On the Optimality of Successive Decoding in Compress-and-Forward Relay Schemes

Xiugang Wu and Liang-Liang Xie
University of Waterloo, Waterloo, ON, Canada N2L 3G1
Email: x23wu@uwaterloo.ca, llxie@uwaterloo.ca

Abstract

In the classical compress-and-forward relay scheme developed by (Cover and El Gamal, 1979), the decoding process operates in a successive way: the destination first decodes the compressed observation of the relay, and then decodes the original message of the source. Recently, two modified compress-and-forward relay schemes were proposed, and in both of them, the destination jointly decodes the compressed observation of the relay and the original message, instead of successively. Such a modification on the decoding process was motivated by realizing that it is generally easier to decode the compressed observation jointly with the original message, and more importantly, the original message can be decoded even without completely decoding the compressed observation. Thus, joint decoding provides more freedom in choosing the compression rate at the relay, i.e., the relay's observation can be compressed at a rate higher than supportable by successive decoding.

However, the question remains whether this freedom of choosing a higher compression rate at the relay improves the achievable rate of the original message. It has been shown in (El Gamal and Kim, 2010) that the answer is negative in the single relay case, and the achievable rate obtained in (Cover and El Gamal, 1979) with successive decoding is still the best. In this paper, we further demonstrate 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 compression rates higher than supportable by successive decoding will actually result in a strictly lower achievable rate for the original message. Therefore, to maximize the achievable rate for the original message, the compression rates should always be chosen to be supportable by successive decoding. The freedom of choosing higher compression rates introduced by joint decoding is actually obtained at the sacrifice of the achievable rate for the original message. This phenomenon is also shown to exist under the repetitive encoding framework recently proposed by (Lim, Kim, El Gamal, and Chung, 2010), which improves the achievable rate in the case of multiple relays compared to the classical encoding framework. Here, another interesting discovery is that the same achievable rate can be obtained without repetitive encoding if the relays encode with memory of previous blocks.

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 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 leads to less loss.

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

- Cumulative encoding/successive decoding,
- Cumulative encoding/joint decoding,
- Repetitive encoding/successive decoding,
- Repetitive encoding/joint decoding,
- Cumulative encoding/relay with memory/successive decoding,
- Cumulative encoding/relay with memory/joint decoding.

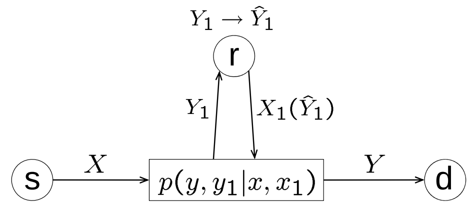

Fig. 1. The single relay channel.

The cumulative encoding/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 [8], where the same information is encoded in each block. The decoding is successive in the sense that the destination first decodes the compressed observation of the relay, and then decodes the original message. The compressed observation $\hat{Y}_1$ can be first recovered at the destination, as long as the following constraint is satisfied:

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

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 successive decoding process requires $\hat{Y}_1$ to be completely decoded. This facilitates the decoding of X, but is not a requirement of the original problem. Recognizing this, a joint decoding process has been proposed in [4]-[7], 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 without completely decoding $\hat{Y}_1$, i.e., only to determine $\hat{Y}_1$ to within a set of possibilities. Thus, with joint decoding, the constraint (1) is not needed, 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, even if $\hat{Y}_1$ is to be completely decoded, it can be more easily done by joint decoding, and instead of (1), we need a less strict constraint:

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

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

Therefore, compared to successive decoding, joint decoding provides more freedom in choosing the compression rate of $\hat{Y}_1$, even at a rate not decodable by the destination. However, the question remains whether joint decoding necessarily achieves higher rates for the original message than successive decoding can do. For the single relay case, it has been proved in [5] that the answer is negative, and any rate achievable by one 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 won't be able to achieve any higher rate either. More interestingly, any compression rates higher than supportable by successive decoding, e.g., violating (1) in the one relay case, will actually result in a strictly lower achievable rate for the original message. Therefore, to maximize the achievable rate for the original message, the compression rates should always be chosen to be supportable by successive decoding, e.g., satisfying (1) in the one relay case.

Recently, a different encoding process was proposed in [8], 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 has the advantage of introducing collaboration among all the blocks, so that in the final decoding, all the blocks are helping each other. This

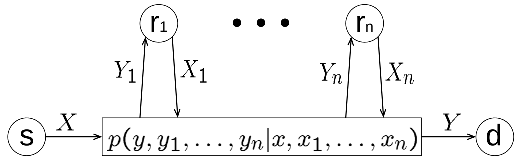

Fig. 2. The multiple-relay channel.

repetitive encoding process was combined with joint decoding in [8], and although no improvement was shown in the single relay case, some interesting improvement can be obtained in the case of multiple relays. In this paper, we consider the combination of repetitive encoding with successive decoding, and similarly demonstrate that successive decoding performs as well as joint decoding in terms of achievable rates for the original message when repetitive encoding is used. Moreover, we also show that any compression rates not supportable by successive decoding will necessarily lead to rate loss of the original message.

As another contribution of this paper, we propose a new compress-and-forward relay strategy where the relays encode with memory of previous blocks. It is found that when such a help to previous blocks is offered by the relays, repetitive encoding is not needed at the source to achieve the same rate. To distinguish from the schemes discussed earlier, we refer to such schemes as cumulative encoding/relay with memory/successive decoding, or cumulative encoding/relay with memory/joint decoding.

Finally, we point out that the optimality of successive decoding is only shown for the case of a single destination in the network. When there are multiple destinations in the network, joint decoding may perform better, since it is more flexible to meet the tradeoff between different destinations.

The remainder of the paper is organized as the following. In the next section, we formally state our problem setup and summarize the main results. Then, in Section III and Section IV, we thoroughly discuss the achievability results with successive decoding and joint decoding, and the optimality of successive decoding, under the frameworks of cumulative encoding and repetitive encoding respectively. Our new scheme of relay with memory is presented in Section IV-A.

