Fall 2016

Source-channel coding for coordination over a noisy two-node network

Sarah A. Obead
New Jersey Institute of Technology

ABSTRACT

SOURCE-CHANNEL CODING FOR COORDINATION OVER A NOISY TWO-NODE NETWORK

by
Sarah A. Obead

Recently, the concept of coordinating actions between distributed agents has emerged in the information theory literature. It was first introduced by Cuff in 2008 for the point-to-point case of coordination. However, Cuff's work and the vast majority of the follow-up research are based on establishing coordination over noise-free communication links. In contrast, this thesis investigates the open problem of coordination over noisy point-to-point links. The aim of this study is to examine Shannon's source-channel separation theorem in the context of coordination. To that end, a general joint scheme to achieve the strong notion of coordination over a discrete memoryless channel is introduced. The strong coordination notion requires that the L_1 distance between the induced joint distribution of action sequences selected by the nodes and a prescribed joint distribution vanishes exponentially fast with the sequence block length. From the general joint scheme, three special cases are constructed, one of which resembles Shannon's separation scheme. As a surprising result, the proposed joint scheme has been found to be able to perform better than a strictly separate scheme. Finally, the last part of the thesis provides simulation results to confirm the presented argument based on comparing the achievable rate regions for the scheme resembling Shannon's separation and a special case of the general joint scheme.

SOURCE-CHANNEL CODING
FOR COORDINATION OVER A NOISY TWO-NODE NETWORK

by
Sarah A. Obead

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Telecommunications

Helen and John C. Hartmann Department of
Electrical and Computer Engineering

January 2017

Blank Page

APPROVAL PAGE

SOURCE-CHANNEL CODING FOR COORDINATION OVER A NOISY TWO-NODE NETWORK

Sarah A. Obead

Dr. Joerg Kliewer, Thesis Advisor Date
Associate Professor of Electrical and Computer Engineering, NJIT

Dr. Ali Abdi, Committee Member Date
Professor of Electrical and Computer Engineering, NJIT

Dr. Osvaldo Simeone, Committee Member Date
Professor of Electrical and Computer Engineering, NJIT

BIOGRAPHICAL SKETCH

Author: Sarah A. Obead

Degree: Master of Science

Date: January 2017

Undergraduate and Graduate Education:

- Master of Science in Telecommunications,
  New Jersey Institute of Technology, Newark, NJ, 2017
- Bachelor of Science in Information Technology,
  University of Benghazi, Benghazi, Libya, 2010

Major: Telecommunications

بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ
(وَ قُلِ اعْمَلُوا فَسَيَرَى ٱللَّهُ عَمَلَكُمْ وَ رَسُولُهُ وَ ٱلْمُؤْمِنُونَ)
صَدَقَ ٱللَّهُ ٱلْعَظِيْمُ

أهدي هذه الرسالة إلى أمي الغالية و أبي الحبيب
وإلى أسرتي جميعاً
ثم إلى اساتذتي الأفاضل و كل من علمني حرفاً

It is with my genuine gratefulness and warmest regard
that I dedicate this work to my beloved parents

ACKNOWLEDGMENT

I would like to first thank my thesis advisor Dr. Joerg Kliewer for his generous support. His guidance helped me in all the time throughout the course of my studies at NJIT. More special thanks to Dr. Badri Vellambi and Dr. Roberto Rojas-Cessa for their kind advise, influence and support. I would like to thank Dr. Osvaldo Simeone and Dr. Ali Abdi, who graciously agreed to serve on my committee.

I must express my very profound gratitude to my parents, Zakia Shoiab and Dr. Ali Obead, and to my siblings for providing me with unfailing support to pursue my aspirations not only through the process of researching and writing this thesis but throughout all my years of study. This accomplishment would not have been possible without them. Thank you.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

Communication and coordination are two essential concepts supporting natural and man-made multi-agent infrastructures. The numerous applications of distributed and event-driven multi-agent platforms have motivated researchers to carry on detailed analysis on many coordinated behaviors. Examples of coordinated behaviors rise in a wide range of infrastructures, from biological-based case like ant colonies to networks of mobile robots. Theoretic and applied studies on multi-agents coordination target several open questions regarding how agents exchange information and how their actions can be correlated to achieve a desired overall behavior. Models to analyze and implement the collective behavior of multi-agent systems, especially those that operate without any central control, are getting a lot of attention.

For instance, [1] studied how harvester ants coordinate their actions and assign workers to different tasks while searching for scattered seeds. The work was motivated by the question of how the actions taken by individual ants add up to form a meaningful behavior without any central planning or decision making. That kind of collective behavior is what makes it possible for ant colonies to tackle survival challenges, reallocate and evolve. This study also sheds a light on fascinating similarities between the coordination behavior of ant colonies and how the bandwidth and the traffic are regulated throughout the Internet. In particular, [1] reported how ants returning from foraging interact with outgoing forgers using brief antennal contact. Based on this interaction, the rate at which outgoing foragers leave the nest changes with the availability of food. If there is plenty of food, more ants are recruited and vice versa. The similarity with communicating over the Internet lies within how the transport layer protocol (TCP) regulates data transmission based on the rate of the receiving acknowledgment messages. If the rate is low, that means there is a scarcity in the bandwidth and thus the data transmission is reduced.

There are two approaches for agents to exchange information in order to coordinate their actions, implicitly and explicitly. Implicit coordination is based on observation or sensing. Examples of implicit coordination from nature include the flocking behavior of birds and how a groups of fish swim in the same direction and redirect simultaneously (i.e., schooling of fish). In both examples, individuals sense variation in air or water current to indicate a change in the direction of the group movement, and then synchronize their movement accordingly. On the other hand, explicit coordination is communication-based. The problem of communication-based coordination of multi-agent systems arises in applications including mobile robot networks and distributed computing platforms such as distributed games, grid computing, and smart traffic control networks [2]. Throughout our discussion, we focus on communication-based coordination.

1.1 Literature Review

To answer the question of how agents can correlate their actions to achieve a desired overall behavior via communication, Cuff in [3] presented the elements of a theory of coordination in networks. In a coordination network model, the collective behavior is described by a joint probability distribution (i.e., a prescribed target distribution) of the action of the agents, and agents are modeled as the nodes in the network. Moreover, for a target joint behavior, the theory classifies coordination into empirical and strong coordination based on different performance metrics. These two metrics determine how well the target distribution is approximated by the coordination network scheme. The empirical coordination criterion requires the histogram of the joint behavior to be asymptotically close to the target distribution. On the other hand, strong coordination has a more strict criterion where the joint distribution must be statistically indistinguishable from the target distribution. Nevertheless, the notion of strong coordination was first studied extensively by Soljanin [4] in the context of quantum compression with unlimited common randomness.

The information theory community has recently devoted much effort to study strong and empirical coordination. As a result, the capacity regions of several classes of both empirical and strong coordination networks have been established [2–7]. The capacity region characterizes the achievable joint distributions and the limits of the communication rates required to establish such joint behavior among the nodes in the networks. Based on the results of [2,3], several extensions and applications have been made. Based on a generalization on [3], new lower and upper bounds for the capacity region of the two-node network as established in [8]. These new bound were generated for strong coordination criterion under the assumption that the nodes communicate in a bidirectional fashion in order to achieve coordination. A smiler framework was adopted and improved in [9].

In [5,10], the authors addressed inner and outer bounds for the capacity region of a three-terminal network with the presence of a relay node. On one hand, [5] focused on the strong coordination network setup when the communication is in one direction and the actions are generated in all nodes. On the other hand, [10] developed inner and outer bounds for the capacity region when the communication is bidirectional. Finally, an inner and outer bound for the capacity region of the three-terminal network setup was studied with respect to empirical coordination in [7]. The work of [5] was later extended in [6,11] to a precise characterization of the strong coordination region for a line network and a multi-hop line network, respectively. [12] also characterized the strong coordination region of multi-hop line network but in the context of secure communication.

Other extensions have focused on developing practical coding schemes based on polar codes. First, based on results from rate-distortion coding with polar codes, [13] presented codes for empirical coordination in cascade network setup. [14] studied the design of polar codes for strong coordination in two-node networks, in which the input action is uniformly distributed over a binary alphabet and the target joint distribution of actions is doubly binary symmetric. Later, the coding scheme of [14] was expanded to include empirical and strong coordination in [15]. However, the later coding scheme achieves the capacity region for doubly non-binary actions in which the input alphabet takes value of prime cardinality.

Most of the efforts mentioned above in achieving strong coordination, is heavily centered around the channel resolvability theorem [16]. The channel resolvability theorem approximates a channel output statistic using a codebook associated with a finite set of uniformly distributed messages. Besides the problem of strong coordination, this theorem is used to study several information theoretic problems including channel synthesis [17] and strong secrecy over wiretap channels [18]. Nevertheless, several exceptions are made in [9], [10], and [19][1] where the achievability proof is based on a random binning technique [20].

1.2 Motivation

Coordination via realistic communication links implies many inevitable physical constraints such as limited bandwidth and the existence of noise. As a result, this motivates questions of significant interest: *How do the noisy communication links affect the ability to achieve coordination? Does communication over such links provide any additional restrictions or benefits?* To gain insight into more practical networks, it make sense to consider coordination networks in which the communication links

[1]During the course of writing this thesis, we discovered an independent work related to this thesis in [19]. Although [19] focuses on simulating a channel using another channel, the notion of strong coordination is applied in a similar way as in this thesis. Moreover, the proposed scheme is also close to the general scheme we propose in Chapter 4. However, there is a significant difference in terms of codebook construction, the network framework, the proof methodology, and the consideration of common and local randomness which highlight our contribution.

are subject to noise. However, the majority of the recent works on coordination have considered noise-free communication channels. Coordination over noisy channels seems to be little explored. Cuff [2] pointed out that the problem of coordination in networks can be recognized as a network source coding problem that branches, to some extend, from traditional source coding. Specifically, the source-channel separation theorem, that was established by Shannon [21], states that the general point-two-point communication system can be broken into two parts. That is, we can separately consider the problem of efficient data compression (i.e., source coding) and the problem of reliable data transmission (i.e., channel coding) and then pair the optimal rates obtained by the two. Shannon showed that, by considering joint coding, benefit in terms of minimum achievable rates is not obtained. As a main contribution of this thesis, we derive achievability results for the two-node coordination network and determine if Shannon's separation theorem also holds in the context of strong coordination over noisy channels.

To provide a suitable background for rest of the thesis, in the following chapter we briefly present the most fundamental tools and quantities of the information theory. The thesis then proceeds as follows: The nature of both empirical and strong coordination problems and the capacity region obtained by coordination over noiseless links are presented in Chapter 3. We then derive achievability results for the two-node coordination network with a noisy communication link in Chapter 4. In Chapter 5 we present a motivating example and simulation results for separate and joint strong coordination schemes to simulating a target distribution given by a doubly binary symmetric source. Finally, Chapter 6 concludes the thesis and suggests future work.

CHAPTER 2

INFORMATION THEORETIC BASICS

In this chapter, we first present the employed notation in the following. We then introduce most of the basic information theoretical entities and tools used during the development of this thesis.

2.1 Notation

Throughout the thesis, we denote a discrete random variable with an upper-case letter (e.g., X) and its realization with the lower case letter (e.g., x). The alphabet size of the random variable X is denoted as $|\mathcal{X}|$. We use X^n to denote the finite sequence $(X_1, X_2, ..., X_n)$. In addition, all logs are taken with respect to base 2, we denote the indicator function by $\mathbf{1}(w)$, which is equal to 1 if the predicate w is true and 0 otherwise, and the counting function $N(w|w^n) = \sum_{i=1}^{n} \mathbf{1}(w_i = w)$.

The probability notation used is as follows: $Pr[A]$ is the probability that the event A occurs; $Pr[A|B]$ is the probability of the event A conditioned on the event B. The probability mass function (pmf) of the discrete random variable X is denoted as $P_X(x)$. However, we sometime use the lower case notation (e.g., $p_X(x)$) to distinguish target pmfs or alternative definitions. The conditional probability distribution of a random variable X given the random variable Y is $P_{X|Y}(x|y)$. Finally, the expectation of the random variable X and conditional expectation given Y are denoted with $E[X]$ and $E[X|Y]$, respectively.

2.2 Basic Information Measures

Now we present the definitions of the information theoretical entities.

Definition 1 (Entropy [22])**.** For a given random variable X drawn from the alphabet $\mathcal{X}$, the entropy measures the amount of information required on average to describe the random variable. It is a function of the distribution of a random variable, measured in bits, and defined as follows:

$$H(X) = -\sum_{x \in \mathcal{X}} P_X(x) \log P_X(x).$$

Given the definition of entropy, we have the following entities:

- Conditional entropy [22]: the entropy of random variable X conditioned on the random variable Y, with alphabet $\mathcal{Y}$, is given by

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

- Binary entropy [22]: the entropy of a random variable that takes on one of two values p or $(1-p)$. We use a lower-case notation to define the binary entropy function as follows:

$$H(p) = -\big[p \log p + (1-p) \log(1-p)\big].$$

Definition 2 (Relative entropy [22])**.** We let $\mathbb{D}(\cdot||\cdot)$ denote the Kullback-Leibler divergence between two distributions $P_X(x)$ and $Q_X(x)$ defined over the sample space $\mathcal{X}$. The Kullback-Leibler divergence measures the distance between the two distributions. In the information theory context it is known as relative entropy [22] and is defined as follows:

$$\mathbb{D}(P_X(x)||Q_X(x)) = \sum_{x \in \mathcal{X}} P_X(x) \log \frac{P_X(x)}{Q_X(x)}.$$

Definition 3 (Mutual information [22])**.** Mutual information is a special case of the Kullback-Leibler divergence between two distributions. It measures the distance between the joint distribution of two random variables and the product of the marginals. Given two random variables X and Y with a joint pmf $P_{XY}(x, y)$ and marginal pmfs $P_X(x)$, $P_Y(y)$ the mutual information is defined as follows:

$$I(X;Y) = \mathbb{D}(P_{XY}(x,y)||P_X(x)P_Y(y)) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{XY}(x,y)\log\frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}.$$

It is also defined in terms of entropy as

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

Following the definition of mutual information we have an extension to conditional mutual information.

- Conditional mutual information [22]: The mutual information between the random variables X and Y conditioned on Z is given by

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

Definition 4 (Total variation)**.** The total variation between two pmfs $P_X(x)$ and $Q_X(x)$ over the same sample space $\mathcal{X}$ is defined by the L_1 distance between the two pmfs as

$$||P_X(x) - Q_X(x)||_{TV} = \frac{1}{2}\sum_{x\in\mathcal{X}} |P_X(x) - Q_X(x)|.$$

Definition 5 (Pinsker's inequality [23])**.** Pinsker's inequality provides an upper bound for the total variation between two pmfs, $P_X(x)$ and $Q_X(x)$, based on the Kullback-Leibler divergence. The bound is defined as follows:

$$||P_X(x) - Q_X(x)||_{TV} \leqslant \sqrt{2\mathbb{D}(P_X(x)||Q_X(x))}$$

Definition 6 (Strongly typical set [24])**.** Consider a pair of sequences (x^n, y^n) drawn according to the joint distribution $P_{XY}(x,y)$ with the marginal pmfs $P_X(x)$ and $P_Y(y)$.

The strongly jointly typical set $\mathcal{T}_\epsilon^{(n)}(P_{XY})$ is defined as

$$\mathcal{T}_\epsilon^{(n)}(P_{XY}) = \left\{ \begin{array}{ll} (x^n, y^n): & \left| \frac{1}{n} N(x|x^n) - P_X(x) \right| < \epsilon, \\ & \left| \frac{1}{n} N(y|y^n) - P_Y(y) \right| < \epsilon, \\ & \left| \frac{1}{n} N(x,y|x^n,y^n) - P_{X,Y}(x,y) \right| < \epsilon \end{array} \right\}, \tag{2.1}$$

for all $x \in \mathcal{X}$ with $P_X(x) > 0$, $y \in \mathcal{Y}$ with $P_Y(y) > 0$, $(x,y) \in \mathcal{X} \times \mathcal{Y}$ with $P_{X,Y}(x,y) > 0$ and for any $\epsilon > 0$.

2.3 Rate-Distortion

The problem of efficiently representing an information source (i.e., data compression) is known as the source coding problem. It was established in Shannon's earlier work [25] that for a finite block length n, data compression is achievable with a fidelity criterion using jointly-typical encoding.

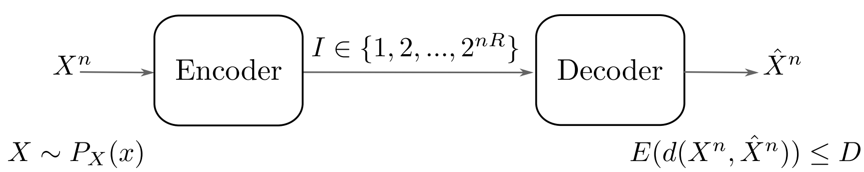

Figure 2.1 Rate distortion encoder and decoder.

A jointly-typical encoder assigns a short description $I \in \{1, 2, ..., 2^{nR}\}$ to the most frequent (i.e., typical) sequences of the source X^n. The decoder, on the other hand, constructs a sequence $\hat{X}^n$ from the description. The fidelity criterion is described as the distortion between the original source sequence X^n and the sequence reconstructed from the description $\hat{X}^n(I)$. The distortion is measured with a function $d(\hat{x}, x)$. Theorem 1 outlines the rate of the description messages required to efficiently represent a source within an average distortion arbitrary close to D. This process is depicted in Figure 2.1 where D is given by

$$D = E(d(X^n, \hat{X}^n)) = \frac{1}{n}\sum_{i=1}^{n} E(d(X_i, \hat{X}_i)).$$

Theorem 1 (Rate-distortion theorem [24])**.** *For a discrete memoryless source X that generates sequence $(X_1, X_2, ..., X_n)$ drawn i.i.d. according to $P_X(x)$ and a distortion measure $d(x, \hat{x})$, the rate distortion function is given by*

$$R(D) = \min_{P_{\hat{X}|X}(\hat{x}|x): E[d(X,\hat{X})] \leq D} I(X; \hat{X}). \tag{2.2}$$

The rate-distortion function provide the minimum rate to describe a source (i.e., fundamental limit of data compression) while maintaining a prescribed level of distortion. Lossless source coding is a special case of rate-distortion theory where the reconstruction of the source sequence is done asymptotically error free (i.e., $D = 0$). The achievability of rate-distortion is proved in [24, Theorem 3.5] using jointly typical encoding based on the strongly typical set given in (2.1). The rate-distortion theorem is closely related to the criterion of empirical coordination. We will introduce this relation in Section 3.1.

