

A New Achievability Scheme for the Relay Channel*

Wei Kang Sennur Ulukus

Department of Electrical and Computer Engineering

University of Maryland, College Park, MD 20742

wkang@umd.edu ulukus@umd.edu

October 4, 2007

Abstract

In this paper, we propose a new coding scheme for the general relay channel. This coding scheme is in the form of a block Markov code. The transmitter uses a superposition Markov code. The relay compresses the received signal and maps the compressed version of the received signal into a codeword conditioned on the codeword of the previous block. The receiver performs joint decoding after it has received all of the B blocks. We show that this coding scheme can be viewed as a generalization of the well-known Compress-And-Forward (CAF) scheme proposed by Cover and El Gamal. Our coding scheme provides options for preserving the correlation between the channel inputs of the transmitter and the relay, which is not possible in the CAF scheme. Thus, our proposed scheme may potentially yield a larger achievable rate than the CAF scheme.

*This work was supported by NSF Grants CCR 03-11311, CCF 04-47613 and CCF 05-14846, and was presented in part at IEEE Information Theory Workshop, Lake Tahoe, CA, September 2007.

1 Introduction

As the simplest model for cooperative communications, relay channel has attracted plenty of attention since 1971, when it was first introduced by van der Meulen [1]. In 1979, Cover and El Gamal proposed two major coding schemes for the relay channel [2]. These two schemes are widely known as Decode-And-Forward (DAF) and Compress-And-Forward (CAF) today; see [3] for a recent review. These two coding schemes represent two different types of cooperation. In DAF, the cooperation is relatively obvious, where the relay decodes the message from the transmitter, and the transmitter and the relay cooperatively transmit the constructed common information to the receiver in the next block. In CAF, the cooperation spirit is less easy to recognize, as the message is sent by the transmitter only once. However, the relay cooperates with the transmitter by compressing and sending its signal to the receiver. The rate gains in these achievable schemes are due to the fact that, through the channel from the transmitter to the relay, *correlation* is created between the transmitter and the relay, and this correlation is utilized to improve the rates.

In the DAF scheme, correlation is created and then utilized in a block Markov coding structure. More specifically, a *full* correlation is created by decoding the message fully at the relay, which enables the transmitter and the relay to create any kind of joint distribution for the channel inputs in the next block. The shortcoming of the DAF scheme is that by forcing the relay to decode the message in its entirety, it limits the overall achievable rate by the rate from the transmitter to the relay. In contrast, by not forcing a full decoding at the relay, the CAF scheme does not limit the overall rate by the rate from the transmitter to the relay, and may yield higher overall rates. The shortcoming of the CAF scheme, on the other hand, is that the correlation offered by the block coding structure is not utilized effectively, since in each block the channel inputs X and X_1 from the transmitter and the relay are independent, as the transmitter sends the message only once.

However, the essence of good coding schemes in multi-user systems with correlated sources (e.g., [4, 5]) is to preserve the correlation of the sources in the channel inputs. Motivated by this basic observation, in this paper, we propose a new coding scheme for the relay channel, that is based on the idea of preserving the correlation in the channel inputs from the transmitter and the relay. We will show that our new coding scheme may be viewed as a more general version of the CAF scheme, and therefore, our new coding scheme may potentially yield larger rates than the CAF scheme. Our proposed scheme can be further combined with the DAF scheme to yield rates that are potentially larger than those offered by both DAF and CAF schemes, similar in spirit to [2, Theorem 7].

Our new achievability scheme for the relay channel may be viewed as a variation of the coding scheme of Ahlswede and Han [5] for the multiple access channel with a correlated helper. In our work, we view the relay as the helper because the receiver does not need to decode the information sent by the relay. Also, we note that the relay is a *correlated helper* as the communication channel from the transmitter to the relay provides relay for free a

correlated version of the signal sent by the transmitter. The key aspects of the Ahlswede-Han [5] scheme are: to preserve the correlation between the channel inputs of the transmitter and the helper (relay), and for the receiver to decode a “virtual” source, a compressed version of the helper, but not the entire signal of the helper.

Our new coding scheme is in the form of block Markov coding. The transmitter uses a superposition Markov code, similar to the one used in the DAF scheme [2], except in the random codebook generation stage, a method similar to the one in [4] is used in order to preserve the correlation between the blocks. Thus, in each block, the fresh information message is mapped into a codeword conditioned on the codeword of the previous block. Therefore, the overall codebook at the transmitter has a tree structure, where the codewords in block l emanate from the codewords in block $l-1$. The depth of the tree is $B-1$. A similar strategy is applied at the relay side where the compressed version of the received signal is mapped into a two-block-long codeword conditioned on the codeword of the previous block. Therefore, the overall codebook at the relay has a tree structure as well. As a result of this coding strategy, we successfully preserve the correlation between the channel inputs of the transmitter and the relay. However, unlike the DAF scheme where a *full* correlation is acquired through decoding at the relay, our scheme provides only a *partially* correlated helper at the relay by not trying to decode the transmitter’s signal fully. From [4, 5], we note that the channel inputs are correlated through the virtual sources in our case, and therefore, the channel inputs between the consecutive blocks are correlated. This correlation between the blocks will surely hurt the achievable rate. The correlation between the blocks is the price we pay for preserving the correlation between the channel inputs of the transmitter and the relay within any given block.

At the decoding stage, we perform joint decoding for the entire B blocks after all of the B blocks have been received, which is different compared with the DAF and CAF schemes. The reason for performing joint decoding at the receiver is that due to the correlation between the blocks, decoding at any time before the end of all the B blocks would decrease the achievable rate. We note that joint decoding increases the decoding complexity and the delay as compared to DAF and CAF, though neither of these is a major concern in an information theoretic context. The only problem with the joint decoding strategy is that it makes the analysis difficult as it requires the evaluation of some mutual information expressions involving the joint probability distributions of up to B blocks of codes, where B is very large.

The analysis of the error events provides us three conditions containing mutual information expressions involving infinite letters of the underlying random process. Evaluation of these mutual information expressions is very difficult, if not impossible. To obtain a computable result, we lower bound these mutual informations by noting some Markov structure in the underlying random process. This operation gives us three conditions to be satisfied by the achievable rates. These conditions involve eleven variables, the two channel inputs from the transmitter and the relay, the two channel outputs at the relay and the receiver

and the compressed version of the channel output at the relay, in two consecutive blocks, and the channel input from the transmitter in the previous block.

We finish our analysis by revisiting the CAF scheme. We develop an equivalent representation for the achievable rates given in [2] for the CAF scheme. We then show that this equivalent representation for the achievable rates for the CAF scheme is a special case of the achievable rates in our new coding scheme, which is obtained by a special selection of the eleven variables mentioned above. We therefore conclude that our proposed coding scheme yields potentially larger rates than the CAF scheme. More importantly, our new coding scheme creates more possibilities, and therefore a spectrum of new achievable schemes for the relay channel through the selection of the underlying probability distribution, and yields the well-known CAF scheme as a special case, corresponding to a particular selection of the underlying probability distribution.

2 The Relay Channel

Consider a relay channel with finite input alphabets \mathcal{X} , \mathcal{X}_1 and finite output alphabets \mathcal{Y} , \mathcal{Y}_1 , characterized by the transition probability $p(y, y_1|x, x_1)$. An n -length block code for the relay channel $p(y, y_1|x, x_1)$ consists of encoders $f, f_i, i = 1, \dots, n$ and a decoder g

$$\begin{aligned} f &: \mathcal{M} \longrightarrow \mathcal{X}^n \\ f_i &: \mathcal{Y}_1^{i-1} \longrightarrow \mathcal{X}_1, \quad i = 1, \dots, n \\ g &: \mathcal{Y}^n \longrightarrow \mathcal{M} \end{aligned}$$

where the encoder at the transmitter sends $x^n = f(m)$ into the channel, where $m \in \mathcal{M} \triangleq \{1, 2, \dots, M\}$; the encoder at the relay at the i th channel instance sends $x_{1i} = f_i(y_1^{i-1})$ into the channel; the decoder outputs $\hat{m} = g(y^n)$. The average probability of error is defined as

$$P_e = \frac{1}{M} \sum_{m \in \mathcal{M}} Pr(\hat{m} \neq m | m \text{ is transmitted}) \quad (1)$$

A rate R is achievable for the relay channel $p(y, y_1|x, x_1)$ if for every $0 < \epsilon < 1$, $\eta > 0$, and every sufficiently large n , there exists an n -length block code (f, f_i, g) with $P_e \leq \epsilon$ and $\frac{1}{n} \ln M \geq R - \eta$.

