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, 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, where 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 where the time-t channel outputs depend on the past inputs and outputs only through the time-t 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 BCs 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) BCs 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 [8], [9] is used in each block to send the fresh data and the update information, at rates outside the no-feedback capacity. 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. Loosely speaking, 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

This paper was in part presented at the *International Symposium on Information Theory 2010*, in Austin, TX, July 2010.

O. Shayevitz is with the Information Theory & Applications Center, University of California, San Diego, USA {email: ofersha@ucsd.edu}. M. Wigger was with the Electrical and Computer Engineering Department at University of California, San Diego. She is now with the Communications and Electronics Department, at Telecom ParisTech, Paris, France {email: michele.wigger@telecom-paristech.fr}. Her research at the University of California, San Diego, was supported by the Swiss National Science Foundation under Grant PBEZP2-125703.

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 [10].

Recently, another single-letter achievable region for general DMBCs with feedback has been proposed [13][1]. Comparing the achievable region in [13] 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

For any real number $M > 1$, we use the notation $[M] \stackrel{\text{def}}{=} \{1, \ldots, \lfloor M \rfloor\}$. The set of positive integers is denoted by $\mathbb{Z}^+$. 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 $[2^{n(r_1+r_2)}]$, 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 [14]. For a finite alphabet $\mathcal{X}$, a sequence $x^n \in \mathcal{X}^n$ is said to be ϵ-typical with respect to (w.r.t.) a distribution P_X on $\mathcal{X}$ if

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

for all $x \in \mathcal{X}$, where π_{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 $\mathcal{T}_\epsilon^n(P_X)$. Similarly, for a law $P_{X_1 \cdots X_k}$ over a product alphabet $\mathcal{X}_1 \times \cdots \times \mathcal{X}_k$, we denote by $\mathcal{T}_\epsilon^n(P_{X_1 \cdots X_k})$ the set of all k-tuples of sequences $(x_1^n \in \mathcal{X}_1^n, \ldots, x_k^n \in \mathcal{X}_k^n)$ that are *jointly* ϵ-typical w.r.t. $P_{X_1 \cdots X_k}$.

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

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

Lemma 2 (Covering Lemma [14])**.** *Let $0 < \epsilon' < \epsilon$, and let X^n satisfy $\Pr(X^n \in \mathcal{T}_{\epsilon'}(P_X)) \to 1$ as $n \to \infty$. Also, for each n, let $M_n \in \mathbb{Z}^+$ be larger than 2^{nr} for some $r \ge 0$, 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*

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

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

Lemma 3 (Packing Lemma [14])**.** *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^{nr} for some $r \ge 0$, 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*

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

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 $M_{i,n} \le 2^{nr_i}$, for $i \in \{1,2,3\}$. Also, let $\{U_i^n(m)\}_{m=1}^{M_{i,n}}$ be a set of P_{U_i}-i.i.d. random vectors such that $\{U_1^n(m_1), U_2^n(m_2), U_3^n(m_3)\}$ are mutually independent for any m_1, m_2, m_3. Then, for any law $P_{U_1U_2U_3}$ with marginals $\{P_{U_i}\}_{i=1}^3$, there exists $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ such that*

$$\lim_{n \to \infty} \Pr\Big(\exists\, m_i \in [M_i] \text{ for } i \in \{1,2,3\} \quad s.t. \quad (U_1^n(m_1), U_2^n(m_2), U_3^n(m_3)) \in T_\epsilon^n(P_{U_1U_2U_3})\Big) = 0$$

if

$$r_1 + r_2 + r_3 < I(U_1;U_2) + I(U_3;U_1,U_2) - \delta(\epsilon). \quad (1)$$

[1]The conference version of [13] has been presented in the same session at ISIT 2010 as the conference version of this paper, see [11] and [12].

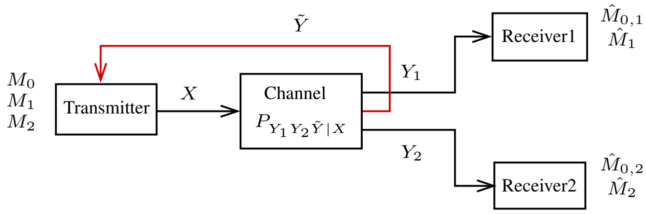

Fig. 1. The two-user discrete memoryless BC with generalized feedback.

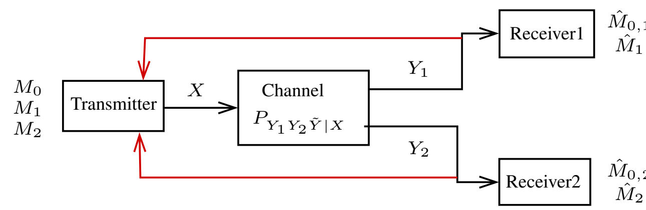

Fig. 2. The two-user DMBC with noise-free feedback from both outputs.

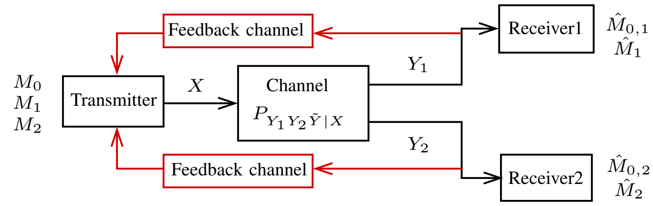

Fig. 3. Example of a two-user DMBC with noisy feedback.

