# Coding Schemes for Multiple-Relay Channels

by

Xiugang Wu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2013

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Abstract

In network information theory, the relay channel models a communication scenario where there is one or more relay nodes that can help the information transmission between the source and the destination. Although the capacity of the relay channel is still unknown even in the single-relay case, two fundamentally different relay schemes have been developed by (Cover and El Gamal, 1979) for such channels, which, depending on whether the relay decodes the information or not, are generally known as Decode-and-Forward (D-F) and Compress-and-Forward (C-F). In the D-F relay scheme, the relay first decodes the message sent by the source and then forwards it to the destination, and the destination decodes the message taking into account the inputs of both the source and the relay. In contrast, the C-F relay scheme is used when the relay cannot decode the message sent by the source, but still can help by compressing its observation into some compressed version, and forwarding this compression into the destination; the destination then either successively or jointly decodes the compression of the relay's observation and the original message of the source. For the single-relay case, it is known that joint compression-message decoding, although providing more freedom in choosing the compression at the relay, cannot achieve higher rates for the original message than successive decoding.

This thesis addresses some fundamental issues in generalizing and unifying the above D-F and C-F relay schemes to the multiple-relay case. We first generalize the C-F scheme to multiple-relay channels, and investigate the question of whether compression-message joint decoding can improve the achievable rate compared to successive decoding in the multiple-relay case. It is demonstrated that in the case of multiple relays, there is no improvement on the achievable rate by joint decoding either. More interestingly, it is discovered that any compressions not supporting successive decoding will actually lead to strictly lower achievable rates for the original message. Therefore, to maximize the achievable rate for the original message, the compressions should always be chosen to support successive decoding. Further-

more, it is shown that any compressions not completely decodable even with joint decoding will not provide any contribution to the decoding of the original message.

We also develop a new C-F relay scheme with block-by-block backward decoding. This new scheme improves the original C-F relay scheme to achieve higher rates in the multiple-relay case as the recently proposed noisy network coding scheme. However, compared to noisy network coding which uses repetitive encoding/all blocks united decoding, our new coding scheme is not only simpler, but also reveals the essential reason for the improvement of the achievable rate, that is, delayed decoding until all the blocks have been finished.

Finally, to allow each relay node the freedom of choosing either the D-F or C-F relay strategy, we propose a unified relay framework, where both the D-F and C-F strategies can be employed simultaneously in the network. This framework employs nested blocks combined with backward decoding to allow for the full incorporation of the best known D-F and C-F relay strategies. The achievable rates under our unified relay framework are found to combine both the best known D-F and C-F achievable rates and include them as special cases. It is also demonstrated through a Gaussian network example that our achievable rates are generally better than the rates obtained with existing unified schemes and with D-F or C-F alone.

## Acknowledgements

First and foremost, I would like to thank my advisor Professor Liang-Liang Xie, who led me into the exciting field of information theory, and worked closely with me during my Ph.D. study. I enjoyed every discussion I had with him in his office, without which this thesis would have been impossible. I appreciate every aspect I have learned or have been trying to learn from him: the enthusiasm in doing fundamental research, the intuition beyond mathematical formulas, the courage to attempt hard problems, the way of presenting the research work interestingly and attractively... Behind every step forward in my academic career is his unselfish sharing of knowledge and experience and continuous encouragement. For all of these, I want to express my deepest gratitude to him.

I am also grateful to the members of my thesis committee. I want to thank Professor En-hui Yang, Professor Oussama Damen, and Professor Chengguo Weng for their valuable comments in my comprehensive exam and their commitment to my Ph.D. defense. I want to thank Professor Wei Yu from the University of Toronto for serving as my external committee member.

Last but not least, I am deeply indebted to my parents for their endless love, support, and encouragement throughout my life.

# Contents

**List of Tables** ix

**List of Figures** x

**1 Introduction** 1

1.1 Information Theory and Network Information Theory . . . . . . . . 1

1.2 Relay Channel . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

1.2.1 Motivations . . . . . . . . . . . . . . . . . . . . . . . . . 6

1.2.2 Contributions . . . . . . . . . . . . . . . . . . . . . . . . 8

1.3 Organization of Thesis . . . . . . . . . . . . . . . . . . . . . . . . 9

**2 Preliminaries** 11

2.1 Basics in Information Theory . . . . . . . . . . . . . . . . . . . . 11

2.1.1 Entropy and Mutual Information . . . . . . . . . . . . . . . 11

2.1.2 Typical Sequences . . . . . . . . . . . . . . . . . . . . . . 12

2.1.3 Channel Capacity . . . . . . . . . . . . . . . . . . . . . . 13

2.2 Several Related Problems in Network Information Theory . . . . . . 15

2.2.1 Slepian-Wolf Coding . . . . . . . . . . . . . . . . . . . . . 15

2.2.2 Multiple Access Channel . . . . . . . . . . . . . . . . . . . 16

2.2.3 Broadcast Channel . . . . . . . . . . . . . . . . . . . . . . 17

2.3 Relay Channel . . . . . . . . . . 18
2.3.1 Single-Relay Case . . . . . . . . . . 18
2.3.2 Multiple-Relay Case . . . . . . . . . . 21
**3 C-F Relay Schemes with Forward Decoding 24**
3.1 Introduction . . . . . . . . . . 24
3.2 Main Results . . . . . . . . . . 27
3.3 Generalization of C-F Schemes with Forward Decoding to Multiple-Relay Case . . . . . . . . . . 29
3.3.1 Proof of Theorem 3.2.1 . . . . . . . . . . 29
3.3.2 Proof of Theorem 3.2.2 . . . . . . . . . . 32
3.4 Optimality of Successive Decoding in Multiple-Relay Case . . . . . 33
3.5 Discussion: An Optimality-Robustness Tradeoff . . . . . . . . . . 38
**4 C-F Relay Schemes with Backward Decoding 42**
4.1 Introduction . . . . . . . . . . 42
4.2 Main Results . . . . . . . . . . 44
4.3 Cumulative encoding/block-by-block backward decoding/compression-message successive decoding and Optimality of Successive Decoding 47
4.4 Cumulative encoding/block-by-block backward decoding/compression-message joint decoding and Necessity of Joint Decodability . . . . . 57
4.5 Proofs of Lemmas 4.3.1-4.3.4 and 4.4.4 . . . . . . . . . . 61
**5 A Unified Relay Framework with Both D-F and C-F Relay Nodes 65**
5.1 Introduction . . . . . . . . . . 65
5.2 Main Results . . . . . . . . . . 69
5.2.1 An Example of Gaussian Networks . . . . . . . . . . 71

5.3 Unified Relay Framework With Block-By-Block Backward Decoding 74

5.3.1 Single D-F Relay Node ($M = 1$) . . . . . . . . . . . . . . . . . 75

5.3.2 Multiple D-F Relay Nodes ($M \geq 2$) . . . . . . . . . . . . . . . 83

5.4 Unified Relay Framework With $B$-Blocks-By-$B$-Blocks Backward Decoding . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 87

5.4.1 Single D-F relay node ($M = 1$) . . . . . . . . . . . . . . . . . 88

5.4.2 Multiple D-F Relay Nodes ($M \geq 2$) . . . . . . . . . . . . . . . 92

5.5 Rates for the AWGN Two-Relay Channel . . . . . . . . . . . . . . . 96

**6 Conclusion and Future Work** **103**

6.1 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 103

6.2 Future Work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 104

**Bibliography** **107**

# List of Tables

5.1 Block-by-Block backward decoding for the single D-F relay node case 78

5.2 $B$-Blocks-by-$B$-Blocks backward decoding for the single D-F relay node case . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89

# List of Figures


1.1 A discrete memoryless channel. . . . . . . . . . . . . . . . . . . . . . . 2

1.2 Multi-hopping vs. Relaying. . . . . . . . . . . . . . . . . . . . . . . . 4

1.3 The single-relay channel. . . . . . . . . . . . . . . . . . . . . . . . . 4

1.4 C-F for the single-relay channel. . . . . . . . . . . . . . . . . . . . . 5

1.5 The multiple-relay channel. . . . . . . . . . . . . . . . . . . . . . . . 7

2.1 Slepian-Wolf coding. . . . . . . . . . . . . . . . . . . . . . . . . . . 15

2.2 The multiple access channel. . . . . . . . . . . . . . . . . . . . . . . 17

2.3 The broadcast channel. . . . . . . . . . . . . . . . . . . . . . . . . . 18

3.1 A Gaussian relay channel with a digital link. . . . . . . . . . . . . . . 39

3.2 Achievable rate for the Gaussian relay channel. . . . . . . . . . . . . 41

5.1 A unified relay framework with both the D-F and C-F relay nodes. 68

5.2 A two-relay channel. . . . . . . . . . . . . . . . . . . . . . . . . . . 72

5.3 Achievable rates for the two-relay channel. . . . . . . . . . . . . . . 73

5.4 An illustration of nested blocks with backward decoding. . . . . . . 76

6.1 A relay channel model with a digital link. . . . . . . . . . . . . . . 104

# Chapter 1

# Introduction

## 1.1 Information Theory and Network Information Theory

In 1948, Claude Shannon, the father of Information Theory, published his revolutionary paper "*A Mathematical Theory of Communication*" [1]. In the paper, the most significant result is arguably given by the channel coding theorem, which answers one of the most fundamental questions in information theory, namely, what is the ultimate transmission rate of communication. Roughly speaking, the channel coding theorem says that there always exists a maximum achievable rate for a channel, called the channel capacity, below which the reliable communication can be implemented while above which the reliable communication is impossible.

Specifically, for the point-to-point channel depicted in Figure 1.1, Shannon showed that the channel capacity can be elegantly expressed as

$$C = \max_{p(x)} I(X;Y), \tag{1.1}$$

where $X$ and $Y$ are respectively the input and the output of the channel, $I(X;Y)$ is the mutual information between $X$ and $Y$ (which will be formally defined in Chapter 2), and the maximum is taken over all input distributions $p(x)$. Information can be transmitted as reliably as desired if and only if the rate is below the capacity $C$.

At first sight, the result is rather counter-intuitive. How can one correct all the

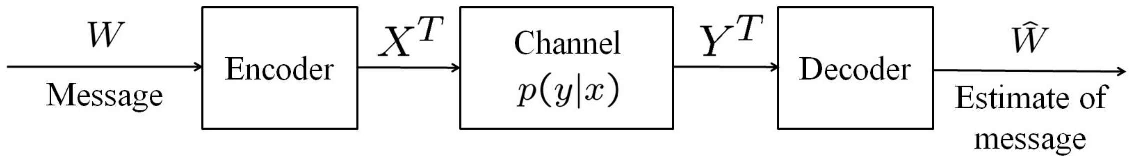



Figure 1.1: A discrete memoryless channel.

errors introduced by the noisy channel and implement the reliable communication? Indeed, proving the channel coding theorem is not trivial at all, and several original ideas were introduced by Shannon in his 1948 paper. Among these original ideas, the genius of Shannon was especially revealed by the random coding argument, where the codebook is randomly generated and the probability of error is calculated averaged over the whole code ensemble. Then, instead of trying to find a specific good code, Shannon showed that the error probability averaged over all the codebooks goes to 0, and thus successfully argued that there exists at least one good code with vanishing probability of error.

While Shannon's classical information theory addresses point-to-point channels and has had profound impact on modern communication systems, network information theory considers communication networks containing multiple senders and receivers, e.g., computer networks and satellite networks, and its main goal is to find the fundamental limits in network communications and the optimal coding schemes to achieve these limits. The presence of multiple senders and receivers in the network brings in many new elements to the communication problems, such as interference, cooperation and feedback, but also makes the problems difficult to solve. To date, there is as yet no unified theory of network information flow, although such a complete theory will undoubtedly have wide implications for today's communication networks, especially the currently ubiquitous wireless communication networks.

Nevertheless, there have been some triumphs in solving a few special problems in network information theory, including distributed lossless source coding [2]-[3], multiple access channels [4]-[5], Gaussian broadcast channels [6]-[10]. Besides, various other special topics have been studied over the past few decades, such as distributed lossy source coding [11], interference channels [12], relay channels [13]-[15], and network coding [16]. In this thesis, we will focus on studying the coding schemes for the relay channels, especially the multiple-relay channels.

## 1.2 Relay Channel

As a fundamental building block in network information theory, the relay channel was introduced by van der Meulen [13],[14]. It models a communication scenario where there is one or more relay nodes that can help the information transmission between the source and the destination. Note that the relay channel discussed here should be distinguished from the multi-hop operation in today's network communications. The main difference between these two communication models lies in how to treat the interference: In the multi-hop operation, each receiver treats the interference introduced by un-intended signals from other transmitters as noise; however, viewed more fundamentally, even the "interference" could carry useful information that may be helpful in decoding the signal from the intended transmitter. Figure 1.2 illustrates this point with a simple three-node network example, where node 0 is the source, which wants to send information to the destination node 2 with the help of node 1. With multi-hopping, one operates the network as if it is a concatenation of two separate point-to-point channels, and the signal transmitted by node 0 causes interference to node 2. However, the fact is that although the signal transmitted by node 0 is intended for node 1, it carries exactly the same information that node 2 wants to decode eventually. This observation has motivated the study of the "relay channel" from an information theoretic point of view, where the decoding at the destination is based on the signals transmitted by both the node 0 and node 1.

Formally, a general information theoretic model for the single-relay channel is depicted in Figure 1.3, where nodes 0, 1, and 2 are the source, the relay, and the destination, respectively; the relay and the source cooperate to resolve the receiver's

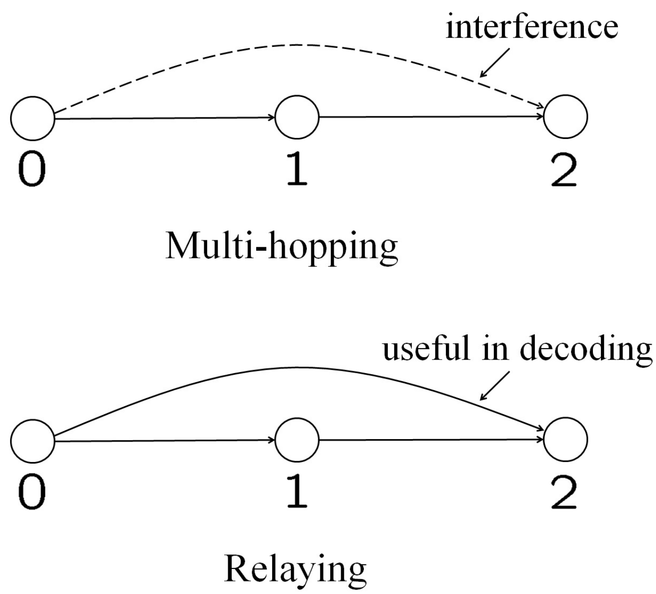



Figure 1.2: Multi-hopping vs. Relaying.

uncertainty. Despite the significant research efforts in the past few decades, the capacity of the relay channel still remains open even in the single-relay case, except for a few special classes, e.g., the physically degraded relay channel and the reversely degraded relay channel [15].

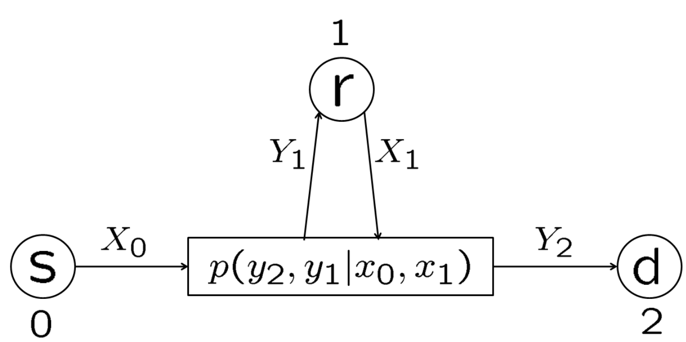



Figure 1.3: The single-relay channel.

Although the capacity of the simplest single-relay channel is still unknown in general, two fundamentally different relay strategies have been developed by Cover and El Gamal for such channels [15]. Depending on whether the relay decodes the information or not, these two strategies are now generally known as *decode-and-forward* (D-F) and *compress-and-forward* (C-F) respectively. In the D-F relay strategy, the relay first decodes the message sent by the source and then forwards it to the destination, and the destination decodes the message taking into account the inputs of both the source and the relay. With the D-F relay strategy, the following rate is achievable:

$$R < \max_{p(x_0,x_1)} \min\{I(X_0;Y_1|X_1), I(X_0,X_1;Y_2)\} \tag{1.2}$$

where, the first condition $R < I(X_0;Y_1|X_1)$ makes node 1 able to decode the message based on the signal transmitted by node 0, and the second condition $R < I(X_0,X_1;Y_2)$ makes node 2 able to decode the message based on the signals transmitted by node 0 and node 1 together. Notably, the maximization in (1.2) is over $p(x_0,x_1)$, rather than $p(x_0)p(x_1)$, which suggests that (1.2) can only be achieved by node 0 and node 1 cooperating with each other when transmitting signals. To accomplish such cooperation, an essential technique called block Markov coding was employed in the D-F coding scheme developed in [15]. Besides, the scheme in [15] also used irregular encoding with codebooks of different sizes at the source and at the relay, random partitioning (binning), and successive decoding. Subsequently, some other D-F coding schemes also achieving (1.2) were found in [17]-[18].

In contrast, the C-F relay strategy is used when the relay cannot decode the message sent by the source, but still can help by compressing its observation $Y_1$ into $\hat{Y}_1$, and forwarding this compressed version to the destination via $X_1$, as illustrated in Figure 1.4. The destination then either successively or jointly decodes the com-

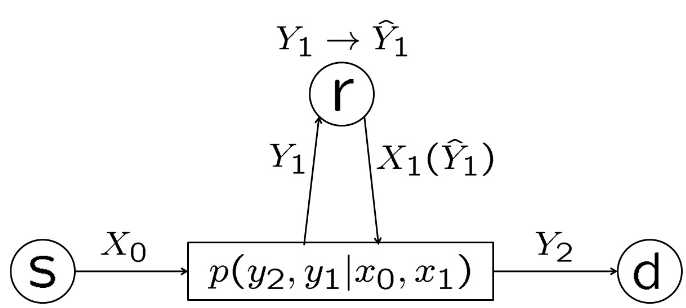



Figure 1.4: C-F for the single-relay channel.

pression of the relay's observation and the original message of the source. In the original C-F scheme of [15], the decoder performs successive compression-message decoding, i.e., it first decodes the compression of the relay's observation, and then decodes the original message of the source, leading to the following achievable rate:

$$R < \max_{p(x_0)p(x_1)p(\hat{y}_1|y_1,x_1)} I(X_0;\hat{Y}_1,Y_2|X_1) \tag{1.3}$$

$$\text{such that} \quad I(Y_1;\hat{Y}_1|X_1,Y_2) \leq I(X_1;Y_2), \tag{1.4}$$

where (1.4) ensures that the compression $\hat{Y}_1$ can be first recovered at the destination, and (1.3) ensures that the destination can decode the original message $X_0$ based on $\hat{Y}_1$ and $Y_2$ together.

The two-step compression-message successive decoding process in [15] requires $\hat{Y}_1$ to be decoded first, which facilitates the decoding of $X_0$, but is not a requirement of the original problem. Recognizing this, a joint compression-message decoding process was proposed in [19], where, instead of successively, the destination decodes $\hat{Y}_1$ and $X_0$ together. It turns out that the decoding of $X_0$ can be helped even if $\hat{Y}_1$ cannot be decoded first. In fact, with joint decoding, the constraint (1.4) is not necessary, and instead of (1.3), the achievable rate is expressed as

$$R < \max_{p(x_0)p(x_1)p(\hat{y}_1|y_1,x_1)} I(X_0;\hat{Y}_1,Y_2|X_1) - \max\{0, I(Y_1;\hat{Y}_1|X_1,Y_2) - I(X_1;Y_2)\}. \tag{1.5}$$

Similar formulas as (1.5) have been derived with different arguments in [20]-[22]. Therefore, compared to successive decoding, joint compression-message decoding provides more freedom in choosing the compression $\hat{Y}_1$. However, the question remains whether joint decoding achieves strictly higher rates for the original message than successive decoding. For the single relay case, it was proved in [22] that the answer is negative, and any rate achievable by either of them can always be achieved by the other, i.e., the achievable rates in (1.3)-(1.4) and (1.5) are essentially the same.

### 1.2.1 Motivations

A natural extension of the single-relay channel in Figure 1.3 is to the case of multiple relays depicted in Figure 1.5, where nodes 0 and $n+1$ are the source and the destination respectively, and nodes $1, 2, \ldots, n$ are the $n$ relay nodes that constitute the relay nodes set, denoted by $\mathcal{N}$. There have been some works on generalizing the

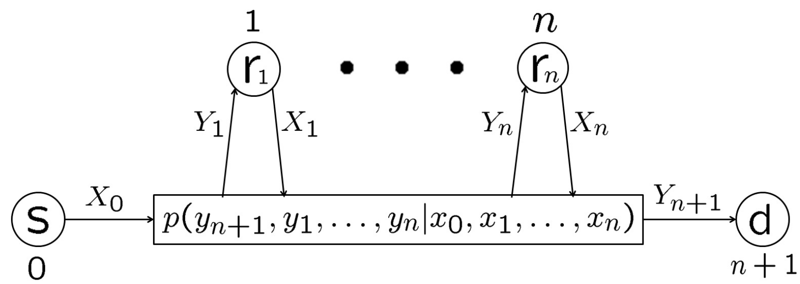



Figure 1.5: The multiple-relay channel.

above D-F and C-F relay strategies to such multi-relay channels [24]-[30]. Specifically, the generalization of D-F to the multi-relay channel has been completely resolved, and the schemes presented in [27]-[30] provided the best D-F rate in the multiple-relay case. However, when generalizing C-F to the multiple-relay case, there are still several fundamental issues left unaddressed, such as, how to generalize the compression-message joint decoding scheme to the multiple-relay case, whether compression-message joint decoding can improve the achievable rate compared to successive decoding in the multiple-relay case, and whether there exists any better C-F relay scheme than the conventional schemes [15], [19] in the case of multiple relays. One of the purposes of this thesis is to address these issues.

Our thesis also addresses the issue of unifying the D-F and C-F relay strategies for general multi-relay channels. In the above discussions, all the relay nodes in the network perform only one type of relay strategy, either D-F or C-F. However, to obtain higher achievable rate, it might be better to let each relay node choose from D-F and C-F its relay strategy depending on the channel condition, e.g., let the relay node close to the source perform D-F while let the relay node close to the destination perform C-F. This has motivated the second line of research in our thesis, namely, the investigation of a unified relay framework where both the D-F and C-F strategies can be employed simultaneously in the network.

### 1.2.2 Contributions

The contributions of this thesis can be summarized as follows. All our results in this thesis have been published or accepted for publication in [33]-[37].

- We generalize the conventional C-F relay scheme to the multiple-relay case. As in the conventional scheme for the single-relay case, our generalized C-F scheme is also based on block-by-block forward decoding. Different from the previous generalization [30], which only considers compression-message successive decoding, our generalization includes two decoding methods: either the original compression-message successive decoding or compression-message joint decoding.

- We further investigate the question of whether the freedom of selecting the compressions in compression-message joint decoding can improve the achievable rate of the original message when C-F is generalized to multiple-relay channels. It is demonstrated that in the case of multiple relays, there is no improvement on the achievable rate by joint decoding either. More interestingly, it is discovered that any compressions not supporting successive decoding will actually lead to strictly lower achievable rates for the original message. Therefore, to maximize the achievable rate for the original message, the compressions should always be chosen to support successive decoding.

- We then develop a new C-F relay scheme with block-by-block backward decoding. To our knowledge, this is the first time of backward decoding being applied to the C-F scheme. It turns out that our new C-F scheme achieves the same rate as the recently proposed "noisy network coding" scheme [32], which improves the achievable rate of the conventional C-F scheme in the case of multiple relays although no improvement is shown in the single relay case. However, compared to the noisy network coding scheme which uses repetitive encoding/all blocks united decoding, our coding scheme is not only simpler, but also reveals the essential reason for the improvement of the achievable rate, that is, delayed decoding until all the blocks have been finished.

- Similarly, our new C-F relay scheme can also be combined with compression-message successive decoding or joint decoding, and it is shown that the achievable rate with our new scheme is maximized only when the compressions are chosen to support successive decoding. Furthermore, it is shown that any compressions not completely decodable even with joint decoding will not provide any contribution to the decoding of the original message.

- Finally, to allow each relay node the freedom of choosing either the D-F or C-F relay strategy, we propose a unified relay framework, where both the D-F and C-F strategies can be employed simultaneously in the network. This framework employs nested blocks combined with backward decoding to allow for the full incorporation of the best known D-F and C-F relay strategies. The achievable rates under our unified relay framework are found to combine both the best known D-F and C-F achievable rates and include them as special cases. It is also demonstrated through a Gaussian network example that our achievable rates are generally better than the rates obtained with existing unified schemes and with D-F or C-F alone.

## 1.3 Organization of Thesis

The remainder of this thesis is organized as follows.

Some preliminaries are first given in Chapter 2. As the prerequisite knowledge for later discussions, we first introduce some classical results in single-user information theory, such as the concepts and properties of typical sequences as well as a formal statement of the channel coding theorem. Then a quick review of several related multi-user communication problems is presented. Finally, some formal definitions on the relay channel, and two fundamental relay schemes, namely, D-F and C-F, are discussed.

In Chapter 3, we discuss the conventional C-F relay scheme with forward decoding. We first generalize the conventional C-F scheme to the multiple-relay case, and then prove the optimality of compression-message successive decoding, in the sense that to maximize the achievable rate, the compressions at the relays must be chosen to support successive decoding. An optimality-robustness tradeoff is discussed with a Gaussian relay channel example.

In Chapter 4, we develop the new C-F relay scheme with backward decoding. In particular, we develop the following two schemes:

- cumulative encoding/block-by-block backward decoding/compression-message successive decoding
- cumulative encoding/block-by-block backward decoding/compression-message joint decoding

and prove that they achieve the equivalent rate as the recently proposed noisy network coding scheme. In proving such a rate equivalence, the optimality of successive decoding and necessity of joint decodability are also demonstrated.

Chapter 5 proposes a unified relay framework where both the D-F and C-F strategies can be employed simultaneously in the network. Nested blocks combined with backward decoding are used to establish the achievable rates, which are found to combine both the best known D-F and C-F achievable rates and include them as special cases. A Gaussian two-relay channel is used as an example to demonstrate that our achievable rates are generally better than the rates obtained with existing unified schemes and with D-F or C-F alone.

Finally, we conclude this thesis and propose some future work in Chapter 6.

# Chapter 2

# Preliminaries

## 2.1 Basics in Information Theory

In this section, we introduce some basic and important tools and results in information theory, which will be used throughout this thesis. We begin with some standard definitions on information measures [44, Ch.2].

### 2.1.1 Entropy and Mutual Information

Let $X$ be a discrete random variable with alphabet $\mathcal{X}$ and probability mass function $p(x) = \Pr(X = x), x \in \mathcal{X}$.

**Definition 2.1.1.** *The entropy $H(X)$ of a discrete random variable $X$ is defined by*

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

The above definition of entropy can be extended to a pair of random variables.

**Definition 2.1.2.** *The joint entropy $H(X, Y)$ of a pair of discrete random variables $(X, Y)$ with a joint distribution $p(x, y)$ is defined as*

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

We can also define the conditional entropy.

**Definition 2.1.3.** *If $(X, Y) \sim p(x, y)$, then the conditional entropy $H(Y|X)$ is defined as*

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

The definition of mutual information is as the following.

**Definition 2.1.4.** *Let $(X, Y)$ be a pair of random variables with a joint probability mass function $p(x, y)$ and marginal probability mass functions $p(x)$ and $p(y)$. The mutual information $I(X;Y)$ is defined as*

$$I(X, Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}.$$

### 2.1.2 Typical Sequences

We now introduce the definitions of typical sequences for discrete random variables, and some of their properties [44], [45].

**Definition 2.1.5.** *A sequence $x^T \in \mathcal{X}^T$ is said to be $\epsilon$-typical with respect to a distribution $p(x)$ on $\mathcal{X}$ if:*

1. *For all $a \in \mathcal{X}$ with $p(a) > 0$, we have*

$$\left| \frac{1}{T} N(a|x^T) - p(a) \right| < \frac{\epsilon}{|\mathcal{X}|}$$

2. *For all $a \in \mathcal{X}$ with $p(a) = 0$, $N(a|x^T) = 0$*

*where $N(a|x^T)$ is the number of occurrences of the symbol $a$ in the sequence $x^T$.*

The set of sequences $x^T \in \mathcal{X}^T$ such that $x^T$ is typical is called the typical set and is denoted by $A_\epsilon^{(T)}(X)$ or $A_\epsilon^{(T)}$ when the random variable is understood from the context.

**Theorem 2.1.1** (Asymptotic Equipartition Property(AEP))**.** *Let $X^T$ be a sequence of length $n$ drawn i.i.d. according to $p(x^T) = \prod_{i=1}^{T} p(x_i)$, then:*

1. $\Pr(A_\epsilon^{(T)}) > 1 - \epsilon$ *for $T$ sufficiently large;*

2. $|A_\epsilon^{(T)}| \leq 2^{T[H(X)+\epsilon']}$*;*

*3.* $|A_\epsilon^{(T)}| \geq (1-\epsilon)2^{T[H(X)-\epsilon']}$ *for* $T$ *sufficiently large;*

*where* $|A|$ *denotes the number of elements in the set* $A$ *and* $\epsilon' \to 0$ *as* $\epsilon \to 0$.

The definition of typicality can be extended to a pair of sequences.

**Definition 2.1.6.** *A pair of sequences* $(x^T, y^T) \in \mathcal{X}^T \times \mathcal{Y}^T$ *is said to be* $\epsilon$*-jointly typical with respect to a distribution* $p(x,y)$ *on* $\mathcal{X} \times \mathcal{Y}$ *if:*

1. *For all* $(a,b) \in \mathcal{X} \times \mathcal{Y}$ *with* $p(a,b) > 0$*, we have*

$$\left| \frac{1}{T} N(a,b|x^T, y^T) - p(a,b) \right| < \frac{\epsilon}{|\mathcal{X}||\mathcal{Y}|} \tag{2.1}$$

2. *For all* $(a,b) \in \mathcal{X} \times \mathcal{Y}$ *with* $p(a,b) = 0$*,* $N(a,b|x^T, y^T) = 0$.

*where* $N(a,b|x^T,y^T)$ *is the number of occurrences of the pair* $(a,b)$ *in the pair of sequences* $(x^T, y^T)$.

From the definition, it follows that if $(x^T, y^T) \in A_\epsilon^{(T)}(X,Y)$, then $x^T \in A_\epsilon^{(T)}(X)$.

### 2.1.3 Channel Capacity

Before formally stating the channel coding theorem, we need a few definitions [44, Ch. 8].

**Definition 2.1.7.** *A discrete channel, denoted by* $(\mathcal{X}, p(y|x), \mathcal{Y})$*, consists of two finite sets* $\mathcal{X}$ *and* $\mathcal{Y}$ *and a collection of probability mass functions* $p(y|x)$*, one for each* $x \in \mathcal{X}$*, such that for every* $x$ *and* $y$*,* $p(y|x) \geq 0$*, and for every* $x$*,* $\sum_y p(y|x) = 1$*, with the interpretation that* $X$ *is the input and* $Y$ *is the output of the channel. The transition probability for the* $T$*-th extension of the discrete memoryless channel (without feedback) is defined as* $p(y^T|x^T) = \prod_{i=1}^T p(y_i|x_i)$.

**Definition 2.1.8.** *An* $(M,T)$ *code for the channel* $(\mathcal{X}, p(y|x), \mathcal{Y})$ *consists of the following:*

1. *An index set* $\{1, 2, \ldots, M\}$.
2. *An encoding function* $X^T : \{1, 2, \ldots, M\} \to \mathcal{X}^T$*, yielding codewords*

$$X^T(1), X^T(2), \ldots, X^T(M).$$

*The set of codewords is called the codebook.*

*3. A decoding function* $g : \mathcal{Y}^T \to \{1, 2, \ldots, M\}$*, which is a deterministic rule which assigns a guess to each possible received vector.*

**Definition 2.1.9** (Probability of error)**.** *Let*

$$\lambda_i = \Pr(g(Y^T) \neq i | X^T = X^T(i)) = \sum_{y^T : g(y^T) \neq i} p(y^T | x^T(i))$$

*be the conditional probability of error given that index i was sent. The maximal probability of error* $\lambda^{(T)}$ *for an* $(M, T)$ *code is defined as*

$$\lambda^{(T)} = \max_{i \in \{1,2,\ldots,M\}} \lambda_i.$$

*The average probability of error* $P_e^{(T)}$ *for an* $(M, T)$ *code is defined as*

$$P_e^{(T)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i.$$

**Definition 2.1.10** (Achievable rate and capacity)**.** *The rate R of an* $(M, T)$ *code is* $R = \log M/T$ *bits per transmission. A rate R is said to be achievable if there exists an sequence of* $(2^{TR}, T)$ *codes such that the maximal probability of error* $\lambda^{(T)}$ *tends to 0 as* $T \to \infty$*. The capacity of a discrete memoryless channel is the supermum of all achievable rates.*

We now formally state Shannon's channel coding theorem.

**Theorem 2.1.2** (The Channel Coding Theorem)**.** *All rates below capacity* $C = \max_{p(x)} I(X;Y)$ *are achievable. Specifically, for every rate* $R < C$*, there exists a sequence of* $(2^{TR}, T)$ *codes with maximum probability error* $\lambda^{(T)} \to 0$.

*Conversely, any sequence of* $(2^{TR}, T)$ *codes with* $\lambda^{(T)} \to 0$ *must have* $R \leq C$.

At first glance, the result is rather counter-intuitive. How can one correct all the errors introduced by the noisy channel and implement the reliable communication? To prove the achievability part, Shannon used a number of new ideas. These original ideas include: i) allowing a vanishing probability of error instead of requiring zero probability of error; ii) using the channel many times instead of only once to put the law of large numbers into effect; iii) randomly generating the codebook and calculating the probability of error averaged over the code ensemble. By showing the error probability averaged over all the codebooks goes to 0, it can be argued that there exists at least one good code with vanishing probability of error.

To prove the converse part, we need the *Fano's Inequality* and the details can be found in [44]. This converse is sometimes called the weak converse to channel coding theorem. It is also possible to prove a strong converse, which states that for rates above capacity, the probability of error goes exponentially to one [45], [46]. Thus, the capacity $C$ is a sharp threshold between perfectly reliable and completely unreliable communication.

## 2.2 Several Related Problems in Network Information Theory

In this section, we consider several related multi-user communication models, including distributed lossless source coding, multiple access channel and broadcast channel. The relay channel will be treated in details in the next section.

### 2.2.1 Slepian-Wolf Coding

We know that a rate $R > H(X)$ is sufficient to encode the source $X$. Now, consider the source coding problem presented in the Figure 2.1, where $X$ and $Y$ are correlated but encoded separately. It is natural to ask what is the sufficient rate pairs for the decoder to reconstruct both $X$ and $Y$.

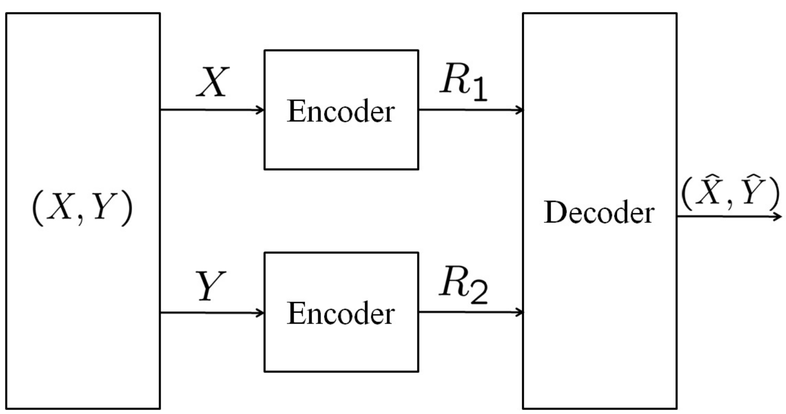



Figure 2.1: Slepian-Wolf coding.

The achievable rate pair is defined as the following.

**Definition 2.2.1.** *A rate pair* $(R_1, R_2)$ *is said to be achievable if there exists a sequence of* $((2^{TR_1}, 2^{TR_2}), T)$ *distributed source codes with probability of error*

$P_e^{(T)} \to 0$, *and the closure of the set of achievable rate pairs is called the achievable rate region.*

The achievable rate region was established by Slepian and Wolf [2].

**Theorem 2.2.1** (Slepian-Wolf Theorem)**.** *For the distributed source coding problem for the source* $(X, Y)$ *drawn i.i.d.* $\sim p(x, y)$*, the achievable rate region is given by*

$$R_1 \geq H(X|Y), \tag{2.2}$$
$$R_2 \geq H(Y|X), \tag{2.3}$$
$$R_1 + R_2 \geq H(X, Y). \tag{2.4}$$

#### Random Binning and Achievability

The Slepian-Wolf Theorem was then extended to jointly ergodic sources by Cover [3]. In his paper, Cover used a binning argument, which has evolved to one of the most significant techniques besides Shannon's random coding.

Briefly speaking, the technique is to randomly assign $2^{TH(X)}$ and $2^{TH(Y)}$ typical sequences to $2^{TR_1}$ and $2^{TR_2}$ indexed bins respectively, such that there are $2^{T(H(X)-R_1)}$ or $2^{T(H(Y)-R_2)}$ sequences in each bin. Given a realization $x^T$ (or $y^T$), the encoder just simply transmits the index of the bin containing $x^T$ (or $y^T$) and the decoder uses the method of jointly decoding, i.e., finding the jointly typical pair $(x^T, y^T)$ contained in the bins corresponding to the received bin index. Readily we can see that, if $R_1 + R_2 > H(X, Y)$, then the probability that there exists another jointly typical sequence pair can be driven to 0 as $T \to \infty$. This is the idea of random binning and the outline of the achievability of Slepian-Wolf Theorem. Note that the binning technique was also employed in the original C-F scheme for the relay channel [15], which will be discussed in the next section.

### 2.2.2 Multiple Access Channel

In the multiple access channel as depicted in Figure 2.2, sender 1 chooses an index $W_1$ uniformly from the set $\{1, 2, \cdots, 2^{TR_1}\}$ and sends the corresponding codeword over the channel and sender 2 does likewise simultaneously.

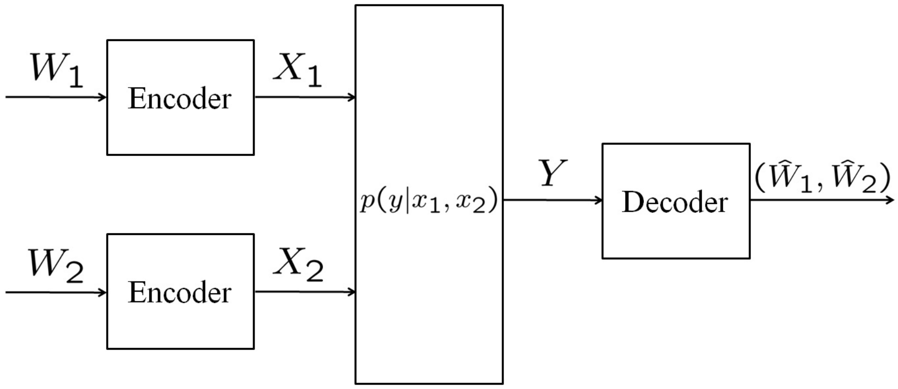



Figure 2.2: The multiple access channel.

A rate pair $(R_1, R_2)$ is said to be achievable for the multiple access channel if there exists a sequence of $((2^{TR_1}, 2^{TR_2}), T)$ codes with $P_e^{(T)} \to 0$, and the closure of the set of achievable rate pairs is called the capacity region. The multiple access channel capacity region was found by Ahlswede [4] and Liao [5]:

**Theorem 2.2.2.** *The capacity region of a multiple access channel is given by the convex hull of all $(R_1, R_2)$ satisfying*

$$R_1 \leq I(X_1; Y|X_2), \tag{2.5}$$

$$R_2 \leq I(X_2; Y|X_1), \tag{2.6}$$

$$R_1 + R_2 \leq I(X_1, X_2; Y), \tag{2.7}$$

*for some product distribution $p(x_1)p(x_2)$ on $\mathcal{X}_1 \times \mathcal{X}_2$.*

### 2.2.3 Broadcast Channel

The broadcast channel was firstly introduced by Cover in [6]. This channel describes the scenario where there is one sender and multiple (two or more) receivers, as illustrated in Figure 2.3. Although the capacity region for general broadcast channels is still unknown, it has been determined for some special classes, for example, the degraded broadcast channels [7], [8], [9].

**Definition 2.2.2.** *A broadcast channel is said to be physically degraded if*

$$p(y_1, y_2|x) = p(y_1|x)p(y_2|y_1). \tag{2.8}$$

*A broadcast channel is said to be stochastically degraded if its conditional marginal distributions are the same as that of a physically degraded broadcast channel.*

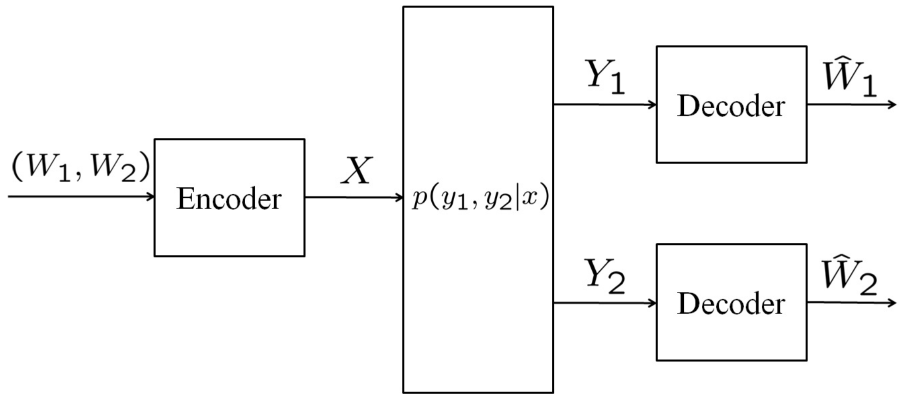



Figure 2.3: The broadcast channel.

Since the capacity region of a broadcast channel depends only on the conditional marginal distributions $p(y_1|x)$ and $p(y_2|x)$ [44], both the physically degraded and the stochastically degraded broadcast channels have the same capacity region if they share the same conditional marginal distributions. The capacity region of degraded broadcast channel was first conjectured by Cover in [6], and then proved to be achievable by Bergmans [7], using the idea of *superposition coding*. Finally Bergmans [9] and Gallager [8] established the converse.

**Theorem 2.2.3.** *The capacity region for sending independent information over the degraded broadcast channel $X \to Y_1 \to Y_2$ is the convex hull of the closure of all $(R_1, R_2)$ satisfying*

$$R_2 \leq I(U; Y_2), \tag{2.9}$$
$$R_1 \leq I(X; Y_1|U) \tag{2.10}$$

*for some joint distribution $p(u)p(x|u)p(y_1, y_2|x)$, where the auxiliary random variable $U$ has cardinality bounded by $|\mathcal{U}| \leq \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\}$.*

## 2.3 Relay Channel

### 2.3.1 Single-Relay Case

The relay channel was introduced by van der Meulen [13],[14]. A general model for the discrete memoryless single-relay channel is depicted in Figure 1.3, which can be viewed as a combination of a broadcast channel (from source to relay and destination) and a multiple access channel (from source and relay to destination). Formally, the channel consists of four finite sets $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_1$ and $\mathcal{Y}_2$, and a collection of probability mass functions $p(\cdot,\cdot|x_0,x_1)$ on $\mathcal{Y}_1 \times \mathcal{Y}_2$, one for each $(x_0,x_1) \in \mathcal{X}_0 \times \mathcal{X}_1$. The interpretation is that $x_0$ is the input to the channel from the source node 0, $y_2$ is the output of the channel to the destination node 2, and $y_1$ is the output received by the relay node 1. The relay sends an input $x_1$ based on what it has received:

$$x_1(t) = r_{1,t}(y_1(t-1), y_1(t-2), \ldots), \quad \text{for every time } t, \tag{2.11}$$

where $r_{1,t}(\cdot)$ can be any causal function.

