On the Optimal Compressions in the Compress-and-Forward Relay Schemes

Xiugang Wu, *Student Member, IEEE*, and Liang-Liang Xie, *Senior Member, IEEE*

Abstract—In the classical compress-and-forward relay scheme developed by Cover and El Gamal, the decoding process operates in a successive way: the destination first decodes the compression of the relay's observation and then decodes the original message of the source. Recently, several modified compress-and-forward relay schemes were proposed, where the destination jointly decodes the compression and the message, instead of successively. Such a modification on the decoding process was motivated by realizing that it is generally easier to decode the compression jointly with the original message, and more importantly, the original message can be decoded even without completely decoding the compression. Thus, joint decoding provides more freedom in choosing the compression at the relay. However, the question remains in these modified compress-and-forward relay schemes—whether this freedom of selecting the compression necessarily improves the achievable rate of the original message. It has been shown by El Gamal and Kim in 2010 that the answer is negative in the single-relay case. In this paper, it is further 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. 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. The above phenomenon is also shown to exist under the repetitive encoding framework recently proposed by Lim *et al.*, which improved the achievable rate in the case of multiple relays. Here, another interesting discovery is that the improvement is not a result of repetitive encoding, but the benefit of delayed decoding after all the blocks have been finished. The same rate is shown to be achievable with the simpler classical encoding process of Cover and El Gamal with a block-by-block backward decoding process.

Index Terms—Backward decoding, compress-and-forward, compression-message joint decoding, compression-message successive decoding, multiple-relay channel.

I. INTRODUCTION

THE relay channel, originally proposed in [1], models a communication scenario where there is a relay node that can help the information transmission between the source and the destination. Two fundamentally different relay strategies have been developed in [2] for such channels, which, depending
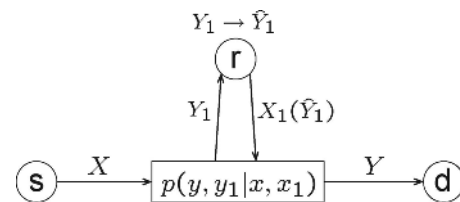
Fig. 1. Single-relay channel.

on whether the relay decodes the information or not, are generally known as *decode-and-forward* and *compress-and-forward*, respectively. The compress-and-forward relay strategy is used when the relay cannot decode the message sent by the source, but still can help by compressing and forwarding its observation to the destination. Specifically, consider the relay channel depicted in Fig. 1. The relay compresses its observation Y_1 into $\hat{Y}_1$ and then forwards $\hat{Y}_1$ to the destination via X_1. To reduce the rate loss caused by the delay, block Markov coding was used in [2], and more blocks lead to less loss.

In this paper, based on the differences in the detailed encoding/decoding processes, the following five different compress-and-forward relay schemes will be considered.

1) Cumulative encoding/block-by-block forward decoding/ compression-message successive decoding.
2) Cumulative encoding/block-by-block forward decoding/ compression-message joint decoding.
3) Repetitive encoding/all blocks united decoding/ compression-message joint decoding.
4) Cumulative encoding/block-by-block backward decoding/ compression-message successive decoding.
5) Cumulative encoding/block-by-block backward decoding/ compression-message joint decoding.

The cumulative encoding/block-by-block forward decoding/ compression-message successive decoding refers to the original compress-and-forward scheme developed in [2]. 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 recently proposed in [3], 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 or by decoding block-by-block backwardly. 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. The compression $\hat{Y}_1$ can be first recovered at the destination, as long as the following constraint is satisfied:

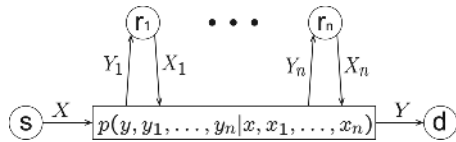$$I(X_1; Y) > I(Y_1; \hat{Y}_1 | X_1, Y). \tag{1}$$

Fig. 2.   Multiple-relay channel.

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

$$R < I(X; \hat{Y}_1, Y | X_1). \tag{2}$$

The above two-step compression-message successive decoding process requires $\hat{Y}_1$ to be decoded first. This facilitates the decoding of X, but is not a requirement of the original problem. Recognizing this, a joint compression-message decoding process was proposed in [4], where, instead of successively, the destination decodes $\hat{Y}_1$ and X together. It turns out that the decoding of X can be helped even if $\hat{Y}_1$ cannot be decoded first. In fact, with joint decoding, constraint (1) is not necessary, and instead of (2), the achievable rate is expressed as

$$R < I(X; \hat{Y}_1, Y | X_1) - \max\{0, I(Y_1; \hat{Y}_1 | X_1, Y) - I(X_1; Y)\}. \tag{3}$$

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 (1), we need a less strict constraint:

$$I(X_1; Y) > I(Y_1; \hat{Y}_1 | X_1, Y, X) \tag{4}$$

where, it is clear to see the assistance provided by X. Similar formulas as (3) have been derived with different arguments in [5]–[7].[1]

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 has been proved in [7] that the answer is negative, and any rate achievable by either of them can always be achieved by the other. In this paper, we are going to further consider the case of multiple relays as depicted in Fig. 2, 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.

Recently, a different encoding process was proposed in [3], where instead of piece by piece, all the information is encoded in each block, and different blocks use independent codebooks to transmit the same information. 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 [3],

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 paper, 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. We will show that either by all blocks united decoding, or by block-by-block backward decoding, the same achievable rate can be obtained. Therefore, in terms of complexity, cumulative encoding/block-by-block backward decoding provides the simplest way to achieve the highest rate in the case of multiple relays.

Similarly, 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 these compress-and-forward relay schemes mentioned above, 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.

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 [3]. 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. Therefore, it is of interest to study how coarser or finer compressions than the optimal affect the achievable rate of the original message [9].

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. Furthermore, to the extreme, when the compression cannot be decoded even with joint decoding, the relay–destination link becomes useless, and the destination would rather simply treat the

---

[1] The formula and proof in [5] missed a Y and were later corrected in [7].

relay's input as purely noise in the decoding, as we will demonstrate in this paper.

The remainder of this paper is organized as the following. In Section II, we formally state our problem setup and summarize the main results. Then, in Sections III and IV, detailed proofs of the achievability results as well as thorough discussions on the optimal choice of the relays' compressions are presented, under the two different frameworks of block-by-block forward decoding and decoding after all the blocks have been finished, respectively. Finally, some concluding remarks are included in Section V.

II. MAIN RESULTS

Consider the multiple-relay channel depicted in Fig. 2, which can be denoted by

$$(\mathcal{X} \times \mathcal{X}_1 \times \cdots \times \mathcal{X}_n,$$
$$p(y, y_1, \ldots, y_n | x, x_1, \ldots, x_n), \mathcal{Y} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n)$$

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

$$x_i(t) = f_{i,t}(y_i(t-1), y_i(t-2), \ldots), \text{ for every time } t \quad (5)$$

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

Before presenting the main results, we introduce some simplified notations. Denote the set $\mathcal{N} = \{1, 2, \ldots, n\}$, and for any subset $\mathcal{S} \subseteq \mathcal{N}$, let $X_{\mathcal{S}} = \{X_i, i \in \mathcal{S}\}$, and use similar notations for other variables. The main results of the paper are presented in the following two different decoding frameworks: 1) block-by-block forward decoding; and 2) decoding after all the blocks have been finished, which includes all blocks united decoding and block-by-block backward decoding.

A. Block-by-Block Forward Decoding

Under the block-by-block forward decoding framework, the achievable rate with successive compression-message decoding and the achievable rate with joint compression-message decoding are presented in Theorems 2.1 and 2.2, respectively. Then, the optimality of successive decoding is stated in Theorem 2.3, and it is shown 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. All the related proofs are presented in Section III.

Theorem 2.1: For the multiple-relay channel depicted in Fig. 2, by the cumulative encoding/block-by-block forward decoding/compression-message successive decoding scheme, a rate $R_{\text{C/F/S}}$ is achievable if for some

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

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

$$\sum_{i \in \mathcal{S}_1} R_i \leq I(X_{\mathcal{S}_1}; Y | X_{\mathcal{S}_1^c}, Q) \quad (6)$$

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

$$I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | \hat{Y}_{\mathcal{S}^c}, Y, X_{\mathcal{N}}, Q) \leq \sum_{i \in \mathcal{S}} R_i \quad (7)$$

and

$$R_{\text{C/F/S}} < I(X; \hat{Y}_{\mathcal{N}}, Y | X_{\mathcal{N}}, Q). \quad (8)$$

Theorem 2.2: For the multiple-relay channel depicted in Fig. 2, by the cumulative encoding/block-by-block forward decoding/compression-message joint decoding scheme, a rate $R_{\text{C/F/J}}$ is achievable if for some

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

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

$$\sum_{i \in \mathcal{S}_1} R_i \leq I(X_{\mathcal{S}_1}; Y | X_{\mathcal{S}_1^c}, Q) \quad (9)$$

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

$$R_{\text{C/F/J}} < I(X; \hat{Y}_{\mathcal{N}}, Y | X_{\mathcal{N}}, Q)$$
$$- I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | \hat{Y}_{\mathcal{S}^c}, Y, X_{\mathcal{N}}, Q) + \sum_{i \in \mathcal{S}} R_i. \quad (10)$$

Let $R_{\text{C/F/S}}^*$ and $R_{\text{C/F/J}}^*$ be the supremum of the achievable rates stated in Theorems 2.1 and 2.2, respectively.

Theorem 2.3: $R_{\text{C/F/S}}^* = R_{\text{C/F/J}}^*$, and $R_{\text{C/F/J}}^*$ can be obtained only when the distribution

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

is chosen such that there exists a rate vector $\{R_i, i = 1, \ldots, n\}$ satisfying (6)–(7).

B. Decoding After all the Blocks Have Been Finished

It was shown in [3] that the original cumulative encoding/block-by-block forward decoding/compression-message successive decoding scheme developed in [2] 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 new compress-and-forward relay scheme [3], 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 Fig. 2, their Theorem 1 in [3] can be restated as the following theorem.