2.4 Packing Lemma

The packing lemma [24] studies the probability of decoding error ($\hat{M} \neq M$) in the channel coding problem illustrated in Figure 2.2. The question that the lemma

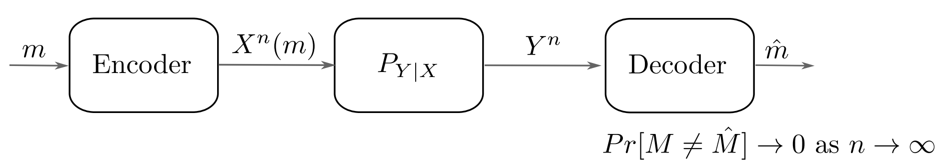

Figure 2.2 Channel coding problem.

investigate is as follows: Given a set of channel codewords $\mathcal{C} = \{X^n(m) \in \mathcal{X}^n :$

$m \in \{2, ..., 2^{nR}\}\}$, a discrete memoryless channel (DMC) $P_{Y|X}$, and a channel output sequence Y^n. What is the probability of observing an output sequence $\tilde{Y}^n$ from sending an independent codeword $X^n(1)$ (i.e., the codeword is not from the set $\mathcal{C}$) such that $(\tilde{Y}^n, X^n(m))$ for any $m \neq 1$ are jointly typical? The intuition behind the packing lemma is depicted in Figure 2.3. The RHS ellipse represents the sample space of the output sequences $\mathcal{Y}^n$. The LHS ellipse represents the sample space of channel codewords. The smaller gray ellipses represent all possible output sequences for a specific codeword $P_{Y|X}(Y^n|x^n)$ which can be decoded. The packing lemma answers the question of the maximum number of codewords $|\mathcal{M}|$ that can be sent over the channel such that the output can be decoded correctly with a probability of decoding error that vanishes asymptotically with the channel use (i.e., as $n \to \infty$).

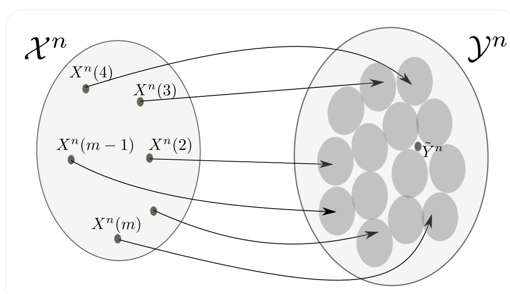

Figure 2.3 Packing lemma of channel coding.

Lemma 2 (Packing lemma [24])**.** *Let (X, Y) be two random variables with the pmf $P_{XY}(x, y) = P_X(x)P_{Y|X}(y|x)$. Let $(\tilde{X}^n, \tilde{Y}^n) \sim P_X(x)P_Y(y)$. Let $X^n(m)$, where $m \in \{1, 2, ..., 2^{nR}\}$, be random sequences. Then, there exists $\delta(\epsilon)$ that tends to zero as $\epsilon \to 0$ with $n \to \infty$ such that*

$$Pr\left[(\tilde{X}^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}(P_{XY})\right] \to 0$$

if

$$R < I(X; Y) - \delta(\epsilon). \tag{2.3}$$

Based on the result of this lemma, the channel capacity [24, Theorem 3.1] is defined as the maximum of all achievable rates for the channel $P_{Y|X}$:

$$C = \max_{p_X(x)} I(X;Y). \tag{2.4}$$

2.5 Source-Channel Separation

Another important information theoretic result, the source-channel separation theorem was also established by Shannon [25]. The theorem combines the rate limits of source coding (2.2) and channel coding (2.4). It states that we can construct an optimal code to represent a source using a finite set of messages, and separately design an optimal channel code to reliably communicate these message while maintaining a desired minimum distortion. Specifically, the theorem argues that the two-stage method of coding depicted in Figure 2.4 (b) is as efficient as any other joint source-channel coding scheme for the problem illustrated in Figure 2.4 (a).

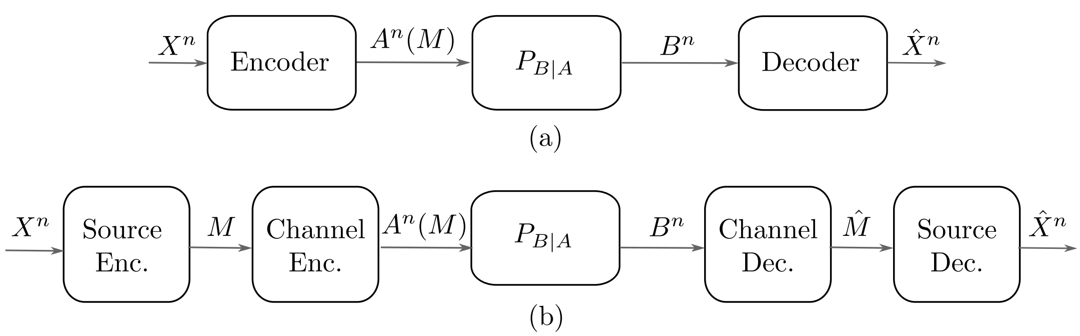

Figure 2.4 Source channel coding: (a) The joint coding problem. (b) The separation between the source coding and the channel coding problems.

Theorem 3 (Source-channel separation theorem [24]). *Given a discrete memoryless source X, a distortion measure $d(x,\hat{x})$ with rate-distortion function $R(D)$, and a DMC $P_{B|A}(b|a)$ with capacity C, the following statement holds:*

$$R(D) < C.$$

That is, we can achieve source compression with any rate $R > R(D)$ satisfying (2.2) while communicating over a noisy channel if and only if $R < C$, thus satisfying (2.4). Note that the setup in Figure 2.4 (b) is translated as

$$R > I(X; \hat{X}), \tag{2.5}$$

$$R < I(A; B). \tag{2.6}$$

This theorem is the starting point of the study conducted in this thesis. As mentioned in Section 1.2, Shannon's source-channel separation has been proved to be optimal for point-to-point communication. The main goal of this work is to verify if Shannon's separation theorem holds in a (strong) coordination context.

2.6 Channel Resolvability

The concept of channel resolvability was first introduced by Wyner [26], then further developed by the resolvability work of Han and Verdú [16]. While the channel coding theorem relies on the packing lemma in Section 2.4, channel resolvability is established based on a soft covering lemma. The question that this lemma investigates is described in Figure 2.5 as follows: Given a set of channel codewords $\mathcal{C} = \{X^n(m) \in \mathcal{X}^n : m \in \{1, ..., 2^{nR}\}\}$ and a discrete memoryless channel (DMC) $P_{Y|X}$ as illustrated in Figure 2.6. What is the minimum rate of codewords that yield jointly typical output sequences which cover the entire sample space $\mathcal{Y}^n$? Lemma 4 states the formulation of the soft-covering lemma.

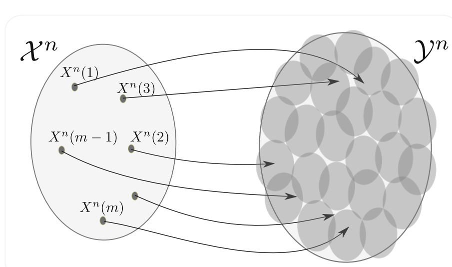

Figure 2.5 Soft covering lemma for channel resolvability.

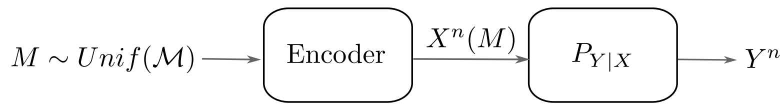

Figure 2.6 Channel resolvability.

Lemma 4 (Soft covering lemma [17])**.** *Let X be a discrete random variable with the pmf $P_X(x)$ and $\mathcal{C} = \{X^n(m) \in \mathcal{X}^n : m \in \{1, ..., 2^{nR}\}\}$. Let M be uniformly distributed over $\{1, 2, ..., 2^{nR}\}$. Then, the induced output distribution defined as*

$$\hat{P}_{Y^n}(y^n) = 2^{-nR} \sum_{m=1}^{2^{nR}} \prod_{k=1}^{n} P_{Y|X}(y_k|X_k(m))$$

is ϵ-close in total variation to the i.i.d. distribution $P_Y^{\otimes n}(y^n) = \prod_{i=1}^{n} P_Y(y_i)$. That is,

$$\left\| \hat{P}_{Y^n}(y^n) - P_Y^{\otimes n}(y^n) \right\|_{TV} \to 0 \text{ in probability.}$$

if

$$R > I(X;Y). \tag{2.7}$$

The result of the lemma is summarized by the fact that we can approximately generate a desired i.i.d. distribution at the output of the DMC by applying a channel resolvability encoder with a codebook $\mathcal{C}$ associated with a finite set of uniformly distributed messages. The significance of channel resolvability is that it simplifies achievability proofs for solving channel synthesis, strong coordination, and secrecy problems. We will utilize channel resolvability in this thesis to generate the achievability results of strong coordination over noisy communication links.

CHAPTER 3

POINT-TO-POINT COORDINATION

A coordination network as originally characterized by Cuff [3] is modeled by several assumptions. First, the action of an agent is modeled as a discrete random variable with a finite alphabet, and the collective behavior of the agents is modeled by a joint distribution. Second, nodes have access to noise-free communication channels and a source of common randomness. Third, only a subset of the nodes receive their action sequence from nature or an external source. Those nodes initiate the coordination process. Fourth, a node is allowed to observe a length-n sequence of actions from the external source before conveying a coordination message to other nodes. Last, each node in the network selects an action sequence based on the common randomness information and the coordination message that it receives.

The straightforward approach for nodes to coordinate their behavior is to notify other nodes with the selected action sequence. This method of coordination is simply a source coding problem. However, many computerized multi-agent infrastructures such as the mobile robot network have limited resources including CPU, memory size, power, and bandwidth. As a result, the goal is to find a resource-efficient coordination scheme. In particular, we are interested in a coordination scheme that minimizes the communication rate while providing other nodes with enough information to select the appropriate set of actions that jointly result in a desired overall behavior.

A comprehensive review of the behavior of coordination networks is beyond the scope of this study. Here we focus only on the two-node network or point-to-point coordination depicted in Figure. 3.1 as understanding this simple model suffices as a building block for bigger, more complicated networks.

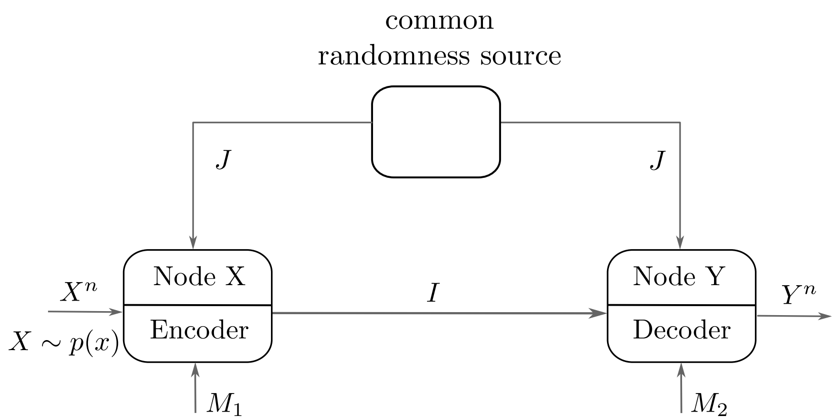

Figure 3.1 Point-to-point coordination network.

As illustrated in Figure 3.1, common randomness J is available to both nodes at a rate of R_o bit/action, that is $J \in \{1, 2, ..., 2^{nR_o}\}$. Node X receives a sequence of actions $X^n = (X_1, X_2, ..., X_n) \in \mathcal{X}^n$ specified by nature. X^n is drawn i.i.d. according to a distribution $p(x)$. Each nodes node possesses local randomness at rate $\rho_k, k = 1, 2$. Node X initiates coordination by communicating a message I to node Y. The message $I \in \{1, 2, ..., 2^{nR}\}$ is generated based on the input action sequence X^n and the common randomness J. Node Y selects its action sequence $Y^n = (Y_1, Y_2, ..., Y_n) \in \mathcal{Y}^n$ according to the received message I and common randomness J. Finally, we assume the common randomness source to be independent of the action specified by nature to X. The local randomness $M_1 \in \{1, 2, ..., 2^{n\rho_1}\}$ is used to generate the message I at node X and the rate of local randomness $M_2 \in \{1, 2, ..., 2^{n\rho_2}\}$ is used to generate the sequence Y^n at node Y.

Next, we outline the performance metrics for coordination [2] that measure how close the resulting joint distribution is to a desired collective behavior.

3.1 Empirical Coordination

As mentioned in the introduction, empirical coordination constrains the induced joint distribution such that the normalized histogram of the joint actions of the two nodes, known as a joint type of actions, is asymptotically equal in total variation to the desired joint behavior $p_{X,Y}(x,y) = p_X(x)p_{Y|X}(y|x)$. That is,

$$\left\| \frac{1}{n} N(x,y|x^n,y^n) - p_{X,Y}(x,y) \right\|_{TV} \to 0 \text{ in probability.}$$

There are two important observations to take into account with respect to the point-to-point empirical coordination. First, common randomness does not play a role in achieving empirical coordination. Second, the rate constraint to achieve empirical coordination relates to lossy compression (i.e., rate-distortion theory) [3].

The intuition behind this statement follows from the nature of the empirical distribution and the definition of the strongly typical set (2.1) (see Definition 6). It can be seen that achieving empirical coordination is a matter of generating sequences Y^n that are strongly typical with X^n (i.e., $(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{XY})$) with high probability. Specifically, the jointly-typical encoder for the rate-distortion setup in Figure 2.1 can be seen as coordinator of a reconstruction sequence Y^n with the source sequence X^n according to a joint distribution $p_X(x)p_{X|Y}(y|x)$. Thus, to achieve such requirement, the capacity region for empirical coordination is given as follows:

$$\underline{\mathcal{C}} = \{R, p(y|x) : R \geq I(X;Y)\}. \tag{3.1}$$

Now, we briefly consider the case of noisy communication link as illustrated in Figure 3.2. This setup differs from the empirical coordination setup demonstrated in Figure 3.1 in the sense that a channel codeword now communicates the message I over a discrete memoryless channel (DMC) $P_{B|A}(b|a)$ to node Y so that it generates a sequence of actions $Y^n = (Y_1, Y_2, ..., Y_n)$.

As mentioned in Section 2.5, the source-channel separation theorem states that we can divide the problem of communicating an information source over a noisy channel into two stages. That is, that we can construct an optimal code to

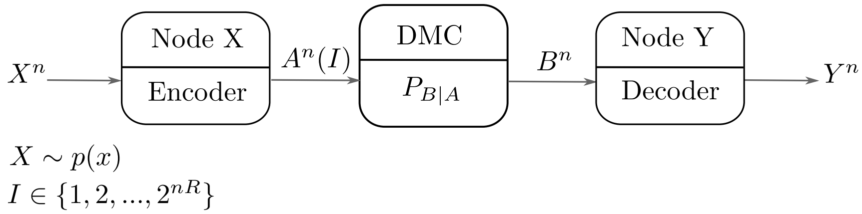

Figure 3.2 Point-to-point empirical coordination over a DMC.

represent a source using a finite set of messages (i.e., lossy source compression), and independently design an optimal channel code to reliably communicate these message while maintaining a desired minimum distortion. The theorem argues that this two-stage coding method is as efficient as any other joint source-channel code. As a result, based on the tight relation between rate-distortion (i.e., lossy source coding) and empirical coordination, we can extend the results of rate-distortion channel code separation to the two-node empirical coordination scheme as outlined in Corollary 4.1.

Corollary 4.1. *The coordination capacity region $\underline{\mathcal{C}}$ for empirical coordination in the two-node network of Figure 4.1 is the set of rate-coordination pairs for which the rate is greater than the mutual information between X and Y and less than the capacity C of the DMC $P_{B|A}(b|a)$. Thus*

$$\underline{\mathcal{C}} \triangleq \left\{ (R, p_{y|x}(y|x) : I(X;Y) < R < C \right\}. \tag{3.2}$$

3.2 Strong Coordination

Strong coordination has a more demanding constraint. Here, the induced joint distribution of time sequence of actions, $\hat{P}_{X^n,Y^n}(x^n, y^n)$ is intended to be statistically indistinguishable in total variation from the desired joint distribution $P_{XY}^{\otimes n}(x^n, y^n) = \prod_{i=1}^{n} p(x_i)p(y_i|x_i)$. That is,

$$\left\| \hat{P}_{X^n,Y^n}(x^n, y^n) - P_{XY}^{\otimes n}(x^n, y^n) \right\|_{TV} \to 0 \text{ in probability.}$$

To achieve such requirement, the capacity region for strong coordination [3] is given as follows:

$$\underline{\mathcal{C}} \triangleq \left\{ \begin{array}{ll} & \exists p(u|x,y) \text{ such that} \\ (R_o, R, p(y|x)) : & |\mathcal{U}| \leq |\mathcal{X}||\mathcal{Y}| + 1, \\ & R \geq I(X;U), \\ & R + R_o \geq I(X,Y;U). \end{array} \right\}. \tag{3.3}$$

The intuition behind this region follows from the channel resolvability theorem in Section 2.6. Consider the setup shown in Figure 2.6. The joint distribution $P_{XY}(x,y)$ is intended to be the output distribution we aim to approximate at the output of the channel. Choose $P_U(u)$ to be an auxiliary random variable to generate the channel resolvability codebook

$$\mathcal{C} = \left\{ U^n(I,J) \in \mathcal{U}^n : I \in \{1, ..., 2^{nR}\}, J \in \{1, ..., 2^{nR_o}\} \right\}.$$

Select $(I,J) \sim \text{Unif}(\mathcal{I} \times \mathcal{J})$. Finally, select the channel in Figure 2.6 to be $P_{XY|U}$ (i.e., the channel has a joint output XY). Then, as a result of Lemma 4, if $R_o + R \geq I(X,Y;U)$

$$\left\| \hat{P}_{X^nY^n}(x^n, y^n) - P_{XY}^{\otimes n}(x^n, y^n) \right\|_{TV} \to 0 \text{ in probability.}$$

According to this result, the channel resolvability theorem provides a feasible way to prove a part of the capacity region of (3.3). However, the structure of the strong coordination problem and the channel resolvability problem is different in terms of the communication direction between node X and the source of common randomness. In the strong coordination problem, node X receives the sequence X^n from nature and generate I according to the message J of common randomness. On the other hand, in the channel resolvability model, node X receives both messages and generates X^n accordingly. The technique to obtain the region is by altering the channel resolvability problem to mimic the behavior of strong coordination structure as depicted in Figure 3.1. To that extend several steps need to be taken as follow:

- Choose the axillary random variable U such that we have the Markov chain [22] $X - U - Y$. As a result $P_{XY|U} = P_{X|U} P_{Y|U}$ and we obtain the *allied* structure in Figure 3.3.

