On the Capacity of the Discrete Memoryless Broadcast Channel with Feedback

Ofer Shayevitz and Michèle Wigger

Abstract—A coding scheme for the discrete memoryless broadcast channel with noiseless, noisy, or generalized feedback is proposed, and the associated achievable region derived. The scheme is based on a block-Markov strategy combining the Marton scheme and a lossy version of the Gray-Wyner scheme with side-information. In each block the transmitter sends fresh data and update information that allows the receivers to improve the channel outputs observed in the previous block. For a generalization of Dueck’s broadcast channel our scheme achieves the noiseless-feedback capacity, which is strictly larger than the no-feedback capacity. For a generalization of Blackwell’s channel and when the feedback is noiseless our new scheme achieves rate points that are outside the no-feedback capacity region. It follows by a simple continuity argument that for both these channels and when the feedback noise is sufficiently low, our scheme improves on the no-feedback capacity even when the feedback is noisy.

I. INTRODUCTION

We consider a broadcast channel (BC) with two receivers, where the transmitter has instantaneous access to a feedback signal. Popular examples of such feedback signals are:

- the channel outputs observed at the two receivers (this setup is called noiseless feedback); or
- a noisy version of these channel outputs (this setup is called noisy feedback).

Here we allow for very general feedback signals, and only require that the time-$t$ feedback signal is obtained by feeding the time-$t$ input and the corresponding time-$t$ outputs into a memoryless feedback channel. This general form of feedback is commonly referred to as generalized feedback [1], [2], [3]. For brevity, here we mostly omit the word generalized. It is easily seen that our setup includes noiseless feedback and noisy feedback as special cases.

We focus on discrete memoryless broadcast channels (DMBCs), namely where the input and output symbols are from finite alphabets and the current channel outputs depend on the past inputs and outputs only through the current input. Our interest lies in the feedback-capacity region of such DMBCs, i.e., in the associated set of rate tuples for which reliable communication is possible.

Most previous results on DMBCs with feedback focus on the case of noiseless feedback. For example, El Gamal [4] proved that when the BC is physically degraded, i.e., one of the two outputs is obtained by processing the other output, then the capacity region with noiseless-feedback coincides with the no-feedback capacity region. In contrast, Dueck [5] and Kramer [6] described some specific examples of DMBCs where the noiseless-feedback capacity region exceeds the no-feedback capacity region. In Dueck’s example, the noiseless-feedback capacity region is known. However, outside these specific examples, determining the capacity region with feedback for (non-physically-degraded) DMBCs is an open problem. In fact, even characterizing the class of DMBCs where feedback enlarges the capacity region seems hard. This is partly because even the no-feedback capacity region is generally unknown, and partly because a computable single-letter achievable region for the DMBC with feedback was missing hitherto. Kramer [6] proposed a multi-letter achievable region for the DMBC with noisy or noiseless feedback.

In this paper we propose a coding scheme for the DMBC with generalized feedback, and present a corresponding single-letter achievable region. Subsequently, we analyze two new examples—a generalization of Dueck’s channel [5] and a noisy version of Blackwell’s channel [7]—where our region is shown to exceed the no-feedback capacity region, even in the presence of feedback noise. Our approach is motivated by Dueck’s example [5], and is based on the following idea. The transmitter uses the feedback to identify update information that is useful to the receivers when decoding their intended messages, and describes this information in subsequent transmissions. More specifically, our scheme adopts a block-Markov strategy, where in each block the transmitter sends a combination of fresh data and compressed update information pertaining to the data sent in the previous block. Marton’s no-feedback scheme [10], [11] is used in each block to send the fresh data and the update information, at rates outside the no-feedback capacity region. The update information sent in a block is essentially an efficient lossy description of the auxiliary inputs in Marton’s scheme from the previous block, taking into account the receivers’ observations and the feedback signal as side-information. The receivers perform backward decoding; starting with the last block, each receiver iteratively performs the following two steps: 1) it decodes its intended data and update information in the current block; and 2) it uses the update information to “improve” the channel outputs in the preceding block, which is processed next. This strategy is gainful whenever the cost of the lossy description (i.e., the rate needed to send the update information) is smaller than the increase in rate it supports (i.e., the increase in capacity of the “improved” channel). Intuitively, this is expected to happen when the descriptions required by the two receivers have a
large common part.

Our scheme has some ideas in common with Lapidoth and Steinberg’s scheme for the MAC with strictly causal state-information at the transmitter [12], [13].

Recently, another single-letter achievable region for general DMBCs with feedback has been proposed [16]. Comparing the achievable region in [16] to ours however seems difficult.

The paper is organized as follows. In Section II, the necessary mathematical background is provided. The channel model is described in Section III. In Section IV, Marton’s scheme for the DMBC without feedback is reviewed in detail. In Section V, a lossy version with side-information of the Gray-Wyner distributed source coding setup is introduced, and an achievable region is obtained. The main result of the paper is introduced in Section VI, where the Marton and the lossy Gray-Wyner schemes are combined into a feedback scheme for general DMBCs, and the associated achievable region is derived. Two new examples are discussed in VII.

A generalization of Dueck’s DMBC, and a noisy version of Blackwell’s DMBC [7]. In both cases, the region achieved by the new scheme is shown to exceed the no-feedback capacity region, using either noiseless feedback or noisy feedback, in the limit of low feedback noise.

II. Preliminaries

A. Notations

We broadly follow the notation in [17]. In particular, for any real number $M > 1$, we use the notation $[M] = \{1, \ldots, [M]\}$. The set of positive integers is denoted by $\mathbb{Z}^+$. Also, we use upper case symbols to denote random variables, e.g., $A$, and lower case symbols for their realizations, e.g., $a$. The corresponding alphabets are denoted by script symbols, e.g., $\mathcal{A}$; and $|\mathcal{A}|$ is used for the cardinality of $\mathcal{A}$. For $n \in \mathbb{Z}^+$ we use $A^n$ and $a^n$ to denote the random sequence $A_1, \ldots, A_n$ and its realization $a_1, \ldots, a_n$.

We think of a product set of the form $[2^{nr_1}] \times [2^{nr_2}]$ as being one-to-one with $[2^{nr_1 + r_2}]$, disregarding the associated integer issues throughout. This assumption does not influence our results, as they concern the asymptotic regime $n \to \infty$. For $\epsilon > 0$, we write $\delta(\epsilon)$ to indicate a general nonnegative function satisfying $\delta(\epsilon) \to 0$ (arbitrarily slow) as $\epsilon \to 0$.

A random sequence $X^n$ is said to be $P_X$-independent-identically-distributed ($P_X$-i.i.d.) if

$$P_{X^n}(x^n) = \prod_{t=1}^{n} P_X(x_t)$$

for all $x^n$. Let $(X^n, Y^n)$ be two jointly distributed random sequences, and let $P_{Y|X}$ be some conditional distribution. We say that $Y^n$ is $P_{Y|X}$-independent given $X^n$ if

$$P_{Y^n|X^n}(y^n|x^n) = \prod_{t=1}^{n} P_{Y|X}(y_t|x_t)$$

for all $y^n$ and $x^n$ with $P_{X^n}(x^n) > 0$.

We use the notion of typicality as defined in [17]. For a finite alphabet $\mathcal{X}$, a sequence $x^n \in \mathcal{X}^n$ is said to be $\epsilon$-typical with respect to (w.r.t.) a distribution $P_X$ on $\mathcal{X}$ if

$$|\pi_{x^n}(x) - P_X(x)| \leq \epsilon \cdot P_X(x)$$

for all $x \in \mathcal{X}$, where $\pi_{x^n}$ is the distribution over $\mathcal{X}$ corresponding to the relative frequency of symbols in $x^n$. The set of all such sequences is denoted $T^n_\epsilon(P_X)$. Similarly, for a law $P_{X_1, \ldots, X_k}$ over a product alphabet $\mathcal{X}_1 \times \cdots \times \mathcal{X}_k$, we denote by $T^n_\epsilon(P_{X_1, \ldots, X_k})$ the set of all $k$-tuples of sequences $(x^n_1 \in \mathcal{X}_1^n, \ldots, x^n_k \in \mathcal{X}_k^n)$ that are jointly $\epsilon$-typical w.r.t. $P_{X_1, \ldots, X_k}$.

Finally, we write $Z \sim \text{Bern}(p)$ for a a binary random variable taking the values 0 and 1 with probabilities $1 - p$ and $p$.

B. Basic Lemmas

The following three lemmas are well known, and used extensively in the sequel.

Lemma 1 (Conditional Typicality Lemma [17]). Let $P_{XY}$ be some joint distribution. Suppose $x^n \in T^n_\epsilon(P_X)$ for some $\epsilon' > 0$, and $Y^n$ is $P_{Y|X}$-independent given $X^n = x^n$. Then for every $\epsilon > \epsilon'$:

$$\lim_{n \to \infty} \Pr \left( (x^n, Y^n) \notin T^n_\epsilon(P_{XY}) \right) = 0.$$

Lemma 2 (Covering Lemma [17]). Let $0 < \epsilon < \epsilon'$, and let $X^n$ satisfy $\Pr(X^n \in T_\epsilon(P_X)) \to 1$ as $n \to \infty$. Also, for each $n$, let $M_n \in \mathbb{Z}^+$ be larger than $2^n r$ for some $r \geq 0$, and let $\{Y^n(m)\}_{m=1}^M$ be a set of $P_Y$-i.i.d. sequences such that $\{X^n, Y^n(m)\}_{m=1}^M$ are mutually independent. Then, for any law $P_{XY}$ with marginals $P_X$ and $P_Y$ there exists $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ such that

$$\lim_{n \to \infty} \Pr \left( \forall m \in [M], (X^n, Y^n(m)) \notin T^n_\epsilon(P_{XY}) \right) = 0$$

if $r > I(X; Y) + \delta(\epsilon)$.

Lemma 3 (Packing Lemma [17]). Let $\epsilon > 0$, and $X^n$ be an arbitrary random sequence. Also, for each $n$, let $M_n \in \mathbb{Z}^+$ be smaller than $2^n r$ for some $r \geq 0$, and let $\{Y^n(m)\}_{m=1}^M$ be a set of $P_Y$-i.i.d. random sequences, where each $Y^n(m)$ is independent of $X^n$. Then, for any law $P_{XY}$ with marginal $P_Y$ there exists $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ such that

$$\lim_{n \to \infty} \Pr \left( \exists m \in [M] \text{ s.t. } (X^n, Y^n(m)) \in T^n_\epsilon(P_{XY}) \right) = 0$$

if $r < I(X;Y) - \delta(\epsilon)$.

The following is a simple multivariate generalization of the packing lemma.

Lemma 4 (Multivariate Packing Lemma). Let $\epsilon > 0$, and for each $n$ let $M_{1,n}, M_{2,n}, M_{3,n} \in \mathbb{Z}^+$ satisfy $M_{i,n} \leq 2^n r_i$ for $i \in \{1,2,3\}$. Also, let $\{U^n_i(m)\}_{m=1}^{M_{i,n}}$ be a set of $P_{U_i}$-i.i.d. random vectors such that $\{U^n_1(m_1), U^n_2(m_2), U^n_3(m_3)\}$ are mutually independent for any $m_1, m_2, m_3$. Then, for any law
The result follows by taking the union bound over $E_{i,j,k}$, and requiring that it tends to zero.  

\section{III. CHANNEL MODEL}

We consider the discrete memoryless broadcast channel with generalized feedback in Figure 1. The goal of the communication is that the transmitter conveys a private Message $M_1$ to a Receiver 1, a private Message $M_2$ to a Receiver 2, and a common message $M_0$ to both receivers. The three messages $M_0, M_1, M_2$ are assumed to be independent and uniformly distributed over the finite sets $[2^{nR_0}], [2^{nR_1}], [2^{nR_2}]$, respectively, where $n$ denotes the blocklength and $R_0, R_1, R_2$ are the corresponding common and private transmission rates.

Communication takes place over a DMBC with generalized feedback. This channel is characterized by a quadruple of finite alphabets $X, Y_1, Y_2$, and $\hat{Y}$, and a conditional probability law $P_{Y_1,Y_2|X}(y_1,y_2,\hat{y}|x)$ where $x \in X$, $y_1 \in Y_1, y_2 \in Y_2$, and $\hat{y} \in \hat{Y}$. Given that at time $t$ the transmitter feeds the symbol $x_t$ to the channel, Receiver 1 and Receiver 2 observe the channel outputs $y_{1,t} \in Y_1$ and $y_{2,t} \in Y_2$, respectively, and the transmitter observes the generalized feedback $\hat{y}_t \in \hat{Y}$, with probability $P_{Y_1,Y_2,\hat{Y}|X}(y_{1,t},y_{2,t},\hat{y}_t|x_t)$.

Thanks to feedback, the transmitter can produce its time-$t$ channel input $X_t$ as a function of the Messages $M_0, M_1, M_2$ and of the previously observed feedback outputs $Y_{\hat{t}}^{t-1} \equiv (Y_1,\ldots,Y_{t-1})$:

$$X_t = \psi^{(n)}_t(M_0,M_1,M_2,Y_{\hat{t}}^{t-1}),$$

for some encoding function $\psi^{(n)}_t$, for $t \in \{1,\ldots,n\}$. The DMBC and its feedback channel are memoryless, which is captured by the following Markov relation for $t \in [n]$:

$$Y_{\hat{t}}^{t+1} \equiv (Y_{\hat{t}}^{t-1},X_t,Y_t)$$

where $Y_{\hat{t}}^{t-1} \equiv (Y_{\hat{t}}^{t+1},Y_{\hat{t}}^{t+1},\ldots,Y_{\hat{t}}^{t+1})$, for $t \in [1,2]$. 