Theorem 2.4: For the multiple-relay channel depicted in Fig. 2, a rate $R_{\mathrm{R/U/J}}$ is achievable if there exists some

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

such that

$$R_{\mathrm{R/U/J}} < \min_{\mathcal{S}\subseteq\mathcal{N}} I(X,X_{\mathcal{S}};\hat{Y}_{\mathcal{S}^c},Y|X_{\mathcal{S}^c},Q)$$
$$- I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X,X_{\mathcal{N}},Y,\hat{Y}_{\mathcal{S}^c},Q). \quad (11)$$

In this paper, 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 2.5: For the multiple-relay channel depicted in Fig. 2, a rate $R_{\mathrm{C/B/S}}$ is achievable if there exists some

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

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

$$I(X_{\mathcal{S}};\hat{Y}_{\mathcal{S}^c},Y|X_{\mathcal{S}^c},Q) - I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_{\mathcal{N}},Y,\hat{Y}_{\mathcal{S}^c},Q) \geq 0 \quad (12)$$

and

$$R_{\mathrm{C/B/S}} < I(X;\hat{Y}_{\mathcal{N}},Y|X_{\mathcal{N}},Q). \quad (13)$$

Let $R^*_{\mathrm{R/U/J}}$ and $R^*_{\mathrm{C/B/S}}$ be the supremum of the achievable rates stated in Theorems 2.4 and 2.5, respectively. The optimality of successive decoding is demonstrated in the following theorem.

Theorem 2.6: $R^*_{\mathrm{R/U/J}} = R^*_{\mathrm{C/B/S}}$, and $R^*_{\mathrm{R/U/J}}$ can be obtained only when the distribution

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

is chosen such that (12) holds.

As mentioned in Section I, 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 2.7: For the multiple-relay channel depicted in Fig. 2, with a given distribution

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

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

$$R_{\mathrm{C/B/J}} < \min_{\mathcal{S}\subseteq\mathcal{D}_{\mathrm{J}}} I(X,X_{\mathcal{S}};\hat{Y}_{\mathcal{D}_{\mathrm{J}}\setminus\mathcal{S}},Y|X_{\mathcal{D}_{\mathrm{J}}\setminus\mathcal{S}},Q)$$
$$- I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X,X_{\mathcal{D}_{\mathrm{J}}},Y,\hat{Y}_{\mathcal{D}_{\mathrm{J}}\setminus\mathcal{S}},Q) \quad (14)$$

where $\mathcal{D}_{\mathrm{J}}$ is the unique largest subset of $\mathcal{N}$ satisfying

$$I(X_{\mathcal{S}};\hat{Y}_{\mathcal{D}_{\mathrm{J}}\setminus\mathcal{S}},Y|X,X_{\mathcal{D}_{\mathrm{J}}\setminus\mathcal{S}},Q)$$
$$- I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X,X_{\mathcal{D}_{\mathrm{J}}},Y,\hat{Y}_{\mathcal{D}_{\mathrm{J}}\setminus\mathcal{S}},Q) > 0 \quad (15)$$

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

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

$$I(X_{\mathcal{S}};\hat{Y}_{\mathcal{D}'_{\mathrm{J}}\setminus\mathcal{S}},Y|X,X_{\mathcal{D}'_{\mathrm{J}}\setminus\mathcal{S}},Q)$$
$$- I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X,X_{\mathcal{D}'_{\mathrm{J}}},Y,\hat{Y}_{\mathcal{D}'_{\mathrm{J}}\setminus\mathcal{S}},Q) \geq 0 \quad (16)$$

for any $\mathcal{S}\subseteq\mathcal{D}'_{\mathrm{J}}$. It will be clear from the proof of Theorem 2.7 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 (11) can be more generally expressed as

$$R_{\mathrm{R/U/J}} < \min_{\mathcal{S}\subseteq\mathcal{M}} I(X,X_{\mathcal{S}};\hat{Y}_{\mathcal{M}\setminus\mathcal{S}},Y|X_{\mathcal{M}\setminus\mathcal{S}},Q)$$
$$- I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X,X_{\mathcal{M}},Y,\hat{Y}_{\mathcal{M}\setminus\mathcal{S}},Q) (17)$$

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 (17), i.e., sometimes, it is better to consider only a subset of relays.

Theorem 2.8: For any $p(q)p(x|q)\prod_{i=1}^{n}p(x_i|q)p(\hat{y}_i|x_i,y_i,q)$, among all the choices of $\mathcal{M}\subseteq\mathcal{N}$, the R.H.S. of (17) is maximized when $\mathcal{M}=\mathcal{D}_{\mathrm{J}}$ or $\mathcal{M}=\mathcal{D}'_{\mathrm{J}}$, but is strictly less than the maximum when $\mathcal{M}\not\subseteq\mathcal{D}'_{\mathrm{J}}$. Here, $\mathcal{D}_{\mathrm{J}}$ and $\mathcal{D}'_{\mathrm{J}}$ are defined as in (15) and (16).

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 (17), i.e., choosing $\mathcal{M}\not\subseteq\mathcal{D}'_{\mathrm{J}}$, will even strictly lower the achievable rate.

By comparing (14) and (17) with $\mathcal{M}=\mathcal{D}_{\mathrm{J}}$, Theorem 2.8 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.

The proofs of Theorems 2.5–2.8 are presented in Section IV.

III. BLOCK-BY-BLOCK FORWARD DECODING

We first prove the achievability results stated in Theorems 2.1 and 2.2 respectively. 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 [7], [12]. The same consideration on Q applies throughout all the achievability proofs of this paper.

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 [2]. The difference between these two schemes is only on the decoding process at the destination: 1) 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 sequence received, and then finds the unique X sequence that is jointly typical with the Y sequence received, and also with the previously recovered $\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n$ sequences. 2) In joint decoding, the destination finds the unique X sequence that is jointly typical with the Y 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 $X_1, X_2, \ldots, X_n$.

A. Proof of Theorem 2.1

The basic idea of the compress-and-forward strategy 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 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.

The encoding process is exactly the same as that in the proof of Theorem 6 of [2]. We only emphasize that the ith 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 $\mathcal{S}_1 \subseteq \mathcal{N}$,

$$\sum_{i \in \mathcal{S}_1} R_i < I(X_{\mathcal{S}_1}; Y | X_{\mathcal{S}_1^c}). \tag{18}$$

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

$$\Big(\underline{Y}(b-1), (\underline{X}_1(b-2), \underline{\hat{Y}}_1(b-1)), \ldots, $$
$$(\underline{X}_n(b-2), \underline{\hat{Y}}_n(b-1)) \Big) \in A_\epsilon(Y, X_{\mathcal{N}}, \hat{Y}_{\mathcal{N}}) \tag{19}$$

where $\underline{Y}(b-1)$ is the Y 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 (18) is satisfied.

Error occurs if the true $\underline{\hat{Y}}_{\mathcal{N}}(b-1)$ does not satisfy (19), or a false $\underline{\hat{Y}}_{\mathcal{N}}(b-1)$ satisfies (19). According to the properties of typical sequences, the true $\underline{\hat{Y}}_{\mathcal{N}}(b-1)$ satisfies (19) 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 $\underline{Y}(b-1)$ and $(\underline{X}_1(b-2), \ldots, \underline{X}_n(b-2))$ can be upper bounded by

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

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

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

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

$$H(\hat{Y}_{\mathcal{S}}|Y, \hat{Y}_{\mathcal{S}^c}, X_{\mathcal{N}}) - \sum_{i \in \mathcal{S}}[H(\hat{Y}_i|Y_i, X_i) + R_i] < 0. \tag{20}$$

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

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

Plugging this into (20), we have $\underline{\hat{Y}}_{\mathcal{N}}(b-1)$ can be decoded at the end of block b if

$$I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}}|\hat{Y}_{\mathcal{S}^c}, Y, X_{\mathcal{N}}) < \sum_{i \in \mathcal{S}} R_i. \tag{21}$$

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

$$R_{\mathrm{C/F/S}} < I(X; \hat{Y}_{\mathcal{N}}, Y | X_{\mathcal{N}}). \tag{22}$$

Combining (18) and (21)–(22), and using the standard technique of time sharing, we conclude that the rate stated in Theorem 2.1 is achievable.[2]

B. Proof of Theorem 2.2

In cumulative encoding/block-by-block forward decoding/compression-message joint decoding, the encoding part is

---

[2]The case of "=" in (6)–(7) can be included since (8) does not include "=." The same consideration applies throughout the paper.

exactly the same as that in the proof of Theorem 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 $\underline{X}(w)$ 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

$$
\begin{aligned}
\Big(\underline{X}(w), \underline{Y}(b-1), (\underline{X}_1(b-2), \underline{\hat{Y}}_1(b-1)), \ldots, \\
(\underline{X}_n(b-2), \underline{\hat{Y}}_n(b-1)) \Big) \in A_\epsilon(X, Y, X_\mathcal{N}, \hat{Y}_\mathcal{N}) \quad (23)
\end{aligned}
$$

where $\underline{Y}(b-1)$, $(\underline{\hat{Y}}_1(b-1), \ldots, \underline{\hat{Y}}_n(b-1))$, and $(\underline{X}_1(b-2), \ldots, \underline{X}_n(b-2))$ have the same interpretations as in (19).

Error occurs if the true $\underline{X}(w)$ does not satisfy (23), or a false $\underline{X}(w')$ satisfies (23). According to the properties of typical sequences, the true $\underline{X}(w)$ satisfies (23) with high probability.

The probability of a false $\underline{X}(w')$ being jointly typical with $\underline{Y}(b-1)$, $(\underline{X}_1(b-2), \ldots, \underline{X}_n(b-2))$, 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

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

There are $2^{T R_{\mathrm{C/F/J}}} - 1$ 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

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

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

$$
R_{\mathrm{C/F/J}} < I(X; \hat{Y}_\mathcal{N}, Y | X_\mathcal{N}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S} | \hat{Y}_{\mathcal{S}^c}, Y, X_\mathcal{N}) + \sum_{i \in \mathcal{S}} R_i. \quad (24)
$$

This combined with the technique of time sharing proves Theorem 2.2.

