On the Relay Channel with Receiver-Transmitter Feedback

Shraga I. Bross, *Member, IEEE*, and Michèle A. Wigger, *Student Member, IEEE*

Abstract—An achievable rate for the discrete memoryless relay channel with receiver-transmitter feedback is proposed based on Block-Markov superposition encoding. The achievable rate can also be extended to Gaussian channels. A second achievable rate for the Gaussian relay channel based on a Schalkwijk-Kailath type scheme is presented. For some channels both achievable rates strictly improve upon all previously known achievable rates. For the discrete memoryless relay channel also a converse result is provided.

Index Terms—Discrete memoryless relay channel, receiver-transmitter feedback.

I. INTRODUCTION

The relay channel was introduced by van der Meulen in [1]. In [3] Cover and El Gamal proposed two coding schemes which are based on the idea of Block-Markov superposition encoding: the decode-and-forward scheme and the compress-and-forward scheme. The decode-and-forward scheme was proved to be optimal for physically degraded relay channels [3]—i.e., for channels where the output observed at the receiver is a degraded version of the channel output at the relay—and for semi-deterministic channels [5]—i.e., for channels where the channel output at the relay is a deterministic function of the channel input at the transmitter and the channel input at the relay. However, for general relay channels the capacity is not known to date. The best known upper bound on capacity is given by the cut-set upper bound [3] and the best known lower bound is due to Chong et. al [13]. The scheme in [13] is a Block-Markov transmission scheme that represents a combination of the decode-and-forward scheme and the compress-and-forward scheme, while the transmitted message is recovered at the receiver by use of backward decoding combined with simultaneous decoding. The use of simultaneous decoding, instead of a sequential form of decoding, results in a relaxed constraint on the compression ratio of the data sent via the compress-and-forward approach and thus (possibly) enlarges the achievable rate. The idea of applying backward decoding to a Block-Markov transmission scheme in order to allow a simultaneous form of decoding was first introduced by Willems and van der Meulen for the multiple-access channel with cribbing encoders [20].

The described upper and lower bounds on the capacity of discrete memoryless relay channels hold also for Gaussian relay channels. In addition, for Gaussian relay channels El Gamal et al. [14] recently proposed coding schemes which

S. Bross is with the School of Engineering, Bar-Ilan University, Ramat Gan, 52900, Israel. Email:brosss@macs.biu.ac.il.

M. Wigger is with the Departement of Electrical Engineering, ETH Zürich, CH-8092 Zurich, Switzerland. Email:wigger@isi.ee.ethz.ch.

apply a simple linear strategy at the relay. Surprisingly, for certain choices of the channel parameters, these simple schemes outperform the more involved Block-Markov schemes in [3], [13].

The relay channel with feedback has first been considered in [3] where it was shown that in the presence of receiver-relay feedback the channel becomes a physically degraded relay channel and hence the capacity can be achieved using the decode-and-forward scheme. For the settings without receiver-relay feedback, but with either receiver-transmitter feedback, relay-transmitter feedback, or both, the capacity is still unknown except for the same special cases as in the no-feedback setting: in the semi-deterministic case and in the physically degraded case. Here a relay-transmitter feedback link [2] and/or a receiver-transmitter feedback link (Observation 1 in this work) does not improve capacity compared to the channel without feedback. However, for general relay channels with either receiver-transmitter feedback or with relay-transmitter feedback the achievable rates reported in [6] are strictly larger than the best known achievable rates without feedback.

Outer bounds for relay channels with either receiver-transmitter feedback or relay-transmitter feedback include the cut-set bound and (for relay-transmitter feedback only) the upper bound in [2].

In this work we focus on relay channels with receiver-transmitter feedback. We study the discrete memoryless relay channel as well as the Gaussian memoryless relay channel. For the discrete memoryless relay channel we propose a new lower bound and a new upper bound on the channel capacity. The lower bound is due to a coding scheme which combines the ideas of restricted decoding used in [21], the nested binning used in [6], and the generalized coding strategy for the relay channel in [3]. The proposed upper bound recovers the existence of an auxiliary random variable which shows up in a similar way in most known lower bounds. In the lower bounds the auxiliary random variable arises when building up correlation between the signals sent at the transmitter and at the relay. However, our upper bound is not tighter than the cut-set upper bound.

The aforementioned bounds apply also to Gaussian relay channels with receiver-transmitter feedback. For some channels the proposed lower bound strictly improves on all previous lower bounds.

For the Gaussian relay channel a second lower bound on the capacity with receiver-transmitter feedback is derived. This lower bound is based on a scheme which builds upon a Schalkwijk-Kailath type strategy [15] at the transmitter and a simple amplify-and-forward strategy at the relay. The achiev-

able rate of this scheme is shown to exceed all previously known achievable rates (including our first lower bound) for some channels, e.g., for channels where the available power at the relay is much smaller than the available power at the encoder. In particular, it demonstrates that for some channels the per-sample estimation based feedback signaling approach of [15] – [19] outperforms the best known Block-Markov feedback encoding approach.

The paper is organized as follows. The next section introduces the general relay channel with receiver-transmitter feedback and recalls some basic results that will be used to establish the achievable rates. In Section III we present our results concerning the time-discrete finite input/output-alphabet memoryless relay channel, while Section IV presents the results concerning the time-discrete memoryless Gaussian relay channel. In Section V we prove the achievability of the rates reported in Sections III-IV, as well as our converse result.

II. Preliminaries

A. Notation

Henceforth, we adopt the following notation conventions. Random variables will be denoted by capital letters, while their realizations will be denoted by the respective lower case letters. Whenever the dimension of a random vector is clear from the context the random vector will be denoted by a bold face letter, that is, $\boldsymbol{X}$ denotes the random vector $(X_1, X_2, \ldots, X_n)$, and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ will designate a specific sample value of $\boldsymbol{X}$. However, in those cases where it is important to emphasize explicitly the dimension of a random vector — $X_1^{n_1}$ shall denote the random vector $(X_{1,1}, X_{1,2}, \ldots, X_{1,n_1})$ and $x_1^{n_1}$ shall denote the sample vector $(x_{1,1}, \ldots, x_{1,n_1})$. The alphabet of a scalar random variable X will be designated by a caligraphic letter $\mathcal{X}$. The n-fold Cartesian power of a generic alphabet $\mathcal{V}$, that is, the set of all n-vectors over $\mathcal{V}$, will be denoted $\mathcal{V}^n$.

B. Setting

The discrete memoryless relay channel is a triple $(\mathcal{X}_1 \times \mathcal{X}_2, p(y, y_1 | x_1, x_2), \mathcal{Y} \times \mathcal{Y}_1)$, where $\mathcal{X}_1$ and $\mathcal{X}_2$ are finite sets corresponding to the input alphabets of the sender and the relay respectively, where $\mathcal{Y}_1$ and $\mathcal{Y}$ are finite sets corresponding to the output alphabets of the relay and the receiver respectively, and where $p(\cdot, \cdot | x_1, x_2)$ is a collection of probability laws on $\mathcal{Y} \times \mathcal{Y}_1$ indexed by the input symbols $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$. The channel's law extends to n-tuples according to the memoryless law

$$p(y_k, y_{1,k} | x_1^k, x_2^k, y^{k-1}, y_1^{k-1}) = p(y_k, y_{1,k} | x_{1,k}, x_{2,k}),$$

where $x_{1,k}, x_{2,k}, y_k$, and $y_{1,k}$ denote the inputs and outputs of the channel at time k, respectively. The investigated communication model is shown in Fig. 1.

The sender has access to an information source which every n channel uses emits a random integer W that is uniformly distributed over the set $\{1, \ldots, M\}$. The goal of the transmission is to communicate the message W to the receiver.
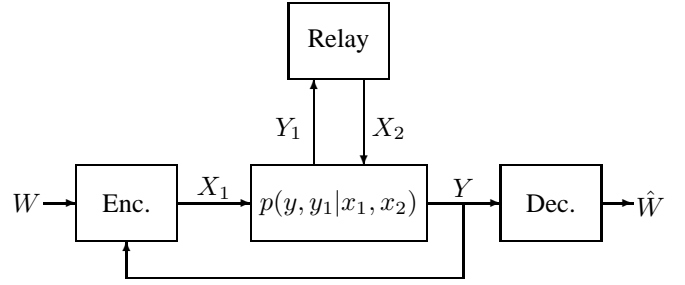

Fig. 1.   The relay channel with receiver-transmitter feedback.

We assume a causal and noiseless feedback link from the receiver to the transmitter. Then both encoders are completely described by corresponding sets of n encoding functions. These functions map the message and the sequence of previous receiver outputs, or the sequence of previous relay outputs, respectively, into the next channel inputs. Specifically,

$$
\begin{align}
x_{1,k} &= f_{1,k}(W, Y_1, Y_2, \ldots, Y_{k-1}), \tag{1}\\
x_{2,k} &= f_{2,k}(Y_{1,1}, Y_{1,2}, \ldots, Y_{1,k-1}), \tag{2}
\end{align}
$$

where $f_{2,1}$ is allowed to be a stochastic function.

The decoder observes the sequence of n channel outputs and estimates W based on that. Formally, the decoder is described by a mapping $\phi : \mathcal{Y}^n \to \{1, \ldots, M\}$ such that

$$\hat{W} = \phi(Y_1, \ldots, Y_n) . \tag{3}$$

An (M, n, ϵ)-code for the relay with receiver-transmitter feedback consists of two sets of n encoding functions as in (1) and (2), and a decoder mapping ϕ such as in (3) such that

$$P_e \triangleq \Pr\left[\hat{W} \neq W\right] \leq \epsilon.$$

A rate R is said to be achievable if for any $\epsilon > 0$ there exists for all n sufficiently large an (M, n, ϵ)-code such that $(1/n) \ln M \geq R$. The supremum over all achievable rates is defined as the capacity.

C. Previous Results

In [6], the following achievability result has been reported for a relay channel in the presence of receiver-transmitter feedback.

Theorem 1 ([6, Theorem 1]): Consider the discrete memoryless relay channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y, y_1 | x_1, x_2), \mathcal{Y} \times \mathcal{Y}_1)$ with receiver-transmitter feedback. Then the rate $\bar{R}_1$ defined by

$$\bar{R}_1 = \sup_{p_{\tilde{V} X_1 X_2 Y Y_1 \hat{Y}_1}} I(X_1; Y \hat{Y}_1 | \tilde{V} X_2), \tag{4}$$

is achievable subject to the constraints

$$
\begin{align}
I(\hat{Y}_1; Y_1 | \tilde{V} X_1 X_2 Y) &\leq I(X_2; Y | \tilde{V} X_1), \tag{5}\\
I(\hat{Y}_1; Y_1 | \tilde{V} X_2 Y) &\leq I(\tilde{V} X_2; Y). \tag{6}
\end{align}
$$

Here the supremum in (4) is taken over all laws on $\tilde{\mathcal{V}} \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y} \times \mathcal{Y}_1 \times \hat{\mathcal{Y}}_1$ of the form

$$
\begin{aligned}
& p_{\tilde{V} X_1 X_2 Y Y_1 \hat{Y}_1}(\tilde{v}, x_1, x_2, y, y_1, \hat{y}_1) \\
& = p_{\tilde{V}}(\tilde{v}) p_{X_1|\tilde{V}}(x_1|\tilde{v}) p_{X_2|\tilde{V}}(x_2|\tilde{v}) p(y, y_1|x_1, x_2) \\
& \quad \cdot p_{\hat{Y}_1|Y_1 X_2 \tilde{V}}(\hat{y}_1|y_1, x_2, \tilde{v}),
\end{aligned} \tag{7}
$$

and the cardinalities of both auxiliary random variables $\tilde{V}$ and $\hat{Y}_1$ can be bounded as follows

$$
\begin{aligned}
\|\tilde{\mathcal{V}}\| &\leq \|\mathcal{X}_1\|\|\mathcal{X}_2\|\|\mathcal{Y}_1\| + 2, \\
\|\hat{\mathcal{Y}}_1\| &\leq \|\tilde{\mathcal{V}}\|\|\mathcal{X}_1\|\|\mathcal{X}_2\|\|\mathcal{Y}_1\| + 2.
\end{aligned}
$$

To date the best known upper bound is given by the cut-set bound derived by Cover and El Gamal in [3].

Theorem 2 ([3, Theorem 4]): Let $(\mathcal{X}_1 \times \mathcal{X}_2, p(y, y_1|x_1, x_2), \mathcal{Y} \times \mathcal{Y}_1)$ be a discrete memoryless relay channel *without feedback*. If

$$
R > \sup_{p_{X_1 X_2 Y Y_1}} \min\{I(X_1 X_2; Y), \ I(X_1; Y Y_1|X_2)\}, \tag{8}
$$

then there exists $\lambda > 0$ such that $P_e > \lambda$ for all n. Here the supremum in (8) is taken over all laws on $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y} \times \mathcal{Y}_1$ of the form

$$
p_{X_1 X_2 Y Y_1}(x_1, x_2, y, y_1) = p_{X_1 X_2}(x_1, x_2) p(y, y_1|x_1, x_2). \tag{9}
$$

In general it is not known whether the cut-set upper bound is achievable. However, El Gamal and Aref showed in [5] that for semi-deterministic relay channels *without feedback* a decode-and-forward strategy achieves the cut-set upper bound.

Theorem 3 ([5]): Let $(\mathcal{X}_1 \times \mathcal{X}_2, p(y, y_1|x_1, x_2), \mathcal{Y} \times \mathcal{Y}_1)$ be a discrete memoryless relay channel *without feedback* and let the channel output at the relay y_1 be a deterministic function of the two channel inputs x_1 and x_2. Then

$$
C = \sup_{p_{X_1 X_2 Y Y_1}} \min\left\{ \begin{array}{c} I(X_1 X_2; Y), \\ I(X_1; Y|X_2 Y_1) + H(Y_1|X_2) \end{array} \right\}, \tag{10}
$$

where the supremum is taken over all joint laws on $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y} \times \mathcal{Y}_1$ of the form

$$
p_{X_1 X_2 Y Y_1}(x_1, x_2, y, y_1) = p_{X_1 X_2}(x_1, x_2) p(y, y_1|x_1, x_2). \tag{11}
$$

III. DISCRETE MEMORYLESS RELAY CHANNEL

Our main result for discrete memoryless relay channels with receiver-transmitter feedback is an achievability result. By combining the generalized coding strategy from [3, Theorem 7], the nested binning technique from [6] and the restricted decoding from [21] we prove the following.

Theorem 4: Consider the discrete memoryless relay channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y, y_1|x_1, x_2), \mathcal{Y} \times \mathcal{Y}_1)$ with receiver-transmitter feedback. Then the rate $\bar{R}_2$ defined by

$$
\bar{R}_2 = \sup_{p_{UV X_1 X_2 Y Y_1 \hat{Y}_1}} \min\left\{ \begin{array}{c} I(X_1; Y\hat{Y}_1|UX_2) + I(U; Y_1|VX_2), \\ I(X_1 X_2; Y) - I(\hat{Y}_1; Y_1|UX_1 X_2 Y) \end{array} \right\}, \tag{12}
$$

is achievable subject to the constraints

$$
I(\hat{Y}_1; Y_1|UX_1 X_2 Y) \leq I(X_2; Y|UVX_1), \tag{13}
$$

$$
I(\hat{Y}_1; Y_1|UX_2) \leq I(\hat{Y}_1; Y|UX_2) + I(VX_2; Y). \tag{14}
$$

The supremum in (12) is taken over all laws on $\mathcal{V} \times \mathcal{U} \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y} \times \mathcal{Y}_1 \times \hat{\mathcal{Y}}_1$ of the form

$$
\begin{aligned}
& p_{UV X_1 X_2 Y Y_1 \hat{Y}_1}(u, v, x_1, x_2, y, y_1, \hat{y}_1) \\
& = p_{UV}(u, v) p_{X_1|U}(x_1|u) p_{X_2|V}(x_2|v) p(y, y_1|x_1, x_2) \\
& \quad \cdot p_{\hat{Y}_1|Y_1 X_2 U}(\hat{y}_1|y_1, x_2, u),
\end{aligned} \tag{15}
$$

and the cardinalities of the auxiliary random variables V and $\hat{Y}_1$ can be bounded as follows

$$
\begin{aligned}
\|\mathcal{V}\| &\leq \|\mathcal{U}\|\|\mathcal{X}_1\|\|\mathcal{X}_2\|\|\mathcal{Y}_1\| + 2, \\
\|\hat{\mathcal{Y}}_1\| &\leq \|\mathcal{V}\|\|\mathcal{U}\|\|\mathcal{X}_1\|\|\mathcal{X}_2\|\|\mathcal{Y}_1\| + 2.
\end{aligned} \tag{16}
$$

Analogously to [6, Theorem 1], the constraint (13) reflects the minimal compression ratio sustainable at the sender as it decodes the compressed data sent from the relay to the receiver. The constraint (14) reflects the minimal compression ratio sustainable at the receiver taking into account the assistance it gets from both the sender and the relay.