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 i-th relay. The i-th relay sends an input x_i based on what it has received:

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

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 this paper are two-fold as the following.

i) **Under the cumulative encoding framework:** In Section III, we first establish the achievable rates for cumulative encoding/successive decoding and cumulative encoding/joint decoding, as stated in Theorems 2.1 and 2.2 respectively; and then demonstrate the optimality of successive decoding in the sense of

Theorem 2.3. Specifically, we show that for the general multiple-relay channel, with the cumulative encoding/joint decoding scheme, the optimal rate can be achieved only if the compression rates at the relays are chosen such that the compressions can be first decoded at the destination, i.e., successive decoding can also be carried out.

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

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

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

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

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

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

and

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

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

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

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

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

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

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

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

Theorem 2.3: $R^*_{\text{C/S}} = R^*_{\text{C/J}}$, and $R^*_{\text{C/J}}$ can be obtained only under the distribution

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

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

$$\sum_{i\in\mathcal{S}_1} R_i \leq I(X_{\mathcal{S}_1};Y|X_{\mathcal{S}_1^c}) \tag{11}$$

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

$$I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|\hat{Y}_{\mathcal{S}^c},Y,X_{\mathcal{N}}) \leq \sum_{i\in\mathcal{S}} R_i. \tag{12}$$

ii) **Under the repetitive encoding framework:** In Section IV, we first establish the achievable rates with successive decoding and joint decoding, and then establish the optimality of successive decoding. Similarly, we show this optimality by proving that the optimal rate with the repetitive encoding/joint decoding scheme can be achieved only if the compression rates at the relays are chosen so that successive decoding can also be carried out.

Theorem 2.4: For the multiple-relay channel depicted in Fig. 2, by the repetitive encoding/successive decoding scheme, a rate $R_{\text{R/S}}$ is achievable if there exists some

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

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

$$I(X_{\mathcal{S}};\hat{Y}_{\mathcal{S}^c},Y|X_{\mathcal{S}^c})-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_{\mathcal{N}},Y,\hat{Y}_{\mathcal{S}^c})\geq 0, \tag{13}$$

and

$$R_{\text{R/S}}<I(X;\hat{Y}_{\mathcal{N}},Y|X_{\mathcal{N}}). \tag{14}$$

Theorem 2.5: For the multiple-relay channel depicted in Fig. 2, by the repetitive encoding/joint decoding scheme, a rate $R_{\text{R/J}}$ is achievable if there exists some

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

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

$$R_{\text{R/J}}<I(X,X_{\mathcal{S}};\hat{Y}_{\mathcal{S}^c},Y|X_{\mathcal{S}^c})-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X,X_{\mathcal{N}},Y,\hat{Y}_{\mathcal{S}^c}). \tag{15}$$

It is interesting to note that

$$\begin{aligned}&I(X;\hat{Y}_{\mathcal{N}},Y|X_{\mathcal{N}})+I(X_{\mathcal{S}};\hat{Y}_{\mathcal{S}^c},Y|X_{\mathcal{S}^c})-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_{\mathcal{N}},Y,\hat{Y}_{\mathcal{S}^c})\\&=I(X,X_{\mathcal{S}};\hat{Y}_{\mathcal{S}^c},Y|X_{\mathcal{S}^c})-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X,X_{\mathcal{N}},Y,\hat{Y}_{\mathcal{S}^c}).\end{aligned}$$

Let $R^*_{\text{R/S}}$ and $R^*_{\text{R/J}}$ be the supremum of the achievable rates stated in Theorems 2.4 and 2.5 respectively.

Theorem 2.6: $R^*_{\text{R/S}}=R^*_{\text{R/J}}$, and $R^*_{\text{R/J}}$ can be obtained only under the distribution

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

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

$$I(X_{\mathcal{S}};\hat{Y}_{\mathcal{S}^c},Y|X_{\mathcal{S}^c})-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_{\mathcal{N}},Y,\hat{Y}_{\mathcal{S}^c})\geq 0. \tag{16}$$

III. Successive Decoding vs. Joint Decoding under Cumulative Encoding Framework

We first prove the achievability results stated in Theorems 2.1 and 2.2 respectively.

In both the cumulative encoding/successive decoding and cumulative encoding/joint decoding schemes, the codebook generation and encoding process is 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: i) In successive decoding, the destination first finds, from the specific bins sent by the relays via $X_1,X_2,\ldots,X_n$, the unique combination of $\hat{Y}_1,\hat{Y}_2,\ldots,\hat{Y}_n$ sequences that is jointly typical with the Y 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. ii) 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. A Simplified Model and the Proof of Theorem 2.1

To make the presentation easier to follow, we introduce a simplified channel model as depicted in Fig. 3, where, the relays are connected to the destination via error-free digital links with capacities $R_1, R_2, \ldots, R_n$, where $(R_1, R_2, \ldots, R_n)$ are chosen based on (6). The i-th digital link plays the same role as the $X_i \to Y$ link in Fig. 2, for any $i = 1, 2, \ldots, n$. Such a replacement will not lead to any essential variation of the original coding scheme, since under the original coding framework, the $X_i \to Y$ link is used as a separate link to forward digital information. The benefit of directly replacing it by a digital link is that the codebook construction for $\hat{Y}_i$ can be simplified, since no X_i needs to be considered. For this simplified model, (7) and (8) simplify to

$$I(Y_{\mathcal{S}}; \hat{Y}_{\mathcal{S}} | \hat{Y}_{\mathcal{S}^c}, Y) \leq \sum_{i \in \mathcal{S}} R_i \tag{17}$$

and

$$R_{\text{C/S}} < I(X; \hat{Y}_{\mathcal{N}}, Y). \tag{18}$$

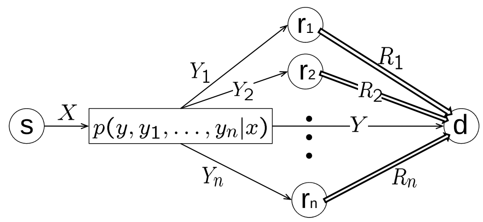

Fig. 3. A simplified multiple-relay model with digital links.

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 i-th relay needs to generate $2^{T(I(Y_i;\hat{Y}_i)+\epsilon)}$ many $\hat{Y}_i$ sequences, and randomly throws them into 2^{TR_i} bins. 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 during the block, and in the next block, inform the destination the index of the bin that contains the $\hat{Y}_i$ sequence.