3 A New Achievability Scheme for the Relay Channel

We adopt a block Markov coding scheme, similar to the DAF and CAF schemes. We have overall B blocks. In each block, we transmit codewords of length n . We denote the variables in the l th block with a subscript of $[l]$. We denote n -letter codewords transmitted in each block with a superscript of n . Following the standard relay channel literature, we denote

the (random) signals transmitted by the transmitter and the relay by X and X_1 , the signals received at the receiver and the relay by Y and Y_1 , and the compressed version of Y_1 at the relay by \hat{Y}_1 . The realizations of these random signals will be denoted by lower-case letters. For example, the n -letter signals transmitted by the transmitter and the relay in the l th block will be represented by $x_{[l]}^n$ and $x_{1[l]}^n$.

Consider the following discrete time stationary Markov process $G_{[l]} \triangleq (X, \hat{Y}_1, X_1, y, Y_1)_{[l]}$ for $l = 0, 1, \dots, B$, with the transition probability distribution

$$\begin{aligned} p((x, \hat{y}_1, x_1, y, y_1)_{[l]} | (x, \hat{y}_1, x_1, y, y_1)_{[l-1]}) \\ = p(x_{[l]} | x_{[l-1]}) p(y_{1[l]}, y_{[l]} | x_{[l]}, x_{1[l]}) p(x_{1[l]} | \hat{y}_{1[l-1]}) p(\hat{y}_{1[l]} | y_{1[l]}, x_{1[l]}) \end{aligned} \quad (2)$$

The codebook generation and the encoding scheme for the l th block, $l = 1, \dots, B - 1$, are as follows.

Random codebook generation: Let $(x_{[l-1]}^n(m_{[l-1]}), x_{1[l-1]}^n, y_{1[l-1]}^n, y_{[l-1]}^n)$ denote the transmitted and the received signals in the $(l - 1)$ st block, where $m_{[l-1]}$ is the message sent by the transmitter in the $(l - 1)$ st block. An illustration of the codebook structure is shown in Figure 1.

1. For each $x_{[l-1]}^n(m_{[l-1]})$ sequence, generate M sequences, where $x_{[l]}^n(m_{[l]})$, the $m_{[l]}$ th sequence, is generated independently according to $\prod_{i=1}^n p(x_{i[l]} | x_{i[l-1]})$. Here, every code-word in the $(l - 1)$ st block expands into a codebook in the l th block. This expansion is indicated by a directed cone from $x_{[l-1]}^n$ to $x_{[l]}^n$ in Figure 1.
2. For each $x_{1[l-1]}^n$ sequence, generate L $\hat{Y}_{1[l-1]}^n$ sequences independently uniformly distributed in the conditional strong typical set¹ $\mathcal{T}_\delta(x_{1[l-1]}^n)$ with respect to the distribution $p(\hat{y}_{1[l-1]} | x_{1[l-1]})$. If $\frac{1}{n} \ln L > I(Y_{1[l-1]}; \hat{Y}_{1[l-1]} | X_{1[l-1]})$, for any given $y_{1[l-1]}^n$ sequence, there exists one $\hat{y}_{1[l-1]}^n$ sequence with high probability when n is sufficiently large such that $(y_{1[l-1]}^n, \hat{y}_{1[l-1]}^n, x_{1[l-1]}^n)$ are jointly typical according to the probability distribution $p(y_{1[l-1]}, \hat{y}_{1[l-1]}, x_{1[l-1]})$. Denote this $\hat{y}_{1[l-1]}^n$ as $\hat{y}_{1[l-1]}^n(y_{1[l-1]}^n, x_{1[l-1]}^n)$. Here, the quantization from $y_{1[l-1]}^n$ to $\hat{y}_{1[l-1]}^n$, parameterized by $x_{1[l-1]}^n$, is indicated in Figure 1 by a directed cone from $y_{1[l-1]}^n$ to $\hat{y}_{1[l-1]}^n$, with a straight line from $x_{1[l-1]}^n$ for the parameterization.
3. For each $\hat{y}_{1[l-1]}^n$, generate one $x_{1[l]}^n$ sequence according to $\prod_{i=1}^n p(x_{1i[l]} | \hat{y}_{1i[l-1]})$. This one-to-one mapping is indicated by a straight line between $\hat{y}_{1[l-1]}^n$ and $x_{1[l]}^n$ in Figure 1.

Encoding: Let $m_{[l]}$ be the message to be sent in this block. If $(x_{[l-1]}^n(m_{[l-1]}), x_{1[l-1]}^n)$ are sent and $y_{1[l-1]}^n$ is received in the previous block, we choose $(x_{[l]}^n(m_{[l]}), \hat{y}_{1[l-1]}^n(y_{1[l-1]}^n, x_{1[l-1]}^n), x_{1[l]}^n)$ according to the code generation method described above and transmit $(x_{[l]}^n(m_{[l]}), x_{1[l]}^n)$. In

¹Strong typical set and conditional strong typical set are defined in [6, Definition 1.2.8, 1.2.9]. For the sake of simplicity, we omit the subscript which is used to indicate the underlying distribution in [6].

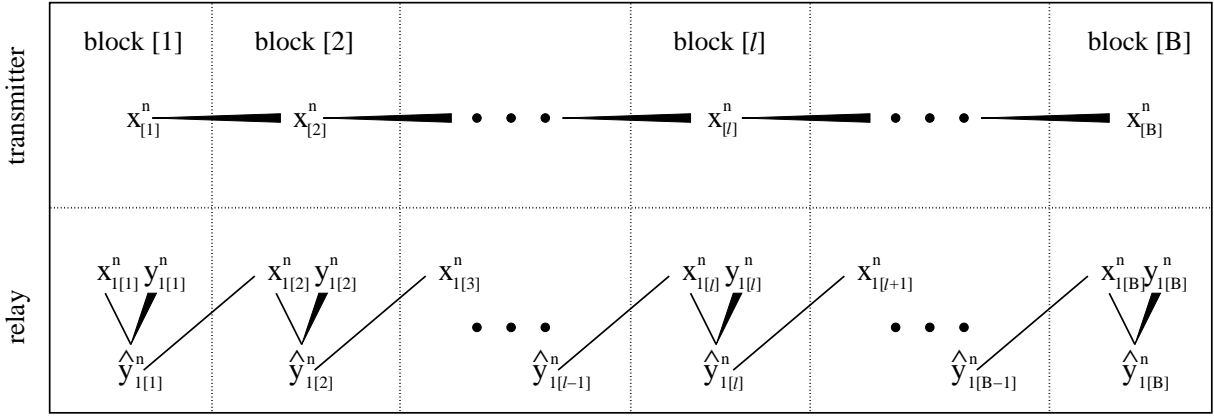


Figure 1: Codebook structure.

the first block, we assume a virtual 0th block, where $(x_{[0]}^n, x_{1[0]}^n, \hat{y}_{1[0]}^n)$, as well as $x_{1[1]}^n$, are known by the transmitter, the relay and the receiver. In the B th block, the transmitter randomly generates one $x_{[B]}^n$ sequence according to $\prod_{i=1}^n p(x_{i[B]} | x_{i[B-1]})$ and sends it into the channel. The relay, after receiving $y_{1[B]}^n$, randomly generates one $\hat{y}_{1[B]}^n$ sequence according to $\prod_{i=1}^n p(\hat{y}_{1[i[B]} | y_{1[i[B]}, x_{1[i[B]})$. We assume that the transmitter and the relay reliably transmit $x_{[B]}^n$ and $\hat{y}_{1[B]}^n$ to the receiver using the next b blocks, where b is some finite positive integer. We note that $B + b$ blocks are used in our scheme, while only the first $B - 1$ blocks carry the message. Thus, the final achievable rate is $\frac{B-1}{B+b} \frac{1}{n} \ln M$ which converges to $\frac{1}{n} \ln M$ for sufficiently large B since b is finite.

Decoding: After receiving B blocks of y^n sequences, i.e., $y_{[1]}^n, \dots, y_{[B]}^n$, and assuming $x_{1[1]}^n, x_{[B]}^n$ and $\hat{y}_{1[B]}^n$ are known at the receiver, we seek $x_{[1]}^n, \dots, x_{[B-1]}^n, \hat{y}_{1[1]}^n, \dots, \hat{y}_{1[B-1]}^n, x_{1[2]}^n, \dots, x_{1[B]}^n$, such that

$$(x_{[1]}^n, \dots, x_{[B]}^n, \hat{y}_{1[1]}^n, \dots, \hat{y}_{1[B]}^n, x_{1[1]}^n, \dots, x_{1[B]}^n, y_{[1]}^n, \dots, y_{[B]}^n) \in \mathcal{T}_\delta$$

according to the stationary distribution of the Markov process $G_{[l]}$ in (2).

The differences between our scheme and the CAF scheme are as follows. At the transmitter side, in our scheme, the fresh message $m_{[l]}$ is mapped into the codeword $x_{[l]}^n$ conditioned on the codeword of the previous block $x_{[l-1]}^n$, while in the CAF scheme, $m_{[l]}$ is mapped into $x_{[l]}^n$, which is generated independent of $x_{[l-1]}^n$. At the relay side, in our scheme, the compressed received signal $\hat{y}_{1[l-1]}^n$ is mapped into the codeword $x_{1[l]}^n$, which is generated according to $p(x_{1[l]} | \hat{y}_{1[l-1]})$, while in the CAF scheme, $x_{1[l]}^n$ is generated independent of $\hat{y}_{1[l-1]}^n$. The aim of our design is to preserve the correlation built in the $(l - 1)$ st block in the channel inputs of the l th block. At the decoding stage, we perform joint decoding for the entire B blocks after all of the B blocks have been received, while in the CAF scheme, the decoding of the message of the $(l - 1)$ st block is performed at the end of the l th block.