Remark 1: The achievable rate in Theorem 4 includes the previously known achievable rate in Theorem 1, i.e., $\bar{R}_1 \leq \bar{R}_2$. This can be seen by deriving an equivalent formulation of the rate $\bar{R}_1$. To this end, notice that for any law $p_{\tilde{V} X_1 X_2 Y Y_1 \hat{Y}_1}$ of the form (7) which satisfies (5) and (6), we have that

$$
\begin{aligned}
& I(X_1; Y\hat{Y}_1|\tilde{V}X_2) \\
& = I(X_1; Y|\tilde{V}X_2) + I(X_1; \hat{Y}_1|\tilde{V}X_2 Y) \\
& \overset{(i)}{=} I(X_1 X_2; Y) - I(\tilde{V}X_2; Y) + I(X_1; \hat{Y}_1|\tilde{V}X_2 Y) \\
& \overset{(ii)}{\leq} I(X_1 X_2; Y) - I(\hat{Y}_1; Y_1|\tilde{V}X_2 Y) + I(X_1; \hat{Y}_1|\tilde{V}X_2 Y) \\
& \overset{(iii)}{=} I(X_1 X_2; Y) + H(\hat{Y}_1|\tilde{V}X_1 X_2 Y Y_1) - H(\hat{Y}_1|\tilde{V}X_1 X_2 Y) \\
& = I(X_1 X_2; Y) - I(\hat{Y}_1; Y_1|\tilde{V}X_1 X_2 Y),
\end{aligned}
$$

where

(i) follows by the Markovian relation $\tilde{V} \oplus (X_1 X_2) \oplus (Y Y_1)$;
(ii) follows by Inequality (6); and
(iii) follows by the Markovian relation $X_1 \oplus (\tilde{V}X_2 Y_1) \oplus \hat{Y}_1$.

Consequently, $\bar{R}_1$ can be expressed as

$$
\bar{R}_1 = \sup_{p_{\tilde{V} X_1 X_2 Y Y_1 \hat{Y}_1}} \min\left\{ \begin{array}{c} I(X_1 X_2; Y) - I(\hat{Y}_1; Y_1|\tilde{V}X_1 X_2 Y), \\ I(X_1; Y\hat{Y}_1|\tilde{V}X_2) \end{array} \right\}, \tag{17}
$$

where the supremum is taken over all laws of the form (7) subject to the constraints (5) and (6). From this formulation it is evident that $\bar{R}_1 \leq \bar{R}_2$ since choosing $U = V$ in Theorem 4 the constraint (13) identifies with (5) and the constraint (14) identifies with (6), while (12) identifies with (17).

Remark 2: The achievable rate expression (12) is identical to the achievable rate expression in [13, Theorem 2] for the

relay channel without feedback. However, the rate expression in [13, Theorem 2] is subject to the constraint

$$I(\hat{Y}_1; Y_1 | U X_2) \leq I(X_2 \hat{Y}_1; Y | U V), \quad (18)$$

whereas our rate expression (12) is subject to the pair of constraints (13) and (14). The constraint (13) is equivalent to (as shown in section V.A)

$$I(\hat{Y}_1; Y_1 | U X_2) \leq I(X_2 \hat{Y}_1; Y | U V) + I(X_2 \hat{Y}_1; X_1 | U V Y), \quad (19)$$

so our first constraint (13) relaxes constraint (18). Consequently, for channels where the supremum in the rate expression (12) is attained for a law which fulfills (14) but violates (18)—i.e., for a law for which $I(X_2; Y | U V) < I(V X_2; Y)$— the achievable rate in Theorem 4 is strictly larger than the achievable rate in [13, Theorem 2]. Hence, for such channels, e.g., some Gaussian channels as is shown in Section IV, our feedback scheme improves upon the no-feedback scheme in [13].

Remark 3: The bound (16) on the alphabet cardinalities necessary for computing the rate $\bar{R}_2$ is partial in the sense that the alphabet cardinalities of V and $\hat{Y}_1$ are upper bounded for a given alphabet $\mathcal{U}$. So far, we've been unable to obtain an upper bound on $\|\mathcal{U}\|$ in terms of $\|\mathcal{X}_1\|$, $\|\mathcal{X}_2\|$, $\|\mathcal{Y}_1\|$, and $\|\mathcal{Y}\|$.

Notice that in Theorems 1 and 4 there exists an auxiliary random variable V which plays a role in building correlation between the transmitter and the relay. In Section V-B we show that one can naturally recover the auxiliary random variable also in the upper bound (see also Observation 1), albeit the resulting expression doesn't provide a tighter upper bound than the cut-set upper bound.

Observation 1: Let $(\mathcal{X}_1 \times \mathcal{X}_2, p(y, y_1 | x_1, x_2), \mathcal{Y} \times \mathcal{Y}_1)$ be a discrete memoryless relay channel with receiver-transmitter feedback. If

$$R > \sup_{p_{V X_1 X_2 Y Y_1}} \min \{I(X_1 X_2; Y), \; I(X_1; Y Y_1 | X_2 V)\}, \quad (20)$$

then there exists $\lambda > 0$ such that $P_e > \lambda$ for all n. The supremum in (20) is taken over all laws on $\mathcal{V} \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y} \times \mathcal{Y}_1$ of the form

$$p_{V X_1 X_2 Y Y_1}(v, x_1, x_2, y, y_1)$$
$$= p_{V X_1 X_2}(v, x_1, x_2) p(y, y_1 | x_1, x_2), \quad (21)$$

and the cardinality of V is bounded by $\|\mathcal{V}\| \leq \|\mathcal{X}_1\| \|\mathcal{X}_2\| \|\mathcal{Y}_1\| + 1$.

That the upper bound in Observation 1 equals the cut-set upper bound follows from the following two observations: the right hand side of (20) is maximized by choosing $V = \emptyset$ because conditioning reduces entropy and because $V \circ X_1 X_2 \circ Y Y_1$ forms a Markov chain; and for $V = \emptyset$ the rate constraint (20) identifies with the rate constraint (8), while (21) identifies with (9).

The advantage of the upper bound in Observation 1 is that it allows for a nice comparison with the lower bounds in Theorem 1, Theorem 4, and [13, Theorem 2]. For example, we see that the gap between the lower bound in Theorem 1 and the upper bound in Observation 1 is due to:

- $\hat{Y}_1 \neq Y_1$; and
- in (20) the supremum is taken over an arbitrary joint law $p_{V X_1 X_2}$ whereas in (17) the supremum is taken just over those laws that satisfy $X_1 \circ \tilde{V} \circ X_2$, and subject to the constraints (5) and (6).

Our last result in this section is on the semi-deterministic relay channel. It is based on Theorem 3 and on observing that the cut-set upper bound in Theorem 2 holds unchanged also for settings with feedback from the receiver to the sender.

Observation 2: Let $(\mathcal{X}_1 \times \mathcal{X}_2, p(y, y_1 | x_1, x_2), \mathcal{Y} \times \mathcal{Y}_1)$ be a discrete memoryless relay channel and let the channel output at the relay y_1 be a deterministic function of the two channel inputs x_1 and x_2. Then the capacity with receiver-transmitter feedback is given by (10) where the supremum is taken over all joint laws of the form (11).

This observation together with the observation in [2] imply that the semi-deterministic relay channel is "degraded" in the sense that neither relay-transmitter feedback nor receiver-transmitter feedback enlarges its no-feedback capacity.

IV. GAUSSIAN RELAY CHANNEL

In this section we focus on the Gaussian relay channel which is described as follows. Let $\{Z_{1,k}\}$ be a sequence of independent identically distributed (i.i.d.) Gaussian random variables of zero mean and variance N_1. Independently thereof let $\{Z_{2,k}\}$ be a sequence of i.i.d. Gaussian random variables of zero mean and variance N_2. The two sequences model the noise on the link from the transmitter to the relay and the noise on the link from the transmitter to the receiver. For given time-k channel inputs at the transmitter and at the relay, $x_{1,k}$ and $x_{2,k}$, the channel outputs at the relay and at the receiver are

$$Y_{1,k} = x_{1,k} + Z_{1,k}, \quad (22)$$
$$Y_k = x_{1,k} + d x_{2,k} + Z_{2,k}. \quad (23)$$

Here d is the gain coefficient of the relay-to-receiver link. The gain coefficients of the other links can be set to one without loss of generality.

As in the previous section we assume a causal noiseless receiver-transmitter feedback link.

We impose an average block power constraint on the input sequences $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$:

$$\frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[\left(X_{1,k}(W, Y^{k-1})\right)^2\right] \leq P_1 \quad (24)$$

and

$$\frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[\left(X_{2,k}(Y_1^{k-1})\right)^2\right] \leq P_2 \quad (25)$$

where $\mathbb{E}$ denotes the expectation operator.

Our main result here demonstrates that for the described Gaussian setting a combination of the Schalkwijk-Kailath signaling method [15] at the transmitter with a naive amplify-and-forward strategy at the relay sometimes (e.g., for $d = 0.5, P_1 = 1, P_2 = 1, N_1 = 1, N_2 = 1$, see Table I) outperforms the best known coding strategies. Let the function $\mathcal{C}(x)$ be defined as $\mathcal{C}(x) = 1/2 \ln(1 + x)$.

Theorem 5: Consider the Gaussian relay channel with receiver-transmitter feedback. Then the rate $\bar{R}_3^{(G)}$ defined by

$$\bar{R}_3^{(G)} = \max_{0 \leq \tilde{P}_2 \leq P_2} \mathcal{C}\left(\frac{P_1\left(1 + d\sqrt{\frac{\tilde{P}_2}{P_1+N_1}}\rho^*\right)^2}{d^2\frac{\tilde{P}_2}{P_1+N_1}N_1 + N_2}\right) \quad (26)$$

is achievable. Here the correlation coefficient ρ^* is given by the unique solution in $[0, 1]$ of the following quartic equation

$$\rho^2\left(P_1 + 2d\sqrt{P_1\tilde{P}_2}\sqrt{\frac{P_1}{P_1+N_1}}\rho + d^2\tilde{P}_2\frac{P_1}{P_1+N_1}\rho^2\right.$$
$$\left. + d^2\frac{N_1}{P_1+N_1}\tilde{P}_2 + N_2\right) = d^2\tilde{P}_2\frac{N_1}{P_1+N_1} + N_2. \quad (27)$$

In Theorem 5 the power $\tilde{P}_2$ denotes the transmit power the relay effectively uses to achieve the rate in (26), and hence it can be chosen arbitrarily between 0 and P_2. The best choice of $\tilde{P}_2$ is in general not the maximum available power P_2 and hence the relay in general might not use all the available transmit power. This is a consequence of the applied sub-optimal amplify-and-forward strategy where the relay not only amplifies the signal from the transmitter to the relay but also amplifies the noise corrupting this signal. Thus the amplification factor for the relay, and hence the used power, should be chosen as a trade-off between aiding the transmission from the transmitter to the receiver and introducing additional noise disturbing this transmission.

Note that for large power P_2 the proposed Schalkwijk-Kailath type scheme should be time-shared with a second coding scheme which can better exploit large available power at the relay. Thus, we propose to use for a fraction $0 \leq \gamma \leq 1$ of time the Schalkwijk-Kailath type scheme with power $\tilde{P}_2 \leq P_2$ at the relay, and use for the remaining fraction of time $1-\gamma$ the chosen block-Markov strategy with power $P_2 + \frac{\gamma}{1-\gamma}(P_2 - \tilde{P}_2)$ at the relay. Figure 2 illustrates the rates achieved by our Schalkwijk-Kailath type scheme, the rates achieved by our Block-Markov scheme, and the rates achieved by time-sharing these two schemes.

The Schalkwijk-Kailath type scheme can also be improved by allowing the relay to send arbitrary linear combinations of the past observed outputs in the spirit of El Gamal et al. [14]. An even more general approach would allow also the transmitter to apply an arbitrary linear strategy similar to the scheme proposed by Butman [22] for Gaussian single-user channels with feedback. In fact, sending maximally informative updates as proposed by Schalkwijk and Kailath can be strictly sub-optimal for multi-terminal settings, as was also pointed out by Ozarow [17] in a broadcast setting. However, both generalizations are difficult to analyze, since the problems of finding the optimal linear combinations are non-convex.

A second achievable rate for the Gaussian relay channel with receiver-transmitter feedback is obtained by evaluating the achievable rate of Theorem 4 for the Gaussian channel. However, we need to exercise some care in doing this, because the proof of Theorem 4 makes use of strong typicality in order to invoke Berger's Markov lemma [4, Lemma 14.8.1]. Since strong typicality does not apply to continuous alphabets, an

extension of this coding result to continuous random variables has to be proved using either the technique presented by Wyner in [9] or by using weak typicality and following the approach of Oohama in [10] (see also [13, Section II.A remark 1]). Consequently, a second achievable rate for the Gaussian relay channel with receiver-transmitter feedback may be obtained by evaluating the achievable rate of Theorem 4 for jointly Gaussian random variables, where the random variable $\hat{Y}_1$ is chosen similarly to [14] based on the Wyner-Ziv source coding (with decoder side information) strategy as $\hat{Y}_1 = \alpha(Y_1 + Z')$ with $Z' \sim \mathcal{N}(0, N')$. This choice is, of course, not necessarily the optimal choice.