After $n$ channel uses Receiver $i$ decodes its intended messages $M_0$ and $M_i$ for $i \in \{1,2\}$. Namely, Receiver $i$ produces the guess:

$$\hat{M}_{0,i} = \Psi^{(n)}_i(Y_{\hat{t}}^{n}), \quad i \in \{1,2\}$$

where $\Psi^{(n)}_i$ denotes Receiver $i$’s decoding function. 

A rate triplet $(R_0, R_1, R_2)$ is called achievable if for every blocklength $n$ there exists a set of encoding functions $\{\psi^{(n)}_t\}_{t=1}^n$ and two decoding functions $\Psi^{(n)}_1$ and $\Psi^{(n)}_2$ such that the probability of decoding error, i.e., the probability that

$$(M_0,M_1) \neq (\hat{M}_{0,1},\hat{M}_{1,1}) \quad \text{or} \quad (M_0,M_2) \neq (\hat{M}_{0,2},\hat{M}_{2,2}),$$

is bounded by $\epsilon$. The capacity-achieving rate $R = (R_0,R_1,R_2)$ is the feedback capacity-region of this setup, and we denote it by $C_{GenFB}$.

The described generalized-feedback setup includes as special cases the no-feedback setup where the feedback outputs are deterministic, e.g., $|\hat{Y}| = 1$; the noiseless-feedback setup where the feedback output coincides with the pair of channel outputs, i.e., $\hat{Y} = (Y_1,Y_2)$ (see Figure 2); and the noisy-feedback setup where the feedback outputs and the channel inputs and outputs satisfy the Markov relation $X_t \rightarrow (Y_{1,t},Y_{2,t}) \rightarrow \hat{Y}_t$ for all $t \in [n]$ (e.g., the setup in Figure 3). In these special cases, we denote the capacity regions by $C_{NoFB}, C_{NoiselessFB}$, and $C_{NoisyFB}$, respectively.

\section{IV. MARTON’S NO-FEEDBACK SCHEME}

We review Marton’s achievable region with a common message $[10], [11], [17]$ and the coding scheme achieving this region. Redescribing the scheme simplifies the description of our feedback scheme in Section VI-B.
A. Marton’s Achievable Region

Let $R_{\text{Marton}}$ be the closure of the set of all nonnegative rate triplets $(R_0, R_1, R_2)$ that for some choice of random variables $U_0, U_1, U_2$ over finite alphabets $\mathcal{U}_0, \mathcal{U}_1, \mathcal{U}_2$ and some function $f : \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \rightarrow \mathcal{X}$ satisfy

\[
R_0 + R_1 < I(U_0; U_1; Y_1) \tag{4a}
\]

\[
R_0 + R_2 < I(U_0; U_2; Y_2) \tag{4b}
\]

\[
R_0 + R_1 + R_2 < I(U_1; Y_1 | U_0) + I(U_2; Y_2 | U_0) + \min_i I(U_i; Y_1 | U_0) - I(U_1; U_2 | U_0) \tag{4c}
\]

\[
2R_0 + R_1 + R_2 < I(U_0; U_1; Y_1) + I(U_0; U_2; Y_2) - I(U_1; U_2 | U_0) \tag{4d}
\]

where $X = f(U_0, U_1, U_2)$,

$$(U_0, U_1, U_2) \rightarrow X \rightarrow (Y_1, Y_2)$$

forms a Markov chain, and $(Y_1, Y_2) \sim P_{Y_1,Y_2|X}$ given $X$.

Theorem 1 (From [10], [11]). $R_{\text{Marton}} \subseteq \mathcal{C}_{\text{NoFB}}$.

B. Marton’s Scheme

We describe the scheme for a DMBC $(X, Y_1, Y_2, P_{Y_1,Y_2|X})$. The scheme has parameters $(\mathcal{U}_0, \mathcal{U}_1, \mathcal{U}_2, P_{U_0,U_1,U_2}, f, R_0, R_1, p, R_{1,q}, R_2, R_{2,e}, R_1', R_2', \epsilon, n)$ where

- $\mathcal{U}_0, \mathcal{U}_1, \mathcal{U}_2$ are auxiliary finite alphabets;
- $P_{U_0,U_1,U_2}$ is a joint law over these auxiliary alphabets;
- $f : \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \rightarrow \mathcal{X}$ is a function mapping the auxiliary inputs into effective inputs;
- $R_0, R_1, p, R_{1,q}, R_2, R_{2,e}, R_1', R_2'$ are nonnegative communication rates where $R_1 = R_{1,q} + R_{1,e}$ and $R_2 = R_{2,e} + R_2'$;
- $R_1', R_2'$ are nonnegative binning rates;
- $\epsilon > 0$ is a small number; and
- $n$ denotes the scheme’s blocklength.

1) Code Construction: Define $R_c \overset{\text{def}}{=} R_0 + R_1 + R_{1,e} + R_2 + R_{2,e}$. The code consists of a single codebook $C_0 : \left\{ \mathcal{C}_1(m_c) \right\}_{m_c=1}^{[2^{n R_c}]}$, and of $[2^{n R_c}]]$ codebooks $\left\{ \mathcal{C}_2(m_c) \right\}_{m_c=1}^{[2^{n R_c}]}$. Codebook $C_0$ consists of $[2^{n R_c}]$ length-$n$ codewords $\left\{ u_0^n(m_c) \right\}_{m_c=1}^{[2^{n R_c}]}$ whose entries are randomly and independently drawn according to $P_{U_0}$. For $i = 1, 2$ and $m_c \in [2^{n R_c}]$, Codebook $\mathcal{C}_i(m_c)$ consists of $[2^{n R_i}]$ codewords where each bin $m_i \in [2^{n R_i}]$ contains $[2^{n R_i}]$ length-$n$ codewords $\left\{ u_i^n(m_c, m_i, \ell_i) \right\}_{\ell_i=1}^{[2^{n R_i}]}$ that are randomly drawn $P_{U_i|m_i}$-independent given $u_0^n(m_c)$.

Reveal all codebooks to the transmitter and codebooks $C_0$ and $\left\{ \mathcal{C}_i(m) \right\}_{m=1}^{[2^{n R_c}]}$ to Receiver $i \in \{1, 2\}$.

2) Encoding: The encoder parses both private messages $M_1 \in [2^{n R_1}]$ and $M_2 \in [2^{n R_2}]$ into pairs of independent submessages $(M_{1,1}, M_{1,2}) \in [2^{n R_1}] \times [2^{n R_1}]$ and $(M_{2,1}, M_{2,2}) \in [2^{n R_2}] \times [2^{n R_2}]$, and forms the new common message $M_c \equiv (M_0, M_{1,1}, M_{2,2})$ of rate $R_c$.

Now, given that $M_c = m_c$, $M_{1,1} = m_{1,1}$, $M_{2,2} = m_{2,2}$, the encoder makes a list of all pairs $(\ell_1, \ell_2)$ such that

\[
(\ell_1, \ell_2) \in \mathcal{T}_{\epsilon/2}^{(n)}(P_{U_0,U_1,U_2}), \tag{5}
\]

and chooses one pair from this list at random. We call the chosen pair $(\ell_1, \ell_2)$. If the list is empty, it chooses $(\ell_1, \ell_2)$ randomly from the set of all indices $[2^{n R_1}] \times [2^{n R_2}]$.

The inputs $x^n$ are obtained from the codewords $u_0^n(m_c)$, $u_1^n(m_c, m_{1,1}, \ell_1)$, $u_2^n(m_c, m_{2,2}, \ell_2)$ by applying the function $f$ componentwise to these three sequences:

\[
x_j = f(u_{0,j}(m_c), u_{1,j}(m_c, m_{1,1, \ell_1}), u_{2,j}(m_c, m_{2,2}, \ell_2)), \quad j \in [n].
\]

3) Decoding: Given that Receiver 1 observes the sequence $y_1^n$, it forms a list of all the tuples $(\hat{m}_c, \hat{m}_{1,1}, \ell_1)$ that satisfy

\[
(x_0^n, u_0^n(m_c), u_1^n(\hat{m}_c, \hat{m}_{1,1}, \hat{\ell}_1), y_1^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{U_0,U_1,Y_1}). \tag{6}
\]

It randomly chooses a tuple $(\hat{m}_c, \hat{m}_{1,1}, \ell_1)$ from this list (if the list is empty, it randomly chooses a pair $(\hat{m}_c, \hat{m}_{1,1})$ from $[2^{n R_1}] \times [2^{n R_2}]$) and parses $m_c$ as $(\hat{m}_{0,1,\hat{m}_{1,1}}, \hat{M}_{2,1,\ell_1})$. It finally produces $\hat{m}_{0,1}$ as its guess of message $M_0$ and $\hat{m}_{1,1} = (\hat{m}_{c,\hat{m}_{1,1}}, \hat{m}_{1,1})$ as its guess of $M_1$.

Receiver 2 produces its guesses $\hat{m}_{0,2}$ and $\hat{m}_{2}$ of the messages $M_0$ and $M_2$ in a similar way.

4) Analysis: See Appendix A

V. LOSSY GRAY-WYNER CODING WITH SIDE-INFORMATION (LGW-SI)

In this section we study a distributed source-coding problem and present an achievable region for this problem. The associated scheme will be used as part of our construction for the DMBC with feedback in Section VII.

Our source coding problem is depicted in Figure 4. Unlike in classical rate-distortion problems where the decoders have to produce sequences that satisfy certain average per-symbol distortion constraints, here, we require that the sequences produced at the decoders are almost jointly-typical with the source sequence. Thus, our problem is a coordination capacity problem [21].

The rate-distortion problem corresponding to our setup is a lossy version of the Gray-Wyner distributed source-coding problem in [19] with additional side-information at the decoders. Our achievable region directly leads to an achievable region for this rate-distortion problem, see [22]. Special cases of this rate-distortion problem have been considered by Heegard and Berger [20], Tian and Diggavi [23], and Steinberg and Merhav [24], and the lossless counterpart by Timo et al. [25], [26].
A. Setup and Achievable Region

Our setup is parameterized by the tuple \((X, Y_1, Y_2, V_1, V_2, P_{XY,Y_1}, P_{V_i|X}, P_{V_2|X}, n)\), where

- \(X, Y_1, Y_2, V_1, V_2\) are discrete finite alphabets;
- \(P_{XY,Y_1}\) is a joint probability distribution over the alphabet \(X \times Y_1 \times Y_2\);
- \(P_{V_i|X}\) and \(P_{V_2|X}\) are conditional probability distributions over \(V_1\) and \(V_2\) given some random variable \(X \in X\);
- \(n\) is the blocklength.

In the following let \(\{(X_i,Y_{1,i},Y_{2,i})(i=1)^n\}\) be an i.i.d. sequence of triplets of discrete random variables, with marginal distribution \(P_{XY,Y_2}\). Consider a distributed source coding setting where a sender observes the source sequence \(X^n\), Receiver 1 observes the side-information \(Y_1^n\), and Receiver 2 observes the side-information \(Y_2^n\). It is assumed that the sender can noiselessly send three rate-limited messages \(K_0, K_1, K_2\) to the receivers: a common message \(K_0\) to both receivers, a private message \(K_1\) to Receiver 1 only, and another private message \(K_2\) to Receiver 2 only. More precisely, the encoding procedure is described by an encoding function \(\lambda^{(n)}: X^n \rightarrow \{2^{nR_0}\} \times \{2^{nR_1}\} \times \{2^{nR_2}\}\), which for a sequence \(X^n\) produces the messages \((K_0, K_1, K_2) = \lambda^{(n)}(X^n)\). Each Receiver \(i\), for \(i \in \{1, 2\}\), produces a reconstruction sequence \(\hat{V}_i^n = \lambda^{(n)}(K_0, K_i, Y_i^n)\) by applying a reconstruction function \(\lambda^{(n)}_j: \{2^{nR_0}\} \times \{2^{nR_1}\} \times \{2^{nR_2}\} \rightarrow \{2^{nR_j}\}\) to the messages \(K_0\) and \(K_i\) and the side-information \(Y_i^n\). The goal of the communication is that for each \(i \in \{1, 2\}\), the reconstruction sequence \(\hat{V}_i^n\) is jointly typical with the source sequence \(X^n\) according to \(P_X \times P_{V_i|X}\).

A rate triplet \((R_0, R_1, R_2)\) is said to be \(\epsilon\)-achievable if there exists a sequence of encoding and reconstruction functions \((\lambda^{(n)}_1, \lambda^{(n)}_2, \lambda^{(n)}_3)\) such that:

\[
\Pr\left( (X^n, \hat{V}_i^n) \notin T^n(P_{XY,Y_i}) \right) \to 0
\]
as \(n \to \infty\), for \(i \in \{1, 2\}\). A triplet is said to be achievable if it is \(\epsilon\)-achievable for all \(\epsilon > 0\). The closure of the set of all achievable rate triplets is denoted \(R_{LGW}\).

Let \(R_{LGW}^{n_{iner}}\) be the closure of the set of all nonnegative rate triplets \((R_0, R_1, R_2)\) satisfying

\[
\begin{align}
R_0 + R_1 &> I(X; V_0, V_1 | Y_1) \\
R_0 + R_2 &> I(X; V_0, V_2 | Y_2) \\
R_0 + R_1 + R_2 &> I(X; V_1 | Y_1, V_2) + I(X; V_2 | Y_2, V_0) + \max_{i \in \{1, 2\}} I(X; V_i | Y_i)
\end{align}
\]

for some choice of the random variable \(V_0\) such that

\[
(V_0, V_1, V_2) \rightarrow X \rightarrow (Y_1, Y_2).
\]

Theorem 2. \(R_{LGW}^{n_{iner}} \subseteq R_{LGW}\). Furthermore, \(R_{LGW}^{n_{iner}}\) is convex.

