strong coordination, which is the fact that under particular circumstances it offers security "for free". In this thesis, we have proved that strong coordination implies strong secrecy when an eavesdropper intercepts the same channel output as the legitimate receiver. However, for the more general model in which the eavesdropper observes a different signal from the legitimate receiver, we were only able to derive an inner bound, and investigating this aspect is one of the most promising directions for further work.

Furthermore, perhaps one of the most intriguing aspects is to use the information theoretic approach that we have developed to study situations in which the agents in a network have diverging (and perhaps selfish) interests, connecting strong coordination to strategical communication [48–51]. This perspective can well adapt to situations in which agents want to optimize their own cost without regard for the welfare of the overall system.

Finally, a whole chapter of this thesis is dedicated to designing polar coding schemes for strong coordination. The construction of the codes relies on the analogy between the properties of source polarization and random binning, but it involves a chaining construction over k blocks: in order to achieve coordination not only the length n of the polarized sequences has to tend to infinity, but also the number of blocks has to go to infinity. Although this makes sense from an information-theoretic point of view, it is not practical for delay-constrained applications. An improvement of our polar coding scheme (or even a different explicit coding scheme) in which no chaining is required is left for future work.

A | Properties and proofs of preliminary results

A.1 Information theoretic properties

In this section, we list the information theoretic definitions and properties the we use throughout this document. Since they are all well-known results, the proofs are omitted.

Definition A.1 - ***Binary entropy function [18]*** *The binary entropy function h is defined as the entropy of a Bernoulli process with probability p:*

$$h(p) := -p \log p - (1-p) \log (1-p).$$

Lemma A.2 - ***Chain rule for entropy [26]*** *Let A^n be a random vector, then*

$$H(A^n) = \sum_{i=1}^{n} H(A_i|A^{i-1}).$$

Lemma A.3 - ***Chain rule for mutual information [26]*** *Given a random vector A^n and a random variable B,*

$$I(A^n; B) = \sum_{i=1}^{n} I(A_i; B|A^{i-1}).$$

Definition A.4 - ***Conditional K-L divergence [20]*** *Given A and B two random variables, and P, P' two probability mass functions, then*

$$\mathbb{D}(P_{B|A} \| P'_{B|A} | P_A) := \sum_{a \in \mathcal{A}} P_A(a) \mathbb{D}(P_{B|A}(\cdot|a) \| P'_{B|A}(\cdot|a))$$

is the conditional K-L divergence.

Lemma A.5 - ***Chain rule for K-L divergence [20]*** *Given A and B two random variables, and P, P' two probability mass functions, then*

$$\mathbb{D}(P_{AB} \| P'_{AB}) = \mathbb{D}(P_{B|A} \| P'_{B|A} | P_A).$$

Definition A.6 - ***Markov Chain [26]*** *We say that $A - B - C$ form a Markov chain if C is condi-*

tionally independent of A given B, that is

$$P_{ABC} = P_A P_{B|A} P_{C|B}$$

or equivalently

$$I(A;C|B) = 0.$$

In particular, $A - B - f(B)$ form a Markov Chain.

Lemma A.7 - ***Data Processing Inequality [26]*** *If $A-B-C$ form a Markov chain, then $I(A;C) \leq I(A;B)$. In particular, this implies that $I(A;f(B)) \leq I(A;B)$.*

Definition A.8 - ***Wyner common information [69]*** *Given a pair of random variables (A,B), the Wyner common information of the pair (A,B) is*

$$C(A;B) := \min_{A-W-B} I(AB;W).$$

Definition A.9 - ***Conditional common information [37, Definition 1]*** *Given a triple of random variables (A,B,Y), the conditional common information of the pair (A,B) given Y is*

$$C(A;B|Y) := \min_{A-(W,Y)-B} I(AB;W|Y).$$

Lemma A.10 - ***Fano's Inequality [27]*** *Given three random variables A, B, and $\hat{A}$, with $\mathcal{A} = \hat{\mathcal{A}}$ and such that $A - B - \hat{A}$ form a Markov chain with $p_e := \mathbb{P}\{A \neq \hat{A}\}$. Then,*

$$H(A|B) \leq H(A|\hat{A}) \leq h(p_e) + p_e \log|\mathcal{A}| \leq 1 + p_e \log|\mathcal{A}|$$

where h is the binary entropy function.

Lemma A.11 - ***Many uses of a DMC do not increase the capacity per transmission [18, Lemma 7.9.2]*** *Given a pair of random vectors (A^n, B^n), where B is the output of a discrete memoryless channel $P_{B|A}$ of capacity C. Then,*

$$I(A^n;B^n) \leq nC \quad \textit{for all } P_{A^n}.$$

Lemma A.12 - ***Csiszár Sum Identity [26]*** *Given two random vectors A^n and B^n, and C a random variable, we have*

$$\sum_{i=1}^{n} I(A_{i+1}^n; B_i | B^{i-1}, C) = \sum_{i=1}^{n} I(B^{i-1}; A_i | A_{i+1}^n, C)$$

where $A_{n+1}^n, B^0 = \emptyset$.

Lemma A.13 - ***Crypto Lemma [8, Lemma 3.1]*** *Let $(\mathcal{G}, \oplus)$ be a compact abelian group with binary operation $\oplus$ and let $X = M \oplus K$, where M and K are random variables over $\mathcal{G}$ and K is* *independent of M and uniform over $\mathcal{G}$. Then X is independent of M and uniform over $\mathcal{G}$.*

Lemma A.14 - ***Pinsker's inequality [59]*** *Given a pair of random variables (A, B) with distribution P_A and P_B respectively, we have*

$$\mathbb{V}(P_A, P_B) \leq \sqrt{2 \log 2} \sqrt{\mathbb{D}(P_A \| P_B)}.$$

Lemma A.15 - ***Coloring Lemma [19, Lemma 1]*** *Given a pair of random variables (A, B) with joint distribution P_{AB}, marginals P_A and P_B and $|\mathcal{A}| \geq 4$, we have*

$$\frac{1}{2 \log 2} \mathbb{V}(P_{AB}, P_A P_B)^2 \leq I(A; B) \leq \mathbb{V}(P_{AB}, P_A P_B) \log \frac{|\mathcal{A}|}{\mathbb{V}(P_{AB}, P_A P_B)}$$

where the left-hand side is due to Pinsker's inequality.

Lemma A.16 - ***[20, Lemma 2.7]*** *Let P and P' two probability mass functions on $\mathcal{A}$ such that $\mathbb{V}(P, P') = \varepsilon \leq 1/2$, then*

$$|H(P) - H(P')| \leq \varepsilon \log \frac{|\mathcal{A}|}{\varepsilon}.$$

Definition A.17 - ***Histogram of a sequence [26]*** *Let A^n be a random vector and let $\mathbf{a} \in A^n$ be a sequence with $a_i \in \mathcal{A}$, $i \in [\![1, n]\!]$. The histogram or empirical distribution of $\mathbf{a}$ is*

$$T_{A^n}(a) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{a_i = a\} \quad \text{for } a \in \mathcal{A}.$$

Now, we introduce the notion of typicality [26].

Definition A.18 - ***Typical sequences [26]*** *Let A^n be a random vector such that, for every $i \in [\![1, n]\!]$, A_i is generated i.i.d. according to P_A. For $\varepsilon \in (0, 1)$, the set of typical sequences $\mathcal{T}_\varepsilon^{(n)}$ with respect to P_A is the set of sequences $\mathbf{a} \in \mathcal{A}^n$ with the property*

$$|P_A(a) - T_{A^n}(a)| \leq \varepsilon P_A(a).$$

Theorem A.19 - ***Properties of typical sequences [26]*** *The following properties hold:*

1. *if $\mathbf{a} \in \mathcal{T}_\varepsilon^{(n)}$, then $2^{-nH(A)(1+\varepsilon)} \leq P_{A^n}(\mathbf{a}) \leq 2^{-nH(A)(1-\varepsilon)}$;*

2. *let A^n be a random vector such that, for every $i = 1, \ldots, n$, A_i is generated i.i.d. according to P_A. Then, with high probability, the empirical distribution does not deviate much from the true distribution and the probability of a sequence being typical is very high:*

$$\lim_{n \to \infty} \mathbb{P}\{A^n \in \mathcal{T}_\varepsilon^{(n)}\} = 1;$$

3. *$|\mathcal{T}_\varepsilon^{(n)}| \leq 2^{nH(A)(1+\varepsilon)}$;*

4. *for n sufficiently large, $|\mathcal{T}_\varepsilon^{(n)}| \geq (1 - \varepsilon) 2^{nH(A)(1-\varepsilon)}$.*

Lemma A.20 - ***Conditional Typicality Lemma [26]*** *Given a pair of random variables* (A, B) *with joint distribution* P_{AB}*, consider* $\mathbf{a} \in \mathcal{T}_{\varepsilon'}^{(n)}$ *and* B^n *generated according to* $P_{B^n|A^n}(\mathbf{b}|\mathbf{a}) = \prod_{i=1}^n P_{B|A}(b_i|a_i)$*. Then for every* $\varepsilon > \varepsilon'$*,*

$$\lim_{n\to\infty} \mathbb{P}\{(\mathbf{a}, B^n) \in \mathcal{T}_{\varepsilon}^{(n)}(A, B)\} = 1.$$

The definition and properties above correspond to the notion of *robust typicality* [26]. Another widely used notion is *weak typicality* or *entropy typicality* [18].

Definition A.21 - ***Entropy typical sequences [18]*** *The set of weakly typical or entropy typical sequences* $\mathcal{A}_{\varepsilon}^{(n)}$ *with respect to* P_A *is the set of sequences* $\mathbf{a} \in \mathcal{A}^n$ *with the property*

$$\left| \frac{1}{n} \log \frac{1}{P_{A^n}(\mathbf{a})} - H(A) \right| \leq \varepsilon.$$

Observe that the latter is a weaker notion, since $T_{\varepsilon'}^{(n)} \subseteq \mathcal{A}_{\delta}^{(n)}$ for $\delta = \varepsilon H(A)$, while the other inclusion does not hold in general. For example, every binary sequence is entropy typical with respect to the uniform distribution $Ber(1/2)$, but not all of them are typical [26].

A.2 Discussion on the boundaries of the strong coordination region

Note that, instead of Definition (2.4), we could have used the slightly different definition of [23] in which the authors define the achievable region as the closure of the set of rates for a given distribution. The latter is a more precise characterization of the achievable set and requires additional precision in the proof. Here, we suppose that we are in the setting of Figure 2.1, but these definitions have obvious generalizations to other networks.

Definition A.22 - ***Achievability for strong coordination and strong coordination region [23]*** *A pair of rates* (R, R_0) *is achievable for synthesizing a memoryless channel* $\bar{P}_{V|U}$ *with input distribution* $\bar{P}_U$ *if there exists a sequence* (f_n, g_n) *of encoders-decoders with rate of common randomness* R_0*, such that*

$$\lim_{n\to\infty} \mathbb{V}\left(P_{U^nV^n}, \bar{P}_{UV}^{\otimes n}\right) = 0$$

where $P_{U^nV^n}$ *is the joint distribution induced by the code.*

Then, the strong coordination region $\tilde{\mathcal{R}}$ *is the closure of the set of achievable pairs* (R, R_0)*.*

In [23, Theorem II.1], the authors prove that in the setting of Figure 2.1, the strong coordination region is

$$\tilde{\mathcal{R}}_{\text{Cuff}} := \left\{ (R, R_0) \;\middle|\; \begin{array}{l} \exists P_{UWV} \in \mathcal{D}_{\text{Cuff}} \\ R \geq I(U;W) \\ R + R_0 \geq I(UV;W) \end{array} \right\} \tag{A.1}$$

where

$$\mathcal{D}_{\text{Cuff}} := \left\{ P_{UWV} \left| \begin{array}{l} (U,V) \text{ is generated according to } \bar{P}_U \bar{P}_{V|U} \\ U - W - V \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{V}| + 1 \end{array} \right. \right\}.$$

To prove the rate constraints in the converse proof, in [23] the authors introduce the following regions

$$\tilde{\mathcal{R}}_{\text{Cuff},\varepsilon} := \left\{ (R, R_0) \left| \begin{array}{l} \exists P_{UWV} \in \mathcal{D}_{\text{Cuff},\varepsilon} \\ R \geq I(U;W) \\ R + R_0 \geq I(UV;W) - f(\varepsilon) \end{array} \right. \right\},$$

$$\mathcal{D}_{\text{Cuff},\varepsilon} := \left\{ P_{UWV} \left| \begin{array}{l} \mathbb{V}(P_{UV}, \bar{P}_U \bar{P}_{V|U}) \leq \varepsilon \\ U - W - V \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{V}| + 1 \end{array} \right. \right\}.$$

The authors show in [23, Lemma VI.4 - *Epsilon Rate Region*] that, if the rate pair (R, R_0) is achievable for channel $\bar{P}_{V|U}$ and source $\bar{P}_U$, then

$$(R, R_0) \in \tilde{\mathcal{R}}_{\text{Cuff},\varepsilon} \quad \forall \varepsilon > 0.$$

Finally, to complete the proof the authors prove in [23, Lemma VI.5 - *Continuity of* $\tilde{\mathcal{R}}_{\mathit{Cuff},\varepsilon}$ *in zero*] that

$$\bigcap_{\varepsilon > 0} \tilde{\mathcal{R}}_{\text{Cuff},\varepsilon} = \tilde{\mathcal{R}}_{\text{Cuff}}.$$

This could seem trivial, since $\tilde{\mathcal{R}}_{\text{Cuff},0} = \tilde{\mathcal{R}}_{\text{Cuff}}$ and therefore $\bigcap_{\varepsilon > 0} \tilde{\mathcal{R}}_{\text{Cuff},\varepsilon} \supset \tilde{\mathcal{R}}_{\text{Cuff}}$, but the other direction is not obvious. In fact, although the inequalities in (A.1) are not strict, $\tilde{\mathcal{R}}_{\text{Cuff},\varepsilon}$ allows a relaxation in both the rate constraints and in the set of distributions. Then, to prove the other inclusion, we would have to prove the continuity of $\tilde{\mathcal{R}}_{\text{Cuff},\varepsilon}$ at zero, which requires a careful discussion. This complication can be avoided by choosing the slightly less precise Definition 2.4, in which the achievable region is defined as the closure of the set of achievable rates and distributions.

A.3 Proof of preliminary results

In Chapter 2 we have stated the key lemmas that we need to prove inner and outer bound of the coordination region. Although similar results may be found in the literature, we need slightly different results which are more convenient to our proofs.