Corollary 1: Consider the Gaussian relay channel with receiver-transmitter feedback. Then the rate $\bar{R}_2^{(G)}$ defined by

$$\bar{R}_2^{(G)} =$$
$$\sup_{\alpha_1, \alpha_2, \rho, N'} \min \left\{ \begin{array}{l} \mathcal{C}\left(\frac{P_1 + d^2P_2 + 2d\sqrt{P_1P_2}\sqrt{\bar{\alpha}_1\bar{\alpha}_2}\rho}{N_2}\right) - \mathcal{C}\left(\frac{N_1}{N'}\right), \\ \mathcal{C}\left(\frac{\alpha_1 P_1}{N_2} + \frac{\alpha_1 P_1}{N_1 + N'}\right) + \mathcal{C}\left(\frac{\bar{\alpha}_1 P_1(1-\rho^2)}{\alpha_1 P_1 + N_1}\right) \end{array} \right\} \quad (28)$$

is achievable subject to the five constraints

$$0 \leq \alpha_1 \leq 1,$$
$$0 \leq \alpha_2 \leq 1,$$
$$0 \leq \rho \leq 1,$$
$$\frac{N_1 N_2}{d^2 \alpha_2 P_2} \leq N',$$
$$\frac{N_1 + \alpha_1 P_1 \frac{N_2}{\alpha_1 P_1 + N_2}}{N'} \leq \frac{\bar{\alpha}_1 \rho^2 P_1 + P_2 d^2 + 2d\sqrt{\bar{\alpha}_1\bar{\alpha}_2 P_1 P_2}\rho}{\alpha_1 P_1 + \bar{\alpha}_1 P_1(1-\rho^2) + N_2}. \quad (29)$$

The achievable rate in Corollary 1 includes the achievable rates in [6] evaluated for the Gaussian relay channel when choosing $\rho = 1$. Table I shows achievable rates for the Gaussian relay channel for various coding strategies with and without receiver-transmitter feedback. The rates are computed for a setting where $P_1 = P_2 = 1$, $N_1 = 2$, and $N_2 = 0.5$. In Table I $R_{\text{EMZ}}^{(G)}$ denotes the El Gamal et al. [14] rate for $k = 2$, $R_{\text{CMG}}^{(G)}$ denotes the achievable rate derived in [13, Theorem 2], $\bar{R}_1^{(G)}$ denotes the achievable rate in [6, Theorem 1], and C_{UP} denotes the upper bound in Theorem 2. From Table I we see that for $d = 0.5, P_1 = 1, P_2 = 1, N_1 = 2$, and $N_2 = 0.5$ the coding scheme in Theorem 5 outperforms the other coding schemes (including that of Corollary 1) whereas for $d = 2.5$ the coding scheme in Corollary 1 outperforms the other coding schemes. For $d = 5$ and $d = 10$ the coding scheme in Corollary 1 and the coding scheme in [13, Theorem 2] perform best.

V. PROOFS

A. Proof of Theorem 4

We propose a coding scheme which is based on Block-Markov superposition encoding and which combines the ideas of nested binning as in [6], restricted decoding as in [21] together with the generalized coding scheme in [3].
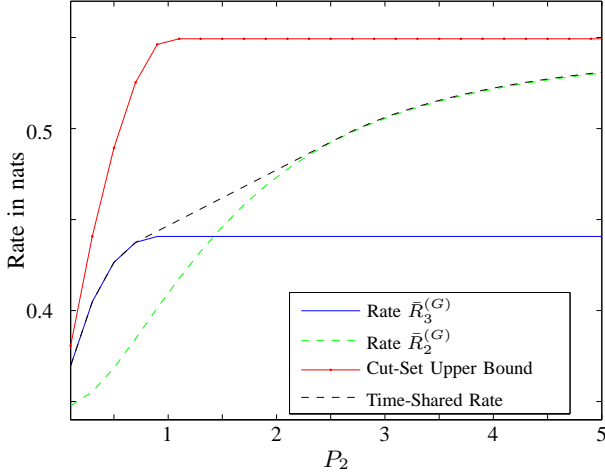
Fig. 2. Bounds on the capacity of the Gaussian relay channel with receiver-transmitter feedback, when $P_1 = 1$, $N_1 = 1$, $N_2 = 1$, and $d = 1$.

d	$R_{\text{EMZ}}^{(G)}$	$R_{\text{CMG}}^{(G)}$	$\bar{R}_1^{(G)}$	$\bar{R}_2^{(G)}$	$\bar{R}_3^{(G)}$	C_{UP}
0.5	0.2027	0.5589	0.5602	0.5602	0.5750	0.6264
2.5	0.2027	0.6097	0.6097	0.6099	0.5750	0.6264
5	0.2027	0.6209	0.6204	0.6209	0.5750	0.6264
10	0.2027	0.6244	0.6237	0.6244	0.5750	0.6264

TABLE I
BOUNDS ON THE CAPACITY OF THE GAUSSIAN RELAY CHANNEL WITH
RECEIVER-TRANSMITTER FEEDBACK. ($P_1 = P_2 = 1$, $N_1 = 2$, $N_2 = 0.5$.)

1) Coding Scheme: We consider $B + 1$ blocks, each of n symbols. We split the message W into a sequence of $B - 1$ sub-messages $W^{(b)}$, for $b = 1, \ldots, B-1$, where $W^{(b)}$ consists of the pair $(W_1^{(b)}, W_2^{(b)})$. Here the sequence $\{W_1^{(b)}\}$ is an i.i.d. sequence of uniform random variables over $\{1, \ldots, e^{nR_1}\}$ and independent thereof $\{W_2^{(b)}\}$ is an i.i.d. sequence of uniform random variables over $\{1, \ldots, e^{nR_2}\}$. As $B \to \infty$, for fixed n, the rate of the message W, $R = (R_1 + R_2)(B-1)/(B+1)$, is arbitrarily close to $R_1 + R_2$.

We assume a tuple of random variables $U \in \mathcal{U}, V \in \mathcal{V}, X_1 \in \mathcal{X}_1, X_2 \in \mathcal{X}_2, Y \in \mathcal{Y}, Y_1 \in \mathcal{Y}_1, \hat{Y}_1 \in \hat{Y}_1$ of joint law

$$
\begin{aligned}
&p_{UVX_1X_2YY_1\hat{Y}_1}(u, v, x_1, x_2, y, y_1, \hat{y}_1) \\
&= p_{UV}(u,v)p_{X_1|U}(x_1|u)p_{X_2|V}(x_2|v)p(y,y_1|x_1,x_2) \\
&\quad \cdot p_{\hat{Y}_1|Y_1X_2U}(\hat{y}_1|y_1, x_2, u). \quad (30)
\end{aligned}
$$

Random coding and partitioning: In each block $b, b = 1, 2, \ldots, B+1$, we shall use the following code.

- Generate $e^{n(\tilde{R}_0 + R_D + R_M)}$ sequences $\boldsymbol{v} = (v_1, \ldots, v_n)$, each with probability $\Pr(\boldsymbol{v}) = \prod_{k=1}^n p_V(v_k)$. Label them $\boldsymbol{v}(\omega_0)$ where $\omega_0 = (\omega_{0,1}, \omega_{0,2}, m)$, $\omega_{0,1} \in \{1, \ldots, e^{n\tilde{R}_0}\}$, $\omega_{0,2} \in \{1, \ldots, e^{nR_D}\}$, and $m \in \{1, \ldots, e^{nR_M}\}$.
- For each $\boldsymbol{v}(\omega_0)$ generate e^{nR_0} sequences $\boldsymbol{x}_2 = (x_{2,1}, x_{2,2}, \ldots, x_{2,n})$, each with probability $\Pr(\boldsymbol{x}_2|\boldsymbol{v}(\omega_0)) = \prod_{k=1}^n p_{X_2|V}(x_{2,k}|v_k(\omega_0))$. Label them $\boldsymbol{x}_2(s, \omega_0), s \in \{1, \ldots, e^{nR_0}\}$.
- For each $\boldsymbol{v}(\omega_0)$ generate e^{nR_1} sequences $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$, each with probability $\Pr(\boldsymbol{u}|\boldsymbol{v}(\omega_0)) = \prod_{k=1}^n p_{U|V}(u_k|v_k(\omega_0))$. Label them $\boldsymbol{u}(w_1, \omega_0), w_1 \in \{1, \ldots, e^{nR_1}\}$.
- For each $\boldsymbol{u}(w_1, \omega_0)$ generate e^{nR_2} sequences $\boldsymbol{x}_1 = (x_{1,1}, x_{1,2}, \ldots, x_{1,n})$, each with probability $\Pr(\boldsymbol{x}_1|\boldsymbol{u}(w_1, \omega_0)) = \prod_{k=1}^n p_{X_1|U}(x_{1,k}|u_k(w_1, \omega_0))$. Label them $\boldsymbol{x}_1(w_2, w_1, \omega_0) w_2 \in \{1, \ldots, e^{nR_2}\}$.
- For each $\boldsymbol{x}_2(s, \omega_0)$ and $\boldsymbol{u}(w_1, \omega_0)$ generate $e^{n\hat{R}}$ sequences $\hat{\boldsymbol{y}}_1$ each with probability $\Pr(\hat{\boldsymbol{y}}_1|\boldsymbol{x}_2(s, \omega_0), \boldsymbol{u}(w_1, \omega_0)) = \prod_{k=1}^n p_{\hat{Y}_1|X_2,U}(\hat{y}_{1,k}|x_{2,k}(s, \omega_0), u_k(w_1, \omega_0))$, where for $x_2 \in \mathcal{X}_2, \hat{y}_1 \in \hat{\mathcal{Y}}_1$, and $u \in \mathcal{U}$ we define

$$
\begin{aligned}
&p_{\hat{Y}_1|X_2U}(\hat{y}_1|x_2, u) \\
&= \frac{\sum_{v,x_1,y,y_1} p_{UVX_1X_2YY_1\hat{Y}_1}(u, v, x_1, x_2, y, y_1, \hat{y}_1)}{\sum_{v,x_1,y,y_1,\hat{y}_1} p_{UVX_1X_2YY_1\hat{Y}_1}(u, v, x_1, x_2, y, y_1, \hat{y}_1)}.
\end{aligned}
$$
$$(31)$$

Label them $\hat{\boldsymbol{y}}_1(z, w_1, s, \omega_0)$, where $s \in \{1, \ldots, e^{nR_0}\}$, $\omega_0 \in \{1, \ldots, e^{(\tilde{R}_0 + R_D + R_M)}\}$, $w_1 \in \{1, \ldots, e^{nR_1}\}$, $z \in \{1, \ldots, e^{n\hat{R}}\}$.

- Partition 1: Randomly partition the set $\{1, \ldots, e^{n\hat{R}}\}$ into e^{nR_0} cells. Label the cells $s \in \{1, \ldots, e^{nR_0}\}$ and let $s(z) = c$ if z belongs to cell c.
- Partition 2: Randomly partition each cell of size $e^{n(\hat{R} - R_0)}$ in Partition 1 into e^{nR_D} subcells. Label the subcells in each such subpartition $\omega_{0,2} \in \{1, \ldots, e^{nR_D}\}$ and let $\omega_{0,2}(z) = c$ if z belongs to subcell c in some subpartition.
- Partition 3: Create a partition over $\{1, \ldots, e^{nR_0}\}$ with $e^{n\tilde{R}_0}$ disjoint cells each containing $e^{n(R_0 - \tilde{R}_0)}$ elements. Label the cells $\omega_{0,1} \in \{1, \ldots, e^{n\tilde{R}_0}\}$, and let $\omega_{0,1}(s) = c$ if s belongs to cell c. This partition referred as a *deterministic* partition will serve later on for the purpose of *restricted decoding* [21].
- Partition 4: Randomly partition the set $\{1, \ldots, e^{nR_1}\}$ into e^{nR_M} cells. Label the cells $m \in \{1, \ldots, e^{nR_M}\}$ and let $m(w_1) = c$ if w_1 belongs to cell c.

Encoding : We denote the realizations of the sequences $\{W^{(b)}\}$, $\{W_1^{(b)}\}$, and $\{W_2^{(b)}\}$ by $\{w^{(b)}\}$, $\{w_1^{(b)}\}$, and $\{w_2^{(b)}\}$. The code builds upon a three-level Block-Markov structure. This implies that Message $w^{(b)} = (w_1^{(b)}, w_2^{(b)})$ is encoded over the three successive blocks $b, (b+1)$, and $(b+2)$, for $= 1, \ldots, B-1$. Furthermore, the code builds upon the following properties: $s^{(b)} = s(z^{(b-1)})$, $\omega_0^{(b)} = (\omega_{0,1}^{(b)}, \omega_{0,2}^{(b)}, m^{(b)})$ where $\omega_{0,1}^{(b)} = \omega_{0,1}(s^{(b-1)})$, $\omega_{0,2}^{(b)} = \omega_{0,2}(z^{(b-2)})$, $m^{(b)} = m(w_1^{(b-1)})$, for $b = 1, \ldots, B+1$. Here the sequence $z^{(1)}, \ldots, z^{(B)}$ will be defined when describing the decoding at the relay and $z^{(-1)} = z^{(0)} = w_1^{(0)} = 1$.

We assume that at the end of block $(b-1)$, $b = 1, \ldots, B+1$

- The sender knows $(w^{(1)}, w^{(2)}, \ldots, w^{(B)})$, and it has available $(\hat{\omega}_{0,E}^{(1)}, \hat{\omega}_{0,E}^{(2)}, \ldots, \hat{\omega}_{0,E}^{(b)})$,

$\left(\hat{s}_E^{(1)}, \hat{s}_E^{(2)}, \ldots, \hat{s}_E^{(b-1)}\right)$, and $\left(\hat{z}_E^{(1)}, \hat{z}_E^{(2)}, \ldots, \hat{z}_E^{(b-2)}\right)$.

- The relay knows $\left(s^{(1)}, s^{(2)}, \ldots, s^{(b)}\right)$ and $\left(z^{(1)}, z^{(2)}, \ldots, z^{(b-1)}\right)$, and it has available $\left(\hat{\omega}_{0,R}^{(1)}, \hat{\omega}_{0,R}^{(2)}, \ldots, \hat{\omega}_{0,R}^{(b)}\right)$ and $\left(\hat{w}_{1,R}^{(1)}, \hat{w}_{1,R}^{(2)}, \ldots, \hat{w}_{1,R}^{(b-1)}\right)$.

Then in block b the relay transmits the codeword

$$\boldsymbol{x}_2(s^{(b)}, \hat{\omega}_{0,R}^{(b)}) =$$
$$\boldsymbol{x}_2\left(s(z^{(b-1)}), \left(\omega_{0,1}(s^{(b-1)}), \omega_{0,2}(z^{(b-2)}), m(\hat{w}_{1,R}^{(b-1)})\right)\right),$$

and the sender transmits the codeword

$$\boldsymbol{x}_1\left(w^{(b)}, \hat{\omega}_{0E}^{(b)}\right) =$$
$$\boldsymbol{x}_1\left(w^{(b)}, \left(\omega_{0,1}(\hat{s}_E^{(b-1)}), \omega_{0,2}(\hat{z}_E^{(b-2)}), m(w_1^{(b-1)})\right)\right).$$

Decoding at the transmitter and at the relay: After the reception of the block-b channel outputs and feedback outputs the transmitter and the relay perform the following decoding steps which enable them for three levels of cooperation.

1) In order to obtain the first level of *cooperation*, after each block b, $b = 1, 2, \ldots, B-1$ the relay upon receiving $\boldsymbol{y}_1^{(b)}$ chooses $\hat{w}_{1,R}^{(b)}$ such that

$$\left(\boldsymbol{v}(\hat{\omega}_{0R}^{(b)}), \boldsymbol{u}(\hat{w}_{1,R}^{(b)}, \hat{\omega}_{0R}^{(b)}), \boldsymbol{x}_2(s^{(b)}, \hat{\omega}_{0R}^{(b)}), \boldsymbol{y}_1^{(b)}\right)$$
$$\in \mathcal{A}_\epsilon(V, U, X_2, Y_1),$$

where $\mathcal{A}_\epsilon(\cdot)$ denotes the strongly typical set (see Appendix A). This determines $m^{(b+1)} = m(\hat{w}_{1,R}^{(b)})$, that the relay transmits in block $(b+1)$.