Proof. Inclusion \(R_{LGW}^{n_{iner}} \subseteq R_{LGW}\) is established in Section V-B. The convexity of \(R_{LGW}^{n_{iner}}\) is proved in Appendix C.

B. Scheme

In this section we describe a scheme achieving the region \(R_{LGW}^{n_{iner}}\). Our scheme is similar to Heegard and Berger’s scheme for the Wyner-Ziv setup with several, differently informed receivers [20]. However, our scheme also uses the double-binning technique for the common codebook proposed in [23], but where here the double-binning is performed in two different ways, one way that is relevant for Receiver 1 and the other way relevant for Receiver 2. This is beneficial when the quality of the side-information at the two receivers is very different.

The scheme we propose has parameters \(V_0, P_{V_0V_iV_j|X}\), \(R_{0,0}, R_{0,1}, R_{0,2}, R_{1,0}, R_{1,1}, R_{1,2}, R_{2,0}, R_{2,1}, R_{2,2}, R'_0, R'_1, R'_2, c, n\), where

- \(V_0\) is an auxiliary alphabet;
- \(P_{V_0V_iV_j|X}\) is a conditional joint probability distribution over \(V_0 \times V_i \times V_j\) given some \(X \in X\) such that its marginals satisfy \(\sum_{v_0,v_1,v_2} P_{V_0V_iV_j|X}(v_0,v_1,v_2|x) = P_{V_i|X}(v_1|x)\) and \(\sum_{v_0,v_1} P_{V_0V_i|X}(v_0,v_1|v_2|x) = P_{V_2|X}(v_2|x)\);
- \(R_{0,0}, R_{0,1}, R_{0,2}, R_{1,0}, R_{1,1}, R_{1,2}, R_{2,0}, R_{2,1}, R_{2,2} \geq 0\) are nonnegative communication rates;
- \(R'_0, R'_1, R'_2 \geq 0\) are nonnegative binning rates, where \(R'_0\) cannot be smaller than \(\max\{R_0, R_1, R_2\}\); and
- \(c > 0\) is a small number; and
- \(n\) is the scheme’s blocklength.

1) Codebook Generation: Generate three codebooks \(C_0, C_1, C_2\) independently of each other in the following way.

Codebook \(C_0\) consists of \([2^{nR_0}]\) superbins, each containing \([2^{nR_0}]\) length-\(n\) codewords whose entries are randomly and independently generated according to the law \(P_{V_0}\). We
make two partitions of the codewords in each superbin, see Figure 5. In the first partition the codewords of each superbin are assigned to \([2^{nR_i}]\) subbins, each containing \([2^{nR_{0_i}-R_{1_i}}]\) codewords; in the second partition they are assigned to \([2^{nR_{0_i}}]\) subbins, each containing \([2^{nR_{0_i}}]\) codewords. There are thus two different ways to refer to a specific codeword in \(C_0\). When we consider the first partition, we denote the codewords in the \(k_{1,0} \in [2^{nR_{0_i}}]\)-th subbin of superbin \(k_{0,0} \in [2^{nR_{0_i}}]\) by

\[
\{v^0_n(1; k_{0,0}, k_{1,0}, \ell_{1,0})\}_{\ell_{1,0}=1}^{2^{n(R_{0_i}-R_{1_i})}};
\]

when we consider the second partition, we denote the codewords in the \(k_{2,0} \in [2^{nR_{2_i}}]\)-th subbin of superbin \(k_{0,0} \in [2^{nR_{0_i}}]\) by

\[
\{v^0_n(2; k_{0,0}, k_{2,0}, \ell_{2,0})\}_{\ell_{2,0}=1}^{2^{n(R_{0_i}-R_{2_i})}}.
\]

Thus, here the first index indicates whether the last two indices refer to the first or the second partition of the superbins.

For \(i \in \{1, 2\}\), Codebook \(C_i\) consists of \([2^{nR_{0_i}}]\) superbins each containing \([2^{nR_i}]\) subbins with \([2^{nR_i}]\) codewords of length \(n\), where all entries of all codewords are randomly and independently drawn according to \(P_{\tilde{X}_i}\). For \(k_{i,i} \in [2^{nR_i}]\), we denote the codewords in the \(k_{i,i}\)-th subbin of superbin \(k_{0,i} \in [2^{nR_{0_i}}]\) by

\[
\{v^0_n (k_{0,i}, k_{i,i}, \ell_{i,i})\}_{\ell_{i,i}=1}^{2^{nR_i}}.
\]

All codebooks are revealed to the sender, and codebooks \(\{C_0, C_i\}\) are revealed to Receiver \(i \in \{1, 2\}\).

2) LGW-SI Encoder: Given that the encoder observes the source sequence \(X^n = x^n\), it searches the codebooks \(C_0, C_1, C_2\) for a triplet of codewords \(v^0_n (k_{0,0}, k_{1,0}, \ell_{1,0}) \in C_0, v^0_n (k_{0,1,1}, \ell_{1,1}) \in C_1, v^0_n (k_{2,2,2}, \ell_{2,2}) \in C_2\) such that for \(i \in \{1, 2\}\):

\[
(X^n, v^0_n (1; k_{0,0}, k_{1,0}, \ell_{1,0}), v^0_n(k_{0,i}, k_{i,i}, \ell_{i,i})) \in T^0_{\epsilon/2}(P_{X^n Y^n V^n}).
\]

(9)

It then forms a list of all tuples of indices \((k_{0,0}, k_{1,0}, \ell_{1,0}, k_{1,1}, \ell_{1,1}, k_{2,2}, \ell_{2,2})\) satisfying (9). If the list is non-empty, the sender chooses one tuple from this list at random; otherwise, it randomly chooses a tuple \((k_{0,0}, k_{1,0}, \ell_{1,0}, k_{0,1,1}, \ell_{1,1}, k_{2,2}, \ell_{2,2})\) from the set \([2^{nR_{0_i}}] \times [2^{nR_{1_i}}] \times [2^{nR_{0_i}}] \times [2^{nR_{1_i}}] \times [2^{nR_{0_i}}] \times [2^{nR_{2_i}}] \times [2^{nR_{2_i}}]\). We denote the chosen indices by \(k^*_0, k^*_{1,0}, k^*_{0,1,1}, \ell^*_1, k^*_{1,1}, \ell^*_1, k^*_{2,2}, \ell^*_2\).

3) LGW-SI Decoder: Receiver \(i \in \{1, 2\}\) first parses the common message \(K_0 = (k^*_0, k^*_{1,0}, k^*_{1,1}, k^*_{2,2})\) to both receivers, the product message \(K_1 = (k^*_0, k^*_1)\) to Receiver 1 only, and the product message \(K_2 = (k^*_0, k^*_{2,2})\) to Receiver 2 only.

3) LGW-SI Decoder: Receiver \(i \in \{1, 2\}\) first parses the common message \(K_0 = (k^*_0, k^*_{1,0}, k^*_{1,1}, k^*_{2,2})\) to both receivers, the product message \(K_1 = (k^*_0, k^*_1)\) to Receiver 1 only, and the product message \(K_2 = (k^*_0, k^*_{2,2})\) to Receiver 2 only.

3) LGW-SI Decoder: Receiver \(i \in \{1, 2\}\) first parses the common message \(K_0 = (k^*_0, k^*_{1,0}, k^*_{1,1}, k^*_{2,2})\) to both receivers, the product message \(K_1 = (k^*_0, k^*_1)\) to Receiver 1 only, and the product message \(K_2 = (k^*_0, k^*_{2,2})\) to Receiver 2 only.
\[ R_0 + R_1 \leq I(U_0, U_1; Y_1, V_1) - I(U_0, U_1, U_2, \tilde{Y}; V_0, V_1|Y_1) \quad (12a) \]
\[ R_0 + R_2 \leq I(U_0, U_2; Y_2, V_2) - I(U_0, U_1, U_2, \tilde{Y}; V_0, V_2|Y_2) \quad (12b) \]
\[ R_0 + R_1 + R_2 \leq I(U_1, U_2, V_1|U_0) + I(U_2; Y_2, V_2|U_0) + \min_{i \in \{1, 2\}} I(U_0; Y_i, V_i) - I(U_1; U_2|U_0) \]
\[ - I(U_0, U_1, U_2, \tilde{Y}; V_1|V_0, Y_1) - I(U_0, U_1, U_2, \tilde{Y}; V_2|V_0, Y_2) - \max_{i \in \{1, 2\}} I(U_0, U_1, U_2, \tilde{Y}; V_0|Y_i) \quad (12c) \]
\[ 2R_0 + R_1 + R_2 \leq I(U_1U_0; Y_1, V_1) + I(U_2; U_0, Y_2, V_2) - I(U_1; U_2|U_0) \]
\[ - I(U_0, U_1, U_2, \tilde{Y}; V_0, V_1|Y_1) - I(U_0, U_1, U_2, \tilde{Y}; V_0, V_2|Y_2) \quad (12d) \]

**Remark 4.** By the previous remark, the following conclusion holds for any DMBC \( P_{Y_1Y_2|X} \) with feedback alphabet \( \bar{Y} = Y_1 \times Y_2 \). Assume that the region \( \mathcal{R}_{inner} \) associated with noiseless feedback (i.e., \( \bar{Y} = (Y_1, Y_2) \)) strictly contains \( \mathcal{C}_{NoFB} \). Now, if we consider a noisy feedback channel \( P_{\bar{Y}|X/Y_1Y_2} \) that is close enough to the noiseless feedback (i.e., \( \bar{Y} \) close to \( (Y_1, Y_2) \)), then also the region \( \mathcal{R}_{inner} \) associated with this noisy feedback strictly contains \( \mathcal{C}_{NoFB} \).

**B. Scheme achieving \( \mathcal{R}_{inner} \)**

Our scheme combines Marton’s no-feedback scheme of Section [LV-B] with our LGW-SI scheme of Section [V-B] using a block-Markov framework. We first present the high-level idea of the scheme, which is also depicted in Figure 6. Transmission takes place over \( B + 1 \) consecutive blocks, where the first \( B \) blocks are of length \( n \) each, and the last block is of length \( \gamma n \) for \( \gamma > 1 \). We denote the input/output/feedback sequences in Block \( b \in [B] \) by \( X_b^n, Y_b^n, \bar{Y}_b \), respectively, and the input/output sequences in Block \( B + 1 \) by \( X_{(B+1)}^n, \bar{Y}_{(B+1)}^n \). The messages to be sent are in a product form \( M_i = (M_{i,(1)}, \ldots, M_{i,(B)}) \), for \( i \in \{0, 1, 2\} \), where each \( M_{i,(b)} \) is uniformly distributed over the set \( [2^n \mathcal{R}_b] \). The effective rates of transmission are thus

\[ \left( \begin{array}{c}
\frac{B}{B + \gamma R_0}, \\
\frac{B}{B + \gamma R_1}, \\
\frac{B}{B + \gamma R_2}
\end{array} \right) \quad (13) \]

and approach \( (R_0, R_1, R_2) \) as the number of blocks \( B \to \infty \).

In each block \( b \) the transmitter uses Marton’s no-feedback scheme to send the Messages \( M_{0,(b)}, M_{1,(b)}, M_{2,(b)} \) together with update information \( K_{0,(b-1)}, K_{1,(b-1)}, K_{2,(b-1)} \) pertaining to the messages sent in the previous block. An exception is the first (resp. last) block where only the message tuple (resp. update information) is sent. The update information is constructed in a way that when \( (K_{0,(b)}, K_{i,(b)}) \) is available at Receiver \( i \), the latter can use it to “improve” its block-\( b \) observations \( \bar{Y}_{i,(b)}^n \). This facilitates decoding of the corresponding messages \( M_{0,(b)}, M_{1,(b)}, M_{2,(b)} \), which otherwise might not have been possible to decode reliably. The update information is generated via the LGW-code described in Section [V-B].

The code is designed for an LGW-setup where the encoder’s “source sequence” consists of the auxiliary Marton-codewords and the feedback signal, and where the receivers’ “side-information” consists of their respective channel outputs.

Each Receiver \( i \), for \( i \in \{1, 2\} \), performs backward decoding. It starts from the last block and decodes the update information \( (K_{0,(b)}, K_{i,(b)}) \) based on \( \bar{Y}_{i,(b+1)}^n \). Denote its guess by \( \tilde{K}_{0,(b)}, \tilde{K}_{i,(b)} \). Then, for each block \( b \in [B] \), starting from block \( B \) and going backwards, it performs the following steps:

1) Using \( (\tilde{K}_{0,b}, \tilde{K}_{1,b}) \), it “improves” its block-\( b \) outputs \( \bar{Y}_{(b)}^n \).

2) Based on these “improved” outputs, it then decodes the data \( (M_{0,(b)}, M_{i,(b)}) \) and the update information \( (K_{0,(b-1)}, K_{i,(b-1)}) \). We denote the corresponding guesses by \( (\tilde{M}_{0,(b)}, \tilde{M}_{i,(b)}) \) and \( (\tilde{K}_{0,(b-1)}, \tilde{K}_{i,(b-1)}) \).