Probability of error: When n is sufficiently large, the probability of error can be made arbitrarily small when the following conditions are satisfied.

1. For all j such that $1 \leq j \leq B - 1$,

$$\begin{aligned} & \frac{1}{n}(B - j) \ln M + (B - j)I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) \\ & < I(X_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[B-1]}, X_{1[j+1]}^{[B]}, Y_{[j]}^{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, X_{1[j]}) \quad (3) \end{aligned}$$

2. For all j, k such that $1 \leq j < k \leq B - 1$,

$$\begin{aligned} & \frac{1}{n}(B - j) \ln M + (B - k)I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) \\ & < I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}, Y_{[j]}^{[B]}, \hat{Y}_{1[B]}, X_{1[B]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j+1]}^{[k]} | X_{[j-1]}, X_{[j]}) \quad (4) \end{aligned}$$

3. For all j, k such that $1 \leq k < j \leq B - 1$,

$$\begin{aligned} & (j - k)I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) + \frac{1}{n}(B - j) \ln M + (B - j)I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) \\ & < I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}, Y_{[k]}^{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[k]}^{[j-1]}, X_{1[k]}) \quad (5) \end{aligned}$$

where the subscript $[l]$ on the left hand sides of (3), (4) and (5) indicates that the corresponding random variables belong to a generic sample $g_{[l]}$ of the underlying random process in (2). The details of the calculation of the probability of error where these conditions are obtained can be found in Appendix A.1. The derivation uses standard techniques from information theory, such as counting error events, etc.

In the above conditions, we used the notation $A_{[j]}^{[B]}$ as a shorthand to denote the sequence of random variables $A_{[j]}, A_{[j+1]}, \dots, A_{[B]}$. Consequently, we note that the mutual informations on the right hand sides of (3), (4) and (5) contain vectors of random variables whose lengths go up to B , where B is very large. In order to simplify the conditions in (3), (4) and (5), we lower bound the mutual information expressions on the right hand sides of (3), (4) and (5) by those that involve random variables that belong to up to three blocks. The detailed derivation of the following lower bounding operation can be found in Appendix A.2. The derivation uses standard techniques from information theory, such as the chain rule of mutual information, and exploiting the Markov structure of the involved random variables.

1. For all j such that $1 \leq j \leq B - 1$,

$$\begin{aligned} & (B - j) \left(\frac{1}{n} \ln M + I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) \right) \\ & < (B - j)I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[l-2]}, X_{1[l-1]}, Y_{[l-1]}) \quad (6) \end{aligned}$$

2. For all j, k such that $1 \leq j < k \leq B - 1$,

$$\begin{aligned} (k - j) \frac{1}{n} \ln M + (B - k) \left(\frac{1}{n} \ln M + I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) \right) \\ < (k - j) I(X_{[l]}; Y_{[l]}, \hat{Y}_{1[l]} | X_{1[l]}, Y_{[l-1]}, \hat{Y}_{1[l-1]}, X_{1[l-1]}, X_{[l-2]}) \\ + (B - k) I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[l-2]}, X_{1[l-1]}, Y_{[l-1]}) \end{aligned} \quad (7)$$

3. For all j, k such that $1 \leq k < j \leq B - 1$,

$$\begin{aligned} (j - k) I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) + (B - j) \left(\frac{1}{n} \ln M + I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) \right) \\ < (j - k) I(Y_{[l]}; \hat{Y}_{1[l]}, X_{1[l]} | X_{[l]}, X_{[l-1]}, X_{1[l-1]}, Y_{[l-1]}) \\ + (B - j) I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[l-2]}, X_{1[l-1]}, Y_{[l-1]}) \end{aligned} \quad (8)$$

We can further derive sufficient conditions for the above three conditions in (6), (7) and (8) as follows. We define the following quantities:

$$C_1 \triangleq \frac{1}{n} \ln M + I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) \quad (9)$$

$$C_2 \triangleq \frac{1}{n} \ln M \quad (10)$$

$$C_3 \triangleq I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) \quad (11)$$

$$D_1 \triangleq I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[l-2]}, X_{1[l-1]}, Y_{[l-1]}) \quad (12)$$

$$D_2 \triangleq I(X_{[l]}; Y_{[l]}, \hat{Y}_{1[l]} | X_{1[l]}, Y_{[l-1]}, \hat{Y}_{1[l-1]}, X_{1[l-1]}, X_{[l-2]}) \quad (13)$$

$$D_3 \triangleq I(Y_{[l]}; \hat{Y}_{1[l]}, X_{1[l]} | X_{[l]}, X_{[l-1]}, X_{1[l-1]}, Y_{[l-1]}) \quad (14)$$

Then, the sufficient conditions in (6), (7) and (8) can also be written as,

1. For all j such that $1 \leq j \leq B - 1$,

$$(B - j)C_1 < (B - j)D_1 \quad (15)$$

2. For all j, k such that $1 \leq j < k \leq B - 1$,

$$(k - j)C_2 + (B - k)C_1 < (k - j)D_2 + (B - k)D_1 \quad (16)$$

3. For all j, k such that $1 \leq k < j \leq B - 1$,

$$(j - k)C_3 + (B - j)C_1 < (j - k)D_3 + (B - j)D_1 \quad (17)$$

We note that the above conditions are implied by the following three conditions,

$$C_1 < D_1 \quad (18)$$

$$C_2 < D_2 \quad (19)$$

$$C_3 < D_3 \quad (20)$$

or in other words, by,

$$R - \eta \leq \frac{1}{n} \ln M < I(X_{[l]}; Y_{[l]}, \hat{Y}_{1[l]} | X_{1[l]}, Y_{[l-1]}, \hat{Y}_{1[l-1]}, X_{1[l-1]}, X_{[l-2]}) \quad (21)$$

$$I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) < I(Y_{[l]}; \hat{Y}_{1[l]}, X_{1[l]} | X_{[l]}, X_{[l-1]}, X_{1[l-1]}, Y_{[l-1]}) \quad (22)$$

$$R - \eta + I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) < I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[l-2]}, X_{1[l-1]}, Y_{[l-1]}) \quad (23)$$

The expressions in (21), (22) and (23) give sufficient conditions to be satisfied by the rate in order for the probability of error to become arbitrarily close to zero. We note that these conditions depend on variables used in three consecutive blocks, l , $l - 1$ and $l - 2$. With this development, we obtain the main result of our paper which is stated in the following theorem.

Theorem 1 *The rate R is achievable for the relay channel, if the following conditions are satisfied*

$$R \leq I(Y, \hat{Y}_1; X | X_1, \tilde{Y}_1, \tilde{Y}, \tilde{X}_1, \tilde{X}) \quad (24)$$

$$I(\hat{Y}_1; Y_1 | X_1, X) < I(Y; \hat{Y}_1, X_1 | X, \tilde{Y}, \tilde{X}, \tilde{X}_1) \quad (25)$$

$$R + I(\hat{Y}_1; Y_1 | X_1, X) \leq I(Y; \hat{Y}_1, X_1, X | \tilde{Y}, \tilde{X}_1, \tilde{X}) \quad (26)$$

where

$$\tilde{\tilde{X}} \longrightarrow (\tilde{X}, \tilde{\tilde{Y}}_1, \tilde{\tilde{X}}_1, \tilde{Y}, \tilde{Y}_1) \longrightarrow (X, \hat{Y}_1, X_1, Y, Y_1) \quad (27)$$

$$p(x, \hat{y}_1, x_1, y, y_1, \tilde{x}) = p(\tilde{x}, \tilde{\tilde{y}}_1, \tilde{\tilde{x}}_1, \tilde{y}, \tilde{y}_1, \tilde{\tilde{x}}) \quad (28)$$

$$p(x, \hat{y}_1, x_1, y, y_1 | \tilde{x}, \tilde{\tilde{y}}_1, \tilde{\tilde{x}}_1, \tilde{y}, \tilde{y}_1) = p(x | \tilde{x}) p(x_1 | \tilde{\tilde{y}}_1) p(y_1, y | x, x_1) p(\hat{y}_1 | y_1, x_1) \quad (29)$$

In the above theorem, the notations $\tilde{\cdot}$ and $\tilde{\tilde{\cdot}}$ are used to denote the signals belonging to the previous block and the block before the previous block, respectively, with respect to a reference block. Therefore, we see that the achievable rate in the relay channel, using our proposed coding scheme, needs to satisfy three conditions that involve mutual information expressions calculated using eleven variables which satisfy the Markov chain constraint in (27), the marginal distribution constraint in (28), and the additional inter-block probability distribution constraint in (29).