C. Optimality of Successive Decoding in Block-by-Block Forward Decoding

Before proceeding to the proof of Theorem 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

$$
I_{\mathcal{A}, \mathcal{B}}(\mathcal{S}) := \sum_{i \in \mathcal{S}} R_i - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S} | \hat{Y}_\mathcal{A}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}}, Y, X_\mathcal{N}), \forall \mathcal{S} \subseteq \mathcal{B} \quad (25)
$$

$$
\begin{aligned}
I_\mathcal{B}(\mathcal{S}) &:= I_{\emptyset, \mathcal{B}}(\mathcal{S}) \\
&= \sum_{i \in \mathcal{S}} R_i - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S} | \hat{Y}_{\mathcal{B} \setminus \mathcal{S}}, Y, X_\mathcal{N}), \forall \mathcal{S} \subseteq \mathcal{B} \quad (26)
\end{aligned}
$$

$$
\begin{aligned}
I(\mathcal{S}) &:= I_\mathcal{N}(\mathcal{S}) \\
&= \sum_{i \in \mathcal{S}} R_i - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S} | \hat{Y}_{\mathcal{S}^c}, Y, X_\mathcal{N}), \forall \mathcal{S} \subseteq \mathcal{N}. \quad (27)
\end{aligned}
$$

Also, in the following proof and the rest of the paper, 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 are given in Appendix A.

Lemma 3.1: 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 3.2: For any $p(x) \prod_{i=1}^n p(x_i) p(\hat{y}_i | y_i, x_i)$ 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 3.3: 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: For any $\mathcal{A}$ and $\mathcal{B}$ with $\mathcal{A} \cap \mathcal{B} = \emptyset$, $I(\mathcal{A}) + I(\mathcal{B}) = I(\mathcal{A} \cup \mathcal{B}) + I(\hat{Y}_\mathcal{A}; \hat{Y}_\mathcal{B} | \hat{Y}_{(\mathcal{A} \cup \mathcal{B})^c}, Y, X_\mathcal{N})$.

We are now ready to prove Theorem 2.3. Still for simplicity of notation, we only prove Theorem 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 2.6 and 2.8.

Proof of Theorem 2.3: With $Q = \emptyset$, $R^*_{\mathrm{C/F/S}}$ and $R^*_{\mathrm{C/F/J}}$ can be, respectively, written as

$$
R^*_{\mathrm{C/F/S}}
$$
$$
= \max_{p(x) \prod_{i=1}^n p(x_i) p(\hat{y}_i | y_i, x_i), \{R_i, i=1,\ldots,n\}} I(X; \hat{Y}_\mathcal{N}, Y | X_\mathcal{N}) \quad (28)
$$

such that
$$
\sum_{i \in \mathcal{S}_1} R_i \leq I(X_{\mathcal{S}_1}; Y | X_{\mathcal{S}_1^c}), \forall \mathcal{S}_1 \subseteq \mathcal{N} \quad (29)
$$

$$
I(Y_\mathcal{S}; \hat{Y}_\mathcal{S} | \hat{Y}_{\mathcal{S}^c}, Y, X_\mathcal{N}) - \sum_{i \in \mathcal{S}} R_i \leq 0, \forall \mathcal{S} \subseteq \mathcal{N} \quad (30)
$$

and

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

such that
$$
\sum_{i \in \mathcal{S}_1} R_i \leq I(X_{\mathcal{S}_1}; Y | X_{\mathcal{S}_1^c}), \forall \mathcal{S}_1 \subseteq \mathcal{N}.
$$

We show $R^*_{\mathrm{C/F/S}} = R^*_{\mathrm{C/F/J}}$ by showing that $R^*_{\mathrm{C/F/S}} \leq R^*_{\mathrm{C/F/J}}$ and $R^*_{\mathrm{C/F/S}} \geq R^*_{\mathrm{C/F/J}}$, respectively. For any $p(x) \prod_{i=1}^n p(x_i) p(\hat{y}_i | y_i, x_i)$ and $\{R_i, i = 1, \ldots, n\}$ satisfying (29)—(30), we have

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

$$
= I(X; \hat{Y}_\mathcal{N}, Y | X_\mathcal{N})
$$

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

To show $R^*_{C/F/S} \geq R^*_{C/F/J}$, it is sufficient to show that $R^*_{C/F/J}$ can be achieved only with $p(x)\prod_{i=1}^n p(x_i)p(\hat{y}_i|y_i,x_i)$ and $\{R_i, i=1,\ldots,n\}$ such that $I(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{N}$. We will show this by two steps as follows: i) We first show that for any $p(x)\prod_{i=1}^n p(x_i)p(\hat{y}_i|y_i,x_i)$ and $\{R_i, i=1,\ldots,n\}$, if $\mathcal{D}^c \neq \emptyset$, then $\mathcal{D}^c \in \arg\min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})$ and $\bigcap_{\mathcal{T}\in\arg\min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})} \mathcal{T} = \mathcal{D}^c$, where $\mathcal{D}$ is defined as in Lemma 3.2 and $\arg\min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S}) :=$ $\{\mathcal{T} \subseteq \mathcal{N} : I(\mathcal{T}) = \min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})\}$. ii) We then argue that under the optimal choice of $p(x)\prod_{i=1}^n p(x_i)p(\hat{y}_i|y_i,x_i)$ 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 $\mathcal{D}$, $I(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{N}$.

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

1) We first show $I(\mathcal{D}^c) < 0$ by using a contradiction argument. Suppose $I(\mathcal{D}^c) \geq 0$, i.e., $I_{\mathcal{D},\mathcal{D}^c}(\mathcal{D}^c) \geq 0$. Then, by Lemma 3.3, we have that there exists some nonempty $\mathcal{B} \subseteq \mathcal{D}^c$ such that $I_{\mathcal{D},\mathcal{B}}(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{B}$. This will further imply, by Part 2) of Lemma 3.1, that $I_{\mathcal{D}\cup\mathcal{B}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{D}\cup\mathcal{B}$. This is contradictory with the definition of $\mathcal{D}$, and thus $I(\mathcal{D}^c) < 0$.

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.,

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

Again by Lemmas 3.3 and 3.1 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, that

$$I(\mathcal{A}) = I(\mathcal{A}_1\cup\mathcal{A}_2)$$
$$= I(\mathcal{A}_1) + I(\mathcal{A}_2) - I(\hat{Y}_{\mathcal{A}_1}; \hat{Y}_{\mathcal{A}_2}|\hat{Y}_{\mathcal{A}^c}, Y, X_\mathcal{N}),$$
$$I(\mathcal{A}_1\cup\mathcal{D}^c) = I(\mathcal{A}_1) + I(\mathcal{D}^c)$$
$$- I(\hat{Y}_{\mathcal{A}_1}; \hat{Y}_{\mathcal{D}^c}|\hat{Y}_{(\mathcal{A}_1\cup\mathcal{D}^c)^c}, Y, X_\mathcal{N}).$$

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

$$I(\hat{Y}_{\mathcal{A}_1}; \hat{Y}_{\mathcal{D}^c}|\hat{Y}_{(\mathcal{A}_1\cup\mathcal{D}^c)^c}, Y, X_\mathcal{N})$$
$$= I(\hat{Y}_{\mathcal{A}_1}; \hat{Y}_{\mathcal{D}^c\setminus\mathcal{A}_2}|\hat{Y}_{(\mathcal{A}_1\cup\mathcal{D}^c)^c}, Y, X_\mathcal{N})$$
$$+ I(\hat{Y}_{\mathcal{A}_1}; \hat{Y}_{\mathcal{A}_2}|\hat{Y}_{(\mathcal{A}_1\cup\mathcal{D}^c)^c}, \hat{Y}_{\mathcal{D}^c\setminus\mathcal{A}_2}, Y, X_\mathcal{N})$$
$$= I(\hat{Y}_{\mathcal{A}_1}; \hat{Y}_{\mathcal{A}_2}|\hat{Y}_{\mathcal{A}^c}, Y, X_\mathcal{N})$$
$$+ I(\hat{Y}_{\mathcal{A}_1}; \hat{Y}_{\mathcal{D}^c\setminus\mathcal{A}_2}|\hat{Y}_{(\mathcal{A}_1\cup\mathcal{D}^c)^c}, Y, X_\mathcal{N})$$
$$\geq I(\hat{Y}_{\mathcal{A}_1}; \hat{Y}_{\mathcal{A}_2}|\hat{Y}_{\mathcal{A}^c}, Y, X_\mathcal{N})$$

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

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

Combining 2)–4), we can conclude that $\mathcal{D}^c \in \arg\min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})$ and $\bigcap_{\mathcal{T}\in\arg\min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})} \mathcal{T} = \mathcal{D}^c$.

ii) We now argue that under the optimal choice of $p(x)\prod_{i=1}^n p(x_i)p(\hat{y}_i|y_i,x_i)$ and $\{R_i, i=1,\ldots,n\}$ that achieves $R^*_{C/F/J}$, if $\mathcal{D}^c \neq \emptyset$, then $R^*_{C/F/J}$ is not optimal; and hence $\mathcal{D}^c$ must be $\emptyset$. The argument is extended from that in [7] and the detailed analysis is as follows.

Suppose $\mathcal{D}^c \neq \emptyset$ at the optimum. Then, $\mathcal{D}^c \in \arg\min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})$ and $\bigcap_{\mathcal{T}\in\arg\min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})} \mathcal{T} = \mathcal{D}^c$. Therefore

$$R^*_{C/F/J} = I(X; \hat{Y}_\mathcal{N}, Y|X_\mathcal{N}) + I(\mathcal{D}^c)$$
$$= I(X; \hat{Y}_\mathcal{D}, Y|X_\mathcal{N}) + I(X; \hat{Y}_{\mathcal{D}^c}|\hat{Y}_\mathcal{D}, Y, X_\mathcal{N})$$
$$+ \sum_{i\in\mathcal{D}^c} R_i - I(X, Y_{\mathcal{D}^c}; \hat{Y}_{\mathcal{D}^c}|\hat{Y}_\mathcal{D}, Y, X_\mathcal{N})$$
$$= I(X; \hat{Y}_\mathcal{D}, Y|X_\mathcal{N}) + \sum_{i\in\mathcal{D}^c} R_i$$
$$- I(Y_{\mathcal{D}^c}; \hat{Y}_{\mathcal{D}^c}|X, \hat{Y}_\mathcal{D}, Y, X_\mathcal{N}) \qquad (31)$$

and similarly

$$R^*_{C/F/J} = I(X; \hat{Y}_\mathcal{N}, Y|X_\mathcal{N}) + I(\mathcal{T})$$
$$= I(X; \hat{Y}_{\mathcal{T}^c}, Y|X_\mathcal{N}) + \sum_{i\in\mathcal{T}} R_i$$
$$- I(Y_\mathcal{T}; \hat{Y}_\mathcal{T}|X, \hat{Y}_{\mathcal{T}^c}, Y, X_\mathcal{N}) \qquad (32)$$

for any $\mathcal{T} \in \arg\min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})$, $\mathcal{T} \neq \mathcal{D}^c$.