We now describe the coding scheme in more detail. Our scheme has parameters \( (U_0, U_1, U_2, V_0, V_1, V_2, P_{U_0U_1U_2}, f, P_{V_0V_1V_2|U_0U_1U_2}, \tilde{Y}, R_0, R_1, R_2, \tilde{R}_1, \tilde{R}_2, \bar{R}_1, \bar{R}_2, \epsilon, \gamma, n, B) \), where:

- \( U_0, U_1, U_2, V_0, V_1, \) and \( V_2 \) are finite auxiliary alphabets;
- \( P_{U_0U_1U_2} \) is a joint probability law over \( U_0 \times U_1 \times U_2 \);
- \( f \) is a function \( f : U_0 \times U_1 \times U_2 \to X \);
- \( P_{V_0V_1V_2|U_0U_1U_2} \tilde{Y} \) is a conditional probability law over \( V_0 \times V_1 \times V_2 \) given a tuple \((U_0, U_1, U_2, \tilde{Y})\);
- \( R_0, R_1, R_2, \tilde{R}_0, \tilde{R}_1, \tilde{R}_2, \bar{R}_0, \bar{R}_1, \bar{R}_2 \) are nonnegative communication rates;
- \( \tilde{R}_1, \tilde{R}_2, \bar{R}_1, \bar{R}_2 \) are nonnegative binning rates;
- \( \epsilon > 0 \) is a small number; and
- \( n, \gamma, \) and \( B \) are positive integers determining the scheme’s blocklength.

1) **Code Construction:** For each block \( b \in [B] \) we construct a Marton code for a DMBC with parameters \( (X, Y_1, Y_2, P_{Y_1Y_2|X}) \) using the code construction in Subsection [V-B]. As parameters of this construction we choose:

- the auxiliary alphabets \( U_0, U_1, U_2 \);
- the joint law \( P_{U_0U_1U_2} \) over these alphabets;
- the function \( f : U_0 \times U_1 \times U_2 \to X \);
- the nonnegative communication rates \( \tilde{R}_0, \tilde{R}_1, \tilde{R}_2 \);
- the nonnegative binning rates \( \bar{R}_1, \bar{R}_2 \);
- the small number \( \epsilon \); and
- the blocklength \( n \).

For block \( B + 1 \), we use a Marton scheme for the DMBC with parameters \((X, Y_1, Y_2, P_{Y_1Y_2|X}) \) of block length \( \gamma n \) where the scheme is chosen as to achieve the rate triplet \((\gamma^{-1} \tilde{R}_0, \gamma^{-1} \tilde{R}_1, \gamma^{-1} \tilde{R}_2)\). To make sure that such a scheme exists, we assume throughout the proof that the single-user
channels $P_{Y_b|X}$ and $P_{Y_b|X}$ both have positive capacities. Under this assumption, it is readily verified that for $\gamma > 1$ large enough such a scheme exists.

In what follows, let $\varphi(b)$, $\Phi_1(b)$, $\Phi_2(b)$ denote the encoding and decoding rules corresponding to the Marton-code in block $b$, for any $b \in \{B+1\}$. Also, let the triplet $(U^n_0(b), U^n_1(b), U^n_2(b))$ denote the auxiliary codewords produced by the block-$b$ Marton encoder $\varphi(b)$, for any $b \in \{B\}$, and let $X^n_0(b), Y^n_1(b), Y^n_2(b)$ denote the corresponding blocks of channel inputs/outputs/feedback outputs.

Then, consider the LGW-SI setup with the following parameters:

- the source alphabet $(U_0 \times U_1 \times U_2 \times \hat{Y})$;
- the decoder side-information alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$;
- the reconstruction alphabets $\mathcal{R}_1$ and $\mathcal{R}_2$;
- the source-side-information law $P(U_0 U_1 U_2 \bar{Y} Y_1 Y_2)$;
- the reconstruction laws $P(U_0 U_1 U_2 \bar{Y} Y_1 Y_2)$, $P(U_0 U_1 U_2 \bar{Y} Y_1 Y_2)$

For this LGW-SI setup we construct for each block $b \in \{B\}$ an LGW-SI code as described in Subsection [V-B1]. Our construction has the following parameters:

- the auxiliary alphabet $\mathcal{Y}_0$;
- the conditional law $P(U_0 Y_0 U_1 Y_1 U_2 Y_2)$;
- the nonnegative rates $\hat{R}_0, \hat{R}_1, \hat{R}_2, \tilde{R}_0, \tilde{R}_1, \tilde{R}_2, \tilde{R}_0, \tilde{R}_1, \tilde{R}_2$;
- the binning rates $\tilde{R}^0, \tilde{R}^1, \tilde{R}^2 \geq 0$, where $\tilde{R}^0$ cannot be smaller than $\max \{\tilde{R}_0, \tilde{R}_1, \tilde{R}_2\}$; and
- the small number $\epsilon/2$; and
- the blocklength $n$.

In what follows, let $\Lambda_{b}(b)$, $\Lambda_{b}(b)$, and $\Lambda_{b}(b)$ denote the LGW-SI encoding and decoding rules corresponding to these codes.

2) Encoding: In the first block $b = 1$, the transmitter forms the product messages $J_0(1) = (M_0(1), I_1(1))$, $J_1(1) = (M_1(1), I_1(1))$, and $J_2(1) = (M_2(1), I_1(1))$, and applies the Marton encoding rule $\varphi_{b}(1)$ to this triplet $(J_0(1), J_1(1), J_2(1))$.

In blocks $b = 2, \ldots, B$ the transmitter first applies the LGW-SI encoding function $\lambda_{b}(1)$ to its “source sequence” $(U^n_0(b-1), U^n_1(b-1), U^n_2(b-1), Y^n_0(b-1))$ to generate the update messages $(R_0(b-1), R_1(b-1), R_2(b-1))$. (Recall that $U^n_0(b-1), U^n_1(b-1), U^n_2(b-1)$ denote the Marton auxiliary codewords produced in the previous encoding step.) The transmitter then generates the messages $J_{b}(b) = (M_{b}(b), K_{b}(b-1))$, and encodes them via the Marton encoding rule $\varphi_{b}(b)$. It finally sends the outcome of this encoding over the channel.

In the last block $B + 1$, the transmitter first applies the LGW-SI encoding function $\lambda_{b}(1)$ to the sequences $(U^n_0(B), U^n_1(B), U^n_2(B), Y^n_0(B))$ to generate the update messages $K_{b}(B)$, for $i \in \{0, 1, 2\}$. It then forms the tuple $J_{b}(B+1) = (1, K_0(B), J_{b}(B+1))$, and $J_{b}(B+1) = (1, K_1(B), K_2(B))$ and encodes them via the Marton encoding rule $\varphi_{b}(B+1)$.

3) Decoding at Receiver $i$: Decoding is performed backwards, starting from the last block. Receiver $i$ first applies the decoding rule $\Phi_{i,c}(B+1)$ to the outputs $Y^n_{i,c}(B+1)$ attempting to decode the indices $(J_{0,i}(B+1), J_{1,i}(B+1))$, and parses its guess $(J_{0,i}(B+1), J_{1,i}(B+1))$ as $J_{0,i}(B+1) = (1, K_{0,i}(B))$ and $J_{1,i}(B+1) = (1, K_{1,i}(B))$.

Now, for every block $b \in \{1, \ldots, B\}$, starting with block $B$ and going backwards, the receiver performs the following steps. It applies the LGW-SI decoder $\Lambda_{i,b}(b)$ to its guess of the update messages $(\hat{K}_{0,i}(b), \hat{K}_{1,i}(b))$ obtained in block $b+1$, and to its “side-information” $Y^n_{i,c}(b)$. It then applies Marton’s decoding rule $\Pi_{i,b}(b)$ to the pair $(\hat{Y}_{i,c}(b), V^n_{i,c}(b))$, where $\hat{V}_{i}(b)$ denotes the reconstruction sequence produced by the LGW-SI decoder $\Lambda_{i,b}(b)$. Finally, it parses the guess produced by Marton’s decoding rule $(\hat{J}_{0,i,c}(b), \hat{J}_{1,i,c}(b))$ as $\hat{J}_{0,i,c}(b) = (\hat{M}_{0,i,c}(b), \hat{K}_{0,i,c}(b-1))$ and $\hat{J}_{1,i,c}(b) = (\hat{M}_{1,i,c}(b), \hat{K}_{1,i,c}(b-1))$.

Receiver $i$’s guess of the messages $M_0$ and $M_i$ are the products $M_{0,i} = (\hat{M}_{0,i,c}(1), \ldots, \hat{M}_{0,i,c}(B))$ and $M_i = (\hat{M}_{1,i,c}(1), \ldots, \hat{M}_{1,i,c}(B))$.

4) Analysis: In Appendix [D] we show that under Constraints [12] the error probability of our scheme tends to 0 as $n \to \infty$. The existence of a deterministic coding scheme with vanishing error probability follows from standard arguments.

VII. EXAMPLES

A. The Generalized Dueck DMBC

In [5] Dueck presented the first example of a DMBC where noise-free feedback increases capacity. In his setup, the channel input consists of three bits, $X = (X_0, X_1, X_2)$, and each of the two outputs of two bits, $Y_1 = (Y_{1,1}, Y_{1,0})$ and $Y_2 = (Y_{2,0}, Y_{2,2})$ where

$Y_{1,0} = Y_{2,0} = X_0$, $Y_{1,1} = X_1 \oplus Z$, $Y_{2,2} = X_2 \oplus Z$.

Here, the noise $Z$ is $Bern(1/2)$ and independent of the inputs, and $\oplus$ denotes addition modulo 2.
Obviously, without feedback, the outputs $Y_{1,1}$ and $Y_{2,2}$ are useless. Thus, the no-feedback-capacity is given by the set of all nonnegative rate triplets $(R_0, R_1, R_2)$ satisfying

$$R_0 + R_1 + R_2 \leq 1.$$  

With noisless feedback, the capacity is increased.

**Theorem 4 (Dueck [5]).** The noiseless feedback capacity of Dueck's DMBC is given by the set of all nonnegative rate triplets $(R_0, R_1, R_2)$ satisfying

$$R_0 + R_1 \leq 1 \quad \text{and} \quad R_0 + R_2 \leq 1.$$  

**Proof.** The converse follows from the cutset bound. The achievability is by the following simple blocklength-$(n + 1)$ scheme. The transmitter sends lossless descriptions of the message pairs $(M_0, M_1)$ and $(M_0, M_2)$ using the inputs $(X_{i,t})_{t=1}^n$ and $(X_{2,t})_{t=1}^n$, respectively. Additionally, for $t = 2, \ldots, (n+1)$, it repeats the previous symbol as $X_{0,t} = Z_{t-1}$. The transmitter knows $Z_{t-1}$ at time $t$ because it is cognizant of the input $X_{1,t-1}$ (or $X_{2,t-1}$) and, through the feedback, also of $Y_{1,t-1} = X_{1,t-1} + Z_{t-1}$ (or $Y_{2,t-1} = X_{2,t-1} + Z_{t-1}$).

Notice that each Receiver $i \in \{1,2\}$ learns the noise sequence $(Z_i)_{t=1}^n$ from its sequence of outputs $(Y_{i,0,i})_{t=1}^{n+1}$. Receiver $i$ can thus compute the channel inputs $X_{i,t} = Y_{i,i-1} - Z_t$, for $t = 1, \ldots, n$, and recover the desired pair of messages $(M_i, M_i)$ whenever the sum-rate $R_0 + R_i$ is smaller than $\frac{n}{n+1}$. Letting the block-length $n$ tend to infinity, we get the desired achievability result.

We generalize Dueck's setup to the DMBC depicted in Figure 7. We assume that all three binary channels are noisy, and the first and third channels are corrupted by different noises. Thus, as before, the channel input consists of three bits, $X = (X_1, X_0, X_2)$, and each output of two bits, $Y_1 = (Y_1, Y_1, 0)$ and $Y_2 = (Y_2, Y_2, 2)$. However, now,

$$Y_{1,0} = Y_{2,0} = X_0 \oplus Z_0,$$

$$Y_{1,1} = X_1 \oplus Z_1,$$

$$Y_{2,2} = X_2 \oplus Z_2,$$

where $Z_0, Z_1, Z_2$ are binary random variables of a given joint law $P_{Z_0Z_1Z_2}$.

**Proposition 1.** The no-feedback capacity region of the generalized Dueck DMBC is the set of all nonnegative rate triplets $(R_0, R_1, R_2)$ that satisfy

$$R_0 + R_1 \leq 2 - H(Z_0, Z_1),$$

$$R_0 + R_2 \leq 2 - H(Z_0, Z_2),$$

$$R_0 + R_1 + R_2 \leq 3 - H(Z_0, Z_1) - H(Z_0, Z_2).$$  

**Proof.** The no-feedback capacity of a DMBC depends on the channel law $P_{Y_1Y_2X}(y_1, y_2 | x)$ only through the marginal laws $P_{Y_1 | X}(y_1 | x)$ and $P_{Y_2 | X}(y_2 | x)$ (see e.g., [27]). We therefore assume in the following that $Z_2 \overset{\text{d}}{=} Z_0 \overset{\text{d}}{=} Z_1$. The converse follows then simply by applying the cutset bound to this modified setup. The achievability follows from Marton's achievable region. More precisely, if in the region in [4] we choose $U_0, U_1, U_2$ to be i.i.d. Bern$(1/2)$, then it evaluates to our region in (15). (Notice that since we choose $U_0, U_1, U_2$ independent, constraint (14) on $2R_0 + R_1 + R_2$ is not active.)