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 combination of $\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n$ sequences that is jointly typical with the Y sequence received, i.e.,

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

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 $\underline{Y}(b-1)$ can be upper bounded by

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

There are $\prod_{i\in\mathcal{S}}(2^{T(I(Y_i;\hat{Y}_i)-R_i+\epsilon)}-1)$ 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

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

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

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

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

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

Plugging this into (20), we obtain (17)[1].

Given that (17) is satisfied for any $\mathcal{S}\subseteq\mathcal{N}$, the destination can recover $\underline{\hat{Y}}_{\mathcal{N}}(b-1)$ at the end of block b. Then, based on $\underline{\hat{Y}}_{\mathcal{N}}(b-1)$ and $\underline{Y}(b-1)$, $\underline{X}(w)$ can be recovered if (18) holds.

B. *Proof of Theorem 2.2*

Similarly, we consider the simplified model as depicted in Fig. 3, where the rates $(R_1,R_2,\dots,R_n)$ are chosen based on (9). Then, (10) simplifies to

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

In cumulative encoding/joint decoding, the encoding part is 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 $b=2,3,\dots$, the destination finds the unique X sequence that is jointly typical with the Y sequence received during block $b-1$, and also with some $\hat{Y}_1,\hat{Y}_2,\dots,\hat{Y}_n$ sequences from the bins forwarded by the relays during block b, i.e.,

$$(\underline{X}(w),\underline{Y}(b-1),\underline{\hat{Y}}_{\mathcal{N}}(b-1))\in A_\epsilon(X,Y,\hat{Y}_1,\dots,\hat{Y}_n). \tag{22}$$

Error occurs if the true $\underline{X}(w)$ does not satisfy (22), or a false $\underline{X}(w')$ satisfies (22). According to the properties of typical sequences, the true $\underline{X}(w)$ satisfies (22) with high probability.

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

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

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

$$\begin{aligned}&2^{TR}2^{T(H(X,Y,\hat{Y}_{\mathcal{N}})+\epsilon)}2^{-T(H(X)-\epsilon)}\\ \times\ &2^{-T(H(Y,\hat{Y}_{\mathcal{S}^c})-\epsilon)}\prod_{i\in\mathcal{S}}2^{-T(H(\hat{Y}_i)-I(Y_i;\hat{Y}_i)+R_i-2\epsilon)}\end{aligned}$$

which tends to zero for sufficiently small ϵ as $T\to\infty$, if (21) holds.

[1]The case of "=" can be included since (18) doesn't include "=". The same consideration applies throughout the paper.

C. Optimality of Successive Decoding under Cumulative Encoding Framework

To make the proof of Theorem 2.3 easier to follow, we still consider the simplified model depicted in Fig. 3. Then, $R^*_{\text{C/S}}$ and $R^*_{\text{C/J}}$ can be respectively written as

$$R^*_{\text{C/S}} = \max_{p(x)\prod_{i=1}^n p(\hat{y}_i|y_i)} I(X;\hat{Y}_{\mathcal{N}},Y) \tag{23}$$

$$\text{such that } I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|\hat{Y}_{\mathcal{S}^c},Y) - \sum_{i\in\mathcal{S}} R_i \leq 0, \forall \mathcal{S}\subseteq\mathcal{N}, \tag{24}$$

and

$$R^*_{\text{C/J}} = \max_{p(x)\prod_{i=1}^n p(\hat{y}_i|y_i)} \min_{\mathcal{S}\subseteq\mathcal{N}} \{I(X;\hat{Y}_{\mathcal{N}},Y) - I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|\hat{Y}_{\mathcal{S}^c},Y) + \sum_{i\in\mathcal{S}} R_i\}. \tag{25}$$

Before proceeding to the proof of Theorem 2.3, we first introduce some useful notations and lemmas. Let

$$I_{\mathcal{A},\mathcal{B}}(\mathcal{S}) := \sum_{i\in\mathcal{S}} R_i - I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|\hat{Y}_{\mathcal{A}},\hat{Y}_{\mathcal{B}\backslash\mathcal{S}},Y), \forall\mathcal{S}\subseteq\mathcal{B}, \tag{26}$$

$$I_{\mathcal{B}}(\mathcal{S}) := I_{\emptyset,\mathcal{B}}(\mathcal{S}) = \sum_{i\in\mathcal{S}} R_i - I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|\hat{Y}_{\mathcal{B}\backslash\mathcal{S}},Y), \forall\mathcal{S}\subseteq\mathcal{B}, \tag{27}$$

$$I(\mathcal{S}) := I_{\mathcal{N}}(\mathcal{S}) = \sum_{i\in\mathcal{S}} R_i - I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|\hat{Y}_{\mathcal{S}^c},Y), \forall\mathcal{S}\subseteq\mathcal{N}. \tag{28}$$

Then, we have the following lemmas, whose proofs will be deferred until we finish the proof of Theorem 2.3.

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 $I_{\mathcal{A}\bigcup\mathcal{B}}(\mathcal{S})\geq 0$, $\forall\mathcal{S}\subseteq\mathcal{A}\bigcup\mathcal{B}$.
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 $I_{\mathcal{A}\bigcup\mathcal{B}}(\mathcal{S})\geq 0$, $\forall\mathcal{S}\subseteq\mathcal{A}\bigcup\mathcal{B}$.

Lemma 3.2: Under any $p(x)\prod_{i=1}^n p(\hat{y}_i|y_i)$, 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 $\mathcal{A}\bigcap\mathcal{B}=\emptyset$, $I(\mathcal{A})+I(\mathcal{B}) = I(\mathcal{A}\bigcup\mathcal{B}) + I(\hat{Y}_{\mathcal{A}};\hat{Y}_{\mathcal{B}}|\hat{Y}_{(\mathcal{A}\bigcup\mathcal{B})^c},Y)$.

We are now ready to prove Theorem 2.3.