Proof of Lemma 2.20 For $\beta > 0$, and $\gamma = (1+\beta)H(B|A)$ we define the following set

$$\mathcal{D}_{\gamma} = \left\{ (\mathbf{a}, \mathbf{b}) \in \mathcal{A}^n \times \mathcal{B}^n \mid P_{B^n|A^n}(\mathbf{b}|\mathbf{a}) < 2^{-n\gamma} \right\}. \tag{A.2}$$

Then,

$$\mathbb{E}[\mathbb{D}(P_{A^nK}\|P_{A^n}Q_K)] = \sum_{\mathbf{a}\in\mathcal{A}^n}\sum_{k\in[\![1,2^{nR}]\!]}\mathbb{E}\left[P_{A^nK}(\mathbf{a},k)\log\frac{P_{A^nK}(\mathbf{a},k)2^{nR}}{P_{A^n}(\mathbf{a})}\right]$$

$$=\sum_{\mathbf{a}\in\mathcal{A}^n}\sum_{\mathbf{b}\in\mathcal{B}^n}\sum_{k\in[\![1,2^{nR}]\!]}\mathbb{E}\left[P_{A^nB^nK}(\mathbf{a},\mathbf{b},k)\log\frac{\sum_{\mathbf{b}'\in\mathcal{B}^n}P_{A^nB^nK}(\mathbf{a},\mathbf{b}',k)2^{nR}}{P_{A^n}(\mathbf{a})}\right]$$

$$=\sum_{\mathbf{a}\in\mathcal{A}^n}\sum_{\mathbf{b}\in\mathcal{B}^n}\sum_{k\in[\![1,2^{nR}]\!]}\mathbb{E}\left[P_{A^nB^n}(\mathbf{a},\mathbf{b})\mathbb{1}\{\varphi_n(\mathbf{b})=k\}\log\frac{\sum_{\mathbf{b}'\in\mathcal{B}^n}P_{A^nB^n}(\mathbf{a},\mathbf{b}')\mathbb{1}\{\varphi_n(\mathbf{b}')=k\}2^{nR}}{P_{A^n}(\mathbf{a})}\right]$$

$$=\sum_{\mathbf{a}\in\mathcal{A}^n}\sum_{\mathbf{b}\in\mathcal{B}^n}\sum_{k\in[\![1,2^{nR}]\!]}P_{A^nB^n}(\mathbf{a},\mathbf{b})\frac{1}{2^{nR}}\mathbb{E}_{\sim\varphi_n(\mathbf{b})}\left[\log\frac{\sum_{\mathbf{b}'\in\mathcal{B}^n}P_{A^nB^n}(\mathbf{a},\mathbf{b}')\mathbb{1}\{\varphi_n(\mathbf{b}')=k\}2^{nR}}{P_{A^n}(\mathbf{a})}\right]$$

$$\overset{(a)}{\leq}\sum_{\mathbf{a}\in\mathcal{A}^n}\sum_{\mathbf{b}\in\mathcal{B}^n}\sum_{k\in[\![1,2^{nR}]\!]}P_{A^nB^n}(\mathbf{a},\mathbf{b})\frac{1}{2^{nR}}\log\mathbb{E}_{\sim\varphi_n(\mathbf{b})}\left[\frac{\sum_{\mathbf{b}'\in\mathcal{B}^n}P_{A^nB^n}(\mathbf{a},\mathbf{b}')\mathbb{1}\{\varphi_n(\mathbf{b}')=k\}2^{nR}}{P_{A^n}(\mathbf{a})}\right]$$

$$\overset{(b)}{=}\sum_{\mathbf{a}\in\mathcal{A}^n}\sum_{\mathbf{b}\in\mathcal{B}^n}\sum_{k\in[\![1,2^{nR}]\!]}P_{A^nB^n}(\mathbf{a},\mathbf{b})\frac{1}{2^{nR}}\log\frac{2^{nR}P_{A^nB^n}(\mathbf{a},\mathbf{b})\mathbb{1}\{\varphi_n(\mathbf{b})=k\}+\sum_{\substack{\mathbf{b}'\in\mathcal{B}^n\\ \mathbf{b}'\neq\mathbf{b}}}P_{A^nB^n}(\mathbf{a},\mathbf{b}')}{P_{A^n}(\mathbf{a})}$$

$$\leq\sum_{\mathbf{a}\in\mathcal{A}^n}\sum_{\mathbf{b}\in\mathcal{B}^n}P_{A^nB^n}(\mathbf{a},\mathbf{b})\log\left(2^{nR}P_{B^n|A^n}(\mathbf{b}|\mathbf{a})+1\right)$$

where $\mathbb{E}_{\sim\varphi_n(\mathbf{b})}$ denotes the average over all the random indices except the index $\varphi_n(\mathbf{b})$, (a) follows from Jensen's inequality, and (b) holds because

$$\mathbb{E}_{\sim\varphi_n(\mathbf{b})}\left[\sum_{\substack{\mathbf{b}'\in\mathcal{B}^n\\ \mathbf{b}'\neq\mathbf{b}}}P_{A^nB^n}(\mathbf{a},\mathbf{b}')\mathbb{1}\{\varphi_n(\mathbf{b}')=k\}\right]=\frac{1}{2^{nR}}\sum_{\substack{\mathbf{b}'\in\mathcal{B}^n\\ \mathbf{b}'\neq\mathbf{b}}}P_{A^nB^n}(\mathbf{a},\mathbf{b}').$$

If $(\mathbf{a},\mathbf{b})$ belongs to $\mathcal{D}_\gamma$, then

$$\log\left(2^{nR}P_{B^n|A^n}(\mathbf{b}|\mathbf{a})+1\right)<\log\left(\frac{2^{nR}}{2^{n\gamma}}+1\right)\leq\frac{2^{nR}}{2^{n\gamma}}=2^{n(R-(1+\beta)H(B|A))}\tag{A.3}$$

because $P_{B^n|A^n}$ is a discrete memoryless channel. On the other hand, if $(\mathbf{a},\mathbf{b})$ is not in $\mathcal{D}_\gamma$,

$$\log\left(2^{nR}P_{B^n|A^n}(\mathbf{b}|\mathbf{a})+1\right)\leq\log\left(2^{nR}+1\right)\tag{A.4}$$

because $P_{B^n|A^n}(\mathbf{b}|\mathbf{a})$ is smaller than 1. By combining (A.3) and (A.4), we have

$$\mathbb{E}[\mathbb{D}(P_{A^nK}\|P_{A^n}Q_K)]\leq\mathbb{P}\left\{(\mathbf{a},\mathbf{b})\notin\mathcal{D}_\gamma\right\}\log\left(2^{nR}+1\right)+\mathbb{P}\left\{(\mathbf{a},\mathbf{b})\in\mathcal{D}_\gamma\right\}2^{n(R-(1+\beta)H(B|A))}.$$

To estimate $\mathbb{P}\left\{(\mathbf{a},\mathbf{b})\notin\mathcal{D}_\gamma\right\}$, observe that it goes to zero because of the properties of typical sequences. Let $(\mathbf{a},\mathbf{b})\in\mathcal{T}_\varepsilon^{(n)}$, then by Property 1 of the typical sequences, $(\mathbf{a},\mathbf{b})$ belongs to

$\mathcal{D}_\gamma$. Hence, $T_\varepsilon^{(n)} \subseteq \mathcal{D}_\gamma$ and

$$\mathbb{P}\left\{(\mathbf{a},\mathbf{b}) \in \mathcal{D}_\gamma\right\} \geq \mathbb{P}\left\{(\mathbf{a},\mathbf{b}) \in \mathcal{T}_\varepsilon^{(n)}\right\} \to 1$$

because of Property 2 of the typical sequences. Then, if $H(B|A) - R > 0$,

$$\mathbb{E}[\mathbb{D}(P_{A^n K} \| P_{A^n} Q_K)] \to 0 \quad \text{exponentially.} \qquad \square$$

Proof of Lemma 2.17 We have

$$I(A_t; A_{\sim t}) = H(A_t) - H(A) + H(A) - H(A_t|A_{\sim t})$$

and we prove separately that

$$\begin{aligned} &H(A) - H(A_t|A_{\sim t}) \leq f(\varepsilon), \\ &H(A_t) - H(A) \leq f(\varepsilon). \end{aligned}$$

We consider the set $\mathcal{E} := \{\mathbf{a} \in \mathcal{A}^{n-1} \mid \mathbb{V}(P_{A_t|A_{\sim t}=\mathbf{a}}, \bar{P}_A) \leq \sqrt{\varepsilon}\}$ and recall the following result.

Lemma A.23 - ***[70, Lemma 3.2]*** *If* $\mathbb{V}(P_A P_{B|A}, P'_A P'_{B|A}) \leq \varepsilon$ *then*

$$\mathbb{P}\{A \in \mathcal{A} \mid \mathbb{V}(P_{B|A=a}, P'_{B|A=a}) \leq \sqrt{\varepsilon}\} \geq 1 - 2\sqrt{\varepsilon}.$$

Then, by Lemma A.23, $\mathbb{P}\{\mathcal{E}\} \geq 1 - 2\sqrt{\varepsilon}$, and we have

$$\begin{aligned} H(A) - H(A_t|A_{\sim t}) &= H(A) - \sum_{\mathbf{a}\in\mathcal{A}^{n-1}} P_{A_{\sim t}}(\mathbf{a}) H(A_t|A_{\sim t} = \mathbf{a}) \\ &= \sum_{\mathbf{a}\in\mathcal{A}^{n-1}} (P_{A_{\sim t}}(\mathbf{a}) H(A) - P_{A_{\sim t}}(\mathbf{a}) H(A_t|A_{\sim t} = \mathbf{a})) \qquad (A.5) \\ &= \sum_{\mathbf{a}\in\mathcal{E}} (P_{A_{\sim t}}(\mathbf{a}) H(A) - P_{A_{\sim t}}(\mathbf{a}) H(A_t|A_{\sim t} = \mathbf{a})) + \sum_{\mathbf{a}\in\mathcal{E}^c} (P_{A_{\sim t}}(\mathbf{a}) H(A) - P_{A_{\sim t}}(\mathbf{a}) H(A_t|A_{\sim t} = \mathbf{a})) \\ &\overset{(a)}{\leq} \sum_{\mathbf{a}\in\mathcal{E}} P_{A_{\sim t}}(\mathbf{a}) \delta + \mathbb{P}\{\mathcal{E}^c\}(H(A_t) + H(A)) \leq \delta + 2\sqrt{\varepsilon}(2H(A) + \delta) \end{aligned}$$

where (a) comes from the fact that, by Lemma A.16, for $\mathbf{a} \in \mathcal{E}$

$$|H(A_t|A_{\sim t} = \mathbf{a}) - H(A)| \leq \varepsilon \log \frac{|\mathcal{A}|}{\varepsilon} := \delta.$$

Lemma A.16 also implies that

$$|H(A_t) - H(A)| \leq \delta. \qquad (A.6)$$

Hence by (A.5) and (A.6), we have $I(A_t; A_{\sim t}) \leq 2\sqrt{\varepsilon}(2H(A) + \delta) + 2\delta$. $\square$

Proof of Lemma 2.18 The proof of (2.4) comes directly from Lemma 2.17:

$$\sum_{t=1}^{n} I(A_t; A^{t-1}B_{\sim t}|B_t) \le \sum_{t=1}^{n} I(A_t; A_{\sim t}B_{\sim t}|B_t) \le \sum_{t=1}^{n} I(A_tB_t; A_{\sim t}B_{\sim t}) \le nf(\varepsilon). \quad \text{(A.7)}$$

To prove (2.5), we have

$$\begin{aligned}
H(C|B^n) \ge I(A^n; C|B^n) &= \sum_{t=1}^{n} I(A_t; C|A^{t-1}B_{\sim t}B_t) \\
&= \sum_{t=1}^{n} I(A_t; CA^{t-1}B_{\sim t}|B_t) - \sum_{t=1}^{n} I(A_t; A^{t-1}B_{\sim t}|B_t) \\
&\ge \sum_{t=1}^{n} I(A_t; CB_{\sim t}|B_t) - \sum_{t=1}^{n} I(A_t; A^{t-1}B_{\sim t}|B_t) \\
&\overset{(a)}{\ge} \sum_{t=1}^{n} I(A_t; CB_{\sim t}|B_t) - nf(\varepsilon) = nI(A_T; CB_{\sim T}|B_TT) - nf(\varepsilon) \\
&= nI(A_T; CB_{\sim T}T|B_T) - nI(A_T; T|B_T) - nf(\varepsilon) \\
&\ge nI(A_T; CB_{\sim T}T|B_T) - nI(A_TB_T; T) - nf(\varepsilon)
\end{aligned}$$

where (a) comes from (A.7). □

B | Proofs of inner and outer bounds for strong coordination

In this chapter, we present the achievability and converse proofs which we omitted in Chapter 3, Chapter 4 and Chapter 6.

B.1 Achievability proofs

Here, we detail the proofs of the inner bound for Theorem 3.10 and Theorem 6.3.

B.1.1 Proof of Lemma 3.6

We denote the event that $\hat{A}^n$ is different from A^n with $\mathcal{E} := \{A^n \neq \hat{A}^n\}$, where $\mathbb{P}\{\mathcal{E}^c\}$ tends to 1. We can write the joint distribution $P_{B^n A^n \hat{A}^n}$ as

$$P_{B^n A^n \hat{A}^n} = \mathbb{P}\left\{\mathcal{E}^c\right\} P_{B^n A^n \hat{A}^n|\mathcal{E}^c} + \mathbb{P}\left\{\mathcal{E}\right\} P_{B^n A^n \hat{A}^n|\mathcal{E}}.$$

Hence, we have

$$\mathbb{V}(P_{B^n A^n \hat{A}^n}, P_{B^n A^n}\mathbb{1}_{\hat{A}^n|A^n}) \leq \mathbb{P}\left\{\mathcal{E}\right\} \left\|P_{B^n A^n \hat{A}^n|\mathcal{E}}\right\|_{L^1} + \left\|\mathbb{P}\left\{\mathcal{E}^c\right\} P_{B^n A^n \hat{A}^n|\mathcal{E}^c} - P_{B^n A^n}\mathbb{1}_{\hat{A}^n|A^n}\right\|_{L^1}$$

where the first term is equal to $(1 - \mathbb{P}\left\{\mathcal{E}^c\right\}) P_{B^n A^n}\mathbb{1}_{\hat{A}^n|A^n}$ and goes to 0 since $\mathbb{P}\left\{\mathcal{E}^c\right\}$ tends to 1 and the second term goes to 0 since $\mathbb{P}\left\{\mathcal{E}\right\}$ does. □

B.1.2 Theorem 3.10

In this section, we generalize the achievability proof of Theorem 3.3 to the broader scenario of channel with state and side information available at the decoder.

Random binning scheme Assume that the sequences U^n, S^n, Z^n, X^n, W^n, Y^n and V^n are jointly i.i.d. with distribution

$$\bar{P}_{U^n S^n Z^n} \bar{P}_{W^n|U^n} \bar{P}_{X^n|W^n U^n} \bar{P}_{Y^n|X^n S^n} \bar{P}_{V^n|W^n Y^n Z^n}.$$

We consider two uniform random binnings for W^n:

- first binning $C = \varphi_1(W^n)$, where $\varphi_1 : \mathcal{W}^n \to [\![1, 2^{nR_0}]\!]$ maps each sequence of $\mathcal{W}^n$ uniformly and independently to the set $[\![1, 2^{nR_0}]\!]$;
- second binning $F = \varphi_2(W^n)$, where $\varphi_2 : \mathcal{W}^n \to [\![1, 2^{nR}]\!]$ maps each sequence of $\mathcal{W}^n$ uniformly and independently to the set $[\![1, 2^{nR}]\!]$.

Note that if $R+R_0 > H(W|YZ)$, by Lemma 2.19, it is possible to recover W^n from Y^n, Z^n and (C, F) with high probability using a Slepian-Wolf decoder via the conditional distribution $P^{\text{SW}}_{\hat{W}^n|CFY^nZ^n}$. This defines a joint distribution:

$$P^{\text{RB}} := \bar{P}_{U^nS^nZ^n}\bar{P}_{W^n|U^n}\bar{P}_{X^n|W^nU^n}\bar{P}_{C|W^n}\bar{P}_{F|W^n}\bar{P}_{Y^n|X^nS^n}\bar{P}_{V^n|W^nY^nZ^n}P^{\text{SW}}_{\hat{W}^n|CFY^nZ^n}.$$

In particular, $P^{\text{RB}}_{W^n|CFU^n}$ is well defined.

Random coding scheme Similarly to Section 3.1.1, we assume that in the setting of Figure 3.3, encoder and decoder have access not only to common randomness C but also to extra randomness F, where C is generated uniformly at random in $[\![1, 2^{nR_0}]\!]$ with distribution Q_C and F is generated uniformly at random in $[\![1, 2^{nR}]\!]$ with distribution Q_F independently of C. Then, the encoder generates W^n according to $\bar{P}_{W^n|CFU^n}$ defined above and X^n according to $\bar{P}_{X^n|U^nW^n}$. The encoder sends X^n through the channel. The decoder obtains (Y^n, Z^n) and (C, F) and reconstructs W^n via the conditional distribution $P^{\text{SW}}_{\hat{W}^n|CFY^nZ^n}$. The decoder then generates V^n letter by letter according to the distribution

$$P^{\text{RC}}_{V^n|\hat{W}^nY^nZ^n}(\hat{\mathbf{u}}|\hat{\mathbf{w}}, \mathbf{y}, \mathbf{z}) = \bar{P}_{V^n|W^nY^nZ^n}(\hat{\mathbf{u}}|\hat{\mathbf{w}}, \mathbf{y}, \mathbf{z}), \tag{B.1}$$

where $\hat{\mathbf{w}}$ is the output of the Slepian-Wolf decoder. This defines a joint distribution:

$$P^{\text{RC}} := Q_CQ_FP^{\text{RC}}_{U^nS^nZ^n}P^{\text{RB}}_{W^n|CFU^n}\bar{P}_{X^n|W^nU^n}\bar{P}_{Y^n|X^nS^n}P^{\text{SW}}_{\hat{W}^n|CFY^nZ^n}P^{\text{RC}}_{V^n|\hat{W}^nY^nZ^n}.$$

Strong coordination of $(U^n, S^n, Z^n, X^n, W^n, Y^n, V^n)$ We want to show that the distribution P^{RB} is achievable for strong coordination:

$$\lim_{n\to\infty} \mathbb{V}(P^{\text{RB}}_{U^nS^nZ^nX^nW^n\hat{W}^nY^nV^n}, P^{\text{RC}}_{U^nS^nZ^nX^nW^n\hat{W}^nY^nV^n}) = 0. \tag{B.2}$$