Proof outline. Let $\mathcal{E}_{ijk} \stackrel{\text{def}}{=} \{U_1^n(i), U_2^n(j), U_3^n(k)) \in T_\epsilon^n(P_{U_1U_2U_3})\}$. We need to show that $\Pr\left(\bigcup_{ijk} \mathcal{E}_{ijk}\right) \to 0$ under Constraint (52). By standard typicality/large deviation arguments we have that

$$\begin{aligned}\Pr(\mathcal{E}_{ijk}) &\le 2^{-n\left(D(P_{U_1U_2U_3}\|P_{U_1}\times P_{U_3}\times P_{U_3})-\delta(\epsilon)\right)}\\ &= 2^{-n\left(D(P_{U_1U_2U_3}\|P_{U_1}\times P_{U_3}\times P_{U_3})-\delta(\epsilon)\right)}\\ &= 2^{-n(D(I(U_1;U_2)+I(U_3;U_1,U_2)-\delta(\epsilon))}.\end{aligned}$$

The result follows by taking the union bound over $\mathcal{E}_{ijk}$, and requiring that it tends to zero. ■

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, and M_2 are assumed to be independent and uniformly distributed over the finite sets $[2^{nR_0}]$, $[2^{nR_1}]$, and $[2^{nR_2}]$ 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 $\mathcal{X}$, $\mathcal{Y}_1, \mathcal{Y}_2$, and $\widetilde{\mathcal{Y}}$, and a conditional probability law $P_{Y_1Y_2\widetilde{Y}|X}(y_1, y_2, \widetilde{y}|x)$ where $x \in \mathcal{X}$, $y_1 \in \mathcal{Y}_1$, $y_2 \in \mathcal{Y}_2$, and $\widetilde{y} \in \widetilde{\mathcal{Y}}$. Given that at time t the transmitter feeds the symbol x_t to the channel, Receiver 1 and Receiver 2 observe the channel outputs $y_{1,t} \in \mathcal{Y}_1$ and $y_{2,t} \in \mathcal{Y}_2$ respectively, and the transmitter observes the generalized feedback $\widetilde{y}_t \in \widetilde{\mathcal{Y}}$, with probability $P_{Y_1Y_2\widetilde{Y}|X}(y_{1,t}, y_{2,t}, \widetilde{y}_t|x_t)$.

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 $\widetilde{Y}^{t-1} \stackrel{\text{def}}{=} (\widetilde{Y}_1, \ldots, \widetilde{Y}_{t-1})$:

$$X_t = \psi_t^{(n)}\left(M_0, M_1, M_2, \widetilde{Y}^{t-1}\right), \tag{2}$$

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

$$(Y_1^{t-1}, Y_2^{t-1}, \widetilde{Y}^{t-1}) -\!\circ\!- X_t -\!\circ\!- (Y_{1,t}, Y_{2,t}, \widetilde{Y}_t)$$

where $Y_i^{t-1} \stackrel{\text{def}}{=} (Y_{i,1}, Y_{i,2}, \ldots, Y_{i,t-1})$, for $i \in \{1, 2\}$.

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:

$$(\hat{M}_{0,i}, \hat{M}_i) = \Psi_i^{(n)}(Y_i^n), \quad i \in \{1, 2\} \tag{3}$$

where $\Psi_i^{(n)}$ 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 n encoding functions $\left\{\psi_t^{(n)}\right\}_{t=1}^n$ and two decoding functions $\Psi_1^{(n)}$ and $\Psi_2^{(n)}$ such that the probability of decoding error, i.e., the probability that

$$(M_0, M_1) \neq (\hat{M}_{0,1}, \hat{M}_1) \quad \text{or} \quad (M_0, M_2) \neq (\hat{M}_{0,2}, \hat{M}_2),$$

tends to 0 as the blocklength n tends to infinity. The closure of the set of achievable rate triplets (R_0, R_1, R_2) is called the *feedback capacity-region* of this setup, and we denote it by $\mathcal{C}_{\text{GenFB}}$.

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

IV. Marton's No-Feedback Scheme

We review the description and the analysis of the Marton coding scheme for the DMBC with a common message in [8], [9], [14]. The reason for repeating the scheme and the analysis is to facilitate the statement and verification of Remarks 1 and 2 at the end of this section and the description of our feedback scheme in Section VI-B.

A. Marton's Achievable Region

Let $\mathcal{R}_{\text{Marton}}$ 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

$f\colon \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$ satisfy

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

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

$$\begin{aligned} R_0 + R_1 + R_2 < {} & I(U_1; Y_1|U_0) + I(U_2; Y_2|U_0) \\ & + \min_i I(U_0; Y_i) - I(U_1; U_2|U_0) \end{aligned} \tag{4c}$$

$$\begin{aligned} 2R_0 + R_1 + R_2 < {} & I(U_0, U_1; Y_1) + I(U_0, U_2; Y_2) \\ & - I(U_1; U_2|U_0) \end{aligned} \tag{4d}$$

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

$$(U_0, U_1, U_2) \multimap\!\!- X \multimap\!\!- (Y_1, Y_2)$$

forms a Markov chain, and $(Y_1, Y_2) \sim P_{Y_1Y_2|X}$ given X.

Theorem 1 (From [8], [9])**.** $\mathcal{R}_{\text{Marton}} \subseteq \mathcal{C}_{\text{NoFB}}$.