Proof of Theorem 2.3: We show $R^*_{\text{C/S}} = R^*_{\text{C/J}}$ by showing that $R^*_{\text{C/S}} \leq R^*_{\text{C/J}}$ and $R^*_{\text{C/S}} \geq R^*_{\text{C/J}}$ respectively. Under any $p(x)\prod_{i=1}^n p(\hat{y}_i|y_i)$ such that $I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|\hat{Y}_{\mathcal{S}^c},Y)\leq\sum_{i\in\mathcal{S}} R_i$, $\forall\mathcal{S}\subseteq\mathcal{N}$, we have

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

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

To show $R^*_{\text{C/S}} \geq R^*_{\text{C/J}}$, it is sufficient to show that $R^*_{\text{C/J}}$ can be achieved only with $p(x)\prod_{i=1}^n p(\hat{y}_i|y_i)$ 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 under any $p(x)\prod_{i=1}^n p(\hat{y}_i|y_i)$, if $\mathcal{D}^c\neq\emptyset$, then $\mathcal{D}^c\in\underset{\mathcal{S}\subseteq\mathcal{N}}{\text{argmin}}\, I(\mathcal{S})$ and $\bigcap_{\mathcal{T}\in\underset{\mathcal{S}\subseteq\mathcal{N}}{\text{argmin}}\, I(\mathcal{S})}\mathcal{T} = \mathcal{D}^c$, where $\mathcal{D}$ is defined as in Lemma 3.2 and $\underset{\mathcal{S}\subseteq\mathcal{N}}{\text{argmin}}\, I(\mathcal{S}) := \{\mathcal{T}\subseteq\mathcal{N} : I(\mathcal{T}) = \min_{\mathcal{S}\subseteq\mathcal{N}} I(\mathcal{S})\}$. ii) We then argue that , under the optimal $p(x)\prod_{i=1}^n p(\hat{y}_i|y_i)$, $\mathcal{D}^c$ must be $\emptyset$, i.e., $\mathcal{D}$ must be $\mathcal{N}$, and thus by the definition of $\mathcal{D}$, $I(\mathcal{S})\geq 0, \forall\mathcal{S}\subseteq\mathcal{N}$.

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

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 $I_{\mathcal{D}\bigcup\mathcal{B}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{D}\bigcup\mathcal{B}$. This is contradictory with the definition of $\mathcal{D}$, and thus $I(\mathcal{D}^c) < 0$.

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

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

Again by Lemma 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 $I_{\mathcal{D}\bigcup\mathcal{B}}(\mathcal{S}) \geq 0, \forall \mathcal{S} \subseteq \mathcal{D}\bigcup\mathcal{B}$, 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

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

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

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

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

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

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

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

ii) We now argue that under the optimal $p(x)\prod_{i=1}^n p(\hat{y}_i|y_i)$ that achieves $R^*_{\text{C/J}}$, if $\mathcal{D}^c \neq \emptyset$, then $R^*_{\text{C/J}}$ is not optimal; and hence $\mathcal{D}^c$ must be $\emptyset$. The argument is extended from that in [5] and the detailed analysis is as follows.

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

$$\begin{aligned} R^*_{\text{C/J}} =& I(X;\hat{Y}_{\mathcal{N}},Y) + I(\mathcal{D}^c) \\ =& I(X;\hat{Y}_{\mathcal{D}},Y) + I(X;\hat{Y}_{\mathcal{D}^c}|\hat{Y}_{\mathcal{D}},Y) + \sum_{i\in\mathcal{D}^c} R_i - I(X,Y_{\mathcal{D}^c};\hat{Y}_{\mathcal{D}^c}|\hat{Y}_{\mathcal{D}},Y) \\ =& I(X;\hat{Y}_{\mathcal{D}},Y) + \sum_{i\in\mathcal{D}^c} R_i - I(Y_{\mathcal{D}^c};\hat{Y}_{\mathcal{D}^c}|X,\hat{Y}_{\mathcal{D}},Y), \end{aligned} \tag{29}$$

and similarly,

$$\begin{aligned} R^*_{\text{C/J}} =& I(X;\hat{Y}_{\mathcal{N}},Y) + I(\mathcal{T}) \\ =& I(X;\hat{Y}_{\mathcal{T}^c},Y) + \sum_{i\in\mathcal{T}} R_i - I(Y_{\mathcal{T}};\hat{Y}_{\mathcal{T}}|X,\hat{Y}_{\mathcal{T}^c},Y), \end{aligned} \tag{30}$$

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

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 $R^*_{\text{C/J}}$. As p decreases from 1, it can be seen from (29) and (30) that both $I(X;\hat{Y}'_{\mathcal{N}},Y) + I(\mathcal{D}^c)$ and $I(X;\hat{Y}'_{\mathcal{N}},Y) + I(\mathcal{T})$ will increase, where $\mathcal{T} \in \underset{\mathcal{S}\subseteq\mathcal{N}}{\operatorname{argmin}} I(\mathcal{S})$, $\mathcal{T} \neq \mathcal{D}^c$. Thus, no matter how $I(X;\hat{Y}'_{\mathcal{N}},Y) + I(\mathcal{S})$ will change as p decreases for $\mathcal{S} \notin \underset{\mathcal{S}\subseteq\mathcal{N}}{\operatorname{argmin}} I(\mathcal{S})$, it is certain that there exists a p^* such that the achievable rate by using $\hat{Y}'_1, \hat{Y}'_2, \ldots, \hat{Y}'_n$ is larger than $R^*_{\text{C/J}}$. This is in contradiction with the optimality of $R^*_{\text{C/J}}$, 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. ∎

Below, we summarize the proofs of Lemma 3.1-3.4.