We argue that higher rate can be achieved. Consider $\hat{Y}_1', \hat{Y}_2', \ldots, \hat{Y}_n'$, 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 $R^*_{C/F/J}$. As p decreases from 1, it can be seen from (31) and (32) that both $I(X; \hat{Y}_\mathcal{N}', Y|X_\mathcal{N}) + I(\mathcal{D}^c)$ and $I(X; \hat{Y}_\mathcal{N}', Y|X_\mathcal{N}) + I(\mathcal{T})$ will increase, where $\mathcal{T} \in \arg\min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})$, $\mathcal{T} \neq \mathcal{D}^c$. Thus, no matter how $I(X; \hat{Y}_\mathcal{N}', Y|X_\mathcal{N}) + I(\mathcal{S})$ will change as p decreases for $\mathcal{S} \notin \arg\min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})$, 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 $R^*_{C/F/J}$. This is in contradiction with the optimality of $R^*_{C/F/J}$, 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 2.3. ∎

IV. Decoding After all Blocks Have Been Finished

In this section, our discussion transfers to the compress-and-forward schemes with decoding after all blocks have been finished. The focus here is on the cumulative encoding/block-by-block backward decoding, since it is the simplest scheme to achieve the highest rate in the general multiple-relay channel, as mentioned before; for the repetitive encoding/all blocks united decoding, see the proof of Theorem 1 in [3].

Cumulative encoding/block-by-block backward decoding can be combined with either compression-message successive decoding or compression-message joint decoding. In the following, we will first present the cumulative encoding/block-by-block backward decoding/compression-message successive decoding scheme to establish the achievable rate in Theorem 2.5 and demonstrate the optimality of successive decoding in the sense of Theorem 2.6. Then, the cumulative encoding/block-by-block backward decoding/compression-message joint decoding scheme will be used to prove Theorem 2.7, and the necessity of joint decodability is demonstrated in the sense that only those relay nodes, whose compressions can be eventually decoded by joint decoding, are helpful to the decoding of the original message.

A. 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 [2] (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 2.5 as follows.

Proof of Theorem 2.5: 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 Bth 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 $\frac{M}{B+M} R_{C/B/S}$ can be made arbitrarily small.

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

For each block $b \in [1 : B + M]$, randomly and independently generate $2^{T R_{C/B/S}}$ sequences $\mathbf{x}_b(m_b), m_b \in [1 : 2^{T R_{C/B/S}}]$; 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 $b \in [1 : B + M]$:

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

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 $\mathbf{x}_b(m_b)$ 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 $\mathbf{l}^B = (\mathbf{l}_1, \ldots, \mathbf{l}_B)$ and some $\mathbf{l}_{B+1}^{B+M} = (\mathbf{l}_{B+1}, \ldots, \mathbf{l}_{B+M})$, where $\mathbf{l}_b = (l_{1,b}, \ldots, l_{n,b})$, $\forall b \in [1 : B + M]$, such that for any $b = 1, \ldots, B + M$

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

$$(\mathbf{X}_{n,b}(l_{n,b-1}), \hat{\mathbf{Y}}_{n,b}(l_{n,b} | l_{n,b-1})), \mathbf{Y}_b \Big) \in A_\epsilon(X_\mathcal{N}, \hat{Y}_\mathcal{N}, Y). \quad (33)$$

Specifically, this can be done backwards as follows:

a) The destination finds the unique $\mathbf{l}_B$ such that there exists some $\mathbf{l}_{B+1}^{B+M} = (\mathbf{l}_{B+1}, \ldots, \mathbf{l}_{B+M})$ satisfying (33) for any $b = B + 1, \ldots, B + M$.

Assume the true $\mathbf{l}_B^{B+M} = \mathbf{1}^{M+1}$, where $\mathbf{1} := (1, \ldots, 1)$ is an n-dimensional all-ones vector. Then, error occurs if $\mathbf{l}_B = \mathbf{1}$ does not satisfy (33) with any $\mathbf{l}_{B+1}^{B+M}$ for any $b = B+1, \ldots, B + M$, or a false $\mathbf{l}_B \neq \mathbf{1}$ satisfies (33) with some $\mathbf{l}_{B+1}^{B+M}$ for any $b = B + 1, \ldots, B + M$. Since $\mathbf{l}_B^{B+M} = \mathbf{1}^{M+1}$ satisfies (33) for any $b = B + 1, \ldots, B + M$ 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 (33) with some $\mathbf{l}_{B+1}^{B+M}$ for any $b = B + 1, \ldots, B + M$. 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 (33). Then, we have

$$\Pr\Big(\bigcup_{\mathbf{l}_B \neq \mathbf{1}} \mathcal{E}_{\mathbf{l}_B} \Big)$$

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

$$= \Pr\Big(\bigcup_{j=1}^{M-1} \bigcup_{\mathbf{l}_{B+1}^{B+M} : \mathbf{l}_{B+j} = \mathbf{1}}$$

$$\bigcup_{\substack{\mathbf{l}_{B+1}^{B+M} : \mathbf{l}_{B+j} \neq \mathbf{1} \\ \forall j \in [1:M-1]}} \bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcap_{b=B+1}^{B+M} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b) \Big)$$

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

$$+ \Pr\Big(\bigcup_{\substack{\mathbf{l}_{B+1}^{B+M} : \mathbf{l}_{B+j} \neq \mathbf{1} \\ \forall j \in [1:M-1]}} \bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcap_{b=B+1}^{B+M} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b) \Big). \quad (34)$$

Let us first consider the second term in (34). 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

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

where $\mathcal{I}(\mathcal{S}_b(\mathbf{l}_{b-1})) = I(X_{\mathcal{S}_b(\mathbf{l}_{b-1})}; \hat{Y}_{\mathcal{S}_b^c(\mathbf{l}_{b-1})}, Y|X_{\mathcal{S}_b^c(\mathbf{l}_{b-1})}) - H(\hat{Y}_{\mathcal{S}_b(\mathbf{l}_{b-1})}|X_\mathcal{N}, \hat{Y}_{\mathcal{S}_b^c(\mathbf{l}_{b-1})}, Y) + \sum_{i \in \mathcal{S}_b(\mathbf{l}_{b-1})} H(\hat{Y}_i|X_i)$ and $\epsilon' \to 0$ as $\epsilon \to 0$. Then, we have inequality (36), given at the bottom of the page, where $\epsilon'' \to 0$ as $\epsilon \to 0$. Thus, as both T and M go to infinity, the second term in (34) goes to 0, if for any nonempty $\mathcal{S} \subseteq \mathcal{N}$, get

$$I(X_\mathcal{S}; \hat{Y}_{\mathcal{S}^c}, Y|X_{\mathcal{S}^c}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S}|X_\mathcal{N}, Y, \hat{Y}_{\mathcal{S}^c}) > 0. \quad (35)$$

Now consider the first term in (34). For any $j \in [1 : M-1]$, we have

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

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 (33) 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

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

$$\Pr\left(\bigcup_{\substack{\mathbf{l}_{B+1}^{B+M} : \mathbf{l}_{B+j} \neq \mathbf{1}, \\ \forall j \in [1:M-1]}} \bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcap_{b=B+1}^{B+M} \mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b) \right) \leq \sum_{\substack{\mathbf{l}_{B+1}^{B+M} : \mathbf{l}_{B+j} \neq \mathbf{1}, \\ \forall j \in [1:M-1]}} \sum_{\mathbf{l}_B \neq \mathbf{1}} \prod_{b=B+1}^{B+M} \Pr(\mathcal{A}_b(\mathbf{l}_{b-1}, \mathbf{l}_b))$$

$$\leq \sum_{\mathbf{l}_{B+M}} \sum_{\substack{\mathbf{l}_{B+1}^{B+M-1} : \mathbf{l}_{B+j} \neq \mathbf{1}, \\ \forall j \in [1:M-1]}} \sum_{\mathbf{l}_B \neq \mathbf{1}} \prod_{b=B+1}^{B+M} 2^{-T(\mathcal{I}(\mathcal{S}_b(\mathbf{l}_{b-1}))-\epsilon')}$$

$$= \sum_{\mathbf{l}_{B+M}} \sum_{\substack{\mathcal{S}_{B+1},\ldots,\mathcal{S}_{B+M} : \\ \mathcal{S}_{B+j} \neq \emptyset, \forall j \in [1:M]}} \sum_{\substack{\mathbf{l}_B^{B+M-1} : \mathcal{S}_b(\mathbf{l}_B^{B+M-1})=\mathcal{S}_b, \\ \forall b \in [B+1:B+M]}} \prod_{b=B+1}^{B+M} 2^{-T(\mathcal{I}(\mathcal{S}_b(\mathbf{l}_{b-1}))-\epsilon')}$$

$$\leq \sum_{\mathbf{l}_{B+M}} \sum_{\substack{\mathcal{S}_{B+1},\ldots,\mathcal{S}_{B+M} : \\ \mathcal{S}_{B+j} \neq \emptyset, \forall j \in [1:M]}} \prod_{b=B+1}^{B+M} 2^{T(\sum_{i \in \mathcal{S}_b}(I(Y_i; \hat{Y}_i|X_i)+\epsilon))} \prod_{b=B+1}^{B+M} 2^{-T(\mathcal{I}(\mathcal{S}_b)-\epsilon')}$$