2) The relay upon receiving $\boldsymbol{y}_1^{(b)}$ decides that $z^{(b)}$ is "received" if

$$\left(\hat{\boldsymbol{y}}_1(z^{(b)}, \hat{w}_{1,R}^{(b)}, s^{(b)}, \hat{\omega}_{0R}^{(b)}), \boldsymbol{y}_1^{(b)}, \boldsymbol{v}(\hat{\omega}_{0R}^{(b)}),\right.$$
$$\left.\boldsymbol{u}(\hat{w}_{1,R}^{(b)}, \hat{\omega}_{0R}^{(b)}), \boldsymbol{x}_2(s^{(b)}, \hat{\omega}_{0R}^{(b)})\right)$$
$$\in \mathcal{A}_\epsilon(\hat{Y}_1, Y_1, V, X_2).$$

3) In order to obtain the second level of *cooperation*, after block b, $b = 1, \ldots, B-1$, the sender chooses $\hat{s}_E^{(b)}$ such that

$$\left(\boldsymbol{v}(\hat{\omega}_{0E}^{(b)}), \boldsymbol{u}(w_1^{(b)}, \hat{\omega}_{0E}^{(b)}), \boldsymbol{x}_1(w^{(b)}, \hat{\omega}_{0E}^{(b)}),\right.$$
$$\left.\boldsymbol{x}_2(\hat{s}_E^{(b)}, \hat{\omega}_{0E}^{(b)}), \boldsymbol{y}^{(b)}\right)$$
$$\in \mathcal{A}_\epsilon(V, U, X_1, X_2, Y).$$

This determines $\omega_{0,1}\left(\hat{s}_E^{(b)}\right)$, that the sender transmits in block $b+1$.

4) In order to obtain the third level of *cooperation*, after each block b, $b = 2, \ldots, B$ the sender forms the set

$\mathcal{L}_E\left(\boldsymbol{y}^{(b-1)}\right)$ of $z_E^{(b-1)}$ such that

$$\mathcal{L}_E\left(\boldsymbol{y}^{(b-1)}\right) = \left\{ z_E^{(b-1)} : \right.$$
$$\left(\boldsymbol{v}(\hat{\omega}_{0E}^{(b-1)}), \boldsymbol{u}(w_1^{(b-1)}, \hat{\omega}_{0E}^{(b-1)}), \boldsymbol{x}_1(w^{(b-1)}, \hat{\omega}_{0E}^{(b-1)}),\right.$$
$$\boldsymbol{x}_2(\hat{s}_E^{(b-1)}, \hat{\omega}_{0E}^{(b-1)}), \boldsymbol{y}^{(b-1)},$$
$$\left.\hat{\boldsymbol{y}}_1(z_E^{(b-1)}, w_1^{(b-1)}, \hat{s}_E^{(b-1)}, \hat{\omega}_{0E}^{(b-1)})\right)$$
$$\left. \in \mathcal{A}_\epsilon\left(V, U, X_1, X_2, Y, \hat{Y}_1\right) \right\}.$$

The sender then declares that $\hat{z}_E^{(b-1)}$ was sent in block $(b-1)$ if and only if there is a unique

$$\hat{z}_E^{(b-1)} \in \mathcal{L}_E\left(\boldsymbol{y}^{(b-1)}\right),$$

such that $s(\hat{z}_E^{(b-1)}) = s_E^{(b)}$.

After this decoding step, the sender and the relay cooperate in the sense that in block $b+1$ the relay transmits $\omega_{0,2}^{(b+1)} = \omega_{0,2}\left(z^{(b-1)}\right)$, while the sender transmits $\omega_{0,2}^{(b+1)} = \omega_{0,2}\left(\hat{z}_E^{(b-1)}\right)$.

Decoding at the receiver:

For the decoding procedure at the receiver starting after the reception of block b we assume that, upon the decoding of block $b-1$, the receiver has available $\left(\hat{w}_{1,D}^{(1)}, \hat{w}_{1,D}^{(2)}, \ldots, \hat{w}_{1,D}^{(b-2)}\right)$, $\left(\hat{w}_{2,D}^{(1)}, \hat{w}_{2,D}^{(2)}, \ldots, \hat{w}_{2,D}^{(b-3)}\right)$, $\left(\hat{\omega}_{0,D}^{(1)}, \hat{\omega}_{0,D}^{(2)}, \ldots, \hat{\omega}_{0,D}^{(b-1)}\right)$, $\left(\hat{s}_D^{(1)}, \hat{s}_D^{(2)}, \ldots, \hat{s}_D^{(b-2)}\right)$, and $\left(\hat{z}_D^{(1)}, \hat{z}_D^{(2)}, \ldots, \hat{z}_D^{(b-3)}\right)$.

Then, after block b the receiver decodes the messages $w_1^{(b-1)}$ and $w_2^{(b-2)}$ as follows.

1) The receiver looks for $\hat{\omega}_{0,D}^{(b)}$ such that

$$\left(\boldsymbol{v}(\hat{\omega}_{0,D}^{(b)}), \boldsymbol{y}^{(b)}\right) \in \mathcal{A}_\epsilon(V, Y), \qquad (32)$$

where $\hat{\omega}_{0,D}^{(b)} = (\hat{\omega}_{0,1,D}^{(b)}, \hat{\omega}_{0,2,D}^{(b)}, \hat{m}_D^{(b)})$.

2) The receiver then considers block $(b-1)$ and chooses $\hat{s}_D^{(b-1)}$ such that

$$\left(\boldsymbol{v}(\hat{\omega}_{0,D}^{(b-1)}), \boldsymbol{x}_2(\hat{s}_D^{(b-1)}, \hat{\omega}_{0,D}^{(b-1)}), \boldsymbol{y}^{(b-1)}\right)$$
$$\in \mathcal{A}_\epsilon(V, X_2, Y),$$

and such that $\omega_{0,1}(\hat{s}_D^{(b-1)}) = \hat{\omega}_{0,1,D}^{(b)}$.

This step is similar to the restricted decoding principle, that has been proposed in [21], for the multiple-access channel with partial feedback.

3) The receiver then considers block $(b-1)$ and forms the set $\mathcal{L}_D^{(1)}\left(\boldsymbol{y}^{(b-1)}\right)$ of $w_{1,D}^{(b-1)}$ such that

$$\mathcal{L}_D^{(1)}\left(\boldsymbol{y}^{(b-1)}\right) = \left\{ w_{1,D}^{(b-1)} : \right.$$
$$\left(\boldsymbol{v}(\hat{\omega}_{0,D}^{(b-1)}), \boldsymbol{x}_2(\hat{s}_D^{(b-1)}, \hat{\omega}_{0,D}^{(b-1)}),\right.$$
$$\left.\boldsymbol{u}(\hat{w}_{1,D}^{(b-1)}, \hat{\omega}_{0,D}^{(b-1)}), \boldsymbol{y}^{(b-1)}\right)$$
$$\left. \in \mathcal{A}_\epsilon\left(V, U, X_2, Y\right) \right\}.$$

The receiver then declares that $\hat{w}_{1,D}^{(b-1)}$ was "received" by the relay in block $(b-1)$ if and only if there is a unique

$$\hat{w}_{1,D}^{(b-1)} \in \mathcal{L}_D^{(1)}\left(\boldsymbol{y}^{(b-1)}\right),$$

such that $m(\hat{w}_{1,D}^{(b-1)}) = \hat{m}_D^{(b)}$.

4) The receiver then considers block $(b-2)$ and forms the set $\mathcal{L}_D^{(2)}\left(\boldsymbol{y}^{(b-2)}\right)$ of $z_D^{(b-2)}$ such that

$$\mathcal{L}_D^{(2)}\left(\boldsymbol{y}^{(b-2)}\right) = \Big\{ z_D^{(b-2)} :$$
$$\Big(\boldsymbol{v}(\hat{\omega}_{0,D}^{(b-2)}), \boldsymbol{x}_2(\hat{s}_D^{(b-2)}, \hat{\omega}_{0,D}^{(b-2)}),$$
$$\boldsymbol{u}(\hat{w}_{1,D}^{(b-2)}, \hat{\omega}_{0,D}^{(b-2)}), \boldsymbol{y}^{(b-2)},$$
$$\hat{\boldsymbol{y}}_1(z_D^{(b-2)}, \hat{w}_{1,D}^{(b-2)}, \hat{s}_D^{(b-2)}, \hat{\omega}_{0,D}^{(b-2)})\Big)$$
$$\in \mathcal{A}_\epsilon\left(V, U, X_2, Y, \hat{Y}_1\right)\Big\}.$$

The receiver then declares that $\hat{z}_D^{(b-2)}$ was sent in block $(b-2)$ if and only if there is a unique

$$\hat{z}_D^{(b-2)} \in \mathcal{L}_D^{(2)}\left(\boldsymbol{y}^{(b-2)}\right),$$

such that $s(\hat{z}_D^{(b-2)}) = \hat{s}_D^{(b-1)}$ and $\omega_{0,2}(\hat{z}_D^{(b-2)}) = \hat{\omega}_{0,2}^{(b)}$.

5) The receiver declares that $\hat{w}_{2,D}^{(b-2)}$ was sent in block $(b-2)$ if

$$\Big(\boldsymbol{v}(\hat{\omega}_{0,D}^{(b-2)}), \boldsymbol{x}_2(\hat{s}_D^{(b-2)}, \hat{\omega}_{0,D}^{(b-2)}),$$
$$\boldsymbol{u}(\hat{w}_{1,D}^{(b-2)}, \hat{\omega}_{0,D}^{(b-2)}), \boldsymbol{x}_1(\hat{w}_{2,D}^{(b-2)}, \hat{w}_{1,D}^{(b-2)}, \hat{\omega}_{0,D}^{(b-2)}),$$
$$\boldsymbol{y}^{(b-2)}, \hat{\boldsymbol{y}}_1(\hat{z}_D^{(b-2)}, \hat{w}_{1,D}^{(b-2)}, \hat{s}_D^{(b-2)}, \hat{\omega}_{0,D}^{(b-2)})\Big)$$
$$\in \mathcal{A}_\epsilon(V, U, X_1, X_2, Y, \hat{Y}_1).$$

2) Bounding the Probability of Error: Genie-aided arguments as in [23] and [24] can be used to show that the probability that the receiver makes a decoding error after block b in the above scheme is upper bounded by the probability that at least one of the following events $E_0^{(b)} - E_9^{(b)}$ happens.

Error events at the receiver:

• $E_0^{(b)}$:

$$\Big(\boldsymbol{v}(\omega_0^{(b)}), \boldsymbol{u}(w_1^{(b)}, \omega_0^{(b)}), \boldsymbol{x}_1(w^{(b)}, \omega_0^{(b)}),$$
$$\boldsymbol{x}_2(s^{(b)}, \omega_0^{(b)}), \boldsymbol{y}^{(b)}, \boldsymbol{y}_1^{(b)},$$
$$\hat{\boldsymbol{y}}_1(z^{(b)}, w_1^{(b)}, s^{(b)}, \omega_0^{(b)})\Big)$$
$$\notin \mathcal{A}_\epsilon(V, U, X_1, X_2, Y, Y_1, \hat{Y}_1).$$

• $E_1^{(b)}$: There exists $\tilde{\omega}_{0,D} \neq \omega_0^{(b)}$ such that

$$\Big(\boldsymbol{v}\left(\tilde{\omega}_{0,D}\right), \boldsymbol{y}^{(b)}\Big) \in \mathcal{A}_\epsilon(V, Y).$$

• $E_2^{(b)}$: There exists $\tilde{s}_D \neq s^{(b-1)}$ such that $\omega_{0,1}(\tilde{s}_D) = \omega_{0,1}^b$ and

$$\Big(\boldsymbol{v}(\omega_0^{(b-1)}), \boldsymbol{x}_2(\tilde{s}_D, \omega_0^{(b-1)}), \boldsymbol{y}^{(b-1)}\Big) \in \mathcal{A}_\epsilon(V, X_2, Y).$$

• $E_3^{(b)}$: There exists $\tilde{w}_{1,D} \neq w_1^{(b-1)}$ such that $m(\tilde{w}_{1,D}) = m^{(b)}$ and

$$\tilde{w}_{1,D} \in \mathcal{L}_D^{(1)}(\boldsymbol{y}^{(b-1)}),$$

• $E_4^{(b)}$: There exists $\tilde{z}_D \neq z^{(b-2)}$ such that $s(\tilde{z}_D) = s^{(b-1)}$, $\omega_{0,2}(\tilde{z}_D) = \omega_{0,2}^{(b)}$, and

$$\tilde{z}_D \in \mathcal{L}_D^{(2)}(\boldsymbol{y}^{(b-2)}).$$

• $E_5^{(b)}$: There exists $\tilde{w}_{2,D} \neq w_2^{(b-2)}$ such that

$$\Big(\boldsymbol{v}(\omega_0^{(b-2)}), \boldsymbol{u}(w_1^{(b-2)}, \omega_0^{(b-2)}),$$
$$\boldsymbol{x}_1(\tilde{w}_{2,D}, w_1^{(b-2)}, \omega_0^{(b-2)}), \boldsymbol{x}_2(s^{(b-2)}, \omega_0^{(b-2)}),$$
$$\boldsymbol{y}^{(b-2)}, \hat{\boldsymbol{y}}_1(z^{(b-2)}, w_1^{(b-2)}, s^{(b-2)}, \omega_0^{(b-2)})\Big)$$
$$\in \mathcal{A}_\epsilon(V, U, X_1, X_2, Y, \hat{Y}_1).$$

Error events at the relay:

• $E_6^{(b)}$: There exists $\tilde{w}_{1,R} \neq w_1^{(b)}$ such that

$$\Big(\boldsymbol{v}(\omega_0^{(b)}), \boldsymbol{u}(\tilde{w}_{1,R}, \omega_0^{(b)}), \boldsymbol{x}_2(s^{(b)}, \omega_0^{(b)}), \boldsymbol{y}_1^{(b)}\Big)$$
$$\in \mathcal{A}_\epsilon(V, U, X_2, Y_1).$$

• $E_7^{(b)}$: There exists no $z^{(b)}$ such that

$$\Big(\hat{\boldsymbol{y}}_1(z^{(b)}, w_1^{(b)}, s^{(b)}, \omega_0^{(b)}), \boldsymbol{y}_1^{(b)}, \boldsymbol{v}(\omega_0^{(b)}),$$
$$\boldsymbol{u}(w_1^{(b)}, \omega_0^{(b)}), \boldsymbol{x}_2(s^{(b)}, \omega_0^{(b)})\Big)$$
$$\in \mathcal{A}_\epsilon(\hat{Y}_1, Y_1, V, U, X_2).$$

Error events at the sender:

• $E_8^{(b)}$: There exists $\tilde{s}_E \neq s^{(b)}$ such that

$$\Big(\boldsymbol{v}(\omega_0^{(b)}), \boldsymbol{u}(w_1^{(b)}, \omega_0^{(b)}), \boldsymbol{x}_1(w^{(b)}, \omega_0^{(b)}),$$
$$\boldsymbol{x}_2(\tilde{s}_E, \omega_0^{(b)}), \boldsymbol{y}^{(b)}\Big)$$
$$\in \mathcal{A}_\epsilon(V, U, X_1, X_2, Y).$$

• $E_9^{(b)}$: There exists $\tilde{z}_E \neq z^{(b-1)}$ such that $s(\tilde{z}_E) = s^{(b)}$.

$$\tilde{z}_E \in \mathcal{L}_E(\boldsymbol{y}^{(b)}).$$

In the following we want to bound the probability that at least one of events $E_0^b, \ldots, E_9^b$ happens. To this end, we bound the probability of error $\bar{P}_e$ averaged over all codebooks and all random partitions. We define the event

$$F^{(b)} \triangleq \bigcup_{j=0}^{9} E_j^{(b)}, \qquad b = 1, \ldots, B+1,$$

which includes the event of a decoding error after block b and which is defined over all choices of the codebooks. Then, we can upper bound the averaged probability of error by

$$\bar{P}_e \leq \sum_{b=1}^{B+1} \Pr\left[F^{(b)} | F^{(1\ldots b-1)^c}\right], \tag{33}$$

where $F^{(1\ldots b-1)^c}$ denotes the complement of the event $F^{(1)} \cup \ldots \cup F^{(b-1)}$. Furthermore, we can upper bound each of the summands as

$$\Pr\left(F^{(b)}|F^{(1\ldots b-1)^c}\right)$$

$$= \Pr\left(\bigcup_{j=0}^{9} E_j^{(b)}|F^{(1\ldots b-1)^c}\right)$$

$$\leq \sum_{k=0}^{5} \Pr\left(E_k^{(b)}|\bigcup_{m=0}^{k-1} E_m^{(b)}, F^{(1\ldots b-1)^c}\right)$$

$$+ \Pr\left(E_6^{(b)}|E_0^{(b)^c}, F^{(1\ldots b-1)^c}\right)$$

$$+ \Pr\left(E_7^{(b)}|E_6^{(b)^c}, E_0^{(b)^c}, F^{(1\ldots b-1)^c}\right)$$

$$+ \Pr\left(E_8^{(b)}|E_0^{(b)^c}, F^{(1\ldots b-1)^c}\right)$$

$$+ \Pr\left(E_9^{(b)}|E_8^{(b)^c}, E_0^{(b)^c}, F^{(1\ldots b-1)^c}\right).$$

In the following we separately examine each of the above summands.