B. Marton's Scheme

We describe the scheme for a DMBC $(\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2, P_{Y_1Y_2|X})$. The scheme has parameters $(\mathcal{U}_0, \mathcal{U}_1, \mathcal{U}_2, P_{U_0U_1U_2}, f, R_0, R_{1,p}, R_{1,c}, R_{2,p}, R_{2,c}, R_1', R_2', \epsilon, n)$ where

- $\mathcal{U}_0, \mathcal{U}_1, \mathcal{U}_2$ are auxiliary finite alphabets;
- $P_{U_0U_1U_2}$ is a joint law over these auxiliary alphabets;
- $f\colon \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$ is a function mapping the auxiliary inputs into effective inputs;
- $R_0, R_{1,p}, R_{2,p}, R_{1,c}, R_{2,c}$ are nonnegative communication rates where $R_1 \stackrel{\text{def}}{=} R_{1,p} + R_{1,c}$ and $R_2 \stackrel{\text{def}}{=} R_{2,p} + R_{2,c}$;
- R_1', R_2' are nonnegative binning rates;
- $\epsilon > 0$ is a small number; and
- n denotes the scheme's blocklength.

As we shall see, the scheme uses a rate-splitting approach: each private message M_i, for $i \in \{1, 2\}$, is split into two independent submessages $M_{i,p}$ and $M_{i,c}$, where $M_{i,p}$ is uniformly distributed over $[2^{nR_{1,p}}]$ and $M_{i,c}$ is uniform over $[2^{nR_{i,c}}]$. Thus, in our scheme, the private messages M_1 and M_2 are of total rates R_1 and R_2. The submessage $M_{i,c}$ will be decoded by both receivers, and hence we call it the common part of Message M_i. The submessage $M_{i,p}$ is only decoded by its intended receiver i, and we call it the private part of M_i.

As we shall see, Message M_0 is of rate R_0, and the total rate of the messages decoded by both receivers equals $R_c \stackrel{\text{def}}{=} R_0 + R_{1,c} + R_{2,c}$. We denote the associated product common message by $M_c \stackrel{\text{def}}{=} (M_0, M_{1,c}, M_{2,c})$.

1) Code Construction: The code consists of a single codebook $\mathcal{C}_0$, of $\lfloor 2^{nR_c} \rfloor$ codebooks $\{\mathcal{C}_1(m_0)\}_{m_0=1}^{\lfloor 2^{nR_c} \rfloor}$, and of $\lfloor 2^{nR_c} \rfloor$ codebooks $\{\mathcal{C}_2(m_0)\}_{m_0=1}^{\lfloor 2^{nR_c} \rfloor}$.

Codebook $\mathcal{C}_0$ consists of $\lfloor 2^{nR_c} \rfloor$ codewords of length n. Each codeword in $\mathcal{C}_0$ is constructed by randomly and independently drawing all its entries according to the distribution P_{U_0}. We denote the m_0-th codeword in $\mathcal{C}_0$ by $u_0^n(m_0)$.

Every codebook $\mathcal{C}_1(m_0)$, for $m_0 \in [2^{nR_c}]$, consists of $\lfloor 2^{nR_{1,p}} \rfloor$ bins, each bin containing $\lfloor 2^{nR_1'} \rfloor$ codewords of length n. We denote the codewords in bin $m_1 \in [2^{nR_{1,p}}]$ by $\{u_1^n(m_0, m_1, \ell_1)\}_{\ell_1=1}^{\lfloor 2^{nR_1'} \rfloor}$. Each codeword $u_1^n(m_0, m_1, \ell_1)$ in $\mathcal{C}_1(m_0)$ is constructed by randomly drawing its j-th entry independent of all other entries in the codebook and according to the law $P_{U_1|U_0}(\cdot|u_{0,j}(m_0))$, where $u_{0,j}(m_0)$ denotes the j-th entry of $u_0^n(m_0)$.

Similarly, every codebook $\mathcal{C}_2(m_0)$, for $m_0 \in [2^{nR_c}]$, consists of $\lfloor 2^{nR_{2,p}} \rfloor$ bins, each bin containing $\lfloor 2^{nR_2'} \rfloor$ codewords of length n. We denote the codewords in bin $m_2 \in [2^{nR_{2,p}}]$ by $\{u_2^n(m_0, m_2, \ell_2)\}_{\ell_2=1}^{\lfloor 2^{nR_2'} \rfloor}$. Each codeword $u_2^n(m_0, m_2, \ell_2)$ in $\mathcal{C}_2(m_0)$ is constructed by randomly drawing its j-th entry independent of all other entries in the codebook and according to the law $P_{U_2|U_0}(\cdot|u_{0,j}(m_0))$.

Reveal all codebooks to the transmitter, codebooks $\mathcal{C}_0$ and $\{\mathcal{C}_1(m_0)\}$ to Receiver 1, and codebooks $\mathcal{C}_0$ and $\{\mathcal{C}_2(m_0)\}$ to Receiver 2.

2) Encoding: As previously described, the encoder first parses both private messages $M_1 \in [2^{nR_1}]$ and $M_2 \in [2^{nR_2}]$ into pairs of independent submessages: $(M_{1,p}, M_{1,c}) \in [2^{nR_{1,p}}] \times [2^{nR_{1,c}}]$ and $(M_{2,p}, M_{2,c}) \in [2^{nR_{2,p}}] \times [2^{nR_{2,c}}]$. Then, it forms the new common message $M_c = (M_0, M_{1,c}, M_{2,c})$.

The encoding takes place as follows. Given that $M_c = m_0$, $M_{1,p} = m_1$, $M_{2,p} = m_2$, the encoder searches the codebooks $\mathcal{C}_1(m_0)$ and $\mathcal{C}_2(m_0)$ for pairs of codewords $u_1^n(m_0, m_1, \ell_1), u_2^n(m_0, m_2, \ell_2)$ that satisfy[2]

$$(u_0^n(m_0), u_1^n(m_0, m_1, \ell_1), u_2^n(m_0, m_2, \ell_2)) \in \mathcal{T}_{\epsilon/32}^{(n)}(P_{U_0U_1U_2}). \tag{5}$$

It prepares a list with all these pairs of indices $\ell_1 \in [2^{nR_1'}]$ and $\ell_2 \in [2^{nR_2'}]$, and chooses one pair of indices from this list at random. We call the chosen pair (ℓ_1^*, ℓ_2^*). If the list is empty then it chooses (ℓ_1^*, ℓ_2^*) randomly from the set of all indices $[2^{nR_1'}] \times [2^{nR_2'}]$.