$$\leq \sum_{\mathbf{l}_{B+M}} \sum_{\substack{\mathcal{S}_{B+1},\ldots,\mathcal{S}_{B+M} : \\ \mathcal{S}_{B+j} \neq \emptyset, \forall j \in [1:M]}} \prod_{b=B+1}^{B+M} 2^{-T(I(X_{\mathcal{S}_b}; \hat{Y}_{\mathcal{S}_b^c}, Y|X_{\mathcal{S}_b^c}) - I(Y_{\mathcal{S}_b}; \hat{Y}_{\mathcal{S}_b}|X_\mathcal{N}, Y, \hat{Y}_{\mathcal{S}_b^c})-\epsilon'')}$$

$$= \sum_{\mathbf{l}_{B+M}} \sum_{\substack{\mathcal{S}_{B+1},\ldots,\mathcal{S}_{B+M} : \\ \mathcal{S}_{B+j} \neq \emptyset, \forall j \in [1:M]}} 2^{-T\sum_{b=B+1}^{B+M}(I(X_{\mathcal{S}_b}; \hat{Y}_{\mathcal{S}_b^c}, Y|X_{\mathcal{S}_b^c}) - I(Y_{\mathcal{S}_b}; \hat{Y}_{\mathcal{S}_b}|X_\mathcal{N}, Y, \hat{Y}_{\mathcal{S}_b^c})-\epsilon'')}$$

$$\leq \sum_{\mathbf{l}_{B+M}} (2^n)^M 2^{-TM(\min_{\mathcal{S} \subseteq \mathcal{N} : \mathcal{S} \neq \emptyset}\{I(X_\mathcal{S}; \hat{Y}_{\mathcal{S}^c}, Y|X_{\mathcal{S}^c}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S}|X_\mathcal{N}, Y, \hat{Y}_{\mathcal{S}^c})-\epsilon''\})}$$

$$\leq 2^{T(\sum_{i \in \mathcal{N}}(I(\hat{Y}_i; Y_i|X_i)+\epsilon))} 2^{nM} 2^{-TM(\min_{\mathcal{S} \subseteq \mathcal{N} : \mathcal{S} \neq \emptyset}\{I(X_\mathcal{S}; \hat{Y}_{\mathcal{S}^c}, Y|X_{\mathcal{S}^c}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S}|X_\mathcal{N}, Y, \hat{Y}_{\mathcal{S}^c})-\epsilon''\})} \quad (36)$$

Then, we have

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

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

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

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

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

$$
=: p_{B+k-1} + p'_{B+k},
$$

where

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

and especially

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

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

$$
p_{B+j} \leq p_{B+j-1} + p'_{B+j} \leq p_{B+j-2} + p'_{B+j-1} + p'_{B+j}
$$

$$
\leq \cdots \leq \sum_{k=1}^{k=j} p'_{B+k}.
$$

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

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

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

$$
\leq \sum_{\mathcal{S}_{B+k}\neq\emptyset} \sum_{\substack{\mathbf{l}_{B+k}=\mathbf{1},\\ \mathbf{l}_{B+k-1}:\mathcal{S}_{B+k}(\mathbf{l}_{B+k-1})=\mathcal{S}_{B+k}}} \Pr(\mathcal{A}_{B+k}(\mathbf{l}_{B+k-1},\mathbf{l}_{B+k}))
$$

$$
\leq \sum_{\mathcal{S}_{B+k}\neq\emptyset} 2^{T(\sum_{i\in\mathcal{S}_{B+k}}(I(Y_i;\hat{Y}_i|X_i)+\epsilon))} 2^{-T(\mathcal{I}(\mathcal{S}_{B+k})-\epsilon')}
$$

$$
\leq 2^n 2^{-T(\min_{\mathcal{S}\subseteq\mathcal{N}:\mathcal{S}\neq\emptyset}\{I(X_\mathcal{S};\hat{Y}_{\mathcal{S}^c},Y|X_{\mathcal{S}^c})-I(Y_\mathcal{S};\hat{Y}_\mathcal{S}|X_\mathcal{N},Y,\hat{Y}_{\mathcal{S}^c})-\epsilon''\})}
$$

and thus $p_{B+j} \to 0$ as $T \to \infty$ if (35) holds. Therefore, if (35) holds, the first term in (34) 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 (33), 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 (33), or a false $\mathbf{l}_{b-1}$ satisfies (33). According to the properties of typical sequences, the true $\mathbf{l}_{b-1}$ satisfies (33) 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 (33) can be upper bounded by

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

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 (35) holds. This combined with a) proves that $\mathbf{l}^B$ can be decoded, if (35) 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 $b = 1, \ldots, B$,

$$
\Big(\mathbf{X}_b(m_b),(\mathbf{X}_{1,b}(l_{1,b-1}),\hat{\mathbf{Y}}_{1,b}(l_{1,b}|l_{sss1,b-1})),\ldots,
$$
$$
(\mathbf{X}_{n,b}(l_{n,b-1}),\hat{\mathbf{Y}}_{n,b}(l_{n,b}|l_{n,b-1})),\mathbf{Y}_b\Big) \in A_\epsilon(X,X_\mathcal{N},\hat{Y}_\mathcal{N},Y). \tag{37}
$$

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 $\mathbf{Y}_b$ in (37) are known to the destination. Thus, from the property of typical sequences, the probability of decoding error will tend to zero if $R_{C/B/S}$ is less than $I(X;X_\mathcal{N},\hat{Y}_\mathcal{N},Y)$, which is equal to $I(X;\hat{Y}_\mathcal{N},Y|X_\mathcal{N})$ noting the independence between X and $X_\mathcal{N}$. ∎

We are now in a position to prove Theorem 2.6. To facilitate the proof, we introduce some notations and lemmas. For any $\mathcal{A},\mathcal{B} \subseteq \mathcal{N}$, let

$$
J_{\mathcal{A},\mathcal{B}}(\mathcal{S}) := I(X_\mathcal{S};\hat{Y}_{\mathcal{B}\setminus\mathcal{S}},\hat{Y}_\mathcal{A},Y|X_\mathcal{A},X_{\mathcal{B}\setminus\mathcal{S}})
$$
$$
- I(Y_\mathcal{S};\hat{Y}_\mathcal{S}|X_\mathcal{A},\hat{Y}_\mathcal{A},Y,X_\mathcal{B},\hat{Y}_{\mathcal{B}\setminus\mathcal{S}}), \forall \mathcal{S}\subseteq\mathcal{B}, \tag{38}
$$
$$
J_\mathcal{B}(\mathcal{S}) := J_{\emptyset,\mathcal{B}}(\mathcal{S})
$$
$$
= I(X_\mathcal{S};\hat{Y}_{\mathcal{B}\setminus\mathcal{S}},Y|X_{\mathcal{B}\setminus\mathcal{S}}) - I(Y_\mathcal{S};\hat{Y}_\mathcal{S}|X_\mathcal{B},\hat{Y}_{\mathcal{B}\setminus\mathcal{S}},Y), \forall \mathcal{S}\subseteq\mathcal{B}, \tag{39}
$$
$$
J(\mathcal{S}) := J_\mathcal{N}(\mathcal{S})
$$
$$
= I(X_\mathcal{S};\hat{Y}_{\mathcal{S}^c},Y|X_{\mathcal{S}^c}) - I(Y_\mathcal{S};\hat{Y}_\mathcal{S}|X_\mathcal{N},Y,\hat{Y}_{\mathcal{S}^c}), \forall \mathcal{S}\subseteq\mathcal{N}. \tag{40}
$$

Then, we have the following lemmas, whose proofs will be presented in Appendix B.

Lemma 4.1: 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 4.2: Under any $p(x)\prod_{i=1}^n p(x_i)p(\hat{y}_i|x_i,y_i)$, 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: 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 4.4: 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

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

The proof of Theorem 2.6 is similar to the proof of Theorem 2.3, and the details are as follows.

Proof of Theorem 2.6: 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

$$
R^*_{\mathrm{C/B/S}} = \max_{p(x) \prod_{i=1}^n p(x_i) p(\hat{y}_i | x_i, y_i)} I(X; \hat{Y}_{\mathcal{N}}, Y | X_{\mathcal{N}}) \quad (41)
$$
$$
\text{such that } J(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{N} \quad (42)
$$

and

$$
R^*_{\mathrm{R/U/J}} = \max_{p(x) \prod_{i=1}^n p(x_i) p(\hat{y}_i | x_i, y_i)}
$$
$$
\min_{\mathcal{S} \subseteq \mathcal{N}} I(X, X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y | X_{\mathcal{S}^c}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X, X_{\mathcal{N}}, Y, \hat{Y}_{\mathcal{S}^c})
$$
$$
= \max_{p(x) \prod_{i=1}^n p(x_i) p(\hat{y}_i | x_i, y_i)} \min_{\mathcal{S} \subseteq \mathcal{N}} I(X; \hat{Y}_{\mathcal{N}}, Y | X_{\mathcal{N}}) + J(\mathcal{S}). \quad (43)
$$

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 $p(x) \prod_{i=1}^n p(x_i) p(\hat{y}_i | x_i, y_i)$ such that $J(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{N}$, we have

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

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 $p(x) \prod_{i=1}^n p(x_i) p(\hat{y}_i | x_i, y_i)$ 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 $p(x) \prod_{i=1}^n p(x_i) p(\hat{y}_i | x_i, y_i)$, if $\mathcal{D}^c \neq \emptyset$, then $\mathcal{D}^c \in \arg\min_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})$ and $\bigcap_{\mathcal{T} \in \arg\min_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})} \mathcal{T} = \mathcal{D}^c$, where $\mathcal{D}$ is defined as in Lemma 4.2 and $\arg\min_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S}) := \{\mathcal{T} \subseteq \mathcal{N} : J(\mathcal{T}) = \min_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})\}$. ii) We then argue that, under the optimal $p(x) \prod_{i=1}^n p(x_i) p(\hat{y}_i | x_i, y_i)$, $\mathcal{D}^c$ 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 $\mathcal{D}^c \in \arg\min_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})$ and $\bigcap_{\mathcal{T} \in \arg\min_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})} \mathcal{T} = \mathcal{D}^c$.