We prove that the random coding scheme possesses all the properties of the initial source coding scheme stated in Section B.1.2. Note that

$$\begin{aligned}
\mathbb{D}(P^{\text{RB}}\|P^{\text{RC}}) &= \mathbb{D}(\bar{P}_{U^nS^nZ^n}\bar{P}_{W^n|U^n}\bar{P}_{X^n|W^nU^n}\bar{P}_{C|W^n}\bar{P}_{F|W^n}\bar{P}_{Y^n|X^nS^n}P^{\text{SW}}_{\hat{W}^n|CFY^nZ^n} \\
&\qquad \|Q_CQ_FP^{\text{RC}}_{U^nS^nZ^n}P^{\text{RB}}_{W^n|CFU^n}\bar{P}_{X^n|W^nU^n}\bar{P}_{Y^n|X^nS^n}P^{\text{SW}}_{\hat{W}^n|CFY^nZ^n}) \\
&\overset{(a)}{=}\mathbb{D}(\bar{P}_{U^nS^nZ^n}\bar{P}_{W^n|U^n}\bar{P}_{C|W^n}\bar{P}_{F|W^n}\|Q_CQ_FP^{\text{RC}}_{U^nS^nZ^n}\bar{P}_{W^n|CFU^n}) \\
&\overset{(b)}{=}\mathbb{D}(P^{\text{RB}}_{U^nS^nZ^nCF}\|P^{\text{RC}}_{U^nS^nZ^n}Q_CQ_F)
\end{aligned} \tag{B.3}$$

where (a) comes from Lemma 2.14. Note that (b) follows from Lemma 2.14 as well, since W^n is generated according to $P^{\text{RB}}_{W^n|CFU^n}$ and because of the Markov chain $W - U - ZS$, W^n is conditionally independent of (Z^n, S^n) given U^n. Then if $R_0 + R < H(W|USZ) = H(W|U)$, we apply Lemma 2.20 to $B^n = W^n$, $K = (C, F)$, $A^n = U^n S^n Z^n$ and claim that there exists a fixed pair of binnings (φ'_1, φ'_2), such that,

$$\mathbb{D}(P^{\text{RB},(\varphi'_1,\varphi'_2)}_{U^nS^nZ^nCF} \| P^{\text{RC}}_{U^nS^nZ^n} Q_C Q_F) = \delta(n), \tag{B.4}$$

which by (3.8) and Lemma A.15 imply

$$\mathbb{D}(P^{\text{RB}} \| P^{\text{RC}}) = \delta(n) \quad \text{and} \quad \mathbb{V}(P^{\text{RB}}, P^{\text{RC}}) = \delta(n). \tag{B.5}$$

Now, we want to prove that we have strong coordination for V^n as well. As in Section 3.1.1, the main difficulty is that in the second coding scheme V^n is generated using the output of the Slepian-Wolf decoder $\hat{W}^n$ and not W^n as in the first scheme. Because of Lemma 2.19, the inequality $\tilde{R} + R_0 > H(W|YZ)$ implies that $\hat{W}^n$ is equal to W^n with high probability and we will use this fact to show that the distributions are close in total variational distance. Then, similarly to Section 3.1.1 we apply Lemma 3.6 and, since $\mathbb{P}\{\hat{W}^n \neq W^n\}$ goes to zero exponentially by Lemma 2.19, we find that

$$\mathbb{V}(P^{\text{RB}}_{U^nS^nZ^nW^nX^nY^nCF}, P^{\text{RB}}_{U^nS^nZ^n\hat{W}^nX^nY^nCF}) = \delta(n),$$

and similarly

$$\mathbb{V}(P^{\text{RC}}_{U^nS^nZ^nW^nX^nY^nCF}, P^{\text{RC}}_{U^nS^nZ^n\hat{W}^nX^nY^nCF}) = \delta(n),$$

that imply respectively

$$\mathbb{V}(P^{\text{RB}}_{U^nS^nZ^nW^n\hat{W}^nX^nY^nCF}, P^{\text{RB}}_{U^nS^nZ^nW^nX^nY^nCF} \mathbb{1}_{\hat{W}^n|W^n}) = \delta(n), \tag{B.6}$$

$$\mathbb{V}(P^{\text{RC}}_{U^nS^nZ^nW^n\hat{W}^nX^nY^nCF}, P^{\text{RC}}_{U^nS^nZ^nW^nX^nY^nCF} \mathbb{1}_{\hat{W}^n|W^n}) = \delta(n). \tag{B.7}$$

By the triangle inequality,

$$\begin{aligned}
\mathbb{V}(P^{\text{RB}}, P^{\text{RC}}) &= \mathbb{V}(P^{\text{RB}}_{U^nS^nZ^nW^n\hat{W}^nX^nY^nCF} \bar{P}_{V^n|W^nY^nZ^n}, P^{\text{RC}}_{U^nS^nZ^nW^n\hat{W}^nX^nY^nCF} P^{\text{RC}}_{V^n|\hat{W}^nY^nZ^n}) \\
&\leq \mathbb{V}(P^{\text{RB}}_{U^nS^nZ^nW^n\hat{W}^nX^nY^nCF} \bar{P}_{V^n|W^nY^nZ^n}, P^{\text{RB}}_{U^nS^nZ^nW^nX^nY^nCF} \mathbb{1}_{\hat{W}^n|W^n} \bar{P}_{V^n|W^nY^nZ^n}) \qquad \text{(B.8)} \\
&\quad + \mathbb{V}(P^{\text{RB}}_{U^nS^nZ^nW^nX^nY^nCF} \mathbb{1}_{\hat{W}^n|W^n} \bar{P}_{V^n|W^nY^nZ^n}, P^{\text{RC}}_{U^nS^nZ^nW^nX^nY^nCF} \mathbb{1}_{\hat{W}^n|W^n} P^{\text{RC}}_{V^n|\hat{W}^nY^nZ^n}) \\
&\quad + \mathbb{V}(P^{\text{RC}}_{U^nS^nZ^nW^nX^nY^nCF} \mathbb{1}_{\hat{W}^n|W^n} P^{\text{RC}}_{V^n|W^nY^nZ^n}, P^{\text{RC}}_{U^nS^nZ^nW^n\hat{W}^nX^nY^nCF} P^{\text{RC}}_{V^n|\hat{W}^nY^nZ^n})
\end{aligned}$$

where first and the third term go to zero exponentially by applying Lemma 2.13 to (B.6) and (B.7) respectively. Then, since by definition of $P^{\text{RC}}_{V^n|\hat{W}^nY^nZ^n}$ (B.1),

$$\mathbb{1}_{\hat{W}^n|W^n} \bar{P}_{V^n|W^nY^nZ^n} = \mathbb{1}_{\hat{W}^n|W^n} P^{\text{RC}}_{V^n|\hat{W}^nY^nZ^n},$$

by using Lemma 2.13 again the second term is equal to

$$\mathbb{V}(P^{\text{RB}}_{U^nS^nZ^nW^nX^nY^nCF}, P^{\text{RC}}_{U^nS^nZ^nW^nX^nY^nCF})$$

and goes to zero by (B.5) and Lemma 2.12. Hence, we have

$$\mathbb{V}(P^{\text{RB}}, P^{\text{RC}}) = \delta(n). \tag{B.9}$$

$$\mathbb{V}(P^{\text{RB}}_{U^nS^nZ^nX^nW^n\hat{W}^nY^nV^n}, P^{\text{RC}}_{U^nS^nZ^nX^nW^n\hat{W}^nY^nV^n}) = \delta(n).$$

Remove the extra randomness F Even though the extra common randomness F is required to coordinate $(U^n, S^n, Z^n, X^n, Y^n, V^n, W^n)$ we will show that we do not need it in order to coordinate only $(U^n, S^n, Z^n, X^n, Y^n, V^n)$. Observe that by Lemma 2.12, equation (B.9) implies that

$$\mathbb{V}(\bar{P}_{U^nS^nZ^nX^nY^nV^nF}, P_{U^nS^nZ^nX^nY^nV^nF}) = \delta(n). \tag{B.10}$$

As in [70], we would like to reduce the amount of common randomness by having the two nodes agree on an instance $F = f$. To do so, we apply Lemma 2.20 again to $B^n = W^n$, $K = F$, and $A^n = U^nS^nZ^nX^nY^nV^n$. If $R < H(W|SUZXYV)$, by Lemma 2.20 there exists a fixed binning φ''_2 such that

$$\mathbb{V}(P^{\text{RB},\varphi''_2}_{U^nS^nZ^nX^nY^nV^nF}, Q_F P^{\text{RB},\varphi''_2}_{U^nS^nZ^nX^nY^nV^n}) = \delta(n). \tag{B.11}$$

Because of (B.10), (B.11) implies

$$\mathbb{V}(P^{\text{RC}}_{U^nS^nZ^nX^nY^nV^nF}, Q_F P^{\text{RB}}_{U^nS^nZ^nX^nY^nV^n}) = \delta(n). \tag{B.12}$$

By Lemma 2.15, there exists an instance $f \in [\![1, 2^{nR}]\!]$ such that

$$\mathbb{V}(P^{\text{RB}}_{U^nS^nZ^nX^nY^nV^n|F=f}, P^{\text{RC}}_{U^nS^nZ^nX^nY^nV^n|F=f}) = \delta(n).$$

Then, by fixing $F = f$ and using common randomness C, we have coordination for $(U^n, S^n, Z^n, X^n, Y^n, V^n)$.

Rate constraints We have imposed the following rate constraints:

$$\begin{aligned} H(W|YZ) < R + R_0 < H(W|U), \\ R < H(W|USZXYV). \end{aligned}$$

Therefore we obtain:

$$\begin{aligned} &R_0 > H(W|YZ) - H(W|USZXYV) = I(W;USXV|YZ), \\ &I(W;U) < I(W;YZ). \end{aligned}$$

□

B.1.3 Theorem 6.3

Here, we prove that the two schemes defined in Section 6.1.1 have the same statistics.

Coordination of $(U^n, X^n, W^n, Y^n, V^n)_{(i)}$ We want to show that the distribution $\hat{P}^{\text{RC}}_{(i)}$ is achievable for strong coordination, i.e.,

$$\lim_{n\to\infty} \mathbb{V}\left(P^{\text{RB}}, \hat{P}^{\text{RC}}_{(i)}\right) = 0. \tag{B.13}$$

Observe that

- By Lemma 2.13 the total variational distance remains the same without $\bar{P}_{Y^n|X^n}$ and $P_{V^n|W^nY^n}$ in both P^{RB} and $\hat{P}^{\text{RC}}_{(i)}$;
- the random binning distribution becomes

$$P^{\text{RB}}_{M_1M_2M_3FX^nU^n}P^{\text{RB}}_{M_4|M_3X^nU^n}P^{\text{RB}}_{W^n|M_3M_4FX^n}$$

  and $P^{\text{RB}}_{X^n|M_1M_2M_3F}P^{\text{RB}}_{M_4|M_3X^nU^n}P^{\text{RB}}_{W^n|M_3M_4FX^n}$ can be removed in both P^{RB} and $\hat{P}^{\text{RC}}_{(i)}$ by Lemma 2.13;
- now, (3.7) is satisfied if

$$\mathbb{V}\left(P^{\text{RB}}_{M_1M_2M_3FU^n}, \hat{P}^{\text{RC}}_{(M_1M_2M_3FU^n)_{(i)}}\right) = \mathbb{V}\left(P^{\text{RB}}_{M_1M_2M_3FU^n}, Q_{M_3}Q_FQ_{M_1}Q_{M_2}P^{\text{RB}}_{U^n}\right)$$

  vanishes. By Lemma 2.20, this would be true if

$$R_1 + R_2 + R_3 + R < H(WX|U). \tag{B.14}$$

  Since we have imposed the rate condition $R_3 + R < H(W|XU)$ and $R_1 + R_2 < H(X)$ and $H(W|XU) + H(X) = H(WX|U)$ because X and U are independent, (B.14) holds and there exists a binning of (W, X) such that

$$\mathbb{V}(P^{\text{RB}}_{M_1M_2M_3FU^n}, Q_{M_3}Q_FQ_{M_1}Q_{M_2}P^{\text{RB}}_{U^n}) \leq \delta(n).$$

  Then we conclude that (B.13) holds.

Coordination of $(U^n, X^n, Y^n, V^n)_{(i)}$ by removing the extra randomness F Even though the extra common randomness F is required to coordinate $(U^n, X^n, Y^n, V^n, W^n)$ we will show that we do not need it in order to coordinate only (U^n, X^n, Y^n, V^n). As in [70], we would like to reduce the amount of common randomness by having the two nodes agree on an instance $F = f$. To do so, we apply Lemma 2.20 again to $B^n = W^n$, $K = F$, and $A^n = U^nX^nY^nV^n$. If $R < H(W|UXYV)$, there exists a fixed binning such that

$$\mathbb{V}(P^{\text{RB}}_{U^nX^nY^nV^nF}, Q_FP^{\text{RB}}_{U^nX^nY^nV^n}) = \delta(n). \tag{B.15}$$

which implies

$$\mathbb{V}(\hat{P}^{\text{RC}}_{(U^nX^nY^nV^nF)_{(i)}}, Q_F P^{\text{RB}}_{U^nX^nY^nV^n}) = \delta(n). \tag{B.16}$$

By Lemma 2.15, there exists an instance $f \in [\![1, 2^{nR}]\!]$ such that

$$\mathbb{V}(P^{\text{RB}}_{U^nX^nY^nV^n|F=f}, \hat{P}^{\text{RC}}_{(U^nX^nY^nV^n)_{(i)}|F_{(i)}=f}) = \delta(n). \tag{B.17}$$

Then, by fixing $F = f$ and using common randomness $C = (M_{1,(i)}, M_{3,(i)}, K_i)$, we have coordination for (U^n, X^n, Y^n, V^n).

Rate of common randomness We have used common randomness to generate M_1,M_3 and the key of the one-time pad, which has the same size of M_4. Then, upon denoting by R_0 the total rate of common randomness, $R_0 := R_1 + R_3 + R_4$ and

$$\begin{aligned}
&R_0 + R > H(X|Y) + H(W|X)\\
&R < H(W|UXYV)
\end{aligned}$$

which implies

$$R_0 > H(X|Y) + H(W|X) - H(W|UXYV). \tag{B.18}$$

Observe that

$$\begin{aligned}
H(WX|Y) &= H(WX) - I(WX;Y)\\
&= H(X) + H(W|X) - I(X;Y)\\
&= H(X|Y) + H(W|X)
\end{aligned}$$

because the Markov chain $W - X - Y$ implies $I(W;Y|X) = 0$ and therefore (B.18) becomes

$$\begin{aligned}
R_0 &> H(WX|Y) - H(W|UXYV)\\
&= H(W|Y) + H(X|WY) - H(W|UXYV)\\
&= I(W;UXV|Y) + H(X|WY).
\end{aligned} \tag{B.19}$$

Coordination of all blocks First, note that two consecutive blocks L_{i-1} and L_i are dependent only through $M_{4,(i-1)}$. In fact, $M_{4,(i-1)}$ is created at time i using $U^n_{(i-1)}$ and $X^n_{(i-1)}$ and it is used to generate $M_{2,(i)}$, which in turn is used at the encoder to generate $X^n_{(i)}$. Hence, since $Y^n_{(i)}$ is the output of the channel and $V^n_{(i)}$ is generated using $Y^n_{(i)}$ and the auxiliary random variable, generated through an estimate of $\hat{M}_{4,(i)}$, uniform common randomness and $X^n_{(i)}$, we can conclude that L_{i-1} and L_i are dependent only through $M'_{2,(i)}$ and therefore $M_{4,(i-1)}$. However, to generate $M_{2,(i)}$, the encoder applies a one-time pad on $M_{4,(i-1)}$ as shown in (6.7), making $M_{4,(i-1)}$ and $M_{2,(i)}$ independent of each other and ensuring the independence of two consecutive blocks.

To conclude the proof we need the following results.

Lemma B.1 *We have*

$$\mathbb{V}\left(P_{L_{1:k-1}}, \prod_{i=1}^{k-1} P_{L_i}\right) \leq \delta(n).$$

Lemma B.2 *We have*

$$\mathbb{V}\left(P_{L_{1:k-1}}, \bar{P}_{UXYV}^{\otimes n(k-1)}\right) \leq \delta(n).$$

We omit the proofs because they are very similar to the proofs of Lemma 5.19 and Lemma 5.20 respectively. □

B.2 Converse proofs

In this section, we prove some technical results which we need to complete the proofs of the outer bounds for the strong coordination region.

B.2.1 Proof of Lemma 3.13.