The inputs x^n are obtained from the codewords $u_0^n(m_0)$, $u_1^n(m_0, m_1, \ell_1^*)$, $u_2^n(m_0, m_2, \ell_2^*)$ by applying the function f componentwise to these three sequences. That means, for each $j \in [n]$ the j-th channel input is given by

$$x_j = f\big(u_{0,j}(m_0), u_{1,j}(m_0, m_1, \ell_1^*), u_{2,j}(m_0, m_2, \ell_2^*)\big)$$

where $u_{0,j}(m_0)$, $u_{1,j}(m_0, m_1, \ell_1^*)$, and $u_{2,j}(m_0, m_2, \ell_2^*)$ denote the j-th components of $u_0^n(m_0)$, $u_1^n(m_0, m_1, \ell_1^*)$, and $u_2^n(m_0, m_2, \ell_2^*)$.

3) Decoding: Given that Receiver 1 observes the sequence y_1^n, it forms a list of all the tuples $(\hat{m}_0, \hat{m}_1, \hat{\ell}_1)$ that satisfy

$$(u_0^n(\hat{m}_0), u_1^n(\hat{m}_0, \hat{m}_1, \hat{\ell}_1), y_1^n) \in \mathcal{T}_\epsilon^{(n)}(P_{U_0U_1Y_1}). \tag{6}$$

It randomly chooses a tuple $(\hat{m}_0, \hat{m}_1, \hat{\ell}_1)$ from this list and produces as its guess $\hat{M}_{c,1} = \hat{m}_0$ and $\hat{M}_{1,p} = \hat{m}_1$. If the list is empty, then it randomly chooses a pair $(\hat{m}_0, \hat{m}_1)$ from $[2^{nR_c}] \times [2^{nR_{1,p}}]$ and guesses $\hat{M}_{c,1} = \hat{m}_0$ and $\hat{M}_{1,p} = \hat{m}_1$.

Receiver 1 finally parses $\hat{M}_{c,1}$ as $(\hat{M}_{0,1}, \hat{M}_{1,c,1}, \hat{M}_{2,c,1})$. Then, $\hat{M}_{0,1}$ is its guess of Message M_0 and $\hat{M}_1 = (\hat{M}_{1,p}, \hat{M}_{1,c,1})$ its guess of Message M_1.

Receiver 2 produces its guesses of the messages M_0 and M_2 in a similar way. We denote these guesses by $\hat{M}_{0,2}$ and $\hat{M}_2$.

[2]The choice of $\epsilon/32$ will be helpful later. Here, any $\epsilon' < \epsilon$ suffices.

4) Analysis: We analyze the average probability of error of the above scheme averaged over the random messages, codebooks, and channel realizations. Recall that an error occurs whenever

$$(\hat{M}_{0,1},\hat{M}_1)\neq(M_0,M_1) \text{ or } (\hat{M}_{0,2},\hat{M}_2)\neq(M_0,M_2).$$

By the symmetry of the code construction this probability of error equals the average (over all codebooks and channel realizations) probability of error conditioned on the event that $M_c = M_{1,p} = M_{2,p} = 1$, i.e.,

$$\Pr[\text{error}] = \Pr[\text{error}|M_c = M_{1,p} = M_{2,p} = 1],$$

which we analyze in the following. To simplify notation we denote the event that $M_c = M_{1,p} = M_{2,p} = 1$ simply by $\mathcal{M} = 1$.

We define the following events. Let

- $\mathcal{E}_0$ be the event that there is no pair $(\ell_1,\ell_2)\in[2^{nR_1'}]\times[2^{nR_2'}]$ satisfying
$$(U_0^n(1),U_1^n(1,1,\ell_1),U_2^n(1,1,\ell_2))\in\mathcal{T}^{(n)}_{\epsilon/32}(P_{U_0U_1U_2}).$$
- $\mathcal{E}_{0i}$ be the event that
$$(U_0^n(1),U_i^n(1,1,L_i^*),Y_i^n)\notin\mathcal{T}^{(n)}_{\epsilon}(P_{U_0U_iY_i}),$$
where L_1^* and L_2^* denote the pair of indices chosen during the encoding step.
- $\mathcal{E}_{1i}$ be the event that there is a $\hat{m}_0\neq 1$ such that
$$(U_0^n(\hat{m}_0),U_i^n(\hat{m}_0,1,L_i^*),Y_i^n)\in\mathcal{T}^{(n)}_{\epsilon}(P_{U_0U_iY_i}).$$
- $\mathcal{E}_{2i}$ be the event that there is a pair $\hat{m}_i\neq 1$ and $\hat{\ell}_i$ such that
$$(U_0^n(1),U_i^n(1,\hat{m}_i,\hat{\ell}_i),Y_i^n)\in\mathcal{T}^{(n)}_{\epsilon}(P_{U_0U_iY_i}).$$
- $\mathcal{E}_{3i}$ be the event that there is a tuple $\hat{m}_0\neq 1$, $\hat{m}_i\neq 1$, and $\hat{\ell}_i$ such that
$$(U_0^n(\hat{m}_0),U_i^n(\hat{m}_0,\hat{m}_i,\hat{\ell}_i),Y_i^n)\in\mathcal{T}^{(n)}_{\epsilon}(P_{U_0U_iY_i}).$$