Although the exact capacity of the general relay channel is still unknown after several decades' effort, two fundamental coding schemes have been developed in [15], which, depending on whether the relay decodes the information or not, are generally known as *decode-and-forward* (D-F) and *compress-and-forward* (C-F) respectively. The coding theorems with D-F and C-F in their original forms [15] are presented respectively in the following.

**Theorem 2.3.1.** *For the single-relay channel, a rate $R$ is achievable with D-F if for some $p(x_0,x_1)$,*

$$R < \min\{I(X_0;Y_1|X_1), I(X_0,X_1|Y_2)\}. \tag{2.12}$$

*Proof. Random coding and binning:* First randomly generate $2^{TR_0}$ i.i.d. sequences according to $p(x_1^T) = \Pi_{t=1}^T p(x_{1t})$, indexed as $x_1^T(s)$, $s \in [1, 2^{TR_0}]$ and for each $x_1^T(s)$, generate $2^{TR}$ conditionally independent sequences $x_0^T(w|s)$, $w \in [1, 2^{TR}]$ according to $p(x_0^T|x_1^T(s)) = \Pi_{t=1}^T p(x_{0t}|x_{1t}(s))$. Then randomly distribute the indexes $1, \cdots, 2^{TR}$ to $2^{TR_0}$ bins $S_1, \cdots, S_{2^{TR_0}}$ such that each message index $w$ is corresponding to a bin index $s$, i.e., contained in the bin $S_s$.

*Encoding:* At block $b$, let $w_b$ be the new index to be sent and assume $w_{b-1} \in S_{s_b}$. The source sends $x_0^T(w_b|s_b)$ while the relay estimates $w_{b-1}$ by $\hat{\hat{w}}_{b-1}$ and sends $x_1^T(\hat{\hat{s}}_b)$ assuming $\hat{\hat{w}}_{b-1} \in S_{\hat{\hat{s}}_b}$.

*Decoding:* At the end of block $b$, the decoding is implemented as follows:

1. Upon estimating $s_b$ by $\hat{\hat{s}}_b$ and receiving $y_1^T(b)$, the relay claims that the message $\hat{\hat{w}}_b = w$ is sent iff there exists a unique $w$ such that $(x_0^T(w|\hat{\hat{s}}_b), y_1^T, x_1^T(\hat{\hat{s}}_b))$ are jointly typical. This decoding error probability can be arbitrarily small if $R < I(X_0;Y_1|X_1)$.

2. Upon receiving $y_2^T(b)$, the receiver claims that the message $\hat{s}_b = s$ is sent iff there exists a unique $s$ such that $(y_2^T, x_1^T(s))$ are jointly typical. This decoding error probability can be arbitrarily small if $R_0 < I(X_1; Y_2)$.

3. The receiver calculates his ambiguity set $\mathcal{L}(y_2^T(b-1))$ consisting of all $w_{b-1}$ such that $(x_0^T(w_{b-1}|\hat{s}_{b-1}), y_2^T(b-1), x_1^T(\hat{s}_{b-1}))$ are jointly typical. Assuming that $s_b$ is decoded successfully, the receiver claims that $\hat{w}_{b-1} = w$ is sent iff there exists a unique $w \in S_{s_b} \cap \mathcal{L}(y_2^T(b-1))$. This decoding error probability can be arbitrarily small if $R < I(X_0; Y_2|X_1) + R_0$. Obviously, the receiver is always one block behind. In $B$ blocks of transmission, a sequence of $B-1$ indices will be sent, resulting in the actual rate $R(B-1)/B$ that is arbitrarily close to $R$ as $B \to \infty$. Combining all the above, we establish Theorem 2.3.1.

□

**Theorem 2.3.2.** *For the single-relay channel ,a rate $R$ is achievable with C-F if for some $p(x_0)p(x_1)p(\hat{y}_1|x_1, y_1)$*

$$R < I(X_0; \hat{Y}_1, Y_2|X_1), \tag{2.13}$$

*and*

$$I(X_1; Y_2) \geq I(Y_1; \hat{Y}_1|Y_2, X_1). \tag{2.14}$$

*Proof. Random coding and binning:* Randomly generate $2^{TR_0}$ sequences according to $p(x_1^T) = \Pi_{t=1}^T p(x_{1t})$, indexed as $x_1^T(s)$, $s \in [1, 2^{TR_0}]$ and $2^{TR}$ sequences according to $p(x_0^T) = \Pi_{t=1}^T p(x_{0t})$, indexed as $x_0^T(w)$, $w \in [1, 2^{TR}]$. For each $x_1^T(s)$, generate $2^{T\hat{R}}$ sequences according to $p(\hat{y}_1^T|x_1^T(s)) = \Pi p(\hat{y}_{1t}|x_{1t}(s))$, where $p(\hat{y}_1|x_1) = \Sigma_{x_0,y_1,y_2} p(\hat{y}_1|x_1, y_1)p(x_0)p(y_1, y_2|x_0, x_1)$, indexed as $\hat{y}_1^T(z|s)$, $z \in [1, 2^{T\hat{R}}]$, $s \in [1, 2^{TR_0}]$. Then randomly distribute the indexes $1, \cdots, 2^{T\hat{R}}$ to $2^{TR_0}$ bins $S_1, \cdots, S_{2^{TR_0}}$.

*Encoding:* At block $b$, let $w_b$ be the new index to be sent and assume

$$(\hat{y}_1^T(z_{b-1}|s_{b-1}), y_1^T(b-1), x_1^T(s_{b-1}))$$

are jointly typical and $z_{b-1} \in S_{s_b}$. The codeword pair $(x_0^T(w_b), x_1^T(s_b))$ are sent.

*Decoding:* At the end of block $b$, the decoding is implemented as follows:

1. Upon receiving $y_2^T(b)$, the receiver claims that the message $\hat{s}_b = s$ is sent iff there exists a unique $s$ such that $(y_2^T, x_1^T(s))$ are jointly typical. This decoding error probability can be arbitrarily small if $R_0 < I(X_1; Y_2)$.

2. The receiver calculates a set $\mathcal{L}(y_2^T(b-1))$ consisting of all $z$ such that

$$(\hat{y}_1^T(z|\hat{s}_{b-1}), x_1^T(\hat{s}_{b-1}), y_2^T(b-1))$$

are jointly typical. The receiver claims that $z_{b-1}$ is sent in block $b-1$ if $\hat{z}_{b-1} \in S_{s_b} \cap \mathcal{L}(y_2^T(b-1))$. This decoding error probability can be arbitrarily small if $\hat{R} < I(\hat{Y}_1; Y_2|X_1) + R_0$.

3. The receiver declares that $\hat{w}_{b-1}$ was sent in block $b-1$ if

$$(x_0^T(\hat{w}_{b-1}), \hat{y}_1^T(\hat{z}_{b-1}|\hat{s}_{b-1}), x_1^T(\hat{s}_{b-1}), y_2^T(b-1))$$

are jointly typical. This decoding error probability can be arbitrarily small if $R < I(X_0; \hat{Y}_1, Y_2|X_1)$.

4. Upon receiving $y_1^T(b)$, the relay declares that $z$ is "received" if

$$(\hat{y}_1^T(z|\hat{s}_b), x_1^T(\hat{s}_b), y_1^T(b))$$

are jointly typical. There will exist such a $z$ if $\hat{R} > I(Y_1; \hat{Y}_1|X_1)$. Combining all the above, we obtain the constraint $I(X_1; Y_2) \geq I(Y_1; \hat{Y}_1|Y_2, X_1)$.

□

Combining the D-F and C-F together, one can further consider the hybrid scheme, where the relay partially decodes the message and compresses the rest of its received signals; see, e.g., [15, Thm 7] for the single-relay case and its extension to the multiple-relay case in [23]. However, such hybrid schemes generally involve superposition coding that induces auxiliary random variables, making the expression and evaluation of the achievable rates rather complicated especially in the case of multiple relays that we will consider in the sequel. Thus, in this thesis, our discussion focuses on the "pure" D-F or C-F strategies only, i.e., the strategies where the relay either completely decodes the message, or does not decode at all but simply compresses and forwards its observation.

### 2.3.2 Multiple-Relay Case

A multiple-relay channel consisting of $n+2$ nodes is depicted in Figure 1.5, where nodes 0 and $n+1$ are the source and the destination respectively, and nodes $1, 2, \ldots, n$ are the $n$ relay nodes that constitute the relay nodes set, denoted by $\mathcal{N}$. Formally, this channel can be denoted by

$$(\mathcal{X}_0 \times \mathcal{X}_1 \times \cdots \times \mathcal{X}_n,\, p(y_{n+1}, y_1, \ldots, y_n|x_0, x_1, \ldots, x_n),\, \mathcal{Y}_{n+1} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n)$$

where, $\mathcal{X}_0, \mathcal{X}_1, \ldots, \mathcal{X}_n$ are the transmitter alphabets of the source and the relays respectively, $\mathcal{Y}_{n+1}, \mathcal{Y}_1, \ldots, \mathcal{Y}_n$ are the receiver alphabets of the destination and the relays respectively, and a collection of probability distributions $p(\cdot, \cdot, \ldots, \cdot | x_0, x_1, \ldots, x_n)$ on $\mathcal{Y}_{n+1} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n$, one for each $(x_0, x_1, \ldots, x_n) \in \mathcal{X}_0 \times \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$. The interpretation is that $x_0$ is the input to the channel from the source, $y_{n+1}$ is the output of the channel to the destination, and $y_i$ is the output received by the $i$-th relay. The $i$-th relay sends an input $x_i$ based on what it has received:

$$x_i(t) = r_{i,t}(y_i(t-1), y_i(t-2), \ldots), \quad \text{for every time } t, \tag{2.15}$$

where $r_{i,t}(\cdot)$ can be any causal function.

There have been some works on generalizing the D-F and C-F relay strategies to such multi-relay channels [24]-[30]. Specifically, the generalization of D-F to the multi-relay channel has been completely resolved, and the schemes presented in [27]-[30] provided the best D-F rate in the multiple-relay case. However, when generalizing C-F to the multiple-relay case, there are still several fundamental issues left unaddressed, such as, how to generalize the compression-message joint decoding scheme to the multiple-relay case, whether compression-message joint decoding can improve the achievable rate compared to successive decoding in the multiple-relay case, and whether there exists any better C-F relay scheme than the conventional schemes [15], [19] in the case of multiple relays. These issues will be addressed in the following chapters of this thesis. Here, in this section, we only introduce the generalization of D-F to the multi-relay channel [27]-[30].

Specifically, in generalizing D-F to the multi-relay channel, [27]-[28] modified the original irregular encoding/successive decoding scheme of [15] to a regular encoding/sliding window decoding scheme to realize the "multi-level" D-F relay strategy. For any fixed permutation $\pi$ on $\{0, 1, \ldots, n+1\}$ with $\pi(1) = 0$ and $\pi(n+2) = n+1$, i.e., any specific ordering of the relay nodes as $\pi(2), \pi(3), \ldots, \pi(n+1)$, their multi-level D-F scheme [27]-[28] achieves the rate stated in Theorem 2.3.3. Later on, it was found in [29]-[30] that the same rate can also be achieved with backward decoding.

**Theorem 2.3.3.** *For the multiple-relay channel, a rate $R$ is achievable if there exists a permutation $\pi$ on $\{0, 1, \ldots, n+1\}$ with $\pi(1) = 0$ and $\pi(n+2) = n+1$, such that*

$$R < \max_{p(x_0, x_1, \ldots, x_n)} \min_{2 \le k \le n+2} I(X_{\pi(1:k-1)}; Y_{\pi(k)} | X_{\pi(k:n+1)}), \tag{2.16}$$

*where* $\pi(k_1 : k_2) := \{\pi(k_1), \pi(k_1+1), \ldots, \pi(k_2)\}$.

The formula (2.16) has a similar interpretation as (1.2)/(2.12). For each node $\pi(k), k = 2, 3, \ldots, n+2$, the corresponding rate constraint is

$$R < I(X_{\pi(1:k-1)}; Y_{\pi(k)} | X_{\pi(k:n+1)}), \tag{2.17}$$

which implies that for the decoding at node $\pi(k)$, the signals transmitted by nodes $\pi(k+1 : n+1)$ are known *a priori*, and the signals transmitted by nodes $\pi(1 : k-1)$ are cooperating in providing the information. A simple explanation of this feasibility is the following. In the multi-level D-F relay strategy, information is passed along the route $\pi(1) \to \pi(2) \to \cdots \to \pi(n+2)$, so that i) any information obtained by the downstream nodes of $\pi(k)$, i.e., nodes $\pi(k+1 : n+1)$, has already been obtained by node $\pi(k)$, and therefore their inputs are predictable by node $\pi(k)$, and ii) by the time the information reaches node $\pi(k)$, all its upstream nodes $\pi(1 : k-1)$ have already obtained the same information and can therefore cooperate with the technique of block Markov coding. The formula (2.16) also demonstrates a remarkable feature of the multi-level D-F relay strategy in [27]-[30], i.e., it completely eliminates the interference in the network: To any node, the signal transmitted by any other node is either a "real" signal that can be used for decoding, or a *priori* known signal that can be subtracted completely.

# Chapter 3

# C-F Relay Schemes with Forward Decoding

## 3.1 Introduction

This chapter discusses the conventional C-F relay schemes with forward decoding. Particularly, the following two C-F coding schemes will be considered.

- Cumulative encoding/block-by-block forward decoding/compression-message successive decoding;
- Cumulative encoding/block-by-block forward decoding/compression-message joint decoding.

The cumulative encoding/block-by-block forward decoding/compression-message successive decoding refers to the original C-F scheme developed in [15]. The encoding is "cumulative" in the sense that in each new block, a new piece of information is encoded at the source. This distinguishes from a "repetitive" encoding process in the recently proposed noisy network coding scheme [32], where the same information is encoded in each block. The decoding is named "block-by-block forward" to distinguish from the other two choices, where the decoding starts only after all the blocks have been finished, either by decoding with all the blocks together as in the noisy network coding scheme [32], or by a block-by-block backward decoding process that will be developed in Chapter 4. The decoding is also called “compression-message successive” in the sense that the destination first decodes the compression of the relay’s observation, and then decodes the original message. In the single-relay case as depicted in Figure 1.4, the compression $\hat{Y}_1$ can be first recovered at the destination, as long as the following constraint is satisfied:

$$I(Y_1; \hat{Y}_1 | X_1, Y_2) \leq I(X_1; Y_2). \tag{3.1}$$

Then, based on $\hat{Y}_1$ and $Y_2$, the destination can decode the original message $X_0$ if the rate of the original message satisfies

$$R < I(X_0; \hat{Y}_1, Y_2 | X_1). \tag{3.2}$$

The above two-step compression-message successive decoding process requires $\hat{Y}_1$ to be decoded first. This facilitates the decoding of $X_0$, but is not a requirement of the original problem. Recognizing this, a joint compression-message decoding process was proposed in [19], where, instead of successively, the destination decodes $\hat{Y}_1$ and $X_0$ together. It turns out that the decoding of $X_0$ can be helped even if $\hat{Y}_1$ cannot be decoded first. In fact, with joint decoding, the constraint (3.1) is not necessary, and instead of (3.2), the achievable rate is expressed as

$$R < I(X_0; \hat{Y}_1, Y_2 | X_1) - \max\{0, I(Y_1; \hat{Y}_1 | X_1, Y_2) - I(X_1; Y_2)\}. \tag{3.3}$$

Moreover, although $\hat{Y}_1$ is not even required to be decoded eventually, it can be more easily decoded by joint decoding, and instead of (3.1), we need a less strict constraint:

$$I(X_1; Y_2) > I(Y_1; \hat{Y}_1 | X_1, Y_2, X_0), \tag{3.4}$$

where, it is clear to see the assistance provided by $X_0$.

Similar formulas as (3.3) have been derived with different arguments in [20]-[22].[1]

Therefore, compared to successive decoding, joint compression-message decoding provides more freedom in choosing the compression $\hat{Y}_1$. However, the question

[1]The formula and proof in [20] missed a $Y$, and were later corrected in [22].

remains whether joint decoding achieves strictly higher rates for the original message than successive decoding. For the single relay case, it has been proved in [22] that the answer is negative, and any rate achievable by either of them can always be achieved by the other. In this chapter, we are going to further consider the case of multiple relays as depicted in Figure 1.5, and demonstrate that joint decoding will not be able to achieve any higher rates either. More interestingly, we will show that any compressions not supporting successive decoding will actually result in strictly lower achievable rates for the original message. Therefore, to optimize the achievable rate, the compressions should always be chosen so that successive decoding can be carried out.

Although the compressions supporting successive decoding can be explicitly characterized as we will show later, it is also of interest to consider other compressions not supporting successive decoding. For example, in a network with multiple destinations, when a relay is simultaneously helping more than one destinations, it is very likely that different destinations require different optimal compressions from the relay. In such a situation, the relay may have to find a tradeoff between these requirements, i.e., adopting a compression which may be too coarse for some destinations, but too fine, thus not supporting successive decoding, for the others. An example of this tradeoff to optimize the sum rate was given for the two-way relay channel in [32]. Another possibility of using too coarse or too fine compressions is when there is channel uncertainty, e.g., in wireless fading channels, so that it is impossible to accurately determine the optimal compressions even with explicit formulas. In these scenarios, compression-message joint decoding introduces flexibility in choosing the compression at the relay, which improves the robustness of the C-F relay scheme. Such an optimality-robustness tradeoff in C-F schemes will be discussed in Section 3.5.

It is not surprising that coarser compressions than the optimal do not fully exploit the capability of the relay, thus leading to lower achievable rates for the original message. However, it may not be so obvious why finer compressions will also lead to lower achievable rates. For this, one needs to realize that a relay's observation not only carries information about the original message, but also reflects the dynamics of the source-relay link, which is unrelated to the original message. Thus, compared to the direct link between the source and the destination, the support by the relay-destination link is not so pure. When the compression is too fine so that only joint compression-message decoding can be carried out, i.e., the direct source-destination link has to sacrifice, the gain does not make up for the loss, thus leading to lower achievable rates.

The remainder of the chapter is organized as the following. We first summarize the main results in Section 3.2. Then, detailed proofs of the achievability results and investigations on the optimal choice of the relays' compressions are presented in Section 3.3 and Section 3.4, respectively. Finally, discussions on the optimality-robustness tradeoff are included in Section 3.5.

***Notation:*** In this chapter and the rest of the thesis, we denote, for any subset $\mathcal{S} \subseteq \mathcal{N}$, $X_{\mathcal{S}} = \{X_i, i \in \mathcal{S}\}$, and $X_{\mathcal{S}^c} = \{X_i, i \in \mathcal{N} \setminus \mathcal{S}\}$. Similar notations are used for other variables.

## 3.2 Main Results

Under the block-by-block forward decoding framework, the achievable rates with successive compression-message decoding and with joint compression-message decoding are presented in Theorems 3.2.1 and 3.2.2 respectively. Then the optimality of successive decoding is stated in Theorem 3.2.3, which shows that the optimal rate can be achieved only if the compressions at the relays are chosen such that they can be first decoded at the destination, i.e., successive compression-message decoding can be carried out. The proofs of Theorems 3.2.1-3.2.2 and 3.2.3 are presented in Section 3.3 and Section 3.4, respectively.

**Theorem 3.2.1.** *For the multiple-relay channel depicted in Figure 1.5, by the cumulative encoding/block-by-block forward decoding/compression-message successive* *decoding scheme, a rate $R_{C/F/S}$ is achievable if for some*

$$p(q)p(x_0|q)p(x_1|q)\cdots p(x_n|q)p(\hat{y}_1|y_1,x_1,q)\cdots p(\hat{y}_n|y_n,x_n,q),$$

*there exists a rate vector $\{R_i, i=1,\ldots,n\}$ satisfying*

$$\sum_{i\in\mathcal{S}_1} R_i \leq I(X_{\mathcal{S}_1};Y_{n+1}|X_{\mathcal{S}_1^c},Q) \tag{3.5}$$

*for any subset $\mathcal{S}_1 \subseteq \mathcal{N}$, such that for any subset $\mathcal{S} \subseteq \mathcal{N}$,*

$$I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|\hat{Y}_{\mathcal{S}^c},Y_{n+1},X_{\mathcal{N}},Q) \leq \sum_{i\in\mathcal{S}} R_i \tag{3.6}$$

*and*

$$R_{C/F/S} < I(X_0;\hat{Y}_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}},Q). \tag{3.7}$$

**Theorem 3.2.2.** *For the multiple-relay channel depicted in Figure 1.5, by the cumulative encoding/block-by-block forward decoding/compression-message joint decoding scheme, a rate $R_{C/F/J}$ is achievable if for some*

$$p(q)p(x_0|q)p(x_1|q)\cdots p(x_n|q)p(\hat{y}_1|y_1,x_1,q)\cdots p(\hat{y}_n|y_n,x_n,q),$$

*there exists a rate vector $\{R_i, i=1,\ldots,n\}$ satisfying*

$$\sum_{i\in\mathcal{S}_1} R_i \leq I(X_{\mathcal{S}_1};Y_{n+1}|X_{\mathcal{S}_1^c},Q) \tag{3.8}$$

*for any subset $\mathcal{S}_1 \subseteq \mathcal{N}$, such that for any subset $\mathcal{S} \subseteq \mathcal{N}$,*

$$R_{C/F/J} < I(X_0;\hat{Y}_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}},Q) - I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|\hat{Y}_{\mathcal{S}^c},Y_{n+1},X_{\mathcal{N}},Q) + \sum_{i\in\mathcal{S}} R_i. \tag{3.9}$$

One can easily check that the achievable rates stated in Theorems 3.2.1 and 3.2.2 include the achievable rates (3.1)-(3.2) and (3.3) for the single-relay channel as special cases.

Let $R^*_{\mathrm{C/F/S}}$ and $R^*_{\mathrm{C/F/J}}$ be the supremum of the achievable rates stated in Theorems 3.2.1 and 3.2.2 respectively.

**Theorem 3.2.3.** *$R^*_{C/F/S} = R^*_{C/F/J}$, and $R^*_{C/F/J}$ can be obtained only when the distribution*

$$p(q)p(x_0|q)p(x_1|q)\cdots p(x_n|q)p(\hat{y}_1|y_1,x_1,q)\cdots p(\hat{y}_n|y_n,x_n,q)$$

*is chosen such that there exists a rate vector $\{R_i, i=1,\ldots,n\}$ satisfying (3.5)-(3.6).*

## 3.3 Generalization of C-F Schemes with Forward Decoding to Multiple-Relay Case

In this section, we prove the achievability results stated in Theorems 3.2.1 and 3.2.2. For simplicity of notation, we consider the case $Q = \emptyset$. Achievability for an arbitrary time-sharing random variable $Q$ can be obtained by using the standard technique of time sharing [44], [22]. The same consideration on $Q$ applies throughout all the achievability proofs of this thesis.

In both the cumulative encoding/block-by-block forward decoding/compression-message successive decoding and the cumulative encoding/block-by-block forward decoding/compression-message joint decoding schemes, the codebook generation and encoding processes are exactly the same as the classical way, i.e., the way in the proof of Theorem 6 of [15]. The difference between these two schemes is only on the decoding process at the destination: i) In successive decoding, the destination first finds, from the specific bins sent by the relays via $X_1, X_2, \ldots, X_n$, the unique combination of $\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n$ sequences that is jointly typical with the $Y_{n+1}$ sequence received, and then finds the unique $X_0$ sequence that is jointly typical with the $Y_{n+1}$ sequence received, and also with the previously recovered $\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n$ sequences. ii) In joint decoding, the destination finds the unique $X_0$ sequence that is jointly typical with the $Y_{n+1}$ sequence received, and also with some combination of $\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n$ sequences from the specific bins sent by the relays via $X_1, X_2, \ldots, X_n$.

### 3.3.1 Proof of Theorem 3.2.1

The basic idea of the C-F scheme is for the relay to compress its observations into some approximations, which can be represented by fewer number of bits, and thus, can be forwarded to the destination. To deal with delay at the relay, block Markov coding was also used, where the total time is divided into a sequence of blocks of equal length $T$, and coding is performed block by block. For example, each relay compresses its observations of each block at the end of the block, and forwards the approximations in the next block. Therefore, to decode the message sent by the source in any block, it is not until the end of the next block, has the destination received the help from the relay.

We now prove that the rate in Theorem 3.2.1 is achievable with cumulative encoding/block-by-block forward decoding/compression-message successive decoding. The encoding process is exactly the same as that in the proof of Theorem 6 of [15]. We only emphasize that the $i$-th relay needs to generate $2^{T(I(Y_i;\hat{Y}_i|X_i)+\epsilon)}$ sequences of $\hat{Y}_i$, and randomly throws them into $2^{TR_i}$ bins, where $\{R_i, i = 1, 2, \ldots, n\}$ are chosen such that for any nonempty subset $\mathcal{S}_1 \subseteq \mathcal{N}$,

$$\sum_{i \in \mathcal{S}_1} R_i < I(X_{\mathcal{S}_1}; Y_{n+1}|X_{\mathcal{S}_1^c}). \tag{3.10}$$

At the end of each block, the relay finds a $\hat{Y}_i$ sequence which is jointly typical with the $Y_i$ sequence it received and the $X_i$ sequence it sent during the block, and in the next block, informs the destination the index of the bin that contains the $\hat{Y}_i$ sequence via $X_i$.

The decoding process operates in a successive way. At the end of each block $b = 2, 3, \ldots$, the destination first finds, from the bins forwarded by the relays during block $b$, the unique $(\underline{\hat{Y}}_1(b-1), \ldots, \underline{\hat{Y}}_n(b-1))$ such that

$$\begin{aligned}&\Big(\underline{Y}_{n+1}(b-1), (\underline{X}_1(b-2), \underline{\hat{Y}}_1(b-1)), \ldots, (\underline{X}_n(b-2), \underline{\hat{Y}}_n(b-1))\Big)\\ &\in A_\epsilon(Y_{n+1}, X_{\mathcal{N}}, \hat{Y}_{\mathcal{N}})\end{aligned} \tag{3.11}$$

where $\underline{Y}_{n+1}(b-1)$ is the $Y_{n+1}$ sequence received during block $b-1$, $(\underline{\hat{Y}}_1(b-1), \ldots, \underline{\hat{Y}}_n(b-1))$ are the $\hat{Y}_1, \ldots, \hat{Y}_n$ sequences from the bins forwarded by the relays during block $b$, and $(\underline{X}_1(b-2), \ldots, \underline{X}_n(b-2))$ are the signals sent by the relays at block $b-1$ which are known to the destination since the multiple-access condition (3.10) is satisfied.

Error occurs if the true $\underline{\hat{Y}}_{\mathcal{N}}(b-1)$ does not satisfy (3.11), or a false $\underline{\hat{Y}}_{\mathcal{N}}(b-1)$ satisfies (3.11). According to the properties of typical sequences, the true $\underline{\hat{Y}}_{\mathcal{N}}(b-1)$ satisfies (3.11) with high probability.

The probability of a false $\underline{\hat{Y}}_{\mathcal{N}}(b-1)$ with some false $\{\underline{\hat{Y}}_i(b-1),\ i \in \mathcal{S}\}$ but true $\{\underline{\hat{Y}}_i(b-1),\ i \in \mathcal{S}^c\}$ being jointly typical with $\underline{Y}_{n+1}(b-1)$ and $(\underline{X}_1(b-2), \ldots, \underline{X}_n(b-2))$ can be upper bounded by

$$2^{T(H(Y_{n+1},\hat{Y}_{\mathcal{N}},X_{\mathcal{N}})+\epsilon)}2^{-T(H(Y_{n+1},\hat{Y}_{\mathcal{S}^c},X_{\mathcal{N}})-\epsilon)}\prod_{i\in\mathcal{S}}2^{-T(H(\hat{Y}_i|X_i)-\epsilon)}.$$

There are $\prod_{i\in\mathcal{S}}(2^{T(I(Y_i;\hat{Y}_i|X_i)-R_i+\epsilon)}-1)$ false $\underline{\hat{Y}}_{\mathcal{S}}(b-1)$ from the bins, thus the probability of finding such a false $\underline{\hat{Y}}_{\mathcal{N}}(b-1)$ can be upper bounded by

$$2^{T(H(Y_{n+1},\hat{Y}_{\mathcal{N}},X_{\mathcal{N}})+\epsilon)}2^{-T(H(Y_{n+1},\hat{Y}_{\mathcal{S}^c},X_{\mathcal{N}})-\epsilon)}\prod_{i\in\mathcal{S}}2^{-T(H(\hat{Y}_i|X_i)-I(Y_i;\hat{Y}_i|X_i)+R_i-2\epsilon)},$$

which tends to zero for sufficiently small $\epsilon$ as $T\to\infty$, if

$$H(\hat{Y}_{\mathcal{S}}|Y_{n+1},\hat{Y}_{\mathcal{S}^c},X_{\mathcal{N}})-\sum_{i\in\mathcal{S}}[H(\hat{Y}_i|Y_i,X_i)+R_i]<0. \tag{3.12}$$

Letting $\mathcal{S}=\{i_j\in\mathcal{N}: j=1,\ldots,|\mathcal{S}|\}$, we have

$$\begin{aligned}\sum_{i\in\mathcal{S}}H(\hat{Y}_i|Y_i,X_i) &= \sum_{j=1,\ldots,|\mathcal{S}|}H(\hat{Y}_{i_j}|Y_{i_j},X_{i_j})\\ &= \sum_{j=1,\ldots,|\mathcal{S}|}H(\hat{Y}_{i_j}|Y_{\mathcal{S}},Y_{n+1},\hat{Y}_{\mathcal{S}^c},X_{\mathcal{N}},\{\hat{Y}_{i_1},\ldots,\hat{Y}_{i_{j-1}}\})\\ &= H(\hat{Y}_{\mathcal{S}}|Y_{\mathcal{S}},Y_{n+1},\hat{Y}_{\mathcal{S}^c},X_{\mathcal{N}}).\end{aligned}$$

Plugging this into (3.12), we have $\underline{\hat{Y}}_{\mathcal{N}}(b-1)$ can be decoded at the end of block $b$ if

$$I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|\hat{Y}_{\mathcal{S}^c},Y_{n+1},X_{\mathcal{N}})<\sum_{i\in\mathcal{S}}R_i. \tag{3.13}$$

Then, based on $\underline{\hat{Y}}_{\mathcal{N}}(b-1)$ and $\underline{Y}_{n+1}(b-1)$, $\underline{X}_0(w)$ can be recovered if

$$R_{\text{C/F/S}}<I(X_0;\hat{Y}_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}}). \tag{3.14}$$

Combining (3.10) and (3.13)-(3.14), and using the standard technique of time sharing, we conclude that the rate stated in Theorem 3.2.1 is achievable.[2]

[2]The case of "=" in (3.5)-(3.6) can be included since (3.7) doesn't include "=". The same consideration applies throughout the thesis.

### 3.3.2 Proof of Theorem 3.2.2

In cumulative encoding/block-by-block forward decoding/compression-message joint decoding, the encoding part is exactly the same as that in the proof of Theorem 3.2.1, and the decoding process operates as the following. At the end of each block $b = 2, 3, \ldots$, the destination finds the unique $\underline{X}_0(w)$ and some $(\underline{\hat{Y}}_1(b-1), \ldots, \underline{\hat{Y}}_n(b-1))$ from the bins forwarded by the relays during block $b$ such that

$$\begin{aligned}&\Big(\underline{X}_0(w), \underline{Y}_{n+1}(b-1), (\underline{X}_1(b-2), \underline{\hat{Y}}_1(b-1)), \ldots, (\underline{X}_n(b-2), \underline{\hat{Y}}_n(b-1))\Big)\\ &\in A_\epsilon(X, Y, X_\mathcal{N}, \hat{Y}_\mathcal{N})\end{aligned} \tag{3.15}$$

where $\underline{Y}_{n+1}(b-1)$, $(\underline{\hat{Y}}_1(b-1), \ldots, \underline{\hat{Y}}_n(b-1))$, and $(\underline{X}_1(b-2), \ldots, \underline{X}_n(b-2))$ have the same interpretations as in (3.11).

Error occurs if the true $\underline{X}_0(w)$ does not satisfy (3.15), or a false $\underline{X}_0(w')$ satisfies (3.15). According to the properties of typical sequences, the true $\underline{X}_0(w)$ satisfies (3.15) with high probability.

The probability of a false $\underline{X}_0(w')$ being jointly typical with $\underline{Y}_{n+1}(b-1)$, $(\underline{X}_1(b-2), \ldots, \underline{X}_n(b-2))$, and some false $\{\underline{\hat{Y}}_i(b-1),\, i \in \mathcal{S}\}$ but true $\{\underline{\hat{Y}}_i(b-1),\, i \in \mathcal{S}^c\}$ can be upper bounded by

$$2^{T(H(X_0, Y_{n+1}, X_\mathcal{N}, \hat{Y}_\mathcal{N})+\epsilon)} 2^{-T(H(X_0)-\epsilon)} 2^{-T(H(Y_{n+1}, X_\mathcal{N}, \hat{Y}_{\mathcal{S}^c})-\epsilon)} \prod_{i\in\mathcal{S}} 2^{-T(H(\hat{Y}_i|X_i)-\epsilon)}.$$

There are $2^{TR_{\mathrm{C/F/J}}} - 1$ false $w'$, and $\prod_{i\in\mathcal{S}}(2^{T(I(Y_i;\hat{Y}_i|X_i)-R_i+\epsilon)} - 1)$ false $\underline{\hat{Y}}_\mathcal{S}(b-1)$ from the bins, thus the probability of finding such a false $\underline{X}(w')$ can be upper bounded by

$$\begin{aligned}&2^{TR_{\mathrm{C/F/J}}} 2^{T(H(X_0, Y_{n+1}, X_\mathcal{N}, \hat{Y}_\mathcal{N})+\epsilon)} 2^{-T(H(X_0)-\epsilon)}\\ &\times 2^{-T(H(Y_{n+1}, X_\mathcal{N}, \hat{Y}_{\mathcal{S}^c})-\epsilon)} \prod_{i\in\mathcal{S}} 2^{-T(H(\hat{Y}_i|X_i)-I(Y_i;\hat{Y}_i|X_i)+R_i-2\epsilon)},\end{aligned}$$

which tends to zero for sufficiently small $\epsilon$ as $T \to \infty$, if

$$R_{\mathrm{C/F/J}} < I(X_0; \hat{Y}_\mathcal{N}, Y_{n+1}|X_\mathcal{N}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S}|\hat{Y}_{\mathcal{S}^c}, Y_{n+1}, X_\mathcal{N}) + \sum_{i\in\mathcal{S}} R_i. \tag{3.16}$$

This combined with the technique of time sharing proves Theorem 3.2.2.

## 3.4 Optimality of Successive Decoding in Multiple-Relay Case

Before proceeding to the proof of Theorem 3.2.3, we first introduce some useful notations and lemmas. For any $\mathcal{A}, \mathcal{B} \subseteq \mathcal{N}$ and $\{R_i, i = 1, \ldots, n\}$, let

$$I_{\mathcal{A},\mathcal{B}}(\mathcal{S}) := \sum_{i \in \mathcal{S}} R_i - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}}, Y_{n+1}, X_{\mathcal{N}}), \forall \mathcal{S} \subseteq \mathcal{B}, \tag{3.17}$$

$$I_{\mathcal{B}}(\mathcal{S}) := I_{\emptyset,\mathcal{B}}(\mathcal{S}) = \sum_{i \in \mathcal{S}} R_i - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | \hat{Y}_{\mathcal{B} \setminus \mathcal{S}}, Y_{n+1}, X_{\mathcal{N}}), \forall \mathcal{S} \subseteq \mathcal{B}, \tag{3.18}$$

$$I(\mathcal{S}) := I_{\mathcal{N}}(\mathcal{S}) = \sum_{i \in \mathcal{S}} R_i - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | \hat{Y}_{\mathcal{S}^c}, Y_{n+1}, X_{\mathcal{N}}), \forall \mathcal{S} \subseteq \mathcal{N}. \tag{3.19}$$

Also, in the following proof and the rest of the thesis, for any two sets $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{A} \cap \mathcal{B}$ or $\mathcal{A}\mathcal{B}$ interchangeably denotes their intersection while $\mathcal{A} \cup \mathcal{B}$ denotes their union. Then, we have the following lemmas, whose proofs will be given until we finish the proof of Theorem 3.2.3.

**Lemma 3.4.1.** *1) If $I_{\mathcal{A}}(\mathcal{S}_1) \geq 0$, $\forall \mathcal{S}_1 \subseteq \mathcal{A}$, and $I_{\mathcal{B}}(\mathcal{S}_2) \geq 0$, $\forall \mathcal{S}_2 \subseteq \mathcal{B}$, then $I_{\mathcal{A} \cup \mathcal{B}}(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{A} \cup \mathcal{B}$.*

*2) If $I_{\mathcal{A}}(\mathcal{S}_1) \geq 0$, $\forall \mathcal{S}_1 \subseteq \mathcal{A}$, and $I_{\mathcal{A},\mathcal{B}}(\mathcal{S}_2) \geq 0$, $\forall \mathcal{S}_2 \subseteq \mathcal{B}$, then $I_{\mathcal{A} \cup \mathcal{B}}(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{A} \cup \mathcal{B}$.*

**Lemma 3.4.2.** *For any $p(x_0) \prod_{i=1}^{n} p(x_i) p(\hat{y}_i | y_i, x_i)$ and $\{R_i, i = 1, \ldots, n\}$, there exists a unique set $\mathcal{D}$, which is the largest subset of $\mathcal{N}$ satisfying*

$$I_{\mathcal{D}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{D}.$$

**Lemma 3.4.3.** *If $I_{\mathcal{A},\mathcal{B}}(\mathcal{B}) \geq 0$ for some nonempty $\mathcal{B}$, then there exists some nonempty $\mathcal{C} \subseteq \mathcal{B}$ such that $I_{\mathcal{A},\mathcal{C}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{C}$.*

**Lemma 3.4.4.** *For any $\mathcal{A}$ and $\mathcal{B}$ with $\mathcal{A} \cap \mathcal{B} = \emptyset$, $I(\mathcal{A}) + I(\mathcal{B}) = I(\mathcal{A} \cup \mathcal{B}) + I(\hat{Y}_{\mathcal{A}}; \hat{Y}_{\mathcal{B}} | \hat{Y}_{(\mathcal{A} \cup \mathcal{B})^c}, Y_{n+1}, X_{\mathcal{N}})$.*

We are now ready to prove Theorem 3.2.3. Still for simplicity of notation, we only prove Theorem 3.2.3 for $Q = \emptyset$, while the proof for an arbitrary $Q$ can be obtained by simple analogy. The same consideration on $Q$ also applies to the proofs of Theorems 4.2.3 and 4.2.5.

*Proof of Theorem 3.2.3.* With $Q = \emptyset$, $R^*_{\text{C/F/S}}$ and $R^*_{\text{C/F/J}}$ can be respectively written as

$$R^*_{\text{C/F/S}} = \max_{p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|y_i,x_i),\{R_i,i=1,\ldots,n\}} I(X_0;\hat{Y}_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}}) \quad (3.20)$$

$$\text{such that } \sum_{i\in\mathcal{S}_1} R_i \le I(X_{\mathcal{S}_1};Y_{n+1}|X_{\mathcal{S}_1^c}), \forall \mathcal{S}_1 \subseteq \mathcal{N} \quad (3.21)$$

$$I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|\hat{Y}_{\mathcal{S}^c},Y_{n+1},X_{\mathcal{N}}) - \sum_{i\in\mathcal{S}} R_i \le 0, \forall \mathcal{S} \subseteq \mathcal{N} \quad (3.22)$$

and

$$R^*_{\text{C/F/J}} = \max_{p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|y_i,x_i),\{R_i,i=1,\ldots,n\}} \min_{\mathcal{S}\subseteq\mathcal{N}}\{I(X_0;\hat{Y}_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}}) - I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|\hat{Y}_{\mathcal{S}^c},Y_{n+1},X_{\mathcal{N}}) + \sum_{i\in\mathcal{S}} R_i\}$$

$$\text{such that } \sum_{i\in\mathcal{S}_1} R_i \le I(X_{\mathcal{S}_1};Y_{n+1}|X_{\mathcal{S}_1^c}), \forall \mathcal{S}_1 \subseteq \mathcal{N}.$$

We show $R^*_{\text{C/F/S}} = R^*_{\text{C/F/J}}$ by showing that $R^*_{\text{C/F/S}} \le R^*_{\text{C/F/J}}$ and $R^*_{\text{C/F/S}} \ge R^*_{\text{C/F/J}}$ respectively. For any $p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|y_i,x_i)$ and $\{R_i, i=1,\ldots,n\}$ satisfying (3.21)-(3.22), we have

$$\min_{\mathcal{S}\subseteq\mathcal{N}}\{I(X_0;\hat{Y}_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}}) - I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|\hat{Y}_{\mathcal{S}^c},Y_{n+1},X_{\mathcal{N}}) + \sum_{i\in\mathcal{S}} R_i\} = I(X_0;\hat{Y}_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}}),$$

and thus $R^*_{\text{C/F/S}} \le R^*_{\text{C/F/J}}$.

To show $R^*_{\text{C/F/S}} \ge R^*_{\text{C/F/J}}$, it is sufficient to show that $R^*_{\text{C/F/J}}$ can be achieved only with $p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|y_i,x_i)$ and $\{R_i, i=1,\ldots,n\}$ such that $I(\mathcal{S}) \ge 0$, $\forall \mathcal{S} \subseteq \mathcal{N}$. We will show this by two steps as follows: i) We first show that for any $p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|y_i,x_i)$ and $\{R_i, i=1,\ldots,n\}$, if $\mathcal{D}^c \ne \emptyset$, then $\mathcal{D}^c \in \operatorname{argmin}_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})$ and $\bigcap_{\mathcal{T}\in\operatorname{argmin}_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})} \mathcal{T} = \mathcal{D}^c$, where $\mathcal{D}$ is defined as in Lemma 3.4.2 and $\operatorname{argmin}_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S}) := \{\mathcal{T} \subseteq \mathcal{N} : I(\mathcal{T}) = \min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})\}$. ii) We then argue that, under the optimal choice of $p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|y_i,x_i)$ and $\{R_i, i=1,\ldots,n\}$, $\mathcal{D}^c$ must be $\emptyset$, i.e., $\mathcal{D}$ must be $\mathcal{N}$, and thus by the definition of $\mathcal{D}$, $I(\mathcal{S}) \ge 0, \forall \mathcal{S} \subseteq \mathcal{N}$.

i) Assuming $\mathcal{D}^c \ne \emptyset$ throughout Part i), we show $\mathcal{D}^c \in \operatorname{argmin}_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})$ and $\bigcap_{\mathcal{T}\in\operatorname{argmin}_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})} \mathcal{T} = \mathcal{D}^c$.

1) We first show $I(\mathcal{D}^c) < 0$ by using a contradiction argument. Suppose $I(\mathcal{D}^c) \ge 0$, i.e., $I_{\mathcal{D},\mathcal{D}^c}(\mathcal{D}^c) \ge 0$. Then, by Lemma 3.4.3, we have that there exists some nonempty $\mathcal{B} \subseteq \mathcal{D}^c$ such that $I_{\mathcal{D},\mathcal{B}}(\mathcal{S}) \ge 0$, $\forall \mathcal{S} \subseteq \mathcal{B}$. This will further imply, by Part 2) of Lemma 3.4.1, that $I_{\mathcal{D}\cup\mathcal{B}}(\mathcal{S}) \ge 0, \forall \mathcal{S} \subseteq \mathcal{D}\cup\mathcal{B}$. This is contradictory with the definition of $\mathcal{D}$, and thus $I(\mathcal{D}^c) < 0$.