We want to prove that $I(Y_tZ_t; C, X_{\sim t}U_{\sim t}S_{\sim t}Y_{\sim t}Z_{\sim t}|X_tU_tS_t) = 0$. We have

$$\begin{aligned}
&I(Y_tZ_t; CX_{\sim t}S_{\sim t}U_{\sim t}Y_{\sim t}Z_{\sim t}|X_tS_tU_t)\\
&\qquad = I(Z_t; CX_{\sim t}S_{\sim t}U_{\sim t}Y_{\sim t}Z_{\sim t}|X_tS_tU_t) + I(Y_t; CX_{\sim t}S_{\sim t}U_{\sim t}Y_{\sim t}Z_{\sim t}|X_tS_tU_tZ_t)\\
&\qquad = I(Z_t; CX_tX_{\sim t}S_{\sim t}U_{\sim t}Y_{\sim t}Z_{\sim t}|S_tU_t) - I(Z_t; X_t|S_tU_t)\\
&\qquad\quad + I(Y_t; CU_tZ_tX_{\sim t}S_{\sim t}U_{\sim t}Y_{\sim t}Z_{\sim t}|X_tS_t) - I(Y_t; U_tZ_t|X_tS_t)\\
&\qquad \leq I(Z_t; CX^nS_{\sim t}U_{\sim t}Y_{\sim t}Z_{\sim t}|S_tU_t) + I(Y_t; CU^nZ^nX_{\sim t}S_{\sim t}Y_{\sim t}|X_tS_t)\\
&\qquad \leq I(Z_t; CX^nY^nS_{\sim t}U_{\sim t}Z_{\sim t}|S_tU_t) + I(Y_t; CU^nZ^nX_{\sim t}S_{\sim t}Y_{\sim t}|X_tS_t)
\end{aligned}$$

where both $I(Z_t; CX^nY^nS_{\sim t}U_{\sim t}Z_{\sim t}|S_tU_t)$ and $I(Y_t; CU^nZ^nX_{\sim t}S_{\sim t}Y_{\sim t}|X_tS_t)$ are equal to zero because by (3.27) the following Markov chains hold:

$$\begin{aligned}
&Z_t - (U_t, S_t) - (C, X^n, Y^n, U_{\sim t}, S_{\sim t}, Z_{\sim t}),\\
&Y_t - (X_t, S_t) - (C, Z^n, U^n, X_{\sim t}, S_{\sim t}, Y_{\sim t}).
\end{aligned}$$

□

B.2.2 Proof of (4.8)

Define the event of error E as follows:

$$E := \begin{cases} 0 & \text{if} \quad U^n = V^n \\ 1 & \text{if} \quad U^n \neq V^n \end{cases}.$$

We note $p_e := \mathbb{P}\{U^n \neq V^n\}$ and recall that by hypothesis the distribution $P_{U^nS^nZ^nX^nY^nV^n}$ is ε-close in total variational distance to the i.i.d. distribution $\bar{P}_{USZXYV}^{\otimes n}$ where the decoder is lossless. Then $\mathbb{V}(P_{U^nV^n}, P_{U^n}\mathbb{1}_{V^n|U^n}) < \varepsilon$ and therefore $p_e < 2\varepsilon$ since

$$\begin{aligned}\mathbb{V}(P_{U^nV^n}, P_{U^n}\mathbb{1}_{V^n|U^n}) &= \frac{1}{2}\sum_{\substack{\mathbf{u}\in\mathcal{U}^n\\ \mathbf{v}\in\mathcal{V}^n}} |P_{U^n}(\mathbf{u})\mathbb{1}_{V^n|U^n}(\mathbf{v}|\mathbf{u}) - P_{U^nV^n}(\mathbf{u},\mathbf{v})| \\ &= \frac{1}{2}\sum_{\mathbf{u}=\mathbf{v}} P_{U^n}(\mathbf{u})\mathbb{1}_{V^n|U^n}(\mathbf{v}|\mathbf{u}) - P_{U^nV^n}(\mathbf{u},\mathbf{u})| + \frac{1}{2}\sum_{\mathbf{u}\neq\mathbf{v}} P_{U^nV^n}(\mathbf{u},\mathbf{v}) \\ &= \frac{\mathbb{P}\{U^n \neq V^n\}}{2}\end{aligned}$$

By Fano's Inequality [27], we have

$$H(U^n|C, Y^n, Z^n) \leq h(p_e) + p_e \log(|\mathcal{U}^n| - 1) \tag{B.20}$$

where h is the binary entropy function. Since $p_e < 2\varepsilon$ vanishes, $h(p_e) < f(\varepsilon)$ and the right-hand side of (B.20) goes to zero. Hence, we have that $H(U^n|C, Y^n, Z^n) \leq (n+1)f(\varepsilon)$, where $f(\varepsilon)$ denotes a function which tends to zero as ε does. □

B.3 Proof of Proposition 4.8

In this section, we characterize the region $\mathcal{R}_{UV\otimes X}$ of Proposition 4.8, defined in (4.18).

B.3.1 Achievability

We show that $\mathcal{R}_{UV\otimes X}$ is contained in the region $\mathcal{R}_{\text{PC}} \cap \mathcal{R}_{\text{SEP}}$ and thus it is achievable.

We consider the subset of $\mathcal{R}_{\text{SEP}}$ when $\bar{P}_{Y|X}(\mathbf{y}|\mathbf{x}) = \mathbb{1}_{Y|X}(\mathbf{y}|\mathbf{x})$ as the union of all $\mathcal{R}_{\text{SEP}}(W)$ with $W = (W_1, W_2)$ that satisfies

$$\begin{aligned}&\bar{P}_{UW_1W_2XV} = \bar{P}_U\bar{P}_{W_2|U}\bar{P}_{V|W_2}\bar{P}_X\bar{P}_{W_1|X},\\ &I(W_1;X) \geq I(W_2;U),\\ &R_0 \geq I(W_2;UV).\end{aligned} \tag{B.21}$$

Similarly, $\mathcal{R}_{\text{PC}}$ is the union of all $\mathcal{R}_{\text{PC}}(W)$ with W that satisfies

$$\begin{aligned}&\bar{P}_{UWXV} = \bar{P}_U\bar{P}_{W|U}\bar{P}_{X|UW}\bar{P}_{V|WX},\\ &H(X) \geq I(WX;U),\\ &R_0 \geq I(W;UV|X).\end{aligned} \tag{B.22}$$

If we choose $W = (W_1, W_2)$ and we add the hypothesis that (W_2, U, V) is independent of (W_1, X), (B.22) becomes

$$\begin{aligned}&\bar{P}_{UW_1W_2XV} = \bar{P}_U\bar{P}_{W_2|U}\bar{P}_{V|W_2}\bar{P}_{W_1}\bar{P}_{X|W_1},\\ &H(X) \geq I(W_1W_2X;U) = I(W_2;U),\\ &R_0 \geq I(W_1W_2;UV|X) = I(W_2;UV).\end{aligned} \tag{B.23}$$

Note that if we identify $W_1 = X$, we have

$$H(X) = I(W_1; X) \quad \text{and} \quad \bar{P}_{W_1}\bar{P}_{X|W_1} = \bar{P}_X\bar{P}_{W_1|X} = \bar{P}_X\mathbb{1}_{W_1|X}.$$

Then, there exists a subset of $\mathcal{R}_{\text{SEP}}$ and $\mathcal{R}_{\text{PC}}$ defined as the union over all W_2 of the distributions $\bar{P}_{UVX}$ satisfying

$$\begin{aligned}
&\bar{P}_{UW_2XV} = \bar{P}_U\bar{P}_{W_2|U}\bar{P}_{V|W_2}\bar{P}_X,\\
&H(X) \geq I(W_2; U),\\
&R_0 \geq I(W_2; UV).
\end{aligned} \tag{B.24}$$

Finally, observe that, by definition of the region (4.18), $\mathcal{R}_{UV\otimes X}$ is the union over all the possible choices for W_2 that satisfy (B.24) and therefore $\mathcal{R}_{UV\otimes X} \subseteq \mathcal{R}_{\text{PC}} \cap \mathcal{R}_{\text{SEP}}$. □

B.3.2 Converse

Consider a code (f_n, g_n) that induces a distribution $P_{U^nX^nV^n}$ that is ε-close in total variational distance to the i.i.d. distribution $\bar{P}_{UV}^{\otimes n}\bar{P}_X^{\otimes n}$. Let T be the random variable defined in Section 3.1.2.

Then, we have

$$\begin{aligned}
nR_0 = H(C) &\overset{(a)}{\geq} I(U^nV^n; C|X^n) = I(U^nV^n; CX^n) - I(U^nV^n; X^n)\\
&\overset{(b)}{\geq} I(U^nV^n; CX^n) - nf(\varepsilon) = \sum_{t=1}^{n} I(U_tV_t; CX^n|U^{t-1}V^{t-1}) - nf(\varepsilon)\\
&= \sum_{t=1}^{n} I(U_tV_t; CX^nU^{t-1}V^{t-1}) - \sum_{t=1}^{n} I(U_tV_t; U^{t-1}V^{t-1}) - nf(\varepsilon)\\
&\overset{(c)}{\geq} \sum_{t=1}^{n} I(U_tV_t; CX^nU^{t-1}V^{t-1}) - 2nf(\varepsilon) \geq \sum_{t=1}^{n} I(U_tV_t; CX^nU^{t-1}) - 2nf(\varepsilon)\\
&= nI(U_TV_T; CX^nU^{T-1}|T) - 2nf(\varepsilon) = nI(U_TV_T; CX^nU^{T-1}T) - nI(U_TV_T; T) - 2nf(\varepsilon)\\
&\overset{(d)}{\geq} nI(U_TV_T; CX^nU^{T-1}T) - 3nf(\varepsilon)
\end{aligned}$$

where (a) follows from basic properties of entropy and mutual information and (b) from the upper bound on the mutual information in Lemma A.15 since we assume $\mathbb{V}(P_{U^nV^nX^n}, \bar{P}_{UV}^{\otimes n}\bar{P}_X^{\otimes n}) \leq \varepsilon$ and $|\mathcal{U} \times \mathcal{V}| \geq 4$. Finally, since the distributions are close to i.i.d. by hypothesis, (c) and (d) come from Lemma 2.17 and [23, Lemma VI.3] respectively.

For the second part of the converse, observe that

$$\begin{aligned}
0 &= H(X^n) - I(X^n; U^nC) - H(X^n|U^nC)\\
&\leq H(X^n) - I(X^n; U^n|C) - H(X^n|U^nC)\\
&\leq \sum_{t=1}^{n} H(X_t) - \sum_{t=1}^{n} I(X^n; U_t|U^{t-1}C) - H(X^n|U^nC)
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{t=1}^{n} H(X_t) - \sum_{t=1}^{n} I(X^n; U_t | U^{t-1} C) \\
&= \sum_{t=1}^{n} H(X_t) - \sum_{t=1}^{n} I(X^n U^{t-1} C; U_t) + \sum_{t=1}^{n} I(U^{t-1} C; U_t) \\
&\overset{(e)}{=} \sum_{t=1}^{n} H(X_t) - \sum_{t=1}^{n} I(X^n U^{t-1} C; U_t) = nH(X_T|T) - nI(X^n U^{T-1} C; U_T|T) \\
&\leq nH(X_T) - nI(X^n U^{T-1} CT; U_T) + nI(T; U_T) = nH(X_T) - nI(X^n U^{T-1} CT; U_T).
\end{aligned}
$$

where (e) follows from the i.i.d. nature of the source $\bar{P}_U$ and the independence of the source from the common randomness.

Then, we identify the auxiliary random variable W_t with (C, X^n, U^{t-1}) for each $t \in [\![1, n]\!]$ and W with $(W_T, T) = (C, X^n, U^{T-1}, T)$. □

B.4 Achievability for secure strong coordination

In this section, we prove the achievability for Theorem 4.15 and Proposition 4.18.

B.4.1 Theorem 4.15

Here, we prove achievability for Theorem 4.15. Let $\bar{P}_X$ a distribution on $\mathcal{X}$, then $\mathcal{S} = \bigcup_{\bar{P}_X} S_{\bar{P}_X}$ where

$$
\mathcal{S}_{\bar{P}_X} := \left\{ (\bar{P}_{UV}, R_0) \;\middle|\; \begin{array}{l} \bar{P}_{UV} = \bar{P}_U \bar{P}_{V|U} \\ \exists\, W \text{ taking values in } \mathcal{W} \\ \bar{P}_{UWV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{V|W} \\ I(W;U) \leq I(X;Y) \\ R_0 \geq I(UV;W) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{V}| + 1 \end{array} \right\}. \tag{B.25}
$$

We prove that for every $\bar{P}_X$ the region $\mathcal{S}_{\bar{P}_X}$ is achievable. We note $\bar{P}_Y$ the distribution on $\mathcal{Y}$, given $\bar{P}_X$ and the channel $\bar{P}_{Y|X}$.

Then, the achievability proof proceeds as follows:

- we introduce an auxiliary random variable $W' = (X, W)$ taking values in $\mathcal{W} \times \mathcal{X}$ such that
$$\bar{P}_{UW'XYV} = \bar{P}_U \bar{P}_{W'|U} \bar{P}_{X|UW'} \bar{P}_{Y|X} \bar{P}_{V|W'Y}$$
where $\bar{P}_{X|UW'}$ is a target i.i.d. conditional distribution;
- we design a random binning and a random coding scheme for the sequence $(U^n, W^n, X^n, Y^n, V^n)$ each of which induces a joint distribution, and we prove that the two schemes have the same statistics;
- since we are not interested in jointly coordinating the whole sequence, by reducing the rate of common randomness we drop the coordination on (W^n, X^n) and we obtain strong coordination on (U^n, Y^n, V^n) only;

- because of the nature of our setting, we specialize the constraints to the special case in which the random variables of the channel are independent from the random variables of the source: (U, W, V) independent of (X, Y);
- we conclude by showing that (4.21) and (4.22) are verified.

Strong coordination of $(U^n, W^n, X^n, Y^n, V^n)$ Here we retrace the same steps of Section 3.1.1. We assume that encoder and decoder have access to common randomness C, generated uniformly at random in $[\![1, 2^{nR_0}]\!]$, and extra randomness F, generated uniformly at random in $[\![1, 2^{n\tilde{R}}]\!]$. The sequences U^n, X^n, W'^n, Y^n and V^n are jointly i.i.d. with distribution

$$\bar{P}_{U^n}\bar{P}_{W'^n|U^n}\bar{P}_{X^n|W'^nU^n}\bar{P}_{Y^n|X^n}\bar{P}_{V^n|W'^nY^n}.$$

Then, we consider the random binning and the random coding schemes defined in Section 3.1.1 and if $\tilde{R} + R_0 > H(W'|Y)$ and $R_0 + \tilde{R} < H(W'|U)$, we have already proved that

$$\mathbb{V}(P^{\text{RB}}_{U^nW'^n\hat{W}'^nX^nY^nCFV^n}, P^{\text{RC}}_{U^nW'^n\hat{W}'^nX^nY^nCFV^n}) = \delta(n), \tag{B.26}$$
$$\mathbb{V}(P^{\text{RB}}_{U^nX^nW'^n\hat{W}'^nY^nV^n}, P^{\text{RC}}_{U^nX^nW'^n\hat{W}'^nY^nV^n}) = \delta(n), \tag{B.27}$$

where $\hat{W}'^n$ is the reconstruction of W'^n at the decoder.

Remove the extra randomness F Similarly to Section 3.1.1, we observe that even though extra common randomness is required to coordinate (U^n, X^n, Y^n, V^n, W^n) we do not need it in order to coordinate only (U^n, Y^n, V^n). Observe that by Lemma 2.12, equation (B.26) implies that

$$\mathbb{V}(P^{\text{RB}}_{U^nX^nY^nV^nF}, P^{\text{RC}}_{U^nX^nY^nV^nF}) = \delta(n) \tag{B.28}$$

and therefore

$$\mathbb{V}(P^{\text{RB}}_{U^nY^nV^nF}, P^{\text{RC}}_{U^nY^nV^nF}) = \delta(n). \tag{B.29}$$

Similarly to [70], we reduce the rate of common randomness by having the two nodes agree on $F = f$. To do so, we apply Lemma 2.20 to $B^n = W'^n$, $K = F$, and $A^n = U^nY^nV^n$. If $\tilde{R} < H(W'|UYV)$, there exists a fixed binning such that

$$\mathbb{V}(P^{\text{RB}}_{U^nY^nV^nF}, Q_F P^{\text{RB}}_{U^nY^nV^n}) = \delta(n). \tag{B.30}$$

By Lemma 2.15, (B.29) implies that there exists $f \in [\![1, 2^{n\tilde{R}}]\!]$ such that

$$\mathbb{V}(P^{\text{RB}}_{U^nY^nV^n|F=f}, P^{\text{RC}}_{U^nY^nV^n|F=f}) = \delta(n). \tag{B.31}$$

Then, by fixing $F = f$ and using common randomness C, we have coordination for (U^n, Y^n, V^n).