In the next section, we will revisit the well-known CAF scheme proposed in [2]. First, we

will develop an equivalent representation for the well-known representation of the achievable rate in the CAF scheme. We will then show that the rates achievable by the CAF scheme can be achieved with our proposed scheme by choosing a certain special structure for the joint probability distribution of the eleven random variables in Theorem 1 while still satisfying the three conditions in (27), (28) and (29).

4 Revisiting the Compress-And-Forward (CAF) Scheme

In [2], the achievable rates for the CAF are characterized as in the following theorem.

Theorem 2 ([2]) *The rate R is achievable for the relay channel, if the following conditions are satisfied*

$$R \leq I(X; Y, \hat{Y}_1 | X_1) \quad (30)$$

$$I(Y_1; \hat{Y}_1 | X_1, Y) < I(X_1; Y) \quad (31)$$

where

$$p(x, x_1, y, y_1, \hat{y}_1) = p(x)p(x_1)p(y, y_1|x, x_1)p(\hat{y}_1|y_1, x_1) \quad (32)$$

In the following theorem, we present three equivalent forms for the rate achievable by the CAF scheme.

Theorem 3 *The following three conditions are equivalent.*

1. For some $p(x, x_1, y, y_1, \hat{y}_1) = p(x)p(x_1)p(y, y_1|x, x_1)p(\hat{y}_1|y_1, x_1)$

$$R - I(X; \hat{Y}_1 | X_1) \leq I(X; Y | \hat{Y}_1, X_1) \quad (33)$$

$$I(Y_1; \hat{Y}_1 | X_1) < I(\hat{Y}_1; Y | X_1) + I(X_1; Y) \quad (34)$$

2. For some $p(x, x_1, y, y_1, \hat{y}_1) = p(x)p(x_1)p(y, y_1|x, x_1)p(\hat{y}_1|y_1, x_1)$

$$R - I(X; \hat{Y}_1 | X_1) \leq I(X; Y | \hat{Y}_1, X_1) \quad (35)$$

$$R - I(X; \hat{Y}_1 | X_1) + I(Y_1; \hat{Y}_1 | X_1) \leq I(X, \hat{Y}_1; Y | X_1) + I(X_1; Y) \quad (36)$$

3. For some $p(x, x_1, y, y_1, \hat{y}_1) = p(x)p(x_1)p(y, y_1|x, x_1)p(\hat{y}_1|y_1, x_1)$

$$R - I(X; \hat{Y}_1 | X_1) \leq I(X; Y | \hat{Y}_1, X_1) \quad (37)$$

$$I(\hat{Y}_1; Y_1 | X_1, X) < I(\hat{Y}_1; Y | X_1, X) + I(X_1; Y | X) \quad (38)$$

$$R - I(X; \hat{Y}_1 | X_1) + I(Y_1; \hat{Y}_1 | X_1) \leq I(X, \hat{Y}_1; Y | X_1) + I(X_1; Y) \quad (39)$$

The proof of the above theorem is given in Appendix A.3.

We rewrite the final equivalent representation in (37), (38) and (39) in the following more compact form in order to compare the rates achievable with our proposed scheme and the rates achievable with the CAF scheme in the next section.

$$R \leq I(X; Y, \hat{Y}_1 | X_1) \quad (40)$$

$$I(\hat{Y}_1; Y_1 | X_1, X) < I(\hat{Y}_1, X_1; Y | X) \quad (41)$$

$$R + I(Y_1; \hat{Y}_1 | X_1, X) \leq I(X, \hat{Y}_1, X_1; Y) \quad (42)$$

5 Comparison of the Achievable Rates with Our Scheme and with the CAF Scheme

We note that the conditions on the achievable rates with our scheme given in Theorem 1, i.e., (24), (25), (26), are very similar to the final equivalent form for the conditions on the achievable rates with the CAF scheme, i.e., (40), (41), (42), except for two differences. First, the channel inputs of the transmitter and the relay, i.e., X and X_1 , in our proposed scheme can be correlated, while in the CAF scheme they are independent, and second, in our scheme there are some extra random variables, which mutual information expressions are conditioned on, e.g., $\tilde{X}, \tilde{X}_1, \tilde{Y}, \tilde{Y}_1, \tilde{\tilde{X}}$. These two differences come from our coding scheme where we introduced correlation between the channel inputs of the transmitter and the relay in a block, and between the variables across the blocks. The correlation between the channel inputs from the transmitter and the relay in any block is an advantage, as for channels which favor correlation, this translates into higher rates. However, the correlation across the blocks is a disadvantage as it decreases the efficiency of transmission, and therefore the achievable rates. In fact, the price we pay for the correlation between the channel inputs in any given block is precisely the correlation we have created across the blocks. For a given correlation structure, it is not clear which of these two opposite effects will overcome the other. That is, the rate of our scheme for a certain correlated distribution may be lower or higher than the rate of the CAF scheme. However, we note that the CAF scheme can be viewed as a special case of our proposed scheme by choosing an independent distribution, i.e., by choosing the following conditional distribution in (29)

$$p(x, \hat{y}_1, x_1, y, y_1 | \tilde{x}, \tilde{\hat{y}}_1, \tilde{x}_1, \tilde{y}, \tilde{y}_1) = p(x)p(x_1)p(y_1, y | x, x_1)p(\hat{y}_1 | x_1, y_1) \quad (43)$$

In this case, the expressions in Theorem 1, i.e., (24), (25), (26), degenerate into the third equivalent form for the CAF scheme in Theorem 3, i.e., (40), (41), (42). The above observation implies that the maximum achievable rate with our proposed scheme over all possible distributions is not less than the achievable rate of the CAF scheme. Thus, we can claim that this paper offers more choices in the achievability scheme than the CAF scheme, and

that these choices may potentially yield larger achievable rates than those offered by the CAF scheme.

A Appendix

A.1 Probability of Error Calculation

The average probability of decoding error can be expressed as follows,

$$P_e = Pr(E_1 \cup E_2) = Pr(E_1) + Pr(E_2 \cap E_1^c) \quad (44)$$

where

$$E_1 \triangleq (x_{[1,\dots,B]}^n, \hat{y}_{1[1,\dots,B]}^n, x_{1[1,\dots,B]}^n, y_{[1,\dots,B]}^n) \notin \mathcal{T}_\delta \quad (45)$$

$$E_2 \triangleq \bigcup_{(\bar{x}_{[1,\dots,B]}^n, \bar{\hat{y}}_{1[1,\dots,B-1]}^n) \neq (x_{[1,\dots,B]}^n, \hat{y}_{1[1,\dots,B-1]}^n)} (\bar{x}_{[1,\dots,B]}^n, \bar{\hat{y}}_{1[1,\dots,B]}^n, \bar{x}_{1[1,\dots,B]}^n, y_{[1,\dots,B]}^n) \in \mathcal{T}_\delta \quad (46)$$

where $(\bar{x}_{[1,\dots,B]}^n, \bar{\hat{y}}_{1[1,\dots,B-1]}^n, \bar{x}_{1[2,\dots,B]}^n)$ is another codeword that is generated according to the rules of our scheme.

From (2), we note the following Markov properties:

1. conditioned on $(\hat{Y}_{1[l]}, X_{[l]}, X_{1[l]}), Y_{[l]}$ is independent of $G_{[\dots,l-1]}$ and $G_{[l,\dots]}$;
2. conditioned on $(X_{[l-1]}, \hat{Y}_{1[l-1]}), G_{[l,\dots]}$ is independent of $G_{[\dots,l-1]}$.

Here, and in the sequel, subscript $[l]$ refers to a generic block within overall B blocks.