When the event $(\mathcal{E}_0^c\cap\mathcal{E}_{0,i}^c\cap\mathcal{E}_{1,i}^c\cap\mathcal{E}_{2,i}^c\cap\mathcal{E}_{3,i}^c)$ occurs, then Receiver $i\in\{1,2\}$ correctly decodes its desired messages M_0 and M_i. Therefore,

$$\begin{aligned}
&\Pr(\text{error}|\mathcal{M}=1)\\
&\le\Pr\left(\mathcal{E}_0\cup\left(\bigcup_{i=1}^{2}\bigcup_{j=1}^{4}\mathcal{E}_{j,i}\right)\middle|\mathcal{M}=1\right)\\
&\le\Pr(\mathcal{E}_0|\mathcal{M}=1)\\
&\quad+\sum_{i=1}^{2}\big(\Pr(\mathcal{E}_{0i}|\mathcal{E}_0^c,\mathcal{M}=1)+\Pr(\mathcal{E}_{1i}|\mathcal{E}_{0i}^c,\mathcal{M}=1)\\
&\qquad+\Pr(\mathcal{E}_{2i}|\mathcal{E}_{0i}^c,\mathcal{M}=1)+\Pr(\mathcal{E}_{3i}|\mathcal{E}_{0i}^c,\mathcal{M}=1)\big).
\end{aligned}$$

We consider each of the terms separately (see also [14], [8]). 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),
$$\lim_{n\to 0}\Pr(\mathcal{E}_0|\mathcal{M}=1)=0, \tag{7}$$
whenever
$$R_1'+R_2' > I(U_1;U_2|U_0)+\delta(\epsilon). \tag{8}$$
- Since the channel outputs Y_i^n is a $P_{Y_i|X}$-i.i.d. sequence given X^n and by the conditional typicality lemma (Lemma 1),
$$\lim_{n\to 0}\Pr(\mathcal{E}_{0i}|\mathcal{E}_0^c,\mathcal{M}=1)=0. \tag{9}$$
- By the code construction and by the packing lemma (Lemma 3),
$$\lim_{n\to 0}\Pr(\mathcal{E}_{1i}|\mathcal{E}_{0i}^c,\mathcal{M}=1)=0, \tag{10}$$
whenever
$$R_0+R_{1,c}+R_{2,c} < I(U_0,U_i;Y_i)-\delta(\epsilon). \tag{11}$$
- By the code construction and by the packing lemma:
$$\lim_{n\to 0}\Pr(\mathcal{E}_{2i}|\mathcal{E}_{0i}^c,\mathcal{M}=1)=0, \tag{12}$$
whenever
$$R_{1,p}+R_i' < I(U_i;Y_i|U_0)-\delta(\epsilon). \tag{13}$$
- Again, by the code construction and by the packing lemma:
$$\lim_{n\to 0}\Pr(\mathcal{E}_{3i}|\mathcal{E}_{0i}^c,\mathcal{M}=1)=0, \tag{14}$$
whenever
$$R_0+R_{1,c}+R_{2,c}+R_{i,p}+R_i' < I(U_0,U_i;Y_i)-\delta(\epsilon). \tag{15}$$

Thus, we conclude that if for $i\in\{1,2\}$

$$\begin{aligned}
R_1'+R_2' &> I(U_1;U_2|U_0)+\delta(\epsilon) && \text{(16a)}\\
R_0+R_{1,c}+R_{2,c} &< I(U_0,U_i;Y_i)-\delta(\epsilon) && \text{(16b)}\\
R_{i,p}+R_i' &< I(U_i;Y_i|U_0)-\delta(\epsilon) && \text{(16c)}\\
R_0+R_{1,c}+R_{2,c}+R_{i,p}+R_i' &< I(U_0,U_i;Y_i)-\delta(\epsilon), && \text{(16d)}
\end{aligned}$$

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

$$I(U_1;Y_1|U_0)+I(U_2;Y_2|U_0)\ge I(U_1;U_2|U_0) \tag{17}$$

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

$$\begin{aligned}
R_0+R_1 &< I(U_0,U_1;Y_1)-\delta(\epsilon) && \text{(18a)}\\
R_0+R_2 &< I(U_0,U_2;Y_2)-\delta(\epsilon) && \text{(18b)}\\
R_0+R_1+R_2 &< I(U_1;Y_1|U_0)+I(U_2;Y_2|U_0)\\
&\quad+\min_{i=1,2} I(U_0;Y_i)-I(U_1;U_2|U_0)-\delta(\epsilon) && \text{(18c)}\\
2R_0+R_1+R_2 &< I(U_0,U_1;Y_1)+I(U_0,U_2;Y_2)\\
&\quad-I(U_1;U_2|U_0)-\delta(\epsilon) && \text{(18d)}
\end{aligned}$$

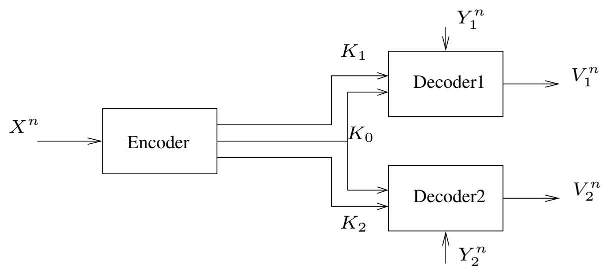

Fig. 4. Lossy Gray-Wyner setup with side-information.

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 (16) holds.

Notice that for every choice of (U_0, U_1, U_2, X) that does not satisfy (17) we can strictly enlarge the rate region (18) 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 (17) 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 (18) is achievable also when (17) is violated.

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

The following two remarks are found useful in the sequel.