Information constraints We have imposed the following rate constraints:

$$\begin{aligned} &H(W'|Y) < \tilde{R} + R_0 < H(W'|U),\\ &\tilde{R} < H(W'|UYV). \end{aligned}$$

Therefore we obtain:

$$\begin{aligned} &R_0 > H(W'|Y) - H(W'|UYV) = I(W';UV|Y),\\ &I(W';U) < I(W';Y). \end{aligned}$$

Separation setting We now consider the case when the random variables of the channel are independent from the random variables of the source: (U, W, V) independent of (X, Y). Since

$$\begin{aligned} I(W';UV|Y) &= I(WX;UV|Y) = I(X;UV|Y) + I(W;UV|YX)\\ &= I(XY;UV) - I(Y;UV) + I(W;UVYX) - I(W;YX)\\ &= I(W;UV), \end{aligned}$$

the target distribution and information constraints become:

$$\begin{aligned} &\bar{P}_U \bar{P}_{W|U} \bar{P}_{V|W} P_X \bar{P}_{Y|X},\\ &I(W;U) < I(X;Y),\\ &R_0 > I(W;UV). \end{aligned}$$

Observe that in this setting the joint distribution is of the form $\bar{P}_{UV}^{\otimes n} \bar{P}_{Y}^{\otimes n}$. Therefore achieving strong coordination means that $\mathbb{V}(\bar{P}_{UV}^{\otimes n} \bar{P}_{Y}^{\otimes n}, P^{\text{RC}}_{U^n V^n Y^n})$ vanishes. Moreover, by the upper bound on the mutual information in Lemma A.15, the strong secrecy condition is verified since we have proved that there exists a sequence of codes such that $\mathbb{V}(\bar{P}_{UV}^{\otimes n} \bar{P}_{Y}^{\otimes n}, P^{\text{RC}}_{U^n V^n Y^n})$ goes to zero exponentially. Moreover, strong coordination for (U^n, Y^n, V^n) implies (4.21). Hence, for every $\bar{P}_X$ we have proved that the region $\mathcal{S}_{\bar{P}_X}$ is achievable. □

B.4.2 Proposition 4.18

Here, we prove the inner bound for the more general model of Proposition 4.18. Note that with the random binning and random coding schemes proposed in the achievability proof of Theorem 4.15, coordinating Z^n as well as $(U^n, X^n, W^n, Y^n, V^n)$ does not require more common randomness:

$$\begin{aligned} &\mathbb{V}(P^{\text{RB}}_{U^n X^n W'^n \hat{W}'^n Y^n Z^n V^n}, P^{\text{RC}}_{U^n X^n W'^n \hat{W}'^n Y^n Z^n V^n})\\ &\quad = \mathbb{V}(\bar{P}_{U^n X^n W'^n \hat{W}'^n Y^n V^n}, P_{U^n X^n W'^n \hat{W}'^n Y^n V^n}) = \delta(n) \end{aligned}$$

by Lemma 2.13 since Z^n is the output of the channel $\bar{P}_{Z|X}$. The only difference with respect to the achievability proof of Theorem 4.15 is that, when we reduce the rate of common randomness, we want to assure the coordination of (U^n, Z^n, V^n). Similarly to the achievability proof of Theorem 4.15, if $\tilde{R} < H(WX|UZV)$, we have strong coordination for the sequence (U^n, Z^n, V^n):

$$\mathbb{V}(P^{\mathrm{RB}}_{U^n Z^n V^n}, P^{\mathrm{RC}}_{U^n Z^n V^n}) = \delta(n). \tag{B.32}$$

As in Theorem 4.15, we suppose that the random variables of the channel are independent from the random variables of the source: (U, W, V) independent of (X, Z, Y). Hence, since we have imposed the following rate constraints:

$$\begin{aligned}
&H(WX|Y) < \tilde{R} + R_0 < H(WX|U),\\
&\tilde{R} < H(WX|UZV),
\end{aligned}$$

we obtain:

$$\begin{aligned}
R_0 &> H(WX|Y) - H(WX|UZV) = I(WX;UZV) - I(WX;Y)\\
&= I(W;UV) + I(X;Z) - I(X;Y)\\
I(W;U) &= I(WX;U) < I(WX;Y) = I(X;Y).
\end{aligned}$$

□

B.5 Proof of cardinality bounds

In this section we prove separately the cardinality bound for all the outer bounds in this paper. Note that since the proofs are basically identical to the cardinality bound of Theorem 3.3 in Section 3.1.2, we omit most details in all the other cases.

B.5.1 Theorem 3.10

Here, we prove the cardinality bound of $\mathcal{R}_{\text{state,out}}$ in Theorem 3.10. Let $\mathcal{A} = \mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}$ and suppose that $h_i(\pi)$, $i = 1, \dots, |\mathcal{A}| + 5$, are real-valued continuous functions of $\pi \in \mathcal{P}$ such that:

$$h_i(\pi) = \begin{cases} \pi(i) & \text{for } i = 1, \dots, |\mathcal{A}| - 1\\ H(U) & \text{for } i = |\mathcal{A}|\\ H(USXV|YZ) & \text{for } i = |\mathcal{A}| + 1\\ H(Y|USX) & \text{for } i = |\mathcal{A}| + 2\\ H(V|YZ) & \text{for } i = |\mathcal{A}| + 3\\ H(V|YZUSX) & \text{for } i = |\mathcal{A}| + 4\\ H(Z|YUSX) & \text{for } i = |\mathcal{A}| + 5 \end{cases}.$$

By the Markov chain $Z - (U, S) - (X, Y, W)$, the mutual information $I(Z; XYW|US)$ is zero and once the distribution $\bar{P}_{USZXYV}$ is preserved, the mutual information $I(Z; XYW|US)$ $= H(Z|US) - H(Z|USXYW)$ only depends on $H(Z|YUSX)$. Therefore there exists an auxiliary random variable W' taking at most $|\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| + 5$ values such that the constraints on the conditional distributions and the information constraints are still verified. □

B.5.2 Theorem 4.1

Here, we prove the cardinality bound of $\mathcal{R}_{\text{PC}}$ in Theorem 4.1. Similarly, let $\mathcal{A} = \mathcal{U} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{V}$ and suppose that $h_i(\pi)$, $i = 1, \dots, |\mathcal{A}| + 4$, are real-valued continuous functions of $\pi \in \mathcal{P}$ such that:

$$h_i(\pi) = \begin{cases} \pi(i) & \text{for } i = 1, \dots, |\mathcal{A}| - 1 \\ H(U|XZ) & \text{for } i = |\mathcal{A}| \\ H(UV|XZ) & \text{for } i = |\mathcal{A}| + 1 \\ H(V|XZ) & \text{for } i = |\mathcal{A}| + 2 \\ H(V|ZUX) & \text{for } i = |\mathcal{A}| + 3 \\ H(Z|UX) & \text{for } i = |\mathcal{A}| + 4 \end{cases}.$$

The information constraint in Theorem 4.1 can be written as

$$\begin{aligned} H(X) + I(W;Z|X) - I(WX;U) &= H(X) + I(WX;Z) - I(Z;X) - I(WX;U) \\ &\overset{(a)}{=} H(X) - I(Z;X) + I(WX;Z) - I(WX;UZ) \\ &= H(X) - I(Z;X) + I(WX;U|Z) \\ &= H(X) - I(Z;X) + H(U|Z) - H(U|WXZ) \geq 0 \end{aligned}$$

where (a) follows from the fact that $I(WX;UZ)$ is equal to $I(WX;U)$ by the Markov chain $Z - U - (W,X)$. By fixing $H(UV|XZ)$ the constraint on the bound for R_0 is satisfied and similarly to the previous cases the Markov chains are still verified. Thus there exists an auxiliary random variable W' taking at most $|\mathcal{U} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{V}| + 4$ values. □

B.5.3 Theorem 4.3

Here, we prove the cardinality bound of $\mathcal{R}_{\text{LD}}$ in Theorem 4.3. First, we rewrite the constraints in the equivalent characterization of the region (4.9) as:

$$\begin{aligned} &H(U) \leq H(YZ) - H(YZ|UW), \\ &R_0 \geq I(W;USX|YZ) + H(U|WYZ) = H(USX|YZ) - H(USX|WYZ) + H(U|WYZ) \\ &\quad = H(USX|YZ) - H(USX) + H(U) + H(SX|U) - H(SX|UWYZ). \end{aligned}$$

Then, let $\mathcal{A} = \mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y}$ and suppose that $h_i(\pi)$, $i = 1, \dots, |\mathcal{A}| + 3$, are real-valued continuous functions of $\pi \in \mathcal{P}$ such that:

$$h_i(\pi) = \begin{cases} \pi(i) & \text{for } i = 1, \dots, |\mathcal{A}| - 1 \\ H(YZ|U) & \text{for } i = |\mathcal{A}| \\ H(SX|UYZ) & \text{for } i = |\mathcal{A}| + 1 \\ H(Y|USX) & \text{for } i = |\mathcal{A}| + 2 \\ H(Z|YUSX) & \text{for } i = |\mathcal{A}| + 3 \end{cases}.$$

and therefore there exists an auxiliary random variable W' taking at most $|\mathcal{U}\times\mathcal{S}\times\mathcal{Z}\times\mathcal{X}\times\mathcal{Y}|+3$ values. □

B.5.4 Theorem 4.5

Now, we prove the cardinality bound of $\mathcal{R}_{\text{SEP}}$ in Theorem 4.5. First , we consider the following equivalent characterization of the information constraints:

$$\begin{aligned}
&0 \leq H(YZ) - H(YZ|W_1W_2) - H(US) + H(US|W_1W_2),\\
&R_0 \geq H(USXV|YZ) - H(USXV|YZW_1W_2).
\end{aligned}$$

In this case we have $W = (W_1, W_2)$. Let $\mathcal{A} = \mathcal{U}\times\mathcal{S}\times\mathcal{Z}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{V}$ and suppose that $h_i(\pi)$, $i = 1, \dots, |\mathcal{A}| + 3$, are real-valued continuous functions of $\pi \in \mathcal{P}$ such that:

$$h_i(\pi) = \begin{cases} \pi(i) & \text{for } i = 1, \dots, |\mathcal{A}| - 1 \\ H(US) & \text{for } i = |\mathcal{A}| \\ H(USXV|YZ) & \text{for } i = |\mathcal{A}| + 1 \\ H(V|Z) & \text{for } i = |\mathcal{A}| + 2 \\ H(V|UZ) & \text{for } i = |\mathcal{A}| + 3 \end{cases}.$$

Then, there exists an auxiliary random variable $W' = (W_1', W_2')$ taking at most $|\mathcal{U}\times\mathcal{S}\times\mathcal{Z}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{V}|$ +3 values. □

B.5.5 Theorem 6.3

The proof of the cardinality bound $\mathcal{R}_{\text{SC,out}}$ in Theorem 6.3 for the case of the strictly causal encoder is nearly identical to the cardinality bound of $\mathcal{R}_{\text{out}}$ proved in Section 3.1.2. Let $\mathcal{A} = \mathcal{U}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{V}$, and suppose that $h_i(\pi)$, $i = 1, \dots, |\mathcal{A}| + 4$, are real-valued continuous functions of $\pi \in \mathcal{P}$ such that:

$$h_i(\pi) = \begin{cases} \pi(i) & \text{for } i = 1, \dots, |\mathcal{A}| - 1 \\ H(U|X) & \text{for } i = |\mathcal{A}| \\ H(UXV|Y) & \text{for } i = |\mathcal{A}| + 1 \\ H(Y|X) & \text{for } i = |\mathcal{A}| + 2 \\ H(V|Y) & \text{for } i = |\mathcal{A}| + 3 \\ H(V|UXY) & \text{for } i = |\mathcal{A}| + 4 \end{cases}.$$

We can rewrite the inequalities and the Markov chains in (6.2) as

$$\begin{aligned}
&H(U) - H(U|WX) \leq I(X;Y),\\
&R_0 \geq H(UXV|Y) - H(UXV|WY),\\
&I(Y;W|X) = H(Y|X) - H(Y|XW) = 0,
\end{aligned}$$

$$I(V;UX|YW) = H(V|YW) - H(V|UXYW) = 0.$$

Then, there exists an auxiliary random variable W' taking at most $|\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| + 4$ values such that the constraints on the conditional distributions and the information constraints are still verified. $\square$

C | Polar coding achievability proofs

In this chapter, we detail the achievability proofs of Chapter 5.

C.1 Proof of Remark 5.8

We define the distributions

$$
\begin{aligned}
P_{Z^n Z^n B^n}(\mathbf{z},\mathbf{z}',\mathbf{b}) :=& P_{Z^n B^n}(\mathbf{z},\mathbf{b}) \mathbb{1}_{Z^n|Z^n}(\mathbf{z}'|\mathbf{z}) = \prod_{i=1}^{n} P_{Z_i|Z^{i-1}B^n}(\mathbf{z}_i|\mathbf{z}^{i-1},\mathbf{b}) \mathbb{1}_{Z|Z}(z_i'|z_i),\\
P_{Z^n \hat{Z}^n B^n}(\mathbf{z},\mathbf{z}',\mathbf{b}) :=& P_{Z^n B^n}(\mathbf{z},\mathbf{b}) P_{\hat{Z}^n|Z^n B^n}(\mathbf{z}'|\mathbf{z},\mathbf{b})\\
=& P_{Z^n B^n}(\mathbf{z},\mathbf{b}) \prod_{i=1}^{n} P_{\hat{Z}_i|\hat{Z}^{i-1}Z^nB^n}(\mathbf{z}_i'|\mathbf{z}'^{i-1},\mathbf{z},\mathbf{b})\\
=& P_{Z^n B^n}(\mathbf{z},\mathbf{b}) \prod_{i\in\mathcal{H}_{A|B}} \mathbb{1}_{Z|Z}(z_i'|z_i) \prod_{i\notin\mathcal{H}_{A|B}} P_{Z_i|Z^{i-1}B^n}(\mathbf{z}_i'|\mathbf{z}'^{i-1},\mathbf{b}).
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
\mathbb{D}(P_{Z^n Z^n B^n} \| P_{Z^n \hat{Z}^n B^n}) &= \sum_{\substack{\mathbf{z},\mathbf{z}'\in\mathcal{Z}^n\\ \mathbf{b}\in\mathcal{B}^n}} P_{Z^n Z^n B^n}(\mathbf{z},\mathbf{z}',\mathbf{b}) \log \frac{P_{Z^n Z^n B^n}(\mathbf{z},\mathbf{z}',\mathbf{b})}{P_{Z^n \hat{Z}^n B^n}(\mathbf{z},\mathbf{z}',\mathbf{b})}\\
&= \sum_{\substack{\mathbf{z},\mathbf{z}'\in\mathcal{Z}^n\\ \mathbf{b}\in\mathcal{B}^n}} P_{Z^n Z^n B^n}(\mathbf{z},\mathbf{z}',\mathbf{b}) \log \left(\prod_{i\in\mathcal{H}^c_{A|B}} \frac{\mathbb{1}_{Z|Z}(z_i'|z_i)}{P_{Z_i|Z^{i-1}B^n}(\mathbf{z}_i'|\mathbf{z}'^{i-1},\mathbf{b})} \right)\\
&= \sum_{i\in\mathcal{H}^c_{A|B}} H(Z_i|Z^{i-1}B^n)\\
&\le |\mathcal{H}^c_{A|B}|\delta_n \le n\delta_n.
\end{aligned}
$$

Therefore, by Pinsker's inequality,

$$
\mathbb{V}(P_{Z^n Z^n B^n}, P_{Z^n \hat{Z}^n B^n}) \le \sqrt{2\log 2}\sqrt{n\delta_n}.
$$

To conclude, note that:

$$\begin{aligned}
\mathbb{V}(P_{Z^n Z^n B^n}, P_{Z^n \hat{Z}^n B^n}) &= \frac{1}{2} \sum_{\substack{\mathbf{z},\mathbf{z}' \in \mathcal{Z}^n \\ \mathbf{b} \in \mathcal{B}^n}} |P_{Z^n Z^n B^n}(\mathbf{z},\mathbf{b}) \mathbb{1}_{Z^n|Z^n}(\mathbf{z}'|\mathbf{z}) - P_{Z^n \hat{Z}^n B^n}(\mathbf{z},\mathbf{z}',\mathbf{b})| \\
&= \frac{1}{2} \sum_{\substack{\mathbf{z} \in \mathcal{Z}^n \\ \mathbf{b} \in \mathcal{B}^n}} |P_{Z^n B^n}(\mathbf{z},\mathbf{b}) - P_{Z^n \hat{Z}^n B^n}(\mathbf{z},\mathbf{z},\mathbf{b})| \\
&\quad + \frac{1}{2} \sum_{\substack{\mathbf{z} \in \mathcal{Z}^n \\ \mathbf{b} \in \mathcal{B}^n}} \sum_{\substack{\mathbf{z}' \in \mathcal{Z}^n \\ \mathbf{z}' \neq \mathbf{z}}} P_{Z^n \hat{Z}^n B^n}(\mathbf{z},\mathbf{z}',\mathbf{b}) \\
&= \frac{\mathbb{P}\{Z^n \neq \hat{Z}^n\} + \mathbb{P}\{Z^n \neq \hat{Z}^n\}}{2}.
\end{aligned}$$

Therefore, $\mathbb{P}\{Z^n \neq \hat{Z}^n\} \to 0$. □

C.2 Proofs of Lemma 5.9

Here, we prove the polar coding counterpart to channel randomness extraction for discrete memoryless sources and channels proved in Lemma 2.20. First, we prove the statement about the K-L divergence. For every $j \in [\![1, n]\!]$, we generate $\hat{Z}^n$ stochastically from the conditional distribution $P_{\hat{Z}_j|\hat{Z}^{j-1}B^n}$ defined as

$$P_{\hat{Z}_j|\hat{Z}^{j-1}B^n} := \begin{cases} Q_Z & \text{if } j \in \mathcal{V}_{A|B}, \\ P_{\hat{Z}_j|\hat{Z}^{j-1}B^n} & \text{if } j \in \mathcal{V}_{A|B}^c. \end{cases}$$

Then, we have

$$\begin{aligned}
\mathbb{D}(P_{Z[\mathcal{V}_{A|B}]B^n} \| Q_{Z[\mathcal{V}_{A|B}]} P_{B^n}) &= \mathbb{D}(P_{Z[\mathcal{V}_{A|B}]B^n} \| P_{\hat{Z}[\mathcal{V}_{A|B}]B^n}) \\
&\overset{(a)}{=} \mathbb{D}(P_{Z[\mathcal{V}_{A|B}]|B^n} \| P_{\hat{Z}[\mathcal{V}_{A|B}]|B^n} | P_{B^n}) \\
&\overset{(b)}{=} \sum_{j=1}^{n} \mathbb{D}(P_{Z_j|Z^{j-1}B^n} \| P_{\hat{Z}_j|\hat{Z}^{j-1}B^n} | P_{Z^{j-1}B^n}) \\
&\overset{(c)}{=} \sum_{j \in \mathcal{V}_{A|B}} \mathbb{D}(P_{Z_j|Z^{j-1}B^n} \| P_{\hat{Z}_j|\hat{Z}^{j-1}B^n} | P_{Z^{j-1}B^n}) \\
&\overset{(d)}{=} \sum_{j \in \mathcal{V}_{A|B}} (1 - H(Z_j|Z^{j-1}B^n)) \overset{(e)}{<} \delta_n |\mathcal{V}_{A|B}| \leq n\delta_n
\end{aligned}$$

where (a), and (b) come from the chain rule for divergence (Lemma A.5), (c), (d) and (e) follow from the definition of $\hat{Z}^n$ and $\mathcal{V}_{A|B}$.