The event $E_0^{(b)^c}$ is independent of the event $F^{(1\ldots b-1)^c}$ and by Lemma 3 (Appendix A) can be made arbitrarily small for sufficiently large n.

Also, by Lemma 4:

- If

$$\tilde{R}_0 + R_D + R_M < I(V; Y), \qquad (34)$$

  then $\Pr\left(E_1^{(b)}|E_0^{(b)^c}, F^{(1\ldots b-1)^c}\right)$ can be made arbitrarily small, provided that n is sufficiently large;
- If

$$R_0 < I(X_2; Y|V) + \tilde{R}_0, \qquad (35)$$

  then $\Pr\left(E_2^{(b)}|E_0^{(b)^c}, E_1^{(b)^c}, F^{(1\ldots b-1)^c}\right)$ can be made arbitrarily small, provided that n is sufficiently large;
- If

$$R_2 < I(X_1; Y, \hat{Y}_1|UX_2), \qquad (36)$$

  then $\Pr\left(E_5^{(b)}|E_0^{(b)^c}, \ldots, E_4^{(b)^c}, F^{(1\ldots b-1)^c}\right)$ can be made arbitrarily small, provided that n is sufficiently large;
- If

$$R_1 < I(U; Y_1|VX_2), \qquad (37)$$

  then $\Pr\left(E_6^{(b)}|E_0^{(b)^c}, F^{(1\ldots b-1)^c}\right)$ can be made arbitrarily small, provided that n is sufficiently large;
- If

$$R_0 < I(X_2; Y|VUX_1), \qquad (38)$$

  then $\Pr\left(E_8^{(b)}|E_0^{(b)^c}, F^{(1\ldots b-1)^c}\right)$ can be made arbitrarily small, provided that n is sufficiently large.

Furthermore, by [11], [7], [12] or [4, Chapter 13], if

$$\hat{R} > I(\hat{Y}_1; Y_1|UX_2), \qquad (39)$$

then $\Pr\left(E_7^{(b)}|E_6^{(b)^c}, E_0^{(b)^c}, F^{(1\ldots b-1)^c}\right)$ can be made arbitrarily small, provided that n is sufficiently large.

The following lemma considers the transmitter error event $E_9^{(b)}$ and shows that its probability can be made arbitrary small if condition (40) is satisfied.

Lemma 1: If

$$\hat{R} < I(\hat{Y}_1; X_1, Y|UX_2) + R_0 - \epsilon_1, \qquad (40)$$

then for sufficiently large n

$$\Pr\left(E_9^{(b)}|E_8^{(b)^c}, E_0^{(b)^c}|F^{(1\ldots b-1)^c}\right) \leq \epsilon/(10(B+1)). \qquad (41)$$

Proof: See Appendix B. ∎

Using a similar argument as that of Lemma 1, finally we obtain:

- If

$$R_1 < I(U; Y|VX_2) + R_M, \qquad (42)$$

  then $\Pr\left(E_3^{(b)}|E_0^{(b)^c}, \ldots, E_2^{(b)^c}, F^{(1\ldots b-1)^c}\right)$ can be made arbitrarily small, provided that n is sufficiently large;
- If

$$\hat{R} < I(\hat{Y}_1; Y|UX_2) + R_0 + R_D, \qquad (43)$$

  then $\Pr\left(E_4^{(b)}|E_0^{(b)^c}, \ldots, E_3^{(b)^c}, F^{(1\ldots b-1)^c}\right)$ can be made arbitrarily small, provided that n is sufficiently large.

Thus, we can make each of the above probabilities smaller than $\epsilon/(10(B+1))$ if (34)—(40), (42), and (43) are satisfied. Then, by (33) the averaged probability of error $\bar{P}_e$ can be upper bounded by ϵ. And we conclude that there must exist at least one possible code of probability of error $P_e < \epsilon$.

3) Further Analysis: To summarize our results so far, with the presented scheme the message W can be transmitted with arbitrary small probability of error if its rate $R = R_1 + R_2$ satisfies (34)—(40), (42), and (43) for some non-negative rates $R_0, \tilde{R}_0, R_D, R_M, \hat{R}$ and for the joint law on $(UVX_1X_2YY_1\hat{Y}_1)$ in (30).

We eliminate the rates $R_0, \tilde{R}_0, R_D$, and $\hat{R}$ from the rate expressions (34)—(40), (42), and (43) by means of the Fourier-Motzkin elimination. To this end, we first eliminate $\tilde{R}_0$ by combining (34) and (35) to obtain

$$R_0 + R_D + R_M < I(VX_2; Y). \qquad (44)$$

Next, we eliminate $\hat{R}$, i.e., we combine (39), (40) and (43), which yields

$$I(\hat{Y}_1; Y_1|UX_2) < I(\hat{Y}_1; X_1Y|UX_2) + R_0 \qquad (45)$$

$$I(\hat{Y}_1; Y_1|UX_2) < I(\hat{Y}_1; Y|UX_2) + R_0 + R_D. \qquad (46)$$

Then, we eliminate R_D by combining (46) together with (44), followed by eliminating R_0 by combining (45) with (38). This yields the two constraints

$$I(\hat{Y}_1; Y_1|UX_2) < I(\hat{Y}_1; X_1Y|UX_2) + I(X_2; Y|VUX_1),$$

$$I(\hat{Y}_1; Y_1|UX_2) < I(\hat{Y}_1; Y|UX_2) + I(VX_2; Y) - R_M,$$

$$\qquad (47)$$

where remember that $R_M \geq 0$. The combination of (36) once with (37) and then with (42) yields

$$R_1 + R_2 < I(X_1; Y\hat{Y}_1|UX_2) + I(U; Y_1|VX_2)$$
$$R_1 + R_2 < I(X_1; Y\hat{Y}_1|UX_2) + I(U; Y|VX_2) + R_M. \quad (48)$$

Finally, the substitution of the upper bound on R_M, which results from the second constraint in (47), into the second upper bound on the sum-rate in (48) yields

$$
\begin{aligned}
R_1 + R_2 &< I(X_1; Y\hat{Y}_1|UX_2) + I(U; Y|VX_2) + I(VX_2; Y), \\
&\quad + I\left(\hat{Y}_1; Y|UX_2\right) - I(\hat{Y}_1; Y_1|UX_2) \\
&= I(X_1; Y\hat{Y}_1|UX_2) + I(VUX_2; Y) \\
&\quad + H(\hat{Y}_1|UX_2Y_1) - H(\hat{Y}_1|UX_2Y) \\
&\stackrel{(a)}{=} I(X_1; Y\hat{Y}_1|UX_2) + I(VUX_2; Y) \\
&\quad - I(\hat{Y}_1; Y_1|UX_2Y) \\
&= I(X_1; Y|UX_2) + I(X_1; \hat{Y}_1|UX_2Y) \\
&\quad + I(VUX_2; Y) - I(\hat{Y}_1; Y_1|UX_2Y) \\
&= I(X_1; Y|UX_2) + I(VUX_2; Y) \\
&\quad + H(\hat{Y}_1|UX_2YY_1) - H(\hat{Y}_1|UX_1X_2Y) \\
&\stackrel{(b)}{=} I(X_1; Y|UX_2) + I(UX_2; Y) \\
&\quad + I(V; Y|UX_2) - I(\hat{Y}_1; Y_1|UX_1X_2Y) \\
&\stackrel{(c)}{=} I(X_1X_2; Y) - I(\hat{Y}_1; Y_1|UX_1X_2Y). \quad (49)
\end{aligned}
$$

Here,

(a) and (b) follow by the Markov relation $(X_1, Y) \ominus (U, X_2, Y_1) \ominus \hat{Y}_1$ and

(c) follows from the Markov relations $V \ominus (X_2, U) \ominus Y$ and $U \ominus (X_1, X_2) \ominus Y$.

Note however that the choice for R_M is only valid if $R_M > 0$ which imposes the following constraint on the input distribution

$$I(\hat{Y}_1; Y_1|UX_2) < I(\hat{Y}_1; Y|UX_2) + I(VX_2; Y). \quad (50)$$

We consider now the first constraint in (47) which reads as follows

$$
\begin{aligned}
H(\hat{Y}_1|UX_1X_2Y) - H(\hat{Y}_1|UX_2Y_1) &< I(X_2; Y|VUX_1) \\
H(\hat{Y}_1|UX_1X_2Y) - H(\hat{Y}_1|UX_1X_2YY_1) &< I(X_2; Y|VUX_1) \\
I(\hat{Y}_1; Y|UX_1X_2Y_1) &< I(X_2; Y|VUX_1), \\
&\quad (51)
\end{aligned}
$$

where the second step follows by the Markov relation $(X_1, Y) \ominus (U, X_2, Y_1) \ominus \hat{Y}_1$.

The combination of the first upper bound in (48) with (49) together with the constraints (50) and (51) proves the achievability of the rate $\bar{R}_2$.

4) Proof of Cardinality Bounds: We start by bounding the cardinality of the auxiliary random variable V. Recalling that the random variables $\hat{Y}_1$ and (Y, X_1) are conditionally
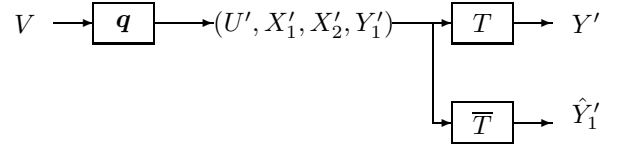

Fig. 3. The first chain generating the random variables in Theorem 4.

independent given (U, X_2, Y_1) we can write

$$
\begin{aligned}
I(X_1X_2; Y) &- I(\hat{Y}_1; Y_1|UX_1X_2Y) \\
&= H(Y) - H(Y|X_1X_2) \\
&\quad - \left[H(\hat{Y}_1|UX_1X_2Y) - H(\hat{Y}_1|UX_2Y_1) \right]. \quad (52)
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
I(X_1; Y\hat{Y}_1|UX_2) &+ I(U; Y_1|VX_2) \\
&= H(Y\hat{Y}_1|UX_2) - H(Y\hat{Y}_1|UX_1X_2) \\
&\quad + H(Y_1|VX_2) - H(Y_1|UVX_2), \quad (53)
\end{aligned}
$$

while the constraint (13) can be expressed as

$$
\begin{aligned}
I(X_2; Y|UVX_1) &- I(\hat{Y}_1; Y_1|UX_1X_2Y) \\
&= H(Y|UVX_1) - H(Y|UVX_1X_2) \\
&\quad - \left[H(\hat{Y}_1|UX_1X_2Y) - H(\hat{Y}_1|UX_2Y_1) \right] \geq 0. \quad (54)
\end{aligned}
$$

Let us be given the sets $\mathcal{U} = \{1, \ldots, J\}$, $\mathcal{X}_1 = \{1, \ldots, K\}$, $\mathcal{X}_2 = \{1, \ldots, L\}$, $\mathcal{Y}_1 = \{1, \ldots, M\}$, $\mathcal{Y} = \{1, \ldots, N\}$ and $\hat{\mathcal{Y}}_1 = \{1, \ldots, S\}$. Then for $j \in \{1, \ldots, J\}$, $k \in \{1, \ldots, K\}$, $l \in \{1, \ldots, L\}$, $m \in \{1, \ldots, M\}$, $n \in \{1, \ldots, N\}$, $s \in \{1, \ldots, S\}$ and any law $P(u, x_1, x_2, y_1, y, \hat{y}_1)$ on $(U, X_1, X_2, Y_1, Y, \hat{Y}_1)$ set

$$
\begin{aligned}
Q_{j,k,l,m} &= \Pr(U = j, X_1 = k, X_2 = l, Y_1 = m) \\
&= \sum_{n,s} \Pr(j, k, l, m, n, s), \quad (55)
\end{aligned}
$$

and define the conditional laws

$$
\begin{aligned}
t_n(k, l, m) &= \Pr(Y = n|U = j, X_1 = k, X_2 = l, Y_1 = m) \\
&= \frac{\sum_s \Pr(i, k, l, m, n, s)}{Q_{j,k,l,m}}, \\
\bar{t}_s(k, l, m) &= \Pr(\hat{Y}_1 = s|U = j, X_1 = k, X_2 = l, Y_1 = m) \\
&= \frac{\sum_n \Pr(k, l, m, n, s)}{Q_{j,k,l,m}}.
\end{aligned}
$$

Let T be the $N \times (J \times K \times L \times M)$ (5-dimensional) matrix with $(n, (j, k, l, m))$-th entry $t_n(j, k, l, m)$ and for $\ell = 1, 2, \ldots$ let Δ_ℓ be the simplex of probability ℓ-vectors. Then $\boldsymbol{Q} = (Q_{1,1,1,1}, Q_{2,1,1,1}, \ldots, Q_{J,K,L,M})^t \in \Delta_{JKLM}$, and $T\boldsymbol{Q} \in \Delta_N$. Thus, T defines a channel with inputs (U, X_1, X_2, Y_1) and output Y. Similarly, let $\overline{T}$ be the $S \times (J \times K \times L \times M)$ (5-dimensional) matrix with $(s, (j, k, l, m))$-th entry $\bar{t}_s(j, k, l, m)$, then $\overline{T}\boldsymbol{Q} \in \Delta_S$. Thus, $\overline{T}$ defines a channel with inputs (U, X_1, X_2, Y_1) and output $\hat{Y}_1$. This chain of generation of the random variables $(U, X_1, X_2, Y_1, Y, \hat{Y}_1)$ is illustrated in Figure 3.

Now let $\{\boldsymbol{q}(v)\}_v \in \mathcal{V}$ be a finite set of vectors in Δ_{JKLM}, indexed by the finite set $\mathcal{V}$. Also let $\{\lambda_v\}$ satisfy

$$\lambda_v \geq 0 \; ; \quad \sum_{v \in \mathcal{V}} \lambda_v = 1.$$

Let V be the random variable which takes the value $v \in \mathcal{V}$ with probability λ_v. Furthermore, suppose that V is the input to a channel with output (U', X_1', X_2', Y_1') taking values in $\mathcal{U} \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}_1$ with transition probability

$$\Pr(U' = j, X_1' = k, X_2' = l, Y_1' = m | V = v) = q_{j,k,l,m}(v),$$
$$1 \leq j \leq J,\ 1 \leq k \leq K,\ 1 \leq l \leq L,\ 1 \leq m \leq M$$

where $q_{j,k,l,m}(v)$ is the (j,k,l,m)-th component of $\boldsymbol{q}(v)$.