Proof of Lemma 3.1:

For any $\mathcal{S} \subseteq \mathcal{A}\bigcup\mathcal{B}$, let $\mathcal{S}_1 = \mathcal{S}\mathcal{A}$ and $\mathcal{S}_2 = \mathcal{S}(\mathcal{B}\setminus\mathcal{A})$. Then,

$$\begin{aligned} I_{\mathcal{A}\bigcup\mathcal{B}}(\mathcal{S}) =& \sum_{i\in\mathcal{S}} R_i - I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|\hat{Y}_{(\mathcal{A}\bigcup\mathcal{B})\setminus\mathcal{S}},Y) \\ =& \sum_{i\in\mathcal{S}_1} R_i - I(Y_{\mathcal{S}_1};\hat{Y}_{\mathcal{S}_1}|\hat{Y}_{(\mathcal{A}\bigcup\mathcal{B})\setminus\mathcal{S}},Y) + \sum_{i\in\mathcal{S}_2} R_i - I(Y_{\mathcal{S}_2};\hat{Y}_{\mathcal{S}_2}|\hat{Y}_{(\mathcal{A}\bigcup\mathcal{B})\setminus\mathcal{S}},\hat{Y}_{\mathcal{S}_1},Y) \\ \geq& \sum_{i\in\mathcal{S}_1} R_i - I(Y_{\mathcal{S}_1};\hat{Y}_{\mathcal{S}_1}|\hat{Y}_{\mathcal{A}\setminus\mathcal{S}_1},Y) + \sum_{i\in\mathcal{S}_2} R_i - I(Y_{\mathcal{S}_2};\hat{Y}_{\mathcal{S}_2}|\hat{Y}_{\mathcal{A}},\hat{Y}_{\mathcal{B}\setminus\mathcal{S}_2},Y) \\ =& I_{\mathcal{A}}(\mathcal{S}_1) + I_{\mathcal{A},\mathcal{B}}(\mathcal{S}_2) & (31) \\ \geq& I_{\mathcal{A}}(\mathcal{S}_1) + I_{\mathcal{B}}(\mathcal{S}_2). & (32) \end{aligned}$$

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 (32), $I_{\mathcal{A}\bigcup\mathcal{B}}(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{A}\bigcup\mathcal{B}$. 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 (31), $I_{\mathcal{A}\bigcup\mathcal{B}}(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{A}\bigcup\mathcal{B}$. ∎

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.,

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

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 $\mathcal{B}$,

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

which proves the lemma. ■

IV. Successive Decoding vs. Joint Decoding under Repetitive Encoding Framework

Specializing Theorem 1 in [8] to the case of single source multiple-relay channel depicted in Fig. 2, we readily have the achievable rate with repetitive encoding/joint decoding, as stated in Theorem 2.5. Below, we focus on demonstrating the achievability result with repetitive encoding/successive decoding, and establishing the optimality of successive decoding under the repetitive encoding framework.

A. *Proof of Theorem 2.4*

In repetitive encoding/successive decoding, the encoding process follows that in the proof of Theorem 1 in [8], but the decoding process operates in a successive way. The details are as follows.

Codebook Generation: Fix $p(x)\prod_{i=1}^{n} p(x_i)p(\hat{y}_i|x_i,y_i)$. Consider $B+M$ blocks, where the source will transmit information in the first B blocks and keep silent in the last M blocks, and $M \ll B$ such that the rate loss can be made arbitrarily small. We randomly and independently generate a codebook for each block.

For each block $b \in [1:B]$, randomly and independently generate $2^{TBR_{\text{R/S}}}$ sequences $\mathbf{x}_b(m)$, $m \in [1:2^{TBR_{\text{R/S}}}]$; for each block $b \in [1:B]$ 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 $b \in [1:B]$,

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

For each block $b \in [B+1 : B+M]$ and each relay node $i \in \mathcal{N}$, randomly and independently generate $2^{T(b-B)\hat{R}_i}$ sequences $\mathbf{x}_{i,b}(l_{i,B}^{b-1})$, where $l_{i,B}^{b-1} = (l_{i,B}, \ldots, l_{i,b-1})$ is a $b-B$ dimensional vector with each component restricted to $[1 : 2^{T\hat{R}_i}]$ and thus $l_{i,B}^{b-1} \in [1 : 2^{T\hat{R}_i}]^{b-B}$; for each relay node $i \in \mathcal{N}$ and each $\mathbf{x}_{i,b}(l_{i,B}^{b-1})$, $l_{i,B}^{b-1} \in [1 : 2^{T\hat{R}_i}]^{b-B}$, randomly and conditionally independently generate $2^{T\hat{R}_i}$ sequences $\hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,B}^{b-1})$, $l_{i,b} \in [1 : 2^{T\hat{R}_i}]$. This defines the codebook for any block $b \in [B+1 : B+M]$,

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

Encoding: Let m be the message to be sent. For any block $b \in [1 : B]$, 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 $(\hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,b-1}), \mathbf{y}_{i,b}, \mathbf{x}_{i,b}(l_{i,b-1})) \in A_\epsilon^{(n)}(X_i, Y_i, \hat{Y}_i)$, where $l_{i,0} = 1$ by convention. The codewords $\mathbf{x}_b(m)$ and $\mathbf{x}_{i,b}(l_{i,b-1}), i \in \mathcal{N}$ are transmitted in block b, $b \in [1 : B]$.

After block B, the source node will be silent and the relay nodes will use additional M blocks to cooperatively transmit $(l_{1,B}, \ldots, l_{n,B})$ to the destination. Specifically, for any block $b \in [B+1 : B+M]$, 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 $(\hat{\mathbf{y}}_{i,b}(l_{i,b}|l_{i,B}^{b-1}), \mathbf{y}_{i,b}, \mathbf{x}_{i,b}(l_{i,B}^{b-1})) \in A_\epsilon^{(n)}(X_i, Y_i, \hat{Y}_i)$. The codeword $\mathbf{x}_{i,b}(l_{i,B}^{b-1}), i \in \mathcal{N}$ is transmitted in block b, $b \in [B+1 : B+M]$.

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 $b = 1, \ldots, B$,

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

and for any $b = B+1, \ldots, B+M$,

$$\Big((\mathbf{X}_{1,b}(l_{1,B}^{b-1}), \hat{\mathbf{Y}}_{1,b}(l_{1,b}|l_{1,B}^{b-1})), \ldots, (\mathbf{X}_{n,b}(l_{n,B}^{b-1}), \hat{\mathbf{Y}}_{n,b}(l_{n,b}|l_{n,B}^{b-1})), \mathbf{Y}_b\Big) \in A_\epsilon^{(n)}(X_\mathcal{N}, \hat{Y}_\mathcal{N}, Y). \tag{34}$$

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 (34) for any $b = B+1, \ldots, B+M$.