Remark 1. *Under conditions* (17) *and* (18) *there exists an associated choice of parameters for our scheme such that the associated auxiliary codewords satisfy*

$$\Pr\Big((U_0^n(M_c), U_1^n(M_c, M_{1,p}, L_1^*), U_2^n(M_c, M_{2,p}, L_2^*)) \in \mathcal{T}_{\epsilon/32}^{(n)}(P_{U_0U_1U_2})\Big) \to 1 \quad \text{as} \quad n \to \infty.$$

Remark 2. *Inspecting the proof, we see that the memoryless channel property has been used only to establish the limit* (9)*. The other limits* (7)*,* (10)*,* (12)*, and* (14) *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* (17) *and* (18)*, 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 $i \in \{1, 2\}$:*

$$\Pr\left((U_0^n(M_c), U_i^n(M_c, M_{i,p}, L_i^*), Y_i^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{U_0U_iY_i})\right) \to 1 \quad \text{as} \quad n \to \infty.$$

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 VI.

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 [16] with additional side-information at the decoders. Our achievable region directly leads to an achievable region for this rate-distortion problem. Special cases of this rate-distortion problem have been considered by Heegard and Berger [17], Tian and Diggavi [18], and Steinberg and Merhav [19], and the lossless counterpart by Timo et al. [20].

A. Setup and Achievable Regions

Our setup is parameterized by the tuple $(\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2, \mathcal{V}_1, \mathcal{V}_2, P_{XY_1Y_2}, P_{V_1|X}, P_{V_2|X}, n)$, where

- $\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2, \mathcal{V}_1, \mathcal{V}_2$ are discrete finite alphabets;
- $P_{XY_1Y_2}$ is a joint probability distribution over the alphabet $\mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2$;
- $P_{V_1|X}$ and $P_{V_2|X}$ are conditional probability distributions over $\mathcal{V}_1$ and $\mathcal{V}_2$ given some random variable $X \in \mathcal{X}$;
- n is the blocklength.

In the following let $\{(X_t, Y_{1,t}, Y_{2,t})\}_{t=1}^n$ be an i.i.d. sequence of triplets of discrete random variables, with marginal distribution $P_{XY_1Y_2}$. 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 $\lambda^{(n)} \colon \mathcal{X}^n \to [2^{nR_0}] \times [2^{nR_1}] \times [2^{nR_2}]$, 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 $\hat{V}_i^n = \Lambda_i^{(n)}(K_0, K_i, Y_i^n)$ by applying a reconstruction function $\Lambda_i^{(n)} \colon [2^{nR_0}] \times [2^{nR_i}] \times \mathcal{Y}_i^n \to \mathcal{V}_i^n$ 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 *ϵ-achievable* if there exists a sequence of encoding and reconstruction functions $(\lambda^{(n)}, \Lambda_1^{(n)}, \Lambda_2^{(n)})$ such that:

$$\Pr\left((X^n, \hat{V}_i^n) \notin \mathcal{T}_\epsilon^n(P_{XV_i})\right) \to 0$$

as $n \to \infty$, for $i \in \{1, 2\}$. A triplet is said to be *achievable* if it is ϵ-achievable for all $\epsilon > 0$. The closure of the set of all achievable rate triplets is denoted $\mathcal{R}_{\text{LGW}}$.

Let $\mathcal{R}_{\text{LGW}}^{\text{inner}}$ be the closure of the set of all nonnegative rate triplets (R_0, R_1, R_2) satisfying

$$R_0 + R_1 > I(X; V_0, V_1|Y_1) \tag{19a}$$

$$R_0 + R_2 > I(X; V_0, V_2|Y_2), \tag{19b}$$

$$R_0 + R_1 + R_2 > I(X; V_1|Y_1, V_0) + I(X; V_2|Y_2, V_0) + \max_{i \in \{1,2\}} I(X; V_0|Y_i) \tag{19c}$$

for some choice of the random variable V_0 such that

$$(V_0, V_1, V_2) -\!\!\circ\!\!- X -\!\!\circ\!\!- (Y_1, Y_2). \tag{20}$$

Theorem 2. $\mathcal{R}_{\text{LGW}}^{\text{inner}} \subseteq \mathcal{R}_{\text{LGW}}$. *Furthermore,* $\mathcal{R}_{\text{LGW}}^{\text{inner}}$ *is convex.*

Proof. Inclusion $\mathcal{R}_{\text{LGW}}^{\text{inner}} \subseteq \mathcal{R}_{\text{LGW}}$ is established in Section V-B. The convexity of $\mathcal{R}_{\text{LGW}}^{\text{inner}}$ is proved in Appendix B. ■

B. Scheme

In this section we describe a scheme achieving the region $\mathcal{R}_{\text{LGW}}^{\text{inner}}$. Our scheme is similar to Heegard and Berger's scheme for the Wyner-Ziv setup with several, differently informed receivers [17, Theorem 2]. However, our scheme also uses the double-binning technique for the common codebook proposed in [18], 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 $\mathcal{V}_0$, $P_{V_0V_1V_2|X}$, $R_{0,0}$, $R_{0,1}$, $R_{0,2}$, $R_{1,0}$, $R_{1,1}$, $R_{2,0}$, $R_{2,2}$, R_0', R_1', R_2', ϵ, n, where

- $\mathcal{V}_0$ is an auxiliary alphabet;
- $P_{V_0V_1V_2|X}$ is a conditional joint probability distribution over $\mathcal{V}_0 \times \mathcal{V}_1 \times \mathcal{V}_2$ given some $X \in \mathcal{X}$ such that its marginals satisfy $\sum_{v_0,v_2} P_{V_0V_1V_2|X}(v_0,v_1,v_2|x) = P_{V_1|X}(v_1|x)$ and $\sum_{v_0,v_1} P_{V_0V_1V_2|X}(v_0,v_1,v_2|x) = P_{V_2|X}(v_2|x)$;
- $R_{0,0}, R_{0,1}, R_{0,2}, R_{1,0}, R_{1,1}, R_{2,0}, R_{2,2} \geq 0$ are nonnegative communication rates;
- $R_0', R_1', R_2' \geq 0$ are nonnegative binning rates, where R_0' cannot be smaller than $\max\{R_{1,0}, R_{2,0}\}$;
- $\epsilon > 0$ is a small number; and
- n is the scheme's blocklength.