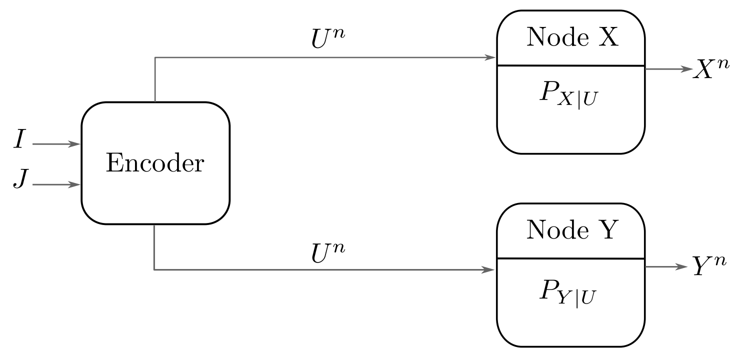

Figure 3.3 Strong coordination allied structure.

- Choose $R > I(X;U)$. This constraint is known as secrecy constraint [18]. It is used to establish close-independence between random variables in a 3-terminal network setup known as wiretap channel [27]. It is needed here to mimic the behavior of the two-node coordination network where the common randomness is independent of the action provided to node X from nature. The condition $R > I(X;U)$ reinforce this constraint in the model of Figure 3.3 and constitutes the second condition in the capacity region (3.3).

In this chapter, we have presented two performance metric for the coordination networks. Moreover, have examined source-channel separation for empirical coordination. In the following chapter, we will discuss achievable schemes for strong coordination over noisy communication links by leveraging the channel resolvability theorem in a similar manner.

CHAPTER 4

STRONG COORDINATION OVER A DMC

The complexity of the strong coordination problem under the total variation constraint mentioned in Section 3.2 makes it difficult to solve as is. However, it has been previously approached by adopting a widely used problem-solving strategy. The strategy starts by transforming the strong coordination problem into another intermediate problem for which a solution exist. Then by applying the solver for the intermediate problem we can recover a solution for the strong coordination problem with some approximation guarantees. In this chapter, we present the problem of strong coordination over a DMC, the corresponding intermediate problem which we will refer to as allied problem, and the approximation steps for the solution. Finally, based on our solution, we characterize joint and separate coordination-channel coding schemes in order to examine code separation for the two-node strong coordination network.

4.1 Setup Under Consideration

In the two-node coordination network depicted in Figure 4.1, we have node X that receives a sequence of actions $X^n = (X_1, X_2, ..., X_n) \in \mathcal{X}^n$ specified by nature. X^n is i.i.d. according to a distribution $p_X(x)$. Both nodes have access to randomness J at rate R_o bit/action from a common source and each node possess local randomness at rate $\rho_k, k = 1, 2$. We want to communicate a message I over the rate-limited DMC $P_{B|A}(b|a)$ to node Y. The message I is constructed based on the input action sequence X^n and the common randomness J. Node Y generates a sequence of actions $Y^n = (Y_1, Y_2, ..., Y_n) \in \mathcal{Y}^n$ based on the received message I and the common randomness J. The goal is to have the induced joint distribution of the time-sequence of actions $\hat{P}_{X^n,Y^n}(x^n, y^n)$ arbitrary close in total variation to the desired distribution

$P_{XY}^{\otimes n}(x^n, y^n) = \prod_{i=1}^{n} p(x_i)p(y_i|x_i)$. That is,

$$\left\| \hat{P}_{X^n,Y^n}(x^n, y^n) - P_{XY}^{\otimes n}(x^n, y^n) \right\|_{TV} \to 0 \text{ in probability.}$$

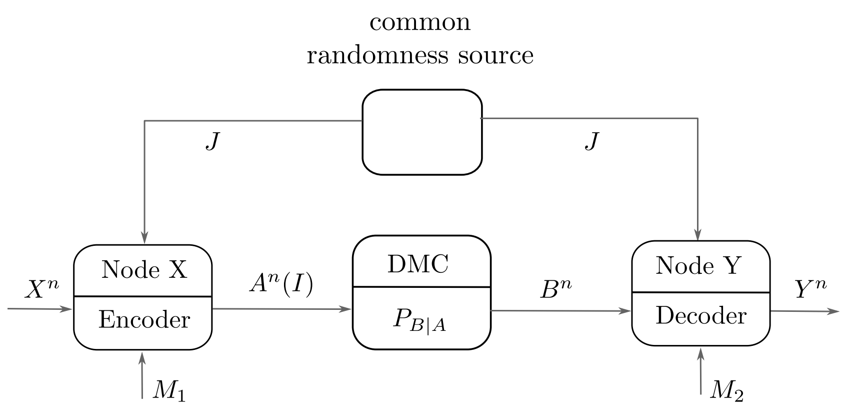

Figure 4.1 *Point-to-point strong coordination over a DMC.* Node X observes an action sequence X^n chosen by nature and common randomness message J. A coordination message I is generated based on X^n and the common randomness message J. The coordination message is then represented by a channel sequence A^n and transmitted to node Y over the DMC. Node Y decodes the received sequence B^n and regenerates $\hat{I}$. Based on the decoded coordination message $\hat{I}$ and the common randomness J, node Y generates the action sequence Y^n.

We assume that the common randomness is independent of the action specified at node X. The local randomness $M_1 \in \{1, 2, ..., 2^{n\rho_1}\}$ is used to generate the message I at node X and the rate of local randomness $M_2 \in \{1, 2, ..., 2^{n\rho_2}\}$ is used to generate the sequence Y^n at node Y.

4.2 General Scheme

Here we first present a general scheme for achieving strong coordination over the DMC channel $P_{B|A}(b|a)$. The scheme follows a similar approach to the work done in [2] and [6] where the coordination code is designed by analyzing an allied channel resolvability problem. Similarly we construct an allied structure that exploits Lemma 4 stated in Section 2.6 to accomplish the strong coordination behavior. First we simulate a memoryless channel $P_{XY|AC}$ from auxiliary random variables A and C to the nodes X and Y. Then, by imposing the Markov chain $X - (A, C) - Y$ the channel $P_{XY|AC}$ is separated into two sub-channels $P_{X|AC}$ and $P_{Y|AC}$ to form the allied scheme. However, due to the presence of a DMC in the strong coordination problem outlined in Section 4.1 we impose further modifications. Specifically, we simulate the sub-channel $P_{Y|AC}$ from the DMC $P_{B|A}$ and a test channel $P_{Y|BC}$. Further details are given below. In Figure 4.2 the allied structure which corresponds to the two-node network in Figure 4.1 is shown.

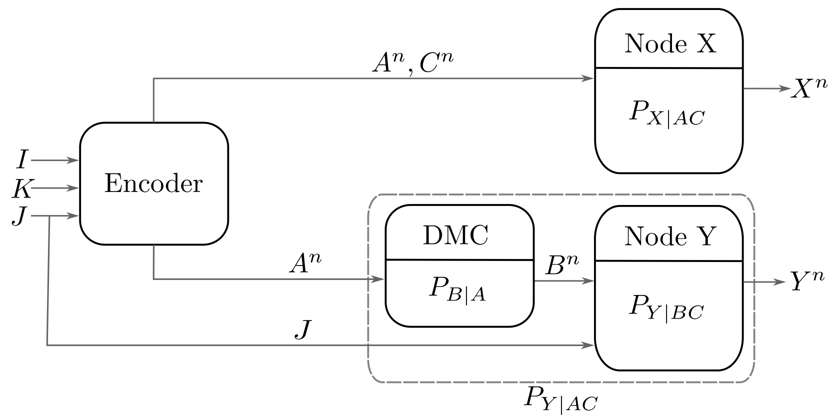

Figure 4.2 *The allied structure of strong coordination over a DMC.* The indices I, J, and K are used to specify the sequences A^n and C^n uniformly at random from a randomly generated codebook of size 2^{nR}. Both sequences are used as channel $P_{X|AC}$ input at node X to generate X^n. Only the sequence A^n is used as a channel input for the DMC. The output sequence B^n is decoded to obtain the index $\hat{I}$. Node Y regenerates C^n using the decoder output and a side information message J. Both the channel output sequence B^n and the regenerated sequence C^n are used as input for the channel with the pmf $P_{Y|BC}$ at node Y to generate Y^n.

The channel resolvability code for the allied problem with parameters (R_o, R_c, R_a, n) consists of the following:

- Two nested codebooks:

  As illustrated in Figure 4.3 the codebook $\mathcal{C}$ of size $2^{n(R_o+R_c)}$ is generated i.i.d. according to a distribution $p_C(c)$. The codebook is arranged into 2^{nR_o} columns and 2^{nR_c} rows. That is, $C^n_{ij} \sim \prod_{l=1}^{n} p_C(c_l)$ for all $(i, j) \in \mathcal{I} \times \mathcal{J}$;

codebook $\mathcal{A}$ is generated accordingly i.e., $A^n_{ijk} \sim \prod_{l=1}^n p_{A|C}(a_l|c_{ij})$ for all $(i,j,k) \in \mathcal{I} \times \mathcal{J} \times \mathcal{K}$.

- Encoding functions:

$$C^n : \{1,2,...,2^{nR_c}\} \times \{1,2,...,2^{nR_o}\} \to \mathcal{C}^n;$$
$$A^n : \{1,2,...,2^{nR_c}\} \times \{1,2,...,2^{nR_o}\} \times \{1,2,...,2^{nR_a}\} \to \mathcal{A}^n.$$

- Index sets:

$$i \in \{1,2,...,2^{nR_c}\},$$
$$j \in \{1,2,...,2^{nR_o}\},$$
$$k \in \{1,2,...,2^{nR_a}\},$$

where

$$R \triangleq R_o + R_c + R_a.$$

These indices are used to specify the pair of sequences C^n_{IJ} and A^n_{IJK} uniformly at random from the codebooks $\mathcal{C}$ and $\mathcal{A}$.

- The selected sequences C^n_{ij} and A^n_{ijk} are then passed through the memoryless channel $P_{X|AC}(x|ac)$ at node X, while at node Y sequence A^n_{ijk} is first passed through the DMC $P_{B|A}(b|a)$. The DMC output B^n is decoded to reconstruct the sequence $C^n_{\hat{i}j}$ and both are then passed through the test channel $P_{Y|BC}(y|bc)$ generated using local randomness at rate ρ_2. As a result, we have the following backward induced joint distribution:

$$\check{P}_{X^nY^nA^nC^n}(x^n,y^n,a^n,c^n) \triangleq \sum_{b^n} 2^{-nR} P_{X|AC}(x^n|A^n_{ijk}C^n_{ij}) P_{B|A}(B^n|A^n) P_{Y|BC}(y^n|B^nC^n_{\hat{i}j}). \tag{4.1}$$

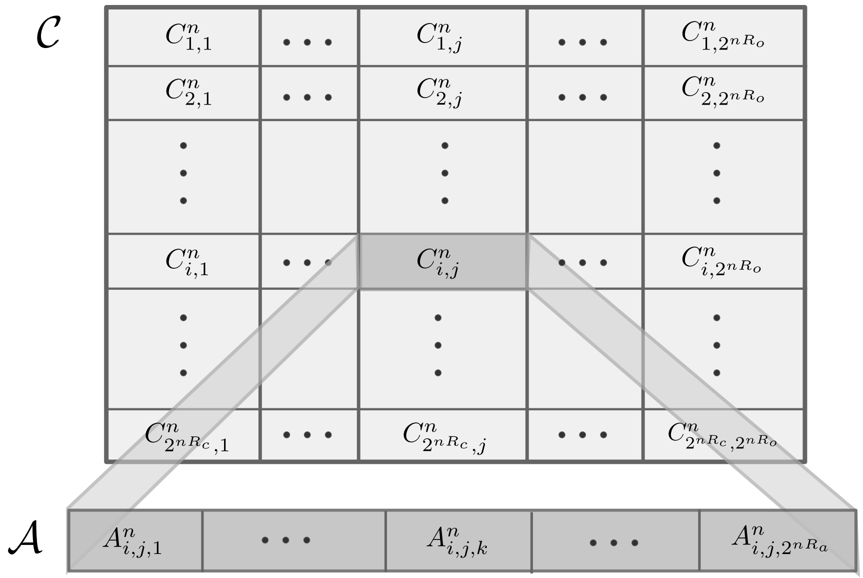

Figure 4.3 *Random codebook structure.* The codebook $\mathcal{C}$ is arranged into I columns and J rows. For each (I, J) codeword in $\mathcal{C}$ there exists 2^{nR_a} codewords in $\mathcal{A}$.

We start deriving the achievable rates for the problem at hand by dividing the allied structure into the following subproblems.

4.2.1 Channel Resolvability Constraints

Here we consider the DMC $P_{B|A}(b|a)$ and the test channel $P_{Y|BC}(y|bc)$ at node Y in the allied problem to form one black box. This box has an input sequence A^n, reconstructs C^n from $\hat{I}$ and side information J and outputs the sequence Y^n. At this point, we assume that C^n is reconstructed perfectly (i.e., $\hat{I} = I$). Consequently, we find the required channel resolvability rates that satisfies the strong coordination constraint. These constraints are characterized in Theorem 5.

Theorem 5 (Resolvability constraints)**.** *For a discrete memoryless channel defined by $P_{B|A}(b|a)$, and a nested codebook structure detailed above, the total variation between the desired i.i.d. distribution $P^{\otimes n}_{XY}(x^n, y^n)$ and the backward induced distribution, induced by the allied scheme, $\check{P}_{X^nY^n}(x^n, y^n)$ goes to zero as n goes to infinity,*

$$\left\|\check{P}_{X^nY^n}(x^n, y^n) - P^{\otimes n}_{XY}(x^n, y^n)\right\|_{TV} \le \epsilon,$$

if the following is satisfied:

$$R_a + R_o + R_c \geq I(XY; AC),$$

$$R_o + R_c \geq I(XY; C).$$

For the following proof, we drop the subscripts from the pmfs; e.g., $P(x^n|A^n, C^n)$ will denote $P_{X|AC}(x^n|A^n, C^n)$. Moreover, we define $\epsilon' > 0$.