2) We show that $\forall \mathcal{A} \subseteq \mathcal{D}^c$ and $\mathcal{A} \neq \mathcal{D}^c$, $I(\mathcal{A}) > I(\mathcal{D}^c)$, and thus $I(\mathcal{A}) > \min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})$. The proof is still by contradiction. Suppose that there exists some $\mathcal{A} \subseteq \mathcal{D}^c$ and $\mathcal{A} \neq \mathcal{D}^c$ such that $I(\mathcal{A}) \leq I(\mathcal{D}^c)$. Then $I(\mathcal{D}^c) - I(\mathcal{A}) \geq 0$, i.e.,

$$\begin{aligned}
&\sum_{i\in\mathcal{D}^c} R_i - I(Y_{\mathcal{D}^c};\hat{Y}_{\mathcal{D}^c}|\hat{Y}_{\mathcal{D}}, Y_{n+1}, X_{\mathcal{N}}) - \sum_{i\in\mathcal{A}} R_i + I(Y_{\mathcal{A}};\hat{Y}_{\mathcal{A}}|\hat{Y}_{\mathcal{A}^c}, Y_{n+1}, X_{\mathcal{N}}) \\
=&\sum_{i\in\mathcal{D}^c\setminus\mathcal{A}} R_i - I(Y_{\mathcal{D}^c\setminus\mathcal{A}};\hat{Y}_{\mathcal{D}^c\setminus\mathcal{A}}|\hat{Y}_{\mathcal{D}}, Y_{n+1}, X_{\mathcal{N}}) \\
=&I_{\mathcal{D},\mathcal{D}^c\setminus\mathcal{A}}(\mathcal{D}^c\setminus\mathcal{A}) \\
\geq&0.
\end{aligned}$$

Again by Lemma 3.4.3 and 3.4.1 successively, we can conclude that there exists some nonempty $\mathcal{B} \subseteq \mathcal{D}^c \setminus \mathcal{A}$, such that $I_{\mathcal{D}\cup\mathcal{B}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{D}\cup\mathcal{B}$, which is in contradiction. Therefore, $I(\mathcal{A}) > I(\mathcal{D}^c) \geq \min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})$.

3) We prove that $\forall\mathcal{A}$ with $\mathcal{A}\mathcal{D} \neq \emptyset$ and $\mathcal{A}\mathcal{D}^c \neq \mathcal{D}^c$, $I(\mathcal{A}) > \min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})$. Let $\mathcal{A}_1 = \mathcal{A}\mathcal{D}$ and $\mathcal{A}_2 = \mathcal{A}\mathcal{D}^c$. Then, we have, by Lemma 3.4.4, that

$$\begin{aligned}
I(\mathcal{A}) =&I(\mathcal{A}_1\cup\mathcal{A}_2) = I(\mathcal{A}_1) + I(\mathcal{A}_2) - I(\hat{Y}_{\mathcal{A}_1};\hat{Y}_{\mathcal{A}_2}|\hat{Y}_{\mathcal{A}^c}, Y_{n+1}, X_{\mathcal{N}}), \\
I(\mathcal{A}_1\cup\mathcal{D}^c) =&I(\mathcal{A}_1) + I(\mathcal{D}^c) - I(\hat{Y}_{\mathcal{A}_1};\hat{Y}_{\mathcal{D}^c}|\hat{Y}_{(\mathcal{A}_1\cup\mathcal{D}^c)^c}, Y_{n+1}, X_{\mathcal{N}}).
\end{aligned}$$

Since $I(\mathcal{A}_2) > I(\mathcal{D}^c)$ by 2) and

$$\begin{aligned}
&I(\hat{Y}_{\mathcal{A}_1};\hat{Y}_{\mathcal{D}^c}|\hat{Y}_{(\mathcal{A}_1\bigcup\mathcal{D}^c)^c}, Y_{n+1}, X_{\mathcal{N}}) \\
=&I(\hat{Y}_{\mathcal{A}_1};\hat{Y}_{\mathcal{D}^c\setminus\mathcal{A}_2}|\hat{Y}_{(\mathcal{A}_1\bigcup\mathcal{D}^c)^c}, Y_{n+1}, X_{\mathcal{N}}) + I(\hat{Y}_{\mathcal{A}_1};\hat{Y}_{\mathcal{A}_2}|\hat{Y}_{(\mathcal{A}_1\bigcup\mathcal{D}^c)^c}, \hat{Y}_{\mathcal{D}^c\setminus\mathcal{A}_2}, Y_{n+1}, X_{\mathcal{N}}) \\
=&I(\hat{Y}_{\mathcal{A}_1};\hat{Y}_{\mathcal{A}_2}|\hat{Y}_{\mathcal{A}^c}, Y_{n+1}, X_{\mathcal{N}}) + I(\hat{Y}_{\mathcal{A}_1};\hat{Y}_{\mathcal{D}^c\setminus\mathcal{A}_2}|\hat{Y}_{(\mathcal{A}_1\bigcup\mathcal{D}^c)^c}, Y_{n+1}, X_{\mathcal{N}}) \\
\geq&I(\hat{Y}_{\mathcal{A}_1};\hat{Y}_{\mathcal{A}_2}|\hat{Y}_{\mathcal{A}^c}, Y_{n+1}, X_{\mathcal{N}}),
\end{aligned}$$

we have $I(\mathcal{A}) > I(\mathcal{A}_1\cup\mathcal{D}^c) \geq \min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})$.

4) We prove that $\forall\mathcal{A}$ with $\mathcal{A}\mathcal{D} \neq \emptyset$ and $\mathcal{A}\mathcal{D}^c = \mathcal{D}^c$, $I(\mathcal{A}) \geq I(\mathcal{D}^c)$. Letting $\mathcal{A}_1 = \mathcal{A}\mathcal{D}$, we have

$$\begin{aligned}
I(\mathcal{A}) =&I(\mathcal{A}_1\cup\mathcal{D}^c) \\
=&I(\mathcal{A}_1) + I(\mathcal{D}^c) - I(\hat{Y}_{\mathcal{A}_1};\hat{Y}_{\mathcal{D}^c}|\hat{Y}_{(\mathcal{A}_1\cup\mathcal{D}^c)^c}, Y_{n+1}, X_{\mathcal{N}}) \\
=&\sum_{i\in\mathcal{A}_1} R_i - I(Y_{\mathcal{A}_1};\hat{Y}_{\mathcal{A}_1}|\hat{Y}_{\mathcal{A}_1^c}, Y_{n+1}, X_{\mathcal{N}}) - I(\hat{Y}_{\mathcal{A}_1};\hat{Y}_{\mathcal{D}^c}|\hat{Y}_{(\mathcal{A}_1\cup\mathcal{D}^c)^c}, Y_{n+1}, X_{\mathcal{N}}) + I(\mathcal{D}^c) \\
=&\sum_{i\in\mathcal{A}_1} R_i - I(\hat{Y}_{\mathcal{A}_1};\hat{Y}_{\mathcal{D}^c}, Y_{\mathcal{A}_1}|\hat{Y}_{(\mathcal{A}_1\cup\mathcal{D}^c)^c}, Y_{n+1}, X_{\mathcal{N}}) + I(\mathcal{D}^c) \\
=&\sum_{i\in\mathcal{A}_1} R_i - I(\hat{Y}_{\mathcal{A}_1};Y_{\mathcal{A}_1}|\hat{Y}_{\mathcal{D}\setminus\mathcal{A}_1}, Y_{n+1}, X_{\mathcal{N}}) + I(\mathcal{D}^c) \\
=&I_{\mathcal{D}}(\mathcal{A}_1) + I(\mathcal{D}^c) \\
\geq&I(\mathcal{D}^c).
\end{aligned}$$

Combining 2)-4), we can conclude that

$$\mathcal{D}^c \in \operatorname*{argmin}_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})$$

and

$$\bigcap_{\mathcal{T}\in \operatorname*{argmin}_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})} \mathcal{T} = \mathcal{D}^c.$$

ii) We now argue that under the optimal choice of $p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|y_i,x_i)$ and $\{R_i, i=1,\dots,n\}$ that achieves $R^*_{\text{C/F/J}}$, if $\mathcal{D}^c \neq \emptyset$, then $R^*_{\text{C/F/J}}$ is not optimal; and hence $\mathcal{D}^c$ must be $\emptyset$. The argument is extended from that in [22] and the detailed analysis is as follows.

Suppose $\mathcal{D}^c \neq \emptyset$ at the optimum. Then,

$$\mathcal{D}^c \in \operatorname*{argmin}_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})$$

and

$$\bigcap_{\mathcal{T}\in \operatorname*{argmin}_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})} \mathcal{T} = \mathcal{D}^c.$$

Therefore,

$$\begin{aligned}
R^*_{\text{C/F/J}} =& I(X_0;\hat{Y}_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}}) + I(\mathcal{D}^c)\\
=& I(X_0;\hat{Y}_{\mathcal{D}},Y_{n+1}|X_{\mathcal{N}}) + I(X_0;\hat{Y}_{\mathcal{D}^c}|\hat{Y}_{\mathcal{D}},Y_{n+1},X_{\mathcal{N}})\\
&+\sum_{i\in\mathcal{D}^c} R_i - I(X_0,Y_{\mathcal{D}^c};\hat{Y}_{\mathcal{D}^c}|\hat{Y}_{\mathcal{D}},Y_{n+1},X_{\mathcal{N}})\\
=& I(X_0;\hat{Y}_{\mathcal{D}},Y_{n+1}|X_{\mathcal{N}}) + \sum_{i\in\mathcal{D}^c} R_i - I(Y_{\mathcal{D}^c};\hat{Y}_{\mathcal{D}^c}|X_0,\hat{Y}_{\mathcal{D}},Y_{n+1},X_{\mathcal{N}}),
\end{aligned} \tag{3.23}$$

and similarly,

$$\begin{aligned}
R^*_{\text{C/F/J}} =& I(X_0;\hat{Y}_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}}) + I(\mathcal{T})\\
=& I(X_0;\hat{Y}_{\mathcal{T}^c},Y_{n+1}|X_{\mathcal{N}}) + \sum_{i\in\mathcal{T}} R_i - I(Y_{\mathcal{T}};\hat{Y}_{\mathcal{T}}|X_0,\hat{Y}_{\mathcal{T}^c},Y_{n+1},X_{\mathcal{N}}),
\end{aligned} \tag{3.24}$$

for any $\mathcal{T} \in \operatorname*{argmin}_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})$, $\mathcal{T}\neq\mathcal{D}^c$.

We argue that higher rate can be achieved. Consider $\hat{Y}'_1,\hat{Y}'_2,\dots,\hat{Y}'_n$, where $\hat{Y}'_i = \hat{Y}_i$ for any $i\in\mathcal{D}$, and $\hat{Y}'_i=\hat{Y}_i$ with probability $p$ and $\hat{Y}'_i=\emptyset$ with probability $1-p$ for any $i\in\mathcal{D}^c$. When $p=1$, the achievable rate with $\hat{Y}'_1,\hat{Y}'_2,\dots,\hat{Y}'_n$ is $R^*_{\text{C/F/J}}$. As $p$ decreases from 1, it can be seen from (3.23) and (3.24) that both $I(X_0;\hat{Y}'_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}})+I(\mathcal{D}^c)$ and $I(X_0;\hat{Y}'_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}})+I(\mathcal{T})$ will increase, where

$\mathcal{T} \in \underset{\mathcal{S}\subseteq\mathcal{N}}{\operatorname{argmin}} I(\mathcal{S})$, $\mathcal{T} \neq \mathcal{D}^c$. Thus, no matter how $I(X_0; \hat{Y}'_{\mathcal{N}}, Y_{n+1}|X_{\mathcal{N}}) + I(\mathcal{S})$ will change as $p$ decreases for $\mathcal{S} \notin \underset{\mathcal{S}\subseteq\mathcal{N}}{\operatorname{argmin}} I(\mathcal{S})$, it is certain that there exists a $p^*$ such that the achievable rate by using $\hat{Y}'_1, \hat{Y}'_2, \ldots, \hat{Y}'_n$ is larger than $R^*_{\mathrm{C/F/J}}$. This is in contradiction with the optimality of $R^*_{\mathrm{C/F/J}}$, and thus at the optimum, $\mathcal{D}^c$ must be $\emptyset$ , i.e., $I(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{N}$. This completes the proof of Theorem 3.2.3. □

We now present the proofs of Lemmas 3.4.1-3.4.4.

*Proof of Lemma 3.4.1.* For any $\mathcal{S} \subseteq \mathcal{A}\cup\mathcal{B}$, let $\mathcal{S}_1 = \mathcal{S}\mathcal{A}$ and $\mathcal{S}_2 = \mathcal{S}(\mathcal{B}\setminus\mathcal{A})$. Then,

$$
\begin{aligned}
I_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}) =& \sum_{i\in\mathcal{S}} R_i - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}}, Y_{n+1}, X_{\mathcal{N}}) \\
=& \sum_{i\in\mathcal{S}_1} R_i - I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1}|\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}}, Y_{n+1}, X_{\mathcal{N}}) \\
&+ \sum_{i\in\mathcal{S}_2} R_i - I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2}|\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}}, \hat{Y}_{\mathcal{S}_1}, Y_{n+1}, X_{\mathcal{N}}) \\
\geq& \sum_{i\in\mathcal{S}_1} R_i - I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1}|\hat{Y}_{\mathcal{A}\setminus\mathcal{S}_1}, Y_{n+1}, X_{\mathcal{N}}) \\
&+ \sum_{i\in\mathcal{S}_2} R_i - I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2}|\hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B}\setminus\mathcal{S}_2}, Y_{n+1}, X_{\mathcal{N}}) \\
=& I_{\mathcal{A}}(\mathcal{S}_1) + I_{\mathcal{A},\mathcal{B}}(\mathcal{S}_2) \qquad (3.25) \\
\geq& I_{\mathcal{A}}(\mathcal{S}_1) + I_{\mathcal{B}}(\mathcal{S}_2). \qquad (3.26)
\end{aligned}
$$

If $I_{\mathcal{A}}(\mathcal{S}_1) \geq 0$, $\forall \mathcal{S}_1 \subseteq \mathcal{A}$, and $I_{\mathcal{B}}(\mathcal{S}_2) \geq 0$, $\forall \mathcal{S}_2 \subseteq \mathcal{B}$, then following (3.26), $I_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{A}\cup\mathcal{B}$. If $I_{\mathcal{A}}(\mathcal{S}_1) \geq 0$, $\forall \mathcal{S}_1 \subseteq \mathcal{A}$, and $I_{\mathcal{A},\mathcal{B}}(\mathcal{S}_2) \geq 0$, $\forall \mathcal{S}_2 \subseteq \mathcal{B}$, then following (3.25), $I_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{A}\cup\mathcal{B}$. □

*Proof of Lemma 3.4.2.* Let $\mathcal{L} := \{\mathcal{F} \subseteq \mathcal{N} : I_{\mathcal{F}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{F}\}$ and $\mathcal{L}_{\max} := \{\mathcal{D} \in \mathcal{L} : |\mathcal{D}| = \max_{\mathcal{F}\in\mathcal{L}} |\mathcal{F}|\}$. Suppose there are more than one element in $\mathcal{L}_{\max}$, say, $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_n$, where $n \geq 2$. Then based on 1) of Lemma 3.4.1, $\mathcal{D} := \bigcup_{i=1}^n \mathcal{D}_i$ also satisfies that $I_{\mathcal{D}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{D}$, which is in contradiction, and hence Lemma 3.4.2 is proved. □

*Proof of Lemma 3.4.3.* If $I_{\mathcal{A},\mathcal{B}}(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{B}$, then this lemma obviously holds. Otherwise, if there exists some $\mathcal{S}_1 \subseteq \mathcal{B}$, $\mathcal{S}_1 \neq \mathcal{B}$, such that $I_{\mathcal{A},\mathcal{B}}(\mathcal{S}_1) < 0$, then we have $I_{\mathcal{A},\mathcal{B}}(\mathcal{B}) - I_{\mathcal{A},\mathcal{B}}(\mathcal{S}_1) \geq 0$, i.e.,

$$
\begin{aligned}
&\sum_{i\in\mathcal{B}} R_i - I(Y_{\mathcal{B}};\hat{Y}_{\mathcal{B}}|\hat{Y}_{\mathcal{A}},Y_{n+1},X_{\mathcal{N}}) - \left(\sum_{i\in\mathcal{S}_1} R_i - I(Y_{\mathcal{S}_1};\hat{Y}_{\mathcal{S}_1}|\hat{Y}_{\mathcal{A}},\hat{Y}_{\mathcal{B}\setminus\mathcal{S}_1},Y_{n+1},X_{\mathcal{N}})\right)\\
=&\sum_{i\in\mathcal{B}\setminus\mathcal{S}_1} R_i - I(Y_{\mathcal{B}\setminus\mathcal{S}_1};\hat{Y}_{\mathcal{B}\setminus\mathcal{S}_1}|\hat{Y}_{\mathcal{A}},Y_{n+1},X_{\mathcal{N}})\\
=&I_{\mathcal{A},\mathcal{B}\setminus\mathcal{S}_1}(\mathcal{B}\setminus\mathcal{S}_1)\\
\geq&0.
\end{aligned}
$$

Now, we arrive at the same situation as in the original assumption with $\mathcal{B}$ replaced by $\mathcal{B}\setminus\mathcal{S}_1$. Continue applying this argument, and we must be able to reach a nonempty $\mathcal{C}\subseteq\mathcal{B}$, such that $I_{\mathcal{A},\mathcal{C}}(\mathcal{S})\geq 0$, $\forall\mathcal{S}\subseteq\mathcal{C}$. □

*Proof of Lemma 3.4.4.* For any disjoint $\mathcal{A}$ and $\mathcal{B}$,

$$
\begin{aligned}
&I(\mathcal{A}\cup\mathcal{B})\\
=&\sum_{i\in\mathcal{A}\cup\mathcal{B}} R_i - I(Y_{\mathcal{A}\cup\mathcal{B}};\hat{Y}_{\mathcal{A}\cup\mathcal{B}}|\hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c},Y_{n+1},X_{\mathcal{N}})\\
=&\sum_{i\in\mathcal{A}} R_i - I(Y_{\mathcal{A}\cup\mathcal{B}};\hat{Y}_{\mathcal{A}}|\hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c},Y_{n+1},X_{\mathcal{N}})\\
&+\sum_{i\in\mathcal{B}} R_i - I(Y_{\mathcal{A}\cup\mathcal{B}};\hat{Y}_{\mathcal{B}}|\hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c},\hat{Y}_{\mathcal{A}},Y_{n+1},X_{\mathcal{N}})\\
=&\sum_{i\in\mathcal{A}} R_i - I(Y_{\mathcal{A}},\hat{Y}_{\mathcal{B}};\hat{Y}_{\mathcal{A}}|\hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c},Y_{n+1},X_{\mathcal{N}})\\
&+\sum_{i\in\mathcal{B}} R_i - I(Y_{\mathcal{B}};\hat{Y}_{\mathcal{B}}|\hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c},\hat{Y}_{\mathcal{A}},Y_{n+1},X_{\mathcal{N}})\\
=&\sum_{i\in\mathcal{A}} R_i - I(\hat{Y}_{\mathcal{B}};\hat{Y}_{\mathcal{A}}|\hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c},Y_{n+1},X_{\mathcal{N}}) - I(Y_{\mathcal{A}};\hat{Y}_{\mathcal{A}}|\hat{Y}_{\mathcal{A}^c},Y_{n+1},X_{\mathcal{N}})\\
&+\sum_{i\in\mathcal{B}} R_i - I(Y_{\mathcal{B}};\hat{Y}_{\mathcal{B}}|\hat{Y}_{\mathcal{B}^c},Y_{n+1},X_{\mathcal{N}})\\
=&I(\mathcal{A})+I(\mathcal{B}) - I(\hat{Y}_{\mathcal{A}};\hat{Y}_{\mathcal{B}}|\hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c},Y_{n+1},X_{\mathcal{N}}),
\end{aligned}
$$

which proves the lemma. □

## 3.5 Discussion: An Optimality-Robustness Trade-off

.

Although the flexibility for the relay to choose its compression may not improve the achievable rate, compression-message joint decoding makes the C-F relay scheme more robust to the variation of channel. To illustrate this point, we consider a Gaussian single-relay channel as depicted in Figure 3.1. In this channel, $Y = X + Z$, $Y_1 = X + Z_1$, where $X$ is of power $P$, and $Z$ and $Z_1$ are independent Gaussian noises, and the relay is connected to the destination via an error-free digital link with capacity $R_1$. The optimal distribution $f(x)f(\hat{y}_1|y_1)$ for the C-F scheme is not known. By convention, we assume $X \sim \mathcal{N}(0, P)$ and $\hat{Y}_1 = Y_1 + \hat{Z}_1$ with $\hat{Z}_1 \sim \mathcal{N}(0, \hat{N}_1)$ and independent of other random variables.

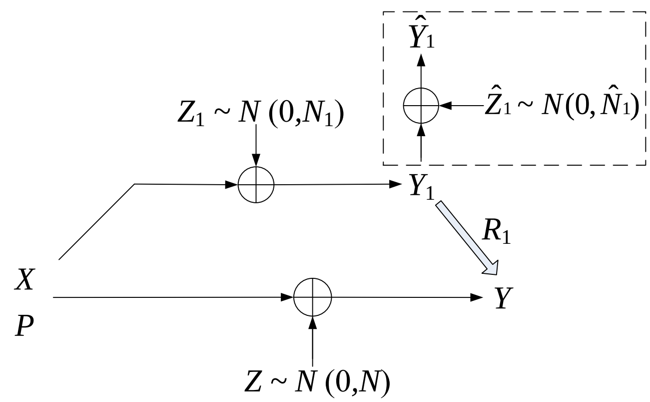



Figure 3.1: A Gaussian relay channel with a digital link.

In the original C-F scheme with compression-message successive decoding, the relay needs to know the value of $I(Y_1; \hat{Y}_1|Y)$ in order to decide upon the appropriate compressed version $\hat{Y}_1$ to choose such that the constraint $I(Y_1; \hat{Y}_1|Y) \leq R_1$ is satisfied. This requires the knowledge of the channel dynamics from $X$ to $Y$, which may be difficult to obtain for the relay, e.g., in wireless communications where the channel could be time varying. In this situation, if $\hat{Y}_1$ is inappropriately chosen such that $I(Y_1; \hat{Y}_1|Y) > R_1$, then the destination can not decode $\hat{Y}_1$, causing a problem to further decode $X$.

However, this problem will not happen with compression-message joint decoding, where the relay can choose any version $\hat{Y}_1$ since $\hat{Y}_1$ is not to be decoded. Despite that the relay may not know the channel dynamics from $X$ to $Y$, it can compress its observation at a constant rate $I(Y_1;\hat{Y}_1)$, where $\hat{Y}_1$ is a pre-specified compressed version and the choice of $\hat{Y}_1$ may be based on the statistics of the channel dynamics. Then, when $I(Y_1;\hat{Y}_1|Y) \leq R_1$, joint decoding can achieve the same rate as successive decoding as in the original C-F scheme. More importantly, when $I(Y_1;\hat{Y}_1|Y) > R_1$, unlike in the original C-F scheme where a decoding problem happens, with joint decoding, the destination can still achieve some rate specified in (3.3), or more generally, Theorem 3.2.2 when there are multiple relays. Clearly, this makes the C-F relay scheme with compression-message joint decoding more robust to the variation of channel, especially when the relay has no access to the channel state information of the $X \to Y$ link.

Specifically, in the setup of Figure 3.1, the information quantities $I(X;\hat{Y}_1,Y)$ and $I(Y_1;\hat{Y}_1|Y)$ reduce to

$$C\left(\frac{P(N+N_1+\hat{N}_1)}{(N_1+\hat{N}_1)N}\right)$$

and

$$C\left(\frac{PN+PN_1+NN_1}{(P+N)\hat{N}_1}\right)$$

respectively, where $C(x) := \frac{1}{2}\log(1+x)$. Thus, the achievable rate stated in (3.3)/Theorem 3.2.2 is given by

$$R < \begin{cases} C\left(\frac{P(N+N_1+\hat{N}_1)}{(N_1+\hat{N}_1)N}\right) & \text{when } \mathcal{C}\left(\frac{PN+PN_1+NN_1}{(P+N)\hat{N}_1}\right) \leq R_1 \\ C\left(\frac{P}{N} - \frac{(P+N)N_1}{(N_1+\hat{N}_1)N}\right) + R_1 & \text{otherwise.} \end{cases} \tag{3.27}$$

Figure 3.2 plots the achievable rate for $P = N = 200$, $N_1 = 10$ and $R_1 = 2$, where the *A*-*B*-*C* curve is the achievable rate with the C-F scheme with joint decoding, while the horizontal straight line is the achievable rate of the main channel $X \to Y$, which is independent of the choice of $\hat{N}_1$.

Point $B$ corresponds to the highest rate in (3.27), which is achieved by choosing the optimal $\hat{N}_1$, namely the $\hat{N}_1$ such that

$$I(Y_1;\hat{Y}_1|Y) = C\left(\frac{PN+PN_1+NN_1}{(P+N)\hat{N}_1}\right) = R_1.$$

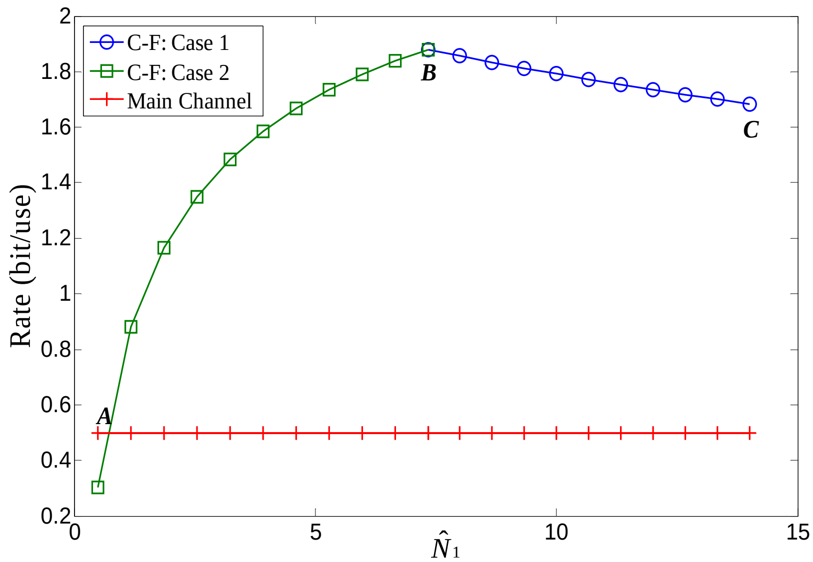



Figure 3.2: Achievable rate for the Gaussian relay channel.

Point $C$ corresponds to the $\hat{N}_1$ such that

$$I(Y_1;\hat{Y}_1) = C\left(\frac{P+N_1}{\hat{N}_1}\right) = R_1,$$

which is the scenario where $\hat{Y}_1$ can always be decoded at the destination no matter how is the channel dynamics from $X$ to $Y$, and the $\hat{Y}_1$ in this scenario is the coarsest compressed version that is of interest. The curve $B$-$C$ describes the achievable rate for the first case in (3.27), i.e., when

$$I(Y_1;\hat{Y}_1|Y) = C\left(\frac{PN+PN_1+NN_1}{(P+N)\hat{N}_1}\right) \leq R_1.$$

In this case, successive decoding and joint decoding achieve the same rate.

Point $A$ is the intersection of the achievable rate curve for the C-F scheme with joint decoding and the rate of the main channel. It corresponds to the scenario where the $\hat{N}_1$ is chosen such that the C-F only achieves the rate of the main channel, and the $\hat{Y}_1$ in this scenario is the finest compressed version of interest. The curve $A$-$B$ describes the achievable rate for the second case in (3.27), i.e., when

$$I(Y_1;\hat{Y}_1|Y) = C\left(\frac{PN+PN_1+NN_1}{(P+N)\hat{N}_1}\right) > R_1.$$

This is the case where the original C-F with successive decoding encounters a decoding problem, but joint decoding can still achieve some rate.

# Chapter 4

# C-F Relay Schemes with Backward Decoding

## 4.1 Introduction

Recently, a noisy network coding scheme was proposed in [32]. Unlike the conventional C-F scheme with cumulative encoding/block-by-block forward decoding, in noisy network coding, different blocks use independent codebooks to repetitively transmit the same information, and the decoding is based on all the blocks together. Compared to cumulative encoding, this repetitive encoding process appears to introduce collaboration among all the blocks, so that all the blocks can unitedly contribute to the decoding of the same message. This repetitive encoding/all blocks united decoding process was combined with joint compression-message decoding in [32], and although no improvement was shown in the single relay case, some interesting improvement on the achievable rate was obtained in the case of multiple relays. In this chapter, we will show that actually it is not necessary to use repetitive encoding to introduce such collaboration among the blocks. The same rate can be achieved with cumulative encoding as long as the decoding starts after all the blocks have been finished. In particular, we will develop the following two schemes:

- cumulative encoding/block-by-block backward decoding/compression-message successive decoding

- cumulative encoding/block-by-block backward decoding/compression-message joint decoding

and show that they achieve the same rate as noisy network coding. To highlight their difference from noisy network coding in the encoding/decoding process, we also use

- repetitive encoding/all blocks united decoding/compression-message joint decoding

to refer to the noisy network coding scheme interchangeably in the thesis.

Since block-by-block backward decoding and compression-message successive decoding are relatively easier to implement than all blocks united decoding and compression-message joint decoding respectively, the cumulative encoding/block-by-block backward decoding/compression-message successive decoding scheme becomes the simplest choice in achieving the highest C-F rate in the case of multiple relays.

For these new encoding/decoding schemes, we will also show that the optimal compressions must be able to support successive compression-message decoding, and any compressions not supporting successive decoding will necessarily lead to strictly lower achievable rates than the optimal. Therefore, for any of the C-F relay schemes studied in the thesis, we can restrict our attention to successive compression-message decoding in the search for the optimal compressions of the relays' observations. Of course, it should be noted that any compressions supporting successive decoding also support joint decoding. Besides the optimality of successive decoding, we will further demonstrate the necessity of joint decodability in the sense that any compressions not completely decodable even with joint decoding will not provide any contribution to the decoding of the original message, and the destination would rather simply treat the corresponding relays' inputs as purely noise in the decoding.

In the rest of the chapter, the main results are first presented in Section 4.2, followed by their proofs in Sections 4.3- 4.5.

## 4.2 Main Results

It was shown in [32] that the original cumulative encoding/block-by-block forward decoding/compression-message successive decoding scheme developed in [15] can be improved to achieve higher rates in the case of multiple relays, although no improvement was obtained in the case of a single relay. In their noisy network coding scheme [32], cumulative encoding was replaced by repetitive encoding, and block-by-block forward decoding was replaced by all blocks united decoding. They also used joint instead of successive compression-message decoding. For the single-source multiple-relay channel depicted in Figure 1.5, their Theorem 1 in [32] can be re-stated as the following theorem.

**Theorem 4.2.1.** *For the multiple-relay channel depicted in Figure 1.5, a rate* $R_{R/U/J}$ *is achievable if there exists some*

$$p(q)p(x_0|q)p(x_1|q)\cdots p(x_n|q)p(\hat{y}_1|y_1,x_1,q)\cdots p(\hat{y}_n|y_n,x_n,q),$$

*such that*

$$R_{R/U/J} < \min_{\mathcal{S}\subseteq\mathcal{N}} I(X_0,X_{\mathcal{S}};\hat{Y}_{\mathcal{S}^c},Y_{n+1}|X_{\mathcal{S}^c},Q) - I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_0,X_{\mathcal{N}},Y_{n+1},\hat{Y}_{\mathcal{S}^c},Q). \quad (4.1)$$

In this chapter, we will show that the improvement is not a result of replacing cumulative encoding by repetitive encoding, but actually, is a benefit obtained when the decoding is delayed, i.e., only starts after all the blocks have been finished. Besides all blocks united decoding, we will show that block-by-block backward decoding also achieves the same improvement since it also starts the decoding after all the blocks have been finished.

Similar to the framework of block-by-block forward decoding, we will also show that for these new schemes with decoding after all the blocks have been finished, the optimal rate can be achieved only when the compressions at the relays are chosen such that successive compression-message decoding can be carried out. Thus, in terms of complexity, cumulative encoding/block-by-block backward decoding/ compression-message successive decoding is the simplest choice in achieving the highest rate in the case of multiple relays. The corresponding achievable rate is presented in the following theorem.

**Theorem 4.2.2.** *For the multiple-relay channel depicted in Figure 1.5, a rate $R_{C/B/S}$ is achievable if there exists some*

$$p(q)p(x_0|q)p(x_1|q)\cdots p(x_n|q)p(\hat{y}_1|y_1,x_1,q)\cdots p(\hat{y}_n|y_n,x_n,q),$$

*such that for any subset $\mathcal{S}\subseteq\mathcal{N}$,*

$$I(X_{\mathcal{S}};\hat{Y}_{\mathcal{S}^c},Y_{n+1}|X_{\mathcal{S}^c},Q)-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_{\mathcal{N}},Y_{n+1},\hat{Y}_{\mathcal{S}^c},Q)\geq 0, \tag{4.2}$$

*and*

$$R_{C/B/S}<I(X_0;\hat{Y}_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}},Q). \tag{4.3}$$

Let $R^*_{\mathrm{R/U/J}}$ and $R^*_{\mathrm{C/B/S}}$ be the supremum of the achievable rates stated in Theorem 4.2.1 and 4.2.2 respectively. The optimality of successive decoding is demonstrated in the following theorem.

**Theorem 4.2.3.** *$R^*_{R/U/J}=R^*_{C/B/S}$, and $R^*_{R/U/J}$ can be obtained only when the distribution*

$$p(q)p(x_0|q)p(x_1|q)\cdots p(x_n|q)p(\hat{y}_1|y_1,x_1,q)\cdots p(\hat{y}_n|y_n,x_n,q)$$

*is chosen such that (4.2) holds.*

As mentioned in Section 3.1, although the optimal rate is achieved only when successive decoding can be supported, there are situations where it is of interest to consider other compressions not supporting successive decoding. Hence, more generally, we will use the cumulative encoding/block-by-block backward decoding/ compression-message joint decoding. The corresponding achievable rate is given in the following theorem.

**Theorem 4.2.4.** *For the multiple-relay channel depicted in Figure 1.5, with a given distribution*

$$p(q)p(x_0|q)p(x_1|q)\cdots p(x_n|q)p(\hat{y}_1|y_1,x_1,q)\cdots p(\hat{y}_n|y_n,x_n,q),$$

*a rate $R_{C/B/J}$ is achievable if*

$$R_{C/B/J} < \min_{\mathcal{S}\subseteq\mathcal{D}_J} I(X_0, X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_J\setminus\mathcal{S}}, Y_{n+1}|X_{\mathcal{D}_J\setminus\mathcal{S}}, Q) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_0, X_{\mathcal{D}_J}, Y_{n+1}, \hat{Y}_{\mathcal{D}_J\setminus\mathcal{S}}, Q), \tag{4.4}$$

*where $\mathcal{D}_J$ is the unique largest subset of $\mathcal{N}$ satisfying*

$$I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_J\setminus\mathcal{S}}, Y_{n+1}|X_0, X_{\mathcal{D}_J\setminus\mathcal{S}}, Q) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_0, X_{\mathcal{D}_J}, Y_{n+1}, \hat{Y}_{\mathcal{D}_J\setminus\mathcal{S}}, Q) > 0, \tag{4.5}$$

*for any nonempty $\mathcal{S} \subseteq \mathcal{D}_J$. In addition, $\hat{Y}_{\mathcal{D}_J}$ can be decoded jointly with $X$.*

There also exists a unique largest subset $\mathcal{D}'_{\mathrm{J}} \subseteq \mathcal{N}$ satisfying

$$I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}'_{\mathrm{J}}\setminus\mathcal{S}}, Y_{n+1}|X_0, X_{\mathcal{D}'_{\mathrm{J}}\setminus\mathcal{S}}, Q) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_0, X_{\mathcal{D}'_{\mathrm{J}}}, Y_{n+1}, \hat{Y}_{\mathcal{D}'_{\mathrm{J}}\setminus\mathcal{S}}, Q) \geq 0, \tag{4.6}$$

for any $\mathcal{S} \subseteq \mathcal{D}'_{\mathrm{J}}$. It will be clear from the proof of Theorem 4.2.4 that the compressions of the relays in $\mathcal{N} \setminus \mathcal{D}'_{\mathrm{J}}$ are not decodable even jointly with the message.

On the other hand, the achievable rate (4.1) can be more generally expressed as

$$R_{\mathrm{R/U/J}} < \min_{\mathcal{S}\subseteq\mathcal{M}} I(X_0, X_{\mathcal{S}}; \hat{Y}_{\mathcal{M}\setminus\mathcal{S}}, Y_{n+1}|X_{\mathcal{M}\setminus\mathcal{S}}, Q) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_0, X_{\mathcal{M}}, Y_{n+1}, \hat{Y}_{\mathcal{M}\setminus\mathcal{S}}, Q) \tag{4.7}$$

if we only consider a subset of relays $\mathcal{M} \subseteq \mathcal{N}$ for the decoding, while treating the other inputs as purely noise. Interestingly, the following theorem implies that $\mathcal{M} = \mathcal{N}$ may not be the optimal choice to maximize the R.H.S. (right-hand-side) of (4.7), i.e., sometimes, it is better to consider only a subset of relays.

**Theorem 4.2.5.** *For any $p(q)p(x_0|q)\prod_{i=1}^{n} p(x_i|q)p(\hat{y}_i|x_i, y_i, q)$, among all the choices of $\mathcal{M} \subseteq \mathcal{N}$, the R.H.S. of (4.7) is maximized when $\mathcal{M} = \mathcal{D}_J$ or $\mathcal{M} = \mathcal{D}'_J$, but is strictly less than the maximum when $\mathcal{M} \nsubseteq \mathcal{D}'_J$. Here, $\mathcal{D}_J$ and $\mathcal{D}'_J$ are defined as in (4.5) and (4.6).*

Therefore, not only the compressions of the relays in $\mathcal{N} \setminus \mathcal{D}'_{\mathrm{J}}$ are not decodable, but also including them in the formula (4.7), i.e., choosing $\mathcal{M} \nsubseteq \mathcal{D}'_{\mathrm{J}}$, will even strictly lower the achievable rate.

By comparing (4.4) and (4.7) with $\mathcal{M} = \mathcal{D}_{\mathrm{J}}$, Theorem 4.2.5 also implies that for any compressions chosen at the relays, the cumulative encoding/block-by-block backward decoding/compression-message joint decoding scheme achieves the same rate as the repetitive encoding/all blocks united decoding/compression-message joint decoding scheme.

## 4.3 Cumulative encoding/block-by-block backward decoding/compression-message successive decoding and Optimality of Successive Decoding

In cumulative encoding/block-by-block backward decoding, the encoding process is similar to that in the proof of Theorem 6 in [15] (except that the binning at the relay is not needed here), but the decoding process operates backwardly. This scheme, combined with compression-message successive decoding, proves Theorem 4.2.2 as follows.

*Proof of Theorem 4.2.2.* Consider $B+M$ blocks, where the source will transmit information in the first $B$ blocks and keep silent in the last $M$ blocks, the relays will compress-and-forward in all the $B+M$ blocks, and the destination will not start decoding until all the $B+M$ blocks have been finished. As we will see in the following proof, the added $M$ blocks are used to ensure the relays' compressions in the $B$-th block can be decoded with the help of the subsequent $M$ blocks. Then, backwardly, the relays' compressions in blocks $B-1$ to 1 can be decoded. Finally, using the recovered relays' compressions in all the first $B$ blocks, the original messages can be decoded. Of course, the added $M$ blocks could introduce decoding delay and thus rate loss, but note that we can always choose $M \ll B$ such that the rate loss $\frac{M}{B+M}R_{\text{C/B/S}}$ can be made arbitrarily small.

*Codebook Generation:* Fix $p(x_0)\prod_{i=1}^{n}p(x_i)p(\hat{y}_i|x_i,y_i)$. We randomly and independently generate a codebook for each block.

For each block $b\in[1:B+M]$, randomly and independently generate $2^{TR_{\text{C/B/S}}}$ sequences $\mathbf{x}_{0,b}(m_b)$, $m_b\in[1:2^{TR_{\text{C/B/S}}}]$; for each block $b\in[1:B+M]$ and each relay node $i\in\mathcal{N}$, randomly and independently generate $2^{T\hat{R}_i}$ sequences $\mathbf{x}_{i,b}(l_{i,b-1})$, $l_{i,b-1}\in[1:2^{T\hat{R}_i}]$, where $\hat{R}_i=I(Y_i;\hat{Y}_i|X_i)+\epsilon$; for each relay node $i\in\mathcal{N}$ and each $\mathbf{x}_{i,b}(l_{i,b-1})$, $l_{i,b-1}\in[1:2^{T\hat{R}_i}]$, randomly and conditionally independently generate $2^{T\hat{R}_i}$ sequences $\hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1})$, $l_{i,b}\in[1:2^{T\hat{R}_i}]$. This defines the codebook for any block $b\in[1:B+M]$,

$$\mathcal{C}_b=\{\mathbf{x}_{0,b}(m_b),\mathbf{x}_{i,b}(l_{i,b-1}),\hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1}): \quad m_b\in[1:2^{TR_{\text{C/B/S}}}],l_{i,b},l_{i,b-1}\in[1:2^{T\hat{R}_i}],i\in\mathcal{N}\}.$$

*Encoding:* Let $\mathbf{m}=(m_1,m_2,\ldots,m_B)$ be the message vector to be sent and let $m_b=1$ be the dummy message for any $b\in[B+1:B+M]$. For any block

$b \in [1:B{+}M]$, each relay node $i \in \mathcal{N}$, upon receiving $\mathbf{y}_{i,b}$ at the end of block $b$, finds an index $l_{i,b}$ such that $(\mathbf{x}_{i,b}(l_{i,b-1}), \mathbf{y}_{i,b}, \hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1})) \in A_\epsilon(X_i, Y_i, \hat{Y}_i)$, where $l_{i,0} = 1$ by convention. The codewords $\mathbf{x}_{0,b}(m_b)$ and $\mathbf{x}_{i,b}(l_{i,b-1}), i \in \mathcal{N}$ are transmitted in block $b$, $b \in [1 : B + M]$.

*Decoding:* i) The destination first finds a unique combination of the relays' compression indices $\mathbf{l}^B = (\mathbf{l}_1, \dots, \mathbf{l}_B)$ and some $\mathbf{l}_{B+1}^{B+M} = (\mathbf{l}_{B+1}, \dots, \mathbf{l}_{B+M})$, where $\mathbf{l}_b = (l_{1,b}, \dots, l_{n,b})$, $\forall b \in [1 : B + M]$, such that for any $b = 1, \dots, B + M$,

$$\Big((\mathbf{X}_{1,b}(l_{1,b-1}), \hat{\mathbf{Y}}_{1,b}(l_{1,b}|l_{1,b-1})), \dots, (\mathbf{X}_{n,b}(l_{n,b-1}), \hat{\mathbf{Y}}_{n,b}(l_{n,b}|l_{n,b-1})), \mathbf{Y}_{n+1,b}\Big) \in A_\epsilon(X_{\mathcal{N}}, \hat{Y}_{\mathcal{N}}, Y_{n+1}). \quad (4.8)$$

Specifically, this can be done backwards as follows:

a) The destination finds the unique $\mathbf{l}_B$ such that there exists some $\mathbf{l}_{B+1}^{B+M} = (\mathbf{l}_{B+1}, \dots, \mathbf{l}_{B+M})$ satisfying (4.8) for any $b = B + 1, \dots, B + M$.