Assume the true $\mathbf{l}_B^{B+M} = \mathbf{1}^{M+1}$. Then, error occurs if $\mathbf{l}_B = \mathbf{1}$ does not satisfy (34) 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 (34) 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 (34) 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 (34) with some $\mathbf{l}_{B+1}^{B+M}$ for any $b = B+1, \ldots, B+M$. For any $\mathbf{l}_B^b$, define $\mathcal{A}_b(\mathbf{l}_B^b)$ as the event that $\mathbf{l}_B^b$ satisfies (34). Then, we have

$$\begin{aligned}
\Pr(\bigcup_{\mathbf{l}_B \neq \mathbf{1}} \mathcal{E}_{\mathbf{l}_B}) &= \Pr(\bigcup_{\mathbf{l}_B \neq \mathbf{1}} \bigcup_{\mathbf{l}_{B+1}^{B+M}} \bigcap_{b=B+1}^{B+M} \mathcal{A}_b(\mathbf{l}_B^b)) \\
&\leq \sum_{\mathbf{l}_B \neq \mathbf{1}} \sum_{\mathbf{l}_{B+1}^{B+M}} \Pr(\bigcap_{b=B+1}^{B+M} \mathcal{A}_b(\mathbf{l}_B^b)) \\
&= \sum_{\mathbf{l}_B \neq \mathbf{1}} \sum_{\mathbf{l}_{B+1}^{B+M}} \prod_{b=B+1}^{B+M} \Pr(\mathcal{A}_b(\mathbf{l}_B^b)) \\
&= \sum_{\mathbf{l}_{B+M}} \sum_{\mathbf{l}_{B+1}^{B+M-1}} \sum_{\mathbf{l}_B \neq \mathbf{1}} \prod_{b=B+1}^{B+M} \Pr(\mathcal{A}_b(\mathbf{l}_B^b)).
\end{aligned} \tag{35}$$

For any $\mathbf{l}_B^{B+M}$, let $\mathcal{S}_b(\mathbf{l}_B^{B+M}) = \{i \in \mathcal{N} : l_{i,B}^{b-1} \neq 1^{b-B}\}$. Note $\mathcal{S}_b(\mathbf{l}_B^{B+M})$ only depends on $\mathbf{l}_B^{b-1}$, so we write it as $\mathcal{S}_b(\mathbf{l}_B^{b-1})$. Define $\mathbf{X}_b(\mathcal{S}_b(\mathbf{l}_B^{b-1}))$ as $\{\mathbf{X}_{i,b}(l_{i,B}^{b-1}), i \in \mathcal{S}_b(\mathbf{l}_B^{b-1})\}$, and similarly define $\mathbf{Y}_b(\mathcal{S}_b(\mathbf{l}_B^{b-1}))$ and $\hat{\mathbf{Y}}_b(\mathcal{S}_b(\mathbf{l}_B^{b-1}))$. Then, $(\mathbf{X}_b(\mathcal{S}_b(\mathbf{l}_B^{b-1})), \hat{\mathbf{Y}}_b(\mathcal{S}_b(\mathbf{l}_B^{b-1})))$ is independent of $(\mathbf{X}_b(\mathcal{S}_b^c(\mathbf{l}_B^{b-1})), \hat{\mathbf{Y}}_b(\mathcal{S}_b^c(\mathbf{l}_B^{b-1})), \mathbf{Y}_b)$, and $\Pr(\mathcal{A}_b(\mathbf{l}_B^b))$ can be upper bounded by

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

where $\mathcal{I}(\mathcal{S}_b(\mathbf{l}_B^{b-1})) = I(X_{\mathcal{S}_b(\mathbf{l}_B^{b-1})}; \hat{Y}_{\mathcal{S}_b^c(\mathbf{l}_B^{b-1})}, Y|X_{\mathcal{S}_b^c(\mathbf{l}_B^{b-1})}) - H(\hat{Y}_{\mathcal{S}_b(\mathbf{l}_B^{b-1})}|X_\mathcal{N}, \hat{Y}_{\mathcal{S}_b^c(\mathbf{l}_B^{b-1})}, Y) + \sum_{i\in\mathcal{S}_b(\mathbf{l}_B^{b-1})} H(\hat{Y}_i|X_i)$ and $\epsilon' \to 0$ as $\epsilon \to 0$. Then, following (35), we have

$$\begin{aligned}
\Pr(\bigcup_{\mathbf{l}_B\neq\mathbf{1}} \mathcal{E}_{\mathbf{l}_B}) &\leq \sum_{\mathbf{l}_{B+M}} \sum_{\mathbf{l}_{B+1}^{B+M-1}} \sum_{\mathbf{l}_B\neq\mathbf{1}} \prod_{b=B+1}^{B+M} \Pr(\mathcal{A}_b(\mathbf{l}_B^b))\\
&\leq \sum_{\mathbf{l}_{B+M}} \sum_{\mathbf{l}_{B+1}^{B+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^{b-1}))-\epsilon')}\\
&\leq \sum_{\mathbf{l}_{B+M}} \sum_{\substack{\mathcal{S}_{B+1},\dots,\mathcal{S}_{B+M}:\\ \mathcal{S}_{B+1}\neq\emptyset, \mathcal{S}_{B+1}\subseteq\mathcal{S}_{B+2}\cdots\subseteq\mathcal{S}_{B+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^{b-1}))-\epsilon')}\\
&\leq \sum_{\mathbf{l}_{B+M}} \sum_{\substack{\mathcal{S}_{B+1},\dots,\mathcal{S}_{B+M}:\\ \mathcal{S}_{B+1}\neq\emptyset, \mathcal{S}_{B+1}\subseteq\mathcal{S}_{B+2}\cdots\subseteq\mathcal{S}_{B+M}}} \prod_{b=B+1}^{B+M} 2^{T(\sum_{i\in\mathcal{S}_b}(I(Y_i;\hat{Y}_i|X_i)+\epsilon))} \prod_{b=B+1}^{B+M} 2^{-T(\mathcal{I}(\mathcal{S}_b)-\epsilon')}\\
&\leq \sum_{\mathbf{l}_{B+M}} \sum_{\substack{\mathcal{S}_{B+1},\dots,\mathcal{S}_{B+M}:\\ \mathcal{S}_{B+1}\neq\emptyset, \mathcal{S}_{B+1}\subseteq\mathcal{S}_{B+2}\cdots\subseteq\mathcal{S}_{B+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'')}\\
&\leq \sum_{\mathbf{l}_{B+M}} \sum_{\substack{\mathcal{S}_{B+1},\dots,\mathcal{S}_{B+M}:\\ \mathcal{S}_{B+1}\neq\emptyset, \mathcal{S}_{B+1}\subseteq\mathcal{S}_{B+2}\cdots\subseteq\mathcal{S}_{B+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''\})}
\end{aligned}$$