Let Y' be the output of the channel defined by T when (U', X_1', X_2', Y_1') is the input. The random variables U', X_1', X_2', Y_1', Y' have a joint law whose marginal on U', X_1', X_2', Y_1' satisfies (55) if and only if

$$\sum_{v \in \mathcal{V}} \lambda_v \boldsymbol{q}(v) = \boldsymbol{Q}. \tag{56}$$

Assuming that (56) is satisfied we can express, for example, the conditional entropies

$$H(Y|UX_1V) \quad = \quad \sum_{v \in \mathcal{V}} \lambda_v H(Y|UX_1V = v),$$

by substituting the expression

$$\Pr(Y = n | U = j, X_1 = k, V = v)$$
$$= \quad \frac{\sum_{l,m} t_n(j,k,l,m) q_{j,k,l,m}(v)}{\sum_{l,m} q_{j,k,l,m}(v)} \triangleq A_n(j,k,v),$$

while $H(\hat{Y}_1 | UX_1X_2Y)$ is expressed via

$$\Pr(Y = n, U = j, X_1 = k, X_2 = l)$$
$$= \quad \sum_m t_n(j,k,l,m) Q_{j,k,l,m},$$

and

$$\Pr(\hat{Y}_1 = s, Y = n, U = j, X_1 = k, X_2 = l)$$
$$= \quad \sum_m t_n(j,k,l,m) \bar{t}_s(j,k,l,m) Q_{j,k,l,m}.$$

Next, define

$$\Gamma_1(\boldsymbol{q}) \quad = \quad H(Y) - H(Y|X_1X_2)$$
$$- \left[H(\hat{Y}_1|UX_1X_2Y) - H(\hat{Y}_1|UX_2Y_1) \right],$$

to conclude that the functional (52) can be expressed as

$$I(X_1X_2; Y) - I(\hat{Y}_1; Y_1|UX_1X_2Y) = \sum_{v \in \mathcal{V}} \lambda_v \Gamma_1(\boldsymbol{q}). \tag{57}$$

Similarly, define

$$\Gamma_2(\boldsymbol{q}) \quad = \quad H(Y\hat{Y}_1|UX_2) - H(Y\hat{Y}_1|UX_1X_2)$$
$$+ H(Y_1|X_2V = v) - H(Y_1|UX_2V = v),$$

to conclude that the functional (53) can be expressed as

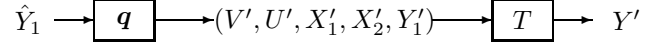$$I(X_1; Y\hat{Y}_1|UX_2) + I(U; Y_1|VX_2) = \sum_{v \in \mathcal{V}} \lambda_v \Gamma_2(\boldsymbol{q}). \tag{58}$$

$$\hat{Y}_1 \longrightarrow \boxed{q} \longrightarrow (V', U', X_1', X_2', Y_1') \longrightarrow \boxed{T} \longrightarrow Y'$$

Fig. 4. The second chain generating the random variables in Theorem 4.

Finally, define

$$\Gamma_3(\boldsymbol{q}) \quad = \quad H(Y|UX_1V = v) - H(Y|UX_1X_2V = v)$$
$$- \left[H(\hat{Y}_1|UX_1X_2Y) - H(\hat{Y}_1|UX_2Y_1) \right],$$

to conclude that the constraint (54) can be expressed as

$$I(X_2; Y|UVX_1) - I(\hat{Y}_1; Y_1|UX_1X_2Y) = \sum_{v \in \mathcal{V}} \lambda_v \Gamma_3(\boldsymbol{q}) \geq 0. \tag{59}$$

Combining (56), (57), (58) and (59) we conclude, based on similar arguments as in [8, Appendix A1], that the cardinality of $\mathcal{V}$ can be bounded by

$$\|\mathcal{V}\| \leq \|\mathcal{U}\| \|\mathcal{X}_1\| \|\mathcal{X}_2\| \|\mathcal{Y}_1\| + 2.$$

We consider next an upper bound on the cardinality of the auxiliary random variable $\hat{Y}_1$. This time we use the chain depicted in Figure 4 in order to show that the cardinality of $\hat{Y}_1$ can be bounded as follows

$$\|\hat{\mathcal{Y}}_1\| \quad \leq \quad \|\mathcal{V}\| \|\mathcal{U}\| \|\mathcal{X}_1\| \|\mathcal{X}_2\| \|\mathcal{Y}_1\| + 2.$$

Finally, we prove the equivalence of constraint (13) with the following constraint

$$I(\hat{Y}_1; Y_1|UX_2) \leq I(X_2\hat{Y}_1; YX_1|UV). \tag{60}$$

This can directly be seen by expressing the right hand side of (60) as

$$I(X_2\hat{Y}_1; X_1Y|UV)$$
$$= I(X_2; X_1Y|UV) + I(\hat{Y}_1; X_1Y|UVX_2)$$
$$\overset{(e)}{=} I(X_2; Y|UVX_1) + H(\hat{Y}_1|UVX_2) - H(\hat{Y}_1|UVX_1X_2Y)$$
$$\overset{(f)}{=} I(X_2; Y|UVX_1) + H(\hat{Y}_1|UVX_2) - H(\hat{Y}_1|UX_1X_2Y) \tag{61}$$

and the left hand side as

$$I(Y_1; \hat{Y}_1|UX_2)$$
$$= H(\hat{Y}_1|UX_2) - H(\hat{Y}_1|UX_2Y_1)$$
$$\overset{(g)}{=} H(\hat{Y}_1|UVX_2) - H(\hat{Y}_1|UX_1X_2YY_1)$$
$$= I(\hat{Y}_1; Y_1|UX_1X_2Y) + H(\hat{Y}_1|UVX_2) - H(\hat{Y}_1|UX_1X_2Y_1). \tag{62}$$

Here,

(e) follows from the Markovity of $X_1 \ominus (V, U) \ominus X_2$ which implies that $I(X_2; X_1|UV) = 0$;

(f) follows from the Markovity of $V \ominus (U, X_1, X_2, Y) \ominus \hat{Y}_1$; and

(g) follows from the Markovity of $V \ominus (U, X_2) \ominus \hat{Y}_1$ and the Markovity of $(X_1, Y) \ominus (U, X_2, Y_1) \ominus \hat{Y}_1$.

B. Derivation of Upper Bound in Observation 1

Suppose there exists an (M, n, ϵ)-code for the relay with receiver-transmitter feedback. The probability mass function on the joint ensemble $(W, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}, \boldsymbol{Y}_1)$ is given by

$$p(w, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}, \boldsymbol{y}_1) =$$
$$\frac{1}{M} \prod_{k=1}^{n} p(x_{1,k}|w, y^{k-1}) p(x_{2,k}|y_1^{k-1}) p(y_k, y_{1,k}|x_{1,k}, x_{2,k}).$$

Now, the Fano inequality yields

$$H(W|\hat{W}) \leq \epsilon \ln M + h(\epsilon) \triangleq n\delta_n(\epsilon), \qquad (63)$$

where, $h(\cdot)$ denotes the binary entropy function and $\delta_n(\epsilon) \to 0$ as $\epsilon \to 0$. From (3) and (63) it follows that

$$H(W|\boldsymbol{Y}) \leq H(W|\hat{W}) \leq n\delta_n(\epsilon). \qquad (64)$$

Consider the identity

$$nR = H(W) = I(W; \boldsymbol{Y}) + H(W|\boldsymbol{Y}).$$

Combining this with (64) we obtain

$$nR \leq I(W; \boldsymbol{Y}) + n\delta_n(\epsilon).$$

We now proceed with a chain of inequalities for $I(W; \boldsymbol{Y})$, where the explanations will follow:

$$I(W; \boldsymbol{Y})$$
$$\leq \quad I(W; \boldsymbol{Y}, \boldsymbol{Y}_1)$$
$$\overset{(a)}{=} \quad \sum_{k=1}^{n} I(W; Y_k Y_{1,k} | Y^{k-1} Y_1^{k-1} X_{2,k})$$
$$= \quad \sum_{k=1}^{n} \Big(H(Y_k Y_{1,k} | Y^{k-1} Y_1^{k-1} X_{2,k})$$
$$\qquad - H(Y_k Y_{1,k} | W Y^{k-1} Y_1^{k-1} X_{2,k}) \Big)$$
$$\overset{(b)}{=} \quad \sum_{k=1}^{n} \Big(H(Y_k Y_{1,k} | Y^{k-1} Y_1^{k-1} X_{2,k})$$
$$\qquad - H(Y_k Y_{1,k} | X_{1,k} X_{2,k}) \Big)$$
$$\overset{(c)}{=} \quad \sum_{k=1}^{n} \Big(H(Y_k Y_{1,k} | Y^{k-1} Y_1^{k-1} X_{2,k})$$
$$\qquad - H(Y_k Y_{1,k} | Y^{k-1} Y_1^{k-1} X_{1,k} X_{2,k}) \Big).$$

Here,

(a) follows by the functional relationship (2);
(b) follows from the functional relationship (1) and from the Markovity of $(WY^{k-1}Y_1^{k-1}) \ominus (X_{1,k} X_{2,k}) \ominus (Y_k Y_{1,k})$; and
(c) follows from the fact that $(Y_k Y_{1,k})$ are conditionally independent of $(Y^{k-1} Y_1^{k-1})$ given $(X_{1,k} X_{2,k})$.

Define

$$V_k \triangleq (Y^{k-1} Y_1^{k-1}),$$

then we have shown that

$$I(W; \boldsymbol{Y}) \quad \leq \quad \sum_{k=1}^{n} I(X_{1,k}; Y_k Y_{1,k} | X_{2,k} V_k).$$

Now, let Z be a random variable independent of $\boldsymbol{V}, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}, \boldsymbol{Y}_1$ uniformly distributed over the set $\{1, \ldots, n\}$, and set

$$X_1 \triangleq X_{1,Z}, \quad X_2 \triangleq X_{2,Z}, \quad Y \triangleq Y_Z, \quad Y_1 \triangleq Y_{1,Z}, \quad V \triangleq V_Z.$$

Then

$$\frac{1}{n} \sum_{k=1}^{n} I(X_{1,k}; Y_k Y_{1,k} | X_{2,k} V_k) = I(X_1; YY_1 | X_2, V, Z). \qquad (65)$$

Next

$$I(X_1; YY_1 | X_2 V Z)$$
$$= \quad H(YY_1 | X_2 V Z) - H(YY_1 | X_1 X_2 V Z)$$
$$\overset{(d)}{=} \quad H(YY_1 | X_2 V Z) - H(YY_1 | X_1 X_2 V)$$
$$\overset{(e)}{\leq} \quad H(YY_1 | X_2 V) - H(YY_1 | X_1 X_2 V)$$
$$= \quad I(X_1; YY_1 | X_2 V). \qquad (66)$$

Here,

(d) follows by the Markovian relation $Z \ominus V \ominus (X_1 X_2) \ominus (YY_1)$; and
(e) follows since conditioning reduces entropy.

The combination of (65) and (66) yields that

$$\frac{1}{n} \sum_{k=1}^{n} I(X_{1,k}; Y_k Y_{1,k} | X_{2,k} V_k) \leq I(X_1; YY_1 | X_2 V). \qquad (67)$$

The inequality

$$I(W; \boldsymbol{Y}) \leq \sum_{k=1}^{n} I(X_{1,k} X_{2,k}; Y_k), \qquad (68)$$

is proved in [3, Lemma 4].

Combining (67) and (68) we conclude that

$$R \leq \sup_{p_{V X_1 X_2 Y Y_1}} \min \left\{ \begin{array}{c} I(X_1 X_2; Y), \\ I(X_1; YY_1 | X_2 V) \end{array} \right\} + \delta_n(\epsilon),$$

where the supremum is taken over all joint laws of the form

$$p_{V X_1 X_2 Y Y_1}(v, x_1, x_2, y, y_1)$$
$$= \quad p_{V X_1 X_2}(v, x_1, x_2) p_{Y Y_1 | X_1 X_2}(y, y_1 | x_1, x_2).$$

Now, a bound on the cardinality of V can be obtained via the technique presented in [8, Appendix A1].

This completes the proof of Proposition 1.

C. Proof of Theorem 5

To prove Theorem 5 we propose a coding scheme where the transmitter sends maximally informative updates similar to [15] and the relay applies a simple amplify-and-forward strategy.

1) Coding Scheme: Prior to transmission the encoder maps the message W onto the real line applying the following one-to-one mapping

$$\theta: \quad w \mapsto \frac{w-1}{M-1} - \frac{1}{2}. \quad (69)$$

Consequently, the random variable $\theta(W)$ is distributed uniformly over M equally spaced values within $\left[-\frac{1}{2}, \frac{1}{2}\right]$. The message point $\theta(W)$ is then transmitted over n channel uses to the receiver. After the reception of the n channel outputs the receiver guesses the transmitted message point $\theta(W)$ and equivalently the message W.

In the remaining of this section we describe the transmission steps and the decoding in detail, followed by an analysis of the performance.

First Transmission Step, $k = 1$: In the first transmission step the encoder transmits a scaled version of the message point $\theta = \theta(W)$

$$X_{1,1} = \sqrt{\frac{P_1}{\text{Var}(\theta)}} \theta.$$

Note that the scaling factor assures that the expected power of the input symbol equals P_1.

The relay stays quiet and the decoder thus receives $Y_1 = \sqrt{\frac{P_1}{\text{Var}(\theta)}} \theta + Z_{2,1}$ and estimates θ as follows

$$\hat{\theta}_1 = \sqrt{\frac{\text{Var}(\theta)}{P_1}} Y_1 = \theta + \sqrt{\frac{\text{Var}(\theta)}{P_1}} Z_{2,1}.$$

As a result, the decoder's estimation error $\epsilon_1 \triangleq \hat{\theta}_1 - \theta = \sqrt{\frac{P_1}{\text{Var}(\theta)}} Z_{2,1}$, is zero-mean Gaussian and of variance

$$\alpha_1 \triangleq \text{Var}(\epsilon_1) = N_2 \frac{\text{Var}(\theta)}{P_1}.$$

Note that this first estimate of θ is sub-optimal in terms of expected mean squared error. However, the advantage is that with this estimate the error ϵ_1 is zero-mean Gaussian, and as we will see later on, this simplifies the analysis. Also, due to the feedback link the encoder observes Y_1 as well and thus with the knowledge of θ it can compute the estimation error ϵ_1.

In the subsequent transmissions the encoder sends resolution information such that the decoder can form a better and better estimate of ϵ_1 and equivalently of θ.

Second Transmission Step, $k = 2$: In the second transmission step the encoder sends a scaled version of the estimation error ϵ_1 while the relay stays again quite. Thus $X_{2,2} = 0$ and

$$X_{1,2} = \sqrt{\frac{P_1}{\alpha_1}} \epsilon_1.$$

Note that here the factor $\sqrt{\frac{P_1}{\alpha_1}}$ is chosen such that the expected power of the transmitted symbol $X_{1,2}$ equals P_1.

The channel output observed at the receiver is given by $Y_2 = \sqrt{\frac{P_1}{\alpha_1}} \epsilon_1 + Z_{2,2}$ and the receiver computes the linear

minimum means squared error (LMMSE) estimate of ϵ_1 based on Y_2

$$\hat{\epsilon}_1 = \frac{\sqrt{\alpha_1 P_1}}{P_1 + N_2} Y_2.$$

Then the receiver updates its estimate of the message point θ as follows

$$\hat{\theta}_2 = \hat{\theta}_1 - \hat{\epsilon}_1,$$

and the new estimation error becomes

$$\epsilon_2 \triangleq \hat{\theta}_2 - \theta = \epsilon_1 - \hat{\epsilon}_1,$$

and is of variance

$$\alpha_2 \triangleq \text{Var}(\epsilon_2) = \alpha_1 \frac{N_2}{P_1 + N_2}.$$

In this second transmission step the relay observes $Y_{1,2} = \sqrt{\frac{P_1}{\alpha_1}} \epsilon_1 + Z_{1,2}$.

Further Transmission Steps, $k = 3, \ldots, n$: Prior to transmission step k the encoder observes the feedback outputs $Y_2, \ldots, Y_{k-1}$ and hence knowing the message point θ it can compute the decoder's LMMSE estimation error $\epsilon_{1,k-1}$. Then in transmission step k the encoder sends the estimation error ϵ_{k-1} scaled by the factor $\sqrt{\frac{P_1}{\alpha_{k-1}}}$ where α_{k-1} denotes the variance of ϵ_{k-1}. Again, the scaling factor assures that the expected power of the input symbol equals P_1.