Then Pinsker's inequality implies

$$\mathbb{V}(P_{Z[\mathcal{V}_{A|B}]B^n}, Q_{\hat{Z}[\mathcal{V}_{A|B}]} P_{B^n}) \leq \sqrt{2 \log 2} \sqrt{n\delta_n}.$$ □

C.3 Empirical coordination

Here we prove that the empirical coordination region can be achieved using polar codes. For brevity, we only detail the results for the case when the encoder and decoder are both non-causal.

C.3.1 Non-causal encoder and decoder

Before describing the polar coding scheme, we recall the definitions of empirical coordination.

Definition C.1 - ***Achievability for empirical coordination [24, 22]*** *A distribution $\bar{P}_{UXYV}$ is achievable for empirical coordination if there exists a sequence (f_n, g_n) of encoders-decoders such that for all $\varepsilon > 0$, there exists $\bar{n}$ such that for all $n \geq \bar{n}$:*

$$\mathbb{P}\left\{\mathbb{V}\left(T_{U^nX^nY^nV^n}, \bar{P}_{UXYV}\right) > \varepsilon\right\} < \varepsilon$$

where

$$T_{U^nX^nY^nV^n}(u,x,y,v) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{(u_i,x_i,y_i,v_i) = (u,x,y,v)\}$$

is the joint histogram of the actions induced by the code (f_n, g_n).

Definition C.2 - ***Empirical coordination region [24, 22]*** *The empirical coordination region $\mathcal{R}_e$ is the closure of the set of achievable distributions $\bar{P}_{UXYV}$.*

As anticipated in Section 3.1, in the case of non-causal encoder and decoder, the problem of characterizing the empirical coordination region is still open, but the following inner bound was proved in [24].

Theorem C.3 - ***Non-causal encoder [24, Theorem 1]*** *Let $\bar{P}_U$ and $\bar{P}_{Y|X}$ be the given source and channel parameters. When the encoder and decoder are allowed to be non-causal, the region $\mathcal{R}_{e,in} \subset \mathcal{R}_e$ defined below is included in the empirical coordination region.*

$$\mathcal{R}_{\text{e,in}} := \left\{ (\bar{P}_{UXYV}) \;\middle|\; \begin{array}{l} \bar{P}_{UXYV} = \bar{P}_U \bar{P}_{X|U} \bar{P}_{Y|X} \bar{P}_{V|UXY} \\ \exists\, W \textit{ taking values in } \mathcal{W} \\ \bar{P}_{UXYWV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{X|UW} \bar{P}_{Y|X} \bar{P}_{V|WY} \\ I(W;U) \leq I(W;Y) \end{array} \right\}. \tag{C.1}$$

For brevity, we only focus on the set of achievable distributions in $\mathcal{R}_{\text{e,in}}$ for which the auxiliary variable W is binary. The scheme can be generalized to the case of a non-binary random variable W using non-binary polar codes.

Theorem C.4 *For all P_{UXYV} for which there exists W taking values in $\mathcal{W} = \{0,1\}$ such that*

$$P_{UWXYV} = \bar{P}_U \bar{P}_{W|U} \bar{P}_{X|UW} \bar{P}_{Y|X} \bar{P}_{V|WY},$$

there exists an explicit polar coding scheme that achieves empirical coordination with rate of *common randomness that goes to zero as n goes to infinity.*

Polar coding scheme We suppose that P_{UXYV} belongs to $\mathcal{R}_{\text{e,in}}$ and show how to achieve empirical coordination with polar codes. Consider the random vectors U^n, W^n, X^n, Y^n and V^n generated i.i.d. according to P_{UWXYV} that satisfies (C.1). Let $Z^n = W^n G_n$ the polarization of W^n, where G_n is the source polarization transform. We consider the sets A_1, A_2, A_3, A_4 defined in (5.3).

Encoding The encoding algorithm is similar to the algorithm in Section 5.3.1, but for empirical coordination it is possible to recycle more common randomness. The encoder generates $W^n_{(i)}$ following Algorithm 5. The chaining construction proceeds as follows:

- The bits in $A_1 \subset \mathcal{V}_{W|U}$ in block $i \in [\![1, k]\!]$ are chosen with uniform probability using a uniform randomness source $\bar{C}_1$ shared with the decoder, and their value is reused over all blocks;
- In the first block the bits in $A_2 \subset \mathcal{V}_{W|U}$ are chosen with uniform probability using M, a local randomness source;
- For the following blocks, let A_3' be a subset of A_2 such that $|A_3'| = |A_3|$. The bits of A_3 in block i are sent to A_3' in the block $i+1$ using a one-time pad with key C_2 and they are uniform thanks to the Crypto Lemma.
- The bits in $A_2 \setminus A_3'$ are chosen with uniform probability using the local randomness source M;
- The bits in A_3 and in A_4 are generated according to the previous bits using successive cancellation encoding as in Definition 5.6. Note that it is possible to sample efficiently from $\bar{P}_{Z_i|Z^{i-1}U^n}$ given U^n.

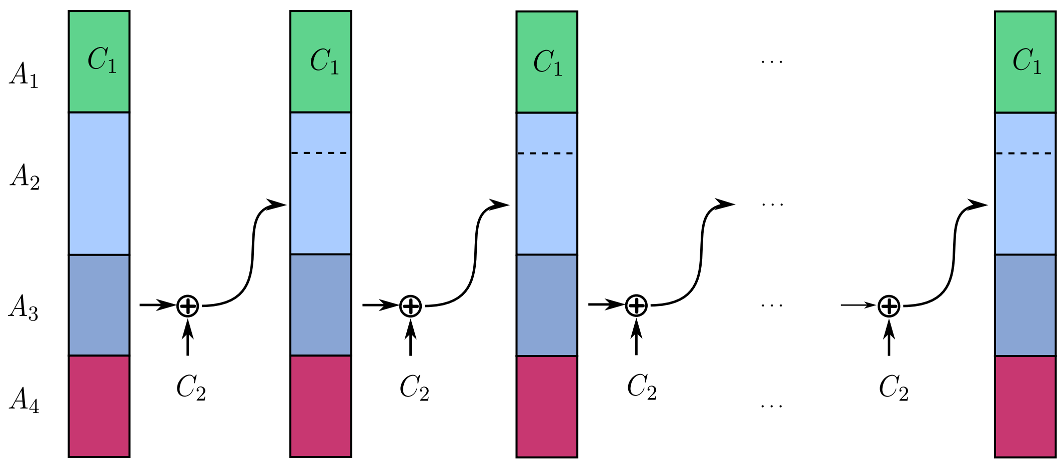

Figure C.1: Chaining construction for block Markov encoding

Algorithm 5: Encoding algorithm at Node 1

Input : $(U^n_{(1)}, \dots, U^n_{(k)})$, M local randomness (uniform random bits) and common randomness $C = (C_1, C_2)$ shared with Node 2: C_1 of size $|A_1|$ and C_2 of size $|A_3|$.

Output: $(Z^n_{(1)}, \dots, Z^n_{(k)})$

if $i = 1$ **then**

  $Z_{(1)}[A_1] \longleftarrow C_1 \qquad Z_{(1)}[A_2] \longleftarrow M$

  for $j \in A_3 \cup A_4$ **do**

    Given $U^n_{(1)}$, succ. draw the bits $Z_{j,(1)}$ according to

$$P_{Z_j|Z^{j-1}U^n}(Z_{j,(1)}|Z^{j-1}_{(1)}U^n_{(1)}) \tag{C.2}$$

  end

end

for $i = 2, \dots, k$ **do**

  $Z_{(i)}[A_1] \longleftarrow C_1 \qquad Z_{(i)}[A'_3] \longleftarrow \tilde{V}_{i-1}[A_3] \oplus C_2$

  $Z_{(i)}[A_2 \setminus A'_3] \longleftarrow M$

  for $j \in A_3 \cup A_4$ **do**

    Given $U^n_{(i)}$, succ. draw the bits $Z_{j,(i)}$ according to

$$P_{Z_j|Z^{j-1}U^n}(Z_{j,(i)}|Z^{j-1}_{(i)}U^n_{(i)}) \tag{C.3}$$

  end

end

Then, the encoder computes $W^n_{(i)} = Z^n_{(i)} G_n$ for $i = 1, \dots, k$ and generates $X^n_{(i)}$ symbol by symbol from $W^n_{(i)}$ and $U^n_{(i)}$ using the conditional distribution

$$\bar{P}_{X_{j,(i)}|W_{j,(i)}U_{j,(i)}}(x|\tilde{w}, u) = \bar{P}_{X|WU}(x|\tilde{w}, u)$$

and sends $X^n_{(i)}$ over the channel.

We use an extra $(k+1)$-th block to send a version of $Z_{(k)}[A_3]$ encoded with a good channel code. In particular, this can be done using the polar code construction for asymmetric channels stated in [33]. Let $S^n = X^n G_n$ be the polarized version of X^n. We place the information $Z_{(k)}[A_3]$ in the positions of S^n indexed by $\mathcal{V}_X \cap \mathcal{H}^c_{X|Y}$. We note that $\mathcal{V}_X \cap \mathcal{H}^c_{X|Y}$ has cardinality approximately equal to $nI(X;Y)$ [33]. We have $|A_3| \leq |A_2| \leq |\mathcal{V}_V \cap \mathcal{H}^c_{V|Y}|$, which is approximately $nI(W;Y)$. By hypothesis, we have the Markov chain $W - X - Y$ and therefore $|A_3| \leq nI(X;Y)$. We can send the bits in A_3 with vanishing error probability. The scheme in [33] requires common randomness, which will have vanishing rate when k is large enough since it's used only in the last block, and uniform messages, which can be achieved using a one-time-pad as before. Finally, $\tilde{X}^n_{(k+1)}$ is the output of the channel code described above.

Remark C.5 - ***Last block is not coordinated.*** Observe that, contrary to the more stringent requirements of strong coordination, when dealing with empirical coordination we do not strictly need to coordinate all $k+1$ blocks. In fact, if k is large enough, the coordination of the first k blocks is enough to ensure empirical coordination, but we still need to send $V_{(k)}[A_3]$ to the decoder in order to initialize the decoding process.

Decoding The decoder observes $(Y^n_{(1)}, \dots, Y^n_{(k+1)})$ and the $(k+1)$-th block allows it to decode in reverse order. The decoding algorithm, described in Algorithm 6, is similar to the one for strong coordination, detailed in Section 5.3.1.

Algorithm 6: Decoding algorithm at Node 2

Input : $(Y^n_{(1)}, \dots, Y^n_{(k+1)})$, $C = (C_1, C_2)$ common randomness shared with Node 1
Output: $(\hat{Z}^n_{(1)}, \dots, \hat{Z}^n_{(k)})$
for $i = k, \dots, 1$ **do**
  $\hat{Z}_{(i)}[A_1] \longleftarrow C_1$
  if $i = k$ **then**
    $\hat{Z}_{(i)}[A_3] \longleftarrow Y^n_{(k+1)}$ as in [33]
  end
  else
    $\hat{Z}_{(i)}[A_3] \longleftarrow \hat{Z}_{(i+1)}[A_3']$
  end
  for $j \in A_2 \cup A_4$ **do**
    Successively draw the bits according to

$$\hat{Z}^j_{(i)} = \begin{cases} 0 & \text{if } L_n(Y^n_{(i)}, Z^{j-1}_{(i)}) \geq 1 \\ 1 & \text{else} \end{cases}$$

$$\text{where} \quad L_n(Y^n_{(i)}, Z^{j-1}_{(i)}) = \frac{P_{Z_{j,(i)}|Z^{j-1}_{(i)}Y^n_{(i)}}\left(0|\hat{Z}^{j-1}_{(i)}Y^n_{(i)}\right)}{P_{Z_{j,(i)}|Z^{j-1}_{(i)}Y^n_{(i)}}\left(1|\hat{Z}^{j-1}_{(i)}Y^n_{(i)}\right)}$$

  end
end

Rate of common randomness The rate of common randomness C is negligible since:

$$\lim_{\substack{n\to\infty \\ k\to\infty}} \frac{|A_1 \cup A_3|}{kn} = \lim_{\substack{n\to\infty \\ k\to\infty}} \frac{|\mathcal{H}_{W|Y}|}{kn} = \lim_{k\to\infty} \frac{H(W|Y)}{k} = 0.$$

Preliminary results We first state a few lemmas that we will need to prove Theorem C.4.

Lemma C.6 *For any $i \in [\![1, k]\!]$, for all $\epsilon_0 > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left\{\mathbb{V}\left(T_{U^n_{(i)}W^n_{(i)}}, \bar{P}_{UW}\right) > \epsilon_0\right\} = 0.$$

Proof. For all $\epsilon_0 > 0$, we define

$$\mathcal{T}_{\epsilon_0}(\bar{P}_{UW}) := \{(U^n, W^n) \mid \mathbb{V}(T_{U^nW^n}, \bar{P}_{UW}) \leq \epsilon_0\}$$

$$\mathbb{P}_{\bar{P}_{UW}}\{(\mathbf{u},\mathbf{w}) \in \mathcal{T}_{\epsilon_0}(\bar{P}_{UW})\} := \sum_{\mathbf{u},\mathbf{w}} \bar{P}_{U^nW^n}(\mathbf{u},\mathbf{w})\mathbb{1}\{(\mathbf{u},\mathbf{w}) \in \mathcal{T}_{\epsilon_0}(\bar{P}_{UW})\}.$$

Note that $\lim_{n\to\infty} \mathbb{P}_{\bar{P}_{UW}}\{(\mathbf{u},\mathbf{w}) \in \mathcal{T}_{\epsilon_0}(\bar{P}_{UW})\} = 1$.