$Pr(E_1)$ can be upper bounded as follows:

$$\begin{aligned} Pr(E_1) &\leq \sum_{l=1}^B \left(Pr \left((x_{[l]}^n, x_{1[l]}^n, y_{[l]}^n, y_{1[l]}^n, g_{[\dots,l-1]}^n) \notin \mathcal{T}_\delta | g_{[\dots,l-1]}^n \in \mathcal{T}_\delta \right) \right. \\ &\quad \left. + Pr \left((\hat{y}_{1[l]}^n, x_{[l]}^n, x_{1[l]}^n, y_{[l]}^n, y_{1[l]}^n, g_{[\dots,l-1]}^n) \notin \mathcal{T}_\delta | (x_{[l]}^n, x_{1[l]}^n, y_{[l]}^n, y_{1[l]}^n, g_{[\dots,l-1]}^n) \in \mathcal{T}_\delta \right) \right) \quad (47) \end{aligned}$$

From the way the code is generated, we have

$$Pr \left((x_{[l]}^n, x_{1[l]}^n, y_{[l]}^n, y_{1[l]}^n, g_{[\dots,l-1]}^n) \notin \mathcal{T}_\delta | g_{[\dots,l-1]}^n \in \mathcal{T}_\delta \right) \leq \epsilon \quad (48)$$

The compression from $y_{1[l]}^n$ to $\hat{y}_{1[l]}^n$ is a conditional version of a rate-distortion code. If $R' > I(Y_1; \hat{Y}_1 | X_1)$, then, when n is sufficiently large, we have

$$Pr \left((\hat{y}_{1[l]}^n, x_{[l]}^n, x_{1[l]}^n, y_{[l]}^n, y_{1[l]}^n, g_{[\dots,l-1]}^n) \notin \mathcal{T}_\delta | (x_{[l]}^n, x_{1[l]}^n, y_{[l]}^n, y_{1[l]}^n, g_{[\dots,l-1]}^n) \in \mathcal{T}_\delta \right) \leq \epsilon \quad (49)$$

Thus,

$$Pr(E_1) \leq 2B\epsilon \quad (50)$$

Now we switch to the error event E_2 .

$$\begin{aligned}
& Pr(E_2 \cap E_1^c) \\
&= \sum_{\substack{(x_{[1,\dots,B]}^n, \hat{y}_{1[1,\dots,B]}^n, x_{1[1,\dots,B]}^n, y_{[1,\dots,B]}^n) \in \mathcal{T}_\delta \\ \times Pr(E_2 | (x_{[1,\dots,B]}^n, \hat{y}_{1[1,\dots,B]}^n, x_{1[1,\dots,B]}^n, y_{[1,\dots,B]}^n) \text{ sent})} \\
&\leq \max_{\substack{(x_{[1,\dots,B]}^n, \hat{y}_{1[1,\dots,B]}^n, x_{1[1,\dots,B]}^n, y_{[1,\dots,B]}^n) \in \mathcal{T}_\delta}} Pr(E_2 | (x_{[1,\dots,B]}^n, \hat{y}_{1[1,\dots,B]}^n, x_{1[1,\dots,B]}^n, y_{[1,\dots,B]}^n) \text{ sent}) \quad (51)
\end{aligned}$$

From our proposed coding scheme, we note that the codebooks at both transmitter and relay have tree structures with $B - 1$ stages. A correct codeword $x_{[1,\dots,B-1]}^n$ can be viewed as a path in the tree-structured codebook at the transmitter. Similarly, for the codeword $\hat{y}_{1[1,\dots,B-1]}^n$ at the relay. An error occurs when we diverge from the correct path at a certain stage in the tree. Thus, the error event E_2 can be decomposed as

$$\begin{aligned}
E_2 = & \bigcup_{\substack{j=2,\dots,B-1 \\ k=2,\dots,B-1}} \bigcup_{\substack{(\bar{x}_{[1]}^n, \dots, \bar{x}_{[j-1]}^n, \bar{y}_{1[1]}^n, \dots, \bar{y}_{1[k-1]}^n) = (x_{[1]}^n, \dots, x_{[j-1]}^n, \hat{y}_{1[1]}^n, \dots, \hat{y}_{1[k-1]}^n) \\ (\bar{x}_{[j]}^n, \bar{y}_{1[k]}^n) \neq (x_{[j]}^n, \hat{y}_{1[k]}^n)}} \\
& (\bar{x}_{[1]}^n, \dots, \bar{x}_{[B]}^n, \bar{y}_{1[1]}^n, \dots, \bar{y}_{1[B]}^n, \bar{x}_{1[1]}^n, \dots, \bar{x}_{1[B]}^n, y_{[1]}^n, \dots, y_{[B]}^n) \in \mathcal{T}_\delta \quad (52)
\end{aligned}$$

where each term in the union in the above equation represents the error event that results when we diverge from the correct paths at the j th stage at the transmitter and at the k th stage at the relay.

Let us define \mathcal{F}_1 to be the set consisting of all feasible codeword pairs $(x_{[j]}^n, \hat{y}_{1[j]}^n)$ for the j th block for a given $x_{[j-1]}^n$ and $x_{1[j]}^n$. Then, we have

$$\begin{aligned}
F_1 \triangleq |\mathcal{F}_1| &\leq M \exp(n(H(\hat{Y}_{1[j]} | X_{[j]}, X_{1[j]}) + 2\epsilon)) \frac{L}{(1 - \epsilon) \exp(n(H(\hat{Y}_{1[j]} | X_{1[j]}) - 2\epsilon))} \\
&\leq M \exp(n(H(\hat{Y}_{1[j]} | X_{[j]}, X_{1[j]}) + 2\epsilon)) \frac{\exp(n(I(\hat{Y}_{1[j]}; Y_{1[j]} | X_{1[j]}) + \epsilon))}{(1 - \epsilon) \exp(n(H(\hat{Y}_{1[j]} | X_{1[j]}) - 2\epsilon))} \\
&\leq M \exp(n(I(\hat{Y}_{1[j]}; Y_{1[j]} | X_{1[j]}, X_{[j]}) + 6\epsilon)) \quad (53)
\end{aligned}$$

We also define \mathcal{F}_2 to be the set consisting of all feasible codewords $x_{[j]}^n$ for the j th block for a given $x_{[j-1]}^n$. Then,

$$F_2 \triangleq |\mathcal{F}_2| = M \quad (54)$$

Similarly, we define \mathcal{F}_3 to be the set consisting of all feasible codewords $\hat{y}_{1[j]}^n$ for the j th block

for a given $x_{[j]}^n$ and $x_{1[j]}^n$. Then,

$$\begin{aligned} F_3 \triangleq |\mathcal{F}_3| &\leq L \frac{\exp(n(H(\hat{Y}_{1[j]}|X_{1[j]}, X_{[j]}) + 2\epsilon))}{(1 - \epsilon) \exp(n(H(\hat{Y}_{1[j]}|X_{1[j]}) - 2\epsilon))} \\ &\leq \exp(n(I(\hat{Y}_{1[j]}; Y_{1[j]}|X_{1[j]}, X_{[j]}) + 6\epsilon)) \end{aligned} \quad (55)$$

We define the error event E_{2jk}

$$\begin{aligned} E_{2jk} \triangleq &\bigcup \\ &(\bar{x}_{[1]}^n, \dots, \bar{x}_{[j-1]}^n, \bar{y}_{1[1]}^n, \dots, \bar{y}_{1[k-1]}^n) = (x_{[1]}^n, \dots, x_{[j-1]}^n, \hat{y}_{1[1]}^n, \dots, \hat{y}_{1[k-1]}^n) \\ &(\bar{x}_{[j]}^n, \bar{y}_{1[k]}^n) \neq (x_{[j]}^n, \hat{y}_{1[k]}^n) \\ &(\bar{x}_{[1]}^n, \dots, \bar{x}_{[B]}^n, \bar{y}_{1[1]}^n, \dots, \bar{y}_{1[B]}^n, \bar{x}_{1[1]}^n, \dots, \bar{x}_{1[B]}^n, y_{[1]}^n, \dots, y_{[B]}^n) \in \mathcal{T}_\delta \end{aligned} \quad (56)$$

Then, we have

$$Pr(E_2 \cap E_1^c) \leq \sum_{j=2}^{B-1} \sum_{k=2}^{B-1} Pr(E_{2jk} \cap E_1^c) \quad (57)$$

and

$$Pr(E_{2jk} \cap E_1^c) \leq |\mathcal{A}_{jk}| \max_{(\bar{x}_{[1]}^n, \dots, \bar{x}_{[B-1]}^n, \bar{y}_{1[1]}^n, \dots, \bar{y}_{1[B-1]}^n) \in \mathcal{A}_{jk}} P_1(\bar{x}_{[1]}^n, \dots, \bar{x}_{[B-1]}^n, \bar{y}_{1[1]}^n, \dots, \bar{y}_{1[B-1]}^n) \quad (58)$$

where

$$\begin{aligned} \mathcal{A}_{jk} \triangleq &\left\{ \begin{array}{l} \text{codeword } (\bar{x}_{[1]}^n, \dots, \bar{x}_{[B-1]}^n, \bar{y}_{1[1]}^n, \dots, \bar{y}_{1[B-1]}^n) : \\ \left(\bar{x}_{[1]}^n, \dots, \bar{x}_{[j-1]}^n, \bar{y}_{1[1]}^n, \dots, \bar{y}_{1[k-1]}^n \right) = \left(x_{[1]}^n, \dots, x_{[j-1]}^n, \hat{y}_{1[1]}^n, \dots, \hat{y}_{1[k-1]}^n \right) \\ \left(\bar{x}_{[j]}^n, \bar{y}_{1[k]}^n \right) \neq \left(x_{[j]}^n, \hat{y}_{1[k]}^n \right) \end{array} \right\} \end{aligned} \quad (59)$$

$$\begin{aligned} &P_1(\bar{x}_{[1]}^n, \dots, \bar{x}_{[B-1]}^n, \bar{y}_{1[1]}^n, \dots, \bar{y}_{1[B-1]}^n) \\ &\triangleq Pr((\bar{x}_{[1]}^n, \dots, \bar{x}_{[B]}^n, \bar{y}_{1[1]}^n, \dots, \bar{y}_{1[B]}^n, \bar{x}_{1[1]}^n, \dots, \bar{x}_{1[B]}^n, y_{[1]}^n, \dots, y_{[B]}^n) \in \mathcal{T}_\delta) \end{aligned} \quad (60)$$

given $(x_{[1]}^n, \dots, x_{[B]}^n, \hat{y}_{1[1]}^n, \dots, \hat{y}_{1[B]}^n, x_{1[1]}^n, \dots, x_{1[B]}^n, y_{[1]}^n, \dots, y_{[B]}^n) \in \mathcal{T}_\delta$.