Our scheme in Section 6-B allows us to obtain the capacity region for the Generalized Dueck DMBC with noisless feedback when

$$H(Z_0, Z_1) \leq 1 \quad \text{and} \quad H(Z_0, Z_2) \leq 1.$$  

**Theorem 5.** Under condition (16) and when no common message is sent, i.e., $R_0 = 0$, the noisess-feedback capacity of the Generalized Dueck DMBC is the set of all nonnegative rate pairs $(R_1, R_2)$ satisfying

$$R_1 \leq 2 - H(Z_0, Z_1),$$

$$R_2 \leq 2 - H(Z_0, Z_2),$$

$$R_1 + R_2 \leq 3 - H(Z_0, Z_1) - H(Z_0, Z_2).$$  

**Proof.** The converse follows from the cutset bound. The direct part follows from Theorem 5 by taking the convex hull of the achievable regions that result when (12) is evaluated for the following two choices: $(U_0, U_1, U_2)$ i.i.d. Bern$(1/2)$; $X_i = U_i$ for $i \in \{0, 1, 2\}$; $V_i = (X_0, X_i)$ for $i \in \{1, 2\}$; and either $V_0 = (Z_0, Z_1)$ or $V_0 = (Z_0, Z_2)$. (Notice that since $U_0, U_1, U_2$ are independent, Constraint (12a) is subsumed by Constraints (12a) and (12b).)

In view of Proposition 1 we have the following corollary to Theorem 5.

**Corollary 1.** If the triplet $(Z_0, Z_1, Z_2)$ satisfies (16) and does not form the Markov chain $Z_1 \overset{\text{d}}{=} Z_0 \overset{\text{d}}{=} Z_2$, then noiseless feedback strictly increases the capacity of our Generalized Dueck DMBC.

Let's briefly consider the case of noisy feedback $\overline{Y} = (Y_{1,1} \oplus W_1, Y_{1,0} \oplus W_0, Y_{2,2} \oplus W_2)$ where $(W_0, W_1, W_2)$ are arbitrary distributed binary random variables, with marginals $W_i \sim \text{Bern}(q_i)$, for $q_0, q_1, q_2 \in (0, 1)$. Evaluating Theorem 3 for this noisy-feedback setup is cumbersome and left out. But from Corollary 1 and the continuity considerations mentioned in Remark 4 we can conclude the following.

**Remark 5.** If the noise triplet $(Z_0, Z_1, Z_2)$ satisfies (16) and does not form the Markov chain $Z_1 \overset{\text{d}}{=} Z_0 \overset{\text{d}}{=} Z_2$, then for any sufficiently small value of $\max\{q_0, q_1, q_2\}$, the noisy feedback introduced above enlarges the capacity region of the Generalized Dueck DMBC.

**B. The Noisy Blackwell DMBC**

Consider the noisy version of the Blackwell DMBC [27] in Figure 8. The input alphabet is ternary $\mathcal{X} = \{0, 1, 2\}$ and both output alphabets are binary $\mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1\}$. Let $Z \sim$
Bern(p), with \(p < \frac{1}{2}\), be independent of \(X\). The channel law \(P_{Y_1,Y_2|X}\) is described as follows.

\[
Y_1 = \begin{cases} 
Z & X = 0 \\
1 - Z & X = 1, 2 
\end{cases} \quad Y_2 = \begin{cases} 
Z & X = 0, 1 \\
1 - Z & X = 2. 
\end{cases}
\]

When \(p = 0\), the described DMBC specializes to Blackwell’s DMBC. For this case the capacity region with and without feedback is given by Marton’s region. We consider noiseless feedback and present an achievable region for this setup based on the region \(R_{\text{inner}}\) in Theorem 3.

Let \(U_0, U_1, U_2\) be binary random variables, where \(U_0 \sim \text{Bern}\left(\frac{1}{2}\right)\), and where given \(U_0 = 0\) the pair \((U_1, U_2)\) has joint conditional law

\[
P_{U_1,U_2|U_0=0}: \quad \begin{tabular}{cccc}
& \(U_2 = 0\) & \(U_2 = 1\) \\
\(U_1 = 0\) & \(\alpha\) & \(\beta\) \\
\(U_1 = 1\) & \(1 - \alpha - \beta\) & 0
\end{tabular}
\]

for some nonnegative \(\alpha, \beta\) satisfying \(\alpha + \beta \leq 1\), and given \(U_0 = 1\) it has joint conditional law

\[
P_{U_1,U_2|U_0=1}: \quad \begin{tabular}{cccc}
& \(U_2 = 0\) & \(U_2 = 1\) \\
\(U_1 = 0\) & \(\beta\) & 0 \\
\(U_1 = 1\) & \(1 - \alpha - \beta\) & \(\alpha\)
\end{tabular}
\]

Set \(X \overset{\text{def}}{=} U_1 + U_2\) (real addition), and let \(V_1 \overset{\text{def}}{=} U_1, V_2 \overset{\text{def}}{=} U_2,\) and \(V_0 \overset{\text{def}}{=} V_1 \oplus Y_1 = Z\). Evaluating the region in (12) for this choice of random variables, we obtain the following theorem.

**Theorem 6.** All nonnegative rate triplets \((R_0, R_1, R_2)\) satisfying

\[
\begin{align*}
R_0 + R_1 &\leq h_b\left(\frac{\alpha + \beta}{2}\right) - h_b(p) \\
R_0 + R_2 &\leq h_b\left(\frac{\alpha + \beta}{2}\right) - h_b(p) \\
R_0 + R_1 + R_2 &\leq h_b\left(\frac{\alpha + \beta}{2}\right) + \frac{1 - \alpha}{2} h_b\left(\frac{\alpha}{1 - \beta}\right) \\
&\quad + \frac{1 - \alpha}{2} h_b\left(\frac{\beta}{1 - \alpha}\right) - h_b(p) \\
2R_0 + R_1 + R_2 &\leq 2h_b\left(\frac{\alpha + \beta}{2}\right) - 2h_b(p) \\
&\quad + H\left(\alpha, \beta, 1 - \alpha - \beta\right) - h_b(\alpha) - h_b(\beta)
\end{align*}
\]

are achievable over the Noisy Blackwell DMBC. Here, \(H[p_1, \ldots, p_m] \overset{\text{def}}{=} \sum_{i=1}^{m} p_i \log \frac{1}{p_i},\) \(h_b(p) \overset{\text{def}}{=} H[p, 1-p]\); and \(\gamma * p \overset{\text{def}}{=} (1-\gamma)p + \gamma (1-p).\)

Let us consider the sum-rates \(R_1 + R_2\) guaranteed by the region above. To that end, we set \(R_0 = 0\) and note it is sufficient to consider only the last two inequalities. We get the following corollary to Theorem 6.

**Corollary 2.** With noiseless feedback, our scheme achieves all nonnegative rate pairs \((R_1, R_2)\) satisfying Inequality (19) shown on top of the next page.

For comparison, let us now upper bound the sum-rates \(R_1 + R_2\) that are achievable without feedback. Since the no-feedback capacity of a DMBC depends only on the marginals \(P_{Y_1|X}, P_{Y_2|X}\), the capacity region for the Noisy Blackwell channel remains the same if in the definitions of \(Y_1\) and \(Y_2\) (see (18) we replace \(Z\) by independent Bern(p) random variables \(Z_1\) and \(Z_2\), respectively. Computing the cut-set upper bound for this latter setting, we obtain that all rate pairs \((R_1, R_2)\) that are achievable without feedback must satisfy

\[
R_1 + R_2 \leq \sup_{\alpha \in (0, 1)} \left\{ H\left(\alpha(\bar{p} - p)^2 + \bar{p}p, \bar{p}^2 + 2\alpha\bar{p}(p - \bar{p}), \alpha(p - \bar{p})^2 + \bar{p}p\right) \right\} - 2h_b(p),
\]

where \(\bar{p} \overset{\text{def}}{=} 1 - p\). Figure 9 depicts the bounds (19) and (20) together with a cut-set upper bound on the sum-rates \(R_1 + R_2\) that are achievable with noiseless feedback. By this Figure 9

**Corollary 3.** Noiseless feedback enlarges the capacity region of the Noisy Blackwell-DMBC.

**Remark 6.** Let \(\tilde{Y} = (Y_1 \oplus W_1, Y_2 \oplus W_2),\) where \((W_1, W_2)\) are jointly distributed binary random variables with marginals \(W_i \sim \text{Bern}(q_i),\) mutually independent of \((X, Y_1, Y_2)\). By the continuity argument in Remark 2 for any \(p \in (0, 1)\) and small enough \(\max\{q_1, q_2\}\), noisy feedback strictly enlarges the capacity region of the Noisy Blackwell-DMBC with noisy feedback.
\[ R_1 + R_2 \geq \sup_{\alpha, \beta \geq 0: \alpha + \beta \leq 1} \min \left\{ h_b \left( \frac{(\alpha + \beta)}{2} \right) + \frac{1 - \beta}{2} h_b \left( \frac{\alpha}{1 - \beta} \right) + \frac{1 - \alpha}{2} h_b \left( \frac{\beta}{1 - \alpha} \right) - h_b(p), \right\} \]

\[ 2h_b \left( \frac{(\alpha + \beta)}{2} \right) + H \left( [\alpha, \beta, 1 - \alpha - \beta] \right) - h_b(\alpha) - h_b(\beta) - 2h_b(p) \]  

(19)

**ACKNOWLEDGEMENT**

The authors thank the Associate Editor, the anonymous reviewers, and Prof. Daniela Tuninetti for the careful reading of the manuscript and their valuable comments. In particular for pointing them to [23].

**APPENDIX A**

**ANALYSIS OF MARTON’S SCHEME**

We analyze the average probability of error of Marton’s scheme averaged over the random messages, codebooks, and channel realizations, see also [10], [11], [17]. Recall that an error occurs whenever \((\hat{M}_{0,1}, \hat{M}_1) \neq (M_0, M_1)\) or \((\hat{M}_{0,2}, \hat{M}_2) \neq (M_0, M_2)\).

By the symmetry of the code construction
\[ \Pr \{ \text{error} \} = \Pr \{ \text{error} | M_\circ = M_{1,p} = M_{2,p} = 1 \}. \]

To shorten notation we denote the event that \(M_\circ = M_{1,p} = M_{2,p} = 1\) by \(M = 1\). Also, let

- \(E_0\) be the event that there is no pair \((\ell_1, \ell_2) \in [2^nR_1] \times [2^nR_2]\) satisfying
  \[ (U^{0^n}_1(1), U^{1^n}_1(1, 1, \ell_1), U^{2^n}_2(1, 1, \ell_2)) \in T^{(n)}(P_{U_0U_1U_2}). \]
- \(E_{0i}\) be the event that
  \[ (U^{0^n}_0(1), U^{1^n}_1(1, 1, \ell_1), U^{2^n}_2(1, 1, \ell_2), Y^n) \not\in T^{(n)}(P_{U_0U_1Y}), \]
  where \(L^1_2\) and \(L^2_2\) denote the pair of indices chosen during the encoding step.
- \(E_{1,i}\) be the event that there is a \(\hat{m}_c \neq 1\) such that
  \[ (U^{0^n}_0(\hat{m}_c), U^{1^n}_1(\hat{m}_c, 1, L^*_i), Y^n) \in T^{(n)}(P_{U_0U_1Y}). \]
- \(E_{2i}\) be the event that there is a pair \(\hat{m}_i \neq 1\) and \(\hat{\ell}_i\) such that
  \[ (U^{0^n}_0(1), U^{1^n}_1(1, \hat{m}_i, \hat{\ell}_i), Y^n) \in T^{(n)}(P_{U_0U_1Y}). \]
- \(E_{3i}\) be the event that there is a tuple \(\hat{m}_c \neq 1, \hat{m}_i \neq 1,\) and \(\hat{\ell}_i\) such that
  \[ (U^{0^n}_0(\hat{m}_c), U^{1^n}_1(\hat{m}_c, \hat{m}_i, \hat{\ell}_i), Y^n) \in T^{(n)}(P_{U_0U_1Y}). \]

When the event \((E_0 \cap E_{0,i} \cap E_{1,i} \cap E_{2,i} \cap E_{3,i})\) occurs, then Receiver \(i \in \{1, 2\}\) correctly decodes its desired messages \(M_0\) and \(M_i\). Therefore,
\[ \Pr (\text{error} | M = 1) = 1 \]
\[ \leq \Pr (E_0 \cup \bigcup_{i=1}^{2} \bigcup_{j=1}^{4} E_{i,j} | M = 1) \]
\[ \leq \Pr (E_0 | M = 1) + \sum_{i=1}^{2} \left( \Pr (E_{0i} | E_{0}, M = 1) + \Pr (E_{1i} | E_{0i}, M = 1) \right) \]
\[ + \Pr (E_{2i} | E_{0i}, M = 1) + \Pr (E_{3i} | E_{0i}, M = 1) \].

We consider each of the terms separately. A nonnegative function \(\delta(\epsilon)\) satisfying \(\delta(\epsilon) \to 0\) as \(\epsilon \to 0\) can be chosen such that the following statements hold.