The relay applies an amplify-and-forward strategy, that is, in transmission step k it transmits a scaled version of the symbol $Y_{1,k-1}$ received in the previous step:

$$\begin{aligned} X_{2,k} &= \sqrt{\frac{\tilde{P}_2}{P_1 + N_1}} Y_{1,k-1} \\ &= \sqrt{\frac{\tilde{P}_2}{P_1 + N_1}} (X_{1,k-1} + Z_{1,k-1}) \\ &= \sqrt{\frac{\tilde{P}_2}{P_1 + N_1}} \left(\sqrt{\frac{P_1}{\alpha_{k-2}}} \epsilon_{k-2} + Z_{1,k-1} \right). \end{aligned}$$

Here, the scaling factor is chosen as $\sqrt{\frac{\tilde{P}_2}{P_1 + N_1}}$ for some $\tilde{P}_2 \in [0, P_2]$ and thus the expected power of the input symbol $X_{2,k}$ equals $\tilde{P}_2 \leq P_2$.

The time-k channel output at the receiver is given by

$$\begin{aligned} Y_k &= X_{1,k} + dX_{2,k} + Z_{2,k} \\ &= \sqrt{\frac{P_1}{\alpha_{k-1}}} \epsilon_{k-1} + d\sqrt{\frac{\tilde{P}_2}{P_1 + N_1}} \sqrt{\frac{P_1}{\alpha_{k-2}}} \epsilon_{k-2} \\ &\quad + d\sqrt{\frac{\tilde{P}_2}{P_1 + N_1}} Z_{1,k-1} + Z_{2,k}. \quad (70) \end{aligned}$$

Using this current and all the previous channel outputs the receiver updates its estimate of the message point θ. It computes $\hat{\epsilon}_{k-1}$, the LMMSE-estimate of ϵ_{k-1} based on the observations $Y_2, \ldots, Y_k$, and subtracts it from the previous estimate $\hat{\theta}_{k-1}$ to obtain the new estimate

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \hat{\epsilon}_{k-1}. \quad (71)$$

So, the remaining task is to compute the LMMSE-estimate $\hat{\epsilon}_{k-1}$. As Y_k is not independent of the previous outputs, the easiest way to do this is to compute the best (in LMMSE-sense) $\hat{Y}_k$ of the channel output Y_k based on $(Y_2, \ldots, Y_{k-1})$ and then subtract it from the original observation Y_k. The resulting innovation $I_k \triangleq Y_k - \hat{Y}_k$ is then independent of the previous observations, and the LMMSE-estimate $\hat{\epsilon}_{k-1}$ is easily computed from this innovation only. Note that since among the summands of Y_k only ϵ_{k-2} depends on the past observations the best predictor $\hat{Y}_k$ is given by a scaled version of the LMMSE-estimate of ϵ_{k-2} based on $(Y_2, \ldots, Y_{k-1})$. Denoting the LMMSE-estimate of ϵ_{k-2} by $\hat{\epsilon}_{k-2}$ the best predictor can be written as

$$\hat{Y}_k = d\sqrt{\frac{\tilde{P}_2}{P_1 + N_1}} \sqrt{\frac{P_1}{\alpha_{k-2}}} \hat{\epsilon}_{k-2}.$$

Based on the predictor the receiver can form the innovation

$$
\begin{aligned}
I_k &\triangleq Y_k - \hat{Y}_k \\
&= \sqrt{\frac{P_1}{\alpha_{k-1}}}\epsilon_{k-1} + d\sqrt{\frac{\tilde{P}_2}{P_1+N_1}}\sqrt{\frac{P_1}{\alpha_{k-2}}}\left(\epsilon_{k-2} - \hat{\epsilon}_{k-2}\right) \\
&\quad + d\sqrt{\frac{\tilde{P}_2}{P_1+N_1}}Z_{1,k-1} + Z_{2,k} \\
&= \sqrt{\frac{P_1}{\alpha_{k-1}}}\epsilon_{k-1} + d\sqrt{\frac{\tilde{P}_2}{P_1+N_1}}\sqrt{\frac{P_1}{\alpha_{k-2}}}\epsilon_{k-1} \\
&\quad + d\sqrt{\frac{\tilde{P}_2}{P_1+N_1}}Z_{1,k-1} + Z_{2,k}
\end{aligned}
$$

which, as already mentioned, is independent of the previous observations $(Y_2, \ldots, Y_{k-1})$. Note that in the second equality we used that

$$\epsilon_{k-1} = \epsilon_{k-2} - \hat{\epsilon}_{k-2}$$

which follows from the definition of $\epsilon_{k-1} \triangleq \theta - \hat{\theta}_{k-1}$ and Recursion (71) for $\hat{\theta}_{k-1}$.

The LMMSE-estimate of ϵ_{k-1} at the receiver based on $(Y_2, \ldots, Y_{k-1}, I_k)$ is given by $\hat{\epsilon}_{k-1} = \frac{\text{Cov}[\epsilon_{k-1}, I_k]}{\text{Var}(I_k)}I_k$, and the new estimation error is

$$\epsilon_k \triangleq \hat{\theta}_k - \theta = \epsilon_{k-1} - \hat{\epsilon}_{k-1} = \epsilon_{k-1} - \frac{\text{Cov}[\epsilon_{k-1}, I_k]}{\text{Var}(I_k)}I_k.$$

The variance of ϵ_k is given by

$$
\begin{aligned}
\alpha_k &= \alpha_{k-1} - \frac{\text{Cov}[\epsilon_{k-1}, I_k]^2}{\text{Var}(I_k)} \\
&= \alpha_{k-1}\left(\frac{\gamma\tilde{P}_2 + N_2}{P_1 + 2\sqrt{\bar{\gamma}P_1\tilde{P}_2}\sqrt{\frac{\alpha_{k-1}}{\alpha_{k-2}}} + \bar{\gamma}\frac{\alpha_{k-1}}{\alpha_{k-2}}\tilde{P}_2 + \gamma\tilde{P}_2 + N_2}\right)
\end{aligned}
$$

where we defined $\gamma \triangleq d^2\frac{N_1}{P_1+N_1}$ and $\bar{\gamma} = d^2 - \gamma$. Defining further $\rho_k = \sqrt{\frac{\alpha_{k-1}}{\alpha_{k-2}}}$ the variance α_k can be expressed as

$$\alpha_k = \alpha_{k-1}\left(\frac{\gamma\tilde{P}_2 + N_2}{P_1 + 2\sqrt{\bar{\gamma}P_1\tilde{P}_2}\rho_{k-1} + \bar{\gamma}\rho_{k-1}^2\tilde{P}_2 + \gamma\tilde{P}_2 + N_2}\right). \quad (72)$$

Equation (72) also leads to a recursive formulation of the sequence $\{\rho_k\}$:

$$\rho_k = \sqrt{\frac{\gamma\tilde{P}_2 + N_2}{\left(\sqrt{P_1} + \rho_{k-1}\sqrt{\bar{\gamma}\tilde{P}_2}\right)^2 + \gamma\tilde{P}_2 + N_2}}, \quad (73)$$

for $k = 3, \ldots, n-1$, and

$$\rho_2 = \sqrt{\frac{N_2}{P_1 + N_2}}.$$

Note that ρ_k equals the correlation coefficient of ϵ_{k-1} and ϵ_k, and thus is proportional to the correlation of the time-$(k+1)$ signal from the encoder to the receiver and the time-$(k+1)$ signal from the relay to the receiver.

Decoding of the Message after Step n: After the n-th transmission step the decoder's estimate of the message point θ is given by $\hat{\theta}_n = \theta + \epsilon_n$. The decoder then guesses that the message $\hat{W} = \hat{w}$ was sent if $\theta(\hat{w})$ is the message point closest to $\hat{\theta}_n$, i.e.,

$$\hat{w} = \arg\min_w |\hat{\theta}_n - \theta(w)|.$$

2) Performance analysis: An error in the decoding occurs only if there is a $w' \neq w$ such that the message point $\theta(w')$ is closer to $\hat{\theta}_n$ than the message point $\theta(w)$. The probability of this event is upper bounded by the probability that the magnitude of ϵ_n is greater than half the distance between adjacent message points. Therefore the probability of error can be upper bounded as

$$P_e \leq \Pr\left[|\epsilon_n| > \frac{1}{2(M-1)}\right] \leq 2Q\left(\frac{1}{2M\sqrt{\alpha_n}}\right), \quad (74)$$

where $Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$ is the tail of the standard Gaussian distribution evaluated at x.

In the above term the variance α_n can be expressed by defining $\gamma \triangleq d^2\frac{N_1}{P_1+N_1}$ and $\bar{\gamma} = d^2 - \gamma$, and by iteratively applying (72)

$$\alpha_n = \alpha_2 \prod_{k=3}^n \left(\frac{\gamma\tilde{P}_2 + N_2}{P_1 + 2\sqrt{\bar{\gamma}P_1\tilde{P}_2}\rho_{k-1} + \bar{\gamma}\tilde{P}_2\rho_{k-1}^2 + \gamma\tilde{P}_2 + N_2}\right).$$

The size of the message set M can be expressed in terms of the transmission rate $R = \frac{1}{n}\ln M$. Then we obtain the upper bound on the probability of a decoding error in (75), on top of the next page. We see that the probability of error tends to 0 when $n \to \infty$ if

$$R < \lim_{n\to\infty}\frac{1}{n}\sum_{k=3}^n \Gamma\left(\frac{P_1 + 2\sqrt{\bar{\gamma}P_1\tilde{P}_2}\rho_{k-1} + \bar{\gamma}\tilde{P}_2\rho_{k-1}^2}{\gamma\tilde{P}_2 + N_2}\right). \quad (76)$$

The convergence of the right hand side of (76) to the bound for R given in Theorem 5 follows by showing that the sequence of correlation coefficients $\{\rho_k\}$ converges to ρ^*, the solution of (27), and then applying Cesáro's Mean Theorem [4, Theorem 4.2.3].

$$P_e \leq 2Q\left(\frac{1}{2\sqrt{\alpha_2}} \cdot \exp\left(\sum_{k=3}^{n} \frac{1}{2}\ln\left(\frac{P_1 + 2\sqrt{\bar{\gamma}P_1\tilde{P}_2}\rho_{k-1} + \bar{\gamma}\tilde{P}_2\rho_{k-1}^2 + d^2\frac{N_1}{P_1+N_1}\tilde{P}_2 + N_2}{d^2\frac{N_1}{P_1+N_1}\tilde{P}_2 + N_2}\right) - nR\right)\right),$$

(75)

In order to prove the convergence of the sequence $\{\rho_k\}$ to ρ^* we need the following lemma.

Lemma 2: Consider the function $f : x \mapsto \sqrt{\frac{a}{a+p(1+bx)^2}}$ defined on the closed interval $[0,1]$ when $a,b,p \geq 0$. For the function $f(\cdot)$ exactly one fixed point x^* exists in $[0,1]$ and for any starting point $x_0 \in [0,1]$ the infinite sequence $x_0, x_1 = f(x_0), x_2 = f(x_1), \ldots$ converges to this fixed point x^*.

Proof: First note that $f(\cdot)$ is continuous and strictly decreasing on the interval $[0,1]$, and that the image of $[0,1]$ under $f(\cdot)$ is a subset of the interval itself. These two properties imply that there is exactly one fixed point of $f(\cdot)$ in $[0,1]$ which we denote by x^*.

Now we proceed to prove the second part of the lemma, i.e., that for arbitrary starting point $x_0 \in [0,1]$ the sequence obtained by iteratively applying the mapping $f(\cdot)$ converges to x^*. To this end, for the chosen starting point define two sequences $\{y_k\}_{k=0}^{\infty}$ and $\{z_k\}_{k=0}^{\infty}$ where the first sequence is defined by $y_0 = x_0, y_1 = f(f(y_0)), y_2 = f(f(y_1)), \ldots$, and the second sequence by $z_0 = f(x_0), z_1 = f(f(z_0)), z_2 = f(f(z_1)), \ldots$ We will show that both sequences converge to the fixed point x^* of $f(\cdot)$, from which then follows that the sequence $x_0, x_1, x_2, \ldots$ also converges to x^*. Note that since x^* is a fixed point of $f(\cdot)$ it is clearly also a fixed point of $(f \circ f)$. Note further that since $f(\cdot)$ is strictly monotonically decreasing on $[0,1]$ either $x_0 = f(x_0) = x^*$, $x_0 < x^* < f(x_0)$, or $f(x_0) < x^* < x_0$. For the first case the lemma follows directly. We will prove the lemma for the second case, the proof of the third case is omitted but follows along the same lines as the proof for the second case.

Thus, in the following we assume that $0 \leq x_0 < x^*$. We start by proving the convergence of the sequence $y_0, y_1, y_2, \ldots$. Note first that since $f(\cdot)$ is strictly monotonically decreasing it follows that $f(y) > x^*$ and $f(f(y)) < x^*$ for all $0 \leq y < x^*$. From the assumption that $y_0 = x_0 < x^*$ then follows that the sequence $y_0, y_1, y_2, \ldots$ is upper bounded by x^*. Next, we show that the sequence is strictly monotonically increasing and that x^* is the only fixed point of $(f \circ f)$ in the interval $[0, x^*]$. Both properties follow by showing that $f(f(y)) > y$ for $y \in [0, x^*[$ or, since y and $f(f(y))$ are non-negative, by

equivalently showing that for $y \in [0, x^*[$:

$$\frac{(f(f(y)))^2}{y^2}$$
$$= \left(\frac{a}{a + p\left(1 + b\sqrt{\frac{a}{a+p(1+by)^2}}\right)^2}\right) \cdot \frac{1}{y^2}$$
$$= \frac{a}{ay^2 + py^2 + 2bpy\sqrt{\frac{ay^2}{a+p(1+by)^2}} + pb^2\frac{ay^2}{a+p(1+by)^2}} > 1.$$

(77)

Note that the expression on the left hand side of the inequality in (77) is strictly monotonically decreasing for $y \geq 0$, and also note that for $y = x^*$ the expression must be equal to 1, since x^* is a fixed point of $(f \circ f)$. Hence for all y larger than 0 and strictly smaller than x^* the above ratio has to be strictly larger than 1.

Concluding we have shown that the sequence $y_0, y_1, y_2, \ldots$ is strictly monotonically increasing and upper bounded by x^* which is the only fixed point of $(f \circ f)$ in $[0, x^*]$. From this follows that the sequence $y_0, y_1, \ldots$ converges to the fixed point x^*.

Similar arguments can be applied to show that also the sequence $z_0, z_1, z_2, \ldots$ converges to x^* which concludes the proof of the lemma. ∎

Applying Lemma 2 to the sequence of correlation coefficients $\{\rho_k\}$ it follows that the sequence converges to the unique fixed point of the Recursion (73) in $[0,1]$ which is given by the unique solution in the interval $[0,1]$ of the following quartic equation in ρ

$$\rho^2\left(P_1 + 2d\sqrt{P_1\tilde{P}_2}\sqrt{\frac{P_1}{P_1+N_1}}\rho + d^2\tilde{P}_2\frac{P_1}{P_1+N_1}\rho^2\right.$$
$$\left. + d^2\frac{N_1}{P_1+N_1}\tilde{P}_2 + N_2\right) = d^2\tilde{P}_2\frac{N_1}{P_1+N_1} + N_2.$$

Substituting γ by $d^2\frac{N_1}{P_1+N_1}$ and $\bar{\gamma}$ by $d^2 - \gamma$ one obtains (27).