Assume the true $\mathbf{l}_B^{B+M} = \mathbf{1}^{M+1}$, where $\mathbf{1} := (1, \dots, 1)$ is an $n$-dimensional all-ones vector. Then, error occurs if $\mathbf{l}_B = \mathbf{1}$ does not satisfy (4.8) with any $\mathbf{l}_{B+1}^{B+M}$ for any $b = B + 1, \dots, B + M$, or a false $\mathbf{l}_B \neq \mathbf{1}$ satisfies (4.8) with some $\mathbf{l}_{B+1}^{B+M}$ for any $b = B{+}1, \dots, B{+}M$. Since $\mathbf{l}_B^{B+M} = \mathbf{1}^{M+1}$ satisfies (4.8) for any $b = B{+}1, \dots, B{+}M$ with high probability according to the properties of typical sequences, we only need to bound $\Pr(\bigcup_{\mathbf{l}_B \neq \mathbf{1}} \mathcal{E}_{\mathbf{l}_B})$, where $\mathcal{E}_{\mathbf{l}_B}$ is defined as the event that $\mathbf{l}_B$ satisfies (4.8) with some $\mathbf{l}_{B+1}^{B+M}$ for any $b = B + 1, \dots, B + M$. For any $(\mathbf{l}_{b-1}, \mathbf{l}_b)$, define $\mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b)$ as the event that $(\mathbf{l}_{b-1}, \mathbf{l}_b)$ satisfies (4.8). Then, we have

$$\begin{aligned}
\Pr(\bigcup_{\mathbf{l}_B \neq \mathbf{1}} \mathcal{E}_{\mathbf{l}_B}) &= \Pr(\bigcup_{\mathbf{l}_{B+1}^{B+M}} \bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcap_{b=B+1}^{B+M} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b)) \\
&= \Pr(\bigcup_{j=1}^{M-1} \bigcup_{\mathbf{l}_{B+1}^{B+M} : \mathbf{l}_{B+j} = \mathbf{1}} \bigcup_{\substack{\mathbf{l}_{B+1}^{B+M} : \mathbf{l}_{B+j} \neq \mathbf{1}, \\ \forall j \in [1:M-1]}} \bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcap_{b=B+1}^{B+M} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b)) \\
&\leq \sum_{j=1}^{M-1} \Pr(\bigcup_{\mathbf{l}_{B+1}^{B+M} : \mathbf{l}_{B+j} = \mathbf{1}} \bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcap_{b=B+1}^{B+M} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b)) \\
&\quad + \Pr(\bigcup_{\substack{\mathbf{l}_{B+1}^{B+M} : \mathbf{l}_{B+j} \neq \mathbf{1}, \\ \forall j \in [1:M-1]}} \bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcap_{b=B+1}^{B+M} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b)). \qquad (4.9)
\end{aligned}$$

Let us first consider the second term in (4.9). For any $\mathbf{l}_B^{B+M}$, let $\mathcal{S}_b(\mathbf{l}_B^{B+M}) = \{i \in \mathcal{N} : l_{i,b-1} \neq 1\}$. Note $\mathcal{S}_b(\mathbf{l}_B^{B+M})$ only depends on $\mathbf{l}_{b-1}$, so we also write it as $\mathcal{S}_b(\mathbf{l}_{b-1})$. Define $\mathbf{X}_b(\mathcal{S}_b(\mathbf{l}_{b-1}))$ as $\{\mathbf{X}_{i,b}(l_{i,b-1}), i \in \mathcal{S}_b(\mathbf{l}_{b-1})\}$, and similarly define

$\mathbf{Y}_b(\mathcal{S}_b(\mathbf{l}_{b-1}))$ and $\hat{\mathbf{Y}}_b(\mathcal{S}_b(\mathbf{l}_{b-1}))$. Then, $(\mathbf{X}_b(\mathcal{S}_b(\mathbf{l}_{b-1})), \hat{\mathbf{Y}}_b(\mathcal{S}_b(\mathbf{l}_{b-1})))$ is independent of $(\mathbf{X}_b(\mathcal{S}_b^c(\mathbf{l}_{b-1})), \hat{\mathbf{Y}}_b(\mathcal{S}_b^c(\mathbf{l}_{b-1})), \mathbf{Y}_b)$, and $\Pr(\mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b))$ can be upper bounded by

$$
\begin{aligned}
&2^{T(H(X_{\mathcal{N}}, \hat{Y}_{\mathcal{N}}, Y_{n+1})+\epsilon)} 2^{-T(H(X_{\mathcal{S}_b^c(\mathbf{l}_{b-1})}, \hat{Y}_{\mathcal{S}_b^c(\mathbf{l}_{b-1})}, Y_{n+1})-\epsilon)} \\
\times\ & 2^{-T(H(X_{\mathcal{S}_b(\mathbf{l}_{b-1})})-\epsilon)} 2^{-T(\sum_{i\in\mathcal{S}_b(\mathbf{l}_{b-1})}(H(\hat{Y}_i|X_i)-\epsilon))} \\
=:& 2^{-T(\mathcal{I}(\mathcal{S}_b(\mathbf{l}_{b-1}))-\epsilon')}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{I}(\mathcal{S}_b(\mathbf{l}_{b-1})) =& I(X_{\mathcal{S}_b(\mathbf{l}_{b-1})}; \hat{Y}_{\mathcal{S}_b^c(\mathbf{l}_{b-1})}, Y_{n+1} | X_{\mathcal{S}_b^c(\mathbf{l}_{b-1})}) \\
&- H(\hat{Y}_{\mathcal{S}_b(\mathbf{l}_{b-1})} | X_{\mathcal{N}}, \hat{Y}_{\mathcal{S}_b^c(\mathbf{l}_{b-1})}, Y_{n+1}) + \sum_{i\in\mathcal{S}_b(\mathbf{l}_{b-1})} H(\hat{Y}_i|X_i)
\end{aligned}
$$

and $\epsilon' \to 0$ as $\epsilon \to 0$. Then, we have

$$
\begin{aligned}
&\Pr\Big(\bigcup_{\substack{\mathbf{l}_{B+1}^{B+M}:\mathbf{l}_{B+j}\neq\mathbf{1},\\ \forall j\in[1:M-1]}} \bigcup_{\mathbf{l}_B\neq\mathbf{1}} \bigcap_{b=B+1}^{B+M} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b)\Big) \\
\leq & \sum_{\substack{\mathbf{l}_{B+1}^{B+M}:\mathbf{l}_{B+j}\neq\mathbf{1},\\ \forall j\in[1:M-1]}} \sum_{\mathbf{l}_B\neq\mathbf{1}} \prod_{b=B+1}^{B+M} \Pr(\mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b)) \\
\leq & \sum_{\mathbf{l}_{B+M}} \sum_{\substack{\mathbf{l}_{B+1}^{B+M-1}:\mathbf{l}_{B+j}\neq\mathbf{1},\\ \forall j\in[1:M-1]}} \sum_{\mathbf{l}_B\neq\mathbf{1}} \prod_{b=B+1}^{B+M} 2^{-T(\mathcal{I}(\mathcal{S}_b(\mathbf{l}_{b-1}))-\epsilon')} \\
= & \sum_{\mathbf{l}_{B+M}} \sum_{\substack{\mathcal{S}_{B+1},\dots,\mathcal{S}_{B+M}:\\ \mathcal{S}_{B+j}\neq\emptyset, \forall j\in[1:M]}} \sum_{\substack{\mathbf{l}_B^{B+M-1}:\mathcal{S}_b(\mathbf{l}_B^{B+M-1})=\mathcal{S}_b,\\ \forall b\in[B+1:B+M]}} \prod_{b=B+1}^{B+M} 2^{-T(\mathcal{I}(\mathcal{S}_b(\mathbf{l}_{b-1}))-\epsilon')} \\
\leq & \sum_{\mathbf{l}_{B+M}} \sum_{\substack{\mathcal{S}_{B+1},\dots,\mathcal{S}_{B+M}:\\ \mathcal{S}_{B+j}\neq\emptyset, \forall j\in[1:M]}} \prod_{b=B+1}^{B+M} 2^{T(\sum_{i\in\mathcal{S}_b}(I(Y_i;\hat{Y}_i|X_i)+\epsilon))} \prod_{b=B+1}^{B+M} 2^{-T(\mathcal{I}(\mathcal{S}_b)-\epsilon')} \\
\leq & \sum_{\mathbf{l}_{B+M}} \sum_{\substack{\mathcal{S}_{B+1},\dots,\mathcal{S}_{B+M}:\\ \mathcal{S}_{B+j}\neq\emptyset, \forall j\in[1:M]}} \prod_{b=B+1}^{B+M} 2^{-T(I(X_{\mathcal{S}_b};\hat{Y}_{\mathcal{S}_b^c}, Y_{n+1}|X_{\mathcal{S}_b^c}) - I(Y_{\mathcal{S}_b};\hat{Y}_{\mathcal{S}_b}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{S}_b^c})-\epsilon'')} \\
= & \sum_{\mathbf{l}_{B+M}} \sum_{\substack{\mathcal{S}_{B+1},\dots,\mathcal{S}_{B+M}:\\ \mathcal{S}_{B+j}\neq\emptyset, \forall j\in[1:M]}} 2^{-T\sum_{b=B+1}^{B+M}(I(X_{\mathcal{S}_b};\hat{Y}_{\mathcal{S}_b^c}, Y_{n+1}|X_{\mathcal{S}_b^c}) - I(Y_{\mathcal{S}_b};\hat{Y}_{\mathcal{S}_b}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{S}_b^c})-\epsilon'')} \\
\leq & \sum_{\mathbf{l}_{B+M}} (2^n)^M 2^{-TM(\min_{\mathcal{S}\subseteq\mathcal{N}:\mathcal{S}\neq\emptyset}\{I(X_{\mathcal{S}};\hat{Y}_{\mathcal{S}^c}, Y_{n+1}|X_{\mathcal{S}^c}) - I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{S}^c})-\epsilon''\})} \\
\leq & 2^{T(\sum_{i\in\mathcal{N}}(I(\hat{Y}_i;Y_i|X_i)+\epsilon))} 2^{nM} 2^{-TM(\min_{\mathcal{S}\subseteq\mathcal{N}:\mathcal{S}\neq\emptyset}\{I(X_{\mathcal{S}};\hat{Y}_{\mathcal{S}^c}, Y_{n+1}|X_{\mathcal{S}^c}) - I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{S}^c})-\epsilon''\})}
\end{aligned}
$$

where $\epsilon'' \to 0$ as $\epsilon \to 0$. Thus, as both $T$ and $M$ go to infinity, the second term in (4.9) goes to 0, if

$$I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{n+1}|X_{\mathcal{S}^c}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{S}^c}) > 0, \forall \mathcal{S} \subseteq \mathcal{N}, \mathcal{S} \neq \emptyset. \tag{4.10}$$

Now consider the first term in (4.9). For any $j \in [1 : M-1]$, we have

$$\Pr(\bigcup_{\mathbf{l}_{B+1}^{B+M}:\mathbf{l}_{B+j}=\mathbf{1}} \bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcap_{b=B+1}^{B+M} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b)) \leq \Pr(\bigcup_{\mathbf{l}_{B+1}^{B+j}:\mathbf{l}_{B+j}=\mathbf{1}} \bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcap_{b=B+1}^{B+j} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b)).$$

Note $\Pr(\bigcup_{\mathbf{l}_{B+1}^{B+j}:\mathbf{l}_{B+j}=\mathbf{1}} \bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcap_{b=B+1}^{B+j} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b))$ is the probability that there exists a false $\mathbf{l}_B \neq \mathbf{1}$ satisfying (4.8) with some $\mathbf{l}_{B+1}^{B+j}$ for any block $b \in [B+1 : B+j]$, where $\mathbf{l}_{B+j} = \mathbf{1}$ is true. Below, we show that this probability goes to 0. The underlying idea is backward decoding, which will also be used in step b).

For any $k \in [1 : j]$, $j \in [1 : M-1]$, denote

$$p_{B+k} := \Pr(\bigcup_{\mathbf{l}_{B+1}^{B+k}:\mathbf{l}_{B+k}=\mathbf{1}} \bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcap_{b=B+1}^{B+k} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b)).$$

Then, we have

$$\begin{aligned}
p_{B+k} &= \Pr(\bigcup_{\mathbf{l}_{B+1}^{B+k}:\mathbf{l}_{B+k}=\mathbf{1}} \bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcap_{b=B+1}^{B+k} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b)) \\
&\leq \Pr(\bigcup_{\mathbf{l}_{B+1}^{B+k}:\mathbf{l}_{B+k}=\mathbf{1},\mathbf{l}_{B+k-1}=\mathbf{1}} \bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcap_{b=B+1}^{B+k} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b)) \\
&\quad + \Pr(\bigcup_{\mathbf{l}_{B+1}^{B+k}:\mathbf{l}_{B+k}=\mathbf{1},\mathbf{l}_{B+k-1}\neq\mathbf{1}} \bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcap_{b=B+1}^{B+k} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b)) \\
&\leq \Pr(\bigcup_{\mathbf{l}_{B+1}^{B+k-1}:\mathbf{l}_{B+k-1}=\mathbf{1}} \bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcap_{b=B+1}^{B+k-1} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b)) \\
&\quad + \Pr(\bigcup_{\mathbf{l}_{B+k}=\mathbf{1},\mathbf{l}_{B+k-1}\neq\mathbf{1}} \mathcal{A}_{B+k}(\mathbf{l}_{B+k-1}, \mathbf{l}_{B+k})) \\
&=: p_{B+k-1} + p'_{B+k},
\end{aligned}$$

where

$$p'_{B+k} := \Pr(\bigcup_{\mathbf{l}_{B+k}=\mathbf{1},\mathbf{l}_{B+k-1}\neq\mathbf{1}} \mathcal{A}_{B+k}(\mathbf{l}_{B+k-1}, \mathbf{l}_{B+k}))$$

and especially

$$p_{B+1} = p'_{B+1} = \Pr(\bigcup_{\mathbf{l}_{B+1}=\mathbf{1},\mathbf{l}_B\neq\mathbf{1}} \mathcal{A}_{B+1}(\mathbf{l}_B, \mathbf{l}_{B+1})).$$

Recursively, for any $j \in [1:M-1]$,

$$p_{B+j} \le p_{B+j-1} + p'_{B+j} \le p_{B+j-2} + p'_{B+j-1} + p'_{B+j} \le \cdots \le \sum_{k=1}^{k=j} p'_{B+k}.$$

For any $k \in [1:j]$, $j \in [1:M-1]$, with $\mathcal{S}_{B+k}(\mathbf{l}_{B+k-1}) := \{i \in \mathcal{N} : l_{i,B+k-1} \neq 1\}$, we have

$$\begin{aligned}
p'_{B+k} &= \Pr\Big(\bigcup_{\mathbf{l}_{B+k}=\mathbf{1},\mathbf{l}_{B+k-1}\neq \mathbf{1}} \mathcal{A}_{B+k}(\mathbf{l}_{B+k-1}, \mathbf{l}_{B+k})\Big) \\
&= \Pr\Big(\bigcup_{\mathbf{l}_{B+k}=\mathbf{1},\mathbf{l}_{B+k-1}:\mathcal{S}_{B+k}(\mathbf{l}_{B+k-1})\neq\emptyset} \mathcal{A}_{B+k}(\mathbf{l}_{B+k-1}, \mathbf{l}_{B+k})\Big) \\
&\le \sum_{\mathcal{S}_{B+k}\neq\emptyset} \sum_{\substack{\mathbf{l}_{B+k}=\mathbf{1},\\ \mathbf{l}_{B+k-1}:\mathcal{S}_{B+k}(\mathbf{l}_{B+k-1})=\mathcal{S}_{B+k}}} \Pr(\mathcal{A}_{B+k}(\mathbf{l}_{B+k-1}, \mathbf{l}_{B+k})) \\
&\le \sum_{\mathcal{S}_{B+k}\neq\emptyset} 2^{T(\sum_{i\in\mathcal{S}_{B+k}}(I(Y_i;\hat{Y}_i|X_i)+\epsilon))} 2^{-T(\mathcal{I}(\mathcal{S}_{B+k})-\epsilon')} \\
&\le 2^n 2^{-T(\min_{\mathcal{S}\subseteq\mathcal{N}:\mathcal{S}\neq\emptyset}\{I(X_{\mathcal{S}};\hat{Y}_{\mathcal{S}^c},Y_{n+1}|X_{\mathcal{S}^c}) - I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_{\mathcal{N}},Y_{n+1},\hat{Y}_{\mathcal{S}^c})-\epsilon''\})},
\end{aligned}$$

and thus $p_{B+j} \to 0$ as $T \to \infty$ if (4.10) holds. Therefore, if (4.10) holds, the first term in (4.9) also goes to 0 as $T \to \infty$, and $\mathbf{l}_B$ can be decoded.

b) Given that $\mathbf{l}_B$ has been recovered, the destination performs the backward decoding as follows. That is, backwards and sequentially from block $b = B$ to block $b = 2$, the destination finds the unique $\mathbf{l}_{b-1}$, such that $(\mathbf{l}_{b-1}, \mathbf{l}_b)$ satisfies (4.8), where $\mathbf{l}_b$ has already been recovered due to the backward property of decoding. At each block $b = B, B-1, \ldots, 2$, error occurs if the true $\mathbf{l}_{b-1}$ does not satisfy (4.8), or a false $\mathbf{l}_{b-1}$ satisfies (4.8). According to the properties of typical sequences, the true $\mathbf{l}_{b-1}$ satisfies (4.8) with high probability.

For a false $\mathbf{l}_{b-1}$ with false $\{l_{i,b-1}, i \in \mathcal{S}\}$ but true $\{l_{i,b-1}, i \in \mathcal{S}^c\}$, $(\mathbf{X}_b(\mathcal{S}), \hat{\mathbf{Y}}_b(\mathcal{S}))$ is independent of $(\mathbf{X}_b(\mathcal{S}^c), \hat{\mathbf{Y}}_b(\mathcal{S}^c), \mathbf{Y}_b)$, and the probability that $(\mathbf{l}_{b-1}, \mathbf{l}_b)$ satisfies (4.8) can be upper bounded by

$$2^{T(H(X_{\mathcal{N}},\hat{Y}_{\mathcal{N}},Y_{n+1})+\epsilon)} 2^{-T(H(X_{\mathcal{S}^c},\hat{Y}_{\mathcal{S}^c},Y_{n+1})-\epsilon)} 2^{-T(H(X_{\mathcal{S}})-\epsilon)} 2^{-T(\sum_{i\in\mathcal{S}}(H(\hat{Y}_i|X_i)-\epsilon))}.$$

Since the number of such false $\mathbf{l}_{b-1}$ is upper bounded by $\prod_{i\in\mathcal{S}} 2^{T(I(Y_i;\hat{Y}_i|X_i)+\epsilon)}$, with the union bound, it is easy to check that the probability of finding such a false $\mathbf{l}_{b-1}$ goes to zero as $T \to \infty$, if (4.10) holds. This combined with a) proves that $\mathbf{l}^B$ can be decoded, if (4.10) holds.

ii) Then, based on the recovered $\mathbf{l}^B$, the destination finds the unique $\mathbf{m} = (m_1, m_2, \ldots, m_B)$ such that for any $b = 1, \ldots, B$,

$$\Big(\mathbf{X}_{0,b}(m_b), (\mathbf{X}_{1,b}(l_{1,b-1}), \hat{\mathbf{Y}}_{1,b}(l_{1,b}|l_{1,b-1})), \ldots, (\mathbf{X}_{n,b}(l_{n,b-1}), \hat{\mathbf{Y}}_{n,b}(l_{n,b}|l_{n,b-1})), \mathbf{Y}_{n+1,b}\Big) \in A_\epsilon(X_0, X_\mathcal{N}, \hat{Y}_\mathcal{N}, Y_{n+1}). \tag{4.11}$$

Note that after $\mathbf{l}^B$ has been recovered,

$$(\mathbf{X}_{1,b}(l_{1,b-1}), \hat{\mathbf{Y}}_{1,b}(l_{1,b}|l_{1,b-1})), \ldots, (\mathbf{X}_{n,b}(l_{n,b-1}), \hat{\mathbf{Y}}_{n,b}(l_{n,b}|l_{n,b-1}))$$

and $\mathbf{Y}_{n+1,b}$ in (4.11) are known to the destination. Thus, from the property of typical sequences, the probability of decoding error will tend to zero if $R_{\mathrm{C/B/S}}$ is less than $I(X_0; X_\mathcal{N}, \hat{Y}_\mathcal{N}, Y)$, which is equal to $I(X_0; \hat{Y}_\mathcal{N}, Y|X_\mathcal{N})$ noting the independence between $X_0$ and $X_\mathcal{N}$. □

We are now in a position to prove Theorem 4.2.3. To facilitate the proof, we introduce some notations and lemmas. For any $\mathcal{A}, \mathcal{B} \subseteq \mathcal{N}$, let

$$\begin{aligned} J_{\mathcal{A},\mathcal{B}}(\mathcal{S}) :=& I(X_\mathcal{S}; \hat{Y}_{\mathcal{B}\backslash\mathcal{S}}, \hat{Y}_\mathcal{A}, Y_{n+1}|X_\mathcal{A}, X_{\mathcal{B}\backslash\mathcal{S}}) \\ &- I(Y_\mathcal{S}; \hat{Y}_\mathcal{S}|X_\mathcal{A}, \hat{Y}_\mathcal{A}, Y_{n+1}, X_\mathcal{B}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}}), \forall\mathcal{S} \subseteq \mathcal{B}, \end{aligned} \tag{4.12}$$

$$J_\mathcal{B}(\mathcal{S}) := J_{\emptyset,\mathcal{B}}(\mathcal{S}) = I(X_\mathcal{S}; \hat{Y}_{\mathcal{B}\backslash\mathcal{S}}, Y_{n+1}|X_{\mathcal{B}\backslash\mathcal{S}}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S}|X_\mathcal{B}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}}, Y_{n+1}), \forall\mathcal{S} \subseteq \mathcal{B}, \tag{4.13}$$

$$J(\mathcal{S}) := J_\mathcal{N}(\mathcal{S}) = I(X_\mathcal{S}; \hat{Y}_{\mathcal{S}^c}, Y_{n+1}|X_{\mathcal{S}^c}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S}|X_\mathcal{N}, Y_{n+1}, \hat{Y}_{\mathcal{S}^c}), \forall\mathcal{S} \subseteq \mathcal{N}. \tag{4.14}$$

Then, we have the following lemmas, whose proofs will be presented in Section 4.5.

**Lemma 4.3.1.** *1) If $J_\mathcal{A}(\mathcal{S}_1) \geq 0$, $\forall\mathcal{S}_1 \subseteq \mathcal{A}$, and $J_\mathcal{B}(\mathcal{S}_2) \geq 0$, $\forall\mathcal{S}_2 \subseteq \mathcal{B}$, then $J_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}) \geq 0$, $\forall\mathcal{S} \subseteq \mathcal{A}\cup\mathcal{B}$.*

*2) If $J_\mathcal{A}(\mathcal{S}_1) \geq 0$, $\forall\mathcal{S}_1 \subseteq \mathcal{A}$, and $J_{\mathcal{A},\mathcal{B}}(\mathcal{S}_2) \geq 0$, $\forall\mathcal{S}_2 \subseteq \mathcal{B}$, then $J_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}) \geq 0$, $\forall\mathcal{S} \subseteq \mathcal{A}\cup\mathcal{B}$.*

**Lemma 4.3.2.** *Under any $p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|x_i, y_i)$, there exists a unique set $\mathcal{D}$, which is the largest subset of $\mathcal{N}$ satisfying*

$$J_\mathcal{D}(\mathcal{S}) \geq 0, \forall\mathcal{S} \subseteq \mathcal{D}.$$

**Lemma 4.3.3.** *If $J_{\mathcal{A},\mathcal{B}}(\mathcal{B}) \geq 0$ for some nonempty $\mathcal{B}$, then there exists some nonempty $\mathcal{C} \subseteq \mathcal{B}$ such that $J_{\mathcal{A},\mathcal{C}}(\mathcal{S}) \geq 0, \forall\mathcal{S} \subseteq \mathcal{C}$.*

**Lemma 4.3.4.** *For any* $\mathcal{A}$ *and* $\mathcal{B}$ *with* $\mathcal{A}\cap\mathcal{B}=\emptyset$*,* $J(\mathcal{A})+J(\mathcal{B})=J(\mathcal{A}\cup\mathcal{B})+J(\mathcal{A}\circ\mathcal{B})$*, where*

$$\begin{aligned}J(\mathcal{A}\circ\mathcal{B}) =& I(X_{\mathcal{A}},\hat{Y}_{\mathcal{A}};X_{\mathcal{B}},\hat{Y}_{\mathcal{B}}|X_{(\mathcal{A}\cup\mathcal{B})^c},\hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c},Y_{n+1})\\ =& I(X_{\mathcal{A}};X_{\mathcal{B}}|X_{(\mathcal{A}\cup\mathcal{B})^c},\hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c},Y_{n+1})+I(X_{\mathcal{A}};\hat{Y}_{\mathcal{B}}|X_{\mathcal{A}^c},\hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c},Y_{n+1})\\ &+I(X_{\mathcal{B}};\hat{Y}_{\mathcal{A}}|X_{\mathcal{B}^c},\hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c},Y_{n+1})+I(\hat{Y}_{\mathcal{A}};\hat{Y}_{\mathcal{B}}|X_{\mathcal{N}},\hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c},Y_{n+1}).\end{aligned}$$

The proof of Theorem 4.2.3 is similar to the proof of Theorem 3.2.3, and the details are as follows.

*Proof of Theorem 4.2.3.* Again, we consider the case $Q=\emptyset$. In this case, $R^*_{\mathrm{C/B/S}}$ and $R^*_{\mathrm{R/U/J}}$ can be respectively written as

$$R^*_{\mathrm{C/B/S}}=\max_{p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|x_i,y_i)} I(X_0;\hat{Y}_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}}) \tag{4.15}$$

$$\text{such that } J(\mathcal{S})\geq 0, \forall\mathcal{S}\subseteq\mathcal{N}, \tag{4.16}$$

and

$$\begin{aligned}&R^*_{\mathrm{R/U/J}}\\ =&\max_{p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|x_i,y_i)}\min_{\mathcal{S}\subseteq\mathcal{N}} I(X_0,X_{\mathcal{S}};\hat{Y}_{\mathcal{S}^c},Y_{n+1}|X_{\mathcal{S}^c})-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_0,X_{\mathcal{N}},Y_{n+1},\hat{Y}_{\mathcal{S}^c})\\ =&\max_{p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|x_i,y_i)}\min_{\mathcal{S}\subseteq\mathcal{N}}\{I(X_0;\hat{Y}_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}})+J(\mathcal{S})\}.\end{aligned} \tag{4.17}$$

We show $R^*_{\mathrm{C/B/S}}=R^*_{\mathrm{R/U/J}}$ by showing that $R^*_{\mathrm{C/B/S}}\leq R^*_{\mathrm{R/U/J}}$ and $R^*_{\mathrm{C/B/S}}\geq R^*_{\mathrm{R/U/J}}$ respectively. Under any $p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|x_i,y_i)$ such that $J(\mathcal{S})\geq 0$, $\forall\mathcal{S}\subseteq\mathcal{N}$, we have

$$\min_{\mathcal{S}\subseteq\mathcal{N}}\{I(X_0;\hat{Y}_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}})+J(\mathcal{S})\}=I(X_0;\hat{Y}_{\mathcal{N}},Y_{n+1}|X_{\mathcal{N}}),$$

and thus $R^*_{\mathrm{C/B/S}}\leq R^*_{\mathrm{R/U/J}}$.

To show $R^*_{\mathrm{C/B/S}}\geq R^*_{\mathrm{R/U/J}}$, it is sufficient to show that $R^*_{\mathrm{R/U/J}}$ can be achieved only with the distribution $p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|x_i,y_i)$ such that $J(\mathcal{S})\geq 0$, $\forall\mathcal{S}\subseteq\mathcal{N}$. We will show this by two steps as follows: i) We first show that under any $p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|x_i,y_i)$, if $\mathcal{D}^c\neq\emptyset$, then

$$\mathcal{D}^c\in\operatorname*{argmin}_{\mathcal{S}\subseteq\mathcal{N}} J(\mathcal{S})$$

and

$$\bigcap_{\mathcal{T}\in\operatorname*{argmin}_{\mathcal{S}\subseteq\mathcal{N}} J(\mathcal{S})}\mathcal{T}=\mathcal{D}^c,$$

where $\mathcal{D}$ is defined as in Lemma 4.3.2 and $\underset{\mathcal{S}\subseteq\mathcal{N}}{\operatorname{argmin}} J(\mathcal{S}) := \{\mathcal{T} \subseteq \mathcal{N} : J(\mathcal{T}) = \min_{\mathcal{S}\subseteq\mathcal{N}} J(\mathcal{S})\}$. ii) We then argue that, under the optimal $p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|x_i, y_i)$, $\mathcal{D}^c$ must be $\emptyset$, i.e., $\mathcal{D}$ must be $\mathcal{N}$, and thus by the definition of $\mathcal{D}$, $J(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{N}$.

i) Assuming $\mathcal{D}^c \neq \emptyset$ throughout Part i), we show $\mathcal{D}^c \in \underset{\mathcal{S}\subseteq\mathcal{N}}{\operatorname{argmin}} J(\mathcal{S})$ and $\bigcap_{\mathcal{T}\in \underset{\mathcal{S}\subseteq\mathcal{N}}{\operatorname{argmin}} J(\mathcal{S})} \mathcal{T} = \mathcal{D}^c$.

1) We first show $J(\mathcal{D}^c) < 0$ by using a contradiction argument. Suppose $J(\mathcal{D}^c) \geq 0$, i.e., $J_{\mathcal{D},\mathcal{D}^c}(\mathcal{D}^c) \geq 0$. Then, by Lemma 4.3.3, we have that there exists some nonempty $\mathcal{B} \subseteq \mathcal{D}^c$ such that $J_{\mathcal{D},\mathcal{B}}(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{B}$. This will further imply, by Part 2) of Lemma 4.3.1, that $J_{\mathcal{D}\cup\mathcal{B}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{D}\cup\mathcal{B}$. This is contradictory with the definition of $\mathcal{D}$, and thus $J(\mathcal{D}^c) < 0$.

2) We show that $\forall \mathcal{A} \subseteq \mathcal{D}^c$ and $\mathcal{A} \neq \mathcal{D}^c$, $J(\mathcal{A}) > J(\mathcal{D}^c)$, and thus $J(\mathcal{A}) > \min_{\mathcal{S}\subseteq\mathcal{N}} J(\mathcal{S})$. The proof is still by contradiction. Suppose that there exists some $\mathcal{A} \subseteq \mathcal{D}^c$ and $\mathcal{A} \neq \mathcal{D}^c$ such that $J(\mathcal{A}) \leq J(\mathcal{D}^c)$. Then $J(\mathcal{D}^c) - J(\mathcal{A}) \geq 0$, i.e.,

$$\begin{aligned}
&I(X_{\mathcal{D}^c}; \hat{Y}_{\mathcal{D}}, Y_{n+1}|X_{\mathcal{D}}) - I(Y_{\mathcal{D}^c}; \hat{Y}_{\mathcal{D}^c}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{D}}) \\
&- I(X_{\mathcal{A}}; \hat{Y}_{\mathcal{A}^c}, Y_{n+1}|X_{\mathcal{A}^c}) + I(Y_{\mathcal{A}}; \hat{Y}_{\mathcal{A}}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{A}^c}) \\
=&I(X_{\mathcal{D}^c\setminus\mathcal{A}}; \hat{Y}_{\mathcal{D}}, Y_{n+1}|X_{\mathcal{D}}) + I(X_{\mathcal{A}}; \hat{Y}_{\mathcal{D}}, Y_{n+1}|X_{\mathcal{A}^c}) \\
&- I(Y_{\mathcal{D}^c\setminus\mathcal{A}}; \hat{Y}_{\mathcal{D}^c\setminus\mathcal{A}}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{D}}) - I(Y_{\mathcal{A}}; \hat{Y}_{\mathcal{A}}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{A}^c}) \\
&- I(X_{\mathcal{A}}; \hat{Y}_{\mathcal{D}}, Y_{n+1}|X_{\mathcal{A}^c}) - I(X_{\mathcal{A}}; \hat{Y}_{\mathcal{D}^c\setminus\mathcal{A}}|\hat{Y}_{\mathcal{D}}, Y_{n+1}, X_{\mathcal{A}^c}) + I(Y_{\mathcal{A}}; \hat{Y}_{\mathcal{A}}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{A}^c}) \\
=&I(X_{\mathcal{D}^c\setminus\mathcal{A}}; \hat{Y}_{\mathcal{D}}, Y_{n+1}|X_{\mathcal{D}}) - H(\hat{Y}_{\mathcal{D}^c\setminus\mathcal{A}}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{D}}) + H(\hat{Y}_{\mathcal{D}^c\setminus\mathcal{A}}|Y_{\mathcal{D}^c\setminus\mathcal{A}}, X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{D}}) \\
&- H(\hat{Y}_{\mathcal{D}^c\setminus\mathcal{A}}|\hat{Y}_{\mathcal{D}}, Y_{n+1}, X_{\mathcal{A}^c}) + H(\hat{Y}_{\mathcal{D}^c\setminus\mathcal{A}}|X_{\mathcal{A}}, \hat{Y}_{\mathcal{D}}, Y_{n+1}, X_{\mathcal{A}^c}) \\
=&I(X_{\mathcal{D}^c\setminus\mathcal{A}}; \hat{Y}_{\mathcal{D}}, Y_{n+1}|X_{\mathcal{D}}) - I(Y_{\mathcal{D}^c\setminus\mathcal{A}}; \hat{Y}_{\mathcal{D}^c\setminus\mathcal{A}}|X_{\mathcal{D}}, X_{\mathcal{D}^c\setminus\mathcal{A}}, Y_{n+1}, \hat{Y}_{\mathcal{D}}) \\
=&J_{\mathcal{D},\mathcal{D}^c\setminus\mathcal{A}}(\mathcal{D}^c \setminus \mathcal{A}) \\
\geq&0.
\end{aligned}$$

Again by Lemma 4.3.3 and 4.3.1 successively, we can conclude that there exists some nonempty $\mathcal{B} \subseteq \mathcal{D}^c \setminus \mathcal{A}$, such that $J_{\mathcal{D}\cup\mathcal{B}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{D}\cup\mathcal{B}$, which is in contradiction. Therefore, $J(\mathcal{A}) > J(\mathcal{D}^c) \geq \min_{\mathcal{S}\subseteq\mathcal{N}} J(\mathcal{S})$.

3) We prove that $\forall \mathcal{A}$ with $\mathcal{A}\mathcal{D} \neq \emptyset$ and $\mathcal{A}\mathcal{D}^c \neq \mathcal{D}^c$, $J(\mathcal{A}) > J(\mathcal{A}\cup\mathcal{D}^c) \geq \min_{\mathcal{S}\subseteq\mathcal{N}} J(\mathcal{S})$. Let $\mathcal{A}_1 = \mathcal{A}\mathcal{D}$ and $\mathcal{A}_2 = \mathcal{A}\mathcal{D}^c$. Then, we have, by Lemma 4.3.4, that

$$\begin{aligned}
J(\mathcal{A}) =&J(\mathcal{A}_1 \cup \mathcal{A}_2) = J(\mathcal{A}_1) + J(\mathcal{A}_2) - J(\mathcal{A}_1 \circ \mathcal{A}_2), \\
J(\mathcal{A}_1 \cup \mathcal{D}^c) =&J(\mathcal{A}_1) + J(\mathcal{D}^c) - J(\mathcal{A}_1 \circ \mathcal{D}^c).
\end{aligned}$$

Since $J(\mathcal{A}_2) > J(\mathcal{D}^c)$ by 2), to show $J(\mathcal{A}) > J(\mathcal{A}\cup\mathcal{D}^c) \geq \min_{\mathcal{S}\subseteq\mathcal{N}} J(\mathcal{S})$, we only need to show $J(\mathcal{A}_1 \circ \mathcal{A}_2) \le J(\mathcal{A}_1 \circ \mathcal{D}^c)$. Let $\mathcal{A}_3 = \mathcal{D}^c \setminus \mathcal{A}_2$. Then, we have

$$\begin{aligned}
&J(\mathcal{A}_1 \circ \mathcal{D}^c) - J(\mathcal{A}_1 \circ \mathcal{A}_2)\\
=&I(X_{\mathcal{A}_1}; X_{\mathcal{A}_2\cup\mathcal{A}_3}|X_{(\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3)^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3)^c}, Y_{n+1})\\
&+ I(X_{\mathcal{A}_1}; \hat{Y}_{\mathcal{A}_2\cup\mathcal{A}_3}|X_{\mathcal{A}_1^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3)^c}, Y_{n+1})\\
&+ I(X_{\mathcal{A}_2\cup\mathcal{A}_3}; \hat{Y}_{\mathcal{A}_1}|X_{(\mathcal{A}_2\cup\mathcal{A}_3)^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3)^c}, Y_{n+1})\\
&+ I(\hat{Y}_{\mathcal{A}_1}; \hat{Y}_{\mathcal{A}_2\cup\mathcal{A}_3}|X_{\mathcal{N}}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3)^c}, Y_{n+1})\\
&- I(X_{\mathcal{A}_1}; X_{\mathcal{A}_2}|X_{(\mathcal{A}_1\cup\mathcal{A}_2)^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2)^c}, Y_{n+1}) - I(X_{\mathcal{A}_1}; \hat{Y}_{\mathcal{A}_2}|X_{\mathcal{A}_1^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2)^c}, Y_{n+1})\\
&- I(X_{\mathcal{A}_2}; \hat{Y}_{\mathcal{A}_1}|X_{\mathcal{A}_2^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2)^c}, Y_{n+1}) - I(\hat{Y}_{\mathcal{A}_1}; \hat{Y}_{\mathcal{A}_2}|X_{\mathcal{N}}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2)^c}, Y_{n+1})\\
=&I(X_{\mathcal{A}_1}; X_{\mathcal{A}_3}|X_{(\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3)^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3)^c}, Y_{n+1})\\
&+ I(X_{\mathcal{A}_1}; X_{\mathcal{A}_2}, \hat{Y}_{\mathcal{A}_3}|X_{(\mathcal{A}_1\cup\mathcal{A}_2)^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3)^c}, Y_{n+1})\\
&+ I(X_{\mathcal{A}_3}; \hat{Y}_{\mathcal{A}_1}|X_{(\mathcal{A}_2\cup\mathcal{A}_3)^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3)^c}, Y_{n+1})\\
&+ I(\hat{Y}_{\mathcal{A}_1}; X_{\mathcal{A}_2}, \hat{Y}_{\mathcal{A}_3}|X_{\mathcal{A}_2^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3)^c}, Y_{n+1})\\
&- I(X_{\mathcal{A}_1}; X_{\mathcal{A}_2}|X_{(\mathcal{A}_1\cup\mathcal{A}_2)^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2)^c}, Y_{n+1}) - I(X_{\mathcal{A}_2}; \hat{Y}_{\mathcal{A}_1}|X_{\mathcal{A}_2^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2)^c}, Y_{n+1})\\
=&I(X_{\mathcal{A}_1}; X_{\mathcal{A}_3}|X_{(\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3)^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3)^c}, Y_{n+1})\\
&+ I(X_{\mathcal{A}_1}; \hat{Y}_{\mathcal{A}_3}|X_{(\mathcal{A}_1\cup\mathcal{A}_2)^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3)^c}, Y_{n+1})\\
&+ I(X_{\mathcal{A}_3}; \hat{Y}_{\mathcal{A}_1}|X_{(\mathcal{A}_2\cup\mathcal{A}_3)^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3)^c}, Y_{n+1}) + I(\hat{Y}_{\mathcal{A}_1}; \hat{Y}_{\mathcal{A}_3}|X_{\mathcal{A}_2^c}, \hat{Y}_{(\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3)^c}, Y_{n+1})\\
\ge&0.
\end{aligned}$$

Thus, we have $J(\mathcal{A}) > J(\mathcal{A}_1 \cup \mathcal{D}^c) \ge \min_{\mathcal{S}\subseteq\mathcal{N}} J(\mathcal{S})$.

4) We prove that $\forall \mathcal{A}$ with $\mathcal{A}\mathcal{D} \ne \emptyset$ and $\mathcal{A}\mathcal{D}^c = \mathcal{D}^c$, $J(\mathcal{A}) \ge J(\mathcal{D}^c)$. Letting $\mathcal{A}_1 = \mathcal{A}\mathcal{D}$, we have

$$J(\mathcal{A}) = J(\mathcal{A}_1 \cup \mathcal{D}^c) = J(\mathcal{A}_1) + J(\mathcal{D}^c) - J(\mathcal{A}_1 \circ \mathcal{D}^c).$$

Thus, to show $J(\mathcal{A}) \ge J(\mathcal{D}^c)$, we only need to show $J(\mathcal{A}_1) - J(\mathcal{A}_1 \circ \mathcal{D}^c) \ge 0$. For this, we have

$$\begin{aligned}
&J(\mathcal{A}_1) - J(\mathcal{A}_1 \circ \mathcal{D}^c)\\
=&I(X_{\mathcal{A}_1}; \hat{Y}_{\mathcal{D}^c}, \hat{Y}_{\mathcal{D}\setminus\mathcal{A}_1}, Y_{n+1}|X_{\mathcal{D}^c}, X_{\mathcal{D}\setminus\mathcal{A}_1}) - I(Y_{\mathcal{A}_1}; \hat{Y}_{\mathcal{A}_1}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{D}^c}, \hat{Y}_{\mathcal{D}\setminus\mathcal{A}_1})\\
&- I(X_{\mathcal{A}_1}, \hat{Y}_{\mathcal{A}_1}; X_{\mathcal{D}^c}, \hat{Y}_{\mathcal{D}^c}|X_{\mathcal{D}\setminus\mathcal{A}_1}, \hat{Y}_{\mathcal{D}\setminus\mathcal{A}_1}, Y_{n+1})\\
=&I(X_{\mathcal{A}_1}; X_{\mathcal{D}^c}, \hat{Y}_{\mathcal{D}^c}, \hat{Y}_{\mathcal{D}\setminus\mathcal{A}_1}, Y_{n+1}|X_{\mathcal{D}\setminus\mathcal{A}_1}) - I(Y_{\mathcal{A}_1}; \hat{Y}_{\mathcal{A}_1}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{D}^c}, \hat{Y}_{\mathcal{D}\setminus\mathcal{A}_1})\\
&- I(X_{\mathcal{A}_1}; X_{\mathcal{D}^c}, \hat{Y}_{\mathcal{D}^c}|X_{\mathcal{D}\setminus\mathcal{A}_1}, \hat{Y}_{\mathcal{D}\setminus\mathcal{A}_1}, Y_{n+1}) - I(\hat{Y}_{\mathcal{A}_1}; X_{\mathcal{D}^c}, \hat{Y}_{\mathcal{D}^c}|X_{\mathcal{D}}, \hat{Y}_{\mathcal{D}\setminus\mathcal{A}_1}, Y_{n+1})\\
=&I(X_{\mathcal{A}_1}; \hat{Y}_{\mathcal{D}\setminus\mathcal{A}_1}, Y_{n+1}|X_{\mathcal{D}\setminus\mathcal{A}_1}) - I(\hat{Y}_{\mathcal{A}_1}; X_{\mathcal{D}^c}, \hat{Y}_{\mathcal{D}^c}, Y_{\mathcal{A}_1}|X_{\mathcal{D}}, \hat{Y}_{\mathcal{D}\setminus\mathcal{A}_1}, Y_{n+1})\\
=&J_{\mathcal{D}}(\mathcal{A}_1)\\
\ge&0,
\end{aligned}$$

and thus $J(\mathcal{A}) \geq J(\mathcal{D}^c)$.

Combining 2)-4), we can conclude that

$$\mathcal{D}^c \in \operatorname*{argmin}_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})$$

and

$$\bigcap_{\mathcal{T} \in \operatorname*{argmin}_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})} \mathcal{T} = \mathcal{D}^c.$$

ii) We now argue that under the optimal $p(x_0) \prod_{i=1}^n p(x_i) p(\hat{y}_i|x_i, y_i)$ that achieves $R^*_{\mathrm{R/U/J}}$, if $\mathcal{D}^c \neq \emptyset$, then $R^*_{\mathrm{R/U/J}}$ is not optimal; and hence $\mathcal{D}^c$ must be $\emptyset$.