- By the code construction and by a conditional version of the covering lemma (Lemma [2]),
  \[ \lim_{n \to 0} \Pr (E_0 | M = 1) = 0, \]
  whenever
  \[ R_1' + R_2' > I(U_1; U_2 | U_0) + \delta(\epsilon). \]
- Since the channel outputs \(Y^n_1\) is a \(P_{Y_1} - \text{i.i.d.}\) sequence given \(X^n\) and by the conditional typicality lemma (Lemma [1]),
  \[ \lim_{n \to 0} \Pr (E_{0i} | E_{0}, M = 1) = 0. \]
- By the code construction and by the packing lemma (Lemma [3]),
  \[ \lim_{n \to 0} \Pr (E_{1i} | E_{0i}, M = 1) = 0, \]
  whenever
  \[ R_0 + R_{1,c} + R_{2,c} < I(U_0, U_1; Y_i) - \delta(\epsilon). \]
- By the code construction and by the packing lemma:
  \[ \lim_{n \to 0} \Pr (E_{2i} | E_{0i}, M = 1) = 0, \]
  whenever
  \[ R_{1,p} + R_i' < I(U_i; Y_i | U_0) - \delta(\epsilon). \]
- Again, by the code construction and by the packing lemma:
  \[ \lim_{n \to 0} \Pr (E_{3i} | E_{0i}, M = 1) = 0, \]
  whenever
  \[ R_0 + R_{1,c} + R_{2,c} + R_{i,p} + R_i' < I(U_0, U_i; Y_i) - \delta(\epsilon). \]
Thus, we conclude that if for $i\in\{1,2\}$

$$R'_1 + R'_2 > I(U_1; U_2|U_0) + \delta(\epsilon) \quad (30a)$$

$$R_{i,p} + R'_i < I(U_i; Y_i|U_0) - \delta(\epsilon) \quad (30b)$$

$$R_0 + R_{1,c} + R_{2,c} + R_{i,p} + R'_i < I(U_0, U_1; Y_i) - \delta(\epsilon), \quad (30c)$$

then the average (over random codebooks, messages, and channel realizations) probability of error of the described scheme tends to 0 as the blocklength $n$ tends to infinity. The existence of a deterministic scheme with average (over messages and channel realizations) probability of error tending to 0 as $n$ tends to infinity follows then from standard arguments.

By the Fourier-Motzkin elimination algorithm we conclude that whenever

$$I(U_1; Y_1|U_0) + I(U_2; Y_2|U_0) \geq I(U_1; U_2|U_0) \quad (31)$$

then for every rate tuple $(R_0, R_1, R_2)$ satisfying

$$R_0 + R_1 < I(U_0, U_1; Y_1) - \delta(\epsilon) \quad (32a)$$

$$R_0 + R_2 < I(U_0, U_2; Y_2) - \delta(\epsilon) \quad (32b)$$

$$R_0 + R_1 + R_2 < I(U_1; Y_1|U_0) + I(U_2; Y_2|U_0) + \min_{i=1,2} I(U_0; Y_i) - I(U_1; U_2|U_0) - \delta(\epsilon) \quad (32c)$$

$$2R_0 + R_1 + R_2 < I(U_0, U_1; Y_1) + I(U_0, U_2; Y_2) - I(U_1; U_2|U_0) - \delta(\epsilon) \quad (32d)$$

for a suitable $\delta(\epsilon) \to 0$ as $\epsilon \to 0$, there exists a choice of the rates $R_{1,p}, R_{1,c}, R_{2,p}, R_{2,c}, R'_1, R'_2 > 0$ such that $R_1 = R_{1,p} + R_{1,c}$ and $R_2 = R_{2,p} + R_{2,c}$ and such that (30) holds.

Notice that for every choice of $(U_0, U_1, U_2, X)$ that does not satisfy (31) we can strictly enlarge the rate region (32) if we replace the random triple $(U_0, U_1, U_2)$ by $(U'_0, U'_1, U'_2)$ where $U'_1$ and $U'_2$ are constants and $U'_0 = (U_0, U_1, U_2)$. The new choice $(U'_0, U'_1, U'_2, X)$ moreover satisfies (31) because both sides are 0. Also, $X$ can be written as a function of the new auxiliaries $U'_0, U'_1, U'_2$. We thus conclude that the rate region in (32) is achievable also when (31) is violated.

Taking $\epsilon \to 0$, now establishes the inclusion $\mathcal{R}_{\text{Marton}} \subseteq \mathcal{C}_{\text{NoFB}}$.

The following two remarks are found useful in the analysis of our feedback scheme in Appendix B.

**Remark 7.** Under conditions (31) and (32) there exists an associated choice of parameters for our scheme such that the associated auxiliary codewords satisfy

$$\Pr\left(\left(U^n_0(M_c), U^n_1(M_c, M_{1,p}, L'_1), U^n_2(M_c, M_{2,p}, L'_2)\right) \in T^n_{(32)}/(P_{U_0U_1U_2})\right) \to 1 \quad \text{as} \quad n \to \infty.$$  

**Remark 8.** Inspecting the proof, we see that the memoryless channel property has been used only to establish the limit (23). The other limits (23), (24), (26), and (28) follow solely from the way we constructed the code. Suppose now we replace the memoryless channel with a general channel $P_{Y^n|X^n}$. Then under conditions (31) and (32), there exists an associated choice of parameters for our scheme such that the average error probability goes to zero as $n \to \infty$, if for $i\in\{1,2\}$:

$$\Pr\left(\left(U^n_0(M_c), U^n_i(M_c, M_{i,p}, L'_i), Y^n_i\right) \in T^n_{(i)}/(P_{U_0U_iY_i})\right) \to 1 \quad \text{as} \quad n \to \infty.$$  

**APPENDIX B**

**ANALYSIS OF THE LOSSY GRAY WYNER SCHEME WITH SIDE-INFORMATION**

We analyze the failure probability $Pr\left(\mathcal{E}^{(i)} \cup \mathcal{E}^{(2)}\right)$ associated with our random coding scheme, where $\mathcal{E}^{(i)}$ is the event that Receiver $i$ fails, i.e., $(X^n, Y^n_i) \notin T^n_{(i)}/(P_{X^nY^n_i})$.

Let $K^n_{0,0}, K^n_{1,0}, K^n_{2,0}, L^n_{1,0}, L^n_{2,0}, K^n_{1,1}, K^n_{1,2}, K^n_{2,2}$, $L^n_{2,2}$ be the tuple of indices chosen by the sender. Also, let

- $\mathcal{E}_0$ be the event that $X^n \notin T^n_{(n)}/(P_X)$;
- $\mathcal{E}_1$ be the event that
  \[ \forall k_{0,0}, k_{1,0}, \ell_{1,0} : \quad (X^n, Y^n_0((1; k_{0,0}, k_{1,0}, \ell_{1,0})) \notin T^n_{(1)}/(P_{X^nY^n_0}) \]
- $\mathcal{E}_{2,i}$, for $i\in\{1,2\}$, be the event that
  \[ \forall k_{0,0}, k_{1,0}, \ell_{1,0} : \quad (X^n, Y^n_0((i; k_{0,0}, k_{1,0}, \ell_{1,0})) \notin T^n_{(i)}/(P_{X^nY^n_0}) \]
- $\mathcal{E}_{3,i}$, for $i\in\{1,2\}$, be the event that
  \[ \exists \ell_{1,0} \neq L_{1,0}^i : \quad (V^n_0((i; K_{0,0}, K_{1,0}, L_{1,0}^i)), V^n_i(K_{1,1}, L_{1,1}^i, L_{1,2}^i), Y^n_i) \in T^n_{(i)}/(P_{X^nY^n_i}) ; \]
- $\mathcal{E}_{4,i}$, for $i\in\{1,2\}$, be the event that
  \[ \exists \ell_{1,0} \neq L_{1,0}^i : \quad (V^n_0((i; K_{0,0}, K_{1,0}, L_{1,0}^i)), V^n_i(K_{1,1}, L_{1,1}^i, \ell_{1,0}), Y^n_i) \in T^n_{(i)}/(P_{X^nY^n_i}) ; \]
- $\mathcal{E}_{5,i}$, for $i\in\{1,2\}$, be the event that
  \[ \exists \ell_{1,0} \neq L_{1,0}^i : \quad (V^n_0((i; K_{0,0}, K_{1,0}, L_{1,0}^i)), V^n_i(K_{1,1}, L_{1,1}^i, \ell_{1,0}), Y^n_i) \in T^n_{(i)}/(P_{X^nY^n_i}) . \]

Notice that whenever event $(\mathcal{E}_0 \cap \mathcal{E}_1 \cap \mathcal{E}_{2,i} \cap \mathcal{E}_{3,i} \cap \mathcal{E}_{4,i} \cap \mathcal{E}_{5,i})$ occurs, then $(X^n, Y^n_i(K_{0,0}, K_{1,0}, L_{1,0}^i)) \in T^n_{(i)}/(P_{X^nY^n_i})$. If additionally also event $(\mathcal{E}_{3,i} \cap \mathcal{E}_{4,i} \cap \mathcal{E}_{5,i})$ occurs, then Receiver $i$ produces $\hat{V}_i^n = V^n_i(K_{0,0}, K_{1,0}, L_{1,0}^i)$. Therefore,

$$\Pr(\mathcal{E}^{(i)}) \leq \Pr(\mathcal{E}_0 \cup \mathcal{E}_1 \cup \mathcal{E}_{2,i} \cup \mathcal{E}_{3,i} \cup \mathcal{E}_{4,i} \cup \mathcal{E}_{5,i}) \leq \Pr(\mathcal{E}_0) + \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_{2,i}) + \Pr(\mathcal{E}_{3,i}) + \Pr(\mathcal{E}_{4,i}) + \Pr(\mathcal{E}_{5,i}).$$

We analyze each of the summands separately. Hereinafter, a nonnegative function $\delta(\epsilon)$ satisfying $\delta(\epsilon) \to 0$ as $\epsilon \to 0$, can be chosen such that the statements hold.

- Since $X^n$ is $P_X$-i.i.d. and by the weak law of large numbers:
  \[ \lim_{n \to \infty} \Pr(\mathcal{E}_0) = 0. \]
• By the code construction and the covering lemma (Lemma 3):
\[
\lim_{n \to \infty} \Pr(\mathcal{E}_1|\mathcal{E}_0^c) = 0
\]  (35)
whenever
\[
R_0' + R_{0,0} > I(X;V_0) + \delta(\epsilon). \tag{36}
\]

• Again, by the code construction and the covering lemma:
\[
\lim_{n \to \infty} \Pr(\mathcal{E}_{2,i}|\mathcal{E}_1^c) = 0 \tag{37}
\]
whenever
\[
R_i' + R_{0,0} + R_{i,i} > I(V_i;X,V_0) + \delta(\epsilon). \tag{38}
\]

• The pair \( (V_0^n(i;K_0^*,K_{i,i}^*,L_i^*) , V_{i,i}^n(K_0^*,K_{i,i}^*,L_i^*) ) \) depends on \( Y_i^n \) only through \( X^n \), i.e., the Markov chain
\[
V_0^n(i;K_0^*,K_{i,i}^*,L_i^*) \rightarrow X^n \rightarrow Y_i^n
\]
holds. Therefore, \( Y_i^n = F_{Y_i|X,V_{i,i}} = F_{Y_i|X} \)-independent given \( (X^n,V_0^n(i;K_0^*,K_{i,i}^*,L_i^*), V_{i,i}^n(K_0^*,K_{i,i}^*,L_i^*)) \) and by the conditional typicality lemma (Lemma 4): 
\[
\lim_{n \to \infty} \Pr(\mathcal{E}_3,|\mathcal{E}_{2,i}^c) = 0. \tag{39}
\]

• Notice that the codewords \( \{V_i^n(K_{0,i,i}^*,K_{i,i}^*,\ell_i)\} \) for \( \ell_i \in [2^{nR_i'}] \setminus \{L_i^*\} \) are not independent and \( P_{V_i,i,i} \)-identical. However, following similar steps as in Appendix 12A, one can prove Inequality (41) on the next page for arbitrary \( (k_{0,i,i}^*,k_{i,i}^*,\ell_i^*) \in [2^{nR_i'}] \times [2^{nR_i'}] \times [2^{nR_i'}] \) which directly yields Inequality (42), also shown on the next page. Here, Equality (40) holds by the symmetry of the code construction. For \( (k_{0,i,i}^*,k_{i,i}^*,\ell_i^*) = (1,1,1) \), Inequality (41) is straightforward; for \( (k_{0,i,i}^*,k_{i,i}^*,\ell_i^*) \neq (1,1,1) \) and \( \ell_i^* \geq 1 \) it follows by this first case and the symmetry of the code construction; and for \( (k_{0,i,i}^*,k_{i,i}^*,\ell_i^*) \neq (1,1,1) \) and \( \ell_i^* \) arbitrary it follows by (43)–(44) which hold again by the symmetry of the code construction and because conditioned on \( K_{0,i,i} = K_{i,i} = L_i^* = 1 \) every set of \( [2^{nR_i'}] - 1 \) codewords \( \{V_i^n(k_{0,i,i},k_{i,i},\ell_i)\} \) for \( (k_{0,i,i},k_{i,i},\ell_i) \neq (1,1,1) \) has the same joint distribution.

Notice that on the right-hand side of (42) we have the probability that one of the \( [2^{nR_i'}] \)-independent and \( P_{V_i,i,i} \)-i.i.d. codewords \( \{V_i^n(1,1,\ell_i)\}_{\ell_i=1}^{[2^{nR_i'}]} \) is jointly \( \epsilon \)-typical with the pair \( V_0^n(i;K_0^*,K_{i,i}^*,L_i^*), Y_i^n \). Thus, by the packing lemma (Lemma 5), the probability on the right-hand side of (42) tends to 0 as \( n \) tends to infinity whenever
\[
R_i' < I(V_i;V_0,Y_1) - \delta(\epsilon). \tag{45}
\]
We thus conclude that
\[
\lim_{n \to \infty} \Pr(\mathcal{E}_{4,i}) = 0 \tag{46}
\]
whenever (45) holds.

4This can be seen with the following simple example. Let the heights of two students \( A_0 \) and \( A_1 \) be uniformly distributed over the interval \([1.7,1.9]\) m and independent of each other. Also, let \( C \) be the index of the student that has height larger than 1.89m if this index is unique; otherwise let \( C \) be Bernoulli(1/2).