1) Codebook Generation: Generate three codebooks $\mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2$ independentely of each other in the following way.

Codebook $\mathcal{C}_0$ consists of $\lfloor 2^{nR_{0,0}} \rfloor$ superbins, each containing $\lfloor 2^{nR_0'} \rfloor$ codewords of length n. All the entries of all the codewords in this codebook $\mathcal{C}_0$ are randomly and independently generated according to the law P_{V_0}.

The codewords in each superbin are partitioned into smaller subbins according to the following two different methods. In the first method each superbin is partitioned into $\lfloor 2^{nR_{1,0}} \rfloor$ subbins, each containing $\lfloor 2^{n(R_0'-R_{1,0})} \rfloor$ codewords, and in the second method it is partitioned into $\lfloor 2^{nR_{2,0}} \rfloor$ subbins, each containing $\lfloor 2^{n(R_0'-R_{2,0})} \rfloor$ codewords. There are thus two different ways to refer to a specific codeword in $\mathcal{C}_0$. When we consider the first partitioning, we denote the codewords in the $k_{1,0} \in [2^{nR_{1,0}}]$-th subbin of superbin $k_{0,0} \in [2^{nR_{0,0}}]$ by $\{v_0^n(1; k_{0,0}, k_{1,0}, \ell_{1,0})\}_{\ell_{1,0}=1}^{\lfloor 2^{n(R_0'-R_{1,0})} \rfloor}$, where we use the first index 1 to indicate that the last two indices refer to the first way of partitioning the superbins. Instead, when we consider the second way of partitioning, we denote the codewords in the $k_{2,0} \in [2^{nR_{2,0}}]$-th subbin of superbin $k_{0,0} \in [2^{nR_{0,0}}]$ by $\{v_0^n(2; k_{0,0}, k_{2,0}, \ell_{2,0})\}_{\ell_{2,0}=1}^{\lfloor 2^{n(R_0'-R_{2,0})} \rfloor}$, where the first index 2 indicates that the last two indices refer to the second way of partitioning the superbins.

The codebooks $\mathcal{C}_1$ and $\mathcal{C}_2$ also consist of nested binning structures. Codebook $\mathcal{C}_i$ consists of $\lfloor 2^{nR_{0,i}} \rfloor$ superbins each containing $\lfloor 2^{nR_{i,i}} \rfloor$ subbins with $\lfloor 2^{nR_i'} \rfloor$ codewords of length n, where all entries of all codewords are randomly and independently drawn according to P_{V_i}. For $k_i \in [2^{nR_{i,i}}]$, we denote the codewords in the k_i-th subbin of superbin $k_{0,i} \in [2^{nR_{0,i}}]$ by $\{v_i^n(k_{0,i}, k_{i,i}, \ell_i)\}_{\ell_i=1}^{\lfloor 2^{nR_i'} \rfloor}$.

All codebooks are revealed to the sender, and codebooks $\{\mathcal{C}_0, \mathcal{C}_i\}$ are revealed to Receiver i, for $i \in \{1,2\}$.

2) LGW-SI Encoder: Given that the encoder observes the source sequence $X^n = x^n$, it searches the codebooks $\mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2$ for a triplet of codewords $v_0^n(1; k_{0,0}, k_{1,0}, \ell_{1,0}) \in \mathcal{C}_0$, $v_1^n(k_{0,1}, k_{1,1}, \ell_1) \in \mathcal{C}_1$, $v_2^n(k_{0,2}, k_{2,2}, \ell_2) \in \mathcal{C}_2$ such that for $i \in \{1,2\}$:

$$(X^n, v_0^n(1; k_{0,0}, k_{1,0}, \ell_{1,0}), v_i^n(k_{0,i}, k_{i,i}, \ell_i)) \in \mathcal{T}_{\epsilon/2}^n(P_{XV_0V_i}). \tag{21}$$

It then forms a list of all tuples of indices $(k_{0,0}, k_{1,0}, \ell_{1,0}, k_{0,1}, k_{1,1}, \ell_1, k_{0,2}, k_{2,2}, \ell_2)$ satisfying (21). If the list is non-empty, the sender chooses one tuple from this list at random. If the list is empty, it randomly chooses a tuple $(k_{0,0}, k_{1,0}, \ell_{1,0}, k_{0,1}, k_{1,1}, \ell_1, k_{0,2}, k_{2,2}, \ell_2)$ from the set $[2^{nR_{0,0}}] \times [2^{nR_{1,0}}] \times [2^{n(R_0'-R_{1,0})}] \times [2^{nR_{0,1}}] \times [2^{nR_{1,1}}] \times [2^{nR_1'}] \times [2^{nR_{0,2}}] \times [2^{nR_{2,2}}] \times [2^{nR_2'}]$. We denote the chosen indices by $k_{0,0}^*, k_{1,0}^*, \ell_{1,0}^*, k_{0,1}^*, k_{1,1}^*, \ell_1^*, k_{0,2}^*, k_{2,2}^*, \ell_2^*$. Also, define $(k_{2,0}^*, \ell_{2,0}^*)$ such that $v_0^n(2; k_{0,0}^*, k_{2,0}^*, \ell_{2,0}^*)$ and $v_0^n(1; k_{0,0}^*, k_{1,0}^*, \ell_{1,0}^*)$ refer to the same codeword in $\mathcal{C}_0$.