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, 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 4.1, 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.,

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

Again by Lemmas 4.3 and 4.1 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 4.4, that

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

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 $J(\mathcal{A}_1 \circ \mathcal{A}_2) \leq J(\mathcal{A}_1 \circ \mathcal{D}^c)$. 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) \\
&\quad + 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) \\
&\quad + 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) \\
&\quad + 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) \\
&\quad - 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) \\
&\quad - 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) \\
&\quad - 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) \\
&\quad - 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) \\
&= 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) \\
&\quad + 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) \\
&\quad + 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) \\
&\quad + 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) \\
&\quad - 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) \\
&\quad - 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) \\
&= 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) \\
&\quad + 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) \\
&\quad + 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) \\
&\quad + 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) \\
&\geq 0.
\end{aligned}
$$

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

4) We prove that $\forall \mathcal{A}$ with $\mathcal{AD} \neq \emptyset$ and $\mathcal{AD}^c = \mathcal{D}^c$, $J(\mathcal{A}) \geq J(\mathcal{D}^c)$. Letting $\mathcal{A}_1 = \mathcal{AD}$, 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 $J(\mathcal{A}) \geq J(\mathcal{D}^c)$, we only need to show $J(\mathcal{A}_1) - J(\mathcal{A}_1 \circ \mathcal{D}^c) \geq 0$. For this, we have

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

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

Combining 2)–4), we can conclude that $\mathcal{D}^c \in \arg\min_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})$ and $\bigcap_{\mathcal{T} \in \arg\min_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})} \mathcal{T} = \mathcal{D}^c$.

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

Suppose $\mathcal{D}^c \neq \emptyset$ at the optimum. Then, $\mathcal{D}^c \in \arg\min_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})$ and $\bigcap_{\mathcal{T} \in \arg\min_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})} \mathcal{T} = \mathcal{D}^c$. Therefore

$$R^*_{\mathrm{R/U/J}} = I(X, X_{\mathcal{D}^c}; \hat{Y}_{\mathcal{D}}, Y | X_{\mathcal{D}}) - I(Y_{\mathcal{D}^c}; \hat{Y}_{\mathcal{D}^c} | X, X_{\mathcal{N}}, Y, \hat{Y}_{\mathcal{D}}) \tag{44}$$

$$= I(X, X_{\mathcal{T}}; \hat{Y}_{\mathcal{T}^c}, Y | X_{\mathcal{T}^c}) - I(Y_{\mathcal{T}}; \hat{Y}_{\mathcal{T}} | X, X_{\mathcal{N}}, Y, \hat{Y}_{\mathcal{T}^c}) \tag{45}$$

for any $\mathcal{T} \in \arg\min_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})$, $\mathcal{T} \neq \mathcal{D}^c$.

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 (44) and (45), both

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

and

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

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

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

will change as p decreases for $\mathcal{S} \notin \arg\min_{\mathcal{S} \subseteq \mathcal{N}} J(\mathcal{S})$, 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 2.6. ∎

B. 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 $\mathcal{A}, \mathcal{B} \subseteq \mathcal{N}$ and any $\mathcal{S} \subseteq \mathcal{B}$, let

$$
\begin{aligned}
K_{\mathcal{A}, \mathcal{B}}(\mathcal{S}) :=\ & I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{B} \setminus \mathcal{S}}, \hat{Y}_{\mathcal{A}}, Y | X, X_{\mathcal{A}}, X_{\mathcal{B} \setminus \mathcal{S}}) \\
& - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A}}, Y, X_{\mathcal{B}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}})
\end{aligned}
\tag{46}
$$

$$
\begin{aligned}
K_{\mathcal{B}}(\mathcal{S}) :=\ & K_{\emptyset, \mathcal{B}}(\mathcal{S}) = I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{B} \setminus \mathcal{S}}, Y | X, X_{\mathcal{B} \setminus \mathcal{S}}) \\
& - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X, X_{\mathcal{B}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}}, Y)
\end{aligned}
\tag{47}
$$

$$
\begin{aligned}
R_{\mathcal{B}}(\mathcal{S}) :=\ & I(X, X_{\mathcal{S}}; \hat{Y}_{\mathcal{B} \setminus \mathcal{S}}, Y | X_{\mathcal{B} \setminus \mathcal{S}}) \\
& - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X, X_{\mathcal{B}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}}, Y).
\end{aligned}
\tag{48}
$$

Lemma 4.5: 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 4.6: Under any $p(x) \prod_{i=1}^{n} p(x_i) p(\hat{y}_i | x_i, y_i)$, there exists a unique set $\mathcal{D}_{\mathrm{J}}$, which is the largest subset of $\mathcal{N}$ satisfying

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

Lemma 4.7: 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 4.8: For any disjoint $\mathcal{A}$ and $\mathcal{B}$, and any $\mathcal{S} \subseteq \mathcal{A} \cup \mathcal{B}$, let $\mathcal{S}_1 = \mathcal{SA}$ and $\mathcal{S}_2 = \mathcal{SB}$. Then, we have the following:
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) Especially, 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 4.5–4.7 can be proved along the same lines as the proofs of Lemmas 4.1–4.3, respectively, while the proof of Lemma 4.8 is given in Appendix C.

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

Proof of Theorem 2.7: The uniqueness of $\mathcal{D}_{\mathrm{J}}$ has been established in Lemma 4.6. Below, we focus on showing that i) the rate in (14) is achievable, and ii) the compressions in the set $\mathcal{D}_{\mathrm{J}}$ can be decoded jointly with X.

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

$$I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y | X, X_{\mathcal{S}^c}) - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X, X_{\mathcal{N}}, Y, \hat{Y}_{\mathcal{S}^c}) > 0 \tag{49}$$

for any nonempty $\mathcal{S} \subseteq \mathcal{N}$ and show that

$$
\begin{aligned}
R_{\mathrm{C/B/J}} < \min_{\mathcal{S} \subseteq \mathcal{N}} \ & I(X, X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c}, Y | X_{\mathcal{S}^c}) \\
& - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X, X_{\mathcal{N}}, Y, \hat{Y}_{\mathcal{S}^c})
\end{aligned}
\tag{50}
$$

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

Fix $p(x)\prod_{i=1}^{n} p(x_i)p(\hat{y}_i|x_i,y_i)$. Assume (49) holds. The codebook generation and encoding process here are exactly the same as those in the proof of Theorem 2.5, 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 $b = 1, \ldots, B+M$

$$\Big(\mathbf{X}_b(m_b), (\mathbf{X}_{1,b}(l_{1,b-1}), \hat{\mathbf{Y}}_{1,b}(l_{1,b}|l_{1,b-1})), \ldots,$$
$$(\mathbf{X}_{n,b}(l_{n,b-1}), \hat{\mathbf{Y}}_{n,b}(l_{n,b}|l_{n,b-1})), \mathbf{Y}_b\Big) \in A_\epsilon(X, X_\mathcal{N}, \hat{Y}_\mathcal{N}, Y)$$
(51)

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 (51) for any $b = B+1, \ldots, B+M$. Through the similar lines as in the proof of Theorem 2.5 with $\mathbf{X}_b(m_b), b \in [B+1 : B+M]$ taken into account and treated as known signals, it follows that $\mathbf{l}_B$ can be decoded if (49) 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 (51), 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 (51) with any $\mathbf{l}_{b-1}$, or a false m_b satisfies (51) with some $\mathbf{l}_{b-1}$. According to the properties of typical sequences, the true $(m_b, \mathbf{l}_{b-1})$ satisfies (51) 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 $\{l_{i,b-1}, i \in \mathcal{S}^c\}$, $\mathbf{X}_b(m_b)$ and $(\mathbf{X}_b(\mathcal{S}), \hat{\mathbf{Y}}_b(\mathcal{S}))$ and $(\mathbf{X}_b(\mathcal{S}^c), \hat{\mathbf{Y}}_b(\mathcal{S}^c), \mathbf{Y}_b)$ are mutually independent, and the probability that $(m_b, \mathbf{l}_{b-1})$ satisfies (51) can be upper bounded by

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

Since the number of such false $(m_b, \mathbf{l}_{b-1})$ is upper bounded by $2^{TR_{C/B/J}}\prod_{i\in\mathcal{S}}2^{T(I(Y_i;\hat{Y}_i|X_i)+\epsilon)}$, 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 (50) holds.

Then, based on the recovered m_b and $\mathbf{l}_b$, again from the proof of Theorem 2.5 with $\mathbf{X}_b(m_b)$ taken into account and treated as known signal, it follows that $\mathbf{l}_{b-1}$ can be decoded if (49) holds.

Combining a) and b), we can conclude that both $\mathbf{m}$ and $\mathbf{l}^B$ can be decoded if both (49) and (50) hold.

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

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

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

$$I(X_\mathcal{S}; \hat{Y}_{\mathcal{D}_J\setminus\mathcal{S}}, Y | X, X_{\mathcal{D}_J\setminus\mathcal{S}}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S} | X, X_{\mathcal{D}_J}, Y, \hat{Y}_{\mathcal{D}_J\setminus\mathcal{S}}) > 0$$

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

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 2.8: Still consider the case $Q = \emptyset$. The uniqueness of $\mathcal{D}_J$ has been treated in Lemma 4.6, while the uniqueness of $\mathcal{D}_J'$ can be established along the same lines. To prove Theorem 2.8, in terms of the notations defined in this section, we will sequentially prove that: i) $\max_{\mathcal{M}\subseteq\mathcal{N}} \min_{\mathcal{S}\subseteq\mathcal{M}} R_\mathcal{M}(\mathcal{S}) = \min_{\mathcal{S}\subseteq\mathcal{D}_J} R_{\mathcal{D}_J}(\mathcal{S})$; ii) $\min_{\mathcal{S}\subseteq\mathcal{M}} R_\mathcal{M}(\mathcal{S}) < \min_{\mathcal{S}\subseteq\mathcal{D}_J'} R_{\mathcal{D}_J'}(\mathcal{S})$, for any $\mathcal{M} \nsubseteq \mathcal{D}_J'$; iii) $\max_{\mathcal{M}\subseteq\mathcal{N}} \min_{\mathcal{S}\subseteq\mathcal{M}} R_\mathcal{M}(\mathcal{S}) = \min_{\mathcal{S}\subseteq\mathcal{D}_J'} R_{\mathcal{D}_J'}(\mathcal{S})$.