Let \( C \) be the index in \([0,1]\) not equal to \( C \). Notice that \( \Pr(A_C \geq 1.89) = \frac{3}{10} \), whereas \( \Pr(A_C \leq 1.89) = \Pr(A_0 \geq 1.89 \text{ and } A_1 \geq 1.89) = \frac{1}{20} \).

Thus, \( A_C \) is not uniform over \([1.7,1.9]\).

• Following similar steps as above, we can prove upper bound (47). Then, by the multivariate packing lemma (Lemma 4):
\[
\lim_{n \to \infty} \Pr(\mathcal{E}_{5,i}) = 0, \tag{48}
\]
whenever
\[
R_0' - R_{i,0} + R_i' < I(V_0;V_i) + I(V_i;V_0,Y_1) - \delta(\epsilon). \tag{49}
\]
Combining (33) with (36), (38), (45), and (49) we obtain that \( \Pr(\mathcal{E}'(1)) \) and \( \Pr(\mathcal{E}'(2)) \) both tend to 0 as \( n \to \infty \) whenever:
\[
R_0' + R_{0,0} > I(X;V_0) + \delta(\epsilon) \tag{50a}
\]
\[
R_i' + R_{i,0} + R_{1,1} > I(V_i;X,V_0) + \delta(\epsilon) \tag{50b}
\]
\[
R_0' + R_{0,0} + R_{2,2} > I(V_2;X,V_0) + \delta(\epsilon) \tag{50c}
\]
\[
R_0' - R_{1,0} + R_{1,1} < I(V_0;V_1) + I(V_i;V_0,Y_1) - \delta(\epsilon) \tag{50d}
\]
\[
R_0' - R_{2,0} + R_2' < I(V_0;Y_2) + I(V_2;V_0,Y_2) - \delta(\epsilon) \tag{50e}
\]
\[
R_0' < I(V_1;V_0,Y_1) - \delta(\epsilon) \tag{50f}
\]
\[
R_2' < I(V_2;V_0,Y_2) - \delta(\epsilon). \tag{50g}
\]

We now argue that with an appropriate choice of the auxiliary rates \( R_0', R_1', R_2', R_{0,0}, R_{0,1}, R_{0,2}, R_{1,0}, R_{1,1}, R_{1,2}, R_{2,0}, R_{2,2} \) > 0 our scheme achieves the region \( R_{\text{inner}} \). We first replace \( R_{i,i} \) by \( R_i - R_{i,0} \), for \( i \in \{1,2\} \) and \( R_0, R_{0,0} \) by \( R_0 - R_{0,1} - R_{0,2} \) to obtain
\[
R_0' + R_{0,0} - R_{1,0} - R_{2,0} > I(X;V_0) + \delta(\epsilon) \tag{51a}
\]
\[
R_i' + R_{i,0} + R_{1,1} - R_{1,0} > I(V_i;X,V_0) + \delta(\epsilon) \tag{51b}
\]
\[
R_2' + R_{0,0} + R_{2,2} > I(V_2;X,V_0) + \delta(\epsilon) \tag{51c}
\]
\[
R_0' - R_{1,0} + R_{1,1} < I(V_0;V_1) + I(V_i;V_0,Y_1) - \delta(\epsilon) \tag{51d}
\]
\[
R_0' - R_{2,0} + R_2' < I(V_0;Y_2) + I(V_2;V_0,Y_2) - \delta(\epsilon) \tag{51e}
\]
\[
R_0' < I(V_1;V_0,Y_1) - \delta(\epsilon) \tag{51f}
\]
\[
R_2' < I(V_2;V_0,Y_2) - \delta(\epsilon). \tag{51g}
\]

Then, employing the Fourier-Motzkin elimination algorithm to eliminate the nuisance variables \( R_0', R_1', R_2', R_{0,1}, R_{0,2}, R_{1,0}, R_{2,0} \), we obtain that if \( (R_0,R_1,R_2) \) satisfies
\[
R_0 + R_1 > I(X;V_0) + I(V_1;V_0,Y_1) - I(V_0;V_1) \tag{52a}
\]
\[
R_0 + R_2 > I(X;V_0) + I(V_2;V_0,Y_2) - I(V_0;V_2) \tag{52b}
\]
\[
R_0 + R_1 + R_2 > I(X;V_0) + I(V_1;V_0,Y_1) + I(V_2;X,V_0) \tag{52c}
\]
\[
R_0 - R_{1,0} - R_{2,0} \geq 0
\]
\[
R_1 - R_{1,0} \geq 0
\]
\[
R_2 - R_{2,0} \geq 0
\]
\[
R_0 - R_{0,1} - R_{0,2} \geq 0.
\]
Due to the Markov chain $(V_0, V_1, V_2) \rightarrow X \rightarrow (Y_1, Y_2)$ the constraints in (52) are equivalent to

\[
\Pr (\mathcal{E}_{4,i}) = \Pr \left( \bigcup_{\ell_i \in [2^nR_i]} (V^n_0(i; K_{0,0}, K_{i,0}, L_{i,0}), V^n_i(K_{0,i}, K_{i,i}, \ell_i), Y^n_i) \in T^n_\epsilon (P_{V_iY_i}) \right)
\]

* \[
= \Pr \left( \bigcup_{\ell_i \neq L_i^*} (V^n_0(i; K_{0,0}, K_{i,0}, L_{i,0}), V^n_i(1,1, \ell_i), Y^n_i) \in T^n_\epsilon (P_{V_iY_i}) \bigg| K_{0,i}^* = K_{i,i}^* = L_i^* = 1 \right) \tag{40}
\]

* \[
\leq \Pr \left( \bigcup_{\ell_i = 1} (V^n_0(i; K_{0,0}, K_{i,0}, L_{i,0}), V^n_i(1,1, \ell_i), Y^n_i) \in T^n_\epsilon (P_{V_iY_i}) \bigg| K_{0,i}^* = K_{i,i}^* = L_i^* = \ell_i^* \right) \tag{41}
\]

Due to the Markov chain $(V_0, V_1, V_2) \rightarrow X \rightarrow (Y_1, Y_2)$ the constraints in (52) are equivalent to

\[
\Pr (\mathcal{E}_{4,i}) \leq \Pr \left( \bigcup_{\ell_i \neq L_i^*} (V^n_0(i; K_{0,0}, K_{i,0}, L_{i,0}), V^n_i(1,1, \ell_i), Y^n_i) \in T^n_\epsilon (P_{V_iY_i}) \right) \tag{42}
\]

Due to the Markov chain $(V_0, V_1, V_2) \rightarrow X \rightarrow (Y_1, Y_2)$ the constraints in (52) are equivalent to

\[
\Pr \left( \bigcup_{\ell_i = 2} (V^n_0(i; K_{0,0}, K_{i,0}, L_{i,0}), V^n_i(1,1, \ell_i), Y^n_i) \in T^n_\epsilon (P_{V_iY_i}) \bigg| K_{0,i}^* = K_{i,i}^* = L_i^* = 1 \right) \tag{43}
\]

\[
= \Pr \left( \bigcup_{\ell_i = 2} (V^n_0(i; K_{0,0}, K_{i,0}, L_{i,0}), V^n_i(1,1, \ell_i), Y^n_i) \in T^n_\epsilon (P_{V_iY_i}) \bigg| K_{0,i}^* = K_{i,i}^* = k_{i,i}^*, L_i^* = \ell_i^* \right) \tag{44}
\]

Thus, we conclude that the region (53) is $\epsilon$-achievable for all choices of the auxiliary random variable $V_0$ satisfying the Markov chain $(V_0, V_1, V_2) \rightarrow X \rightarrow (Y_1, Y_2)$. Letting $\epsilon \rightarrow 0$, the achievability of $R_{1GW}^{\text{max}}$ is established.

The following remark is found useful in the analysis of the feedback scheme in Appendix [D]

**Remark 9.** In our error analysis, only Limits (34) and (39) rely on the assumption that $(X^n, Y^n_1, Y^n_2)$ are $P_{XY_1Y_2}$-i.i.d. It is easy to check that replacing this assumption with the more general assumptions

\[
(i) \text{ Pr}(X^n \in T^n_\epsilon (P_X)) \rightarrow 1 \text{ as } n \rightarrow \infty.
\]

\[
(ii) \text{ (} Y^n_1, Y^n_2 \text{) is } P_{Y_1Y_2|X} \text{-independent given } X^n.
\]

still guarantees the existence of associated parameters such that the scheme above $\epsilon$-achieves the region (53). In particular,

\[
\Pr \left( (X^n, Y^n_1(K_{0,i}^*, K_{i,i}^*, L_i^*)) \notin T^n_\epsilon (P_{XV_i}) \right) \rightarrow 0
\]

\[
\text{and}
\]

\[
\Pr \left( \hat{V}^n_i \neq V^n_i(K_{0,i}^*, K_{i,i}^*, L_i^*) \right) \rightarrow 0,
\]

for $i \in \{1, 2 \}$, as $n \rightarrow \infty$.

**APPENDIX C**

**CONVEXITY IN THEOREM 2**

Let $\{V_{0,j}, V_{1,j}, V_{2,j}, X_j, Y_{1,j}, Y_{2,j}\} \in (0,1)$ be two sets of mutually independent random variables for $j \in \{1, 2 \}$, where

- $(X_j, Y_{1,j}, Y_{2,j}) \sim P_{XY_1Y_2}$;
- $(V_{0,j}, V_{1,j}, V_{2,j}) \rightarrow X \rightarrow (Y_{1,j}, Y_{2,j})$. 
• $P_{i,j}|X_i = P_{V_i|X}$ for $i \in \{1,2\}$.

Let $Q \sim \text{Bern}(\alpha)$ be independent of the union of the two sets, and define $V_0 \triangleq V_{i,0}, V_i \triangleq V_{i,i}, X \triangleq X_{i,i}, Y_i \triangleq Y_{i,i}$, for $i \in \{1,2\}$. Notice that as the law of $(X_1, Y_{1,1}, Y_{2,1})$ and the law of $(X_2, Y_{1,2}, Y_{2,2})$ are the same, the “time-sharing” random variable $Q$ is independent of the triplet $(X, Y_1, Y_2)$. Therefore, and since by assumption

$$ (V_0, V_1, V_2) \sim (X, Y_1, Y_2), $$

we conclude that defining $V_0 \triangleq Q$, $V_0$ we have the Markov chain

$$ (V_0, V_1, V_2) \sim (X, Y_1, Y_2). $$

(54)

We further notice that for $i \in \{1,2\}$:

$$ I(X; V_i | V_0, Y_i) = I(X; V_i | Y_i). $$

(55)

and

$$ I(X; V_0 | Y_i, Q) = I(X; V_0 | Y_i). $$

(56)

where the first equality holds because of the independence of $Q$ and $(X, Y_i)$. Moreover, by (55) and (56)

$$ I(X; V_0, V_i | Y_i) = I(X; V_i | Y_i). $$

(57)

Combining these inequalities with the Markov condition, we conclude that the region $R_{\text{LGTW}}$ is convex.

**APPENDIX D**

**ERROR ANALYSIS FOR THE FEEDBACK SCHEME**

We bound the average probability of error (where the average is over the random messages, codes, and channel realizations). Let $\mathcal{E}$ be the error event:

$$ \mathcal{E} \triangleq \bigcup_{i=1}^{2} \bigcup_{b=1}^{B} \left\{ \left( \tilde{M}_{0,i,b}, \tilde{M}_{i,b} \right) \neq \left( M_{0,i,b}, M_{i,b} \right) \right\}. $$

Moreover, for each $b \in [B + 1]$, let $\mathcal{F}_b$ be the error event of the Marton code in block $b$:

$$ \mathcal{F}_b \triangleq \bigcup_{i=1}^{B+1} \left\{ \left( \tilde{J}_{0,i,b}, \tilde{J}_{i,b} \right) \neq \left( J_{0,i,b}, J_{i,b} \right) \right\}. $$

Then,

$$ \Pr(\mathcal{E}) \leq \Pr \left( \bigcup_{b=1}^{B+1} \mathcal{F}_b \right) \leq \sum_{b=1}^{B} \Pr(\mathcal{F}_b | \mathcal{F}_{b+1}) + \Pr(\mathcal{F}_{B+1}). $$

By construction, we have that $\Pr(\mathcal{F}_{B+1}) \to 0$ as $n \to \infty$. Let us now analyze the probability $\Pr(\mathcal{F}_b | \mathcal{F}_{b+1})$ for a fixed $b \in [B]$. In light of Remark [2] we see that if

$$ I(U_1; Y_1, V_1 | U_0) + I(U_2; Y_2, V_2 | U_0) \geq I(U_1; U_2 | U_0), $$

and

$$ \tilde{R}_0 + \tilde{R}_1 < I(U_0, U_1; Y_1, V_1) - \delta(\epsilon) $$

(59a)

$$ \tilde{R}_0 + \tilde{R}_2 < I(U_0, U_2; Y_2, V_2) - \delta(\epsilon) $$

(59b)

$$ \tilde{R}_0 + \tilde{R}_1 + \tilde{R}_2 < I(U_1; Y_1, V_1 | U_0) + I(U_2; Y_2, V_2 | U_0) + \min_i I(U_i; Y_i, V_i) - I(U_1; U_2 | U_0) - \delta(\epsilon); $$

(59c)

$$ 2\tilde{R}_0 + \tilde{R}_1 + \tilde{R}_2 < I(U_0, U_1; Y_1, V_1) + I(U_0, U_2; Y_2, V_2) - I(U_1; U_2 | U_0) - \delta(\epsilon); $$

(59d)

and for $i \in \{1,2\}$:

$$ \Pr\left( (U^n_{i,0,b}, U^n_{i,1,b}, U^n_{i,2,b}, V^n_{i,0,b}) \notin T^n_{\epsilon}(P_{U_0 U_i Y_i V_i}) \right) \to 0 $$

(60)

as $n \to \infty$, then there exists a choice of the parameters such that $\Pr\left( \mathcal{F}_b | \mathcal{F}_{b+1} \right) \to 0$ as $n \to \infty$.