Proof.

$$\begin{aligned}
&E[\mathbb{D}(\check{P}_{X^nY^n}(x^n, y^n)||P_{XY}^{\otimes n}(x^n, y^n))] \\
&\quad = E\Bigg[\sum_{x^n,y^n}\Big(\sum_{i,j,k}\frac{P(x^n|A^n_{ijk}, C^n_{ij})P(y^n|A^n_{ijk}, C^n_{ij})}{2^{nR}}\Big) \\
&\qquad\qquad \log\Bigg(\sum_{i',j',k'}\frac{P(x^n|A^n_{i'j'k'}, C^n_{i'j'})P(y^n|A^n_{i'j'k'}, C^n_{i'j'})}{2^{nR}P_{XY}^{\otimes n}(x^n, y^n)}\Bigg)\Bigg] \\
&\quad = \sum_{x^n,y^n} E\Bigg[\Big(\sum_{i,j,k}\frac{P(x^n|A^n_{ijk}, C^n_{ij})P(y^n|A^n_{ijk}, C^n_{ij})}{2^{nR}}\Big) \\
&\qquad\qquad \log\Bigg(\sum_{i',j',k'}\frac{P(x^n|A^n_{i'j'k'}, C^n_{i'j'})P(y^n|A^n_{i'j'k'}, C^n_{i'j'})}{2^{nR}P_{XY}^{\otimes n}(x^n, y^n)}\Bigg)\Bigg] \\
&\quad \overset{(a)}{=} \sum_{x^n,y^n} E_{A^n_{ijk}C^n_{ij}}\Bigg[\Big(\sum_{i,j,k}\frac{P(x^n|A^n_{ijk}, C^n_{ij})P(y^n|A^n_{ijk}, C^n_{ij})}{2^{nR}}\Big) \\
&\qquad\qquad E_{rest}\Big[\log\Big(\sum_{i',j',k'}\frac{P(x^n|A^n_{i'j'k'}, C^n_{i'j'})P(y^n|A^n_{i'j'k'}, C^n_{i'j'})}{2^{nR}P_{XY}^{\otimes n}(x^n, y^n)}\Big)\Big|A^n_{ijk}C^n_{ij}\Big]\Bigg] \\
&\quad \overset{(b)}{\leq} \sum_{x^n,y^n} E_{A^n_{ijk}C^n_{ij}}\Bigg[\Big(\sum_{i,j,k}\frac{P(x^n|A^n_{ijk}, C^n_{ij})P(y^n|A^n_{ijk}, C^n_{ij})}{2^{nR}}\Big) \\
&\qquad\qquad \log\Big(E_{rest}\Big[\sum_{i',j',k'}\frac{P(x^n|A^n_{i'j'k'}, C^n_{i'j'})P(y^n|A^n_{i'j'k'}, C^n_{i'j'})}{2^{nR}P_{XY}^{\otimes n}(x^n, y^n)}\Big|A^n_{ijk}C^n_{ij}\Big]\Big)\Bigg] \\
&\quad = \sum_{x^n,y^n}\sum_{a^n_{ijk},c^n_{ij}} P(a^n_{ijk}, c^n_{ij})\Big(\sum_{i,j,k}\frac{P(x^n|A^n_{ijk}, C^n_{ij})P(y^n|A^n_{ijk}, C^n_{ij})}{2^{nR}}\Big) \\
&\qquad\qquad \log\Big(\sum_{i',j',k'} E_{rest}\Big[\frac{P(x^n|A^n_{i'j'k'}, C^n_{i'j'})P(y^n|A^n_{i'j'k'}, C^n_{i'j'})}{2^{nR}P_{XY}^{\otimes n}(x^n, y^n)}\Big|A^n_{ijk}C^n_{ij}\Big]\Big)
\end{aligned}$$

$$
\begin{aligned}
&\overset{(c)}{=} \sum_{x^n,y^n} \sum_{a^n_{ijk},c^n_{ij}} \sum_{i,j,k} \frac{P(x^n,y^n,a^n_{ijk},c^n_{ij})}{2^{nR}} \\
&\qquad \log\Bigg(\sum_{\substack{i',j',k':\\(i',j',k')=(i,j,k)}} E_{A^n_{ijk}C^n_{ij}}\Big[\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})P(y^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{nR}P^{\otimes n}_{XY}(x^n,y^n)}\Big|A^n_{ijk}C^n_{ij}\Big] \\
&\qquad\quad + \sum_{\substack{i',j',k':\\(i',j')=(i,j),(k'\neq k)}} E_{A^n_{ijk}C^n_{ij}}\Big[\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})P(y^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{nR}P^{\otimes n}_{XY}(x^n,y^n)}\Big|A^n_{ijk}C^n_{ij}\Big] \\
&\qquad\quad + \sum_{\substack{i',j',k'':\\(i',j')\neq(i,j)}} E_{A^n_{ijk}C^n_{ij}}\Big[\frac{P(x^n|A^n_{i'j'k'},C^n_{i'j'})P(y^n|A^n_{i'j'k'},C^n_{i'j'})}{2^{nR}P^{\otimes n}_{XY}(x^n,y^n)}\Big|A^n_{ijk}C^n_{ij}\Big]\Bigg) \\
&\overset{(d)}{=} \sum_{x^n,y^n} \sum_{a^n_{ijk},c^n_{ij}} \sum_{i,j,k} \frac{P(x^n,y^n,a^n_{ijk},c^n_{ij})}{2^{nR}} \log\Bigg(\frac{P(x^n,y^n|a^n_{ijk},c^n_{ij})}{2^{nR}P^{\otimes n}_{XY}(x^n,y^n)} \\
&\qquad\quad + \sum_{i',j',k':(i',j')=(i,j),(k'\neq k)} \frac{P(x^n,y^n|c^n_{ij})}{2^{nR}P^{\otimes n}_{XY}(x^n,y^n)} + \sum_{i',j',k':(i',j')\neq(i,j)} \frac{P^{\otimes n}_{XY}(x^n,y^n)}{2^{nR}P^{\otimes n}_{XY}(x^n,y^n)}\Bigg) \\
&\overset{(e)}{\leq} \sum_{x^n,y^n,a^n_{ijk},c^n_{ij}} P(x^n,y^n,a^n_{ijk},c^n_{ij}) \log\Bigg(\frac{P(x^n,y^n|a^n_{ijk},c^n_{ij})}{2^{nR}P^{\otimes n}_{XY}(x^n,y^n)} + (2^{R_a})\frac{P(x^n,y^n|c^n_{ij})}{2^{nR}P^{\otimes n}_{XY}(x^n,y^n)} + 1\Bigg) \\
&= \sum_{x^n,y^n,a^n_{ijk},c^n_{ij}} P(x^n,y^n,a^n_{ijk},c^n_{ij}) \log\Bigg(\frac{P(x^n,y^n|a^n_{ijk},c^n_{ij})}{2^{nR}P^{\otimes n}_{XY}(x^n,y^n)} + \frac{P(x^n,y^n|c^n_{ij})}{2^{n(R_o+R_c)}P^{\otimes n}_{XY}(x^n,y^n)} + 1\Bigg) \\
&\overset{(f)}{\leq} \Bigg[\sum_{\substack{x^n,y^n,a^n_{ijk},c^n_{ij}:\\(x^n,y^n,a^n,c^n)\in\mathcal{T}^n_\epsilon(p_{XYAC})}} P(x^n,y^n,a^n_{ijk},c^n_{ij}) \log\Bigg(\frac{2^{-nH(XY|AC)(1-\epsilon)}}{2^{nR}2^{-nH(XY)(1+\epsilon)}} + \frac{2^{-nH(XY|C)(1-\epsilon)}}{2^{n(R_o+R_c)}2^{-nH(XY)(1+\epsilon)}} + 1\Bigg)\Bigg] \\
&\qquad\quad + Pr\big((x^n,y^n,a^n_{ijk},c^n_{ij}) \notin \mathcal{T}^n_\epsilon(p_{XYAC})\big)\log(2\mu^{-n}_{XY}+1) \\
&\overset{(g)}{\leq} \Bigg[\sum_{\substack{x^n,y^n,a^n_{ijk},c^n_{ij}:\\(x^n,y^n,a^n,c^n)\in\mathcal{T}^n_\epsilon(p_{XYAC})}} P(x^n,y^n,a^n_{ijk},c^n_{ij}) \log\Bigg(\frac{2^{n(I(XY;AC)+\delta(\epsilon))}}{2^{nR}} + \frac{2^{n(I(XY;C)+\delta(\epsilon))}}{2^{n(R_o+R_c)}} + 1\Bigg)\Bigg] \\
&\qquad\quad + \big(2|\mathcal{X}||\mathcal{Y}||\mathcal{A}||\mathcal{C}|e^{-n\epsilon^2\mu_{XYAC}}\big)\log(2\mu^{-n}_{XY}+1) \\
&\overset{(h)}{\leq} \epsilon'
\end{aligned}
$$

where:

(a^n_{ijk}, c^n_{ij}) are the codewords corresponding to the indices (i,j,k), and $(a^n_{i'j'k'}, c^n_{i'j'})$ are the codewords corresponding to the indices (i',j',k');

(a) follows from the law of iterated expectation;

(b) follows from Jensen's inequality [22];

(c) follows from dividing the inner summation over the indices (i', j', k') into three subsets based on the indices (i, j, k) from the outer summation;

(d) follows as a result from taking the expectation within the subsets in (c) such that when:

- $(i', j') = (i, j), (k' \neq k)$: $a^n_{i'j'k'}$ is conditionally independent of a^n_{ijk} following the nature of the codebook construction (i.e., i.i.d. at random);

- $(i', j') \neq (i, j)$: both codewords (a^n_{ijk}, c^n_{ij}) are independent of $(a^n_{i'j'k'}, c^n_{i'j'})$ regardless of the value of k. As a result, the expected value of the induced distribution with respect to the input codebooks is the desired distribution $P^{\otimes n}_{XY}$ [2].

(e) follows from

- $(i', j', k') = (i, j, k)$: there is only one pair of codewords (a^n_{ijk}, c^n_{ij});

- when $(k' \neq k)$ while $(i', j') = (i, j)$ there are $(2^{nR_a} - 1)$ indices in the sum;

- $(i', j') \neq (i, j)$: the number of the indices is at most 2^{nR}.

(f) results from splitting the outer summation: The first summation contains typical sequences and is bounded by using the probabilities of the typical set. The second summation contains the tuple of sequences when the pair of actions sequences x^n, y^n and codewords c^n, a^n are not ϵ-jointly typical (i.e., $(x^n, y^n, a^n, c^n) \notin \mathcal{T}^n_\epsilon(p_{XYAC})$). This sum is upper bounded following [5] with $\mu_{XY} = \min_{x,y} \big(p_{XY}(x, y)\big)$.

(g) following the Chernoff bound of the probability that a sequence is not strongly typical [28] where $\mu_{XYAC} = \min_{x,y,a,c} \big(p_{XYAC}(x, y, a, c)\big)$.

(h) consequently, the contribution of typical sequences can be asymptotically made small if

$$R_a + R_o + R_c \geq I(XY; AC),$$

$$R_o + R_c \geq I(XY; C),$$

while the second term converges to zero exponentially fast with n [28].

Finally, by applying Pinsker's inequality [23] we have

$$||\check{P}_{X^nY^n}(x^n,y^n) - P_{XY}^{\otimes n}(x^n,y^n)||_{TV} \leqslant \sqrt{2\mathbb{D}(\check{P}_{X^nY^n}(x^n,y^n)||P_{XY}^{\otimes n}(x^n,y^n))},$$
$$||\check{P}_{X^nY^n}(x^n,y^n) - P_{XY}^{\otimes n}(x^n,y^n)||_{TV} \leqslant \sqrt{2\epsilon'}.$$

■

After deriving the rates achieving strong coordination over the allied structure, in Section 4.2.2 and 4.2.3 we will look into the process that will force the allied structure to match the behavior of the original point-to-point coordination scheme over a DMC.

4.2.2 Decodability Constraint

Here we look in details to the operation of the DMC $P_{B|A}(b|a)$ and the test channel $P_{Y|BC}(y|bc)$ at node Y, mentioned as the black-box in Section 4.2.1. We want to send the sequence A^n_{ijk} over the channel $P_{B|A}(b|a)$ so that node Y can reconstruct the correlated sequence $C^n_{\hat{i}j}$ using the channel output B^n and the the common randomness information J as shown in Figure 4.4. In other words, we want to estimate the message I from the output of the DMC with a negligible probability of error. Theorem 6 outlines the necessary rates for this to occur.

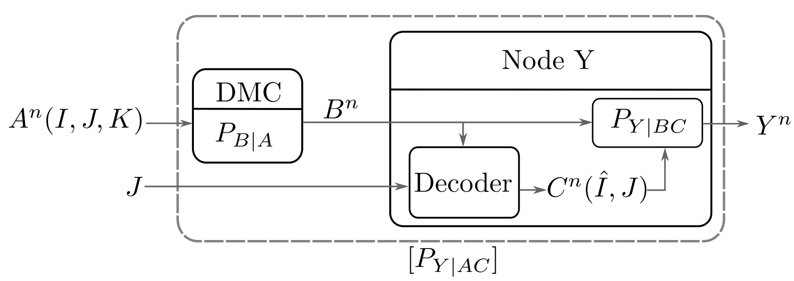

Figure 4.4 Allied structure; simulation of $P_{Y|AC}$.

Theorem 6 (Decodability constraint). *Given a nested codebook of the size $2^{n(R_o+R_c)}$ and $2^{n(R_o+R_c+R_a)}$ producing the codewords C^n_{IJ} and A^n_{IJK}, respectively. Given the channel output sequence B^n obtained from passing the codeword A^n_{IJK} through a DMC $P_{B|A}(b|a)$. A jointly typical decoder with side information J can estimate the index I to reconstruct the sequence $C^n_{\hat{I}J}$ from B^n with an estimation error that goes to zero as n goes to infinity (i.e., $Pr(\hat{I} \neq I) \rightarrow 0$), if the codebook rate for the index I satisfies $R_c < I(B;C)$.*

Proof. We begin the proof by constructing the sets

$$\begin{aligned} S_j \triangleq & \{i : \exists k \text{ such that } (A^n_{ijk}, B^n, C^n_{ij}) \in \mathcal{T}^n_\epsilon(P_{ABC})\}, \\ \hat{S}_j \triangleq & \{i : (B^n, C^n_{ij}) \in \mathcal{T}^n_\epsilon(P_{BC})\}. \end{aligned}$$

S_j is the set of the indices $i \in I$ such that for a given index $J = j$ the sequences A^n_{ijk}, B^n, C^n_{ij} are jointly-typical. $\hat{S}_j$ is a subset containing the indices $i \in I$ where only B^n and C^n_{ij} are jointly-typical. Following the jointly typical decoding method [22], the receiver, node Y, declares that the index I was sent if the following conditions are met:

1. For $I = i$ and $J = j$ the sequence B^n is jointly typical with C^n. That is $Pr[I \in \hat{S}_j] \rightarrow 1$.

$$\begin{aligned} Pr[I \in \hat{S}_j] &= Pr[I \in \hat{S}_j|^{I=i}_{J=j}] \\ &\overset{(a)}{=} Pr[I \in \hat{S}_j|^{I=1}_{J=1}] \\ &= \sum_{a^n,b^n,c^n} P(C^n_{ij} = C^n)P(A^n|C^n)P(B^n|A^n)\mathbf{1}\big((C^n, B^n) \in \mathcal{T}^n_\epsilon(P_{BC})\big) \\ &= \sum_{a^n,c^n} P(A^n, C^n) \sum_{b^n} P(B^n|A^n)\mathbf{1}\big((C^n, B^n) \in \mathcal{T}^n_\epsilon(P_{BC})\big) \end{aligned}$$

$$
\begin{aligned}
&\overset{(b)}{\geq} \sum_{a^n,c^n} P(A^n,C^n) \sum_{b^n} P(B^n|A^n)\mathbf{1}\big((A^n,C^n,B^n) \in \mathcal{T}_\epsilon^n(P_{ABC})\big) \\
&= \sum_{a^n,b^n,c^n} P(A^n,B^n,C^n)\mathbf{1}\big((A^n,C^n,B^n) \in \mathcal{T}_\epsilon^n(P_{ABC})\big) \\
&\overset{(c)}{\geq} 1-\delta(\epsilon)
\end{aligned}
$$

where

(a) results from the symmetry of codebook construction,

(b) follows from the fact that $\hat{S}_j \supseteq S_j$,

(c) follows from the probability of the jointly typical set [28], where $\delta(\epsilon) \to 0$ as $n \to \infty$.

2. There is no other index $I' \neq I$ such that $(C^n_{I'J}, B^n) \in \mathcal{T}_\epsilon^n(P_{BC})$. That is $Pr[|\hat{S}_j| = 1] \to 1$

$$
\begin{aligned}
Pr[|\hat{S}_j| = 1] &= Pr[\hat{S}_j \cap \{1,2,..,I-1,I+1,...,2^{nR_c}\} = \emptyset] \\
&= \sum_{I,J} P(i,j) Pr[\hat{S}_j \cap \{1,2,..,I-1,I+1,...,2^{nR_c}\} = \emptyset|_{J=j}^{I=i}] \\
&= 2^{-n(R_o+R_c)} \sum_{I,J} \left(1 - \sum_{i' \neq i} Pr[i' \in \hat{S}_j|_{J=j}^{I=i}]\right) \\
&= 2^{-n(R_o+R_c)} \sum_{I,J} \left(1 - \sum_{i' \neq i} Pr[(C^n_{i'j}, B^n) \in \mathcal{T}_\epsilon^n(P_{BC})]\right)
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(a)}{\geq} 2^{-n(R_o+R_c)} \sum_{I,J} \left(1 - \sum_{i' \neq i} 2^{-n(I(B;C)-\delta(\epsilon))}\right) \\
&= 2^{-n(R_o+R_c)} \sum_{I,J} \left(1 - (2^{nR_c}-1)2^{-n(I(B;C)-\delta(\epsilon))}\right) \\
&= 1 - (2^{nR_c}-1)2^{-n(I(B;C)-\delta(\epsilon))} \\
&= 1 - 2^{-n(I(B;C)-R_c-\delta(\epsilon))} + 2^{-nI(B;C)} \\
&\overset{(b)}{\geq} 1-\delta(\epsilon')
\end{aligned}
$$

where

(a) follows from the packing lemma from Section 2.4 [24],

(b) results if $R_c < I(B;C) - \delta(\epsilon')$ and for sufficiently large n: $\delta(\epsilon') \to 0$.

■

4.2.3 Secrecy Constraint

We complete modifying the allied structure to mimic to the original problem with a final step. By assumption, we have a natural independence between the action sequence X^n and the common randomness source J. As a result, the joint distribution over X^n and J in the original problem is a product of the marginal distributions $P_X^{\otimes n}(x)$ and $P_J(j)$. To mimic this behavior in the allied structure we need to force that independence by minimizing the mutual information between the two. This process is outlined in Theorem 7.