where $\epsilon'' \to 0$ as $\epsilon \to 0$. Thus, as both T and M go to infinity, $\Pr(\bigcup_{\mathbf{l}_B\neq\mathbf{1}} \mathcal{E}_{\mathbf{l}_B}) \to 0$ if $I(X_\mathcal{S}; \hat{Y}_{\mathcal{S}^c}, Y|X_{\mathcal{S}^c}) - I(Y_\mathcal{S}; \hat{Y}_\mathcal{S}|X_\mathcal{N}, Y, \hat{Y}_{\mathcal{S}^c}) > 0$, for any nonempty $\mathcal{S} \subseteq \mathcal{N}$.

b) 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 backwards property of decoding.

At each block $b = B, B-1, \dots, 2$, 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

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

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

Combining a) and b), we can conclude that $\mathbf{l}^B$ can be decoded if for any nonempty $\mathcal{S}\subseteq\mathcal{N}$,

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

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

$$\Big(\mathbf{X}_b(m),(\mathbf{X}_{1,b}(l_{1,b-1}),\hat{\mathbf{Y}}_{1,b}(l_{1,b}|l_{1,b-1})),\dots,(\mathbf{X}_{n,b}(l_{n,b-1}),\hat{\mathbf{Y}}_{n,b}(l_{n,b}|l_{n,b-1})),\mathbf{Y}_b\Big)\in A_\epsilon^{(n)}(X,X_{\mathcal{N}},\hat{Y}_{\mathcal{N}},Y). \tag{36}$$

Obviously, the probability of decoding error will tend to zero if $R_{\text{R/S}}<I(X;\hat{Y}_{\mathcal{N}},Y|X_{\mathcal{N}})$.

■

Remark 4.1: It should be noted that in the proof of Theorem 2.4, our encoding process is not exactly the same as that in [8]. We add M blocks at the end, during which, the relays encode with memory of the previous blocks, i.e., trying to forward $l_{i,B}^{b-1}$ instead of $l_{i,b-1}$ alone. This ensures that $l_{i,B}$ can be decoded with the help of the subsequent blocks. Then backwardly, all previous $l_{i,B-1},l_{i,B-2},\dots,l_{i,1}$ can be decoded.

Remark 4.2: It is interesting to point out that in the proof of Theorem 2.4, repetitive encoding can be replaced by cumulative encoding, while the same rate can be achieved. Specifically, the source can transmit the message vector $(m_1,m_2,\dots,m_B)$ in the first B blocks and the destination finds the unique message vector $(m_1,m_2,\dots,m_B)$ such that for any $b=1,\dots,B$,

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

One can easily check that all the above analysis still applies. Hence, when the relays encode with memory of previous blocks, collaboration among the blocks is introduced, which has the same effect of improving the achievable rate as using repetitive encoding.

B. Optimality of Successive Decoding under Repetitive Encoding Framework

The proof of Theorem 2.6 is analogous to that of Theorem 2.3. Some useful notations and lemmas paralleled with those in III-C are as follows. The proofs of these lemmas are deferred until we finish the proof of Theorem 2.6.

Let

$$J_{\mathcal{A},\mathcal{B}}(\mathcal{S}):=I(X_{\mathcal{S}};\hat{Y}_{\mathcal{B}\backslash\mathcal{S}},\hat{Y}_{\mathcal{A}},Y|X_{\mathcal{A}},X_{\mathcal{B}\backslash\mathcal{S}})-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_{\mathcal{A}},\hat{Y}_{\mathcal{A}},Y,X_{\mathcal{B}},\hat{Y}_{\mathcal{B}\backslash\mathcal{S}}),\forall\mathcal{S}\subseteq\mathcal{B}, \tag{38}$$

$$J_{\mathcal{B}}(\mathcal{S}):=J_{\emptyset,\mathcal{B}}(\mathcal{S})=I(X_{\mathcal{S}};\hat{Y}_{\mathcal{B}\backslash\mathcal{S}},Y|X_{\mathcal{B}\backslash\mathcal{S}})-I(Y_{\mathcal{S}};\hat{Y}_{\mathcal{S}}|X_{\mathcal{B}},\hat{Y}_{\mathcal{B}\backslash\mathcal{S}},Y),\forall\mathcal{S}\subseteq\mathcal{B}, \tag{39}$$

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

Lemma 4.1: 1) If $J_{\mathcal{A}}(\mathcal{S}_1)\ge 0,\forall\mathcal{S}_1\subseteq\mathcal{A}$, and $J_{\mathcal{B}}(\mathcal{S}_2)\ge 0,\forall\mathcal{S}_2\subseteq\mathcal{B}$, then $J_{\mathcal{A}\bigcup\mathcal{B}}(\mathcal{S})\ge 0,\forall\mathcal{S}\subseteq\mathcal{A}\bigcup\mathcal{B}$.
2) If $J_{\mathcal{A}}(\mathcal{S}_1)\ge 0,\forall\mathcal{S}_1\subseteq\mathcal{A}$, and $J_{\mathcal{A},\mathcal{B}}(\mathcal{S}_2)\ge 0,\forall\mathcal{S}_2\subseteq\mathcal{B}$, then $J_{\mathcal{A}\bigcup\mathcal{B}}(\mathcal{S})\ge 0,\forall\mathcal{S}\subseteq\mathcal{A}\bigcup\mathcal{B}$.

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

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

Lemma 4.3: If $J_{\mathcal{A},\mathcal{B}}(\mathcal{B})\ge 0$ for some nonempty $\mathcal{B}$, then there exists some nonempty $\mathcal{C}\subseteq\mathcal{B}$ such that $J_{\mathcal{A},\mathcal{C}}(\mathcal{S})\ge 0,\forall\mathcal{S}\subseteq\mathcal{C}$.