In order to have the probability of such error events go to zero, we need the following conditions to hold.

When $j = k$, from the structure of the block Markov code and (53), we have

$$|\mathcal{A}_{jk}| = F_1^{B-j} \leq M^{B-j} \exp(n(B-j)(I(\hat{Y}_{1[j]}; Y_{1[j]}|X_{1[j]}, X_{[j]}) + 6\epsilon)) \quad (61)$$

and

$$\begin{aligned}
& P_1(\bar{x}_{[1]}^n, \dots, \bar{x}_{[B-1]}^n, \bar{y}_{[1]}^n, \dots, \bar{y}_{[B-1]}^n) \\
& \leq \exp(n(H(X_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[B-1]}, X_{1[j+1]}^{[B]} | Y_{[j]}^{[B]}, \hat{Y}_{1[B]}, X_{[B]}, X_{[j-1]}, X_{1[j]} + 2\epsilon)) \\
& \quad \times \exp(-n(H(X_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[B-1]}, X_{1[j+1]}^{[B]} | X_{[j-1]}, X_{1[j]} - 2\epsilon)) \\
& = \exp(n(-I(X_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[B-1]}, X_{1[j+1]}^{[B]}; Y_{[j]}^{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, X_{1[j]} + 4\epsilon)) \quad (62)
\end{aligned}$$

When $j < k$, we have

$$|\mathcal{A}_{jk}| = F_2^{k-j} F_1^{B-k} \leq M^{B-j} \exp(n(B-k)(I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]} + 6\epsilon)) \quad (63)$$

and

$$\begin{aligned}
& P_1(\bar{x}_{[1]}^n, \dots, \bar{x}_{[B-1]}^n, \bar{y}_{[1]}^n, \dots, \bar{y}_{[B-1]}^n) \\
& \leq \exp(n(H(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]} | Y_{[j]}^{[B]}, \hat{Y}_{1[B]}, X_{[B]}, \hat{Y}_{1[j]}^{[k-1]}, X_{[j-1]}, X_{1[j]}^{[k]} + 2\epsilon)) \\
& \quad \times \exp(-n(H(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]} | X_{[j-1]}, X_{1[j]} - 2\epsilon)) \\
& = \exp(n(-I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[j]}^{[B]}, \hat{Y}_{1[B]}, X_{[B]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j+1]}^{[k]} | X_{[j-1]}, X_{1[j]} + 4\epsilon)) \quad (64)
\end{aligned}$$

When $j > k$, we have

$$\begin{aligned}
|\mathcal{A}_{jk}| & = F_3^{j-k} F_1^{B-j} \leq \exp(n(j-k)(I(\hat{Y}_{1[j]}; Y_{1[j]} | X_{1[j]}, X_{[j]} + 6\epsilon)) \\
& \quad \times M_l^{B-k} \exp(n(B-k)(I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]} + 6\epsilon)) \quad (65)
\end{aligned}$$

and

$$\begin{aligned}
& P_1(\bar{x}_{[1]}^n, \dots, \bar{x}_{[B-1]}^n, \bar{y}_{[1]}^n, \dots, \bar{y}_{[B-1]}^n) \\
& \leq \exp(n(H(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]} | Y_{[k]}^{[B]}, \hat{Y}_{1[B]}, X_{[B]}, X_{[k]}^{[j-1]}, X_{1[k]} + 2\epsilon)) \\
& \quad \times \exp(-n(H(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]} | X_{[k]}^{[j-1]}, X_{1[k]} - 2\epsilon)) \\
& = \exp(n(-I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[k]}^{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[k]}^{[j-1]}, X_{1[k]} + 4\epsilon)) \quad (66)
\end{aligned}$$

Thus, when n is sufficiently large, using (58) and (61) through (66), we have

$$Pr(E_{2jk} \cap E_1^c) \leq \epsilon, \quad j, k = 2, \dots, B-1 \quad (67)$$

if the following conditions are satisfied:

1. For all j such that $1 \leq j \leq B - 1$,

$$\begin{aligned} & \frac{1}{n}(B - j) \ln M + (B - j)I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) \\ & < I(X_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[B-1]}, X_{1[j+1]}^{[B]}; Y_{[j]}^{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, X_{1[j]}) \end{aligned} \quad (68)$$

2. For all j, k such that $1 \leq j < k \leq B - 1$,

$$\begin{aligned} & \frac{1}{n}(B - j) \ln M + (B - k)I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) \\ & < I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[j]}^{[B]}, \hat{Y}_{1[B]}, X_{1[B]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j+1]}^{[k]} | X_{[j-1]}, X_{1[j]}) \end{aligned} \quad (69)$$

3. For all j, k such that $1 \leq k < j \leq B - 1$,

$$\begin{aligned} & (j - k)I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) + \frac{1}{n}(B - j) \ln M + (B - j)I(\hat{Y}_{1[l]}; Y_{1[l]} | X_{1[l]}, X_{[l]}) \\ & < I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[k]}^{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[k]}^{[j-1]}, X_{1[k]}) \end{aligned} \quad (70)$$

Therefore, we have

$$P_e = Pr(E_1) + Pr(E_2 \cap E_1^c) \leq (2B + B^2)\epsilon \quad (71)$$

When n is sufficiently large, $(2B + B^2)\epsilon$ can be made arbitrarily small.

A.2 Lower Bounding the Mutual Informations in (3), (4), (5)

For the right hand side of (3), we have

$$\begin{aligned}
& I(X_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[B-1]}, X_{1[j+1]}^{[B]}; Y_{[j]}^{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, X_{1[j]}) \\
& \stackrel{1}{=} \sum_{l=j}^{B-1} I(X_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[B-1]}, X_{1[j+1]}^{[B]}; Y_{[l]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[l-1]}) \\
& \quad + I(X_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[B-1]}, X_{1[j+1]}^{[B]}; Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[B-1]}) \\
& \stackrel{2}{=} I(Y_{[j]}; X_{[j]}, \hat{Y}_{1[j]} | X_{1[j]}, X_{[j-1]}) + \sum_{l=j+1}^{B-1} I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[l-1]}) \\
& \quad + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{1[B]}, X_{[B-1]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[B-1]}) \\
& \stackrel{3}{=} \sum_{l=j+1}^{B-1} I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[l-1]}) + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]} | X_{1[B]}, X_{[B-1]}) \\
& \quad + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{1[B]}, X_{[B-1]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[B-1]}) \\
& \stackrel{4}{\geq} \sum_{l=j+1}^{B-1} I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[l-1]}) \\
& \quad + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]} | X_{1[B]}, X_{[B-1]}, X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[B-1]}) \\
& \quad + I(Y_{[B]}; X_{1[B]}, X_{[B-1]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[B-1]}) \\
& = \sum_{l=j+1}^{B-1} I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[l-1]}) \\
& \quad + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]}, X_{[B-1]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[B-1]}) \\
& \stackrel{5}{=} \sum_{l=j+1}^B I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[j-1]}, X_{1[j]}, Y_{[j]}^{[l-1]}) \\
& \stackrel{6}{\geq} (B-j)I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[l-2]}, X_{1[l-1]}, Y_{[l-1]}) \tag{72}
\end{aligned}$$

where

1. follows from the chain rule;
2. because of Markov properties 1 and 2;
3. because of the stationarity of the random process and the property that conditioning reduces entropy;
4. because of Markov property 2;
5. because of Markov property 1;
6. because of Markov property 2 and the stationarity of the random process.