In the original structure of Figure 4.1, we are given the input action sequence x^n and the index j available from the common randomness source to implement the codewords C_{ij}^n, A_{ijk}^n respectively. The sequence A_{ijk}^n is then communicated over the DMC channel $P_{B|A}(b|a)$ to node Y as pointed out in Section 4.2.2. As a result, the induced joint distribution by the original problem is given as

$$\hat{P}_{X^nY^nA^nC^n}(x^n, y^n, a^n, c^n) \triangleq \sum_{b^n} P_X^{\otimes n}(x)P(j)P_{AC|XJ}(C_{ij}^n, A^n ijk|x^n, j)P_{B|A}(B^n|A^n)P_{Y|BC}(y^n|B^n, C_{\hat{i}j}^n). \quad (4.2)$$

Similar to the strong coordination scheme in [2] the joint distribution induced by the original problem and the backward joint distribution obtained from the allied structure are derived from a different joint distribution over X^n and J passed through the same channel

$$P_{AC|XJ}(C_{ij}^n, A_{ijk}^n|x^n, j)P_{B|A}(B^n|A^n)P_{Y|BC}(y^n|B^n, C_{\hat{i}j}^n).$$

We identify the distributions of the action sequence X^n, the test channels of the allied structure, and the encoder function for the original problem with the following notation:

$$
\begin{aligned}
&P_J(j) = p(j), \\
&P_X^{\otimes n}(x) = p(x^n), \\
&P_{X|AC}(x^n|A^n_{ijk}, C^n_{ij}) \triangleq \hat{p}(x^n|i,j,k), \\
&P_{Y|AC}(y^n|A^n_{ijk}, C^n_{ij}) \triangleq \hat{p}(y^n|i,j,k), \\
&P_{AC|XJ}(A^n_{ijk}, C^n_{ij}|x^n, j) \triangleq \hat{p}(i|x^n,j)\hat{p}(k|i,j) = \hat{p}(i,k|x^n,j).
\end{aligned}
$$

Accordingly, with these definitions we re-write the joint distribution induced by the original problem as

$$\hat{P}_{X^nY^nA^nC^n}(x^n, y^n, a^n, c^n) = p(x^n)p(j)\hat{p}(i,k|x^n,j)\hat{p}(y^n|\hat{i},j,k)$$

and the joint distribution induced by the allied structure as

$$
\begin{aligned}
\check{P}_{X^nY^nA^nC^n}(x^n, y^n, a^n, c^n) &= 2^{-nR}\hat{p}(x^n|i,j,k)\hat{p}(y^n|\hat{i},j,k) \\
&= \hat{p}(i,j,k)\hat{p}(x^n|i,j,k)\hat{p}(y^n|\hat{i},j,k) \\
&= \hat{p}(x^n,i,j,k)\hat{p}(y^n|\hat{i},j,k) \\
&= \hat{p}(x^n,j)\hat{p}(i,k|x^n,j)\hat{p}(y^n|\hat{i},j,k).
\end{aligned}
$$

Theorem 7 (Secrecy constraint)**.** *Given the input sequence $X^n = (X_1, X_2, ..., X_n)$, channel resolvability code with parameters (R_o, R_c, R_a, n) and the common randomness source, the dependency between the action sequence x^n and the common randomness J converges to zero exponentially fast as n goes to infinity,*

$$||\hat{p}(x^n, j) - p(x^n)p(j))||_{TV} \leq \epsilon,$$

if the code rates satisfy:

$$R_a + R_c \geq I(X;AC),$$
$$R_c \geq I(X;C).$$

The proof of Theorem 7 is built on the results of Section 4.2.2 (i.e., $\hat{I} = I$ with high probability). The bound on $||\hat{p}(x^n, j) - p(x^n)p(j))||_{TV}$ is obtained in a similar manner as the channel resolvability constraint proof in Section 4.2.1 as follows.

Proof.

$$
\begin{aligned}
E[\mathbb{D}(\hat{p}(x^n,j)||p(x^n)p(j))] &= E\Bigg[\sum_{x^n,j}\Big(\sum_{i,k}\frac{\hat{p}(x^n|i,j,k)}{2^{nR}}\Big)\log\Big(\sum_{i',k'}\frac{\hat{p}(x^n|i',j',k')}{2^{nR}p(x^n)p(j)}\Big)\Bigg] \\
&= \sum_{x^n,i,j,k} E\Bigg[\Big(\frac{\hat{p}(x^n|i,j,k)}{2^{nR}}\Big)\log\Big(\sum_{i',k'}\frac{\hat{p}(x^n|i',j',k')}{2^{n(R_a+R_c)}p(x^n)}\Big)\Bigg] \\
&\overset{(a)}{=} \sum_{x^n,i,j,k} E_{i,j,k}\Bigg[\Big(\frac{\hat{p}(x^n|i,j,k)}{2^{nR}}\Big)E_{rest}\Big[\log\Big(\sum_{i',k'}\frac{\hat{p}(x^n|i',j',k')}{2^{n(R_a+R_c)}p(x^n)}\Big)\Big|i,j,k\Big]\Bigg] \\
&\overset{(b)}{\leq} \sum_{x^n,i,j,k} E_{i,j,k}\Bigg[\Big(\frac{\hat{p}(x^n|j,j,k)}{2^{nR}}\Big)\log\Big(E_{rest}\Big[\sum_{i',k'}\frac{\hat{p}(x^n|i',j',k')}{2^{n(R_a+R_c)}p(x^n)}\Big|i,j,k\Big]\Big)\Bigg] \\
&= \sum_{x^n,i,j,k}\hat{p}(x^n,i,j,k)\Big(\log\Big(\sum_{i',k'}E_{rest}\Big[\frac{\hat{p}(x^n|i',j',k')}{2^{n(R_a+R_c)}p(x^n)}\Big|i,j,k\Big]\Big) \\
&\overset{(c)}{=} \sum_{x^n,i,j,k}\hat{p}(x^n,i,j,k)\log\Bigg(\sum_{\substack{i',k':\\(i',j',k')=(i,j,k)}} E_{i,j,k}\Big[\frac{\hat{p}(x^n|i',j',k')}{2^{n(R_a+R_c)}p(x^n)}\Big|i,j,k\Big] \\
&\quad + \sum_{\substack{i',k':\\(i',j')=(i,j),(k'\neq k)}} E_{i,j,k}\Big[\frac{\hat{p}(x^n|i',j',k')}{2^{n(R_a+R_c)}p(x^n)}\Big|i,j,k\Big] \\
&\quad + \sum_{\substack{i',j',k':\\(i',j',k')\neq(i,j,k)}} E_{i,j,k}\Big[\frac{\hat{p}(x^n|i',j,'k')}{2^{n(R_a+R_c)}p(x^n)}\Big|i,j,k\Big]\Bigg)
\end{aligned}
$$

$$\overset{(d)}{=} \sum_{x^n,i,j,k} \hat{p}(x^n,i,j,k) \log\left(\frac{\hat{p}(x^n|i,j,k)}{2^{n(R_a+R_c)}p(x^n)} + \sum_{\substack{i',k':\\(i',j')=(i,j),(k'\neq k)}} \frac{\hat{p}(x^n|i,j)}{2^{n(R_a+R_c)}p(x^n)} \right.$$

$$\left. + \sum_{\substack{i',k':\\(i',j',k')\neq(i,j,k)}} \frac{\hat{p}(x^n)}{2^{n(R_a+R_c)}p(x^n)} \right)$$

$$\overset{(e)}{\leq} \sum_{x^n,i,j,k} \hat{p}(x^n,i,j,k) \log\left(\frac{\hat{p}(x^n|i,j,k)}{2^{n(R_a+R_c)}p(x^n)} + (2^{R_a}) \frac{\hat{p}(x^n|i,j)}{2^{n(R_a+R_c)}p(x^n)} + 1 \right)$$

$$\overset{(f)}{\leq} \left[\sum_{\substack{x^n,i,j,k:\\(x^n,a^n(i,j,k),c^n(i,j)\in\mathcal{T}_\epsilon^n(p_{XAC})}} P(x^n,i,j,k) \log\left(\frac{2^{-nH(X|AC)(1-\epsilon)}}{2^{n(R_a+R_o)}2^{-nH(X)(1+\epsilon)}} + \frac{2^{-nH(X|C)(1-\epsilon)}}{2^{nR_c}2^{-nH(X)(1+\epsilon)}} \right.\right.$$

$$\left.\left. + 1 \right)\right] + Pr\big((x^n,a^n(i,j,k),c^n(i,j)) \notin \mathcal{T}_\epsilon^n(p_{XAC})\big) \log(2\mu_X^{-n}+1)$$

$$\overset{(g)}{\leq} \left[\sum_{\substack{x^n,i,j,k:\\(x^n,a^n(i,j,k),c^n(i,j))\in\mathcal{T}_\epsilon^n(p_{XAC})}} P(x^n,i,j,k) \log\left(\frac{2^{n(I(X;AC)+\delta(\epsilon))}}{2^{n(R_c+R_a)}} + \frac{2^{n(I(X;C)+\delta(\epsilon))}}{2^{n(R_c)}} + 1 \right)\right]$$

$$+ \big(2|\mathcal{X}||\mathcal{A}||\mathcal{C}|e^{-n\epsilon^2\mu_{XAC}}\big) \log(2\mu_X^{-n}+1)$$

$$\overset{(h)}{\leq} \epsilon',$$

where $\epsilon' > 0$;

(a) follows from the law of iterated expectation;

(b) follows from Jensen's inequality [22];

(c) follows from dividing the inner summation over the indices (i', k') into three subsets based on the indices (i, j, k) from the outer summation;

(d) results from taking the conditional expectation within the subsets in (c);

(e) follows from

- $(i', j', k') = (i, j, k)$: there is only one pair of codewords represented by the indices A^n_{ijk}, C^n_{ij} corresponding to x^n;
- when $(k' \neq k)$ and $(i', j') = (i, j)$ there are $(2^{nR_a} - 1)$ indices in the sum;

- $(i', j', k') \neq (i, j, k)$: the number of the indices is at most $2^{n(R_a+R_c)}$. Moreover, $\hat{p}(x)$ is less than ϵ close to $p(x^n)$ as a consequence of Theorem 5 and [2, Lemma 16].

(f) results from splitting the outer summation: The first summation contains typical sequences and is bounded by using the probabilities of the typical set. The second summation contains the tuple of sequences when the action sequence x^n and codewords (c^n, a^n) represented here by the indices (i, j, k) are not ϵ-jointly typical (i.e., $(x^n, a^n, c^n) \notin \mathcal{T}_\epsilon^n(p_{XAC})$). This sum is upper bounded following [5] using $\mu_X = \min_x \big(p_X(x)\big)$.

(g) following from the Chernoff bound on the probability of a non-typical sequence [28].

(h) consequently, the contribution of typical sequences can be asymptotically made small if

$$
\begin{aligned}
R_a + R_c &\geq I(X; AC), \\
R_c &\geq I(X; C).
\end{aligned}
$$

The second term converges to zero exponentially fast with n [28] and following Pinsker's inequality [23] we have

$$
\begin{aligned}
||\hat{p}(x^n, j) - p(x^n)p(j))||_{TV} &\leqslant \sqrt{2\mathbb{D}(\hat{p}(x^n, j)||p(x^n)p(j))}, \\
||\hat{p}(x^n, j) - p(x^n)p(j))||_{TV} &\leqslant \sqrt{2\epsilon'}.
\end{aligned}
$$

Finally, we combine the results of Theorem 5, Theorem 6, and Theorem 7. From [2] we obtain the following:

$$
\begin{aligned}
||\hat{P}_{X^nY^n}(x^n, y^n) - P^{\otimes n}{}_{XY}(x^n, y^n)||_{TV} & \\
&\overset{(a)}{\leq} ||\hat{P}_{X^nY^n}(x^n, y^n) - \check{P}_{X^nY^n}(x^n, y^n)||_{TV} \\
&\qquad + ||\check{P}_{X^nY^n}(x^n, y^n) - P^{\otimes n}_{XY}(x^n, y^n)||_{TV} \\
&\overset{(b)}{\leq} ||\hat{P}_{X^nY^nA^nC^n}(x^n, y^n, a^n, c^n) - \check{P}_{X^nY^nA^nC^n}(x^n, y^n, a^n, c^n)||_{TV} \\
&\qquad + ||\check{P}_{X^nY^n}(x^n, y^n) - P^{\otimes n}_{XY}(x^n, y^n)||_{TV} \\
&\overset{(c)}{=} ||\hat{p}(x^n, j) - p(x^n)p(j)||_{TV} + ||P_{X^nY^n}(x^n, y^n) - P^{\otimes n}_{XY}(x^n, y^n)||_{TV} \\
&\overset{(d)}{\leq} \sqrt{2\epsilon'} + \sqrt{2\epsilon'},
\end{aligned}
$$

where

$\epsilon' > 0$, $\epsilon' \to 0$ as $n \to \infty$;

(a) follows from the triangle inequality;

(b) follows from [2, Lemma 16];

(c) follows from [2, Lemma 17];

(d) the first term is the bound obtained as a result of Theorem 7 and the second term is the result of Theorem 5 and Theorem 6 combined.

■

As a conclusion, we can achieve strong coordination for the two-node network of Figure 4.1 using the (R_o, R_c, R_a, n)-channel resolvability code, as constructed in the presented scheme, given the conditions on the rates as

$$
\left\{
\begin{aligned}
R_a + R_o + R_c &\geq I(XY; AC), \\
R_o + R_c &\geq I(XY; C), \\
R_a + R_c &\geq I(X; AC), \\
R_c &\geq I(X; C), \\
R_c &< I(B; C).
\end{aligned}
\right\}
$$

4.2.4 Local Randomness Rates

In the original coordination network characterized by Cuff [2], each node possesses unlimited amount of local randomization to probabilistically map its input arguments onto the range of the encoding and decoding functions. Here we quantify the local randomness, $(\rho_k, k \in \{1,2\})$, possessed at each node. At node X, local randomness at rate ρ_1 is needed to randomize the selection of indices (i,k) by synthesizing the channel $P_{IK|X^nJ}(i,k|x^n,j)$. On the other hand, node Y utilizes local randomness at rate ρ_2 to generate the action sequence y^n by simulating the channel $P_{Y|BC}(y^n|B^n,C^n)$.

At node X, according to [6], to achieve the probabilistic selection of indices I,K the rate of local randomness must satisfy

$$\rho_1 \geq R_a + R_c - I(X;AC).$$

Here we have $(2^{R_c+R_a})$ (i.e., $(i,k) \in \mathcal{I} \times \mathcal{K}$) as this is the size of the selection list for any (i,k) pair for a given action sequence X^n and common randomness index J. On the other hand, $I(X;AC)$ is the minimum rate for (R_c+R_a) to establish independency between X and the common randomness source.

However, for the generation of the action sequence at node Y the necessary local randomness is bounded by the conditional entropy representing the channel $P_{Y|BC}(y^n|B^n,C^n)$. That is,

$$\rho_2 \geq H(Y|BC).$$

Moreover, as stated in [6, Lemma 2], if strong coordination is achievable we can apply the rate-transfer argument. That is, for any $1 \leq l \leq 2$, $\delta_k \geq 0, k \in \{1,2\}$ and $\delta_k \leq \rho_l$ the amount of δ_k is transferable between the node possessing the local randomness rate ρ_l and the rate of common randomness R_o. Accordingly, we can obtain the following list of rate conditions:

$$\left\{\begin{aligned} R_a + R_o + R_c &\geq I(X;AC) + \delta_1 + \delta_2, \\ R_o + R_c &\geq I(XY;C) + \delta_1 + \delta_2, \\ R_a + R_c &\geq I(X;AC), \\ R_c &\geq I(X;C), \\ R_c &< I(B;C), \\ \rho_1 &\geq R_a + R_c - I(X;AC) - \delta_1, \\ \rho_2 &\geq H(Y|BC) - \delta_2. \end{aligned}\right\} \tag{4.3}$$

Finally, the general scheme is considered as a joint coordination-channel coding scheme because it achieves strong coordination over the DMC channel in one shot. Since the objective of this section is to analyze joint versus separate coordination-channel coding over point-to-point coordination links we now proceed to separation-based coding schemes.

4.3 Basic Separate Coordination-Channel Coding Scheme

Now we consider the straight forward separation of coordination and channel codes by combining the main results of the strong coordination over a two-node network in [2] and the channel coding theorem [22]. We entitle this scheme as the basic separation scheme, and it is designed as follows:

- Coordination encoder: describes the action sequence X^n using a coordination code $(2^{nR_c}, 2^{nR_o}, n)$ with the codebook $\mathcal{U}$ as detailed in Section 3.2. This coordination code is achievable when the rates satisfies

$$\left\{\begin{aligned} R_c + R_o &\geq I(XY;U), && \text{channel resolvability constraint} \\ R_c &\geq I(X;U), && \text{secrecy constraint} \end{aligned}\right\}.$$

- Channel encoder: uses a channel code $(2^{mR_a}, m)$ with the codebook $\mathcal{A}$ to map the coordination message i with a codeword according to the encoding function

$$A^m : \{1, 2, ..., 2^{mR_a}\} \rightarrow \mathcal{A}^m,$$

where

$$R_c = \lambda R_a,$$

and λ is the transmission ratio given by m/n.

- Channel decoder: estimate the transmitted message $\hat{i}$ according to

$$g(B^m) : \mathcal{B}^m \rightarrow \{1, 2, ..., 2^{mR_a}\}.$$

According the channel coding theorem [22], this channel code is achievable and can communicate a block of length m with probability of decoding error (i.e., $Pr(\hat{I} \neq I)$) that goes to zero as $m \rightarrow \infty$ for any rate R_a less than the channel capacity C; resulting in

$$\begin{aligned} nR_c &= mR_a, \\ R_c &= \lambda R_a, \\ R_c &\leq \lambda C, \\ R_c &\leq \lambda I(A; B). \end{aligned}$$

- Coordination decoder: receives $\hat{i}$ from the channel decoder. Chooses the coordination codeword $u^n(\hat{i}, j)$ accordingly and passes it through a memoryless channel $P_{Y|U}$ to generate the action sequence y^n.