From this point forward we assume that Conditions (58) and (59) hold, and prove that if additionally

$$ \tilde{R}_0 + \tilde{R}_1 > I(U_0, U_1, U_2; \tilde{Y}, V_0, V_1) + \delta(\epsilon) $$

(61a)

$$ \tilde{R}_0 + \tilde{R}_2 > I(U_0, U_1, U_2; \tilde{Y}, V_0, V_2) + \delta(\epsilon) $$

(61b)

$$ \tilde{R}_0 + R_1 + \tilde{R}_2 > I(U_0, U_1, U_2; \tilde{Y}, V_0, V_2) + \max_i I(U_0, U_i, U_2; \tilde{Y}, V_0 | V_i) + \delta(\epsilon) $$

(61c)

then the limit (60) holds. We notice that

$$ \Pr\left( (U^n_{i,0,b}, U^n_{i,1,b}, Y^n_{i,0,b}, Y^n_{i,1,b}) \notin T^n_{\epsilon}(P_{U_0 U_i Y_i V_i}) \right) \leq \Pr\left( (U^n_{i,0,b}, U^n_{i,1,b}, V^n_{i,0,b}, V^n_{i,1,b}) \notin T^n_{\epsilon}(P_{U_0 U_i Y_i V_i}) \right) $$

$$ + \Pr\left( \tilde{V}^n_{i,0,b} \neq V^n_{i,0,b}, \tilde{V}^n_{i,1,b} \neq V^n_{i,1,b} \right), $$

(62)

where $V^n_{i,0,b}$ and $V^n_{i,1,b}$ denote the codewords chosen by the LGW-SI encoding rule $\lambda_i(b)$. We now verify that under conditions (61), both terms on the right-hand side of (62) vanish as $n \to \infty$.

Since the input $X^n_i(b)$ is a component-wise function of $(U^n_{i,0,b}, U^n_{i,1,b}, U^n_{i,2,b})$ and the channel is memoryless, $(Y^n_{i,0,b}, Y^n_{i,1,b}, Y^n_{i,2,b})$ is $P_{Y_i} Y_i Y_{i,i} | U_0 U_1 U_2$-independent given $(U^n_{i,0,b}, U^n_{i,1,b}, U^n_{i,2,b})$. Furthermore, from Marton’s code construction and in light of Remark [7] we have that under conditions (58) and (59)

$$ \Pr\left( (U^n_{0,0,b}, U^n_{1,1,b}, U^n_{2,2,b}) \notin T^n_{\epsilon/2}(P_{U_0 U_1 U_2}) \right) \to 0. $$

Therefore, by the conditional typicality Lemma, also

$$ \Pr\left( (U^n_{0,0,b}, U^n_{1,1,b}, U^n_{2,2,b}, \tilde{V}^n_{i,0,b}, \tilde{V}^n_{i,1,b}) \notin T^n_{\epsilon/2}(P_{U_0 U_i U_2 \tilde{Y}_i}) \right) \to 0 $$

as $n \to \infty$.

Thus, by Remark [2] (recall we have used the parameter $\epsilon/2$ for the LGW-SI code) and under conditions (61)

$$ \Pr\left( \tilde{V}^n_{i,0,b} \neq V^n_{i,0,b} \right) \to 0 $$

(63)

$$ \Pr\left( (U^n_{0,0,b}, U^n_{1,1,b}, U^n_{2,2,b}, \tilde{V}^n_{i,0,b}, \tilde{V}^n_{i,1,b}) \notin T^n_{\epsilon/2}(P_{U_0 U_i U_2 \tilde{Y}_i}) \right) \to 0 $$

(64)

as $n \to \infty$.

Now, since $Y^n_{i,0,b}$ is $P_{Y} Y_i Y_{i,i} | U_0 U_i U_2$-independent given $(U^n_{i,0,b}, U^n_{i,1,b}, U^n_{i,2,b})$, and the Markov condition

$$ V^n_{i,0,b} \rightarrow (U^n_{i,0,b}, U^n_{i,1,b}, U^n_{i,2,b}, Y^n_{i,0,b}) \rightarrow Y^n_{i,1,b} $$

holds, (63) and the conditional typicality Lemma imply that

$$ \Pr\left( (U^n_{i,0,b}, U^n_{i,1,b}, U^n_{i,2,b}, Y^n_{i,0,b}) \notin T^n_{\epsilon}(P_{U_0 U_i U_2 \tilde{Y}_i}) \right) \to 0. $$

(65)

With (62) and (63) this establishes (60). We thus proved that whenever (58), (59), and (61) are satisfied, then the probability of error tends to 0 as $n \to \infty$, for any $\epsilon$ small enough.
Employing the Fourier-Motzkin elimination algorithm on constraints \([59]\) and \([61]\) where we replaced \(R_i\) by \(R_i + R_i^\prime\), and letting \(\epsilon\) tend to 0, we obtain that under the set of constraints \([12]\) and when \([58]\) holds, then there exists a choice of the parameters such that the probability of error of our scheme tends to 0 as \(n \to \infty\). Notice however, that when a triplet \((U_0, U_1, U_2)\) does not satisfy \([58]\), then the rate region \([12]\) is strictly enlarged if we replace this triplet by \((U_0', U_1', U_2')\) where \(U_1'\) and \(U_2'\) are constants and \(U_0' = \langle U_0, U_1, U_2\rangle\). The new choice \((U_0', U_1', U_2')\) moreover satisfies \([58]\) because both sides are 0. It also satisfies the Markov chain \([21]\) and \(X\) can be expressed as a function of the new auxiliaries \(U_0', U_1', U_2'\). We can thus ignore constraint \([58]\) in the statement of the achievable region.

We conclude that since by \([13]\) the effective rates of transmission tend to \((R_0, R_1, R_2)\) as \(B \to \infty\), any rate triplet satisfying the constraints \([12]\) is achievable by our scheme.

REFERENCES

[1] A. B. Carleial, “Multiple-access channels with different generalized feedback signals,” IEEE Trans. on Inf. Theory, vol. 28, no. 6, pp. 841–850, Nov. 1982.

[2] F. M. J. Willems, E. C. van der Meulen, and J. P. M. Schalkwijk, “Generalized feedback for the discrete memoryless multiple access channel,” Proc. 21st Annual Allerton Conference on Communication, Control, and Computing, Monticello (IL), USA, October 5–7, 1983, pp. 284–292.

[3] D. Tuninetti, On interference channels with generalized feedback, Proc. of ISIT 2007, Nice, France, June 24–29, 2007, pp. 2861–2865.

[4] A. El Gamal, “The feedback capacity of degraded broadcast channels,” IEEE Trans. on Inf. Theory, vol. 24, no. 3, pp. 379–381, May 1978.

[5] G. Dueck, “Partial feedback for two-way and broadcast channels,” Problems of Information and Control, vol. 46, no. 1, pp. 1–15, 1980.

[6] G. Kramer, “Capacity results for the discrete memoryless network,” IEEE Trans. on Inf. Theory, vol. 49, no. 1, pp. 4–21, Jan. 2003.

[7] E. C. van der Meulen, “A survey of multi-way channels in information theory: 1961–1976,” IEEE Trans. on Inf. Theory, vol. 23, no. 1, pp. 1–37, Jan. 1977.

[8] L. Georgiadis and L. Tassiulas, “Broadcast erasure channel with feedback—capacity and algorithms, Proc. of 2009 Workshop on Network Coding, Theory and Applications, Lausanne, Switzerland, June 15–16, 2009, pp. 54–61.

[9] C. C. Wang, “Capacity of 1-to-k broadcast erasure channel with channel output feedback,” Proc. of 48st Annual Allerton Conference on Communication, Control, and Computing, Monticello (IL), USA, Sep. 29–Oct. 1, 2010, pp. 1347–1354.

[10] K. Marton, “A coding theorem for the discrete memoryless broadcast channel,” IEEE Trans. on Inf. Theory, vol. 25, pp. 306–311, no. 3, May 1979.

[11] S. I. Gelfand and M. S. Pinsker, “Capacity of a broadcast channel with one deterministic component,” Probl. Inf. Transm., vol. 16, no. 1, pp. 17–25, Jan.–Mar. 1980.

[12] A. Lapidoth and Y. Steinberg, “The Multiple Access Channel with Causal and Strictly Causal Side Information at the Encoders,” Proc. Int. Zurich Seminar on Communications (IZS), Zurich, Switzerland, March 3–5, 2010, pp. 13–16.

[13] A. Lapidoth and Y. Steinberg, “The multiple access channel with two independent states each known causally to one encoder,” Proc. ISIT 2010, Austin, Texas, June 13–18, 2010, pp. 480–484.

[14] O. Shayevitz and M. Wigger, “An achievable region for the discrete memoryless broadcast channel with feedback,” Proc. ISIT 2010, Austin, Texas, June 13–18, 2010.

[15] R. Venkataraman and S.S. Pradhan, “Achievable rates for the broadcast channel with feedback,” Proc. ISIT 2010, Austin, Texas, June 13–18, 2010.

[16] R. Venkataraman and S. S. Pradhan, “An achievable rate region for the broadcast channel with feedback,” submitted to IEEE Trans. on Inf. Theory, May, 2011, available at http://arxiv.org/abs/1105.2311.

[17] A. El Gamal and Y.-H Kim, “Lecture notes on network information theory,” Stanford University and UCSD, 2009. Available online: http://arxiv.org/abs/1001.3404

[18] Y. Liang, G. Kramer, and H. V. Poor, “Equivalence of two inner bounds on the capacity region of the broadcast channel,” Proc. 49th Annual Allerton Conference on Communication, Control, and Computing, Monticello (IL), USA, Sept. 2008.

[19] R. Gray and A. Wyner, “Source Coding for a Simple Network,” Bell System Tech. J. vol. 48, pp. 1681–1721, Nov. 1974.

[20] C. Heegard and T. Berger, “Rate distortion when side information may be absent,” IEEE Transactions on Information Theory, vol. 31, no. 6, pp. 727–734, March 1985.

[21] P. W. Cuff and H. H. Permuter and T. M. Cover, “Coordination capacity,” IEEE Trans. on Inf. Theory, vol. 56, no. 9, pp. 4181–4206, Sept. 2010.

[22] M. Wigger, “Source Coding with Side-Information at the Receivers and an Application,” Proc. of 2012 International Zurich Seminar on Communications, Zurich, Switzerland, February 29–March 2, 2012.

[23] C. Tian and S. Diggavi, “Side-information scalable source coding,” IEEE Trans. on Inf. Theory, vol. 54, no. 12, pp. 5591–5608, Dec. 2008.

[24] Y. Steinberg and N. Merhav, “On successive refinement for the Wyner-Ziv problem,” IEEE Trans. on Inf. Theory, vol. 50, no. 8, pp. 1636–1654, Aug. 2004.

[25] R. Tamo and A. Grant and T. Chan and G. Kramer, “Source coding for a simple network with receiver side information,” in Proc. of ISIT 2008, Toronto, Canada, July 6–11, 2008.

[26] R. Tamo and T. J. Oechtering and M. Wigger, “Source Coding for Conditionally Noisy Side-Information,” in Proc. of ITW 2012, Lausanne, Switzerland, September 3–7, 2012.

[27] H. Sato, “An outer bound to the capacity region of broadcast channels,” IEEE Trans. on Info. Theory, vol. 24, no. 3, pp. 374–377, May 1978.

[28] R. W. Yeung, “A First Course in Information Theory”, Kluwer, New York, 2002.

Ofer Shayevitz
Ofer Shayevitz received the B.Sc. degree (summa cum laude) from the Technion Institute of Technology, Haifa, Israel, in 1997 and the M.Sc. and Ph.D. degrees from the Tel-Aviv University, Tel-Aviv, Israel, in 2004 and 2009, respectively, all in electrical engineering.

He is currently a quantitative analyst with the D. E. Shaw group. Before that, he was a Postdoctoral Fellow in the Information Theory and Applications (ITA) Center at the University of California, San Diego. Prior to his graduate studies, he served as an Engineer and Team Leader in the Israeli Defense Forces (1997–2003), and as an Algorithms Engineer with CellGuide, a high-tech company developing low-power GPS navigation solutions (2003–2004). Dr. Shayevitz is the recipient of the ITA Postdoctoral Fellowship (2009–2011), the Adams Fellowship (2006–2008) awarded by the Israel Academy of Sciences and Humanities, the Advanced Communication Center (ACC) Feder Family award for an outstanding Ph.D. thesis (2009), and the Weinstein Prize (2006-2009) for research and publications in signal processing.

Michie Wigger
(S’05, M’09) received the M.Sc. degree in electrical engineering (with distinction) and the Ph.D. degree in electrical engineering both from ETH Zurich in 2003 and 2008, respectively. In 2009, she was a Postdoctoral Researcher at the ITA Center, University of California, San Diego. Since December 2009, she has been an Assistant Professor at Telecom ParisTech, Paris, France. Her research interests are in information and communications theory; in particular in wireless networks, feedback channels, and channels with states.