For the right hand side of (4), we have

$$\begin{aligned}
& I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[j]}^{[B]}, \hat{Y}_{1[B]}^{[B]}, X_{[B]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j+1]}^{[k]} | X_{[j-1]}, X_{1[j]}) \\
& \stackrel{1}{=} I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[j]}, \hat{Y}_{1[j]}^{[B]} | X_{[j-1]}, X_{1[j]}) \\
& \quad + \sum_{l=j+1}^{k-1} I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[l]}, \hat{Y}_{1[l]}^{[B]}, X_{1[l]} | X_{[j-1]}, Y_{[j]}^{[l-1]}, \hat{Y}_{1[j]}^{[l-1]}, X_{1[j]}^{[l-1]}) \\
& \quad + I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[k]}, X_{1[k]} | X_{[j-1]}, Y_{[j]}^{[k-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k-1]}) \\
& \quad + \sum_{l=k+1}^{B-1} I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[l]} | X_{[j-1]}, Y_{[j]}^{[l-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \quad + I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[B]}, \hat{Y}_{1[B]}^{[B]}, X_{[B]} | X_{[j-1]}, Y_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \stackrel{2}{\geq} I(X_{[j]}; Y_{[j]}, \hat{Y}_{1[j]} | X_{[j-1]}, X_{1[j]}) + \sum_{l=j+1}^{k-1} I(X_{[l]}; Y_{[l]}, \hat{Y}_{1[l]} | X_{[j-1]}, Y_{[j]}^{[l-1]}, \hat{Y}_{1[j]}^{[l-1]}, X_{1[j]}^{[l]}) \\
& \quad + I(X_{[k]}, \hat{Y}_{1[k]}; Y_{[k]} | X_{[j-1]}, Y_{[j]}^{[k-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \quad + \sum_{l=k+1}^{B-1} I(X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]}; Y_{[l]} | X_{[j-1]}, Y_{[j]}^{[l-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \quad + I(X_{[B-1]}, X_{1[B]}; Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, Y_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \stackrel{3}{=} \sum_{l=j+1}^{k-1} I(X_{[l]}; Y_{[l]}, \hat{Y}_{1[l]} | X_{[j-1]}, Y_{[j]}^{[l-1]}, \hat{Y}_{1[j]}^{[l-1]}, X_{1[j]}^{[l]}) \\
& \quad + \sum_{l=k+1}^{B-1} I(X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]}; Y_{[l]} | X_{[j-1]}, Y_{[j]}^{[l-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \quad + I(X_{[B]}; Y_{[B]}, \hat{Y}_{1[B]} | X_{[B-1]}, X_{1[B]}) \\
& \quad + I(X_{[B]}, \hat{Y}_{1[B]}; Y_{[B]} | X_{[j-1+B-k]}, Y_{[j+B-k]}^{[B-1]}, \hat{Y}_{1[j+B-k]}^{[B-1]}, X_{1[j+B-k]}^{[B]}) \\
& \quad + I(X_{[B-1]}, X_{1[B]}; Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, Y_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \stackrel{4}{\geq} \sum_{l=j+1}^{k-1} I(X_{[l]}; Y_{[l]}, \hat{Y}_{1[l]} | X_{[j-1]}, Y_{[j]}^{[l-1]}, \hat{Y}_{1[j]}^{[l-1]}, X_{1[j]}^{[l]}) \\
& \quad + \sum_{l=k+1}^{B-1} I(X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]}; Y_{[l]} | X_{[j-1]}, Y_{[j]}^{[l-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \quad + I(X_{[B]}; Y_{[B]}, \hat{Y}_{1[B]} | X_{1[B]}, S) + I(X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]}; Y_{[B]} | S) \\
& \stackrel{5}{\geq} (k-j)I(X_{[l]}; Y_{[l]}, \hat{Y}_{1[l]} | X_{1[l]}, Y_{[l-1]}, \hat{Y}_{1[l-1]}, X_{1[l-1]}, X_{[l-2]}) \\
& \quad + (B-k)I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[l-2]}, X_{1[l-1]}, Y_{[l-1]})
\end{aligned} \tag{73}$$

where

$$S \triangleq (X_{[j-1+B-k]}, Y_{[j+B-k]}^{[B-1]}, \hat{Y}_{1[j+B-k]}^{[B-1]}, X_{1[j+B-k]}^{[B-1]}, X_{[j-1]}, Y_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \quad (74)$$

and

1. follows from the chain rule;
2. because of Markov properties 1 and 2;
3. because of the stationarity of the random process;
4. because of the following derivation

$$\begin{aligned}
& I(X_{[B]}; Y_{[B]}, \hat{Y}_{1[B]} | X_{[B-1]}, X_{1[B]}) \\
& \quad + I(X_{[B]}, \hat{Y}_{1[B]}; Y_{[B]} | X_{[j-1+B-k]}, Y_{[j+B-k]}^{[B-1]}, \hat{Y}_{1[j+B-k]}^{[B-1]}, X_{1[j+B-k]}^{[B]}) \\
& \quad + I(X_{[B-1]}, X_{1[B]}; Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[j-1]}, Y_{[j]}^{[B-1]}, \hat{Y}_{1[j]}^{[k-1]}, X_{1[j]}^{[k]}) \\
& \geq I(X_{[B]}; Y_{[B]}, \hat{Y}_{1[B]} | X_{[B-1]}, X_{1[B]}, S) + I(X_{[B]}, \hat{Y}_{1[B]}; Y_{[B]} | X_{1[B]}, S) \\
& \quad + I(X_{[B-1]}, X_{1[B]}; Y_{[B]}, \hat{Y}_{1[B]} | S) \\
& \geq I(X_{[B]}; Y_{[B]}, \hat{Y}_{1[B]} | X_{[B-1]}, X_{1[B]}, S) + I(X_{[B]}, \hat{Y}_{1[B]}; Y_{[B]} | X_{1[B]}, S) \\
& \quad + I(X_{[B-1]}; Y_{[B]}, \hat{Y}_{1[B]} | X_{1[B]}, S) + I(X_{1[B]}; Y_{[B]} | S) \\
& = I(X_{[B]}; Y_{[B]}, \hat{Y}_{1[B]} | X_{1[B]}, S) + I(X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]}; Y_{[B]} | S) \quad (75)
\end{aligned}$$

5. because of Markov property 1 and 2 and the stationarity of the random process.

For the right hand side of (5), we have

$$\begin{aligned}
& I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[k]}^{[B]}, \hat{Y}_{1[B]}^{[B]}, X_{[B]} | X_{[k]}^{[j-1]}, X_{1[k]}) \\
& \stackrel{1}{=} \sum_{l=k}^{B-1} I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[l]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[l-1]}) \\
& \quad + I(X_{[j]}^{[B-1]}, \hat{Y}_{1[k]}^{[B-1]}, X_{1[k+1]}^{[B]}; Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[B-1]}) \\
& \stackrel{2}{\geq} I(Y_{[k]}; \hat{Y}_{1[k]} | X_{[k]}, X_{1[k]}) + \sum_{l=k+1}^{j-1} I(Y_{[l]}; \hat{Y}_{1[l]}, X_{1[l]} | X_{[k]}^{[l]}, X_{1[k]}, Y_{[k]}^{[l-1]}) \\
& \quad + I(Y_{[j]}; X_{[j]}, \hat{Y}_{1[j]}, X_{1[j]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[j-1]}) \\
& \quad + \sum_{l=j+1}^{B-1} I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[l-1]}) \\
& \quad + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{[j]}^{[B-1]}, X_{1[B]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[B-1]}) \\
& \stackrel{3}{=} \sum_{l=k+1}^{j-1} I(Y_{[l]}; \hat{Y}_{1[l]}, X_{1[l]} | X_{[k]}^{[l]}, X_{1[k]}, Y_{[k]}^{[l-1]}) \\
& \quad + \sum_{l=j+1}^{B-1} I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[l-1]}) + I(Y_{[B]}; \hat{Y}_{1[B]} | X_{[B]}, X_{1[B]}) \\
& \quad + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]} | X_{[k+B-j]}^{[B-1]}, X_{1[k+B-j]}, Y_{[k+B-j]}^{[B-1]}) \\
& \quad + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{[j]}^{[B-1]}, X_{1[B]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[B-1]}) \\
& \stackrel{4}{\geq} \sum_{l=k+1}^{j-1} I(Y_{[l]}; \hat{Y}_{1[l]}, X_{1[l]} | X_{[j]}^{[l]}, X_{1[k]}, Y_{[k]}^{[l-1]}) \\
& \quad + \sum_{l=j+1}^{B-1} I(Y_{[l]}; X_{[l]}, \hat{Y}_{1[l]}, X_{1[l]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[l-1]}) \\
& \quad + I(Y_{[B]}; \hat{Y}_{1[B]}, X_{1[B]} | X_{[j]}^{[B]}, S') + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]} | S') \\
& \stackrel{5}{\geq} (j-k)I(Y_{[j]}; \hat{Y}_{1[j]}, X_{1[j]} | X_{[j]}, X_{1[j-1]}, Y_{[j-1]}) \\
& \quad + (B-j)I(Y_{[j]}; X_{[j]}, \hat{Y}_{1[j]}, X_{1[j]} | X_{[j-2]}, X_{1[j-1]}, Y_{[j-1]}) \tag{76}
\end{aligned}$$

where

$$S' \triangleq (X_{1[k+B-j]}, Y_{[k]}^{[B-1]}, X_{[k]}^{[j-1]}, X_{1[k]}) \tag{77}$$

and

1. follows from the chain rule;
2. because of Markov properties 1 and 2;
3. because of the stationarity of the random process;