Therefore, we can achieve strong coordination with communication over the DMC $P_{B|A}(b|a)$ when we have coordination rate R_c satisfying

$$I(X; U) < R_c \leq \lambda I(A; B).$$

Next, similar to the joint scheme, we also quantify the local randomness at both nodes ρ_1 and ρ_2, apply the rate transfer Lemma [6, Lemma 2], and assume $\lambda = 1$ for simplicity. As a result we obtain the following rate region:

$$\left\{\begin{array}{rl} R_c + R_o \geq I(XY;U) + \delta_1 + \delta_2, & \text{channel resolvability constraint} \\ R_c \geq I(X;U), & \text{secrecy constraint} \\ R_c < I(A;B), & \text{decodability constraint} \\ \rho_1 \geq R_c - I(X;U) - \delta_1, & \text{index randomization at node X} \\ \rho_2 \geq H(Y|U) - \delta_2, & \text{sequence generation at node Y} \end{array}\right\}. \tag{4.4}$$

4.4 Improved Separate Coordination-Channel Coding Schemes

Now we first design an improved separate coordination-channel coding scheme by modifying the code for the allied problem depicted in Figure 4.5. In this scheme we imply separation by decoding the DMC input sequence A^n and separately use the decoder output to randomly generate Y^n. We construct the improved separate scheme as follows:

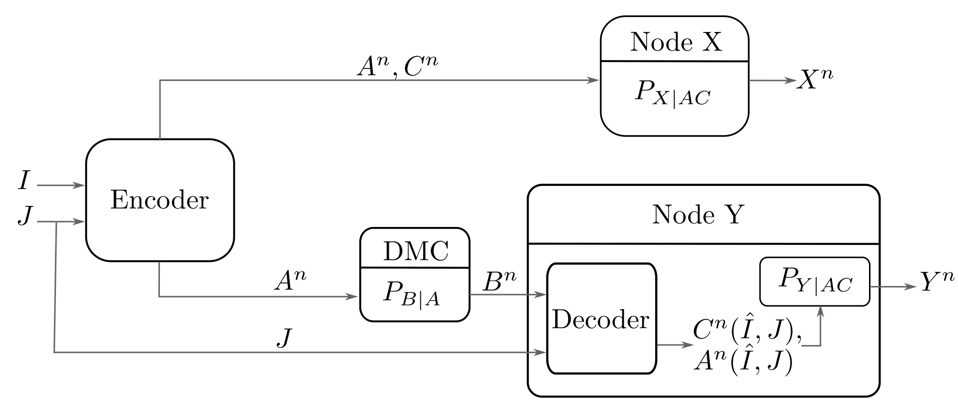

Figure 4.5 Allied structure of separation-based code

- Let the channel codebook $\mathcal{A}$ be the same size as coordination codebook $\mathcal{C}$. That can be obtained by setting $R_a = 0, |K| = 1$ (i.e., a one-by-one mapping between the codebook $\mathcal{A}$ and $\mathcal{C}$ in Figure 4.3).

- At node Y, estimate the index $\hat{I}$ from the output of the DMC B^n. The decoding is based on jointly typical sequences with respect to the channel input A^n and side information from the common randomness message J. Use the estimated index $\hat{I}$ to reconstruct both codewords $A^n_{\hat{i}j}$ and $C^n_{\hat{i}j}$.

- Pass the reconstructed codewords through a memoryless channel $P_{Y|AC}(y^n|a^n, c^n)$ to obtain the output sequence Y^n.

As a consequence, this will induce the following distribution

$$\begin{aligned} P_{X^nY^n}(x^n, y^n) &\triangleq \sum_{b^n,i,j} 2^{-nR} P_{X|AC}(x^n|A^n_{ij}, C^n_{ij}) P_{B|A}(B^n|A^n) P_{Y|AC}(y^n|A^n_{\hat{i}j}, C^n_{\hat{i}j}) \\ &= \sum_{i,j} 2^{-nR} P_{X|AC}(x^n|A^n_{ij}, C^n_{ij}) P_{Y|AC}(y^n|A^n_{\hat{i}j}, C^n_{\hat{i}j}), \end{aligned} \tag{4.5}$$

where $R \triangleq R_o + R_c$.

Clearly, since this joint distribution is similar to the one for the general scheme, it can be shown that the following rate conditions are sufficient to achieve strong coordination:

$$\left\{ \begin{array}{ll} R_o + R_c \geq I(XY;AC) + \delta_1 + \delta_2, & \text{channel resolvability constraint} \\ R_c \geq I(X;AC), & \text{secrecy constraint} \\ R_c < I(A;B), & \text{decodability constraint} \\ \rho_1 \geq R_c - I(X;AC) - \delta_1, & \text{index randomization at node X} \\ \rho_2 \geq H(Y|AC) - \delta_2, & \text{sequence generation at node Y} \end{array} \right\}. \tag{4.6}$$

4.5 Hybrid Coordination-Channel Coding Scheme

Finally, a hybrid scheme is formed by altering the allied problem of the general scheme similarly to the improved separate scheme as depicted in Figure 4.6. However, this scheme is considered as hybrid because it benefits from the DMC output B^n. Compared with the improved separate scheme in Section 4.4, the test channel at node Y incorporates the output of the channel B^n in generating the action sequence Y^n. As a result, the channel coding layer is not completely separated. This is accomplished by passing the DMC output sequence B^n along with the reconstructed codewords $A^n_{\hat{i}j}$ and $C^n_{\hat{i}j}$ through the channel $P_{Y|ABC}(y^n|A^n, B^n, C^n)$. The induced joint distribution is as follows:

$$P_{X^nY^n}(x^n, y^n) \triangleq \sum_{b^n,i,j} 2^{-nR} P_{X|AC}(x^n|A^n_{ij}, C^n_{ij}) P_{B|A}(B^n|A^n) P_{Y|ABC}(y^n|A^n_{\hat{i}j}, B^n, C^n_{\hat{i}j}), \tag{4.7}$$

where $R \triangleq R_o + R_c$.

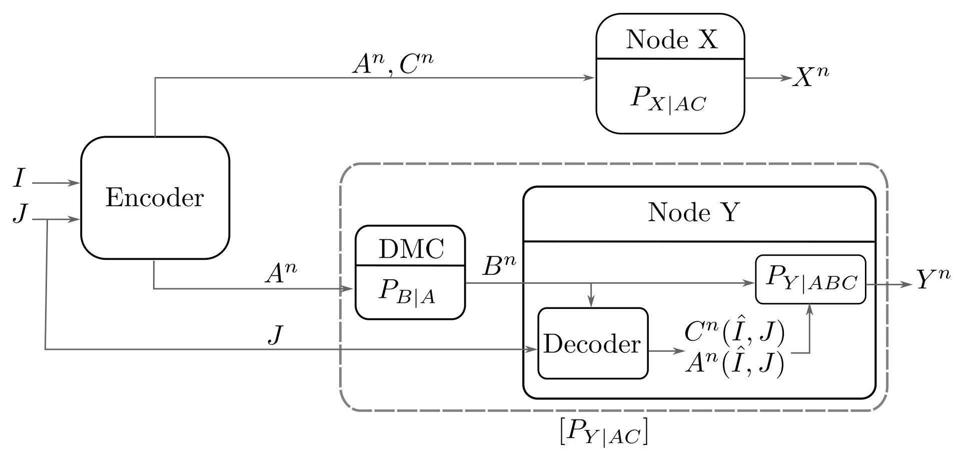

Figure 4.6 Structure of the hybrid coordination-channel coding scheme.

This process only affects the action generation function at node Y. Specifically, the amount of local randomness required to simulate the channel $P_{Y|ABC}(y^n|a^n, b^n, c^n)$

is given as:

$$\left\{ \begin{array}{rl} R_o + R_c \geq I(XY;AC) + \delta_1 + \delta_2, & \text{channel resolvability constraint} \\ R_c \geq I(X;AC), & \text{secrecy constraint} \\ R_c < I(B;C), & \text{decodability constraint} \\ \rho_1 \geq R_c - I(X;AC) - \delta_1, & \text{index randomization at node X} \\ \rho_2 \geq H(Y|ABC) - \delta_2 & \text{sequence generation at node Y} \end{array} \right\}. \tag{4.8}$$

4.6 Observations

Here, we will prove that the two-stage basic separate coordination-channel coding method discussed in Section 4.3 is not necessarily as good as the joint coding method proposed by the general scheme. In other words, we prove that the general scheme is able to strictly provide smaller rates to achieve strong coordination. To prove that we show that the basic scheme is a subset of the proposed joint scheme, and hence, the coordination rate region of the former is a subregion of the latter scheme. Theorem 8 outlines this inclusion of the rate regions.

Theorem 8 (Rate region inclusion)**.** *For the point-to-point coordination over a DMC, the relation between the rate region of the joint, hybrid, improved separate and basic separate schemes can be expressed as*

$$\underline{\mathcal{R}}_{basic\ sep.} \subseteq \underline{\mathcal{R}}_{improved\ sep.} \subseteq \underline{\mathcal{R}}_{hybrid} \subseteq \underline{\mathcal{R}}_{joint} \tag{4.9}$$

Proof. First we see that the basic scheme is constructed as a special case of the improved separate coordination-channel coding scheme by choosing

$$
\left\{
\begin{array}{ccc}
\text{Separate/Hybrid} & & \text{Basic} \\
X - (A, C) - Y & & X - U - Y \\
A & = & A^* \\
C & = & U \\
P_{AC} & = & P_A P_U \\
P_{X|AC} & = & P_{X|U} \\
P_{Y|AC} & = & P_{Y|U}
\end{array}
\right\}. \tag{4.10}
$$

Similarly, we can construct both separate and hybrid schemes from the joint coordination-channel coding scheme. Choose

$$
\left\{
\begin{array}{ccc}
\text{Joint} & & \text{Separate/Hybrid} \\
|K| \geq 1 & & |K| = 1 \\
A' & = & A \\
C' & = & (A, C)
\end{array}
\right\}. \tag{4.11}
$$

Based on the pair-wise construction of schemes, this suggests that the basic separation scheme is constructed as a special case of the presented joint scheme. As a result, the rates that are achievable by the basic separate scheme are included in the rate region of the proposed joint scheme. ■

Finally, we derive the criteria to compare the rate regions of the presented coding schemes. Although the rate constraints specified by the equation sets (4.3), (4.4), (4.6), and (4.8) incorporate six variables representing rates, in this study, we are in particular interested in the randomness rates, both local and common. Therefore, we deploy Fourier–Motzkin elimination [29] on the rate constraints of the mentioned coding schemes to obtain rate constraints only involving the three randomness rates. Appendix A details the Fourier-Motzkin elimination steps to obtain the rate regions shown in Table 4.1.

Table 4.1 Randomness Rate Regions for the Coordination-channel Coding Schemes

	Joint coding	*Hybrid coding*	*Improved separation*	*Basic separation*
$\lambda_1 \geq$	$I(Y;A'C'\|X)$ $+H(Y\|B'C')$	$I(Y;AC\|X)$ $+H(Y\|BC)$	$I(Y;AC\|X)$ $+H(Y\|AC)$	$H(Y\|X)$
$\lambda_2 \geq$	$I(XY;C')$ $+H(Y\|B'C')$ $-I(B';C')$	$I(XY;AC)$ $+H(Y\|BC)$ $-I(A;B)$	$I(XY;AC)$ $+H(Y\|AC)$ $-I(A;B)$	$I(XY;U)$ $+H(Y\|U)$ $-I(A;B)$
$\lambda_3 \geq$	$I(Y;A'C'\|X)$	$I(Y;AC\|X)$	$I(Y;AC\|X)$	$I(Y;U\|X)$
$\lambda_4 \geq$	$I(XY;C')$ $-I(B';C')$	$I(XY;AC)$ $-I(A;B)$	$I(XY;AC)$ $-I(A;B)$	$I(XY;U)$ $-I(A;B)$
	$I(X;C')$ $< I(B';C')$	$I(X;AC)$ $< I(A;B)$	$I(X;AC)$ $< I(A;B)$	$I(X;U)$ $< I(A^*;B^*)$

where

$$\left\{ \begin{aligned} &\lambda_1 \triangleq R_0 + \rho_1 + \rho_2, \\ &\lambda_2 \triangleq R_o + \rho_2, \\ &\lambda_3 \triangleq R_o + \rho_1, \\ &\lambda_4 \triangleq R_0. \end{aligned} \right\}.$$

In principle, Theorem 8 indicates that Shannon's separation is not necessarily optimal (i.e., by providing smaller rates for the schemes on the LHS of equation (4.9) compared to schemes further on the RHS of this equation). In the following Chapter 5, we will illustrate the mentioned rate region inclusion of Theorem 8 with an example.

CHAPTER 5
EXAMPLE AND DISCUSSION

In this chapter, we compute the rate region for the hybrid scheme and the basic separation scheme with an example, and then demonstrate simulation results highlighting the observations made in Section 4.6.

5.1 Binary Symmetric Action Distribution over Binary Symmetric Common Channel

For a Bernoulli $(\frac{1}{2})$ source X, a joint distribution $p_{XY}(x,y)$, and a common communication channel $P_{B|A}$, we derive the common and local randomness region for the basic strong coordination-channel separation scheme associated with synthesizing the target joint distribution. Here, the communication channel between the two nodes is a BSC with a cross over probability p_o and the joint target behavior is $p_{XY}(x,y) = p_X(x)p_{Y|X}(y|x)$ where $p_{Y|X}(y|x)$ is described as Binary Symmetric Channel (BSC) with cross over probability p. For this setup, we assume the alphabet of the auxiliary random variable C used in generating the coordination codebook is binary[2] (*i.e.*, $|\mathcal{C}| = 2$). Accordingly, we construct the basic separation scheme to acquire the randomness rate region and, for comparison, the hybrid coordination-channel coding scheme where we derive a single operating point on its randomness rate region to demonstrate that the latter scheme is able to provide strictly better rates than the basic separation scheme. Both the hybrid scheme and the basic separation schemes are built following the guidelines given in (4.10).

5.1.1 Basic Separation Scheme

To build the basic separation scheme for the setup, we follow a similar approach as in [2]. Here, the joint distribution $P_X(x)P_{U|X}(u|x)P_{Y|U}(y|u)$ that produce the

[2]Although the alphabet size of the auxiliary random variable C may not be optimal, it provides a reasonable basis for comparing the two mentioned schemes.

strong coordination capacity region is formed by cascading two BSC as illustrated in Figure 5.1 forming the Markov chain $X - U - Y$ where

$$p_2 \in [0, p], \quad p_1 = \frac{p - p_2}{1 - 2p_2}.$$

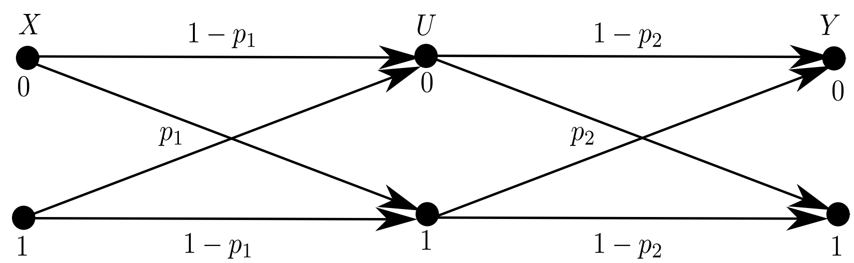

Figure 5.1 Decomposition of the joint target behavior $BSC(p)$ for the basic separation coding scheme.

The mutual information terms for the capacity region from the equation set (4.4) can be written as follows

$$\begin{aligned} I(X;U) =& 1 - H(p_1), \\ I(XY;U) =& 1 + H(p) - H(p_1) - H(p_2), \\ I(A;B) =& 1 - H(p_o). \end{aligned}$$

By applying these quantities to the local and common randomness region from Table 4.1 we obtain the following rate constraints:

$$R_o + \rho_1 + \rho_2 \geq H(p), \tag{5.1}$$

$$R_o + \rho_1 \geq H(p) - H(p_2), \tag{5.2}$$

$$R_o + \rho_2 > H(p) - H(p_1) + H(p_o), \tag{5.3}$$

$$R_o + \rho_2 > H(p) - H(p_1) - H(p_2) + H(p_o), \tag{5.4}$$

$$H(p_1) > H(p_o). \tag{5.5}$$

5.1.2 Hybrid Scheme

On the other hand, the hybrid scheme described in Section 4.5 is constructed in two stages. First, we find a joint distribution $P_{AC}(a, c)$ over the channel codebook $\mathcal{A}$ and the coordination codebook $\mathcal{C}$. The hybrid scheme correlates the codebooks such that the decodability constraint (i.e., $R_c < I(B; C)$) from (4.6) is satisfied. Specifically, we find the joint distribution $P_C(c)P_{A|C}(a|c)$ that maximizes the mutual information between the coordination codeword C_{ij}^n and the channel output sequence B^n corresponding to the channel input codeword A_{ij}^n. Illustrated in Figure 5.2 is the channel associated with such a distribution where

$$\begin{aligned}\text{Capacity(decodability channel)} &= \max_{P_C(c)P_{A|C}(a|c)} I(B; C) \\ &= \max_{P_C(c)P_{A|C}(a|c)} \big(H(B) - H(B|C)\big) \\ &= 1 - H(p_o) - H(\tilde{p}).\end{aligned} \tag{5.6}$$

The maximum in (5.6) is achieved when $P_C(c)$ is uniformly distributed and if $\tilde{p} = 0$.

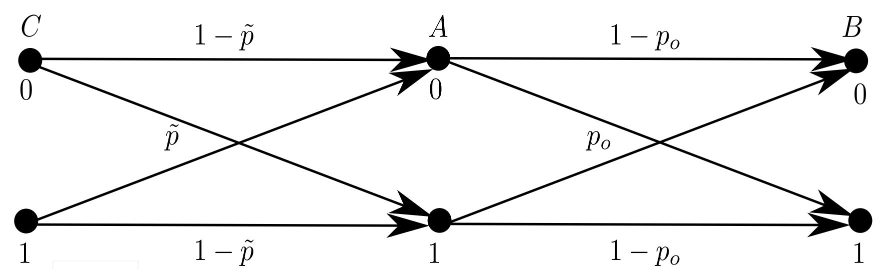

Figure 5.2 Codebook correlation channel.

Then, the joint distribution $P_X(x)P_{CA|X}(c, a|x)P_{B|A}(b|a)P_{Y|BC}(y|b, c)$ from (4.7) that produces the boundary of the strong coordination region for the hybrid scheme is formed by cascading two BSCs and a uniformly focusing channel. Recall that in the hybrid scheme, we only transmit the codeword A_{ij}^n over the common channel and then perform full decoding to obtain $C_{\hat{i},j}^n$ at the other end as described in Section

4.5. This process forms the Markov chain $X - (C, A) - (C, B) - Y$ characterized in Figure 5.3 where $\alpha \in [0, 1], \quad \beta \in [0, 1]$.