Let $i \in [\![1,k]\!]$, we have:

$$\begin{aligned}
&\mathbb{P}_{\bar{P}_{UW}}\{\mathbb{V}(T_{U^n_{(i)}W^n_{(i)}}, \bar{P}_{UW}) > \epsilon_0\} \\
&= \sum_{\mathbf{u},\mathbf{w}} \bar{P}_{U^n_{(i)}W^n_{(i)}}(\mathbf{u},\mathbf{w})\mathbb{1}\{(\mathbf{u},\mathbf{w}) \notin \mathcal{T}_{\epsilon_0}(\bar{P}_{UW})\} \\
&= \sum_{\mathbf{u},\mathbf{w}} (\bar{P}_{U^n_{(i)}W^n_{(i)}}(\mathbf{u},\mathbf{w}) - \bar{P}_{U^nW^n}(\mathbf{u},\mathbf{w}) + \bar{P}_{U^nW^n}(\mathbf{u},\mathbf{w}))\mathbb{1}\{(\mathbf{u},\mathbf{w}) \notin \mathcal{T}_{\epsilon_0}(\bar{P}_{UW})\} \\
&\leq \mathbb{V}(\bar{P}_{U^nW^n}, \bar{P}_{U^nW^n}) + \mathbb{P}_{\bar{P}_{UW}}\{(\mathbf{u},\mathbf{w}) \notin \mathcal{T}_{\epsilon_0}(\bar{P}_{UW})\}
\end{aligned}$$

which tends to 0 thanks to a typicality argument and the fact that for any $i \in [\![1,k]\!]$, let $\delta_n = 2^{-n^\beta}$ for some $0 < \beta < 1/2$

$$\mathbb{V}(\bar{P}_{U^nW^n}, \bar{P}_{U^n_{(i)}W^n_{(i)}}) \leq \sqrt{2\log 2}\sqrt{n\delta_n} \tag{C.4}$$

which follows from (5.7). □

Lemma C.7 *Let P_A a distribution, A^n a random vector, B^n a random vector generated from A^n with i.i.d. conditional distribution $P_{B|A}$ and suppose $\lim_{n\to\infty} \mathbb{P}\{\mathbb{V}(T_{A^n}, P_A) > \epsilon\} = 0$. Then, for all $\epsilon' > \epsilon$ we have:*

$$\lim_{n\to\infty} \mathbb{P}\{\mathbb{V}(T_{A^nB^n}, P_{AB}) > \epsilon'\} = 0.$$

Proof. We have:

$$\begin{aligned}
\mathbb{P}\{\mathbb{V}(T_{A^nB^n}, P_{AB}) > \epsilon'\} &\leq \mathbb{P}\{\mathbb{V}(T_{A^n}, P_A) > \epsilon\} \\
&\quad + \mathbb{P}\{\mathbb{V}(T_{A^n}, P_A) \leq \epsilon\}\mathbb{P}\{\mathbb{V}(T_{A^nB^n}, P_{AB}) > \epsilon' \mid \mathbb{V}(T_{A^n}, P_A) \leq \epsilon\}.
\end{aligned}$$

Then as n goes to infinity, the first term tends to zero by the conditional typicality lemma and the second tends to zero by hypothesis. □

Lemma C.8 *Let X^n, $\tilde{X}^n$ two possibly dependent random sequences taking values in $\mathcal{X}^n$ and define*

$$T_{(X^n,\tilde{X}^n)}(x) := \frac{1}{2n}\sum_{i=1}^{n}(\mathbb{1}\{X_i = x\} + \mathbb{1}\{\tilde{X}_i = x\}).$$

Then for any distribution P on $\mathcal{X}$,

$$\mathbb{V}(T_{(X^n,\tilde{X}^n)}, P) \leq \frac{1}{2}\mathbb{V}(T_{X^n}, P) + \frac{1}{2}\mathbb{V}(T_{\tilde{X}^n}, P).$$

Proof. The statement follows from the inequalities:

$$\left|T_{(X^n,\tilde{X}^n)}(x) - P(x)\right| = \left|\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\mathbb{1}\{X_i = x\}}{n} + \frac{\mathbb{1}\{\tilde{X}_i = x\}}{n}\right) - \frac{P(x)}{2} - \frac{P(x)}{2}\right|$$

$$\leq \frac{1}{2}\left|\sum_{i=1}^{n}\frac{\mathbb{1}\{X_i = x\}}{n} - P(x)\right| + \frac{1}{2}\left|\sum_{i=1}^{n}\frac{\mathbb{1}\{\tilde{X}_i = x\}}{n} - P(x)\right|. \qquad \square$$

Lemma C.9 $\mathbb{V}(T_{X^n}, P_X) \leq \mathbb{V}(T_{X^nY^n}, P_{XY})$.

The proof of Lemma C.9 is straightforward and thus omitted.

Achievability proof Now, we want to show that the polar coding scheme achieves empirical coordination. Given $\epsilon > 0$, we want to prove that:

$$\lim_{\substack{n\to\infty \\ k\to\infty}} \mathbb{P}\{\mathbb{V}(T_{U^n_{(1:k+1)}X^n_{(1:k+1)}Y^n_{(1:k+1)}V^n_{(1:k+1)}}, \bar{P}_{UXYV}) > \epsilon\} = 0.$$

In order to simplify the notation, we set the joint types as

$$T := T_{U^n_{(1:k+1)}W^n_{(1:k+1)}X^n_{(1:k+1)}Y^n_{(1:k+1)}V^n_{(1:k+1)}},$$
$$T_i := T_{U^n_{(i)}W^n_{(i)}X^n_{(i)}Y^n_{(i)}V^n_{(i)}} \quad i \in [\![1, k+1]\!].$$

Lemma C.6 states that for $i \in [\![1, k]\!]$ and for all $\epsilon_0 > 0$,

$$\lim_{n\to\infty} \mathbb{P}\{\mathbb{V}(T_{U^n_{(i)}W^n_{(i)}}, \bar{P}_{UW}) > \epsilon_0\} = 0.$$

Then, because of Lemma C.7, we have that for all $\epsilon' > \epsilon_0$

$$\lim_{n\to\infty} \mathbb{P}\{\mathbb{V}(T_{U^n_{(i)}W^n_{(i)}X^n_{(i)}Y^n_{(i)}}, \bar{P}_{UWXY}) > \epsilon'\} = 0.$$

We can apply Lemma C.7 again and add V, but since V is generated by $\hat{W}$ and not by $\tilde{W}$, we need the conditional probability: $\forall \epsilon > \epsilon'$ for $i \in [\![1, k]\!]$ we have

$$\lim_{n\to\infty} \mathbb{P}\{\mathbb{V}(T_i, \bar{P}_{UWXYV}) > \epsilon \mid \hat{W}^{1:n}_{(i)} = \tilde{W}^n_{(i)}\} = 0$$

We can write:

$$\begin{aligned}
&\mathbb{P}\{\mathbb{V}(T_i, \bar{P}_{UWXYV}) > \epsilon\} \\
&\quad = \mathbb{P}\{\mathbb{V}(T_i, \bar{P}_{UWXYV}) > \epsilon \mid \hat{W}^n_{(i)} = \tilde{W}^n_{(i)}\}\mathbb{P}\{\hat{W}^n_{(i)} = \tilde{W}^n_{(i)}\} \\
&\quad + \mathbb{P}\{\mathbb{V}(T_i, \bar{P}_{UWXYV}) > \epsilon \mid \hat{W}^n_{(i)} \neq \tilde{W}^n_{(i)}\}\mathbb{P}\{\hat{W}^n_i \neq \tilde{W}^n_i\}.
\end{aligned}$$

Note that the last term tends to 0 since $\tilde{W}^n$ is equal to $\hat{W}^n$ with high probability because of Theorem 5.5. Hence for $i \in [\![1, k]\!]$ we have

$$\lim_{n\to\infty} \mathbb{P}\{\mathbb{V}(T_i, \bar{P}_{UWXYV}) > \epsilon\} = 0.$$

The convergence in probability of T to P_{UWXYV} follows from the convergence in probability of T_i to P_{UWXYV} for $i \in [\![1,k]\!]$ (coordination in the first k blocks). In fact, observe that by Lemma C.8,

$$\mathbb{V}(T, \bar{P}_{UWXYV}) \leq \frac{1}{k+1} \sum_{i=1}^{k+1} \mathbb{V}(T_i, \bar{P}_{UWXYV}).$$

This implies that:

$$\mathbb{E}_T \left[\mathbb{V} \left(T, \bar{P}_{UWXYV} \right) \right] \leq \frac{1}{k+1} \sum_{i=1}^{k+1} \mathbb{E}_T \left[\mathbb{V} \left(T_i, \bar{P}_{UWXYV} \right) \right]. \tag{C.5}$$

The right hand side in (C.5) goes to zero since:

- for $i \in [\![1,k]\!]$ we already have the convergence in probability of $\mathbb{V}\left(T_i, \bar{P}_{UWXYV}\right)$ to zero, therefore the convergence in mean since $\mathbb{V}\left(T_i, \bar{P}_{UWXYV}\right)$ is bounded for all i;
- for $i = k+1$, since T_{k+1} and $\bar{P}_{UWXYV}$ are probability distributions, $\mathbb{V}\left(T_{k+1}, \bar{P}_{UWXYV}\right) \leq 2$. For k large enough $2/(k+1)$ goes to zero, then $\mathbb{E}[2]/(k+1) = 2/(k+1)$ goes to zero and empirical coordination still holds.

Then, the left hand side in (C.5) goes to zero and because convergence in mean implies convergence in probability, we have the convergence in probability of $\mathbb{V}\left(T, \bar{P}_{UWXYV}\right)$ to zero. To complete the proof we recall that because of Lemma C.9, $\mathbb{V}\left(T, \bar{P}_{UWXYV}\right) < \epsilon$ implies that

$$\mathbb{V}\left(T_{U^n_{(1:k+1)} X^n_{(1:k+1)} Y^n_{(1:k+1)} V^n_{(1:k+1)}}, \bar{P}_{UXYV}\right) < \epsilon. \qquad \square$$

C.3.2 Strictly causal encoder

As anticipated in Section 6.1, in the case of strictly causal encoder and non-causal decoder, [24] characterizes the empirical coordination region.

Theorem C.10 - *Strictly causal encoder [24, Theorem 3]* *Let $\bar{P}_U$ and $\bar{P}_{Y|X}$ be the given source and channel parameters. When the encoder is strictly causal, the empirical coordination region $\mathcal{R}_{SC,e}$ is given by*

$$\mathcal{R}_{SC,e} := \left\{ \bar{P}_{UXYV} \;\middle|\; \begin{array}{l} \bar{P}_{UXYV} = \bar{P}_U \bar{P}_X \bar{P}_{Y|X} \bar{P}_{V|UXY} \\ \exists\, W \text{ taking values in } \mathcal{U} \\ P_{UWXYV} = \bar{P}_U \bar{P}_X \bar{P}_{W|XU} \bar{P}_{Y|X} \bar{P}_{V|WY} \\ I(X,W;U) \leq I(X,W;Y) \\ |\mathcal{W}| \leq |\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| + 4 \end{array} \right\}. \tag{C.6}$$

Theorem C.11 *For all $\bar{P}_{UXYV} \in \mathcal{R}_{SC,e}$ such that $\mathcal{W} = \{0,1\}$, there exists an explicit polar coding scheme that achieves empirical coordination with vanishing rate of common randomness.*

Remark C.12 - *Cardinality of $\mathcal{W}$*. Since W is binary we only achieve a subset of $\mathcal{R}_{\text{SC,e}}$. The proof can be generalized to the case where $|\mathcal{W}|$ is a prime number using non-binary polar codes.

Polar coding scheme Let $S^n = X^n G_n$ be the polarization of X^n, and $Z^n = W^n G_n$ be the polarization of W^n and consider the sets A_1, A_2, A_3, A_4 and B_1, B_3 and B_4 as defined in (6.15) and (6.17) respectively.

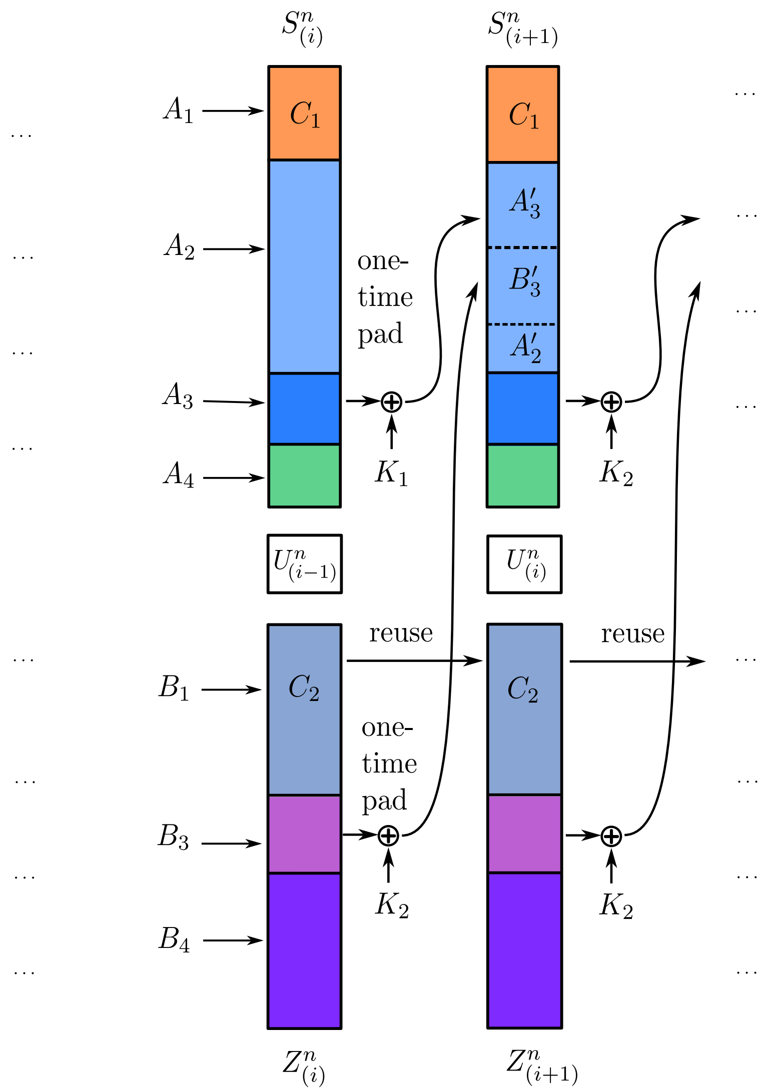

Figure C.2: Chaining construction for block Markov encoding

Encoding The encoding algorithm, detailed in Algorithm 7, is similar to the one in Section 6.2.1, but, as in Section C.3.1, for empirical coordination it is possible to recycle more common randomness, as shown in Figure C.3.2 The encoder observes $U^n_{(0:k)} := (U^n_{(0)}, U^n_{(1)}, \dots, U^n_{(k)})$, where $U^n_{(0)}$ is a uniform random sequence and $U^n_{(i)}$ for $i \in [\![1, k]\!]$ are k blocks of the source. Then, the chaining construction, detailed in Algorithm 7, proceeds as follows:

- The bits in $A_1 \subset \mathcal{V}_X$ in block $i \in [\![1, k]\!]$ are chosen with uniform probability using a uniform randomness source C_1 shared with the decoder;

Algorithm 7: Encoding algorithm at Node 1

Input : $(U^n_{(0)}, \dots, U^n_{(k)})$, local randomness (uniform random bits) M and common randomness $C = (C_1, C_2, K_1, K_2)$ shared with Node 2: C_1 of size $|A_1|$ and K_1 of size $|A_3|$, C_2 of size $|B_1|$, and K_2 of size $|B_3|$

Output: $(S^n_{(1)}, \dots, S^n_{(k)})$, $(Z^n_{(1)}, \dots, Z^n_{(k)})$

if $i = 1$ **then**

$S_{(1)}[A_1] \longleftarrow C_1, \quad S_{(1)}[A_2] \longleftarrow M$

for $j \in A_3 \cup A_4$ **do**

Successively draw the bits $S_{j,(1)}$ according to

$$\bar{P}_{S_j|S^{j-1}}(S_{(i),j}|S^{j-1}_{(i)}) \tag{C.7}$$

$Z_{(1)}[B'_1] \longleftarrow C_2$

for $j \in B_3 \cup B_4$ **do**

Given $U^n_{(1)}$, successively draw the bits $Z^j_{(1)}$ according to

$$\bar{P}_{Z_j|Z^{j-1}X^nU^n}(Z_{(i),j}|Z^{j-1}_{(i)}X^n_{(i)}U^n_{(i-1)}) \tag{C.8}$$

for $i = 2, \dots, k$ **do**

$$S_{(i)}[A_1] \longleftarrow C_1, \quad S_{(i)}[A'_2] \longleftarrow M,$$
$$S_{(i)}[B'_3] \longleftarrow Z_{(i-1)}[B_3] \oplus K_2, \quad S_{(i)}[A'_3] \longleftarrow S_{(i-1)}[A_3] \oplus K_1$$

for $j \in A_3 \cup A_4$ **do**

Successively draw the bits $S_{(i),j}$ according to (C.7)

$Z_{(i)}[B'_1] \longleftarrow C_2$

for $j \in B_3 \cup B_4$ **do**

Succ. draw the bits $Z_{(i),j}$ according to (C.8)

- In the first block the bits in $A_2 \subset \mathcal{V}_X$ are chosen with uniform probability using a local randomness source M;
- The bits in $B_1 \subset \mathcal{V}_{W|XU}$ in block $i \in [\![1, k]\!]$ are chosen with uniform probability using a uniform randomness source C_2 shared with the decoder, and their value is reused over all blocks;
- The bits in $A_3 \cup A_4$ and $B_3 \cup B_4$ are generated according to the previous bits using successive cancellation encoding as in Definition 5.6. Note that it is possible to sample efficiently from $\bar{P}_{S_j|S^{j-1}}$ and $\bar{P}_{Z_j|Z^{j-1}X^nU^n}$ (given U^n and X^n) respectively;
- From the second block, the encoder generates the bits of A_2 in the following way. Let A'_3 and B'_3 be two disjoint subsets of A_2 of cardinality $|A'_3| = |A_3|$ and $|B'_3| = |B_3|$. The existence of those disjoint subsets is guaranteed by Remark 6.9 and Remark 6.10. The bits of A_3 and B_3 in block i are used as A'_3 and B'_3 in block $i + 1$ using one-time pads with keys K_1 and K_2 respectively and are uniform thanks to the Crypto Lemma. Finally, the bits in $A'_2 := A_2 \setminus (A'_3 \cup B'_3)$ are chosen with uniform probability using the local

randomness source M.

The encoder then computes $X_i^n = S_i^n G_n$ for $i = 1, \dots, k$ and sends it over the channel. We use an extra $(k+1)$-th block to send a version of $S_{(k)}[A_3]$ encoded with a good channel code as in Section C.3.1.