4. because of the following derivation

$$\begin{aligned}
& I(Y_{[B]}; \hat{Y}_{1[B]} | X_{[B]}, X_{1[B]}) + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]} | X_{[k+B-j]}^{[B-1]}, X_{1[k+B-j]}, Y_{[k+B-j]}^{[B-1]}) \\
& \quad + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{[j]}^{[B-1]}, X_{1[B]} | X_{[k]}^{[j-1]}, X_{1[k]}, Y_{[k]}^{[B-1]}) \\
& \geq I(Y_{[B]}; \hat{Y}_{1[B]} | X_{[B]}, X_{1[B]}, S') + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]} | X_{[j]}^{[B-1]}, S') \\
& \quad + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{[j]}^{[B-1]}, X_{1[B]} | S') \\
& = I(Y_{[B]}; \hat{Y}_{1[B]} | X_{[B]}, X_{1[B]}, S') + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]} | X_{[j]}^{[B-1]}, S') \\
& \quad + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{1[B]} | X_{[j]}^{[B-1]}, S') + I(Y_{[B]}, \hat{Y}_{1[B]}, X_{[B]}; X_{[j]}^{[B-1]} | S') \\
& \geq I(Y_{[B]}; \hat{Y}_{1[B]} | X_{[B]}, X_{1[B]}, X_{[j]}^{[B-1]}, S') + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]} | X_{[j]}^{[B-1]}, S') \\
& \quad + I(Y_{[B]}; X_{1[B]} | X_{[B]}, X_{[j]}^{[B-1]}, S') + I(Y_{[B]}; X_{[j]}^{[B-1]} | S') \\
& = I(Y_{[B]}; \hat{Y}_{1[B]}, X_{1[B]} | X_{[j]}^{[B]}, S') + I(Y_{[B]}; X_{[B]}, \hat{Y}_{1[B]}, X_{1[B]} | S') \tag{78}
\end{aligned}$$

5. because of Markov property 1 and 2 and the stationarity of the random process.

A.3 Proof of Theorem 3

First, we note that condition 1 is equivalent to the expression in Theorem 2. We also note that condition 2 is seemingly weaker than condition 1 because (36) is implied by (33) and (34), and condition 3 is seemingly stronger than condition 2 because condition 3 consists of every element in condition 2 plus (38). Even though they seem different, these three conditions are indeed equivalent. The equivalence of conditions 2 and 3 is shown in [5]. Here, we use a similar proof technique to show the equivalence of conditions 1 and 2 as follows². For a given distribution $p(x, x_1, y, y_1, \hat{y}_1)$, condition 1 is stronger than condition 2, which means that an arbitrary rate R satisfying condition 1 will also satisfy condition 2. Conversely, for a rate R satisfying condition 2, if (34) is satisfied, then condition 1 is satisfied. If (34) is not satisfied, i.e.,

$$I(Y_1; \hat{Y}_1 | X_1) \geq I(\hat{Y}_1; Y | X_1) + I(X_1; Y) \tag{79}$$

we know that $R \in [0, R^*]$, where

$$R^* - I(X; \hat{Y}_1 | X_1) \leq I(X; Y | \hat{Y}_1, X_1) \tag{80}$$

$$R^* - I(X; \hat{Y}_1 | X_1) + I(Y_1; \hat{Y}_1 | X_1) = I(X, \hat{Y}_1; Y | X_1) + I(X_1; Y) \tag{81}$$

²A similar result is given in [7] by means of time-sharing.

That is, R^* is defined such that (36) is satisfied with equality. We may rewrite (80) and (81) as

$$R^* \leq I(X; Y|X_1) + I(X; \hat{Y}_1|Y, X_1) \quad (82)$$

$$R^* = I(X, X_1; Y) - I(Y_1; \hat{Y}_1|X, X_1, Y) \quad (83)$$

We define a new random variable \hat{Y}'_1 such that \hat{Y}'_1 has the same marginal distribution as \hat{Y}_1 and $\hat{Y}'_1 \rightarrow \hat{Y}_1 \rightarrow (Y_1, X, X_1, Y)$. Due to the continuity of mutual information, there exists a choice of \hat{Y}'_1 such that $I(X; \hat{Y}'_1|Y, X_1) = A$ for any $A \in [0, I(X; \hat{Y}_1|Y, X_1)]$. If $R^* - I(X; Y|X_1) > 0$, we choose \hat{Y}'_1 such that $R^* = I(X; Y|X_1) + I(X; \hat{Y}'_1|Y, X_1)$. We note that, in this case, $I(Y_1; \hat{Y}'_1|X, X_1, Y) \geq I(Y_1; \hat{Y}_1|X, X_1, Y)$. Thus,

$$R^* = I(X; Y|X_1) + I(X; \hat{Y}'_1|Y, X_1) \quad (84)$$

$$R^* \leq I(X, X_1; Y) - I(Y_1; \hat{Y}'_1|X, X_1, Y) \quad (85)$$

which means that R^* satisfies condition 1 with joint distribution $p(x, x_1, y, y_1, \hat{y}'_1)$ and so does any $R \leq R^*$. If $R^* - I(X; Y|X_1) \leq 0$, we choose \hat{Y}'_1 independent of $(\hat{Y}_1, X, X_1, Y_1, Y)$. In this case,

$$R^* \leq I(X; Y|X_1) + I(X; \hat{Y}'_1|Y, X_1) = I(X; Y|X_1) \quad (86)$$

$$0 = I(Y_1; \hat{Y}'_1|X_1) \leq I(\hat{Y}'_1; Y|X_1) + I(X_1; Y) \quad (87)$$

Therefore, in this case, R^* satisfies condition 1 with joint distribution $p(x, x_1, y, y_1, \hat{y}'_1)$ and so does any $R \leq R^*$.

As we mentioned above the equivalence between condition 2 and 3 is shown in [5]. For completeness, we restate their proof here as follows. For a given distribution $p(x, x_1, y, y_1, \hat{y}_1)$, condition 3 is stronger than condition 2, which means that an arbitrary rate R satisfying condition 3 will also satisfy condition 2. Conversely, for a rate R satisfying condition 2, if (38) is satisfied, then condition 3 is satisfied. If (38) is not satisfied, i.e., the following inequalities are satisfied

$$R - I(X; \hat{Y}_1|X_1) \leq I(X; Y|\hat{Y}_1, X_1) \quad (88)$$

$$I(\hat{Y}_1; Y_1|X_1, X) \geq I(\hat{Y}_1; Y|X_1, X) + I(X_1; Y|X) \quad (89)$$

$$R - I(X; \hat{Y}_1|X_1) + I(Y_1; \hat{Y}_1|X_1) \leq I(X, \hat{Y}_1; Y|X_1) + I(X_1; Y) \quad (90)$$

then the following inequalities are satisfied also, since we simply drop the first inequality,

$$I(\hat{Y}_1; Y_1|X_1, X) \geq I(\hat{Y}_1; Y|X_1, X) + I(X_1; Y|X) \quad (91)$$

$$R - I(X; \hat{Y}_1|X_1) + I(Y_1; \hat{Y}_1|X_1) \leq I(X, \hat{Y}_1; Y|X_1) + I(X_1; Y) \quad (92)$$

By combining (91) and (92), we have

$$\begin{aligned}
R &\leq I(X; \hat{Y}_1|X_1) - I(Y_1; \hat{Y}_1|X_1) + I(\hat{Y}_1; Y_1|X_1, X) \\
&\quad + I(X, \hat{Y}_1; Y|X_1) + I(X_1; Y) - I(\hat{Y}_1; Y|X_1, X) - I(X_1; Y|X) \\
&\leq I(X; Y|X_1) - (I(X_1; Y|X) - I(X_1; Y)) \\
&\leq I(X; Y|X_1)
\end{aligned} \tag{93}$$

which implies condition 3, i.e., (37), (38) and (39), with \hat{Y}_1 set to be a constant.

References

- [1] E. C. van der Meulen. Three-terminal communication channels. *Adv. App. Prob.*, 3:120–154, 1971.
- [2] T. M. Cover and A. El Gamal. Capacity theorems for the relay channel. *IEEE Trans. Inform. Theory*, 25:572–584, Sep. 1979.
- [3] G. Kramer, M. Gastpar, and P. Gupta. Cooperative strategies and capacity theorems for relay networks. *IEEE Trans. Inform. Theory*, 51(9):3037–3063, September 2005.
- [4] T. M. Cover, A. El Gamal, and M. Salehi. Multiple access channel with arbitrarily correlated sources. *IEEE Trans. Inform. Theory*, 26:648–657, Nov. 1980.
- [5] R. Ahlswede and T. S. Han. On source coding with side information via a multiple-access channel and related problems in multi-user information theory. *IEEE Trans. Inform. Theory*, 29(3):396–412, 1983.
- [6] I. Csiszar and J. Korner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.
- [7] R. Dabora and S. Servetto. On the role of estimate-and-forward with time-sharing in cooperative communications. *Submitted to IEEE Transactions on Information Theory*, 2006, <http://cn.ece.cornell.edu/publications/papers/20060529/pp.pdf>.