Suppose $\mathcal{D}^c \neq \emptyset$ at the optimum. Then,

$$\mathcal{D}^c \in \operatorname*{argmin}_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})$$

and

$$\bigcap_{\mathcal{T} \in \operatorname*{argmin}_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})} \mathcal{T} = \mathcal{D}^c.$$

Therefore,

$$\begin{aligned} R^*_{\mathrm{R/U/J}} =& I(X_0, X_{\mathcal{D}^c}; \hat{Y}_{\mathcal{D}}, Y_{n+1}|X_{\mathcal{D}}) - I(Y_{\mathcal{D}^c}; \hat{Y}_{\mathcal{D}^c}|X_0, X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{D}}) && (4.18) \\ =& I(X_0, X_{\mathcal{T}}; \hat{Y}_{\mathcal{T}^c}, Y_{n+1}|X_{\mathcal{T}^c}) - I(Y_{\mathcal{T}}; \hat{Y}_{\mathcal{T}}|X_0, X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{T}^c}), && (4.19) \end{aligned}$$

for any $\mathcal{T} \in \operatorname*{argmin}_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})$, $\mathcal{T} \neq \mathcal{D}^c$.

We argue that higher rate can be achieved. Consider $\hat{Y}'_1, \hat{Y}'_2, \ldots, \hat{Y}'_n$, where $\hat{Y}'_i = \hat{Y}_i$ for any $i \in \mathcal{D}$, and $\hat{Y}'_i = \hat{Y}_i$ with probability $p$ and $\hat{Y}'_i = \emptyset$ with probability $1-p$ for any $i \in \mathcal{D}^c$. When $p = 1$, the achievable rate with $\hat{Y}'_1, \hat{Y}'_2, \ldots, \hat{Y}'_n$ is $R^*_{\mathrm{R/U/J}}$. As $p$ decreases from 1, in (4.18) and (4.19), both

$$I(X_0, X_{\mathcal{D}^c}; \hat{Y}_{\mathcal{D}}, Y_{n+1}|X_{\mathcal{D}}) - I(Y_{\mathcal{D}^c}; \hat{Y}_{\mathcal{D}^c}|X_0, X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{D}})$$

and

$$I(X_0, X_{\mathcal{T}}; \hat{Y}_{\mathcal{T}^c}, Y_{n+1}|X_{\mathcal{T}^c}) - I(Y_{\mathcal{T}}; \hat{Y}_{\mathcal{T}}|X_0, X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{T}^c})$$

will increase, where $\mathcal{T} \in \operatorname*{argmin}_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})$, $\mathcal{T} \neq \mathcal{D}^c$. Thus, no matter how

$$I(X_0, X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y_{n+1}|X_{\mathcal{S}^c}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_0, X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{S}^c})$$

will change as $p$ decreases for $\mathcal{S} \notin \operatorname*{argmin}_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})$, it is certain that there exists a $p^*$ such that the achievable rate by using $\hat{Y}'_1, \hat{Y}'_2, \ldots, \hat{Y}'_n$ is larger than $R^*_{\mathrm{R/U/J}}$. This is in contradiction with the optimality of $R^*_{\mathrm{R/U/J}}$, and thus at the optimum, $\mathcal{D}^c$ must be $\emptyset$ , i.e., $J(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{N}$. This completes the proof of Theorem 4.2.3. □

## 4.4 Cumulative encoding/block-by-block backward decoding/compression-message joint decoding and Necessity of Joint Decodability

Some notations and lemmas are introduced to facilitate the later discussion. For any $\mathcal{A}, \mathcal{B} \subseteq \mathcal{N}$, let

$$K_{\mathcal{A},\mathcal{B}}(\mathcal{S}) := I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{B}\setminus\mathcal{S}}, \hat{Y}_{\mathcal{A}}, Y_{n+1} | X_0, X_{\mathcal{A}}, X_{\mathcal{B}\setminus\mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}}, Y_{n+1}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{B}\setminus\mathcal{S}}), \forall \mathcal{S} \subseteq \mathcal{B}, \tag{4.20}$$

$$K_{\mathcal{B}}(\mathcal{S}) := K_{\emptyset,\mathcal{B}}(\mathcal{S}) = I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{B}\setminus\mathcal{S}}, Y_{n+1} | X_0, X_{\mathcal{B}\setminus\mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_{\mathcal{B}}, \hat{Y}_{\mathcal{B}\setminus\mathcal{S}}, Y_{n+1}), \forall \mathcal{S} \subseteq \mathcal{B}, \tag{4.21}$$

$$R_{\mathcal{B}}(\mathcal{S}) := I(X_0, X_{\mathcal{S}}; \hat{Y}_{\mathcal{B}\setminus\mathcal{S}}, Y_{n+1} | X_{\mathcal{B}\setminus\mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_{\mathcal{B}}, \hat{Y}_{\mathcal{B}\setminus\mathcal{S}}, Y_{n+1}), \forall \mathcal{S} \subseteq \mathcal{B}. \tag{4.22}$$

**Lemma 4.4.1.** *1) If $K_{\mathcal{A}}(\mathcal{S}_1) > 0$, for any nonempty $\mathcal{S}_1 \subseteq \mathcal{A}$, and $K_{\mathcal{B}}(\mathcal{S}_2) > 0$, for any nonempty $\mathcal{S}_2 \subseteq \mathcal{B}$, then $K_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}) > 0$, for any nonempty $\mathcal{S} \subseteq \mathcal{A} \cup \mathcal{B}$.*

*2) If $K_{\mathcal{A}}(\mathcal{S}_1) > 0$, for any nonempty $\mathcal{S}_1 \subseteq \mathcal{A}$, and $K_{\mathcal{A},\mathcal{B}}(\mathcal{S}_2) > 0$, for any nonempty $\mathcal{S}_2 \subseteq \mathcal{B}$, then $K_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}) > 0$, for any nonempty $\mathcal{S} \subseteq \mathcal{A} \cup \mathcal{B}$.*

**Lemma 4.4.2.** *Under any $p(x_0)\prod_{i=1}^{n} p(x_i)p(\hat{y}_i|x_i, y_i)$, there exists a unique set $\mathcal{D}_J$, which is the largest subset of $\mathcal{N}$ satisfying*

$$K_{\mathcal{D}_J}(\mathcal{S}) > 0, \forall \mathcal{S} \subseteq \mathcal{D}_J, \mathcal{S} \neq \emptyset.$$

**Lemma 4.4.3.** *If $K_{\mathcal{A},\mathcal{B}}(\mathcal{B}) > 0$ for some nonempty $\mathcal{B}$, then there exists some nonempty $\mathcal{C} \subseteq \mathcal{B}$ such that $K_{\mathcal{A},\mathcal{C}}(\mathcal{S}) > 0$, for any nonempty $\mathcal{S} \subseteq \mathcal{C}$.*

**Lemma 4.4.4.** *For any disjoint $\mathcal{A}$ and $\mathcal{B}$, and any $\mathcal{S} \subseteq \mathcal{A} \cup \mathcal{B}$, let $\mathcal{S}_1 = \mathcal{S}\mathcal{A}$ and $\mathcal{S}_2 = \mathcal{S}\mathcal{B}$. Then, we have:*

*1) $R_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}) \geq R_{\mathcal{A}}(\mathcal{S}_1) + K_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}_2)$.*

*2) Specially, when $\mathcal{S}_2 = \mathcal{B}$, $R_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}) = R_{\mathcal{A}}(\mathcal{S}_1) + K_{\mathcal{A},\mathcal{B}}(\mathcal{B})$.*

Lemmas 4.4.1-4.4.3 can be proved along the same lines as the proofs of Lemmas 4.3.1-4.3.3 respectively, while the proof of Lemma 4.4.4 is given in Section 4.5.

The cumulative encoding/block-by-block backward decoding/compression-message joint decoding scheme is presented in the following proof.

*Proof of Theorem 4.2.4.* The uniqueness of $\mathcal{D}_\mathrm{J}$ has been established in Lemma 4.4.2. Below, we focus on showing that i) the rate in (4.4) is achievable, and ii) the compressions in the set $\mathcal{D}_\mathrm{J}$ can be decoded jointly with $X_0$.

To make the presentation easier to follow, we first consider the case when $\mathcal{D}_\mathrm{J} = \mathcal{N}$, i.e., the case when

$$I(X_\mathcal{S}; \hat{Y}_{\mathcal{S}^c}, Y_{n+1}|X_0, X_{\mathcal{S}^c}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S}|X_0, X_\mathcal{N}, Y_{n+1}, \hat{Y}_{\mathcal{S}^c}) > 0, \forall \mathcal{S} \subseteq \mathcal{N}, \mathcal{S} \neq \emptyset, \tag{4.23}$$

and show that

$$R_{\mathrm{C/B/J}} < \min_{\mathcal{S}\subseteq\mathcal{N}} I(X_0, X_\mathcal{S}; \hat{Y}_{\mathcal{S}^c}, Y_{n+1}|X_{\mathcal{S}^c}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S}|X_0, X_\mathcal{N}, Y_{n+1}, \hat{Y}_{\mathcal{S}^c}) \tag{4.24}$$

is achievable. The case of $\mathcal{D}_\mathrm{J} \neq \mathcal{N}$ will follow immediately after the case of $\mathcal{D}_\mathrm{J} = \mathcal{N}$ is treated.

Fix $p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|x_i, y_i)$. Assume (4.23) holds. The codebook generation and encoding process here are exactly the same as those in the proof of Theorem 4.2.2, and hence omitted. For the decoding, the destination finds the unique message vector $\mathbf{m} = (m_1, m_2, \ldots, m_B)$ and some $\mathbf{l}^{B+M} = (\mathbf{l}_1, \ldots, \mathbf{l}_{B+M})$ such that for any $b = 1, \ldots, B+M$,

$$\Big(\mathbf{X}_{0,b}(m_b), (\mathbf{X}_{1,b}(l_{1,b-1}), \hat{\mathbf{Y}}_{1,b}(l_{1,b}|l_{1,b-1})), \ldots, (\mathbf{X}_{n,b}(l_{n,b-1}), \hat{\mathbf{Y}}_{n,b}(l_{n,b}|l_{n,b-1})), \mathbf{Y}_{n+1,b}\Big) \in A_\epsilon(X_0, X_\mathcal{N}, \hat{Y}_\mathcal{N}, Y_{n+1}), \tag{4.25}$$

where $m_b = 1$ is dummy message for all $b \in [B+1 : B+M]$.

Again, this can be done backwardly as follows.

a) The destination first finds the unique $\mathbf{l}_B$ such that there exists some $\mathbf{l}_{B+1}^{B+M} = (\mathbf{l}_{B+1}, \ldots, \mathbf{l}_{B+M})$ satisfying (4.25) for any $b = B+1, \ldots, B+M$. Through the similar lines as in the proof of Theorem 4.2.2 with $\mathbf{X}_{0,b}(m_b), b \in [B+1 : B+M]$ taken into account and treated as known signals, it follows that $\mathbf{l}_B$ can be decoded if (4.23) holds.

b) Backwards and sequentially from block $b = B$ to block $b = 1$, the destination finds the unique pair $(m_b, \mathbf{l}_{b-1})$, such that $(m_b, \mathbf{l}_{b-1})$ satisfies (4.25), where $\mathbf{l}_b$ has already been recovered due to the backward property of decoding.

At each block $b = B, B-1, \ldots, 1$, error occurs with $m_b$ if the true $m_b$ does not satisfy (4.25) with any $\mathbf{l}_{b-1}$, or a false $m_b$ satisfies (4.25) with some $\mathbf{l}_{b-1}$. According to the properties of typical sequences, the true $(m_b, \mathbf{l}_{b-1})$ satisfies (4.25) with high probability.

For a false $m_b$ and a $\mathbf{l}_{b-1}$ with false $\{l_{i,b-1}, i \in \mathcal{S}\}$ but true $\{l_{i,b-1}, i \in \mathcal{S}^c\}$, $\mathbf{X}_{0,b}(m_b)$ and $(\mathbf{X}_b(\mathcal{S}), \hat{\mathbf{Y}}_b(\mathcal{S}))$ and $(\mathbf{X}_b(\mathcal{S}^c), \hat{\mathbf{Y}}_b(\mathcal{S}^c), \mathbf{Y}_{n+1,b})$ are mutually independent, and the probability that $(m_b, \mathbf{l}_{b-1})$ satisfies (4.25) can be upper bounded by

$$\begin{aligned} &2^{T(H(X_0,X_\mathcal{N},\hat{Y}_\mathcal{N},Y_{n+1})+\epsilon)}2^{-T(H(X_0)-\epsilon)} \\ &\times 2^{-T(H(X_{\mathcal{S}^c},\hat{Y}_{\mathcal{S}^c},Y_{n+1})-\epsilon)}2^{-T(H(X_\mathcal{S})-\epsilon)}2^{-T(\sum_{i\in\mathcal{S}}(H(\hat{Y}_i|X_i)-\epsilon))}. \end{aligned}$$

Since the number of such false $(m_b, \mathbf{l}_{b-1})$ is upper bounded by

$$2^{TR_{\mathrm{C/B/J}}}\prod_{i\in\mathcal{S}}2^{T(I(Y_i;\hat{Y}_i|X_i)+\epsilon)},$$

with the union bound, it is easy to check that the probability of finding a false $m_b$ goes to zero as $T \to \infty$, if (4.24) holds.

Then, based on the recovered $m_b$ and $\mathbf{l}_b$, again from the proof of Theorem 4.2.2 with $\mathbf{X}_{0,b}(m_b)$ taken into account and treated as known signal, it follows that $\mathbf{l}_{b-1}$ can be decoded if (4.23) holds.

Combining a) and b), we can conclude that both $\mathbf{m}$ and $\mathbf{l}^B$ can be decoded if both (4.23) and (4.24) hold.

If under $p(x_0)\prod_{i=1}^n p(x_i)p(\hat{y}_i|x_i,y_i)$, $\mathcal{D}_\mathrm{J} \neq \mathcal{N}$, then through the same line as above with $\mathcal{N}$ replaced by $\mathcal{D}_\mathrm{J}$, it readily follows that

$$R_{\mathrm{C/B/J}} < \min_{\mathcal{S}\subseteq\mathcal{D}_\mathrm{J}} I(X_0, X_\mathcal{S}; \hat{Y}_{\mathcal{D}_\mathrm{J}\setminus\mathcal{S}}, Y_{n+1}|X_{\mathcal{D}_\mathrm{J}\setminus\mathcal{S}}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S}|X_0, X_{\mathcal{D}_\mathrm{J}}, Y_{n+1}, \hat{Y}_{\mathcal{D}_\mathrm{J}\setminus\mathcal{S}})$$

is achievable; and $\hat{Y}_{\mathcal{D}_\mathrm{J}}$, or more strictly, $\{l_i^B, i \in \mathcal{D}_\mathrm{J}\}$, can be decoded jointly with $X_0$ since

$$I(X_\mathcal{S}; \hat{Y}_{\mathcal{D}_\mathrm{J}\setminus\mathcal{S}}, Y_{n+1}|X_0, X_{\mathcal{D}_\mathrm{J}\setminus\mathcal{S}}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S}|X_0, X_{\mathcal{D}_\mathrm{J}}, Y_{n+1}, \hat{Y}_{\mathcal{D}_\mathrm{J}\setminus\mathcal{S}}) > 0,$$

for any nonempty $\mathcal{S} \subseteq \mathcal{D}_\mathrm{J}$. □

Now, we demonstrate that only those relay nodes, whose compressions can be eventually decoded, are helpful to the decoding of the original message.

*Proof of Theorem 4.2.5.* Still consider the case $Q = \emptyset$. The uniqueness of $\mathcal{D}_\mathrm{J}$ has been treated in Lemma 4.4.2, while the uniqueness of $\mathcal{D}'_\mathrm{J}$ can be established along the same lines. To prove Theorem 4.2.5, in terms of the notations defined in this section, we will sequentially prove that: i) $\max_{\mathcal{M}\subseteq\mathcal{N}}\min_{\mathcal{S}\subseteq\mathcal{M}} R_\mathcal{M}(\mathcal{S}) = \min_{\mathcal{S}\subseteq\mathcal{D}_\mathrm{J}} R_{\mathcal{D}_\mathrm{J}}(\mathcal{S})$; ii) $\min_{\mathcal{S}\subseteq\mathcal{M}} R_\mathcal{M}(\mathcal{S}) < \min_{\mathcal{S}\subseteq\mathcal{D}'_\mathrm{J}} R_{\mathcal{D}'_\mathrm{J}}(\mathcal{S})$, for any $\mathcal{M} \nsubseteq \mathcal{D}'_\mathrm{J}$; iii) $\max_{\mathcal{M}\subseteq\mathcal{N}}\min_{\mathcal{S}\subseteq\mathcal{M}} R_\mathcal{M}(\mathcal{S}) = \min_{\mathcal{S}\subseteq\mathcal{D}'_\mathrm{J}} R_{\mathcal{D}'_\mathrm{J}}(\mathcal{S})$.

i) We prove $\max_{\mathcal{M}\subseteq\mathcal{N}}\min_{\mathcal{S}\subseteq\mathcal{M}} R_\mathcal{M}(\mathcal{S}) = \min_{\mathcal{S}\subseteq\mathcal{D}_\mathrm{J}} R_{\mathcal{D}_\mathrm{J}}(\mathcal{S})$ by proving that: 1) For any $\mathcal{M} \cap \mathcal{D}_\mathrm{J} = \mathcal{D}_\mathrm{J}$, $\mathcal{M} \neq \mathcal{D}_\mathrm{J}$, $\min_{\mathcal{S}\subseteq\mathcal{M}} R_\mathcal{M}(\mathcal{S}) \leq \min_{\mathcal{S}\subseteq\mathcal{D}_\mathrm{J}} R_{\mathcal{D}_\mathrm{J}}(\mathcal{S})$. 2) For any $\mathcal{M} \cap \mathcal{D}_{\rm J} \neq \mathcal{D}_{\rm J}$, $\min_{\mathcal{S}\subseteq\mathcal{M}} R_{\mathcal{M}}(\mathcal{S}) \leq \min_{\mathcal{S}\subseteq\mathcal{M}\cup\mathcal{D}_{\rm J}} R_{\mathcal{M}\cup\mathcal{D}_{\rm J}}(\mathcal{S})$, and thus by 1), $\min_{\mathcal{S}\subseteq\mathcal{M}} R_{\mathcal{M}}(\mathcal{S}) \leq \min_{\mathcal{S}\subseteq\mathcal{D}_{\rm J}} R_{\mathcal{D}_{\rm J}}(\mathcal{S})$ . The details are as follows.

1) Assume $\mathcal{M}\cap\mathcal{D}_{\rm J} = \mathcal{D}_{\rm J}$, $\mathcal{M} \neq \mathcal{D}_{\rm J}$. We show $\min_{\mathcal{S}\subseteq\mathcal{M}} R_{\mathcal{M}}(\mathcal{S}) \leq \min_{\mathcal{S}\subseteq\mathcal{D}_{\rm J}} R_{\mathcal{D}_{\rm J}}(\mathcal{S})$ by showing that for any $\mathcal{S} \subseteq \mathcal{D}_{\rm J}$, $R_{\mathcal{M}}(\mathcal{S} \cup (\mathcal{M} \setminus \mathcal{D}_{\rm J})) \leq R_{\mathcal{D}_{\rm J}}(\mathcal{S})$.

For any $\mathcal{S} \subseteq \mathcal{D}_{\rm J}$, by Part 2) of Lemma 4.4.4, we have

$$R_{\mathcal{M}}(\mathcal{S} \cup (\mathcal{M} \setminus \mathcal{D}_{\rm J})) = R_{\mathcal{D}_{\rm J}\cup(\mathcal{M}\setminus\mathcal{D}_{\rm J})}(\mathcal{S} \cup (\mathcal{M} \setminus \mathcal{D}_{\rm J})) = R_{\mathcal{D}_{\rm J}}(\mathcal{S}) + K_{\mathcal{D}_{\rm J},\mathcal{M}\setminus\mathcal{D}_{\rm J}}(\mathcal{M} \setminus \mathcal{D}_{\rm J}).$$

We argue $K_{\mathcal{D}_{\rm J},\mathcal{M}\setminus\mathcal{D}_{\rm J}}(\mathcal{M}\setminus\mathcal{D}_{\rm J}) \leq 0$ by contradiction. Suppose $K_{\mathcal{D}_{\rm J},\mathcal{M}\setminus\mathcal{D}_{\rm J}}(\mathcal{M}\setminus\mathcal{D}_{\rm J}) > 0$. Then, by Lemma 4.4.3, we have that there exists some nonempty $\mathcal{C} \subseteq \mathcal{M} \setminus \mathcal{D}_{\rm J}$ such that $K_{\mathcal{D}_{\rm J},\mathcal{C}}(\mathcal{S}) > 0$, for any nonempty $\mathcal{S} \subseteq \mathcal{C}$. This will further imply, by Part 2) of Lemma 4.4.1, that $K_{\mathcal{D}_{\rm J}\cup\mathcal{C}}(\mathcal{S}) > 0$, for any nonempty $\mathcal{S} \subseteq \mathcal{D}_{\rm J} \cup \mathcal{C}$, which is in contradiction with the definition of $\mathcal{D}_{\rm J}$. Thus, we must have $K_{\mathcal{D}_{\rm J},\mathcal{M}\setminus\mathcal{D}_{\rm J}}(\mathcal{M}\setminus\mathcal{D}_{\rm J}) \leq 0$, and $R_{\mathcal{M}}(\mathcal{S} \cup (\mathcal{M} \setminus \mathcal{D}_{\rm J})) \leq R_{\mathcal{D}_{\rm J}}(\mathcal{S})$.

2) Assume $\mathcal{M} \cap \mathcal{D}_{\rm J} \neq \mathcal{D}_{\rm J}$. For any $\mathcal{S} \subseteq \mathcal{M} \cup \mathcal{D}_{\rm J}$, let $\mathcal{S}_1 = \mathcal{S}\mathcal{M}$ and $\mathcal{S}_2 = \mathcal{S}(\mathcal{D}_{\rm J} \setminus \mathcal{M})$. By Part 1) of Lemma 4.4.4, we have

$$R_{\mathcal{M}\cup\mathcal{D}_{\rm J}}(\mathcal{S}) = R_{\mathcal{M}\cup(\mathcal{D}_{\rm J}\setminus\mathcal{M})}(\mathcal{S}) \geq R_{\mathcal{M}}(\mathcal{S}_1) + K_{\mathcal{M}\cup\mathcal{D}_{\rm J}}(\mathcal{S}_2),$$

and then,

$$\begin{aligned}
\min_{\mathcal{S}\subseteq\mathcal{M}\cup\mathcal{D}_{\rm J}} R_{\mathcal{M}\cup\mathcal{D}_{\rm J}}(\mathcal{S}) &\geq \min_{\mathcal{S}\subseteq\mathcal{M}\cup\mathcal{D}_{\rm J}} \{R_{\mathcal{M}}(\mathcal{S}_1) + K_{\mathcal{M}\cup\mathcal{D}_{\rm J}}(\mathcal{S}_2)\} \\
&\geq \min_{\mathcal{S}\subseteq\mathcal{M}\cup\mathcal{D}_{\rm J}} \{R_{\mathcal{M}}(\mathcal{S}_1) + K_{\mathcal{D}_{\rm J}}(\mathcal{S}_2)\} \\
&= \min_{\mathcal{S}_1\subseteq\mathcal{M},\mathcal{S}_2\subseteq\mathcal{D}_{\rm J}\setminus\mathcal{M}} \{R_{\mathcal{M}}(\mathcal{S}_1) + K_{\mathcal{D}_{\rm J}}(\mathcal{S}_2)\} \\
&= \min_{\mathcal{S}_1\subseteq\mathcal{M}} R_{\mathcal{M}}(\mathcal{S}_1) + \min_{\mathcal{S}_2\subseteq\mathcal{D}_{\rm J}\setminus\mathcal{M}} K_{\mathcal{D}_{\rm J}}(\mathcal{S}_2) \\
&\geq \min_{\mathcal{S}_1\subseteq\mathcal{M}} R_{\mathcal{M}}(\mathcal{S}_1),
\end{aligned}$$

where the last inequality follows from the fact that $K_{\mathcal{D}_{\rm J}}(\mathcal{S}_2) > 0$, for any nonempty $\mathcal{S}_2 \subseteq \mathcal{D}_{\rm J}$.

ii) We can prove $\min_{\mathcal{S}\subseteq\mathcal{M}} R_{\mathcal{M}}(\mathcal{S}) < \min_{\mathcal{S}\subseteq\mathcal{D}'_{\rm J}} R_{\mathcal{D}'_{\rm J}}(\mathcal{S})$, for any $\mathcal{M} \nsubseteq \mathcal{D}'_{\rm J}$ by two similar steps as follows.

1) Through the similar lines as in Step 1) of Part i), we can prove that for any $\mathcal{M} \cap \mathcal{D}'_{\rm J} = \mathcal{D}'_{\rm J}$, $\mathcal{M} \neq \mathcal{D}'_{\rm J}$, $\min_{\mathcal{S}\subseteq\mathcal{M}} R_{\mathcal{M}}(\mathcal{S}) < \min_{\mathcal{S}\subseteq\mathcal{D}'_{\rm J}} R_{\mathcal{D}'_{\rm J}}(\mathcal{S})$. The only difference is that here the inequality is strict, but it can be easily justified by noting that "=" is included in the definition of $\mathcal{D}'_{\rm J}$.

2) From Step 2) of Part i), it can be similarly proved that for any $\mathcal{M}\cap\mathcal{D}'_{\rm J} \neq \mathcal{D}'_{\rm J}$, $\min_{\mathcal{S}\subseteq\mathcal{M}} R_{\mathcal{M}}(\mathcal{S}) \leq \min_{\mathcal{S}\subseteq\mathcal{M}\cup\mathcal{D}'_{\rm J}} R_{\mathcal{M}\cup\mathcal{D}'_{\rm J}}(\mathcal{S})$. Therefore, if, further, $\mathcal{M} \nsubseteq \mathcal{D}'_{\rm J}$, then by 1) we have

$$\min_{\mathcal{S}\subseteq\mathcal{M}} R_{\mathcal{M}}(\mathcal{S}) \leq \min_{\mathcal{S}\subseteq\mathcal{M}\cup\mathcal{D}'_{\rm J}} R_{\mathcal{M}\cup\mathcal{D}'_{\rm J}}(\mathcal{S}) < \min_{\mathcal{S}\subseteq\mathcal{D}'_{\rm J}} R_{\mathcal{D}'_{\rm J}}(\mathcal{S}).$$

iii) From Part ii), we have 1) $\min_{\mathcal{S}\subseteq\mathcal{M}} R_{\mathcal{M}}(\mathcal{S}) < \min_{\mathcal{S}\subseteq\mathcal{D}'_{\mathrm{J}}} R_{\mathcal{D}'_{\mathrm{J}}}(\mathcal{S})$, for any $\mathcal{M}\cap\mathcal{D}'_{\mathrm{J}} = \mathcal{D}'_{\mathrm{J}}$, $\mathcal{M} \neq \mathcal{D}'_{\mathrm{J}}$, and 2) for any $\mathcal{M}\cap\mathcal{D}'_{\mathrm{J}} \neq \mathcal{D}'_{\mathrm{J}}$, $\min_{\mathcal{S}\subseteq\mathcal{M}} R_{\mathcal{M}}(\mathcal{S}) \leq \min_{\mathcal{S}\subseteq\mathcal{M}\cup\mathcal{D}'_{\mathrm{J}}} R_{\mathcal{M}\cup\mathcal{D}'_{\mathrm{J}}}(\mathcal{S}) \leq \min_{\mathcal{S}\subseteq\mathcal{D}'_{\mathrm{J}}} R_{\mathcal{D}'_{\mathrm{J}}}(\mathcal{S})$. Thus, it follows immediately that

$$\min_{\mathcal{S}\subseteq\mathcal{D}'_{\mathrm{J}}} R_{\mathcal{D}'_{\mathrm{J}}}(\mathcal{S}) = \max_{\mathcal{M}\subseteq\mathcal{N}} \min_{\mathcal{S}\subseteq\mathcal{M}} R_{\mathcal{M}}(\mathcal{S}).$$

This completes the proof of Theorem 4.2.5. □

## 4.5 Proofs of Lemmas 4.3.1-4.3.4 and 4.4.4

*Proof of Lemma 4.3.1.* For any $\mathcal{S}\subseteq\mathcal{A}\cup\mathcal{B}$, let $\mathcal{S}_1 = \mathcal{S}\mathcal{A}$ and $\mathcal{S}_2 = \mathcal{S}(\mathcal{B}\setminus\mathcal{A})$. Then,

$$\begin{aligned}
&J_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S})\\
=&I(X_{\mathcal{S}};\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}},Y_{n+1}|X_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}}) - I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_{\mathcal{A}\cup\mathcal{B}},\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}},Y_{n+1})\\
=&I(X_{\mathcal{S}_1};\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}},Y_{n+1}|X_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}}) + I(X_{\mathcal{S}_2};\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}},Y_{n+1}|X_{\mathcal{S}_1},X_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}})\\
&- I(Y_{\mathcal{S}_1};\hat{Y}_{\mathcal{S}_1}|X_{\mathcal{A}\cup\mathcal{B}},\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}},Y_{n+1}) - I(Y_{\mathcal{S}_2};\hat{Y}_{\mathcal{S}_2}|X_{\mathcal{A}\cup\mathcal{B}},\hat{Y}_{\mathcal{S}_1},\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}},Y_{n+1})\\
=&I(X_{\mathcal{S}_1};\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}},Y_{n+1}|X_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}}) + I(X_{\mathcal{S}_2};\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}},Y_{n+1}|X_{\mathcal{S}_1},X_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}})\\
&- [I(Y_{\mathcal{S}_1};\hat{Y}_{\mathcal{S}_1}|X_{\mathcal{A}},\hat{Y}_{\mathcal{A}\setminus\mathcal{S}_1},Y_{n+1}) - I(\hat{Y}_{\mathcal{S}_1};X_{\mathcal{B}\setminus\mathcal{A}},\hat{Y}_{\mathcal{B}\mathcal{A}^c\setminus\mathcal{S}_2}|X_{\mathcal{A}},\hat{Y}_{\mathcal{A}\setminus\mathcal{S}_1},Y_{n+1})]\\
&- I(Y_{\mathcal{S}_2};\hat{Y}_{\mathcal{S}_2}|X_{\mathcal{A}\cup\mathcal{B}},\hat{Y}_{\mathcal{S}_1},\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}},Y_{n+1})\\
=&[I(X_{\mathcal{S}_1};\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}},Y_{n+1}|X_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}}) - I(Y_{\mathcal{S}_1};\hat{Y}_{\mathcal{S}_1}|X_{\mathcal{A}},\hat{Y}_{\mathcal{A}\setminus\mathcal{S}_1},Y_{n+1})]\\
&+ [I(X_{\mathcal{S}_2};\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}},Y_{n+1}|X_{\mathcal{S}_1},X_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}}) + I(\hat{Y}_{\mathcal{S}_1};X_{\mathcal{B}\setminus\mathcal{A}},\hat{Y}_{\mathcal{B}\mathcal{A}^c\setminus\mathcal{S}_2}|X_{\mathcal{A}},\hat{Y}_{\mathcal{A}\setminus\mathcal{S}_1},Y_{n+1})]\\
&- I(Y_{\mathcal{S}_2};\hat{Y}_{\mathcal{S}_2}|X_{\mathcal{A}},X_{\mathcal{B}},\hat{Y}_{\mathcal{A}},\hat{Y}_{\mathcal{B}\setminus\mathcal{S}_2},Y_{n+1})\\
\geq&[I(X_{\mathcal{S}_1};\hat{Y}_{\mathcal{A}\setminus\mathcal{S}_1},Y_{n+1}|X_{\mathcal{A}\setminus\mathcal{S}_1}) - I(Y_{\mathcal{S}_1};\hat{Y}_{\mathcal{S}_1}|X_{\mathcal{A}},\hat{Y}_{\mathcal{A}\setminus\mathcal{S}_1},Y_{n+1})]\\
&+ [I(X_{\mathcal{S}_2};\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}},Y_{n+1}|X_{\mathcal{S}_1},X_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}}) + I(\hat{Y}_{\mathcal{S}_1};X_{\mathcal{B}\setminus\mathcal{A}},\hat{Y}_{\mathcal{B}\mathcal{A}^c\setminus\mathcal{S}_2}|X_{\mathcal{A}},\hat{Y}_{\mathcal{A}\setminus\mathcal{S}_1},Y_{n+1})]\\
&- I(Y_{\mathcal{S}_2};\hat{Y}_{\mathcal{S}_2}|X_{\mathcal{A}},X_{\mathcal{B}},\hat{Y}_{\mathcal{A}},\hat{Y}_{\mathcal{B}\setminus\mathcal{S}_2},Y_{n+1})\\
=&[I(X_{\mathcal{S}_2};\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}},Y_{n+1}|X_{\mathcal{S}_1},X_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}})\\
&+ I(\hat{Y}_{\mathcal{S}_1};X_{\mathcal{S}_2},X_{\mathcal{B}\mathcal{A}^c\setminus\mathcal{S}_2},\hat{Y}_{\mathcal{B}\mathcal{A}^c\setminus\mathcal{S}_2}|X_{\mathcal{A}},\hat{Y}_{\mathcal{A}\setminus\mathcal{S}_1},Y_{n+1})]\\
&- I(Y_{\mathcal{S}_2};\hat{Y}_{\mathcal{S}_2}|X_{\mathcal{A}},X_{\mathcal{B}},\hat{Y}_{\mathcal{A}},\hat{Y}_{\mathcal{B}\setminus\mathcal{S}_2},Y_{n+1}) + J_{\mathcal{A}}(\mathcal{S}_1)\\
\geq&[I(X_{\mathcal{S}_2};\hat{Y}_{(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{S}},Y_{n+1}|X_{\mathcal{A}},X_{\mathcal{B}\setminus\mathcal{S}_2}) + I(\hat{Y}_{\mathcal{S}_1};X_{\mathcal{S}_2}|X_{\mathcal{A}},X_{\mathcal{B}\mathcal{A}^c\setminus\mathcal{S}_2},\hat{Y}_{\mathcal{B}\mathcal{A}^c\setminus\mathcal{S}_2},\hat{Y}_{\mathcal{A}\setminus\mathcal{S}_1},Y_{n+1})]\\
&- I(Y_{\mathcal{S}_2};\hat{Y}_{\mathcal{S}_2}|X_{\mathcal{A}},X_{\mathcal{B}},\hat{Y}_{\mathcal{A}},\hat{Y}_{\mathcal{B}\setminus\mathcal{S}_2},Y_{n+1}) + J_{\mathcal{A}}(\mathcal{S}_1)\\
=&I(X_{\mathcal{S}_2};\hat{Y}_{\mathcal{A}},\hat{Y}_{\mathcal{B}\setminus\mathcal{S}_2},Y_{n+1}|X_{\mathcal{A}},X_{\mathcal{B}\setminus\mathcal{S}_2}) - I(Y_{\mathcal{S}_2};\hat{Y}_{\mathcal{S}_2}|X_{\mathcal{A}},X_{\mathcal{B}},\hat{Y}_{\mathcal{A}},\hat{Y}_{\mathcal{B}\setminus\mathcal{S}_2},Y_{n+1}) + J_{\mathcal{A}}(\mathcal{S}_1)\\
=&J_{\mathcal{A}}(\mathcal{S}_1) + J_{\mathcal{A},\mathcal{B}}(\mathcal{S}_2) \qquad (4.26)\\
\geq&J_{\mathcal{A}}(\mathcal{S}_1) + J_{\mathcal{B}}(\mathcal{S}_2). \qquad (4.27)
\end{aligned}$$

If $J_{\mathcal{A}}(\mathcal{S}_1) \geq 0$, $\forall\mathcal{S}_1\subseteq\mathcal{A}$, and $J_{\mathcal{B}}(\mathcal{S}_2) \geq 0$, $\forall\mathcal{S}_2\subseteq\mathcal{B}$, then following (4.27), $J_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}) \geq 0$, $\forall\mathcal{S}\subseteq\mathcal{A}\cup\mathcal{B}$. If $J_{\mathcal{A}}(\mathcal{S}_1) \geq 0$, $\forall\mathcal{S}_1\subseteq\mathcal{A}$, and $J_{\mathcal{A},\mathcal{B}}(\mathcal{S}_2) \geq 0$, $\forall\mathcal{S}_2\subseteq\mathcal{B}$, then following (4.26), $J_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}) \geq 0$, $\forall\mathcal{S}\subseteq\mathcal{A}\cup\mathcal{B}$. □

*Proof of Lemma 4.3.2.* Let $\mathcal{L} := \{\mathcal{F} \subseteq \mathcal{N} : J_{\mathcal{F}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{F}\}$ and $\mathcal{L}_{\max} := \{\mathcal{D} \in \mathcal{L} : |\mathcal{D}| = \max_{\mathcal{F}\in\mathcal{L}} |\mathcal{F}|\}$. Suppose there are more than one elements in $\mathcal{L}_{\max}$, say, $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_n$, where $n \geq 2$. Then based on 1) of Lemma 4.3.1, $\mathcal{D} := \bigcup_{i=1}^{n} \mathcal{D}_i$ also satisfies that $J_{\mathcal{D}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{D}$, which is in contradiction, and hence Lemma 4.3.2 is proved. □

*Proof of Lemma 4.3.3.* If $J_{\mathcal{A},\mathcal{B}}(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{B}$, then this lemma obviously holds. Otherwise, if there exists some $\mathcal{S}_1 \subseteq \mathcal{B}$, $\mathcal{S}_1 \neq \mathcal{B}$, such that $J_{\mathcal{A},\mathcal{B}}(\mathcal{S}_1) < 0$, then we have $J_{\mathcal{A},\mathcal{B}}(\mathcal{B}) - J_{\mathcal{A},\mathcal{B}}(\mathcal{S}_1) \geq 0$, i.e.,

$$\begin{aligned}
&I(X_{\mathcal{B}}; \hat{Y}_{\mathcal{A}}, Y_{n+1}|X_{\mathcal{A}}) - I(Y_{\mathcal{B}}; \hat{Y}_{\mathcal{B}}|X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}}, Y_{n+1}, X_{\mathcal{B}}) \\
&- I(X_{\mathcal{S}_1}; \hat{Y}_{\mathcal{B}\setminus\mathcal{S}_1}, \hat{Y}_{\mathcal{A}}, Y_{n+1}|X_{\mathcal{A}}, X_{\mathcal{B}\setminus\mathcal{S}_1}) + I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1}|X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}}, Y_{n+1}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{B}\setminus\mathcal{S}_1}) \\
=&I(X_{\mathcal{B}\setminus\mathcal{S}_1}; \hat{Y}_{\mathcal{A}}, Y_{n+1}|X_{\mathcal{A}}) + I(X_{\mathcal{S}_1}; \hat{Y}_{\mathcal{A}}, Y_{n+1}|X_{\mathcal{A}}, X_{\mathcal{B}\setminus\mathcal{S}_1}) \\
&- I(Y_{\mathcal{B}\setminus\mathcal{S}_1}; \hat{Y}_{\mathcal{B}\setminus\mathcal{S}_1}|X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}}, Y_{n+1}, X_{\mathcal{B}}) - I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1}|X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}}, Y_{n+1}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{B}\setminus\mathcal{S}_1}) \\
&- I(X_{\mathcal{S}_1}; \hat{Y}_{\mathcal{A}}, Y_{n+1}|X_{\mathcal{A}}, X_{\mathcal{B}\setminus\mathcal{S}_1}) - I(X_{\mathcal{S}_1}; \hat{Y}_{\mathcal{B}\setminus\mathcal{S}_1}|\hat{Y}_{\mathcal{A}}, Y_{n+1}, X_{\mathcal{A}}, X_{\mathcal{B}\setminus\mathcal{S}_1}) \\
&+ I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1}|X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}}, Y_{n+1}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{B}\setminus\mathcal{S}_1}) \\
=&I(X_{\mathcal{B}\setminus\mathcal{S}_1}; \hat{Y}_{\mathcal{A}}, Y_{n+1}|X_{\mathcal{A}}) - I(Y_{\mathcal{B}\setminus\mathcal{S}_1}; \hat{Y}_{\mathcal{B}\setminus\mathcal{S}_1}|\hat{Y}_{\mathcal{A}}, Y_{n+1}, X_{\mathcal{A}}, X_{\mathcal{B}\setminus\mathcal{S}_1}) \\
=&J_{\mathcal{A},\mathcal{B}\setminus\mathcal{S}_1}(\mathcal{B} \setminus \mathcal{S}_1) \\
\geq&0.
\end{aligned}$$

Now, we arrive at the same situation as in the original assumption with $\mathcal{B}$ replaced by $\mathcal{B} \setminus \mathcal{S}_1$. Continue applying this argument, and we must be able to reach a nonempty $\mathcal{C} \subseteq \mathcal{B}$, such that $J_{\mathcal{A},\mathcal{C}}(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{C}$. □