D. Proof of Corollary 1

We apply the coding scheme which achieves the rate $\bar{R}_2$ of Theorem 4 for the Gaussian relay channel. Even though a jointly Gaussian distribution does not necessarily maximize (12) subject to the constraints (13) and (15) we let $W \sim \mathcal{N}(0, \rho)$, $\tilde{U} \sim \mathcal{N}(0, 1-\rho)$ and $\tilde{V} \sim \mathcal{N}(0, 1-\rho)$ where $W, \tilde{U}$ and $\tilde{V}$ are independent and form $U = W + \tilde{U}$ and $V = W + \tilde{V}$. Independently of these random variables and independently of each other we let $\tilde{X}_1 \sim \mathcal{N}(0, \alpha_1 P_1)$ and $\tilde{X}_2 \sim \mathcal{N}(0, \alpha_2 P_2)$ and set $X_1 = \sqrt{\bar{\alpha}_1 P_1}U + \tilde{X}_1$ and $X_2 = \sqrt{\bar{\alpha}_2 P_2}V + \tilde{X}_2$ where

$\bar{\alpha}_1 = 1 - \alpha_1$ and $\bar{\alpha}_2 = 1 - \alpha_2$. Finally, we choose for the quantization step $\hat{Y}_1 = \alpha(Y_1 + Z')$ where $Z' \sim \mathcal{N}(0, N')$ and is independent of $(W, \tilde{U}, \tilde{V}, Z_1, Z_2)$.

Then the following rate is achievable

$$R = \min \left\{ \begin{array}{l} \mathcal{C}\left(\frac{P_1 + d^2 P_2 + 2d\sqrt{P_1 P_2}\sqrt{\bar{\alpha}_1 \bar{\alpha}_2}\rho}{N_2}\right) - \mathcal{C}\left(\frac{N_1}{N'}\right), \\ \mathcal{C}\left(\frac{\alpha_1 P_1}{N_2} + \frac{\alpha_1 P_1}{N_1 + N'}\right) + \mathcal{C}\left(\frac{\bar{\alpha}_1 P_1 (1 - \rho^2)}{\alpha_1 P_1 + N_1}\right) \end{array} \right\}$$

where the parameters α_1, α_2, ρ and N' must satisfy

$$0 \leq \alpha_1 \leq 1,$$
$$0 \leq \alpha_2 \leq 1,$$
$$0 \leq \rho \leq 1,$$
$$\frac{N_1 N_2}{d^2 \alpha_2 P_2} \leq N',$$
$$\frac{N_1 + \alpha_1 P_1 \frac{N_2}{\alpha_1 P_1 + N_2}}{N'} \leq \frac{\bar{\alpha}_1 \rho^2 P_1 + P_2 d^2 + 2d\sqrt{\bar{\alpha}_1 \bar{\alpha}_2 P_1 P_2}\rho}{\alpha_1 P_1 + \bar{\alpha}_1 P_1 (1 - \rho^2) + N_2}.$$

ACKNOWLEDGMENT

APPENDIX

A. Strong Typicality

Let $\left\{ X^{(1)}, X^{(2)}, \ldots, X^{(k)} \right\}$ denote a finite collection of discrete random variables with some joint distribution $P\left(x^{(1)}, x^{(2)}, \ldots, x^{(k)}\right)$ with $\left(x^{(1)}, x^{(2)}, \ldots, x^{(k)}\right) \in \mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \ldots \times \mathcal{X}^{(k)}$. Let S denote an ordered nonempty subset of these random variables and consider n independent copies of S. Thus, with $\boldsymbol{S} \triangleq (S_1, S_2, \ldots, S_n)$,

$$\Pr\{\boldsymbol{S} = \boldsymbol{s}\} = \prod_{j=1}^n \Pr\{S_j = s_j\}.$$

Let $N(s; \boldsymbol{s})$ be the number of indices $j \in \{1, 2, \ldots, n\}$ such that $S_j = s$. By the law of large numbers, for any subset S of random variables and for all $s \in S$,

$$\frac{1}{n} N(s; \boldsymbol{s}) \to P(s), \tag{78}$$

as well as

$$-\frac{1}{n} \ln P(s_1, s_2, \ldots, s_n) = -\frac{1}{n} \sum_{j=1}^n \ln P(s_j) \to H(S). \tag{79}$$

The convergence in (78) and (79) takes place simultaneously with probability one for all nonempty subsets S [3].

Definition 1: The set $\mathcal{A}_\epsilon$ of ϵ-strongly typical n-sequences is defined by (see [4, Chapter 3,12,13])

$$\begin{aligned} \mathcal{A}_\epsilon & \triangleq \mathcal{A}_\epsilon\left(X^{(1)}, X^{(2)}, \ldots, X^{(k)}\right) \\ & \triangleq \left\{ \left(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(k)}\right) : \right. \\ & \quad \left| \frac{1}{n} N\left(x^{(1)}, x^{(2)}, \ldots, x^{(k)}; \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(k)}\right) \right. \\ & \quad \left. - P\left(x^{(1)}, x^{(2)}, \ldots, x^{(k)}\right) \right| \\ & \quad < \frac{\epsilon}{\|\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \ldots \times \mathcal{X}^{(k)}\|}, \\ & \quad \left. \forall \left(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(k)}\right) \in \mathcal{X}^{(1)} \times \ldots \times \mathcal{X}^{(k)} \right\}, \end{aligned}$$

where $\|\mathcal{X}\|$ is the cardinality of the set $\mathcal{X}$.

Let $\mathcal{A}_\epsilon(S)$ be defined similar to $\mathcal{A}_\epsilon$ but now with constraints corresponding to all nonempty subsets of S. We recall now two basic lemmas (for the proofs we refer to [4]).

Lemma 3: For any $\epsilon > 0$ the following statements hold for every integer $n \geq 1$:

1) If $\boldsymbol{s} \in \mathcal{A}_\epsilon(S)$, then $\exp\left(-n(H(S) + \epsilon)\right) \leq \Pr\{\boldsymbol{S} = \boldsymbol{s}\} \leq \exp\left(-n(H(S) - \epsilon)\right)$.
2) If $S_1, S_2 \subseteq \{X_1, X_2, \ldots, X_k\}$ and $(\boldsymbol{s}_1, \boldsymbol{s}_2) \in \mathcal{A}_\epsilon(S_1 \cup S_2)$, then

$$\begin{aligned} \exp\left(-n(H(S_1|S_2) + 2\epsilon)\right) & \leq \Pr\{\boldsymbol{S}_1 = \boldsymbol{s}_1 | \boldsymbol{S}_2 = \boldsymbol{s}_2\} \\ & \leq \exp\left(-n(H(S_1|S_2) - 2\epsilon)\right). \end{aligned}$$

Moreover, the following statements hold for every sufficiently large n:

3) $\Pr\{\mathcal{A}_\epsilon(S)\} \geq 1 - \epsilon$,
4) $(1 - \epsilon) \exp(n(H(S) - \epsilon)) \leq \|\mathcal{A}_\epsilon(S)\| \leq \exp(n(H(S) + \epsilon))$.

Lemma 4: Let the discrete random variables X, Y, Z have joint distribution $P_{X,Y,Z}(x, y, z)$. Let X' and Y' be conditionally independent given Z, with the marginals

$$\begin{aligned} P_{X'|Z}(x|z) & = \sum_y P_{X,Y,Z}(x, y, z)/P_Z(z), \\ P_{Y'|Z}(y|z) & = \sum_x P_{X,Y,Z}(x, y, z)/P_Z(z). \end{aligned}$$

Let $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) \sim \prod_{k=1}^n P_{X,Y,Z}(x_k, y_k, z_k)$ and $(\boldsymbol{X}', \boldsymbol{Y}', \boldsymbol{Z}) \sim \prod_{k=1}^n P_{X'|Z}(x_k|z_k) P_{Y'|Z}(y_k|z_k) P_Z(z_k)$. Then

$$\Pr\left\{(\boldsymbol{X}', \boldsymbol{Y}', \boldsymbol{Z}) \in \mathcal{A}_\epsilon(X, Y, Z)\right\} \leq \exp(-n[I(X; Y|Z) - \epsilon]).$$

B. Proof of Lemma 1

We assume that the decoding of the previous blocks was successful, i.e, we assume the event $F^{(1\ldots b-1)^c}$, and we assume that the relay has sent $s^{(b)}$. Then, we define $\Psi\left(z_E | \boldsymbol{Y}^{(b)}\right)$ in (80) on top of the next page. The cardinality of $\mathcal{L}_E\left(\boldsymbol{Y}^{(b)}\right)$ is the random variable

$$\left\| \mathcal{L}_E\left(\boldsymbol{Y}^{(b)}\right) \right\| = \sum_{z_E} \Psi\left(z_E | \boldsymbol{Y}^{(b)}\right),$$

$$\Psi\left(z_E|\boldsymbol{Y}^{(b)}\right) = \begin{cases} 1 & \left(\boldsymbol{V}(\omega_0^{(b)}), \boldsymbol{U}(w_1^{(b)}, \omega_0^{(b)}), \boldsymbol{X}_1(w^{(b)}, \omega_0^{(b)}), \boldsymbol{X}_2(s^{(b)}, \omega_0^{(b)}), \boldsymbol{Y}^{(b)}, \hat{\boldsymbol{Y}}_1(z_E, w_1^{(b)}, s^{(b)}, \omega_0^{(b)})\right) \in \mathcal{A}_\epsilon \\ 0 & \text{otherwise.} \end{cases} \quad (80)$$

and

$$\mathbb{E}\left\{\left\|\mathcal{L}_E\left(\boldsymbol{Y}^{(b)}\right)\right\| |F^{(b)^c}\right\} = \mathbb{E}\left\{\Psi\left(z^{(b)}|\boldsymbol{y}^{(b)}\right)|F^{(1...b-1)^c}\right\}$$
$$+ \sum_{z_E \neq z^{(b)}} \mathbb{E}\left\{\Psi\left(z_E|\boldsymbol{y}^{(b)}\right)|F^{(1...b-1)^c}\right\},$$

where $\mathbb{E}$ denotes the expectation operator.

Now by Lemma 4 for each $z_E \neq z^{(b)}$, $z_E \in \left\{1, \ldots, e^{n\hat{R}}\right\}$

$$\mathbb{E}\left\{\Psi\left(z_E|\boldsymbol{Y}^{(b)}\right)|F^{(1...b-1)^c}\right\} \leq e^{-n\left(I(\hat{Y}_1;X_1Y|UX_2)-\epsilon\right)}.$$

Thus,

$$\mathbb{E}\left\{\left\|\mathcal{L}_E\left(\boldsymbol{y}^{(b)}\right)\right\| |F^{(1...b-1)^c}\right\}$$
$$\leq 1 + \left(e^{n\hat{R}} - 1\right)\left(e^{-n\left(I(\hat{Y}_1;X_1Y|UX_2)-\epsilon\right)}\right).$$

Furthermore conditioning on the event $E_8^{(c)}$ implies that the sender has decoded $s^{(b-1)}$ correctly and thus the sender only declares that $z_E^{(b-2)}$ has been transmitted by the relay in block $b-1$ if $s(z_E^{(b-2)}) = s^{(b-1)}$. Hence, we can upper bound the probability of the event $E_9^{(b)}$ conditioned on that there was decoding error in the previous blocks as follows:

$$\Pr\left(E_9^{(b)}\big|E_8^{(b)^c}, E_0^{(b)^c}, F^{(1...b-1)^c}\right)$$
$$\leq \Pr\left(E_9^{(b)}\big|F^{(1...b-1)^c}\right)$$
$$\leq \mathbb{E}\left\{\sum_{z_E \neq z^{(b)}} \Pr\left(z_E \in \mathcal{L}_E\left(\boldsymbol{Y}\right) \cap s(z_E) = s^{(b+1)}\big|\right.\right.$$
$$\left.\left. E_8^{(b)^c}, F^{(1...b-1)^c}\right)\right\}$$
$$\leq \mathbb{E}\left\{\left\|\mathcal{L}_E\left(\boldsymbol{y}^{(b)}\right)\right\| e^{-nR_0}|F^{(1,...,b-1)^c}\right\}$$
$$\leq e^{-nR_0}\left(1 + e^{n\left(\hat{R}-I(\hat{Y}_1;X_1Y|UX_2)+\epsilon\right)}\right).$$

Apparently for $\hat{R} < R_0 + I(\hat{Y}_1; X_1Y|UX_2) - \epsilon_1$ and sufficiently large n, the claim (41) follows.

REFERENCES

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

[2] S. I. Bross, "A converse result for the discrete memoryless relay channel with relay-transmitter feedback", *IEEE Trans. Inform. Theory,* vol. IT-52, no. 8, pp. 3789-3792, Sep. 2006.

[3] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory,* vol. IT-25, no. 5, pp. 572-584, Sep. 1979.

[4] T. M. Cover and J. A. Thomas, "*Elements of Information Theory,*" Wiley, 1991.

[5] A. El Gamal and M. Aref, "The capacity of the semi-deterministic relay channel," *IEEE Trans. Inform. Theory,* vol. IT-28, no. 3, p. 536, May 1982.

[6] Y. Gabbai and S. I. Bross, "Achievable rates for the discrete memoryless relay channel with partial feedback configurations," *IEEE Trans. Inform. Theory,* vol. IT-52, no. 11, pp. 4989-5007, Nov. 2006.

[7] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inform. Theory,* vol. IT-21, no. 6, pp. 294-300, May 1975.

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

[9] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Inform. Contr,* vol. 38, pp. 60-80, July 1978.

[10] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inform. Theory,* vol. IT-43, no. 6, pp. 1912-1923, Nov. 1997.

[11] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inform. Theory,* vol. IT-21, no. 6, pp. 629-637, Nov. 1975.

[12] T. Berger, "Multiterminal source coding." Lecture notes presented at the 1977 CISM Summer School, Udine, Italy, July 18-20, 1977, Springer-Verlag.

[13] H. F. Chong, M. Motani and H. K. Garg, "Generalized backward decoding strategies for the relay channel," *IEEE Trans. Inform. Theory,* vol. IT-53, no. 1, pp. 394-401, Jan. 2007.

[14] 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, April 2006.

[15] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–I: No bandwidth constraint, " *IEEE Trans. Inform. Theory,* vol. IT-12, no. 2, pp. 172-182, Apr. 1966.

[16] L. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Inform. Theory,* vol. IT-30, no. 4, pp. 623-629, Jul. 1984.

[17] L. Ozarow, "An achievable region and outer bound for the Gaussian broadcast channel with feedback", *IEEE Trans. Inform. Theory,* vol. IT-30, no. 4, pp. 667-671, Jul. 1984.

[18] G. Kramer, "Feedback strategies for white Gaussian interference networks", *IEEE Trans. Inform. Theory,* vol. IT-48, no. 6, pp. 1423-1438, Jun. 2002.

[19] N. Merhav and T. Weissman, "Coding for the feedback Gel'fand-Pinsker channel and the feedforward Wyner-Ziv source", *IEEE Trans. Inform. Theory,* vol. IT-52, no. 9, pp. 4207-4211, Sep. 2006.

[20] F.M.J. Willems and E.C. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders", *IEEE Trans. Inform. Theory,* vol. IT-31, no. 3, pp. 313-327, May 1985.

[21] F.M.J. Willems and E.C. van der Meulen, "Partial feedback for the discrete memoryless multiple-access channel", *IEEE Trans. Inform. Theory,* vol. IT-29, no. 2, pp. 287-290, Mar. 1983.

[22] S. Butman, "A general formulation of linear feedback communication systems with solutions", *IEEE Trans. Inform. Theory,* vol. IT-15, no. 3, pp.392-400, May 1969.

[23] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel", *IEEE Trans. Inform. Theory,* vol. IT-42, no. 2, pp. 364 – 375, Mar 1996.

[24] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering.* John Wiley & Sons, 1965.