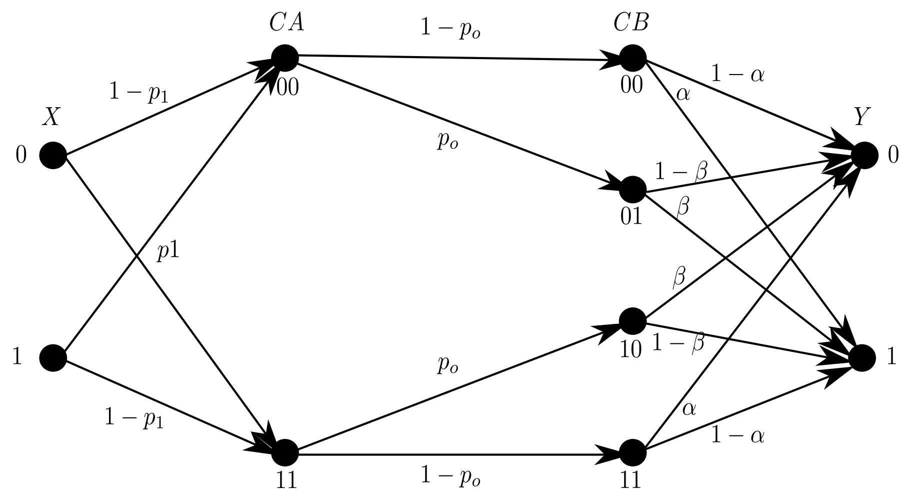

Figure 5.3 Hybrid scheme channel simulation of $BSC(p)$.

We can simplify the channel synthesis for the hybrid scheme to two cascading BSCs similarly to Figure 5.1 by setting

$$p_2 \triangleq (1 - p_o)\alpha + p_o\beta.$$

Then, the mutual information terms for the capacity region from (4.6) are written as

$$\begin{aligned}
I(X; AC) &= 1 - H(p_1), \\
I(XY; AC) &= 1 + H(p) - H(p_1) - H(p_2), \\
I(B; C) &= 1 - H(p_o).
\end{aligned}$$

As a result, the sum of the achievable rates for the hybrid scheme rate region from Table 4.1 becomes

$$\begin{aligned} R_o + \rho_1 + \rho_2 &\geq I(Y;AC|X) + H(Y|BC) \\ &= H(Y|X) - H(Y|XAC) + H(Y|BC) \\ &\overset{(a)}{=} H(Y|X) - H(Y|AC) + H(Y|BC) \\ &= H(p) - H(p_2) + (1-p_o)H(\alpha) + p_o H(\beta), \end{aligned} \tag{5.7}$$

where (a) is a result of the Markov chain $X - (A, C) - Y$.

Similarly we obtain the rest of the rate constraints as follows:

$$R_o + \rho_2 > H(p) - H(p_1) - H(p_2) + H(p_o) + (1-p_o)H(\alpha) + p_o H(\beta), \tag{5.8}$$

$$R_o + \rho_1 \geq H(p) - H(p_2), \tag{5.9}$$

$$R_o > H(p) - H(p_1) - H(p_2) + H(p_o), \tag{5.10}$$

$$H(p_1) > H(p_o). \tag{5.11}$$

On the one hand, the overall achievable randomness from (5.7) for coordinating the target p_{XY} distribution is characterized by the parameters p_2, α, and β. To find the minimum achievable sum rate we optimize over the selection of the above mentioned parameters representing $P_{CA|X}(c, a|x)$ and $P_{Y|BC}(y|b, c)$ from (4.7), respectively as follows:

$$\begin{aligned} R_o + \rho_1 + \rho_2 &\geq \min_{p_2,\alpha,\beta} \big(H(p) - H(p_2) + (1-p_o)H(\alpha) + p_o H(\beta)\big) \\ &= H(p) - \max_{p_2,\alpha,\beta} \big(H(p_2) - (1-p_o)H(\alpha) - p_o H(\beta)\big). \end{aligned} \tag{5.12}$$

On the other hand, for the basic separation scheme from (5.1) we have

$$R_o + \rho_1 + \rho_2 \geq H(p).$$

By assumption, the hybrid scheme parameters p_2, α, and β can also be selected to simulate separation. Specifically, by selecting:

$$\alpha = \beta,$$

then

$$\begin{aligned} p_2 =& (1-p_o)\beta + p_o\beta = \beta, \\ (R_o+\rho_1+\rho_2)_{\text{Hybrid}} =& H(p) - H(\beta) + (1-p_o)H(\beta) + p_oH(\beta) \\ =& H(p) \\ =& (R_o+\rho_1+\rho_2)_{\text{Basic Separation}}. \end{aligned}$$

As a result, at this point where $\alpha = \beta = p_2$ the hybrid scheme performance will be identical to the basic separation scheme. This supports our claim in Section 4.6 that the rate region for the basic separation scheme is strictly included in the rate region of the hybrid scheme by comparing (5.1) and (5.12).

5.1.3 Simulation Results

For an example suppose that a doubly binary symmetric source $p_{XY}(x, y)$ with $p_{Y|X} = BSC(0.4)$ and a common noisy channel of $P_{B|A} = BSC(0.3)$ are selected to investigate both the obtained randomness region for the basic separate scheme and a single operating point of the hybrid scheme which provides the minimum in (5.12). This sample is collected by performing an exhaustive search over p_2, α, and β such that the second term on the RHS of (5.12) is maximized. Figure 5.4 shows both the boundary of the basic separation scheme region and the single operating point of the hybrid scheme, and Figure 5.5 presents a projection to the $(R_o\ , \rho_2)$ plane of Figure 5.4.

As can be seen from Figure 5.4, the hybrid scheme sample provides a smaller sum rate $(R_o + \rho_1 + \rho_2)$. Comparing the single operating point of the hybrid scheme with the corresponding operating point (obtained by the same value of the parameter

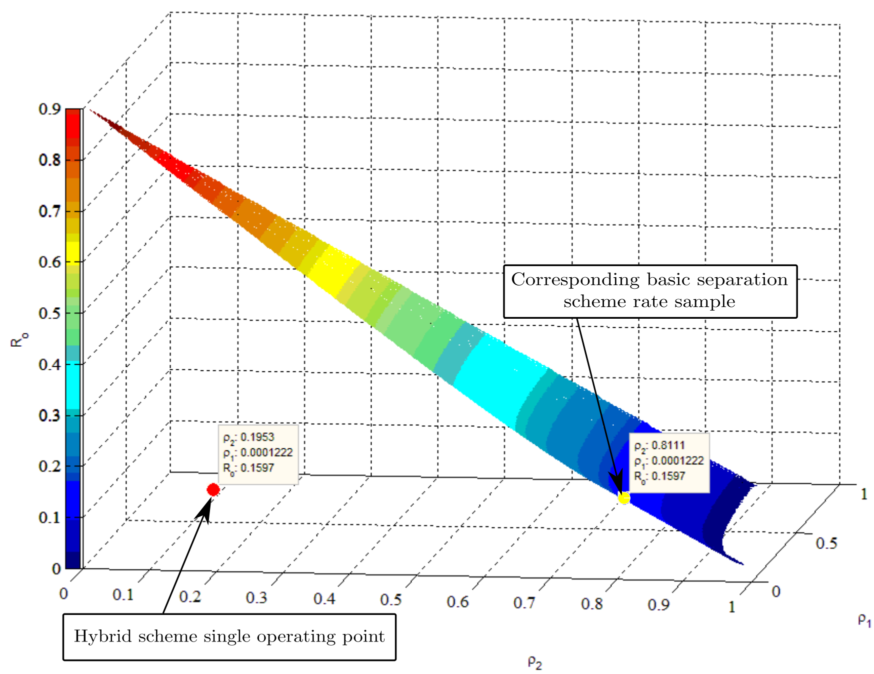

Figure 5.4 Channel randomness region for the separate scheme for a target distribution modeled as $p_{Y|X} = BSC(0.4)$.

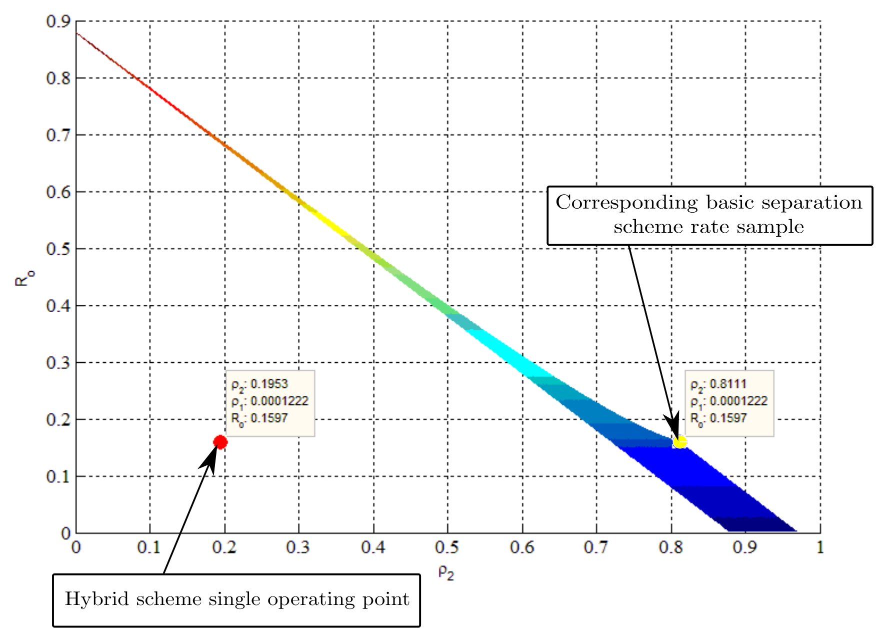

Figure 5.5 Channel randomness region for the separate scheme and $p_{Y|X} = BSC(0.4)$ projected on the $(R_o\ ,\ \rho_2)$ plane of Figure 5.4.

p_2) for the separation based scheme yields a difference of about 0.6158 bits in terms of sum rate. The values of β, α, and p_2 returned by the exhaustive search producing this hybrid scheme sample are $\frac{5}{6}, 0$, and $\frac{1}{4}$, respectively. Hence, the hybrid scheme provides better rates than the basic separation scheme and that breaks down the coordination-channel separation.

This simulation has been repeated for a couple of communication BSCs with different values of the cross over probability p_o for a fixed $p = 0.4$. Figure 5.6 and Figure 5.7 present a comparison between the minimum local randomness ρ_2 required at node Y achieved by both hybrid and separate schemes verses the sum rate $R_o + \rho_1$.

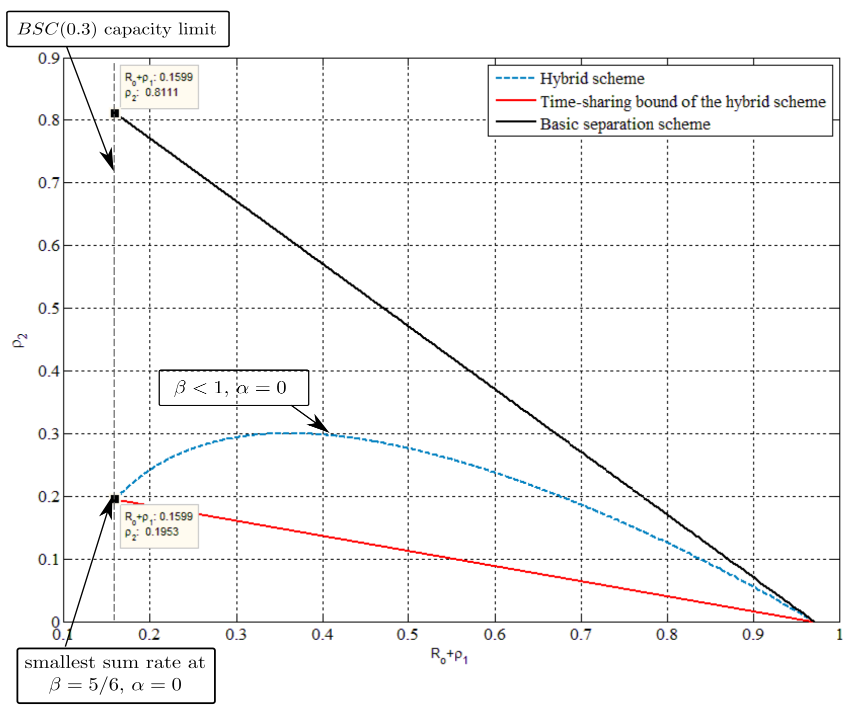

Figure 5.6 Combined rates $R_o + \rho_1$ vs. local randomness rate ρ_2 for $p_{Y|X} = BSC(0.4)$ and $P_{B|A} = BSC(0.3)$.

Comparing between the two randomness regions of the hybrid scheme for different BSCs depicted in Figures 5.6 and 5.7, it can be noticed that for bigger BSC cross over probabilities the minimum sum rate does not necessary decrease. This behavior is due to the capacity limitation imposed by the communication channel.

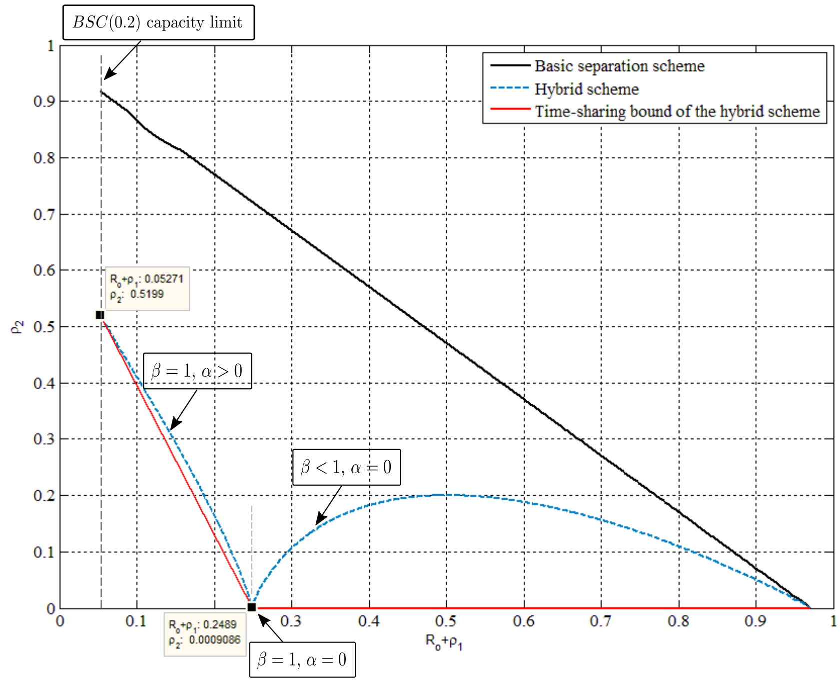

Figure 5.7 Combined rates $R_o + \rho_1$ vs. local randomness rate ρ_2 for $p_{Y|X} = BSC(0.4)$ and $P_{B|A} = BSC(0.2)$.

However, time-sharing can be used to achieve any point on the chord between two sum rate achieving pairs as illustrated with the red curve in Figures 5.6 and 5.7.

Finally, Figure 5.8 demonstrates how the cross over probability p_o affects the minimum randomness sum rate $(R_0 + \rho_1 + \rho_2)$ for the basic separation and hybrid schemes and $p_{Y|X} = BSC(0.4)$. It can be noticed that as p_o increases, the sum rate of the hybrid scheme decreases until the point where $p_o = \frac{1-\sqrt{1-2P}}{2} = 0.2764$. After this point, the channel provides too much randomness for simulating the given doubly binary symmetric source. As a result, the parameter β in Figure 5.3 is decreased (i.e, $\beta < 1$) to reduce the channel randomness propagating to node Y. As the cross over probability p_o increases after this point, the required total randomness of the hybrid scheme approaches the one for the basic separation scheme again.

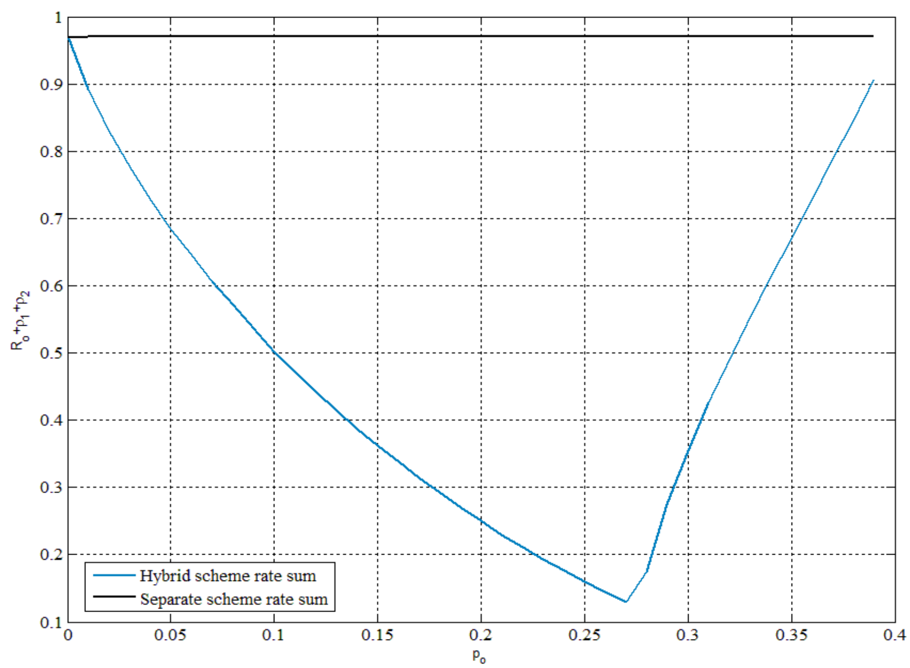

Figure 5.8 Comparison of required randomness sum rates for separate and hybrid scheme for $p_{Y|X} = BSC(0.4)$ versus the BSC cross over probability p_o.