*Proof of Lemma 4.3.4.* For any disjoint $\mathcal{A}$ and $\mathcal{B}$,

$$\begin{aligned}
&J(\mathcal{A} \circ \mathcal{B}) \\
=&J(\mathcal{A}) + J(\mathcal{B}) - J(\mathcal{A} \cup \mathcal{B}) \\
=&I(X_{\mathcal{A}}; \hat{Y}_{\mathcal{A}^c}, Y_{n+1}|X_{\mathcal{A}^c}) - I(Y_{\mathcal{A}}; \hat{Y}_{\mathcal{A}}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{A}^c}) \\
&+ I(X_{\mathcal{B}}; \hat{Y}_{\mathcal{B}^c}, Y_{n+1}|X_{\mathcal{B}^c}) - I(Y_{\mathcal{B}}; \hat{Y}_{\mathcal{B}}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{B}^c}) \\
&- I(X_{\mathcal{B}}; \hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c}, Y_{n+1}|X_{(\mathcal{A}\cup\mathcal{B})^c}) - I(X_{\mathcal{A}}; \hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c}, Y_{n+1}|X_{\mathcal{A}^c}) \\
&+ I(Y_{\mathcal{A}}; \hat{Y}_{\mathcal{A}}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c}) + I(Y_{\mathcal{B}}; \hat{Y}_{\mathcal{B}}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{\mathcal{B}^c}) \\
=&I(X_{\mathcal{A}}; \hat{Y}_{\mathcal{B}}|X_{\mathcal{A}^c}, \hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c}, Y_{n+1}) + I(X_{\mathcal{B}}; X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}}|X_{(\mathcal{A}\cup\mathcal{B})^c}, \hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c}, Y_{n+1}) \\
&+ I(\hat{Y}_{\mathcal{A}}; \hat{Y}_{\mathcal{B}}|X_{\mathcal{N}}, Y_{n+1}, \hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c}) \\
=&I(X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}}; \hat{Y}_{\mathcal{B}}|X_{\mathcal{A}^c}, \hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c}, Y_{n+1}) + I(X_{\mathcal{B}}; X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}}|X_{(\mathcal{A}\cup\mathcal{B})^c}, \hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c}, Y_{n+1}) \\
=&I(X_{\mathcal{B}}, \hat{Y}_{\mathcal{B}}; X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}}|X_{(\mathcal{A}\cup\mathcal{B})^c}, \hat{Y}_{(\mathcal{A}\cup\mathcal{B})^c}, Y_{n+1}),
\end{aligned}$$

which proves the lemma. □

*Proof of Lemma 4.4.4.* For any disjoint $\mathcal{A}$ and $\mathcal{B}$, and any $\mathcal{S} \subseteq \mathcal{A} \cup \mathcal{B}$, let $\mathcal{S}_1 = \mathcal{S}\mathcal{A}$ and $\mathcal{S}_2 = \mathcal{S}\mathcal{B}$. Then, we have

$$\begin{aligned}
& R_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}) \\
=& I(X_0, X_{\mathcal{S}}; \hat{Y}_{(\mathcal{A}\cup\mathcal{B})\backslash\mathcal{S}}, Y_{n+1}|X_{(\mathcal{A}\cup\mathcal{B})\backslash\mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_0, X_{\mathcal{A}\cup\mathcal{B}}, \hat{Y}_{(\mathcal{A}\cup\mathcal{B})\backslash\mathcal{S}}, Y_{n+1}) \\
=& I(X_0, X_{\mathcal{S}_1}, X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1}|X_{\mathcal{A}\backslash\mathcal{S}_1}, X_{\mathcal{B}\backslash\mathcal{S}_2}) \\
& - I(Y_{(\mathcal{S}_1\cup\mathcal{S}_2)}; \hat{Y}_{(\mathcal{S}_1\cup\mathcal{S}_2)}|X_0, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1}) \\
=& I(X_0, X_{\mathcal{S}_1}; \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1}|X_{\mathcal{A}\backslash\mathcal{S}_1}, X_{\mathcal{B}\backslash\mathcal{S}_2}) \\
& + I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1}|X_0, X_{\mathcal{A}}, X_{\mathcal{B}\backslash\mathcal{S}_2}) \\
& - [I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1}|X_0, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1}) \\
& + I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2}|X_0, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1})] \\
=& I(X_0, X_{\mathcal{S}_1}; \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1}|X_{\mathcal{A}\backslash\mathcal{S}_1}) + I(X_0, X_{\mathcal{S}_1}; X_{\mathcal{B}\backslash\mathcal{S}_2}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}|X_{\mathcal{A}\backslash\mathcal{S}_1}, \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1}) \\
& - [I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1}|X_0, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1}) - I(X_{\mathcal{B}}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_1}|X_0, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1})] \\
& + I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1}|X_0, X_{\mathcal{A}}, X_{\mathcal{B}\backslash\mathcal{S}_2}) \\
& - I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2}|X_0, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1}) \\
=& [I(X_0, X_{\mathcal{S}_1}; \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1}|X_{\mathcal{A}\backslash\mathcal{S}_1}) - I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1}|X_0, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1})] \\
& + I(X_0, X_{\mathcal{S}_1}; X_{\mathcal{B}\backslash\mathcal{S}_2}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}|X_{\mathcal{A}\backslash\mathcal{S}_1}, \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1}) \\
& + I(X_{\mathcal{B}}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_1}|X_0, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1}) \\
& + I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1}|X_0, X_{\mathcal{A}}, X_{\mathcal{B}\backslash\mathcal{S}_2}) \\
& - I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2}|X_0, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1}) \\
=& R_{\mathcal{A}}(\mathcal{S}_1) + I(X_0, X_{\mathcal{S}_1}; X_{\mathcal{B}\backslash\mathcal{S}_2}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}|X_{\mathcal{A}\backslash\mathcal{S}_1}, \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1}) \\
& + I(X_{\mathcal{B}}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_1}|X_0, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1}) \\
& + I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1}|X_0, X_{\mathcal{A}}, X_{\mathcal{B}\backslash\mathcal{S}_2}) \\
& - I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2}|X_0, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1}). \qquad (4.28)
\end{aligned}$$

When $\mathcal{S}_2 = \mathcal{B}$, following (4.28), we have

$$\begin{aligned}
R_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}) =& R_{\mathcal{A}}(\mathcal{S}_1) + I(X_{\mathcal{B}}; \hat{Y}_{\mathcal{S}_1}|X_0, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1}) \\
& + I(X_{\mathcal{B}}; \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1}|X_0, X_{\mathcal{A}}) - I(Y_{\mathcal{B}}; \hat{Y}_{\mathcal{B}}|X_0, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, Y_{n+1}) \\
=& R_{\mathcal{A}}(\mathcal{S}_1) + I(X_{\mathcal{B}}; \hat{Y}_{\mathcal{A}}, Y_{n+1}|X_0, X_{\mathcal{A}}) - I(Y_{\mathcal{B}}; \hat{Y}_{\mathcal{B}}|X_0, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, Y_{n+1}) \\
=& R_{\mathcal{A}}(\mathcal{S}_1) + K_{\mathcal{A},\mathcal{B}}(\mathcal{B}).
\end{aligned}$$

Generally, for any $\mathcal{S}_2 \subseteq \mathcal{B}$, continuing (4.28), we have

$$
\begin{aligned}
&R_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}) \\
\geq &R_{\mathcal{A}}(\mathcal{S}_1) + I(X_{\mathcal{B}}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_1} | X_0, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1}) \\
&+ I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1} | X_0, X_{\mathcal{A}}, X_{\mathcal{B}\backslash\mathcal{S}_2}) \\
&- I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X_0, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1}) \\
= &R_{\mathcal{A}}(\mathcal{S}_1) + I(X_{\mathcal{B}\backslash\mathcal{S}_2}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_1} | X_0, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1}) \\
&+ I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_1} | X_0, X_{\mathcal{A}}, X_{\mathcal{B}\backslash\mathcal{S}_2}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1}) \\
&+ I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1} | X_0, X_{\mathcal{A}}, X_{\mathcal{B}\backslash\mathcal{S}_2}) \\
&- I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X_0, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1}) \\
= &R_{\mathcal{A}}(\mathcal{S}_1) + I(X_{\mathcal{B}\backslash\mathcal{S}_2}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_1} | X_0, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}\backslash\mathcal{S}_1}, Y_{n+1}) \\
&+ I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1} | X_0, X_{\mathcal{A}}, X_{\mathcal{B}\backslash\mathcal{S}_2}) - I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X_0, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1}) \\
\geq &R_{\mathcal{A}}(\mathcal{S}_1) + I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1} | X_0, X_{\mathcal{A}}, X_{\mathcal{B}\backslash\mathcal{S}_2}) \\
&- I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X_0, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B}\backslash\mathcal{S}_2}, Y_{n+1}) \\
= &R_{\mathcal{A}}(\mathcal{S}_1) + K_{\mathcal{A}\cup\mathcal{B}}(\mathcal{S}_2).
\end{aligned}
$$

This completes the proof of Lemma 4.4.4. □

# Chapter 5

# A Unified Relay Framework with Both D-F and C-F Relay Nodes

## 5.1 Introduction

So far, we have discussed the generalization of D-F and C-F to multiple-relay channels, with the best known D-F ([27]-[30]) and C-F ([32]-[34])[1] rates summarized in Theorem 2.3.3 and in Theorems 4.2.1, 4.2.2, 4.2.4, respectively. In these discussions, all the relay nodes in the network perform only one type of relay strategy, either D-F or C-F. However, to obtain higher achievable rate, it might be better to let each relay node choose from D-F and C-F its relay strategy depending on the channel condition, e.g., let the relay node close to the source perform D-F while let the relay node close to the destination perform C-F. This invokes a unified relay framework that includes both the D-F and C-F relay nodes in the network. In developing such a framework, one naturally wants to combine the advantages of both the best known D-F and C-F schemes, i.e., the multi-level D-F schemes in [27]-[30] and the recent advances on C-F schemes [32]-[34] that have been discussed in last chapter.

Some attempts towards this unified relay framework have been made in [30, Thm 4], [41], [42]. The work [30], however, had been done before noisy network coding was proposed, and thus the recent progress on C-F schemes was not reflected in it.

[1] Part of the results in [33]-[34] have also been recognized in [38]-[40].

Moreover, in the scheme of [30, Thm 4], the relay nodes are not fully cooperating in the sense that the D-F relay nodes do not utilize the help of the C-F relay nodes. [41] improves [30, Thm 4] by combing noisy network coding to their scheme, but still does not allow the D-F relay nodes exploit the help of the C-F relay nodes. In [42], the authors incorporate noisy network coding to their scheme and let the D-F relay nodes exploit the help of the C-F relay nodes. Nevertheless, [42] does not use the multi-level D-F schemes as in [27]-[30]. Instead, all the D-F relay nodes in [42] are at the same level, and thus the decoding at each D-F relay node cannot exploit the help of other D-F relay nodes. Besides, in [42], although the destination performs backward decoding to fully exploit the help of the C-F relay nodes, the decoding at each D-F relay node is based on two consecutive blocks only and thus does not fully utilize the help of the C-F relay nodes as in [32]-[34].

Indeed, it turns out that, to incorporate in full the advantages of both the best known D-F and C-F relay strategies into a unified framework is nontrivial due to the following major challenge: For the D-F relay nodes to fully utilize the help of the C-F relay nodes as in [32]-[34], decoding at the D-F relay nodes should not be conducted until all the blocks have been finished; however, to perform the multi-level D-F strategy as in [27]-[30], the upstream nodes have to decode prior to the downstream nodes in order to help, which makes simultaneous decoding at all the D-F relay nodes after all the blocks have been finished inapplicable.

To tackle this problem, nested blocks ([29]-[30], [43]) combined with backward decoding are used in our framework, so that the D-F relay nodes at different levels can perform backward decoding at different frequencies: the closer to the source in the information passing route, the higher decoding frequency. As such, the upstream D-F relay nodes can decode before the downstream D-F relay nodes and the destination, and the use of backward decoding at each D-F relay node ensures the full exploitation of the help of both the other D-F relay nodes and the C-F relay nodes.

Specifically, we partition the relay nodes set $\mathcal{N}$ into two sets, $\mathcal{M}$ with $|\mathcal{M}| = M$

and $\mathcal{N}\backslash\mathcal{M}$, as depicted in Figure 5.1, and fix some permutation $\pi$ on $\{0\}\bigcup\mathcal{M}\bigcup\{n+1\}$ with $\pi(1)=0$ and $\pi(M+2)=n+1$. Let the relay nodes in $\mathcal{M}$ perform the multi-level D-F cooperatively along the route $\pi(1)\to\pi(2)\to\cdots\to\pi(M+2)$, while let each node $i\in\mathcal{N}\setminus\mathcal{M}$ perform C-F as in [32]-[34] independently. Then, a total of $B^{M+1}$ blocks will be used and the length of a "virtual" block for node $\pi(k), k=2,3,\ldots,M+2$, will be $B^{k-2}$ blocks. The backward decoding at the destination, i.e., node $\pi(M+2)$, will happen at the end of all $B^{M+1}$ blocks, while the backward decoding at the D-F relay node $\pi(k), k=2,3,\ldots,M+1$, will happen whenever it has received $B$ new "virtual" blocks, i.e., at the end of each block $b=vB^{k-1}, v\in[1:B^{M+1}/B^{k-1}]$. Also, both the D-F relay nodes and the destination will perform compression-message joint decoding, which is in general necessary since the compressions of the C-F relay nodes may not be chosen to support successive decoding at all the D-F relay nodes and the destination.

Under the above described framework, for any given distribution

$$p(x_0)p(x_{\mathcal{M}}|x_0)\prod_{i\in\mathcal{N}\backslash\mathcal{M}}p(x_i)p(\hat{y}_i|y_i,x_i),$$

the following rate is achievable:

$$\begin{aligned}R<\min_{2\le k\le M+2}\min_{\mathcal{S}\subseteq\mathcal{D}_k}&I(X_{\pi(1:k-1)},X_{\mathcal{S}};\hat{Y}_{\mathcal{D}_k\backslash\mathcal{S}},Y_{\pi(k)}|X_{\mathcal{D}_k\backslash\mathcal{S}},X_{\pi(k:M+1)})\\&-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_{\pi(1:M+1)},X_{\mathcal{D}_k},Y_{\pi(k)},\hat{Y}_{\mathcal{D}_k\backslash\mathcal{S}}),\end{aligned}\tag{5.1}$$

where $\mathcal{D}_k$ is the unique largest subset of $\mathcal{N}\setminus\mathcal{M}$ satisfying

$$I(X_{\mathcal{S}};\hat{Y}_{\mathcal{D}_k\backslash\mathcal{S}},Y_{\pi(k)}|X_{\pi(1:M+1)},X_{\mathcal{D}_k\backslash\mathcal{S}})-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_{\pi(1:M+1)},X_{\mathcal{D}_k},Y_{\pi(k)},\hat{Y}_{\mathcal{D}_k\backslash\mathcal{S}})>0,\tag{5.2}$$

for any nonempty $\mathcal{S}\subseteq\mathcal{D}_k$.

(5.1) has the flavors of both (2.16) and (4.4). Specifically, for each node $\pi(k), k=2,3,\ldots,M+2$, the corresponding rate constraint is

$$\begin{aligned}R<\min_{\mathcal{S}\subseteq\mathcal{D}_k}&I(X_{\pi(1:k-1)},X_{\mathcal{S}};\hat{Y}_{\mathcal{D}_k\backslash\mathcal{S}},Y_{\pi(k)}|X_{\mathcal{D}_k\backslash\mathcal{S}},X_{\pi(k:M+1)})\\&-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_{\pi(1:M+1)},X_{\mathcal{D}_k},Y_{\pi(k)},\hat{Y}_{\mathcal{D}_k\backslash\mathcal{S}}),\end{aligned}\tag{5.3}$$

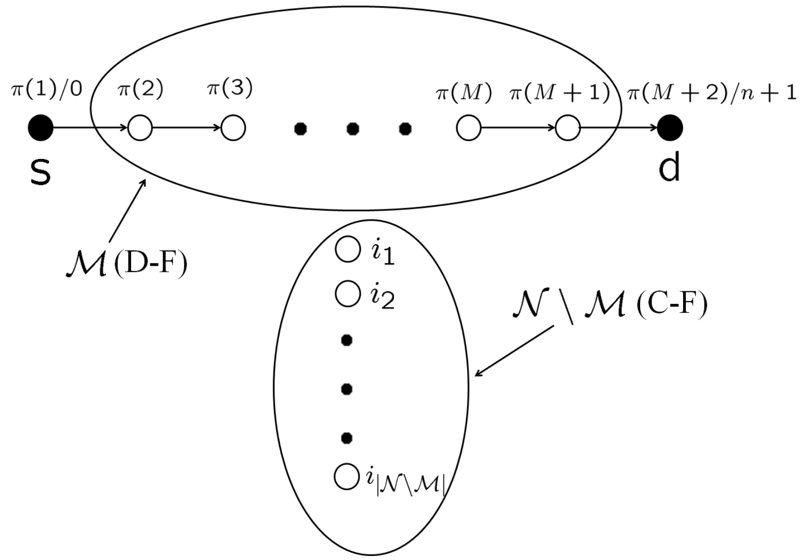



Figure 5.1: A unified relay framework with both the D-F and C-F relay nodes.

which is in a form similar to (4.4) but with the appearance of $X_{\pi(1:k-1)}$, $X_{\pi(k:M+1)}$ and $X_{\pi(1:M+1)}$. (5.3) has the similar form as (4.4) since node $\pi(k)$ uses the help of the C-F relay nodes as in [33]-[34]. $X_{\pi(1:k-1)}$, $X_{\pi(k:M+1)}$ and $X_{\pi(1:M+1)}$ appear in (5.3) because node $\pi(k)$ also utilizes the help of other D-F relay nodes as in [27]-[30] so that the signals of its upstream nodes, i.e., $X_{\pi(1:k-1)}$, are cooperatively providing the information while the signals of its downstream nodes and itself $X_{\pi(k:M+1)}$ are known at $\pi(k)$. Also, the set $\mathcal{D}_k$ defined in (5.2) has a similar interpretation as the set $\mathcal{D}$ defined in (4.5), i.e., the "jointly decodable" C-F relay nodes set at node $\pi(k)$ such that the compressions of the relays in this set are decodable jointly with $X_{\pi(1:k-1)}$ given that $X_{\pi(k:M+1)}$ are known at node $\pi(k)$.

It can be easily seen that (5.1) includes the achievable rates in (2.16) and (4.4) as special cases: When $\mathcal{M} = \mathcal{N}$, i.e., all the relays perform D-F, $\mathcal{D}_k = \emptyset$ and (5.1) reduces to (2.16); When $\mathcal{M} = \emptyset$, i.e., all the relays perform C-F, (5.1) reduces to (4.4).

Finally, it should be noted that, the achievable rate (5.1) is proved by using the block-by-block backward decoding scheme in [34]. We can also modify the all blocks united decoding scheme in [32] to a $B$-blocks-by-$B$-blocks backward decoding scheme, to fit it into our unified relay framework and prove the following achievable rate:

$$R < \min_{2\le k\le M+2}\max_{\mathcal{T}_k\subseteq\mathcal{N}\setminus\mathcal{M}}\min_{\mathcal{S}\subseteq\mathcal{T}_k} I(X_{\pi(1:k-1)}, X_{\mathcal{S}}; \hat{Y}_{\mathcal{T}_k\setminus\mathcal{S}}, Y_{\pi(k)}|X_{\mathcal{T}_k\setminus\mathcal{S}}, X_{\pi(k:M+1)}) \\ - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_{\pi(1:M+1)}, X_{\mathcal{T}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{T}_k\setminus\mathcal{S}}). \tag{5.4}$$

Similarly to the equivalence between (4.4) and (4.7), here (5.1) and (5.4) are also equivalent. One can also easily check that (5.4) includes the achievable rates in (2.16) and (4.7) as special cases by letting $\mathcal{M} = \mathcal{N}$ and $\mathcal{M} = \emptyset$ respectively. Notably, again, in terms of complexity, block-by-block backward decoding is relatively easier to implement since $B$-blocks-by-$B$-blocks backward decoding involves $B$ blocks united decoding.

The remainder of the chapter is organized as the following. The main results, as well as their applications to Gaussian networks, are first summarized in Section 5.2. Then, in Section 5.3 and Section 5.4, our unified relay framework with block-by-block backward decoding and with $B$-blocks-by-$B$-blocks backward decoding will be presented in detail respectively. The details on the evaluation of the rates for the Gaussian network are finally included in Section 5.5.

## 5.2 Main Results

Before presenting the main results, we introduce some simplified notations. As in Chapters 3-4, denote the set $\mathcal{N} = \{1, 2, \ldots, n\}$. For any subset $\mathcal{S} \subseteq \{0, 1, \ldots, n+1\}$, let $X_{\mathcal{S}} = \{X_i, i \in \mathcal{S}\}$, and use similar notations for other variables. For any $\mathcal{M} \subseteq \mathcal{N}$ with $|\mathcal{M}| = M$, let $\pi(\{0, \mathcal{M}, n+1\})$ be a permutation on $\{0\}\bigcup\mathcal{M}\bigcup\{n+1\}$ with $\pi(1) = 0$ and $\pi(M+2) = n+1$, and let $\pi(k_1 : k_2) = \{\pi(k_1), \pi(k_1+1), \ldots, \pi(k_2)\}$.

Under our unified relay framework as described in the Introduction, the following Theorems 5.2.1 and 5.2.2 present the achievable rates by block-by-block backward decoding and $B$-blocks-by-$B$-blocks backward decoding respectively. The coding schemes used to prove these theorems constitute the key contributions of this chapter, and will be presented in detail in Sections 5.3 and 5.4 respectively.

**Theorem 5.2.1.** *For the multiple-relay channel, a rate $R$ is achievable if for some $\mathcal{M} \subseteq \mathcal{N}$ with $|\mathcal{M}| = M$, there exists a permutation $\pi(\{0, \mathcal{M}, n+1\})$ and some*

$$p(q)p(x_0|q)p(x_{\mathcal{M}}|x_0, q) \prod_{i \in \mathcal{N} \setminus \mathcal{M}} p(x_i|q)p(\hat{y}_i|y_i, x_i, q),$$

*such that for any $k = 2, 3, \ldots, M+2$,*

$$R < \min_{\mathcal{S} \subseteq \mathcal{D}_k} I(X_{\pi(1:k-1)}, X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_k \setminus \mathcal{S}}, Y_{\pi(k)} | X_{\mathcal{D}_k \setminus \mathcal{S}}, X_{\pi(k:M+1)}, Q) \\ - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_{\pi(1:M+1)}, X_{\mathcal{D}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{D}_k \setminus \mathcal{S}}, Q), \tag{5.5}$$

*where $\mathcal{D}_k$ is the unique largest subset of $\mathcal{N} \setminus \mathcal{M}$ satisfying*

$$I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_k \setminus \mathcal{S}}, Y_{\pi(k)} | X_{\pi(1:M+1)}, X_{\mathcal{D}_k \setminus \mathcal{S}}, Q) \\ - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_{\pi(1:M+1)}, X_{\mathcal{D}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{D}_k \setminus \mathcal{S}}, Q) > 0, \tag{5.6}$$

*for any nonempty $\mathcal{S} \subseteq \mathcal{D}_k$.*

**Theorem 5.2.2.** *For the multiple-relay channel, a rate $R$ is achievable if for some $\mathcal{M} \subseteq \mathcal{N}$ with $|\mathcal{M}| = M$, there exists a permutation $\pi(\{0, \mathcal{M}, n+1\})$ and some*

$$p(q)p(x_0|q)p(x_{\mathcal{M}}|x_0, q) \prod_{i \in \mathcal{N} \setminus \mathcal{M}} p(x_i|q)p(\hat{y}_i|y_i, x_i, q),$$

*such that for any $k = 2, 3, \ldots, M+2$,*

$$R < \max_{\mathcal{T}_k \subseteq \mathcal{N} \setminus \mathcal{M}} \min_{\mathcal{S} \subseteq \mathcal{T}_k} I(X_{\pi(1:k-1)}, X_{\mathcal{S}}; \hat{Y}_{\mathcal{T}_k \setminus \mathcal{S}}, Y_{\pi(k)} | X_{\mathcal{T}_k \setminus \mathcal{S}}, X_{\pi(k:M+1)}, Q) \\ - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_{\pi(1:M+1)}, X_{\mathcal{T}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{T}_k \setminus \mathcal{S}}, Q). \tag{5.7}$$

The following theorem establishes the equivalence between the achievable rates in Theorems 5.2.1 and 5.2.2. The proof of this theorem can be immediately obtained by analogy to the proof of Theorem 4.2.5 and will be omitted here.

**Theorem 5.2.3.** *For any $\mathcal{M} \subseteq \mathcal{N}$ with $|\mathcal{M}| = M$, any permutation $\pi(\{0, \mathcal{M}, n+1\})$, any distribution*

$$p(q)p(x_0|q)p(x_{\mathcal{M}}|x_0, q) \prod_{i \in \mathcal{N} \setminus \mathcal{M}} p(x_i|q)p(\hat{y}_i|y_i, x_i, q),$$

*and any $k = 2, 3, \ldots, M+2$, the maximum in the R.H.S. of (5.7) is attained when $\mathcal{T}_k = \mathcal{D}_k$, where $\mathcal{D}_k$ is as defined in (5.6).*

**Remark 5.2.1.** *Finally, we point out that Theorems 5.2.1 and 5.2.2 can also be applied to multiple-destination problems, by choosing the D-F relay nodes set $\mathcal{M}$ to include the other destinations.*

### 5.2.1 An Example of Gaussian Networks

We now apply the above results to additive white Gaussian noise (AWGN) networks and compare the achievable rates under our unified framework to the rates with D-F or C-F alone and with other unified schemes [30], [42]. In particular, we consider an AWGN two-relay channel with the geometry shown in Figure 5.2, where nodes 0 and 3 are the source and destination respectively, and nodes 1 and 2 are the relay nodes. This channel is described by

$$
\begin{aligned}
Y_1 &= g_{01}X_0 + g_{21}X_2 + Z_1, \\
Y_2 &= g_{02}X_0 + g_{12}X_1 + Z_2, \\
Y_3 &= g_{03}X_0 + g_{13}X_1 + g_{23}X_2 + Z_3,
\end{aligned}
$$

in which, the channel gains are $g_{03} = 1$, $g_{01} = g_{23} = d^{-\alpha/2}$, $g_{02} = g_{13} = (1-d)^{-\alpha/2}$, $g_{12} = g_{21} = (1-2d)^{-\alpha/2}$, where $\alpha$ is the path-loss exponent; the noises $Z_1, Z_2, Z_3$ are zero-mean unit-variance Gaussian random variables that are independent of each other and the channel inputs. We further assume uniform average power constraint $P$ on each transmitter $X_i, i = 0, 1, 2$, and noncoherent transmission among the transmitters, i.e., $X_0, X_1, X_2$ are independent.[2] The achievable rates with our unified schemes and various other schemes are evaluated for this two-relay channel in Section 5.5. Figure 5.3 plots these rates for $P = 10$, $\alpha = 2$, and $d \in (0, 0.5)$. As can be seen, the rates with our unified schemes are generally better than the rates with existing unified schemes and with D-F or C-F alone. Detailed analyses based on Figure 5.3 are as follows.

*D-F vs. C-F:* The D-F and C-F achievable rates plotted in Figure 5.3 are the multi-level D-F rates [27]-[30] and noisy network coding rates [32]-[34] respectively. Roughly speaking, when $d$ is relatively large, D-F performs better than C-F; when $d$ is relatively small, C-F outperforms D-F. This is because, relatively large $d$ means

[2] The assumption of noncoherent transmission is made mainly to simplify the evaluation of the achievable rates; nevertheless, it does also reflect some realistic wireless communication scenarios, e.g., the phase-fading channels with the phase information unknown to the transmitter, so that coherent beamforming cannot be achieved [30].

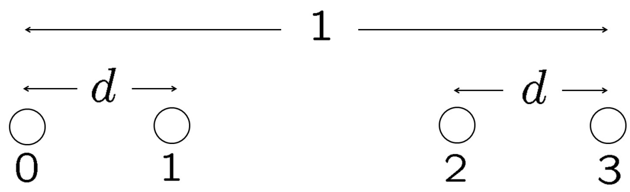
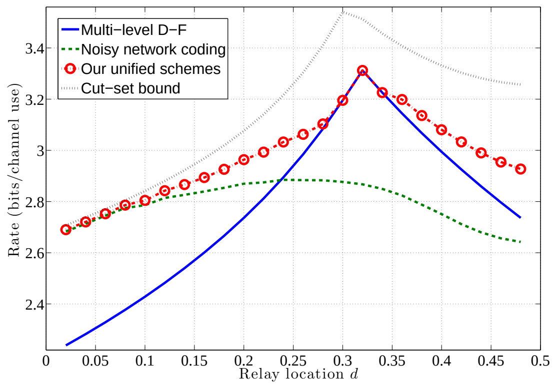


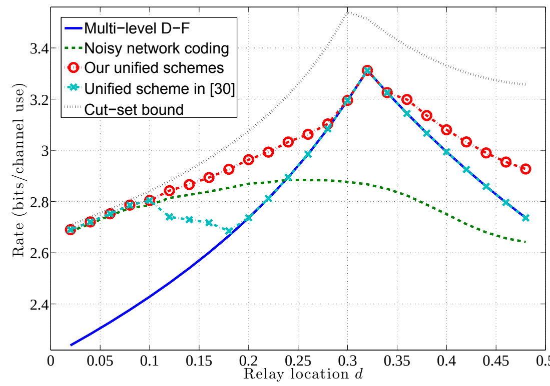

Figure 5.2: A two-relay channel.

that both relay nodes are not far from the midpoint, which facilitates the use of multi-level D-F; while when $d$ is small, i.e., when relay nodes 1 and 2 are respectively close to the source and destination, the D-F scheme still requires node 2 to be able to decode the message, thus limiting the achievable rate for the channel compared to the C-F scheme which has no requirement of decoding at either relay node. Also note here the highest D-F rate is achieved when $d$ is around 1/3, i.e., when relay nodes 1 and 2 are approximately evenly distributed along the line.

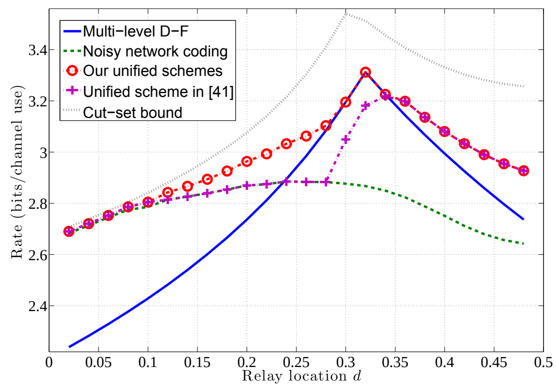
*Our unified schemes vs. D-F or C-F alone:* With multi-level D-F and noisy network coding included as their special cases, our unified schemes achieve the same rate as multi-level D-F when $d$ is around 1/3, and provide considerable gains on the achievable rates against using D-F or C-F alone as $d$ decreases or increases from around 1/3. The interpretation is that, in our unified schemes, for $d$ around 1/3, nodes 1 and 2 should perform multi-level D-F; otherwise, a better choice is to let node 1 perform D-F while let node 2 perform C-F. It is not surprising that as $d$ decreases from around 1/3, i.e., as nodes 1 and 2 move towards the source and destination respectively, introducing the freedom of choosing nodes 1 and 2 to be the D-F and C-F node respectively can improve the achievable rates, compared to performing D-F or C-F alone at both relay nodes. However, it may not be so obvious why this choice of the D-F and C-F node also leads to better achievable rates than multi-level D-F as $d$ increases from around 1/3, i.e., as relay nodes 1 and 2 move towards each other. For this, one has to realize that in our schemes, D-F node 1, when decoding the message from source node 0, also exploits the help of C-F node 2, which could result in a looser rate constraint for decoding at D-F node 1 than in multi-level D-F, thus potentially improving the achievable rate for




(a) Our schemes vs. D-F or C-F alone




(b) Our schemes vs. Unified scheme in [30]




(c) Our schemes vs. Unified scheme in [42]

Figure 5.3: Achievable rates for the two-relay channel.

the channel. In contrast, if the D-F node does not utilize the help of the C-F node, which is the case in the unified scheme of [30], then, as $d$ increases from around 1/3, this gain will not happen.

*Our unified schemes vs. unified scheme in [30]:* Generally, the unified scheme in [30] also improves D-F alone by introducing the flexibility of choosing node 2 to be C-F node, but cannot achieve the same rates as our unified schemes. The reasons are two-fold as mentioned in Section 5.1, i.e., i) it incorporates an inferior C-F scheme instead of noisy network coding, and ii) it does not allow the D-F node to utilize the help of the C-F node. For the first reason, it cannot achieve the noisy network coding rates for $d$ roughly between 0.1 and 0.24. Due to the second reason, it cannot provide any gain against using multi-level D-F as $d$ increases from around 1/3, in contrast to our unified schemes.

*Our unified schemes vs. unified scheme in [42]:* Our unified schemes also outperform the unified scheme in [42]. Recall that there are two drawbacks in the scheme of [42]. First, in the scheme of [42], all the D-F relay nodes are at the same level, and thus it cannot achieve the same rates as multi-level D-F for $d \in (0.24, 0.34)$. Second, it does not allow the D-F node to perform backward decoding to fully utilize the help of the C-F node as in our schemes, which, together with the first drawback, leads to lower achievable rates than our schemes for $d \in (0.1, 0.36)$.

## 5.3 Unified Relay Framework With Block-By-Block Backward Decoding

To prove Theorem 5.2.1, we incorporate the multi-level D-F scheme in [27]-[30] and the cumulative encoding/block-by-block backward decoding/comression-message joint decoding C-F scheme in [34] into the unified relay framework described in the Introduction.

Specifically, we divide the relay set $\mathcal{N}$ into two sets, $\mathcal{M}$ with $|\mathcal{M}| = M$ and $\mathcal{N} \setminus \mathcal{M}$, as shown in Fig. 5.1, and fix some permutation $\pi(\{0, \mathcal{M}, n+1\})$ with $\pi(1) = 0$ and $\pi(M+2) = n+1$. The source performs cumulative encoding, in the sense that a new message is encoded at the source in each new block; the nodes in $\mathcal{M}$ perform the multi-level D-F cooperatively, along the route $\pi(1) \rightarrow \pi(2) \rightarrow \cdots \rightarrow \pi(M+2)$, in a similar manner with [27]-[30]; each node $i \in \mathcal{N} \setminus \mathcal{M}$ performs C-F independently in the same way as [32]-[34]; both the D-F relay nodes and the destination node, i.e., nodes $\pi(2 : M+2)$, perform compression-message joint decoding in a block-by-block backward manner. (Note here, the nodes $\pi(2 : M+2)$ will be treated as multiple destinations with respect to the C-F relay nodes, and thus compression-message joint decoding is generally necessary at the these nodes, as mentioned in the Introduction.) A total of $B^{M+1}$ blocks will be used and the length of a "virtual" block for node $\pi(k), k = 2, 3 \ldots, M+2$, will be $B^{k-2}$ blocks. The backward decoding at the destination, i.e., node $\pi(M+2)$, will happen at the end of all $B^{M+1}$ blocks, while the backward decoding at the D-F relay node

$\pi(k), k = 2, 3, \ldots, M+1$, will happen at the end of every $B^{k-1}$ blocks, i.e., at the end of block $b = vB^{k-1}, v \in [1 : B^{M+1}/B^{k-1}]$.

To make the presentation of the detailed coding scheme easier to follow, we first consider the case of single D-F relay node, i.e., when $M = 1$, and then present the extension to the general case of multiple D-F relay nodes, i.e., when $M \geq 2$.

### 5.3.1 Single D-F Relay Node ($M = 1$)

Assume that, among the relay nodes set $\mathcal{N}$, only node 1 is the D-F relay node, and all other relay nodes are the C-F relay nodes. Denote $\tilde{\mathcal{N}} = \mathcal{N} \setminus \{1\}$. Specializing Theorem 5.2.1 to this case, we have that a rate $R$ is achievable, if there exists some

$$p(q)p(x_0|q)p(x_1|x_0, q) \prod_{i \in \tilde{\mathcal{N}}} p(x_i|q)p(\hat{y}_i|y_i, x_i, q),$$

such that

$$R < \min \left\{ \begin{array}{l} \min_{\mathcal{S} \subseteq \mathcal{D}_1} I(X_0, X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}}, Y_1 | X_1, X_{\mathcal{D}_1 \setminus \mathcal{S}}, Q) \\ \qquad\qquad - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\mathcal{D}_1}, Y_1, \hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}}, Q) \\ \min_{\mathcal{S} \subseteq \mathcal{D}_{n+2}} I(X_0, X_1, X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_{n+2} \setminus \mathcal{S}}, Y_{n+2} | X_{\mathcal{D}_{n+2} \setminus \mathcal{S}}, Q) \\ \qquad\qquad - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\mathcal{D}_{n+2}}, Y_{n+2}, \hat{Y}_{\mathcal{D}_{n+2} \setminus \mathcal{S}}, Q) \end{array} \right\} \tag{5.8}$$

where $\mathcal{D}_1$ is the unique largest subset of $\tilde{\mathcal{N}}$ satisfying

$$I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}}, Y_1 | X_0, X_1, X_{\mathcal{D}_1 \setminus \mathcal{S}}, Q) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\mathcal{D}_1}, Y_1, \hat{Y}_{\mathcal{D}_1 \setminus \mathcal{S}}, Q) > 0, \tag{5.9}$$

for any nonempty $\mathcal{S} \subseteq \mathcal{D}_1$, and $\mathcal{D}_{n+2}$ is the unique largest subset of $\tilde{\mathcal{N}}$ satisfying

$$\begin{aligned} & I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_{n+2} \setminus \mathcal{S}}, Y_{n+2} | X_0, X_1, X_{\mathcal{D}_{n+2} \setminus \mathcal{S}}, Q) \\ & \qquad - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\mathcal{D}_{n+2}}, Y_{n+2}, \hat{Y}_{\mathcal{D}_{n+2} \setminus \mathcal{S}}, Q) > 0, \end{aligned} \tag{5.10}$$

for any nonempty $\mathcal{S} \subseteq \mathcal{D}_{n+2}$.

The uniqueness of $\mathcal{D}_1$ and $\mathcal{D}_{n+2}$ can be immediately obtained by analogy to the proof of Theorem 4.2.4. Below, we focus on proving the achievablity of the rate in

(5.8). For simplicity of notation, we only prove the achievability for the case $Q = \emptyset$. Achievability for an arbitrary time-sharing random variable $Q$ can be obtained by using the standard technique of time sharing [44], [22]. The same consideration on $Q$ applies throughout all the proofs of this paper.

In the case of single D-F relay node, a total of $B^2$ blocks will be used; the backward decoding at the destination node $n+2$ will happen at the end of all $B^2$ blocks, while the backward decoding at the D-F relay node 1 will happen at the end of every $B$ blocks, i.e., at the end of block $b = vB, v \in [1 : B]$. See Figure 5.4 for an illustration. Note here, in order to fully utilize the help of the C-F nodes as in [32]-[34], even the only D-F relay node 1, has to perform backward decoding, which is different from the situation arising in [29]-[30] and [43], where there is no issue of exploiting the help of the C-F nodes and node 1 can decode at the end of every block. The detailed codebook generation and encoding/decoding process are as follows, which can be understood with the help of Table 5.1.

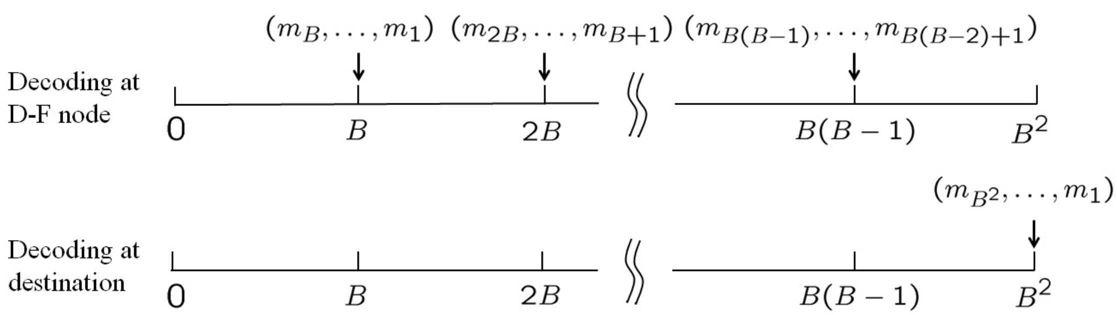



Figure 5.4: An illustration of nested blocks with backward decoding.

*Codebook Generation:* Fix $p(x_0)p(x_1|x_0)\prod_{i\in\tilde{\mathcal{N}}} p(x_i)p(\hat{y}_i|y_i, x_i)$. We randomly and independently generate a codebook for each block.

i) First consider the codebook generation for the source node 0 and the D-F relay node 1. A joint codebook for these two nodes will be generated in a backward manner similar to [28] for each block. Specifically, for each block $b \in [1 : B^2]$, randomly generate $2^{TR}$ independent sequences $\mathbf{x}_{1,b}(m_{b-B})$ for node 1, and randomly generate $2^{TR}$ conditionally independent sequences $\mathbf{x}_{0,b}(m_b|m_{b-B})$ for node 0, where $m_b, m_{b-B} \in [1 : 2^{TR}]$. As in [28], the codebook is generated in the backward manner because the source node 0 knows what the D-F relay node 1 is going to transmit, and therefore can adjust its own transmission accordingly, but not the converse. The difference from [28] is that here the delay between the messages transmitted by node 1 and node 0 is $B$ blocks, instead of 1 block in [28], since in our framework node 1 has to wait for every $B$ blocks to perform backward decoding for exploiting the help of the C-F relay nodes.

ii) Then we generate the codebooks for the C-F relay nodes in the same way as in [32]-[34]. For each block $b \in [1 : B^2]$ and each relay node $i \in \tilde{\mathcal{N}}$, randomly and independently generate $2^{T\hat{R}_i}$ sequences $\mathbf{x}_{i,b}(l_{i,b-1})$, $l_{i,b-1} \in [1 : 2^{T\hat{R}_i}]$, where $\hat{R}_i = I(Y_i; \hat{Y}_i | X_i) + \epsilon$; for each relay node $i \in \tilde{\mathcal{N}}$ and each $\mathbf{x}_{i,b}(l_{i,b-1})$, $l_{i,b-1} \in [1 : 2^{T\hat{R}_i}]$, randomly and conditionally independently generate $2^{T\hat{R}_i}$ sequences $\hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1})$, $l_{i,b} \in [1 : 2^{T\hat{R}_i}]$.

The combination of i) and ii) defines the codebook for any block $b \in [1 : B^2]$,

$$\mathcal{C}_b = \Big\{ \mathbf{x}_{1,b}(m_{b-B}), \mathbf{x}_{0,b}(m_b|m_{b-B}) : m_b, m_{b-B} \in [1 : 2^{TR}]; \\ \mathbf{x}_{i,b}(l_{i,b-1}), \hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1}) : l_{i,b}, l_{i,b-1} \in [1 : 2^{T\hat{R}_i}], i \in \tilde{\mathcal{N}} \Big\}. \tag{5.11}$$

*Encoding:* Let $\mathbf{m} = (m_1, m_2, \ldots, m_{B^2})$ be the message vector to be sent and let $m_b = 1$ be the dummy message for any

$$b \in \cup_{w=1}^{B} [wB - L + 1 : wB] \bigcup [(B-1)B + 1 : B^2] \tag{5.12}$$

and for any $b \leq 0$. As we will see, these dummy messages are inserted to ensure the start of block-by-block backward decoding. Due to these dummy messages, the actually achievable rate becomes $\frac{(B-L)(B-1)}{B^2} R$, which, however, can be made arbitrarily close to $R$ by choosing $L \ll B$, i.e., the rate loss $\frac{B(L+1)-L}{B^2} R$ can always be made arbitrarily small.

Table 5.1: Block-by-Block backward decoding for the single D-F relay node case

| Block | 1 | 2 | $\cdots$ | $B-L$ | $B-L+1$ | $\cdots$ | $B$ | $\cdots$ |
|---|---|---|---|---|---|---|---|---|
| $X_0$ | $\mathbf{x}_{0,1}(m_1\|1)$ | $\mathbf{x}_{0,2}(m_2\|1)$ | $\cdots$ | $\mathbf{x}_{0,B-L}(m_{B-L}\|1)$ | $\mathbf{x}_{0,B-L+1}(1\|1)$ | $\cdots$ | $\mathbf{x}_{0,B}(1\|1)$ | $\cdots$ |
| $Y_1$ | $\emptyset$ | $\emptyset$ | $\cdots$ | $\emptyset$ | $\emptyset$ | $\cdots$ | $(m_1, m_2, \ldots, m_B)$ | $\cdots$ |
| $X_1$ | $\mathbf{x}_{1,1}(1)$ | $\mathbf{x}_{1,2}(1)$ | $\cdots$ | $\mathbf{x}_{1,B-L}(1)$ | $\mathbf{x}_{1,B-L+1}(1)$ | $\cdots$ | $\mathbf{x}_{1,B}(1)$ | $\cdots$ |
| $Y_{\tilde{\mathcal{N}}}$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},1}(l_{\tilde{\mathcal{N}},1}\|\mathbf{1})$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},2}(l_{\tilde{\mathcal{N}},2}\|l_{\tilde{\mathcal{N}},1})$ | $\cdots$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B-L}(l_{\tilde{\mathcal{N}},B-L}\|l_{\tilde{\mathcal{N}},B-L-1})$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B-L+1}(l_{\tilde{\mathcal{N}},B-L+1}\|l_{\tilde{\mathcal{N}},B-L})$ | $\cdots$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B}(l_{\tilde{\mathcal{N}},B}\|l_{\tilde{\mathcal{N}},B-1})$ | $\cdots$ |
| $X_{\tilde{\mathcal{N}}}$ | $\mathbf{x}_{\tilde{\mathcal{N}},1}(\mathbf{1})$ | $\mathbf{x}_{\tilde{\mathcal{N}},2}(l_{\tilde{\mathcal{N}},1})$ | $\cdots$ | $\mathbf{x}_{\tilde{\mathcal{N}},B-L}(l_{\tilde{\mathcal{N}},B-L-1})$ | $\mathbf{x}_{\tilde{\mathcal{N}},B-L+1}(l_{\tilde{\mathcal{N}},B-L})$ | $\cdots$ | $\mathbf{x}_{\tilde{\mathcal{N}},B}(l_{\tilde{\mathcal{N}},B-1})$ | $\cdots$ |
| $Y_{n+2}$ | $\emptyset$ | $\emptyset$ | $\cdots$ | $\emptyset$ | $\emptyset$ | $\cdots$ | $\emptyset$ | $\cdots$ |

| Block | $B^2-B+1$ | $\cdots$ | $B^2-L$ | $B^2-L+1$ | $\cdots$ | $B^2$ |
|---|---|---|---|---|---|---|
| $X_0$ | $\mathbf{x}_{0,B^2-B+1}(1\|m_{B^2-2B+1})$ | $\cdots$ | $\mathbf{x}_{0,B^2-L}(1\|m_{B^2-B-L})$ | $\mathbf{x}_{0,B^2-L+1}(1\|1)$ | $\cdots$ | $\mathbf{x}_{0,B^2}(1\|1)$ |
| $Y_1$ | $\emptyset$ | $\cdots$ | $\emptyset$ | $\emptyset$ | $\cdots$ | $(m_{B^2-B+1}, \ldots, m_{B^2})$ |
| $X_1$ | $\mathbf{x}_{1,B^2-B+1}(m_{B^2-2B+1})$ | $\cdots$ | $\mathbf{x}_{1,B^2-L}(m_{B^2-B-L})$ | $\mathbf{x}_{1,B^2-L+1}(1)$ | $\cdots$ | $\mathbf{x}_{1,B^2}(1)$ |
| $Y_{\tilde{\mathcal{N}}}$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2-B+1}(l_{\tilde{\mathcal{N}},B^2-B+1}\|l_{\tilde{\mathcal{N}},B^2-B})$ | $\cdots$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2-L}(l_{\tilde{\mathcal{N}},B^2-L}\|l_{\tilde{\mathcal{N}},B^2-L-1})$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2-L+1}(l_{\tilde{\mathcal{N}},B^2-L+1}\|l_{\tilde{\mathcal{N}},B^2-L})$ | $\cdots$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2}(l_{\tilde{\mathcal{N}},B^2}\|l_{\tilde{\mathcal{N}},B^2-1})$ |
| $X_{\tilde{\mathcal{N}}}$ | $\mathbf{x}_{\tilde{\mathcal{N}},B^2-B+1}(l_{\tilde{\mathcal{N}},B^2-B})$ | $\cdots$ | $\mathbf{x}_{\tilde{\mathcal{N}},B^2-L}(l_{\tilde{\mathcal{N}},B^2-L-1})$ | $\mathbf{x}_{\tilde{\mathcal{N}},B^2-L+1}(l_{\tilde{\mathcal{N}},B^2-L})$ | $\cdots$ | $\mathbf{x}_{\tilde{\mathcal{N}},B^2}(l_{\tilde{\mathcal{N}},B^2-1})$ |
| $Y_{n+2}$ | $\emptyset$ | $\cdots$ | $\emptyset$ | $\emptyset$ | $\cdots$ | $(m_1, m_2, \ldots, m_{B^2})$ |

i) First consider the encoding process for nodes 0 and 1.