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

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

For any $\mathcal{S} \subseteq \mathcal{D}_J$, by Part 2) of Lemma 4.8, we have

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

We argue $K_{\mathcal{D}_J,\mathcal{M}\setminus\mathcal{D}_J}(\mathcal{M} \setminus \mathcal{D}_J) \leq 0$ by contradiction. Suppose $K_{\mathcal{D}_J,\mathcal{M}\setminus\mathcal{D}_J}(\mathcal{M} \setminus \mathcal{D}_J) > 0$. Then, by Lemma 4.7, we have that there exists some nonempty $\mathcal{C} \subseteq \mathcal{M}\setminus\mathcal{D}_J$ such that $K_{\mathcal{D}_J,\mathcal{C}}(\mathcal{S}) > 0$, for any nonempty $\mathcal{S} \subseteq \mathcal{C}$. This will further imply, by Part 2) of Lemma 4.5, that $K_{\mathcal{D}_J\cup\mathcal{C}}(\mathcal{S}) > 0$, for any nonempty $\mathcal{S} \subseteq \mathcal{D}_J\cup\mathcal{C}$, which is in contradiction with the definition of $\mathcal{D}_J$. Thus, we must have $K_{\mathcal{D}_J,\mathcal{M}\setminus\mathcal{D}_J}(\mathcal{M}\setminus\mathcal{D}_J) \leq 0$, and $R_\mathcal{M}(\mathcal{S}\cup(\mathcal{M}\setminus\mathcal{D}_J)) \leq R_{\mathcal{D}_J}(\mathcal{S})$.

2) Assume $\mathcal{M} \cap \mathcal{D}_J \neq \mathcal{D}_J$. For any $\mathcal{S} \subseteq \mathcal{M} \cup \mathcal{D}_J$, let $\mathcal{S}_1 = \mathcal{S}\mathcal{M}$ and $\mathcal{S}_2 = \mathcal{S}(\mathcal{D}_J \setminus \mathcal{M})$. By Part 1) of Lemma 4.8, we have

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

and then

$$\min_{\mathcal{S}\subseteq\mathcal{M}\cup\mathcal{D}_J} R_{\mathcal{M}\cup\mathcal{D}_J}(\mathcal{S}) \geq \min_{\mathcal{S}\subseteq\mathcal{M}\cup\mathcal{D}_J}\{R_\mathcal{M}(\mathcal{S}_1) + K_{\mathcal{M}\cup\mathcal{D}_J}(\mathcal{S}_2)\}$$
$$\geq \min_{\mathcal{S}\subseteq\mathcal{M}\cup\mathcal{D}_J}\{R_\mathcal{M}(\mathcal{S}_1) + K_{\mathcal{D}_J}(\mathcal{S}_2)\}$$
$$= \min_{\mathcal{S}_1\subseteq\mathcal{M},\mathcal{S}_2\subseteq\mathcal{D}_J\setminus\mathcal{M}}\{R_\mathcal{M}(\mathcal{S}_1) + K_{\mathcal{D}_J}(\mathcal{S}_2)\}$$
$$= \min_{\mathcal{S}_1\subseteq\mathcal{M}} R_\mathcal{M}(\mathcal{S}_1) + \min_{\mathcal{S}_2\subseteq\mathcal{D}_J\setminus\mathcal{M}} K_{\mathcal{D}_J}(\mathcal{S}_2)$$
$$\geq \min_{\mathcal{S}_1\subseteq\mathcal{M}} R_\mathcal{M}(\mathcal{S}_1)$$

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

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

1) Through the similar lines as in Step 1) of Part i), we can prove $\min_{\mathcal{S} \subseteq \mathcal{M}} R_{\mathcal{M}}(\mathcal{S}) < \min_{\mathcal{S} \subseteq \mathcal{D}'_J} R_{\mathcal{D}'_J}(\mathcal{S})$, for any $\mathcal{M} \cap \mathcal{D}'_J = \mathcal{D}'_J$, $\mathcal{M} \neq \mathcal{D}'_J$. 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 $\mathcal{D}'_J$.

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

$$\min_{\mathcal{S} \subseteq \mathcal{M}} R_{\mathcal{M}}(\mathcal{S}) \leq \min_{\mathcal{S} \subseteq \mathcal{M} \cup \mathcal{D}'_J} R_{\mathcal{M} \cup \mathcal{D}'_J}(\mathcal{S}) < \min_{\mathcal{S} \subseteq \mathcal{D}'_J} R_{\mathcal{D}'_J}(\mathcal{S})$$

iii) From Part ii), we have 1) $\min_{\mathcal{S} \subseteq \mathcal{M}} R_{\mathcal{M}}(\mathcal{S}) < \min_{\mathcal{S} \subseteq \mathcal{D}'_J} R_{\mathcal{D}'_J}(\mathcal{S})$, for any $\mathcal{M} \cap \mathcal{D}'_J = \mathcal{D}'_J$, $\mathcal{M} \neq \mathcal{D}'_J$, and 2) for any $\mathcal{M} \cap \mathcal{D}'_J \neq \mathcal{D}'_J$, $\min_{\mathcal{S} \subseteq \mathcal{M}} R_{\mathcal{M}}(\mathcal{S}) \leq \min_{\mathcal{S} \subseteq \mathcal{M} \cup \mathcal{D}'_J} R_{\mathcal{M} \cup \mathcal{D}'_J}(\mathcal{S}) \leq \min_{\mathcal{S} \subseteq \mathcal{D}'_J} R_{\mathcal{D}'_J}(\mathcal{S})$. Thus, it follows immediately that $\min_{\mathcal{S} \subseteq \mathcal{D}'_J} R_{\mathcal{D}'_J}(\mathcal{S}) = \max_{\mathcal{M} \subseteq \mathcal{N}} \min_{\mathcal{S} \subseteq \mathcal{M}} R_{\mathcal{M}}(\mathcal{S})$. This completes the proof of Theorem 2.8. ∎

V. CONCLUSION

Joint compression-message decoding introduced more freedom in selecting the compressions at the relays. Motivated by it, we have investigated the problem of finding the optimal compressions in maximizing the achievable rate of the original message. We have studied several different compress-and-forward 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 scheme. Compared to the repetitive encoding/all blocks united decoding scheme recently proposed in [3], 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.

Finally, we need to point out that our discussion of optimality is restricted to the few selected compress-and-forward relay schemes. In generalizing the classical compress-and-forward relay scheme in [2] to the case of multiple relays, there could be many other choices of coding considerations [10]. Even for the single-relay case, the optimality of the original compression method used in [2] remains an open question (see [6] and [11]).

APPENDIX A
PROOFS OF LEMMAS 3.1–3.4

Proof of Lemma 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

$$
\begin{aligned}
I_{\mathcal{A} \cup \mathcal{B}}(\mathcal{S}) &= \sum_{i \in \mathcal{S}} R_i - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y, X_{\mathcal{N}}) \\
&= \sum_{i \in \mathcal{S}_1} R_i - I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1} | \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y, X_{\mathcal{N}}) \\
&\quad + \sum_{i \in \mathcal{S}_2} R_i - I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, \hat{Y}_{\mathcal{S}_1}, Y, X_{\mathcal{N}}) \\
&\geq \sum_{i \in \mathcal{S}_1} R_i - I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1} | \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y, X_{\mathcal{N}}) + \sum_{i \in \mathcal{S}_2} R_i \\
&\quad - I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y, X_{\mathcal{N}}) \\
&= I_{\mathcal{A}}(\mathcal{S}_1) + I_{\mathcal{A}, \mathcal{B}}(\mathcal{S}_2) \qquad\qquad (52) \\
&\geq I_{\mathcal{A}}(\mathcal{S}_1) + I_{\mathcal{B}}(\mathcal{S}_2). \qquad\qquad (53)
\end{aligned}
$$

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 (53), $I_{\mathcal{A} \cup \mathcal{B}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{A} \cup \mathcal{B}$. 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 (52), $I_{\mathcal{A} \cup \mathcal{B}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{A} \cup \mathcal{B}$. ∎

Proof of Lemma 3.2: 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 3.1, $\mathcal{D} := \bigcup_{i=1}^{n} \mathcal{D}_i$ also satisfies that $I_{\mathcal{D}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{D}$, which is in contradiction, and hence Lemma 3.2 is proved. ∎

Proof of Lemma 3.3: 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.,

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

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: For any disjoint $\mathcal{A}$ and $\mathcal{B}$,