CHAPTER 6
CONCLUSION

The source-channel separation theorem states that we can construct an optimal code to represent a source using a finite set of messages (i.e., lossy source compression), and independently design an optimal channel code to reliably communicate these message while maintaining a desired minimum distortion. This theorem is a classical result of the information theory that was established by Shannon and an essential concept for reliable point-to-point communication. In this thesis, we have investigated a fundamental question regarding communication-based coordination: Is Shannon's separation theorem optimal in the context of point-to-point coordination networks? In other words, is designing a coordination code separately from a channel code the best strategy? For this purpose, we have studied the problem of coordination over noisy communication links. In particular, we considered the two-node network setup with a discrete memoryless channel as the communication link. We have examined the strong coordination notation where the induced joint distribution of action sequences selected by the nodes is close to a target distribution. Specifically, the induced distribution is statistically indistinguishable from the target distribution. To achieve strong coordination, the coding scheme must induce a distribution that simulate the target joint distribution with tolerance of an arbitrary small error measured by total variation.

To that extent, we derive achievability for the two-node coordination network with a noisy communication link and construct a general joint coding scheme. The proof technique underlying our general coding scheme is based on channel resolvability, a technique which is widely used in analyzing strong coordination problems. We derive the necessary constraints over the rates of communication, local randomness, and common randomness to achieve strong coordination over the mentioned setup. In particular, we focus on the trade-off between the required rates of common and local randomness.

Moreover, we have shown that from our general scheme, three special cases can be constructed. Specifically, a hybrid coding scheme, an improved separation scheme and the basic separation scheme that follows Shannon's separation principle. Based on this construction, we have established that the set of achievable rates of the basic separation scheme is a subset of the rates achieved by the joint scheme. As a result, we conclude that the general scheme which jointly solves the coordination-channel problem is able, in fact, to provide better performance in terms of a larger achievable rate region. Specifically, a smaller sum of common and local randomness rate is obtained by joint encoding and decoding. Surprisingly, this result breaks the classical information theoretic concept that was established by Shannon and summarizes the main contribution of this thesis. Finally, a simple example has been studied showing the construction of a hybrid scheme and a basic separation scheme for coordinating a doubly binary symmetric source over a binary symmetric communication channel. The example is followed by some simulation results. The simulation results validate our claim by showing that the hybrid scheme is capable of providing a smaller randomness sum rate compared with the basic separation scheme.

Although this work yields some insight in coordination over noisy communication links, there are still open questions that can be addressed. For example, a convex proof that would establish that the presented joint coding scheme is indeed the optimal scheme is absent from this thesis. Further, this work can be extended to more complicated three-terminal network setups including relay, broadcast, and multiple access networks. Finally, the work can be extended to code implementation. In particular, based on the success of poler codes in achieving strong coordination via channel resolvability, one natural extension is implementing polar codes for the noisy two-node network following our general construction and to determine whether the same rates can be guaranteed asymptotically.

APPENDIX A
FOURIER-MOTZKIN ELIMINATION

For the purpose of comparing between the achievable schemes from Chapter 4, we characterize each scheme in terms of local and common randomness rates. In this appendix we perform the Fourier-Motzkin variable elimination technique [29] in order to present the randomness rate region sets as listed in Table 4.1 by eliminating the rate-transfer variables δ_1, δ_2 and the communication rate R_c from the rate regions as follows

A.1 Randomness Rate Region for the Basic Separation Scheme

From the equation set (4.4) we have

$$R_o + R_c \geq I(XY;U) + \delta_1 + \delta_2, \tag{A.1}$$

$$R_c \geq I(X;U), \tag{A.2}$$

$$-R_c > -I(A;B), \tag{A.3}$$

$$\rho_1 \geq R_c - I(X;U) - \delta_1, \tag{A.4}$$

$$\rho_2 \geq H(Y|U) - \delta_2, \tag{A.5}$$

where

$$\delta_1 \geq 0, \tag{A.6}$$

$$\delta_2 \geq 0. \tag{A.7}$$

Steps:

1. Add (A.4) and (A.5) to (A.1):

$$R_o + R_c + \rho_1 + \rho_2 \geq I(XY;U) + \delta_1 + \delta_2, +R_c - I(X;U) - \delta_1 + H(Y|U) - \delta_2,$$

$$R_o + \rho_1 + \rho_2 \geq I(XY;U) - I(X;U) + H(Y|U),$$

$$= I(Y;U|X) + H(Y|U),$$
$$= H(Y|X) - H(Y|UX) + H(Y|U),$$
$$= H(Y|X) - H(Y|U) + H(Y|U),$$
$$= H(Y|X). \tag{A.8}$$

2. Add (A.6), (A.7), and (A.3) to (A.1):

$$R_o + R_c + \delta_1 + \delta_2 - R_c > I(XY;U) + \delta_1 + \delta_2 - I(A;B),$$
$$R_o > I(XY;U) - I(A;B). \tag{A.9}$$

3. Add (A.4) and (A.7) to (A.1):

$$R_o + R_c + \rho_1 + \delta_2 \geq I(XY;U) + \delta_1 + \delta_2 + R_c - I(X;U) - \delta_1,$$
$$R_o + \rho_1 \geq I(XY;U) - I(X;U),$$
$$= I(Y;U|X). \tag{A.10}$$

4. Add (A.5), (A.6), and (A.3) to (A.1):

$$R_o + R_c + \rho_2 + \delta_1 - R_c > I(XY;U) + \delta_1 + \delta_2 + H(Y|U) - \delta_2 - I(A;B),$$
$$R_o + \rho_2 > I(XY;U) + H(Y|U) - I(A;B). \tag{A.11}$$

5. Add (A.2) to (A.3):

$$R_c - R_c > I(X;U) - I(A;B),$$
$$I(A;B) > I(X;U). \tag{A.12}$$

A.2 Randomness Rate Region for the Separate Scheme

For the separate scheme we have from the equation set (4.6)

$$R_o + R_c \geq I(XY; AC) + \delta_1 + \delta_2, \tag{A.13}$$

$$R_c \geq I(X; AC), \tag{A.14}$$

$$-R_c > -I(A; B), \tag{A.15}$$

$$\rho_1 \geq R_c - I(X; AC) - \delta_1, \tag{A.16}$$

$$\rho_2 \geq H(Y|AC) - \delta_2. \tag{A.17}$$

Steps:

1. Add (A.16) and (A.17) to (A.13):

$$R_o + R_c + \rho_1 + \rho_2 \geq I(XY; AC) + \delta_1 + \delta_2 + R_c - I(X; AC) - \delta_1 + H(Y|BC) - \delta_2,$$

$$\begin{aligned} R_o + \rho_1 + \rho_2 &\geq I(XY; AC) - I(X; AC) + H(Y|AC), \\ &= I(Y; AC|X) + H(Y|AC), \\ &= H(Y|X). \end{aligned} \tag{A.18}$$

2. Add (A.6), (A.7), and (A.15) to (A.13):

$$R_o + R_c + \delta_1 + \delta_2 - R_c > I(XY; AC) + \delta_1 + \delta_2 - I(A; B),$$

$$R_o > I(XY; AC) - I(A; B). \tag{A.19}$$

3. Add (A.16) and (A.7) to (A.13):

$$R_o + R_c + \rho_1 + \delta_2 \geq I(XY; AC) + \delta_1 + \delta_2 + R_c - I(X; AC) - \delta_1,$$

$$\begin{aligned} R_o + \rho_1 &\geq I(XY; AC) - I(X; AC), \\ &= I(Y; AC|X). \end{aligned} \tag{A.20}$$

4. Add (A.17), (A.6) and (A.15) to (A.13):

$$R_o + R_c + \rho_2 + \delta_1 - R_c > I(XY;AC) + \delta_1 + \delta_2 + H(Y|AC) - \delta_2 - I(A;B),$$

$$R_o + \rho_2 > I(XY;AC) + H(Y|AC) - I(A;B). \tag{A.21}$$

5. Add (A.14) to (A.15):

$$R_c - R_c > I(X;AC) - I(A;B),$$

$$I(A;B) > I(X;AC). \tag{A.22}$$

A.3 Randomness Rate Region for the Hybrid Scheme

Similarly for the hybrid scheme we have from the equation set (4.6)

$$R_o + R_c \geq I(XY;AC) + \delta_1 + \delta_2, \tag{A.23}$$

$$R_c \geq I(X;AC), \tag{A.24}$$

$$-R_c > -I(A;B), \tag{A.25}$$

$$\rho_1 \geq R_c - I(X;AC) - \delta_1, \tag{A.26}$$

$$\rho_2 \geq H(Y|BC) - \delta_2. \tag{A.27}$$

Steps:

1. Add (A.25) and (A.26) to (A.23):

$$R_o + R_c + \rho_1 + \rho_2 \geq I(XY;AC) + \delta_1 + \delta_2 + R_c - I(X;AC) - \delta_1 + H(Y|BC) - \delta_2,$$

$$\begin{aligned} R_o + \rho_1 + \rho_2 &\geq I(XY;AC) - I(X;AC) + H(Y|BC), \\ &= I(Y;AC|X) + H(Y|BC), \\ &= H(Y|X) - H(Y|AC) + H(Y|BC). \end{aligned} \tag{A.28}$$

2. Add (A.6), (A.7), and (A.25) to (A.23):

$$
\begin{aligned}
R_o + R_c + \delta_1 + \delta_2 - R_c &> I(XY;AC) + \delta_1 + \delta_2 - I(A;B), \\
R_o &> I(XY;AC) - I(A;B).
\end{aligned}
\tag{A.29}
$$

3. Add (A.25) and (A.7) to (A.23):

$$
\begin{aligned}
R_o + R_c + \rho_1 + \delta_2 &\geq I(XY;AC) + \delta_1 + \delta_2 + R_c - I(X;AC) - \delta_1, \\
R_o + \rho_1 &\geq I(XY;AC) - I(X;AC), \\
&= I(Y;AC|X).
\end{aligned}
\tag{A.30}
$$

4. Add (A.26), (A.6), and (A.25) to (A.23):

$$
\begin{aligned}
R_o + R_c + \rho_2 + \delta_1 - R_c &> I(XY;AC) + \delta_1 + \delta_2 + H(Y|CB) - \delta_2 - I(A;B), \\
R_o + \rho_2 &> I(XY;AC) + H(Y|CB) - I(A;B).
\end{aligned}
\tag{A.31}
$$

5. Add (A.24) to (A.25):

$$
\begin{aligned}
&R_c - R_c > I(X;AC) - I(A;B), \\
&I(A;B) > I(X;AC).
\end{aligned}
\tag{A.32}
$$

A.4 Randomness Rate Region for the Joint Scheme

Finally, for the joint scheme, from the equation set 4.3 we have

$$R_o + R_c + R_a \geq I(XY;AC) + \delta_1 + \delta_2, \tag{A.33}$$

$$R_o + R_c \geq I(XY;C) + \delta_1 + \delta_2, \tag{A.34}$$

$$R_c + R_a \geq I(X;AC), \tag{A.35}$$

$$R_c \geq I(X;C), \tag{A.36}$$

$$-R_c > -I(B;C), \tag{A.37}$$

$$\rho_1 \geq R_c + R_a - I(X;AC) - \delta_1, \tag{A.38}$$

$$\rho_2 \geq H(Y|BC) - \delta_2. \tag{A.39}$$

Steps:

1. To eliminate δ_1 and δ_2 from (A.33) add (A.38) and (A.39) to (A.33):

$$\begin{aligned} R_o + R_c + R_a + \rho_1 + \rho_2 &\geq I(XY;AC) + \delta_1 + \delta_2 + R_c + R_a - I(X;AC) - \delta_1 \\ &\quad + H(Y|BC) - \delta_2, \\ R_o + \rho_1 + \rho_2 &\geq I(XY;AC) - I(X;AC) + H(Y|BC), \\ &= I(Y;AC|X) + H(Y|BC), \\ &= H(Y|X) - H(Y|AC) + H(Y|BC). \end{aligned} \tag{A.40}$$

2. Add (A.7), (A.38) to (A.33):

$$\begin{aligned} R_o + R_c + R_a + \rho_1 + \delta_2 &\geq I(XY;AC) + \delta_1 + \delta_2 + R_c + R_a - I(X;AC) - \delta_1, \\ R_o + \rho_1 &\geq I(XY;AC) - I(X;AC), \\ &= I(Y;AC|X). \end{aligned} \tag{A.41}$$

3. Add (A.6), (A.37), and (A.39) to (A.34):

$$R_o + R_c + \rho_2 + \delta_1 - R_c > I(XY;C) + \delta_1 + \delta_2 + H(Y|CB) - \delta_2 - I(B;C),$$
$$R_o + \rho_2 > I(XY;C) + H(Y|CB) - I(B;C). \tag{A.42}$$

4. Add (A.6), (A.7) and (A.37) to (A.34):

$$R_o + R_c + \delta_2 + \delta_1 - R_c > I(XY;C) + \delta_1 + \delta_2 - I(B;C),$$
$$R_o > I(XY;C) - I(B;C). \tag{A.43}$$

5. Add (A.36) to (A.37):

$$R_c - R_c > I(X;C) - I(B;C),$$
$$I(B;C) > I(X;C). \tag{A.44}$$

Notice that the Fourier-Motzkin elimination steps concerning the rest of the rate region inequalities for the joint scheme result in a redundant set and thus they are not mentioned here.

REFERENCES

[1] B. Prabhakar, K. Dektar, and D. M. Gordon, "The regulation of ant colony foraging activity without spatial information," *PLOS Computational Biology*, vol. 8, no. 8, pp. 1–7, Aug. 2012.

[2] P. Cuff, "Communication in networks for coordinating behavior," Ph.D. dissertation, Stanford University, 2009.

[3] P. W. Cuff, H. H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4181–4206, 2010.

[4] E. Soljanin, "Compressing quantum mixed-state sources by sending classical information," *IEEE Transactions on Information Theory*, vol. 48, no. 8, pp. 2263–2275, Aug. 2002.

[5] M. R. Bloch and J. Kliewer, "Strong coordination over a three-terminal relay network," in *Proc. IEEE Information Theory Workshop*, Hobart, Australia, Nov. 2014, pp. 646–650.

[6] B. N. Vellambi, J. Kliewer, and M. R. Bloch, "Strong coordination over multi-hop line networks," in *Proc. IEEE Information Theory Workshop*, Jeju Island, Korea, Oct. 2015, pp. 192–196.

[7] A. Bereyhi, M. Bahrami, M. Mirmohseni, and M. R. Aref, "Empirical coordination in a triangular multi-terminal network," in *Proc. IEEE International Symposium on Information Theory*, Istanbul, Turkey, Jul. 2013, pp. 2149–2153.

[8] A. A. Gohari and V. Anantharam, "Generating dependent random variables over networks," in *Proc. IEEE Information Theory Workshop*, Paraty, Brazil, Oct. 2011, pp. 698–702.

[9] M. H. Yassaee, A. Gohari, and M. R. Aref, "Channel simulation via interactive communications," *IEEE Transactions on Information Theory*, vol. 61, no. 6, pp. 2964–2982, 2015.

[10] F. Haddadpour, M. H. Yassaee, A. Gohari, and M. R. Aref, "Coordination via a relay," in *Proc. IEEE International Symposium on Information Theory*, Cambridge, MA, USA, Jul. 2012, pp. 3048–3052.

[11] M. R. Bloch and J. Kliewer, "Strong coordination over a line network," in *Proc. IEEE International Symposium on Information Theory*, Istanbul, Turkey, Jul. 2013, pp. 2319–2323.

[12] S. Satpathy and P. Cuff, "Secure cascade channel synthesis," in *Proc. IEEE International Symposium on Information Theory Proc.*, Istanbul, Turkey, Jul. 2013, pp. 2955–2959.

[13] R. Blasco-Serrano, R. Thobaben, and M. Skoglund, "Polar codes for coordination in cascade networks," in *International Zurich Seminar on Communications (IZS)*. Eidgenössische Technische Hochschule Zürich, 2012, pp. 55–58.

[14] M. R. Bloch, L. Luzzi, and J. Kliewer, "Strong coordination with polar codes," in *50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, IL, USA, Oct. 2012, pp. 565–571.

[15] R. A. Chou, M. Bloch, and J. Kliewer, "Empirical and strong coordination via soft covering with polar codes," *arXiv preprint arXiv:1608.08474*, 2016.

[16] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, 1993.

[17] P. Cuff, "Distributed channel synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.

[18] M. R. Bloch, "Achieving secrecy: capacity vs. resolvability," in *Proc. IEEE International Symposium on Information Theory*, St. Petersburg, Russia, Jul. 2011, pp. 633–637.

[19] F. Haddadpour, M. H. Yassaee, S. Beig, A. Gohari, and M. R. Aref, "When is it possible to simulate a DMC channel from another?" *CoRR*, vol. abs/1305.5901, 2013. [Online]. Available: http://arxiv.org/abs/1305.5901

[20] M. H. Yassaee, M. R. Aref, and A. Gohari, "Achievability proof via output statistics of random binning," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6760–6786, 2014.

[21] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[22] T. M. Cover and J. Thomas, *Elements of Information Theorey*, 2nd ed., ser. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience, 2006.

[23] M. Pinsker, "Information and information stability of random variables and processes," *London-Amsterdam xii*, vol. 243, 1964, translated and edited by Amiel Feinstein Holden-Day, Inc., San Francisco, Calif.

[24] A. El Gamal and Y. Kim, *Network Information Theory.* Cambridge University Press, 2011.

[25] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec*, vol. 4, no. 142-163, p. 1, 1959.

[26] A. Wyner, "The common information of two dependent random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 163–179, 1975.

[27] A. D. Wyner, "The wire-tap channel," *The Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[28] G. Kramer, "Topics in multi-user information theory," *Foundations and Trends in Communications and Information Theory*, vol. 4, no. 45, pp. 265–444, 2008. [Online]. Available: http://dx.doi.org/10.1561/0100000028

[29] G. Dantzig and B. C. Eaves, "Fourier-Motzkin elimination and its dual," *Journal of Combinatorial Theory, Series A*, vol. 14, no. 3, pp. 288–297, 1973.