- In block $b \in [1 : B^2]$, the source node 0 transmits $\mathbf{x}_{0,b}(m_b|m_{b-B})$.
- At the end of block $vB, v \in [1 : B-1]$, the D-F relay node 1 has decoded messages

$$(m_{vB-B+1}, m_{vB-B+2}, \ldots, m_{vB})$$

using backward decoding (see the decoding part). In the next $B$ blocks, i.e., in block $b \in [vB+1 : (v+1)B]$, the relay node 1 transmits $\mathbf{x}_{1,b}(m_{b-B})$, where $m_{b-B}$ for any $b \in [vB+1 : (v+1)B]$) has been decoded by block $vB$.

ii) For any block $b \in [1 : B^2]$, each relay node $i \in \tilde{\mathcal{N}}$, upon receiving $\mathbf{y}_{i,b}$ at the end of block $b$, finds an index $l_{i,b}$ such that

$$(\mathbf{x}_{i,b}(l_{i,b-1}), \mathbf{y}_{i,b}, \hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1})) \in A_\epsilon(X_i, Y_i, \hat{Y}_i),$$

where $l_{i,0} = 1$ by convention. In block $b \in [1 : B^2]$, the relay node $i \in \tilde{\mathcal{N}}$ transmits $\mathbf{x}_{i,b}(l_{i,b-1})$.

*Decoding:* We present the decoding process at the D-F relay node 1 and at the destination node $n+2$ separately.

i) At the end of block $b = vB, v \in [1 : B]$, the D-F relay node 1 decodes messages

$$(m_{b-B+1}, m_{b-B+2}, \ldots, m_b)$$

using block-by-block backward decoding. In fact, among these messages,

$$(m_{b-L+1}, m_{b-L+2}, \ldots, m_b)$$

are dummy messages according to (5.12) and only

$$(m_{b-B+1}, m_{b-B+2}, \ldots, m_{b-L})$$

need decoding.

- a) Node 1 first finds the unique $l_{\mathcal{D}_1,b-L} = \{l_{i,b-L}, i \in \mathcal{D}_1\}$ such that there exists some $l^b_{\mathcal{D}_1,b-L+1}$ satisfying that for any block $j = b-L+1, b-L+2, \ldots, b$,

$$(\mathbf{X}_{0,j}(m_j|m_{j-B}), \mathbf{X}_{1,j}(m_{j-B}), \{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{D}_1\}, \mathbf{Y}_{1,j}) \in A_\epsilon(X_0, X_1, X_{\mathcal{D}_1}, \hat{Y}_{\mathcal{D}_1}, Y_1). \quad (5.13)$$

Note in (5.13), for any $j = b-L+1, b-L+2, \ldots, b$, $m_j$ and $m_{j-B}$ are both dummy messages according to (5.12), and both $\mathbf{X}_{0,j}(m_j|m_{j-B})$ and $\mathbf{X}_{1,j}(m_{j-B})$ are known at node 1. Then, it follows from the proof of Theorem 4.2.4 that $l_{\mathcal{D}_1,b-L}$ can be decoded if

$$I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_1\setminus\mathcal{S}}, Y_1|X_0, X_1, X_{\mathcal{D}_1\setminus\mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_0, X_1, X_{\mathcal{D}_1}, Y_1, \hat{Y}_{\mathcal{D}_1\setminus\mathcal{S}}) > 0, \quad (5.14)$$

for any nonempty $\mathcal{S} \subseteq \mathcal{D}_1$.

- b) Backwardly and sequentially from block $j = b-L$ to $j = b-B+1$, node 1 finds the unique pair $(m_j, l_{\mathcal{D}_1,j-1})$ satisfying (5.13), where $l_{\mathcal{D}_1,j}$ has already been recovered due to the backward property of decoding, and $m_{j-B}$ has been decoded by block $b-B$.

At each block $j = b-L, b-L-1, \ldots, b-B+1$, error occurs with $m_j$ if the true $m_j$ does not satisfy (5.13) with any $l_{\mathcal{D}_1,j-1}$, or a false $m_j$ satisfies (5.13) with some $l_{\mathcal{D}_1,j-1}$. According to the properties of typical sequences, the true $(m_j, l_{\mathcal{D}_1,j-1})$ satisfies (5.13) with high probability.

For a false $m_j$ and a $l_{\mathcal{D}_1,j-1}$ with false $\{l_{i,j-1}, i \in \mathcal{S}\}$ but true $\{l_{i,j-1}, i \in \mathcal{D}_1 \setminus \mathcal{S}\}$, $\mathbf{X}_{0,j}(m_j|m_{j-B})$ is conditionally independent of $\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{D}_1\}$ and $\mathbf{Y}_{1,j}$ given $\mathbf{X}_{1,j}(m_{j-B})$; and $\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{S}\}$ are independent of $\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{D}_1 \setminus \mathcal{S}\}$, $\mathbf{X}_{1,j}(m_{j-B})$ and $\mathbf{Y}_{1,j}$.

Therefore, the probability that such false $(m_j, l_{\mathcal{D}_1,j-1})$ satisfies (5.13) can be upper bounded by

$$2^{T(H(X_0,X_1,X_{\mathcal{D}_1},\hat{Y}_{\mathcal{D}_1},Y_1)+\epsilon)}2^{-T(H(X_1,X_{\mathcal{D}_1\setminus\mathcal{S}},\hat{Y}_{\mathcal{D}_1\setminus\mathcal{S}},Y_1)-\epsilon)}$$
$$\times 2^{-T(H(X_0|X_1)-\epsilon)}2^{-T(H(X_{\mathcal{S}})-\epsilon)}2^{-T(\sum_{i\in\mathcal{S}}(H(\hat{Y}_i|X_i)-\epsilon))}.$$

Since the number of such false $(m_j, l_{\mathcal{D}_1,j-1})$ is upper bounded by

$$2^{TR}\prod_{i\in\mathcal{S}}2^{T(I(Y_i;\hat{Y}_i|X_i)+\epsilon)},$$

with the union bound, it is easy to check that the probability of finding a false $m_j$ goes to zero as $T\to\infty$, if

$$R<\min_{\mathcal{S}\subseteq\mathcal{D}_1} I(X_0,X_{\mathcal{S}};\hat{Y}_{\mathcal{D}_1\setminus\mathcal{S}},Y_1|X_1,X_{\mathcal{D}_1\setminus\mathcal{S}})-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_0,X_1,X_{\mathcal{D}_1},Y_1,\hat{Y}_{\mathcal{D}_1\setminus\mathcal{S}}). \tag{5.15}$$

Then, based on the recovered $m_j$, $m_{j-B}$, and $l_{\mathcal{D}_1,j}$, again from the proof of Theorem 4.2.4, it follows that $l_{\mathcal{D}_1,j-1}$ can be decoded if (5.14) holds.

By a) and b) together, at the end of block $b=vB, v\in[1:B]$, the D-F relay node 1 can decode messages $(m_{b-B+1},m_{b-B+2},\ldots,m_b)$ if both (5.14) and (5.15) hold.

ii) At the end of all $B^2$ block, the destination node $n+2$ decodes messages $(m_1,m_2,\ldots,m_{B^2})$ using block-by-block backward decoding. Similarly, we only consider the decoding of $(m_1,m_2,\ldots,m_{B^2-B-L})$, since $(m_{B^2-B-L+1},m_{B^2-B-L+2},\ldots,m_{B^2})$ are all dummy messages according to (5.12).

- a) Node $n+2$ first finds the unique $l_{\mathcal{D}_{n+2},B^2-L}=\{l_{i,B^2-L}, i\in\mathcal{D}_{n+2}\}$ such that there exists some $l^{B^2}_{\mathcal{D}_{n+2},B^2-L+1}$ satisfying that for any block $j=B^2-L+1,B^2-L+2,\ldots,B^2$,

$$(\mathbf{X}_{0,j}(m_j|m_{j-B}),\mathbf{X}_{1,j}(m_{j-B}),\{(\mathbf{X}_{i,j}(l_{i,j-1}),\hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})):i\in\mathcal{D}_{n+2}\},\mathbf{Y}_{n+2,j})$$
$$\in A_\epsilon(X_0,X_1,X_{\mathcal{D}_{n+2}},\hat{Y}_{\mathcal{D}_{n+2}},Y_{n+2}), \tag{5.16}$$

where, similarly, $m_j$ and $m_{j-B}$ are both dummy messages according to (5.12), and $\mathbf{X}_{0,j}(m_j|m_{j-B})$ and $\mathbf{X}_{1,j}(m_{j-B})$ are both known at node $n+2$. Still, from the proof of Theorem 4.2.4, $l_{\mathcal{D}_{n+2},B^2-L}$ can be decoded if

$$I(X_{\mathcal{S}};\hat{Y}_{\mathcal{D}_{n+2}\setminus\mathcal{S}},Y_{n+2}|X_0,X_1,X_{\mathcal{D}_{n+2}\setminus\mathcal{S}}) \\ -I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_0,X_1,X_{\mathcal{D}_{n+2}},Y_{n+2},\hat{Y}_{\mathcal{D}_{n+2}\setminus\mathcal{S}})>0, \tag{5.17}$$

for any nonempty $\mathcal{S}\subseteq\mathcal{D}_{n+2}$.

- b) Backwardly and sequentially from block $j=B^2-L$ to $j=1$, node $n+2$ finds the unique pair $(m_{j-B},l_{\mathcal{D}_{n+2},j-1})$ satisfying (5.16), where $l_{\mathcal{D}_{n+2},j}$ has already been recovered due to the backward property of decoding, and $m_j$ either is a dummy message (for $j=B^2-L,B^2-L-1,\dots,B^2-B-L+1$) or has been decoded due to the backward property of decoding (for $j=B^2-B-L,B^2-B-L-1,\dots,1$).

At each block $j=B^2-L,B^2-L-1,\dots,1$, error occurs with $m_{j-B}$ if the true $m_{j-B}$ does not satisfy (5.16) with any $l_{\mathcal{D}_{n+2},j-1}$, or a false $m_{j-B}$ satisfies (5.16) with some $l_{\mathcal{D}_{n+2},j-1}$. According to the properties of typical sequences, the true $(m_{j-B},l_{\mathcal{D}_{n+2},j-1})$ satisfies (5.16) with high probability.

For a false $m_{j-B}$ and a $l_{\mathcal{D}_{n+2},j-1}$ with false $\{l_{i,j-1},i\in\mathcal{S}\}$ but true $\{l_{i,j-1},i\in\mathcal{D}_{n+2}\setminus\mathcal{S}\}$, $\mathbf{X}_{0,j}(m_j|m_{j-B})$ and $\mathbf{X}_{1,j}(m_{j-B})$ are independent of

$$\{(\mathbf{X}_{i,j}(l_{i,j-1}),\hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})):i\in\mathcal{D}_{n+2}\}\text{ and }\mathbf{Y}_{n+2,j};$$

and $\{(\mathbf{X}_{i,j}(l_{i,j-1}),\hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})):i\in\mathcal{S}\}$ are independent of

$$\{(\mathbf{X}_{i,j}(l_{i,j-1}),\hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})):i\in\mathcal{D}_{n+2}\setminus\mathcal{S}\}\text{ and }\mathbf{Y}_{n+2,j}.$$

Therefore, the probability that such false $(m_j,l_{\mathcal{D}_{n+2},j-1})$ satisfies (5.16) can be upper bounded by

$$2^{T(H(X_0,X_1,X_{\mathcal{D}_{n+2}},\hat{Y}_{\mathcal{D}_{n+2}},Y_{n+2})+\epsilon)}2^{-T(H(X_{\mathcal{D}_{n+2}\setminus\mathcal{S}},\hat{Y}_{\mathcal{D}_{n+2}\setminus\mathcal{S}},Y_{n+2})-\epsilon)} \\ \times 2^{-T(H(X_0,X_1)-\epsilon)}2^{-T(H(X_{\mathcal{S}})-\epsilon)}2^{-T(\sum_{i\in\mathcal{S}}(H(\hat{Y}_i|X_i)-\epsilon))}.$$

Since the number of such false $(m_j, l_{\mathcal{D}_{n+2},j-1})$ is upper bounded by

$$2^{TR}\prod_{i\in\mathcal{S}} 2^{T(I(Y_i;\hat{Y}_i|X_i)+\epsilon)},$$

with the union bound, it is easy to check that the probability of finding a false $m_j$ goes to zero as $T \to \infty$, if

$$R < \min_{\mathcal{S}\subseteq\mathcal{D}_1} I(X_0, X_1, X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_{n+2}\backslash\mathcal{S}}, Y_{n+2}|X_{\mathcal{D}_{n+2}\backslash\mathcal{S}}) \\ - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_0, X_1, X_{\mathcal{D}_{n+2}}, Y_{n+2}, \hat{Y}_{\mathcal{D}_{n+2}\backslash\mathcal{S}}). \quad (5.18)$$

Then, similarly, based on the recovered $m_j$, $m_{j-B}$, and $l_{\mathcal{D}_{n+2},j}$, $l_{\mathcal{D}_{n+2},j-1}$ can be decoded if (5.17) holds.

By a) and b) together, at the end of all $B^2$ block, the destination node $n+2$ can decode messages $(m_1, m_2, \ldots, m_{B^2})$ if both (5.17) and (5.18) hold.

Combining i) and ii), and using the standard technique of time sharing, we conclude that the rate described in (5.8)-(5.10) is achievable.

### 5.3.2 Multiple D-F Relay Nodes ($M \geq 2$)

When there are multiple D-F relay nodes, i.e., $M \geq 2$, a total of $B^{M+1}$ blocks will be used. The detailed codebook generation and encoding/decoding process are as follows.

*Codebook Generation:* Fix $p(x_0)p(x_{\mathcal{M}}|x_0)\prod_{i\in\mathcal{N}\backslash\mathcal{M}} p(x_i)p(\hat{y}_i|y_i, x_i)$. We randomly and independently generate a codebook for each block.

i) First consider the codebook generation for nodes $\pi(1 : M+1)$.

- For each block $b \in [1 : B^{M+1}]$, backwardly and sequentially for each relay node $\pi(k), k = M+1, M, \ldots, 2$, randomly generate $2^{TR}$ conditionally independent sequences

$$\mathbf{x}_{\pi(k),b}(m_{b-B^{k-1}}|m_{b-B^k}, \ldots, m_{b-B^M}),$$

where $m_{b-B^{k-1}}, m_{b-B^k}, \ldots, m_{b-B^M} \in [1 : 2^{TR}]$;

- For each block $b \in [1 : B^{M+1}]$ and node $\pi(1)$, i.e., the source node 0, randomly generate $2^{TR}$ conditionally independent sequences $\mathbf{x}_{0,b}(m_b|m_{b-B}, \ldots, m_{b-B^M})$, where

$$m_b, m_{b-B}, \ldots, m_{b-B^M} \in [1 : 2^{TR}].$$

ii) The codebook generation for the nodes in $\mathcal{N} \setminus \mathcal{M}$ is the same as that in the case of $M = 1$. For each block $b \in [1 : B^{M+1}]$ and each relay node $i \in \mathcal{N} \setminus \mathcal{M}$, randomly and independently generate $2^{T\hat{R}_i}$ sequences $\mathbf{x}_{i,b}(l_{i,b-1})$, $l_{i,b-1} \in [1 : 2^{T\hat{R}_i}]$, where $\hat{R}_i = I(Y_i; \hat{Y}_i|X_i) + \epsilon$; for each relay node $i \in \mathcal{N} \setminus \mathcal{M}$ and each $\mathbf{x}_{i,b}(l_{i,b-1})$, $l_{i,b-1} \in [1 : 2^{T\hat{R}_i}]$, randomly and conditionally independently generate $2^{T\hat{R}_i}$ sequences $\hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1})$, $l_{i,b} \in [1 : 2^{T\hat{R}_i}]$.

The combination of i) and ii) defines the codebook for any block $b \in [1 : B^{M+1}]$,

$$\begin{aligned}\mathcal{C}_b = \Big\{&\mathbf{x}_{\pi(k),b}(m_{b-B^{k-1}}|m_{b-B^k}, \ldots, m_{b-B^M}) :\\ &m_{b-B^{k-1}}, m_{b-B^k}, \ldots, m_{b-B^M} \in [1 : 2^{TR}], k = M+1, M, \ldots, 2;\\ &\mathbf{x}_{0,b}(m_b|m_{b-B}, \ldots, m_{b-B^M}) : m_b, m_{b-B}, \ldots, m_{b-B^M} \in [1 : 2^{TR}];\\ &\mathbf{x}_{i,b}(l_{i,b-1}), \hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1}) : l_{i,b}, l_{i,b-1} \in [1 : 2^{T\hat{R}_i}], i \in \mathcal{N} \setminus \mathcal{M}\Big\}. \end{aligned} \tag{5.19}$$

*Encoding:* Let $\mathbf{m} = (m_1, m_2, \ldots, m_{B^{M+1}})$ be the message vector to be sent and let $m_b = 1$ be the dummy message for any

$$b \in \cup_{w=1}^{B^M}[wB - L + 1 : wB] \bigcup \cup_{u=1}^{M} \cup_{v=1}^{B^{M-u}} [v(B-1)B^u + 1 : vB^{u+1}], \tag{5.20}$$

and for any $b \leq 0$. Now, the actually achievable rate is $\frac{B-L}{B}(\frac{B-1}{B})^M R$ due to the dummy messages, which can still be made arbitrarily close to $R$ by choosing $L \ll B$, similarly as in the proof for the single D-F relay node case in Subsection 5.3.1.

i) We still first consider the encoding process for nodes $\pi(1 : M+1)$.

- In block $b \in [1 : B^{M+1}]$, node $\pi(1)$, i.e., the source node 0, transmits $\mathbf{x}_{0,b}(m_b|m_{b-B}, \ldots, m_{b-B^M})$.

- By the end of block $vB^{k-1}, v \in [1 : \frac{B^{M+1}}{B^{k-1}} - 1]$, the D-F relay node $\pi(k), k = 2, \ldots, M+1$, has decoded messages $(m_1, m_2, \ldots, m_{vB^{k-1}})$ using backward decoding (see the decoding part). In the next $B^{k-1}$ blocks, i.e., in block $b \in [vB^{k-1}+1 : (v+1)B^{k-1}]$, node $\pi(k), k = 2, \ldots, M+1$, transmits $\mathbf{x}_{\pi(k),b}(m_{b-B^{k-1}}|m_{b-B^k}, \ldots, m_{b-B^M})$, where

$$(m_{b-B^{k-1}}, m_{b-B^k}, \ldots, m_{b-B^M}), b \in [vB^{k-1}+1 : (v+1)B^{k-1}]$$

have all been decoded by block $vB^{k-1}$.

ii) The encoding process for the nodes in $\mathcal{N} \setminus \mathcal{M}$ is still the same as that in the case of $M = 1$. For any block $b \in [1 : B^{M+1}]$, each relay node $i \in \mathcal{N} \setminus \mathcal{M}$, upon receiving $\mathbf{y}_{i,b}$ at the end of block $b$, finds an index $l_{i,b}$ such that

$$(\mathbf{x}_{i,b}(l_{i,b-1}), \mathbf{y}_{i,b}, \hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1})) \in A_\epsilon(X_i, Y_i, \hat{Y}_i),$$

where $l_{i,0} = 1$ by convention. In block $b \in [1 : B^{M+1}]$, the relay node $i \in \mathcal{N} \setminus \mathcal{M}$ transmits $\mathbf{x}_{i,b}(l_{i,b-1})$.

*Decoding:* At the end of block $b = vB^{k-1}, v \in [1 : B^{M+1}/B^{k-1}]$, the node $\pi(k), k = 2, \ldots, M+2$, decodes messages $(m_{b-B^{k-1}+1}, \ldots, m_b)$ using block-by-block backward decoding as follows.

i) The node $\pi(k), k = 2, \ldots, M+2$, first finds the unique $l_{\mathcal{D}_k,b-L} = \{l_{i,b-L}, i \in \mathcal{D}_k\}$ such that there exists some $l^b_{\mathcal{D}_k,b-L+1}$ satisfying that for any block $j = b-L+1, b-L+2, \ldots, b$,

$$\begin{aligned}
&(\mathbf{X}_{0,j}(m_j|m_{j-B}, \ldots, m_{j-B^M}),\\
&\{\mathbf{X}_{\pi(s),j}(m_{j-B^{s-1}}|m_{j-B^s}, \ldots, m_{j-B^M}), s = 2, \ldots, k-1, k, k+1, \ldots, M+1\},\\
&\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{D}_k\}, \mathbf{Y}_{\pi(k),j}) \in A_\epsilon(X_0, X_{\mathcal{M}}, X_{\mathcal{D}_k}, \hat{Y}_{\mathcal{D}_k}, Y_{\pi(k)}).
\end{aligned} \tag{5.21}$$

Note in (5.21), $(m_j, m_{j-B}, \ldots, m_{j-B^M}), j = b-L+1, b-L+2, \ldots, b$ are all dummy messages according to (5.20), and thus $\mathbf{X}_{\pi(s),j}, s = 1, \ldots, M+1$ are all known at node $\pi(k)$. Then, it follows from the proof of Theorem 4.2.4 that $l_{\mathcal{D}_k,b-L}$ can be decoded if

$$I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_k \setminus \mathcal{S}}, Y_{\pi(k)}|X_{\pi(1:M+1)}, X_{\mathcal{D}_k \setminus \mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_{\pi(1:M+1)}, X_{\mathcal{D}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{D}_k \setminus \mathcal{S}}) > 0, \tag{5.22}$$

for any nonempty $\mathcal{S} \subseteq \mathcal{D}_k$.

ii) Backwardly and sequentially from block $j = b - L$ to $j = b - B^{k-1} + 1$, node $\pi(k), k = 2, 3, \ldots, M+2$, jointly decodes the message transmitted by its immediate upstream node $\pi(k-1)$, and the compressions of the C-F relay nodes. Specifically, node $\pi(k), k = 3, 4, \ldots, M+2$ finds the unique pair $(m_{j-B^{k-2}}, l_{\mathcal{D}_k,j-1})$ satisfying (5.21), and node $\pi(k), k = 2$, finds the unique pair $(m_j, l_{\mathcal{D}_k,j-1})$ satisfying (5.21). Here the exception for node $\pi(2)$ arises because the source node $\pi(1)$ transmits $m_j$ rather than $m_{j-1}$ in block $j$, but the ideas of the decoding processes at all $\pi(k), k = 2, 3, \ldots, M+2$, are exactly the same. Thus, below, we only present the decoding at node $\pi(k), k = 3, 4, \ldots, M+2$, while the decoding at node $\pi(2)$ can be easily obtained by analogy. The same consideration also applies to the proof in 5.4.2.

In (5.21), $l_{\mathcal{D}_k,j}$ has already been recovered due to the backward property of decoding, and among the messages $(m_j, m_{j-B}, \ldots, m_{j-B^M})$, only $m_{j-B^{k-2}}$ is the unknown message at node $\pi(k)$ that needs to be decoded in block $j$. In fact, $(m_{j-B^{k-1}}, \ldots, m_{j-B^M})$ have been decoded by block $b-B^{k-1}$, while $(m_j, \ldots, m_{j-B^{k-3}})$ either are dummy messages according to (5.20) (for block $j = b-L, b-L-1 \ldots, b-B^{k-2}-L+1$) or have been decoded due to the backward property of decoding (for block $j = b - B^{k-2} - L, b - B^{k-2} - L - 1, \ldots, b - B^{k-1} + 1$).

At each block $j = b-L, b-L-1, \ldots, b-B^{k-1}+1$, error occurs with $m_{j-B^{k-2}}$ if the true $m_{j-B^{k-2}}$ does not satisfy (5.21) with any $l_{\mathcal{D}_k,j-1}$, or a false $m_{j-B^{k-2}}$ satisfies (5.21) with some $l_{\mathcal{D}_k,j-1}$. According to the properties of typical sequences, the true $(m_{j-B^{k-2}}, l_{\mathcal{D}_k,j-1})$ satisfies (5.21) with high probability.

For a false $m_{j-B^{k-2}}$ and a $l_{\mathcal{D}_k,j-1}$ with false $\{l_{i,b-1}, i \in \mathcal{S}\}$ but true $\{l_{i,b-1}, i \in \mathcal{D}_k \setminus \mathcal{S}\}$,

$$\{\mathbf{X}_{\pi(1),j}(m_j|m_{j-B}, \ldots, m_{j-B^M}), \mathbf{X}_{\pi(s),j}(m_{j-B^{s-1}}|m_{j-B^s}, \ldots, m_{j-B^M}), s = 2, \ldots, k-1\}$$

are conditionally independent of $\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{D}_k\}$ and $\mathbf{Y}_{\pi(k),j}$ given

$$\{\mathbf{X}_{\pi(s),j}(m_{j-B^{s-1}}|m_{j-B^s}, \ldots, m_{j-B^M}), s = k, \ldots, M+1\};$$

and $\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{S}\}$ are independent of

$$\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \mathcal{D}_k \setminus \mathcal{S}\},$$
$$\{\mathbf{X}_{\pi(s),j}(m_{j-B^{s-1}}|m_{j-B^s}, \ldots, m_{j-B^M}), s \in [k : M+1]\}, \mathbf{Y}_{\pi(k),j}.$$

Therefore, the probability that such false $(m_{j-B^{k-2}}, l_{\mathcal{D}_k,j-1})$ satisfies (5.21) can be upper bounded by

$$2^{T(H(X_{\pi(1:M+1)},X_{\mathcal{D}_k},\hat{Y}_{\mathcal{D}_k},Y_{\pi(k)})+\epsilon)}2^{-T(H(X_{\pi(k:M+1)},X_{\mathcal{D}_k\setminus\mathcal{S}},\hat{Y}_{\mathcal{D}_k\setminus\mathcal{S}},Y_{\pi(k)})-\epsilon)}$$
$$\times 2^{-T(H(X_{\pi(1:k-1)}|X_{\pi(k:M+1)})-\epsilon)}2^{-T(H(X_{\mathcal{S}})-\epsilon)}2^{-T(\sum_{i\in\mathcal{S}}(H(\hat{Y}_i|X_i)-\epsilon))}.$$

Since the number of such false $(m_{j-B^{k-2}}, l_{\mathcal{D}_k,j-1})$ is upper bounded by

$$2^{TR}\prod_{i\in\mathcal{S}}2^{T(I(Y_i;\hat{Y}_i|X_i)+\epsilon)},$$

with the union bound, it is easy to check that the probability of finding a false $m_{j-B^{k-2}}$ goes to zero as $T \to \infty$, if

$$R < \min_{\mathcal{S}\subseteq\mathcal{D}_k} I(X_{\pi(1:k-1)}, X_{\mathcal{S}}; \hat{Y}_{\mathcal{D}_k\setminus\mathcal{S}}, Y_{\pi(k)}|X_{\mathcal{D}_k\setminus\mathcal{S}}, X_{\pi(k:M+1)})$$
$$- I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|X_{\pi(1:M+1)}, X_{\mathcal{D}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{D}_k\setminus\mathcal{S}}). \tag{5.23}$$

Then, based on the recovered $(m_j, m_{j-B}, \ldots, m_{j-B^M})$ and $l_{\mathcal{D}_k,j}$, from the proof of Theorem 4.2.4, it follows that $l_{\mathcal{D}_k,j-1}$ can be decoded if (5.22) holds.

Combining i) and ii), using the technique of time sharing, we obtain the achievable rate (5.5)-(5.6).

## 5.4 Unified Relay Framework With $B$-Blocks-By-$B$-Blocks Backward Decoding

Under the unified relay framework using nested blocks and backward decoding, we can also consider combining the noisy network coding scheme [32] with the multi-level D-F scheme. However, since noisy network coding uses repetitive encoding/all blocks united decoding, to make it fit into our framework, a modification is needed.

Specifically, assume some fixed $\mathcal{M} \subseteq \mathcal{N}$ with $|\mathcal{M}| = M$ and $\pi(\{0, \mathcal{M}, n+1\})$, and a total of $B^{M+1}$ blocks are used. The source can still repetitively encode intra-$B$-blocks as in [32], but inter-$B$-blocks, the source has to cumulatively encode to allow for the operation of D-F strategy; Correspondingly, both the D-F relay nodes and the destination will perform $B$-blocks-by-$B$-blocks backward decoding, which is essentially a combination of backward decoding and $B$ blocks united decoding. Same as in Section 5.3, the backward decoding at node $\pi(k), k = 2, 3, \ldots, M+2$, will happen at the end of every $B^{k-1}$ blocks, i.e., at the end of block $b = vB^{k-1}, v \in [1 : B^{M+1}/B^{k-1}]$, and both the D-F relay nodes and the destination node perform compression-message joint decoding. Below, we still first consider the case of single D-F relay node ($M = 1$) to illustrate the main idea, and then extend it to the general case of multiple D-F relay nodes ($M \geq 2$).

### 5.4.1 Single D-F relay node ($M = 1$)

Still assume that only node 1 is the D-F relay node, and all other relay nodes are the C-F relay nodes, and let $\tilde{\mathcal{N}} := \mathcal{N} \setminus \{1\}$. Specializing Theorem 5.2.2 to this case, we have that a rate $R$ is achievable, if there exists some

$$p(q)p(x_0|q)p(x_1|x_0, q) \prod_{i \in \tilde{\mathcal{N}}} p(x_i|q)p(\hat{y}_i|y_i, x_i, q),$$

such that

$$R < \min \left\{ \begin{aligned} &\max_{\mathcal{T}_1 \subseteq \tilde{\mathcal{N}}} \min_{\mathcal{S} \subseteq \mathcal{T}_1} I(X_0, X_\mathcal{S}; \hat{Y}_{\mathcal{T}_1 \setminus \mathcal{S}}, Y_1 | X_1, X_{\mathcal{T}_1 \setminus \mathcal{S}}, Q) \\ &\qquad\qquad - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S} | X_0, X_1, X_{\mathcal{T}_1}, Y_1, \hat{Y}_{\mathcal{T}_1 \setminus \mathcal{S}}, Q), \\ &\max_{\mathcal{T}_{n+2} \subseteq \tilde{\mathcal{N}}} \min_{\mathcal{S} \subseteq \mathcal{T}_{n+2}} I(X_0, X_1, X_\mathcal{S}; \hat{Y}_{\mathcal{T}_{n+2} \setminus \mathcal{S}}, Y_{n+2} | X_{\mathcal{T}_{n+2} \setminus \mathcal{S}}, Q) \\ &\qquad\qquad - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S} | X_0, X_1, X_{\mathcal{T}_{n+2}}, Y_{n+2}, \hat{Y}_{\mathcal{T}_{n+2} \setminus \mathcal{S}}, Q). \end{aligned} \right\} \tag{5.24}$$

Still, a total of $B^2$ blocks will be used. The detailed codebook generation and encoding/decoding process are as follows, which can be understood with the help of Table 5.2.

Table 5.2: $B$-Blocks-by-$B$-Blocks backward decoding for the single D-F relay node case

| Block | 1 | 2 | $\cdots$ | $B-1$ | $B$ | $\cdots$ |
|---|---|---|---|---|---|---|
| $X_0$ | $\mathbf{x}_{0,1}(m_1\|1)$ | $\mathbf{x}_{0,2}(m_1\|1)$ | $\cdots$ | $\mathbf{x}_{0,B-1}(m_1\|1)$ | $\mathbf{x}_{0,B}(m_1\|1)$ | $\cdots$ |
| $Y_1$ | $\emptyset$ | $\emptyset$ | $\cdots$ | $\emptyset$ | $m_1$ | $\cdots$ |
| $X_1$ | $\mathbf{x}_{1,1}(1)$ | $\mathbf{x}_{1,2}(1)$ | $\cdots$ | $\mathbf{x}_{1,B-1}(1)$ | $\mathbf{x}_{1,B}(1)$ | $\cdots$ |
| $Y_{\tilde{\mathcal{N}}}$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},1}(l_{\tilde{\mathcal{N}},1}\|\mathbf{1})$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},2}(l_{\tilde{\mathcal{N}},2}\|l_{\tilde{\mathcal{N}},1})$ | $\cdots$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B-1}(l_{\tilde{\mathcal{N}},B-1}\|l_{\tilde{\mathcal{N}},B-2})$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B}(l_{\tilde{\mathcal{N}},B}\|l_{\tilde{\mathcal{N}},B-1})$ | $\cdots$ |
| $X_{\tilde{\mathcal{N}}}$ | $\mathbf{x}_{\tilde{\mathcal{N}},1}(\mathbf{1})$ | $\mathbf{x}_{\tilde{\mathcal{N}},2}(l_{\tilde{\mathcal{N}},1})$ | $\cdots$ | $\mathbf{x}_{\tilde{\mathcal{N}},B-1}(l_{\tilde{\mathcal{N}},B-2})$ | $\mathbf{x}_{\tilde{\mathcal{N}},B}(l_{\tilde{\mathcal{N}},B-1})$ | $\cdots$ |
| $Y_{n+2}$ | $\emptyset$ | $\emptyset$ | $\cdots$ | $\emptyset$ | $\emptyset$ | $\cdots$ |

| Block | $B^2-B+1$ | $B^2-B+2$ | $\cdots$ | $B^2-1$ | $B^2$ |
|---|---|---|---|---|---|
| $X_0$ | $\mathbf{x}_{0,B^2-B+1}(1\|m_{B-1})$ | $\mathbf{x}_{0,B^2-B+2}(1\|m_{B-1})$ | $\cdots$ | $\mathbf{x}_{0,B^2-1}(1\|m_{B-1})$ | $\mathbf{x}_{0,B^2}(1\|m_{B-1})$ |
| $Y_1$ | $\emptyset$ | $\emptyset$ | $\cdots$ | $\emptyset$ | $m_B$ |
| $X_1$ | $\mathbf{x}_{1,B^2-B+1}(m_{B-1})$ | $\mathbf{x}_{1,B^2-B+2}(m_{B-1})$ | $\cdots$ | $\mathbf{x}_{1,B^2-1}(m_{B-1})$ | $\mathbf{x}_{1,B^2}(m_{B-1})$ |
| $Y_{\tilde{\mathcal{N}}}$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2-B+1}(l_{\tilde{\mathcal{N}},B^2-B+1}\|l_{\tilde{\mathcal{N}},B^2-B})$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2-B+2}(l_{\tilde{\mathcal{N}},B^2-B+2}\|l_{\tilde{\mathcal{N}},B^2-B+1})$ | $\cdots$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2-1}(l_{\tilde{\mathcal{N}},B^2-1}\|l_{\tilde{\mathcal{N}},B^2-2})$ | $\hat{\mathbf{y}}_{\tilde{\mathcal{N}},B^2}(l_{\tilde{\mathcal{N}},B^2}\|l_{\tilde{\mathcal{N}},B^2-1})$ |
| $X_{\tilde{\mathcal{N}}}$ | $\mathbf{x}_{\tilde{\mathcal{N}},B^2-B+1}(l_{\tilde{\mathcal{N}},B^2-B})$ | $\mathbf{x}_{\tilde{\mathcal{N}},B^2-B+2}(l_{\tilde{\mathcal{N}},B^2-B+1})$ | $\cdots$ | $\mathbf{x}_{\tilde{\mathcal{N}},B^2-1}(l_{\tilde{\mathcal{N}},B^2-2})$ | $\mathbf{x}_{\tilde{\mathcal{N}},B^2}(l_{\tilde{\mathcal{N}},B^2-1})$ |
| $Y_{n+2}$ | $\emptyset$ | $\emptyset$ | $\cdots$ | $\emptyset$ | $(m_1, m_2, \ldots, m_B)$ |

*Codebook Generation:* Fix $p(x_0)p(x_1|x_0)\prod_{i\in\tilde{\mathcal{N}}} p(x_i)p(\hat{y}_i|y_i,x_i)$. We randomly and independently generate a codebook for each block.

i) First consider the codebook generation for the source node 0 and the D-F relay node 1. Denote $f(b) := \lceil \frac{b}{B} \rceil$, i.e., the smallest integer greater than or equal to $\frac{b}{B}$. For each block $b\in[1:B^2]$, randomly generate $2^{TBR}$ independent sequences $\mathbf{x}_{1,b}(m_{f(b-B)})$ for node 1, and randomly generate $2^{TBR}$ conditionally independent sequences $\mathbf{x}_{0,b}(m_{f(b)}|m_{f(b-B)})$ for node 0, where $m_{f(b)}, m_{f(b-B)} \in [1:2^{TBR}]$.

ii) The codebook generation for the C-F relay nodes is exactly the same as that in Section 5.3. For each block $b\in[1:B^2]$ and each relay node $i\in\tilde{\mathcal{N}}$, randomly and independently generate $2^{T\hat{R}_i}$ sequences $\mathbf{x}_{i,b}(l_{i,b-1})$, $l_{i,b-1}\in[1:2^{T\hat{R}_i}]$, where $\hat{R}_i = I(Y_i;\hat{Y}_i|X_i)+\epsilon$; for each relay node $i\in\tilde{\mathcal{N}}$ and each $\mathbf{x}_{i,b}(l_{i,b-1})$, $l_{i,b-1}\in[1:2^{T\hat{R}_i}]$, randomly and conditionally independently generate $2^{T\hat{R}_i}$ sequences $\hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1})$, $l_{i,b}\in[1:2^{T\hat{R}_i}]$.

The combination of i) and ii) defines the codebook for any block $b\in[1:B^2]$,

$$\begin{aligned}\mathcal{C}_b = \Big\{&\mathbf{x}_{1,b}(m_{f(b-B)}), \mathbf{x}_{0,b}(m_{f(b)}|m_{f(b-B)}) : m_{f(b)}, m_{f(b-B)} \in [1:2^{TBR}];\\ &\mathbf{x}_{i,b}(l_{i,b-1}), \hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1}) : l_{i,b}, l_{i,b-1}\in[1:2^{T\hat{R}_i}], i\in\tilde{\mathcal{N}}\Big\}. \end{aligned} \quad (5.25)$$

*Encoding:* Let the message vector to be sent be

$$\mathbf{m} = (\underbrace{m_1, m_1, \ldots, m_1}_{B}, \underbrace{m_2, m_2, \ldots, m_2}_{B}, \ldots, \underbrace{m_B, m_B, \ldots, m_B}_{B}).$$

Let $m_B = 1$ be the dummy message, i.e., $m_{f(b)} = 1$ for any

$$b \in [(B-1)B+1 : B^2], \quad (5.26)$$

and for any $b \le 0$. The actually achievable rate is $\frac{B-1}{B}R$ due to the dummy messages, which, however, can be made arbitrarily close to $R$ by letting $B\to\infty$.

i) First consider the encoding process for nodes 0 and 1.

- In block $b\in[1:B^2]$, the source node 0 transmits $\mathbf{x}_{0,f(b)}(m_{f(b)}|m_{f(b-B)})$.

- At the end of block $vB, v \in [1 : B-1]$, the D-F relay node 1 has decoded message $m_v$ using $B$ blocks united decoding (see the decoding part). In the next $B$ blocks, i.e., in block $b \in [vB+1 : (v+1)B]$, the relay node 1 transmits $\mathbf{x}_{1,b}(m_{f(b-B)})$, where $m_{f(b-B)}$ for any $b \in [vB+1 : (v+1)B]$ is corresponding to $m_v$ that has been decoded by block $vB$.

ii) For any block $b \in [1 : B^2]$, each relay node $i \in \tilde{\mathcal{N}}$, upon receiving $\mathbf{y}_{i,b}$ at the end of block $b$, finds an index $l_{i,b}$ such that

$$(\mathbf{x}_{i,b}(l_{i,b-1}), \mathbf{y}_{i,b}, \hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1})) \in A_\epsilon(X_i, Y_i, \hat{Y}_i),$$

where $l_{i,0} = 1$ by convention. In block $b \in [1 : B^2]$, the relay node $i \in \tilde{\mathcal{N}}$ transmits $\mathbf{x}_{i,b}(l_{i,b-1})$.