Decoding Similarly to the decoding algorithm for strong coordination of the previous section, the decoder proceeds in reverse order. Here, we detail the procedure in Algorithm 8.

Algorithm 8: Decoding algorithm at Node 2

Input : $(Y_{(1)}^n, \dots, Y_{(k)}^n)$, $S_{(k)}[A_3] \cup Z_{(k)}[B_3]$ and $C = (C_1, C_2, K_1, K_2)$ common randomness shared with Node 1

Output: $(\hat{S}_{(1)}^n, \dots, \hat{S}_{(k)}^n)$, $(\hat{Z}_{(1)}^n, \dots, \hat{Z}_{(k)}^n)$

for $i = k, \dots, 1$ **do**

$$\hat{S}_{(i)}[A_1] \longleftarrow C_1, \quad \hat{Z}_{(i)}[B_1] \longleftarrow C_2$$

if $i \neq k$ **then**

$$\hat{S}_{(i)}[A_3] \leftarrow \hat{S}_{(i+1)}[A_3'] \oplus K_1, \quad \hat{Z}_{(i)}[B_3] \leftarrow \hat{S}_{(i+1)}[B_3'] \oplus K_2$$

for $j \in A_2 \cup A_4$ **do**

Successively draw the bits according to

$$\hat{S}_{(i),j} = \begin{cases} 0 & \text{if } L_n(Y_{(i)}^n, \hat{S}_{(i)}^{j-1}) \geq 1 \\ 1 & \text{else} \end{cases}$$

$$\text{where} \quad L_n(Y_{(i)}^n, \hat{S}_{(i)}^{j-1}) = \frac{\bar{P}_{S_j|S^{j-1}Y^n}(0|\hat{S}_{(i)}^{j-1} Y_{(i)}^n)}{\bar{P}_{S_j|S^{j-1}Y^n}(1|\hat{S}_{(i)}^{j-1} Y_{(i)}^n)}$$

for $j \in B_4$ **do**

Successively draw the bits according to

$$\hat{Z}_{(i),j} = \begin{cases} 0 & \text{if } L_n(X_{(i+1)}^n, \hat{Z}_{(i)}^{j-1}) \geq 1 \\ 1 & \text{else} \end{cases}$$

Rate of common randomness The rate of common randomness is negligible since:

$$\lim_{\substack{n\to\infty \\ k\to\infty}} \frac{|A_1 \cup A_3| + |B_1 \cup B_3|}{kn} = \lim_{\substack{n\to\infty \\ k\to\infty}} \frac{|\mathcal{H}_{X|Y}| + |\mathcal{V}_{W|X}|}{kn}$$
$$= \lim_{k\to\infty} \frac{H(X|Y) + H(W|X)}{k} = 0.$$

Achievability proof We omit the proof since it follows the same steps as the achievability proof for the non-causal encoder, but we use (6.20) instead of (5.7). □

BIBLIOGRAPHY

[1] E. Arıkan. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, 55(7):3051–3073, 2009.

[2] E. Arıkan. Source polarization. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pages 899–903, 2010.

[3] R. J. Aumann, M. Maschler, and R. E. Stearns. *Repeated games with incomplete information*. MIT press, 1995.

[4] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Transactions on Information Theory*, 48(10):2637–2655, 2002.

[5] R. Blasco-Serrano, R. Thobaben, and M. Skoglund. Polar codes for coordination in cascade networks. In *Proc. of International Zurich Seminar on Communications*, pages 55–58, 2012.

[6] M. R. Bloch. Covert communication over noisy channels: A resolvability perspective. *IEEE Transactions on Information Theory*, 62(5):2334–2354, 2016.

[7] M. R. Bloch. *Physical-layer security, Information-theoretic and coding mechanisms for security*. in preparation.

[8] M. R. Bloch and J. Barros. *Physical-layer security: from information theory to security engineering*. Cambridge University Press, 2011.

[9] M. R. Bloch and J. Kliewer. Strong coordination over a line network. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pages 2319–2323, 2013.

[10] M. R. Bloch and J. Kliewer. Strong coordination over a three-terminal relay network. In *Information Theory Workshop (ITW), 2014 IEEE*, pages 646–650. IEEE, 2014.

[11] M. R. Bloch, L. Luzzi, and J. Kliewer. Strong coordination with polar codes. In *Proc. of Allerton Conference on Communication, Control and Computing*, pages 565–571, 2012.

[12] R. A. Chou and M. R. Bloch. Polar coding for the broadcast channel with confidential messages: A random binning analogy. *IEEE Transactions on Information Theory*, 62(5):2410–2429, 2016.

[13] R. A. Chou, M. R. Bloch, and J. Kliewer. Empirical and strong coordination via soft covering with polar codes. to appear in *IEEE Transactions on Information Theory*. URL http://arxiv.org/abs/1608.08474.

[14] R. A. Chou, M. R. Bloch, and E. Abbe. Polar coding for secret-key generation. *IEEE Transactions on Information Theory*, 61(11):6213–6237, 2015.

[15] R. A. Chou, M. R. Bloch, and J. Kliewer. Polar coding for empirical and strong coordination via distribution approximation. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pages 1512–1516, 2015.

[16] C. Choudhuri, Y.-H. Kim, and U. Mitra. Capacity-distortion trade-off in channels with state. In *Proc. of Allerton Conference on Communication, Control and Computing*, pages 1311–1318, 2010.

[17] C. Choudhuri, Y.-H. Kim, and U. Mitra. Causal state amplification. In *Proc. of IEEE International Sympo-*

sium on Information Theory (ISIT), pages 2110–2114. IEEE, 2011.

[18] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[19] I. Csiszár. Almost independence and secrecy capacity. *Problems of Information Transmission*, 32(1):48–57, 1996.

[20] I. Csiszár and J. Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[21] P. Cuff. Communication requirements for generating correlated random variables. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pages 1393–1397, 2008.

[22] P. Cuff. *Communication in Networks for Coordinating Behavior*. PhD thesis, Stanford University, 2009.

[23] P. Cuff. Distributed channel synthesis. *IEEE Transactions on Information Theory*, 59(11):7071–7096, 2013.

[24] P. Cuff and C. Schieler. Hybrid codes needed for coordination over the point-to-point channel. In *Proc. of Allerton Conference on Communication, Control and Computing*, pages 235–239, 2011.

[25] P. W. Cuff, H. H. Permuter, and T. M. Cover. Coordination capacity. *IEEE Transactions on Information Theory*, 56(9):4181–4206, 2010.

[26] A. El Gamal and Y. H. Kim. *Network information theory*. Cambridge University Press, 2011.

[27] R. M. Fano. *Transmission of Information: A Statistical Theory of Communications*. The M.I.T. Press and John Wiley & Sons, Inc., 1961.

[28] O. Gossner and T. Tomala. Repeated games with complete information. In *Encyclopedia of complexity and systems science*, pages 7616–7630. Springer, 2009.

[29] O. Gossner, P. Hernandez, and A. Neyman. Optimal use of communication resources. *Econometrica*, pages 1603–1636, 2006.

[30] F. Haddadpour, M. H. Yassaee, A. Gohari, and M. R. Aref. Coordination via a relay. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pages 3048–3052, 2012.

[31] F. Haddadpour, M. H. Yassaee, S. Beigi, A. Gohari, and M. R. Aref. Simulation of a channel with another channel. *IEEE Transactions on Information Theory*, 63(5):2659–2677, 2017.

[32] S. H. Hassani and R. Urbanke. Universal polar codes. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pages 1451–1455, 2014.

[33] J. Honda and H. Yamamoto. Polar coding without alphabet extension for asymmetric models. *IEEE Transactions on Information Theory*, 59(12):7829–7838, 2013.

[34] S. B. Korada and R. L. Urbanke. Polar codes are optimal for lossy source coding. *IEEE Transactions on Information Theory*, 56(4):1751–1768, 2010.

[35] G. Kramer and S. A. Savari. Quantum data compression with commuting density operators. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, page 44. IEEE, 2002.

[36] G. Kramer and S. A. Savari. Communicating probability distributions. *IEEE Transactions on Information Theory*, 53(2):518–525, 2007.

[37] A. Lapidoth and M. Wigger. Conditional and relevant common information. In *Proc. of IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, pages 1–5, 2016.

[38] B. Larrousse and S. Lasaulce. Coded power control: Performance analysis. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pages 3040–3044, 2013.

[39] B. Larrousse, S. Lasaulce, and M. Bloch. Coordination in distributed networks via coded actions with application to power control. 2015. URL http://arxiv.org/abs/1501.03685.

[40] B. Larrousse, S. Lasaulce, and M. Wigger. Coordinating partially-informed agents over state-dependent

networks. In *Proc. of IEEE Information Theory Workshop (ITW)*, pages 1–5, 2015.

[41] B. Larrousse, S. Lasaulce, and M. Wigger. Coordination in state-dependent distributed networks: The two-agent case. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pages 979–983. IEEE, 2015.

[42] M. Le Treust. Correlation between channel state and information source with empirical coordination constraint. In *Proc. of IEEE Information Theory Workshop (ITW)*, pages 272–276, 2014.

[43] M. Le Treust. Empirical coordination with channel feedback and strictly causal or causal encoding. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pages 471–475. IEEE, 2015.

[44] M. Le Treust. Empirical coordination with two-sided state information and correlated source and state. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pages 466–470, 2015.

[45] M. Le Treust. Coding theorems for empirical coordination. Technical report, 2015. URL https://cloud.ensea.fr/index.php/s/X9e5x8EzJfI7I4Q.

[46] M. Le Treust. Joint empirical coordination of source and channel. *IEEE Transactions on Information Theory*, 63(8):5087–5114, 2017.

[47] M. Le Treust and M. R. Bloch. State leakage and coordination of actions: Core of the receiver's knowledge. 2018. URL http://arxiv.org/abs/1812.07026.

[48] M. Le Treust and T. Tomala. Information design for strategic coordination of autonomous devices with non-aligned utilities. In *Proc. of Allerton Conference on Communication, Control and Computing*, pages 233–242, 2016.

[49] M. Le Treust and T. Tomala. Persuasion with limited communication capacity. 2017. URL http://arxiv.org/abs/1711.04474.

[50] M. Le Treust and T. Tomala. Information-theoretic limits of strategic communication. 2018. URL http://arxiv.org/abs/1807.05147.

[51] M. Le Treust and T. Tomala. Strategic coordination with state information at the decoder. In *2018 International Zurich Seminar on Information and Communication*, 2018.

[52] E. L. Lehmann and J. P. Romano. *Testing statistical hypotheses*. Springer Science & Business Media, 2006.

[53] T. Lindvall. *Lectures on the Coupling Method*. John Wiley & Sons, Inc., 1992. Reprint: Dover paperback edition, 2002.

[54] J. Mertens, S. Sorin, and S. Amir. *Repeated Games*. Cambridge University Press, 2015.

[55] M. Mondelli, S. H. Hassani, I. Sason, and R. Urbanke. Achieving Marton's region for broadcast channels using polar codes. *IEEE Transactions on Information Theory*, 61(2):783–800, 2015.

[56] S. A. Obead, J. Kliewer, and B. N. Vellambi. Joint coordination-channel coding for strong coordination over noisy channels based on polar codes. In *Proc. of Allerton Conference on Communication, Control and Computing*, 2017.

[57] S. A. Obead, B. N. Vellambi, and J. Kliewer. Strong coordination over noisy channels. 2018. URL http://arxiv.org/abs/1808.05475.

[58] A. J. Pierrot and M. R. Bloch. Joint channel intrinsic randomness and channel resolvability. In *Proc. of IEEE Information Theory Workshop (ITW)*, pages 1–5, 2013.

[59] M. S. Pinsker. *Information and Information Stability of Random Variables and Random Processes*. Holden-Day, 1964, originally published in Russian in 1960.

[60] E. Şaşoğlu. Polar codes for discrete alphabets. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, pages 2137–2141, 2012.

[61] E. Şaşoğlu. Polarization and polar codes. *Foundations and Trends® in Communications and Information*

Theory, 8(4):259–381, 2012.

[62] S. Satpathy and P. Cuff. Secure cascade channel synthesis. *IEEE Transactions on Information Theory*, 62 (11):6081–6094, 2016.

[63] C. E. Shannon. Communication theory of secrecy systems. *Bell system technical journal*, 28(4):656–715, 1949.

[64] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on information Theory*, 19(4):471–480, 1973.

[65] E. Soljanin. Compressing quantum mixed-state sources by sending classical information. *IEEE Transactions on Information Theory*, 4(8):2263–2275, 2002.

[66] B. N. Vellambi, J. Kliewer, and M. R. Bloch. Strong coordination over multi-hop line networks. In *Proc. of IEEE Information Theory Workshop-Fall (ITW)*, pages 192–196, 2015.

[67] B. N. Vellambi, J. Kliewer, and M. R. Bloch. Strong coordination over a line when actions are markovian. In *Proc. of Annual Conference on Information Science and Systems (CISS)*, pages 412–417, 2016.

[68] A. Winter. Compression of sources of probability distributions and density operators. 2002. URL http://arxiv.org/abs/quant-ph/0208131.

[69] A. Wyner. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179, 1975.

[70] M. H. Yassaee, M. R. Aref, and A. Gohari. Achievability proof via output statistics of random binning. *IEEE Transactions on Information Theory*, 60(11):6760–6786, 2014.

[71] M. H. Yassaee, A. Gohari, and M. R. Aref. Channel simulation via interactive communications. *IEEE Transactions on Information Theory*, 61(6):2964–2982, 2015.

INDEX

Coordination of autonomous devices over noisy channels: capacity results and coding techniques

Abstract : 5G networks will be characterized by machine to machine communication and the Internet of Things, a unified network of connected objects. In this context, communicating devices are autonomous decision-makers that cooperate, coordinate their actions, and reconfigure dynamically according to changes in the environment. To do this, it is essential to develop effective techniques for coordinating the actions of the nodes in the network.
Information theory allows us to study the long-term behavior of the devices through the analysis of the joint probability distribution of their actions. In particular, we are interested in *strong coordination*, which requires the joint distribution of sequences of actions to converge to an i.i.d. target distribution in L^1 distance.
We consider a two-node network comprised of an information source and a noisy channel, and we require the coordination of the signals at the input and at the output of the channel with the source and the reconstruction. We assume that the encoder and decoder share a common source of randomness and we introduce a state capturing the effect of the environment.
The first objective of this work is to characterize the strong coordination region, i.e. the set of achievable joint behaviors and the required minimal rates of common randomness. We prove inner and outer bounds for this region. Then, we characterize the exact coordination region in three particular cases: when the channel is perfect, when the decoder is lossless and when the random variables of the channel are separated from the random variables of the source. The study of the latter case allows us to show that the joint source-channel separation principle does not hold for strong coordination. Moreover, we prove that strong coordination offers "free" security guarantees at the physical layer.
The second objective of this work is to develop practical codes for coordination: by exploiting the technique of source polarization, we design an explicit coding scheme for coordination, providing a constructive alternative to random coding proofs.

Coordination d'appareils autonomes sur canaux bruités: régions de capacité et algorithmes de codage

Résumé : Les réseaux de 5ème génération se caractérisent par la communication directe entre machines (M2M) et l'Internet des Objets, un réseau unifié d'objets connectés. Dans ce contexte, les appareils communicants sont des décideurs autonomes qui coopérent, coordonnent leurs actions et se reconfigurent de manière dynamique en fonction de leur environnement. L'enjeu est de développer des algorithmes efficaces pour coordonner les actions des appareils autonomes constituant le réseau.
La théorie de l'information nous permet d'étudier le comportement à long-terme des appareils grâce aux distributions de probabilité conjointes. En particulier, nous sommes intéressés par la *coordination forte*, qui exige que la distribution induite sur les suites d'actions converge en distance L^1 vers une distribution i.i.d. cible.
Nous considérons un modèle point-à-point composé d'une source d'information, d'un encodeur, d'un canal bruité, d'un décodeur, d'une information commune et nous cherchons à coordonner les signaux en entrée et en sortie du canal avec la source et sa reconstruction.
Nos premiers résultats sont des bornes intérieures et extérieures pour la région de coordination forte, c'est-à-dire l'ensemble des distributions de probabilité conjointes réalisables et la quantité d'information commune requise. Ensuite, nous caractérisons cette région de coordination forte dans trois cas particuliers: lorsque le canal est parfait, lorsque le décodeur est sans perte et lorsque les variables aléatoires du canal sont indépendantes des variables aléatoires de la source. L'étude de ce dernier cas nous permet de remettre en cause le principe de séparation source-canal pour la coordination forte. Nous démontrons également que la coordination forte offre "gratuitement" des garanties de sécurité au niveau de la couche physique.
Par ailleurs, nous étudions la coordination sous l'angle du codage polaire afin de développer des algorithmes de codage implémentables. Nous appliquons la polarisation de la source de manière à créer un schéma de codage explicite qui offre une alternative constructive aux preuves de codage aléatoires.