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

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

which proves the lemma. ∎

APPENDIX B
PROOFS OF LEMMAS 4.1–4.4

Proof of Lemma 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

$$J_{\mathcal{A} \cup \mathcal{B}}(\mathcal{S}) = I(X_{\mathcal{S}}; \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y | X_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}})$$
$$- I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X_{\mathcal{A} \cup \mathcal{B}}, \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y)$$
$$= I(X_{\mathcal{S}_1}; \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y | X_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}})$$
$$+ I(X_{\mathcal{S}_2}; \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y | X_{\mathcal{S}_1}, X_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}})$$
$$- I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1} | X_{\mathcal{A} \cup \mathcal{B}}, \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y)$$
$$- I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X_{\mathcal{A} \cup \mathcal{B}}, \hat{Y}_{\mathcal{S}_1}, \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y)$$
$$= I(X_{\mathcal{S}_1}; \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y | X_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}})$$
$$+ I(X_{\mathcal{S}_2}; \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y | X_{\mathcal{S}_1}, X_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}})$$
$$- [I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1} | X_{\mathcal{A}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y)$$
$$- I(\hat{Y}_{\mathcal{S}_1}; X_{\mathcal{B} \setminus \mathcal{A}}, \hat{Y}_{\mathcal{B}\mathcal{A}^c \setminus \mathcal{S}_2} | X_{\mathcal{A}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y)]$$
$$- I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X_{\mathcal{A} \cup \mathcal{B}}, \hat{Y}_{\mathcal{S}_1}, \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y)$$
$$= [I(X_{\mathcal{S}_1}; \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y | X_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}})$$
$$- I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1} | X_{\mathcal{A}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y)]$$
$$+ [I(X_{\mathcal{S}_2}; \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y | X_{\mathcal{S}_1}, X_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}})$$
$$+ I(\hat{Y}_{\mathcal{S}_1}; X_{\mathcal{B} \setminus \mathcal{A}}, \hat{Y}_{\mathcal{B}\mathcal{A}^c \setminus \mathcal{S}_2} | X_{\mathcal{A}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y)]$$
$$- I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y)$$
$$\geq [I(X_{\mathcal{S}_1}; \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y | X_{\mathcal{A} \setminus \mathcal{S}_1}) - I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1} | X_{\mathcal{A}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y)]$$
$$+ [I(X_{\mathcal{S}_2}; \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y | X_{\mathcal{S}_1}, X_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}})$$
$$+ I(\hat{Y}_{\mathcal{S}_1}; X_{\mathcal{B} \setminus \mathcal{A}}, \hat{Y}_{\mathcal{B}\mathcal{A}^c \setminus \mathcal{S}_2} | X_{\mathcal{A}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y)]$$
$$- I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y)$$
$$= [I(X_{\mathcal{S}_2}; \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y | X_{\mathcal{S}_1}, X_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}})$$
$$+ I(\hat{Y}_{\mathcal{S}_1}; X_{\mathcal{S}_2}, X_{\mathcal{B}\mathcal{A}^c \setminus \mathcal{S}_2}, \hat{Y}_{\mathcal{B}\mathcal{A}^c \setminus \mathcal{S}_2} | X_{\mathcal{A}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y)]$$
$$- I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y) + J_{\mathcal{A}}(\mathcal{S}_1)$$
$$\geq [I(X_{\mathcal{S}_2}; \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y | X_{\mathcal{A}}, X_{\mathcal{B} \setminus \mathcal{S}_2})$$
$$+ I(\hat{Y}_{\mathcal{S}_1}; X_{\mathcal{S}_2} | X_{\mathcal{A}}, X_{\mathcal{B}\mathcal{A}^c \setminus \mathcal{S}_2}, \hat{Y}_{\mathcal{B}\mathcal{A}^c \setminus \mathcal{S}_2}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y)]$$
$$- I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y) + J_{\mathcal{A}}(\mathcal{S}_1)$$
$$= I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y | X_{\mathcal{A}}, X_{\mathcal{B} \setminus \mathcal{S}_2})$$
$$- I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y) + J_{\mathcal{A}}(\mathcal{S}_1)$$
$$= J_{\mathcal{A}}(\mathcal{S}_1) + J_{\mathcal{A}, \mathcal{B}}(\mathcal{S}_2) \tag{54}$$
$$\geq J_{\mathcal{A}}(\mathcal{S}_1) + J_{\mathcal{B}}(\mathcal{S}_2). \tag{55}$$

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 (55), $J_{\mathcal{A} \cup \mathcal{B}}(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{A} \cup \mathcal{B}$. 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 (54), $J_{\mathcal{A} \cup \mathcal{B}}(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{A} \cup \mathcal{B}$. ∎

Proof of Lemma 4.2: 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 4.1, $\mathcal{D} := \bigcup_{i=1}^{n} \mathcal{D}_i$ also satisfies that $J_{\mathcal{D}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{D}$, which is in contradiction, and hence Lemma 4.2 is proved. ∎

Proof of Lemma 4.3: 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.,

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

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 4.4: For any disjoint $\mathcal{A}$ and $\mathcal{B}$

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

which proves the lemma. ∎

APPENDIX C
PROOF OF LEMMA 4.8

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, X_{\mathcal{S}}; \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y | X_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}) \\
&\quad - I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | X, X_{\mathcal{A} \cup \mathcal{B}}, \hat{Y}_{(\mathcal{A} \cup \mathcal{B}) \setminus \mathcal{S}}, Y) \\
&= I(X, X_{\mathcal{S}_1}, X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y | X_{\mathcal{A} \setminus \mathcal{S}_1}, X_{\mathcal{B} \setminus \mathcal{S}_2}) \\
&\quad - I(Y_{(\mathcal{S}_1 \cup \mathcal{S}_2)}; \hat{Y}_{(\mathcal{S}_1 \cup \mathcal{S}_2)} | X, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y) \\
&= I(X, X_{\mathcal{S}_1}; \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y | X_{\mathcal{A} \setminus \mathcal{S}_1}, X_{\mathcal{B} \setminus \mathcal{S}_2}) \\
&\quad + I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y | X, X_{\mathcal{A}}, X_{\mathcal{B} \setminus \mathcal{S}_2}) \\
&\quad - [I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1} | X, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y) \\
&\quad + I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y)] \\
&= I(X, X_{\mathcal{S}_1}; \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y | X_{\mathcal{A} \setminus \mathcal{S}_1}) \\
&\quad + I(X, X_{\mathcal{S}_1}; X_{\mathcal{B} \setminus \mathcal{S}_2}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2} | X_{\mathcal{A} \setminus \mathcal{S}_1}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y) \\
&\quad - [I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1} | X, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y) \\
&\quad - I(X_{\mathcal{B}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}; \hat{Y}_{\mathcal{S}_1} | X, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y)] \\
&\quad + I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y | X, X_{\mathcal{A}}, X_{\mathcal{B} \setminus \mathcal{S}_2}) \\
&\quad - I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y) \\
&= [I(X, X_{\mathcal{S}_1}; \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y | X_{\mathcal{A} \setminus \mathcal{S}_1}) \\
&\quad - I(Y_{\mathcal{S}_1}; \hat{Y}_{\mathcal{S}_1} | X, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y)] \\
&\quad + I(X, X_{\mathcal{S}_1}; X_{\mathcal{B} \setminus \mathcal{S}_2}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2} | X_{\mathcal{A} \setminus \mathcal{S}_1}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y) \\
&\quad + I(X_{\mathcal{B}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}; \hat{Y}_{\mathcal{S}_1} | X, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y) \\
&\quad + I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y | X, X_{\mathcal{A}}, X_{\mathcal{B} \setminus \mathcal{S}_2}) \\
&\quad - I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y) \\
&= R_{\mathcal{A}}(\mathcal{S}_1) + I(X, X_{\mathcal{S}_1}; X_{\mathcal{B} \setminus \mathcal{S}_2}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2} | X_{\mathcal{A} \setminus \mathcal{S}_1}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y) \\
&\quad + I(X_{\mathcal{B}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}; \hat{Y}_{\mathcal{S}_1} | X, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y) \\
&\quad + I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y | X, X_{\mathcal{A}}, X_{\mathcal{B} \setminus \mathcal{S}_2}) \\
&\quad - I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y). \quad (56)
\end{aligned}
$$

When $\mathcal{S}_2 = \mathcal{B}$, following (56), we have

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

Generally, for any $\mathcal{S}_2 \subseteq \mathcal{B}$, continuing (56), we have

$$
\begin{aligned}
R_{\mathcal{A} \cup \mathcal{B}}(\mathcal{S}) &\geq R_{\mathcal{A}}(\mathcal{S}_1) + I(X_{\mathcal{B}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}; \hat{Y}_{\mathcal{S}_1} | X, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y) \\
&\quad + I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y | X, X_{\mathcal{A}}, X_{\mathcal{B} \setminus \mathcal{S}_2}) \\
&\quad - I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y) \\
&= R_{\mathcal{A}}(\mathcal{S}_1) + I(X_{\mathcal{B} \setminus \mathcal{S}_2}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}; \hat{Y}_{\mathcal{S}_1} | X, X_{\mathcal{A}}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y) \\
&\quad + I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_1} | X, X_{\mathcal{A}}, X_{\mathcal{B} \setminus \mathcal{S}_2}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, Y) \\
&\quad + I(X_{\mathcal{S}_2}; \hat{Y}_{\mathcal{A} \setminus \mathcal{S}_1}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y | X, X_{\mathcal{A}}, X_{\mathcal{B} \setminus \mathcal{S}_2}) \\
&\quad - I(Y_{\mathcal{S}_2}; \hat{Y}_{\mathcal{S}_2} | X, X_{\mathcal{A}}, X_{\mathcal{B}}, \hat{Y}_{\mathcal{A}}, \hat{Y}_{\mathcal{B} \setminus \mathcal{S}_2}, Y)
\end{aligned}
$$

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

This completes the proof of Lemma 4.8.

REFERENCES

[1] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–154, 1971.

[2] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.

[3] S. H. Lim, Y.-H. Kim, A. E. Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[4] L.-L. Xie, An Improvement of Cover/EL Gamal's Compress-and-Forward Relay Scheme Aug. 2009 [Online]. Available: http://arxiv.org/abs/0908.0163

[5] A. E. Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1545–1561, Apr. 2006.

[6] Y.-H. Kim, "Coding techniques for primitive relay channels," presented at the 45th Annu. Allerton Conf. Commun., Control Comput., Monticello, IL, USA, Sep. 2007.

[7] A. E. Gamal and Y.-H. Kim, Lecture Notes on Network Information Theory Jan. 2010 [Online]. Available: http://arxiv.org/abs/1001.3404

[8] X. Wu and L.-L. Xie, "On the optimality of successive decoding in compress-and-forward relay schemes," presented at the 48th Annu. Allerton Conf. Commun., Control, Comput., Monticello, IL, USA, September 29-October 1, 2010.

[9] X. Wu and L.-L. Xie, "An optimality-robustness tradeoff in the compress-and-forward relay scheme," presented at the IEEE 72nd Veh. Technol. Conf., Ottawa, ON, Canada, Sep. 2010.

[10] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[11] X. Wu and L.-L. Xie, Asymptotic equipartition property of output when rate is above capacity Aug. 2009 [Online]. Available: http://arxiv.org/abs/0908.4445

[12] T. Cover and J. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

Xiugang Wu (S'08) received the B.Eng. degree with honors in electronics and information engineering from Tongji University, Shanghai, China, in 2007, and the M.A.Sc degree in electrical and computer engineering from the University of Waterloo, Waterloo, Ontario, Canada, in 2009.

He is currently pursuing the Ph.D. degree in the Department of Electrical and Computer Engineering from the University of Waterloo. His research interests are in information theory and wireless networks.

Liang-Liang Xie (M'03–SM'09) received the B.S. degree in mathematics from Shandong University, Jinan, China, in 1995 and the Ph.D. degree in control theory from the Chinese Academy of Sciences, Beijing, China, in 1999.

He did postdoctoral research with the Automatic Control Group, Linköping University, Linköping, Sweden, during 1999–2000 and with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, during 2000–2002. He is currently a Professor at the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include wireless networks, information theory, adaptive control, and system identification.