*Decoding:* We present the decoding process at the D-F relay node 1 and at the destination node $n+2$ separately.

i) At the end of block $b = vB, v \in [1 : B]$, the D-F relay node 1 decodes messages $m_v$ using $B$ blocks united decoding, i.e., it finds the unique $m_v$, such that there exists some $l_{\tilde{\mathcal{N}},(v-1)B+1}^{vB}$ satisfying that for any block $j = (v-1)B+1, (v-1)B+2, \dots, vB$,

$$\begin{aligned}&(\mathbf{X}_{0,j}(m_{f(j)}|m_{f(j-B)}), \mathbf{X}_{1,j}(m_{f(j-B)})\\ &\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \tilde{\mathcal{N}}\}, \mathbf{Y}_{1,j}) \in A_\epsilon(X_0, X_1, X_{\tilde{\mathcal{N}}}, \hat{Y}_{\tilde{\mathcal{N}}}, Y_1),\end{aligned} \quad (5.27)$$

where $m_{f(j-B)}$ is corresponding to $m_{v-1}$ and has been decoded by the end of block $(v-1)B$, and $m_{f(j)}$ is corresponding to $m_v$. From [32, Thm 1] and its proof (see also Theorem 4.2.1), we have that $m_v$ can be decoded if

$$R < \min_{\mathcal{S} \subseteq \tilde{\mathcal{N}}} I(X_0, X_{\mathcal{S}}; \hat{Y}_{\tilde{\mathcal{N}} \setminus \mathcal{S}}, Y_1 | X_{\tilde{\mathcal{N}} \setminus \mathcal{S}}, X_1) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\tilde{\mathcal{N}}}, Y_1, \hat{Y}_{\tilde{\mathcal{N}} \setminus \mathcal{S}}). \quad (5.28)$$

Note, (5.28) can be improved by considering only a subset $\mathcal{T}_1 \subseteq \tilde{\mathcal{N}}$ for the decoding while treating the inputs of other C-F relay nodes as purely noise, leading to following more general rate constraint:

$$R < \max_{\mathcal{T}_1 \subseteq \tilde{\mathcal{N}}} \min_{\mathcal{S} \subseteq \mathcal{T}_1} I(X_0, X_{\mathcal{S}}; \hat{Y}_{\mathcal{T}_1 \setminus \mathcal{S}}, Y_1 | X_{\mathcal{T}_1 \setminus \mathcal{S}}, X_1) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\mathcal{T}_1}, Y_1, \hat{Y}_{\mathcal{T}_1 \setminus \mathcal{S}}). \quad (5.29)$$

ii) At the end of all $B^2$ block, the destination node decodes all messages

$$(m_1, m_2, \ldots, m_B)$$

using $B$-blocks-by-$B$-blocks backward decoding. In fact, since $m_B = 1$ is dummy message, only $(m_1, m_2, \ldots, m_{B-1})$ need to be decoded. For this, backwardly and sequentially for $g = B-1, B-2, \ldots, 1$, node $n+2$ finds the unique $m_g$ such that there exists some $l_{\tilde{\mathcal{N}},gB+1}^{gB+B}$ satisfying that for any block $j = gB+1, gB+2, \ldots, gB+B$,

$$\begin{aligned}&(\mathbf{X}_{0,j}(m_{f(j)}|m_{f(j-B)}), \mathbf{X}_{1,j}(m_{f(j-B)})\\ &\{(\mathbf{X}_{i,j}(l_{i,j-1}), \hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i \in \tilde{\mathcal{N}}\}, \mathbf{Y}_{n+2,j}) \in A_\epsilon(X_0, X_1, X_{\tilde{\mathcal{N}}}, \hat{Y}_{\tilde{\mathcal{N}}}, Y_{n+2}).\end{aligned} \tag{5.30}$$

Note in (5.30), for $j = gB+1, gB+2, \ldots, gB+B$, only $m_{f(j-B)}$, corresponding to $m_g$, needs decoding; and $m_{f(j)}$, corresponding to $m_{g+1}$, either is a dummy message (for $g = B-1$, i.e., $j = (B-1)B+1, (B-1)B+2, \ldots, B^2$), or has been decoded due to the backward property of decoding (for $g = B-2, \ldots, 1$). Thus, $X_0$ and $X_1$ are cooperatively transmitting the message $m_g$, and similarly as above, $m_g$ can be decoded if

$$R < \min_{\mathcal{S} \subseteq \tilde{\mathcal{N}}} I(X_0, X_1, X_{\mathcal{S}}; \hat{Y}_{\tilde{\mathcal{N}} \setminus \mathcal{S}}, Y_{n+2} | X_{\tilde{\mathcal{N}} \setminus \mathcal{S}}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\tilde{\mathcal{N}}}, Y_{n+2}, \hat{Y}_{\tilde{\mathcal{N}} \setminus \mathcal{S}}). \tag{5.31}$$

Also, (5.31) can be improved by considering only a subset $\mathcal{T}_{n+2}$ for the decoding, leading to the following rate constraint:

$$\begin{aligned}R < \max_{\mathcal{T}_{n+2} \subseteq \tilde{\mathcal{N}}} \min_{\mathcal{S} \subseteq \mathcal{T}_{n+2}} & I(X_0, X_1, X_{\mathcal{S}}; \hat{Y}_{\mathcal{T}_{n+2} \setminus \mathcal{S}}, Y_{n+2} | X_{\mathcal{T}_{n+2} \setminus \mathcal{S}})\\ & - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_0, X_1, X_{\mathcal{T}_{n+2}}, Y_{n+2}, \hat{Y}_{\mathcal{T}_{n+2} \setminus \mathcal{S}}).\end{aligned} \tag{5.32}$$

Combining (5.29) and (5.32) and using the technique of time sharing, we have that the rate in (5.24) is achievable.

### 5.4.2 Multiple D-F Relay Nodes ($M \geq 2$)

*Codebook Generation:* Fix $p(x_0)p(x_{\mathcal{M}}|x_0)\prod_{i \in \mathcal{N} \setminus \mathcal{M}} p(x_i)p(\hat{y}_i|y_i, x_i)$. We randomly and independently generate a codebook for each block. The codebook generation for the C-F relay nodes is exactly the same as that in 5.3 and 5.4.1, and hence omitted. We only present the codebook generation for nodes $\pi(1:M+1)$. Still, denote $f(b) := \lceil \frac{b}{B} \rceil$, i.e., the smallest integer greater than or equal to $\frac{b}{B}$.

- For each block $b \in [1 : B^{M+1}]$, backwardly and sequentially for each relay node $\pi(k), k = M+1, M, \dots, 2$, randomly generate $2^{TBR}$ conditionally independent sequences

$$\mathbf{x}_{\pi(k),b}(m_{f(b-B^{k-1})}|m_{f(b-B^{k})}, \dots, m_{f(b-B^{M})}),$$

  where $m_{f(b-B^{k-1})}, m_{f(b-B^{k})}, \dots, m_{f(b-B^{M})} \in [1 : 2^{TBR}]$.

- For each block $b \in [1 : B^{M+1}]$ and node $\pi(1)$, i.e., the source node 0, randomly generate $2^{TBR}$ conditionally independent sequences

$$\mathbf{x}_{0,b}(m_{f(b)}|m_{f(b-B)}, \dots, m_{f(b-B^{M})}),$$

  where $m_{f(b)}, m_{f(b-B)}, \dots, m_{f(b-B^{M})} \in [1 : 2^{TBR}]$.

The above, together with the codebook generation for the C-F relay nodes, defines the codebook for any block $b \in [1 : B^{M+1}]$,

$$\begin{aligned}\mathcal{C}_b = \{&\mathbf{x}_{\pi(k),b}(m_{f(b-B^{k-1})}|m_{f(b-B^{k})}, \dots, m_{f(b-B^{M})}) : \\ &m_{f(b-B^{k-1})}, \dots, m_{f(b-B^{M})} \in [1 : 2^{TBR}], k = M+1, M, \dots, 2; \\ &\mathbf{x}_{0,b}(m_{f(b)}|m_{f(b-B)}, \dots, m_{f(b-B^{M})}) : m_{f(b)}, m_{f(b-B)}, \dots, m_{f(b-B^{M})} \in [1 : 2^{TBR}]; \\ &\mathbf{x}_{i,b}(l_{i,b-1}), \hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1}) : l_{i,b}, l_{i,b-1} \in [1 : 2^{T\hat{R}_i}], i \in \mathcal{N} \setminus \mathcal{M}\}.\end{aligned}$$

*Encoding:* Let the message vector to be sent be

$$\mathbf{m} = (\underbrace{m_1, m_1, \dots, m_1}_{B}, \underbrace{m_2, m_2, \dots, m_2}_{B}, \dots, \underbrace{m_{B^M}, m_{B^M}, \dots, m_{B^M}}_{B}).$$

Let $m_{f(b)} = 1$ be the dummy message for any

$$b \in \cup_{u=1}^{M} \cup_{v=1}^{B^{M-u}} [v(B-1)B^u + 1 : vB^{u+1}], \tag{5.33}$$

and for any $b \le 0$. The actually achievable rate is $(\frac{B-1}{B})^M R$ due to the dummy messages, which can still be made arbitrarily close to $R$ by letting $B \to \infty$ for any $M$.

The encoding process for the C-F relay nodes is still exactly the same as that in 5.3 and 5.4.1, and hence omitted. We only present the encoding process for nodes $\pi(1 : M+1)$.

- In block $b \in [1 : B^{M+1}]$, the source node 0 transmits

$$\mathbf{x}_{0,b}(m_{f(b)}|m_{f(b-B)}, \ldots, m_{f(b-B^M)}).$$

- At the end of block $vB^{k-1}, v \in [1 : \frac{B^{M+1}}{B^{k-1}} - 1]$, the relay node $\pi(k), k = 2, \ldots, M+1$, has decoded messages $(m_1, m_2, \ldots, m_{vB^{k-2}})$ using backward decoding (see the decoding part). In the next $B^{k-1}$ blocks, i.e., in block $b \in [vB^{k-1}+1 : (v+1)B^{k-1}]$, the relay node $\pi(k), k = 2, \ldots, M+1$, transmits

$$\mathbf{x}_{\pi(k),b}(m_{f(b-B^{k-1})}|m_{f(b-B^k)}, \ldots, m_{f(b-B^M)}),$$

  where $(m_{f(b-B^{k-1})}, m_{f(b-B^k)}, \ldots, m_{f(b-B^M)})$ for any $b \in [vB^{k-1} + 1 : (v + 1)B^{k-1}]$ have all been decoded by block $vB^{k-1}$.

*Decoding:* At the end of every $B^{k-1}$ blocks, the node $\pi(k), k = 2, \ldots, M+2$ decodes $B^{k-2}$ messages using $B$-Blocks-By-$B$-Blocks backward decoding. (Note every $B^{k-1}$ blocks carry $B^{k-2}$ messages.) Specifically, at the end of block $b = vB^{k-1}, v \in [1 : B^{M+1}/B^{k-1}]$, the node $\pi(k), k = 2, \ldots, M+2$, decodes messages $(m_{(v-1)B^{k-2}+1}, \ldots, m_{vB^{k-2}})$. In fact, $(m_{vB^{k-2}-B^{k-3}+1}, \ldots, m_{vB^{k-2}})$ are dummy messages according to (5.33), and only $(m_{(v-1)B^{k-2}+1}, \ldots, m_{vB^{k-2}-B^{k-3}})$ need decoding. For this, backwardly and sequentially for $g = vB^{k-2} - B^{k-3}, vB^{k-2} - B^{k-3} - 1, \ldots, (v-1)B^{k-2} + 1$, node $\pi(k)$ finds the unique $m_g$ such that there exists some $l^{gB+B^{k-2}}_{\mathcal{N}\setminus\mathcal{M},(g-1)B+B^{k-2}+1}$ satisfying that for any block $j = (g-1)B + B^{k-2}+1, (g-1)B+$

$B^{k-2}+2,\ldots,gB+B^{k-2}$,

$$(\mathbf{X}_{0,j}(m_{f(j)}|m_{f(j-B)},\ldots,m_{f(j-B^M)}),$$
$$\{\mathbf{X}_{\pi(s),j}(m_{f(j-B^{s-1})}|m_{f(j-B^s)},\ldots,m_{f(j-B^M)}), s=2,\ldots,k-1,k,k+1,\ldots,M+1\},$$
$$\{(\mathbf{X}_{i,j}(l_{i,j-1}),\hat{\mathbf{Y}}_{i,j}(l_{i,j}|l_{i,j-1})) : i\in\mathcal{N}\setminus\mathcal{M}\},\mathbf{Y}_{\pi(k),j})\in A_\epsilon(X_0,X_{\mathcal{M}},X_{\mathcal{N}\setminus\mathcal{M}},\hat{Y}_{\mathcal{N}\setminus\mathcal{M}},Y_{\pi(k)}), \tag{5.34}$$

where $(m_{f(j)},m_{f(j-B)},\ldots,m_{f(j-B^{k-3})},m_{f(j-B^{k-2})},m_{f(j-B^{k-1})},\ldots,m_{f(j-B^M)})$ are corresponding to

$$(m_{g+B^{k-3}},m_{g+B^{k-3}-1},\ldots,m_{g+B^{k-3}-B^{k-4}},m_g,m_{g+B^{k-3}-B^{k-2}},\ldots,m_{g+B^{k-3}-B^{M-1}}). \tag{5.35}$$

Among the messages in (5.35), only $m_g$, corresponding to $m_{f(j-B^{k-2})}$, is the unknown message at node $\pi(k)$ that needs to be decoded. In fact,

$(m_{g+B^{k-3}-B^{k-2}},\ldots,m_{g+B^{k-3}-B^{M-1}})$, corresponding to $(m_{f(j-B^{k-1})},\ldots,m_{f(j-B^M)})$,

have been decoded by block $b-B^{k-1}$, while

$$(m_{g+B^{k-3}},m_{g+B^{k-3}-1},\ldots,m_{g+B^{k-3}-B^{k-4}}),$$

corresponding to $(m_{f(j)},m_{f(j-B)},\ldots,m_{f(j-B^{k-3})})$,

either are dummy messages according to (5.33) (for $g=vB^{k-2}-B^{k-3},\ldots,vB^{k-2}-2B^{k-3}+1$) or have been decoded due to the backward property of decoding (for $g=vB^{k-2}-2B^{k-3},vB^{k-2}-2B^{k-3}-1,\ldots,(v-1)B^{k-2}+1$). Therefore, in (5.34),

$$\{\mathbf{X}_{\pi(s),j}, s=k,k+1,\ldots,M+1\}$$

are known at node $\pi(k)$, while

$$\{\mathbf{X}_{\pi(s),j}, s=1,\ldots,k-1\}$$

are cooperatively transmitting the message $m_g$. Having noted this fact, from [32, Thm 1] and its proof (see also Theorem 4.2.1), we have that $m_g$ can be decoded if

$$R<\min_{\mathcal{S}\subseteq\mathcal{N}\setminus\mathcal{M}} I(X_{(1:k-1)},X_{\mathcal{S}};\hat{Y}_{(\mathcal{N}\setminus\mathcal{M})\setminus\mathcal{S}},Y_{\pi(k)}|X_{(\mathcal{N}\setminus\mathcal{M})\setminus\mathcal{S}},X_{(k:M+1)})$$
$$-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_{(1:M+1)},X_{\mathcal{N}\setminus\mathcal{M}},Y_{\pi(k)},\hat{Y}_{(\mathcal{N}\setminus\mathcal{M})\setminus\mathcal{S}}). \tag{5.36}$$

By considering only a subset $\mathcal{T}_k \subseteq \mathcal{N} \setminus \mathcal{M}$ for the decoding at node $\pi(k)$ while treating the inputs of other C-F relay nodes as purely noise, and using the technique of time sharing, (5.36) can be improved to

$$\begin{aligned} R < \max_{\mathcal{T}_k \subseteq \mathcal{N} \setminus \mathcal{M}} \min_{\mathcal{S} \subseteq \mathcal{T}_k} & I(X_{\pi(1:k-1)}, X_{\mathcal{S}}; \hat{Y}_{\mathcal{T}_k \setminus \mathcal{S}}, Y_{\pi(k)} | X_{\mathcal{T}_k \setminus \mathcal{S}}, X_{\pi(k:M+1)}, Q) \\ & - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_{\pi(1:M+1)}, X_{\mathcal{T}_k}, Y_{\pi(k)}, \hat{Y}_{\mathcal{T}_k \setminus \mathcal{S}}, Q), \end{aligned} \tag{5.37}$$

which proves Theorem 5.2.2.

## 5.5 Rates for the AWGN Two-Relay Channel

With $C(x) := \frac{1}{2}\log_2(1+x)$, various rates are evaluated for the AWGN two-relay channel as follows.

***Multi-level D-F:*** The best achievable D-F rates are the multi-level D-F rates [27]-[30]

$$R_a = \max\{R_{a_1}, R_{a_2}\}$$

where

$$\begin{aligned} R_{a_1} &= \min\{a_{11}, a_{12}, a_{13}\} \\ R_{a_2} &= \min\{a_{21}, a_{22}, a_{23}\} \end{aligned}$$

with

$$\begin{aligned} a_{11} &= I(X_0; Y_1 | X_1, X_2) \\ &= C(g_{01}^2 P) \\ a_{12} &= I(X_0, X_1; Y_2 | X_2) \\ &= C((g_{02}^2 + g_{12}^2)P) \\ a_{13} &= I(X_0, X_1, X_2; Y_3) \\ &= C((g_{03}^2 + g_{13}^2 + g_{23}^2)P) \\ a_{21} &= I(X_0; Y_2 | X_1, X_2) \\ &= C(g_{02}^2 P) \end{aligned}$$

$$
\begin{aligned}
a_{22} &= I(X_0, X_2; Y_1|X_1) \\
&= C((g_{01}^2 + g_{21}^2)P) \\
a_{23} &= I(X_0, X_1, X_2; Y_3) \\
&= C((g_{03}^2 + g_{13}^2 + g_{23}^2)P).
\end{aligned}
$$

***Noisy network coding:*** The best achievable C-F rates are the noisy network coding rates [32]-[34]

$$
R_b = \min\{b_1, b_2, b_3, b_4\}
$$

with

$$
\begin{aligned}
b_1 &= I(X_0; \hat{Y}_1, \hat{Y}_2, Y_3|X_1, X_2) \\
&= C\left(\frac{g_{01}^2 P}{1+\hat{\sigma}_1^2} + \frac{g_{02}^2 P}{1+\hat{\sigma}_2^2} + g_{03}^2 P\right) \\
b_2 &= I(X_0, X_1; \hat{Y}_2, Y_3|X_2) - I(Y_1; \hat{Y}_1|X_0, X_1, X_2, \hat{Y}_2, Y_3) \\
&= C\left((g_{03}^2 + g_{13}^2)P + \frac{(g_{02}^2 + g_{12}^2)P}{1+\hat{\sigma}_2^2} + \frac{(g_{02}g_{13} - g_{12}g_{03})^2 P^2}{1+\hat{\sigma}_2^2}\right) - C\left(\frac{1}{\hat{\sigma}_1^2}\right) \\
b_3 &= I(X_0, X_2; \hat{Y}_1, Y_3|X_1) - I(Y_2; \hat{Y}_2|X_0, X_1, X_2, \hat{Y}_1, Y_3) \\
&= C\left((g_{03}^2 + g_{23}^2)P + \frac{(g_{01}^2 + g_{21}^2)P}{1+\hat{\sigma}_1^2} + \frac{(g_{01}g_{23} - g_{21}g_{03})^2 P^2}{1+\hat{\sigma}_1^2}\right) - C\left(\frac{1}{\hat{\sigma}_2^2}\right) \\
b_4 &= I(X_0, X_1, X_2; Y_3) - I(Y_1, Y_2; \hat{Y}_1, \hat{Y}_2|X_0, X_1, X_2, Y_3) \\
&= C((g_{03}^2 + g_{13}^2 + g_{23}^2)P) - C\left(\frac{1+\hat{\sigma}_1^2+\hat{\sigma}_2^2}{\hat{\sigma}_1^2\hat{\sigma}_2^2}\right)
\end{aligned}
$$

where the optimal $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are determined numerically.

***Our unified schemes:*** The rates under our unified framework are

$$
R_c = \max\{R_a, R_b, R_{c_1}, R_{c_2}\}
$$

where

$$
\begin{aligned}
R_{c_1} &= \min\{\max\{c_{11}, \min\{c_{12}, c_{13}\}\}, \max\{c_{14}, \min\{c_{15}, c_{16}\}\}\} \\
R_{c_2} &= \min\{\max\{c_{21}, \min\{c_{22}, c_{23}\}\}, \max\{c_{24}, \min\{c_{25}, c_{26}\}\}\}
\end{aligned}
$$

with

$$
\begin{aligned}
c_{11} &= I(X_0; Y_1|X_1) \\
&= C\left(\frac{g_{01}^2 P}{1 + g_{21}^2 P}\right) \\
c_{12} &= I(X_0; \hat{Y}_2, Y_1|X_1) \\
&= C\left(g_{01}^2 P + \frac{g_{02}^2 P}{1 + \hat{\sigma}_2^2}\right) \\
c_{13} &= I(X_0, X_2; Y_1|X_1) - I(Y_2; \hat{Y}_2|X_0, X_1, X_2, Y_1) \\
&= C((g_{01}^2 + g_{21}^2)P) - C\left(\frac{1}{\hat{\sigma}_2^2}\right) \\
c_{14} &= I(X_0, X_1; Y_3) \\
&= C\left(\frac{(g_{03}^2 + g_{13}^2)P}{1 + g_{23}^2 P}\right) \\
c_{15} &= I(X_0, X_1; \hat{Y}_2, Y_3|X_2) \\
&= C\left((g_{03}^2 + g_{13}^2)P + \frac{(g_{02}^2 + g_{12}^2)P}{1 + \hat{\sigma}_2^2} + \frac{(g_{02}g_{13} - g_{12}g_{03})^2 P^2}{1 + \hat{\sigma}_2^2}\right) \\
c_{16} &= I(X_0, X_1, X_2; Y_3) - I(Y_2; \hat{Y}_2|X_0, X_1, X_2, Y_3) \\
&= C((g_{03}^2 + g_{13}^2 + g_{23}^2)P) - C\left(\frac{1}{\hat{\sigma}_2^2}\right) \\
c_{21} &= I(X_0; Y_2|X_2) \\
&= C\left(\frac{g_{02}^2 P}{1 + g_{12}^2 P}\right) \\
c_{22} &= I(X_0; \hat{Y}_1, Y_2|X_2) \\
&= C\left(g_{02}^2 P + \frac{g_{01}^2 P}{1 + \hat{\sigma}_1^2}\right) \\
c_{23} &= I(X_0, X_1; Y_2|X_2) - I(Y_1; \hat{Y}_1|X_0, X_1, X_2, Y_2) \\
&= C((g_{02}^2 + g_{12}^2)P) - C\left(\frac{1}{\hat{\sigma}_1^2}\right)
\end{aligned}
$$

$$
\begin{aligned}
c_{24} &= I(X_0, X_2; Y_3) \\
&= C\left(\frac{(g_{03}^2 + g_{23}^2)P}{1 + g_{13}^2 P}\right) \\
c_{25} &= I(X_0, X_2; \hat{Y}_1, Y_3 | X_1) \\
&= C\left((g_{03}^2 + g_{23}^2)P + \frac{(g_{01}^2 + g_{21}^2)P}{1 + \hat{\sigma}_1^2} + \frac{(g_{01}g_{23} - g_{21}g_{03})^2 P^2}{1 + \hat{\sigma}_1^2}\right) \\
c_{26} &= I(X_0, X_1, X_2; Y_3) - I(Y_1; \hat{Y}_1 | X_0, X_1, X_2, Y_3) \\
&= C((g_{03}^2 + g_{13}^2 + g_{23}^2)P) - C\left(\frac{1}{\hat{\sigma}_1^2}\right).
\end{aligned}
$$

***Unified scheme in [30]:*** The rates in [30, Thm 4] (with $U_i = \emptyset, i = 1, 2$) are

$$R_d = \max\{R_a, R_{d_1}, R_{d_2}, R_{d_3}\}$$

with

$$
\begin{aligned}
R_{d_1} &= d_{11} \quad \text{s.t. } d_{12} \le d'_{12}, d_{13} \le d'_{13}, d_{14} \le d'_{14} \\
R_{d_2} &= \min\{d_{21}, d_{22}\} \quad \text{s.t. } d_{23} \le d'_{23} \\
R_{d_3} &= \min\{d_{31}, d_{32}\} \quad \text{s.t. } d_{33} \le d'_{33}
\end{aligned}
$$

where

$$
\begin{aligned}
d_{11} &= I(X_0; \hat{Y}_1, \hat{Y}_2, Y_3 | X_1, X_2) \\
&= b_1 \\
d_{12} &= I(Y_1; \hat{Y}_1 | X_1, X_2, \hat{Y}_2, Y_3) + I(\hat{Y}_1; X_2 | X_1) \\
&= C\left(\frac{1}{\hat{\sigma}_1^2} + \frac{g_{01}^2 P}{\hat{\sigma}_1^2 (1 + \frac{g_{02}^2 P}{1 + \hat{\sigma}_2^2} + g_{03}^2 P)}\right) + C\left(\frac{g_{21}^2 P}{g_{01}^2 P + 1 + \hat{\sigma}_1^2}\right) \\
d'_{12} &= I(X_1; Y_3 | X_2) \\
&= C\left(\frac{g_{13}^2 P}{1 + g_{03}^2 P}\right) \\
d_{13} &= I(Y_2; \hat{Y}_2 | X_1, X_2, \hat{Y}_1, Y_3) + I(\hat{Y}_2; X_1 | X_2) \\
&= C\left(\frac{1}{\hat{\sigma}_2^2} + \frac{g_{02}^2 P}{\hat{\sigma}_2^2 (1 + \frac{g_{01}^2 P}{1 + \hat{\sigma}_1^2} + g_{03}^2 P)}\right) + C\left(\frac{g_{12}^2 P}{g_{02}^2 P + 1 + \hat{\sigma}_2^2}\right)
\end{aligned}
$$

$$
\begin{aligned}
d'_{13} &= I(X_2;Y_3|X_1) \\
&= C\left(\frac{g_{23}^2 P}{1+g_{03}^2 P}\right) \\
d_{14} &= I(Y_1,Y_2;\hat{Y}_1,\hat{Y}_2|X_1,X_2,Y_3) + I(\hat{Y}_1;X_2|X_1) + I(\hat{Y}_2;X_1|X_2) \\
&= C\left(\frac{1+\hat{\sigma}_1^2+\hat{\sigma}_2^2}{\hat{\sigma}_1^2\hat{\sigma}_2^2} + \frac{g_{01}^2(1+\hat{\sigma}_2^2)P + g_{02}^2(1+\hat{\sigma}_1^2)P}{\hat{\sigma}_1^2\hat{\sigma}_2^2(1+g_{03}^2P)}\right) \\
&\quad + C\left(\frac{g_{21}^2 P}{g_{01}^2 P + 1 + \hat{\sigma}_1^2}\right) + C\left(\frac{g_{12}^2 P}{g_{02}^2 P + 1 + \hat{\sigma}_2^2}\right) \\
d'_{14} &= I(X_1,X_2;Y_3) \\
&= C\left(\frac{(g_{13}^2+g_{23}^2)P}{1+g_{03}^2P}\right) \\
d_{21} &= I(X_0;Y_1|X_1) \\
&= c_{11} \\
d_{22} &= I(X_0,X_1;\hat{Y}_2,Y_3|X_2) \\
&= c_{15} \\
d_{23} &= I(Y_2;\hat{Y}_2|X_2,Y_3) \\
&= C\left(\frac{1}{\hat{\sigma}_2^2} + \frac{(g_{02}g_{13}-g_{12}g_{03})^2P^2}{\hat{\sigma}_2^2((g_{03}^2+g_{13}^2)P+1)} + \frac{(g_{02}^2+g_{12}^2)P}{\hat{\sigma}_2^2((g_{03}^2+g_{13}^2)P+1)}\right) \\
d'_{23} &= I(X_2;Y_3) \\
&= C\left(\frac{g_{23}^2P}{1+(g_{03}^2+g_{13}^2)P}\right) \\
d_{31} &= I(X_0;Y_2|X_2) \\
&= c_{21} \\
d_{32} &= I(X_0,X_2;\hat{Y}_1,Y_3|X_1) \\
&= c_{25} \\
d_{33} &= I(Y_1;\hat{Y}_1|X_1,Y_3) \\
&= C\left(\frac{1}{\hat{\sigma}_1^2} + \frac{(g_{01}g_{03}-g_{21}g_{23})^2P^2}{\hat{\sigma}_1^2((g_{03}^2+g_{23}^2)P+1)} + \frac{(g_{01}^2+g_{21}^2)P}{\hat{\sigma}_1^2((g_{03}^2+g_{23}^2)P+1)}\right) \\
d'_{33} &= I(X_1;Y_3) \\
&= C\left(\frac{g_{13}^2P}{1+(g_{03}^2+g_{23}^2)P}\right)
\end{aligned}
$$

***Unified scheme in [42]:*** The rates in [42] are

$$R_e = \max\{R_{e_1}, R_b, R_{e_2}, R_{e_3}\}$$

where

$$\begin{aligned} R_{e_1} &= \min\{e_{11}, e_{12}, e_{13}\} \\ R_{e_2} &= \min\{e_{21}, \max\{e_{22}, \min\{e_{23}, e_{24}\}\}\} \\ R_{e_3} &= \min\{e_{31}, \max\{e_{32}, \min\{e_{33}, e_{34}\}\}\} \end{aligned}$$

with

$$\begin{aligned} e_{11} &= a_{11} \\ e_{12} &= a_{21} \\ e_{13} &= a_{13} \\ e_{21} &= \begin{cases} \min\{c_{12}, c_{13}\} & \text{if} \quad C\left(\frac{1}{\hat{\sigma}_2^2}\right) \leq C\left(\frac{g_{21}^2 P}{1+g_{01}^2 P}\right) \\ c_{11} & \text{otherwise} \end{cases} \\ e_{22} &= c_{14} \\ e_{23} &= c_{15} \\ e_{24} &= c_{16} \\ e_{31} &= \begin{cases} \min\{c_{22}, c_{23}\} & \text{if} \quad C\left(\frac{1}{\hat{\sigma}_1^2}\right) \leq C\left(\frac{g_{12}^2 P}{1+g_{02}^2 P}\right) \\ c_{21} & \text{otherwise} \end{cases} \\ e_{32} &= c_{24} \\ e_{33} &= c_{25} \\ e_{34} &= c_{26}. \end{aligned}$$

***Cut-set bound:*** Finally, the cut-set bound is given by

$$R_f = \min\{f_1, f_2, f_3, f_4\}$$

where

$$
\begin{aligned}
f_1 &= I(X_0; Y_1, Y_2, Y_3 | X_1, X_2) \\
&= C((g_{01}^2 + g_{02}^2 + g_{03}^2)P) \\
f_2 &= I(X_0, X_1; Y_2, Y_3 | X_2) \\
&= C((g_{02}g_{13} - g_{12}g_{03})^2 P^2 + (g_{02}^2 + g_{12}^2 + g_{03}^2 + g_{13}^2)P) \\
f_3 &= I(X_0, X_2; Y_1, Y_3 | X_1) \\
&= C((g_{03}g_{21} - g_{23}g_{01})^2 P^2 + (g_{01}^2 + g_{21}^2 + g_{03}^2 + g_{23}^2)P) \\
f_4 &= I(X_0, X_1, X_2; Y_3) \\
&= C((g_{03}^2 + g_{13}^2 + g_{23}^2)P).
\end{aligned}
$$

# Chapter 6

# Conclusion and Future Work

## 6.1 Conclusion

In C-F relay schemes, joint compression-message decoding introduces more freedom in selecting the compressions at the relays. Motivated by it, we have investigated, in the setup of general multiple-relay channels, the problem of finding the optimal compressions in maximizing the achievable rate of the original message. We have studied several different C-F relay schemes, and the unanimous conclusion is that the optimal compressions should always support successive compression-message decoding. In situations where compressions not supporting successive decoding have to be used, we have found that only those that can be jointly decoded are helpful to the decoding of the original message.

We have also developed a backward block-by-block decoding C-F relay scheme. Compared to the repetitive encoding/all blocks united decoding scheme recently proposed in [32], which improved the achievable rate in the multiple-relay case, we have realized that the key to the improvement comes from delaying the decoding until all the blocks have been finished. In retrospect, the multiple-relay case is different from the single-relay case in that it may take multiple blocks for the relays to help each other before their compressions can finally reach the destination. Hence, the block-by-block forward decoding scheme, which is sufficient for the single-relay case, may not work satisfactorily for multiple relays in general.

Last but not least, we have proposed a unified relay framework with both the D-F and C-F relay nodes for multiple-relay channels. This framework employs nested blocks combined with backward decoding to allow for the full incorporation of the best known D-F and C-F relay strategies. The achievable rates obtained under such a framework turn out to combine both the best known D-F and C-F achievable rates and include them as special cases. It is also demonstrated through a Gaussian network example that our achievable rates are generally better than the rates obtained with existing unified schemes and with D-F or C-F alone.

## 6.2 Future Work

While the recent research on relay channels, including this thesis, centers on developing the achievability coding schemes, little progress has been made for the converse over the past few decades. Motivated by this, the first line of our future research is to study the converse problem for relay channels, especially to study the issue of the optimality of the C-F relay scheme. To this end, we have actually conducted some preliminary investigations [47]-[50]. To continue our research in this direction, particularly, we will study an open problem on the capacity of the relay channel, posed by Cover [51] more than two decades ago.

Consider the relay channel as depicted in Figure 6.1, where the source's input $X$ is received by the relay $Z$ and the destination $Y$ through a channel $p(y,z|x)$, and the relay can communicate to the destination via an error-free digital link with rate $R_0$. We wish to communicate a message index $W \in \{1, 2, \ldots, 2^{TR}\}$ reliably

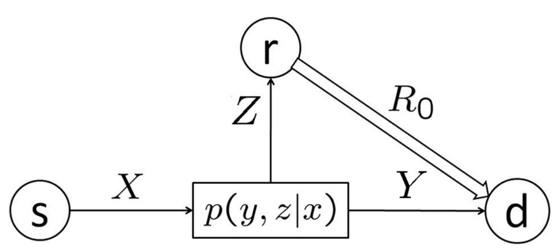



Figure 6.1: A relay channel model with a digital link.

over this relay channel. A $(2^{TR}, T)$ code for this channel consists of

1) an encoding function $X^T : \{1, \ldots, 2^{TR}\} \to \mathcal{X}^T$,

2) a relay function $r : \mathcal{Z}^T \to \{1, \ldots, 2^{TR_0}\}$,

3) a decoding function $g$: $\mathcal{Y}^T \times \{1, \ldots, 2^{TR_0}\} \to \{1, \ldots, 2^{TR}\}$.

The probability of error is defined by

$$P_e^{(T)} = \Pr(g(r(Z^T), Y^T) \neq W),$$

where $W$ is uniformly distributed over $\{1, \ldots, 2^{TR}\}$. The capacity of $C(R_0)$ is the supremum of all achievable rates $R$, i.e., the rates $R$ for which there exists a sequence of $(2^{TR}, T)$ codes such that $P_e^{(T)}$ tends to zero as $T \to \infty$.

We note the following facts:

1) $C(0) = \max_{p(x)} I(X; Y)$.

2) $C(\infty) = \max_{p(x)} I(X; Y, Z)$.

3) $C(R_0)$ is a nondecreasing function of $R_0$.

Instead of demanding a full characterization of $C(R_0)$, Cover posed the question as "what is the critical value of $R_0$ such that $C(R_0)$ first equals $C(\infty)$". Equivalently, we are interested in finding

$$R_0^* := \inf\{R_0 : C(R_0) = C(\infty) = \max_{p(x)} I(X; Y, Z)\},$$

namely, the smallest rate needed for the relay-destination communication so that the maximum information rate $C(\infty) = \max_{p(x)} I(X; Y, Z)$ can be achieved.

It is clear that, by the simple C-F scheme with Slepian-Wolf binning used, the rate

$$R_{\text{conj}} = \min_{p(x): I(X;Y,Z) = C(\infty)} H(Z|Y)$$

is enough such that $C(R_{\text{conj}}) = C(\infty) = \max_{p(x)} I(X; Y, Z)$, but is the C-F scheme optimal in this problem? To answer this question, we have carefully studied the binary symmetric case, i.e., the case where the source-relay and source-destination links are both binary symmetric channels. Although a formal proof is still out of reach at this moment, we strongly believe that in the binary symmetric case, the simple C-F scheme is indeed optimal so that $R_0^* = R_{\text{conj}}$. As our future work, we will continue this research, attempting to first obtain a conclusive result in the binary symmetric case and then extend it to the general case.

The second line of the future work is to further the research in this thesis. One direction is to extend our work on the unified relay framework to multi-source networks. Here, a challenge is how to coordinate the D-F nodes to relay the messages for different sources. In this regard, a multi-source, multi-relay D-F coding scheme has been developed in [43], which also employed backward decoding. It is of interest to study the problem of incorporating the D-F scheme in [43] to our unified relay framework for general multi-source multi-relay multi-destination networks.

Recently, there have been some results on achieving the capacity of the relay networks to within a constant gap based on quantize-map-and-forward [31] and noisy network coding [32]. However, a limitation in these results is that the gap grows with the number of nodes in the network. This may be because both quantize-map-and-forward and noisy network coding belong to the C-F relay schemes in nature, and thus the noise in the these schemes cannot be eliminated as in the D-F relay scheme but will be accumulated. Recall that the unified relay scheme proposed in this thesis combines the best known D-F and C-F relay schemes and endows each node the freedom of choosing either D-F or C-F so that higher and easy-to-evaluate rates can be achieved. Thus, as another direction of the future research, it is worth exploring whether our unified scheme can achieve a better gap or even a universal gap independent of the number of nodes in the network.

# Bibliography


[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pt. I, pp. 379–423, 1948; pt. II, pp. 623–656, 1948.

[2] D. Slepian and J. K.Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Info. Theory*, vol. IT-19, pp. 471–480, Jul. 1973.

[3] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Info. Theory*, vol. 21, pp. 226–228, Mar. 1975.

[4] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Information Theory*, pp. 23–52, Budapest, Hungary: Hungarian Acad. Sci., 1971.

[5] H. Liao, "Multiple access channels," Ph.D. dissertation, Dept. Elec. Eng., Univ. of Hawaii, Honolulu, 1972.

[6] T. M. Cover, "Broadcast channels," *IEEE Trans. Info. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.

[7] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Info. Theory*, vol. IT-19, pp. 197–207, Mar. 1973.

[8] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredacy Informaccii*, vol. 10, no. 3, pp. 3–14, Jul.–Sep. 1974.

[9] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Info. Theory*, vol. IT-20, pp. 279–280, Mar. 1974.

[10] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," *IEEE Trans. Info. Theory*, vol. 52, pp. 3936–3964, Sep. 2006.

[11] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the receiver," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–11, Jan. 1976.

[12] T. S. Han and K. Kobayashi, "New achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 49–60, Jan. 1981.

[13] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[14] E. C. van der Meulen, "Transmission of information in a t-terminal discrete memoryless channel," Ph.D. dissertation, Dept. of Statistics, Univ. of California, Berkeley, 1968.

[15] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[16] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, pp. 1204–1216, July 2000.

[17] A. B. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 6, pp. 841–850, Nov. 1982.

[18] F. M. J. Willems and E. C. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inform. Theory*, vol. IT-31, no. 3, pp. 313–327, May 1985.

[19] L.-L. Xie, "An improvement of Cover/El Gamal's compress-and-forward relay scheme," August 2009, available online at http://arxiv.org/abs/0908.0163.

[20] A. El Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Trans. Inform. Theory*, vol. IT-52, no. 4, pp. 1545–1561, 2006.

[21] Y.-H. Kim, "Coding techniques for primitive relay channels," in *Proc. Forty-Fifth Annual Allerton Conf. Commun., Contr. Comput.*, Monticello, IL, Sep. 2007.

[22] A. El Gamal and Y.-H. Kim, "Lectures notes on network information theory," 2010, available online at http://arxiv.org/abs/1001.3404.

[23] P. Rost and G. Fettweis, "Analysis of a mixed strategy for multiple relay networks," *IEEE Trans. Inform. Theory*, vol. 55, pp. 174–189, Jan. 2009.

[24] M. R. Aref, "Information Flow in Relay Networks," Ph.D. dissertation, Stanford University, Stanford, CA, 1980.

[25] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: An achievable rate region," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1877–1894, Aug. 2003.

[26] A. Reznik, S. R. Kulkarni, and S. Verdú, "Degraded Gaussian multirelay channel: Capacity and optimal power allocation," *IEEE Trans. Inform. Theory*, vol. 50, pp. 3037–3046, Dec. 2004.

[27] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communication: scaling laws and optimal operation," *IEEE Trans. Inform. Theory*, vol. 50, pp. 748–767, May 2004.

[28] L.-L. Xie and P. R. Kumar, " An achievable rate for the multiple-level relay channel," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1348–1358, April 2005.

[29] G. Kramer, M. Gastpar, and P. Gupta, "Capacity theorems for wireless relay channels," in *Proc. 41th Annual Allerton Conf. Commun., Contr. Comput.*, Monticello, IL, Oct. 2003.

[30] G. Kramer, M. Gastpar, and P. Gupta, “Cooperative strategies and capacity theorems for relay networks,” *IEEE Trans. Inform. Theory*, vol. 51, pp. 3037–3063, September 2005.

[31] A. Avestimehr, S. Diggavi, and D. Tse, “Wireless network information flow: A deterministic approach,” *IEEE Trans. Inform. Theory*, vol. 57, no. 4, pp. 1872–1905, April 2011.

[32] S. H. Lim, Y.-H. Kim, A. El Gamal, S.-Y. Chung, “Noisy network coding,” *IEEE Trans. Inform. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[33] X. Wu and L.-L. Xie, “On the optimality of successive decoding in compress-and-forward relay schemes,” in *Proc. 48th Annual Allerton Conf. Commun., Contr. Comput.*, pp. 534–541, Monticello, IL, Sep. 29-Oct. 1, 2010.

[34] X. Wu and L.-L. Xie, “On the optimal compressions in the compress-and-forward relay schemes,” *IEEE Trans. Inform. Theory*, vol. 59, no. 5, pp. 2613-2628, May 2013.

[35] X. Wu and L.-L. Xie, “An optimality-robustness tradeoff in the compress-and-forward relay scheme,” in *Proc. IEEE 72nd Vehicular Technology Conference*, Ottawa, Sep. 2010.

[36] X. Wu and L.-L. Xie, “A unified relay framework with both D-F and C-F relay nodes,” in *Proc. of the 13th Canadian Workshop on Information Theory*, Toronto, Canada, June 18-21, 2013.

[37] X. Wu and L.-L. Xie, “A unified relay framework with both D-F and C-F relay nodes,” to appear in *IEEE Trans. Inform. Theory*, available online at http://arxiv.org/abs/1209.4889.

[38] G. Kramer and J. Hou, “Short-message quantize-forward network coding,” in *Proc. of 8th Int. Workshop on Multi-Carrier Systems & Solutions*, Herrsching, Germany, 2011.

[39] G. Kramer and J. Hou, "On message lengths for noisy network coding," in *Proc. IEEE Inf. Theory Workshop*, Paraty, Brazil, 2011.

[40] A. Raja and P. Viswanath, "Compress-and-forward scheme for relay networks: backword decoding and connection to bisubmodular flows," available online at http://arxiv.org/abs/1012.0416v3.

[41] J. Hou and G. Kramer, "Short message noisy network coding with a decode-forward option," available online at http://arxiv.org/abs/1304.1692.

[42] A. Behboodi and P. Piantanida, "Selective coding strategy for unicast composite networks," in *Proc. of the IEEE International Symposium on Information Theory*, Cambridge, MA, USA, July 2012.

[43] L.-L. Xie and P. R. Kumar, "Multi-source, multi-destination, multi-relay wireless networks," *IEEE Transactions on Information Theory, Special Issue on Models, Theory and Codes for Relaying and Cooperation in Communication Networks*, vol. 53, pp. 3586–3595, October 2007.

[44] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[45] I. Csiszár and J. Körner, *Information Theory: Coding Theorem for Discrete Memoryless Systems*. Academic Press, New York, 1981.

[46] R. G. Gallager, *Information Theory and Reliable Communication*. New York, Wiley, 1968.

[47] X. Wu and L.-L. Xie, "AEP of output when rate is above capacity," in *Proc. of the 11th Canadian Workshop on Information Theory*, Ottawa, Canada, May 13-15, 2009.

[48] X. Wu and L.-L. Xie, "Asymptotic Equipartition Property of output when rate is above capacity," available online at http://arxiv.org/abs/0908.4445.

[49] X. Wu and L.-L. Xie, "AEP of output when rate is above capacity: the Gaussian case," in *Proc. of IEEE International Symposium on Informaiton Theory 2010*, Austin, Texas, U.S.A., June 13-18, 2010.

[50] X. Wu and L.-L. Xie, "Codebook information does not reduce output entropy when rate is above capacity". in *Proc. of 48th Annual Allerton Conference on Communication, Control, and Computing*, Allerton Retreat Center, Monticello, Illinois, Sep. 29 - Oct. 1, 2010.

[51] T. M. Cover, "The capacity of the relay channel," in *Open Problems in Communication and Computation*, T. M. Cover and B. Gopinath, Eds. New York: Springer-Verlag, 1987, pp. 72–73.