Lemma 4.4: For any $\mathcal{A}$ and $\mathcal{B}$ with $\mathcal{A}\bigcap\mathcal{B}=\emptyset$, $J(\mathcal{A})+J(\mathcal{B})=J(\mathcal{A}\bigcup\mathcal{B})+J(\mathcal{A}\circ\mathcal{B})$, where

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

We are now ready to present the proof of Theorem 2.6.

Proof of Theorem 2.6: $R^*_{\text{R/S}}$ and $R^*_{\text{R/J}}$ can be respectively written as

$$R^*_{\text{R/S}}=\max_{p(x)\prod_{i=1}^n p(x_i)p(\hat{y}_i|x_i,y_i)} I(X;\hat{Y}_{\mathcal{N}},Y|X_{\mathcal{N}}) \tag{41}$$

$$\text{such that } J(\mathcal{S})\geq 0,\forall\mathcal{S}\subseteq\mathcal{N}, \tag{42}$$

and

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

To show $R^*_{\text{R/S}}=R^*_{\text{R/J}}$, it is sufficient to show that $R^*_{\text{R/J}}$ can be achieved only with $p(x)\prod_{i=1}^n p(x_i)p(\hat{y}_i|x_i,y_i)$ such that $J(\mathcal{S})\geq 0$, $\forall\mathcal{S}\subseteq\mathcal{N}$. Similarly to the proof of Theorem 2.3, this can be proved by two steps and the details are 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 $\mathcal{D}^c\in\underset{\mathcal{S}\subseteq\mathcal{N}}{\operatorname{argmin}}\,J(\mathcal{S})$ and $\bigcap_{\mathcal{T}\in\underset{\mathcal{S}\subseteq\mathcal{N}}{\operatorname{argmin}}J(\mathcal{S})}\mathcal{T}=\mathcal{D}^c$, where $\mathcal{D}$ is defined as in Lemma 4.2 and $\underset{\mathcal{S}\subseteq\mathcal{N}}{\operatorname{argmin}}\,J(\mathcal{S}):=\{\mathcal{T}\subseteq\mathcal{N}:J(\mathcal{T})=\min_{\mathcal{S}\subseteq\mathcal{N}}J(\mathcal{S})\}$.

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

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

Again by Lemma 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 $J_{\mathcal{D}\bigcup\mathcal{B}}(\mathcal{S})\geq 0,\forall\mathcal{S}\subseteq\mathcal{D}\bigcup\mathcal{B}$, 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 $\mathcal{A}\mathcal{D}^c\neq\mathcal{D}^c$, $J(\mathcal{A})>J(\mathcal{A}\bigcup\mathcal{D}^c)\geq\min_{\mathcal{S}\subseteq\mathcal{N}}J(\mathcal{S})$. 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

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

Since $I(\mathcal{A}_2) > I(\mathcal{D}^c)$ by 2), to show $J(\mathcal{A}) > J(\mathcal{A}\bigcup\mathcal{D}^c) \geq \min_{\mathcal{S}\subseteq\mathcal{N}} J(\mathcal{S})$, 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) + 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)\\
&+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) + 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)\\
&-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) - 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)\\
&-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(\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) + 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)\\
&+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) + 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)\\
&-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) - 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) + 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)\\
&+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) + 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 $J(\mathcal{A}) > J(\mathcal{A}_1\bigcup\mathcal{D}^c) \geq \min_{\mathcal{S}\subseteq\mathcal{N}} J(\mathcal{S})$.

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

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

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

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

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

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

ii) Applying exactly the same argument as in Part ii) of the proof of Theorem 2.3, we can obtain that, at the optimum, $\mathcal{D}^c$ must be $\emptyset$, i.e., $\mathcal{D}$ must be $\mathcal{N}$, and thus by the definition of $\mathcal{D}$, $J(\mathcal{S}) \geq 0, \forall\mathcal{S}\subseteq\mathcal{N}$. ■

Below, we summarize the proofs of Lemmas 4.1-4.4.

Proof of Lemma 4.1:

For any $\mathcal{S} \subseteq \mathcal{A} \bigcup \mathcal{B}$, let $\mathcal{S}_1 = \mathcal{S}\mathcal{A}$ and $\mathcal{S}_2 = \mathcal{S}(\mathcal{B} \setminus \mathcal{A})$. Then,

$$
\begin{aligned}
J_{\mathcal{A}\bigcup\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) \qquad (44) \\
\geq& J_{\mathcal{A}}(\mathcal{S}_1) + J_{\mathcal{B}}(\mathcal{S}_2). \qquad (45)
\end{aligned}
$$

If $J_{\mathcal{A}}(\mathcal{S}_1) \geq 0$, $\forall \mathcal{S}_1 \subseteq \mathcal{A}$, and $J_{\mathcal{B}}(\mathcal{S}_2) \geq 0$, $\forall \mathcal{S}_2 \subseteq \mathcal{B}$, then following (45), $J_{\mathcal{A}\bigcup\mathcal{B}}(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{A} \bigcup \mathcal{B}$. 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 (44), $J_{\mathcal{A}\bigcup\mathcal{B}}(\mathcal{S}) \geq 0$, $\forall \mathcal{S} \subseteq \mathcal{A} \bigcup \mathcal{B}$. ■

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.,

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

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

Proof of Lemma 4.4: For any disjoint $\mathcal{A}$ and $\mathcal{B}$,

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

which proves the lemma. ■

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. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, pp. 572–584, 1979.

[3] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the receiver," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 1, pp. 1–11, Jan. 1976.

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

[5] A. El Gamal and Y.-H. Kim, "Lecture notes on network information theory," 2010, available online at http://arxiv.org/abs/1001.3404.

[6] Y.-H. Kim, "Coding techniques for primitive relay channels," in *Proc. Forty-Fifth Annual Allerton Conf. Commun., Contr. Comput.*, Monticello, IL, Sep. 2007.

[7] X. Wu and L.-L. Xie, "An optimality-robustness tradeoff in the compress-and-forward relay scheme," in *Proc. IEEE 72nd Vehicular Technology Conference*, Ottawa, Sep. 2010.

[8] S. H. Lim, Y.-H. Kim, A. El Gamal, S.-Y. Chung, "Noisy Network Coding," submitted to *IEEE Trans. Inform. Theory*, available online at http://arxiv.org/abs/1002.3188.