The encoder then sends the product message $K_0 = (k_{0,0}^*, k_{0,1}^*, k_{0,2}^*)$ to both receivers, the product message $K_1 = (k_{1,0}^*, k_{1,1}^*)$ to Receiver 1 only, and the product message $K_2 = (k_{2,0}^*, k_{2,2}^*)$ to Receiver 2 only.

3) LGW-SI Decoder: Receiver $i \in \{1,2\}$ first parses the common message K_0 as $(K_{0,0}, K_{0,1}, K_{0,2})$ and its private message K_i as $K_i = (K_{i,0}, K_{i,i})$. Then, given that Receiver i's side-information is $Y_i^n = y_i^n$ and that $K_{0,0} = k_{0,0}$, $K_{0,i} = k_{0,i}$, $K_{i,0} = k_{i,0}$, and $K_{i,i} = k_{i,i}$, Receiver i seeks a codeword $v_0^n(i; k_{0,0}, k_{i,0}, \ell_{i,0})$ in codebook $\mathcal{C}_0$ and a codeword $v_i^n(k_{0,i}, k_{i,i}, \ell_i)$ in codebook $\mathcal{C}_i$ such that $(v_0^n(i; k_{0,0}, k_{i,0}, \ell_{i,0}), v_i^n(k_{0,i}, k_{i,i}, \ell_i), y_i^n) \in \mathcal{T}_\epsilon^n(P_{V_0V_iY_i})$. If exactly one such pair of codewords exists, Receiver i produces as its reconstruction sequence $\hat{V}_i^n = v_i^n(k_{0,i}, k_{i,i}, \ell_i)$. Otherwise, it randomly chooses a triplet $(k_{0,i}', k_{i,i}', \ell_i')$ from the set $[2^{nR_{0,i}}] \times [2^{nR_{i,i}}] \times [2^{nR_i'}]$ and produces as its reconstruction sequence $\hat{V}_i^n = v_i^n(k_{0,i}', k_{i,i}', \ell_i')$.

4) Analysis: We analyze the failure probability $\Pr\left(\mathcal{E}^{(1)} \cup \mathcal{E}^{(2)}\right)$ associated with the above random coding scheme, where $\mathcal{E}^{(i)}$ is the event where Receiver i fails, i.e., where $(X^n, \hat{V}_i^n) \notin \mathcal{T}_\epsilon^n(P_{XV_i})$.

In what follows, let $K_{0,0}^*$, $K_{1,0}^*$, $K_{2,0}^*$, $L_{1,0}^*$, $L_{2,0}^*$, $K_{0,1}^*$, $K_{1,1}^*$, L_1^*, $K_{0,2}^*$, $K_{2,2}^*$, L_2^* be the tuple of indices chosen by the sender. Also, let

- $\mathcal{E}_0$ be the event that $X^n \notin T_{\epsilon/8}^n(P_X)$;
- $\mathcal{E}_1$ be the event that
$$\forall k_{0,0}, k_{1,0}, \ell_{1,0} : \\ \left(X^n, V_0^n(1; k_{0,0}, k_{1,0}, \ell_{1,0})\right) \notin T_{\epsilon/4}^n(P_{XV_0});$$
- $\mathcal{E}_{2,i}$, for $i \in \{1,2\}$, be the event that
$$\forall k_{0,i}, k_{i,i}, \ell_i : \\ (X^n, V_0^n(i; K_{0,0}^*, K_{i,0}^*, L_{i,0}^*), V_i^n(k_{0,i}, k_{i,i}, \ell_i)) \\ \notin T_{\epsilon/2}^n(P_{XV_0V_i});$$

- $\mathcal{E}_{3,i}$, for $i \in \{1,2\}$, be the event that
$$(V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),V_i^n(K_{0,i}^*,K_{i,i}^*,L_i^*),Y_i^n) \notin T_\epsilon^n(P_{V_0V_iY_i});$$
- $\mathcal{E}_{4,i}$, for $i \in \{1,2\}$, be the event that
$$\exists \ell_i \neq L_i^*: (V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),V_i^n(K_{0,i}^*,K_{i,i}^*,\ell_i),Y_i^n) \in T_\epsilon^n(P_{V_0V_iY_i});$$
- $\mathcal{E}_{5,i}$, for $i \in \{1,2\}$, be the event that
$$\exists \ell_{i,0} \neq L_{i,0}^*, \ell_i \neq L_i^*: (V_0^n(i;K_{0,0}^*,K_{i,0}^*,\ell_{i,0}),V_i^n(K_{0,i}^*,K_{i,i}^*,\ell_i),Y_i^n) \in T_\epsilon^n(P_{V_0V_iY_i}).$$

Notice that whenever event $(\mathcal{E}_0^c \cap \mathcal{E}_1^c \cap \mathcal{E}_{2,i}^c)$ occurs, then $(X^n, V_i^n(K_{0,i}^*,K_{i,i}^*,L_i^*)) \in T_\epsilon^n(P_{XV_i})$. If additionally also event $(\mathcal{E}_{3,i}^c \cap \mathcal{E}_{4,i}^c \cap \mathcal{E}_{5,i}^c)$ occurs, then Receiver i produces $\hat{V}_i^n = V_i^n(K_{0,i}^*,K_{i,i}^*,L_i^*)$. Therefore, denoting by $\mathcal{E}^{(i)}$ the event where Receiver i fails, we have:

$$\begin{aligned}\Pr(\mathcal{E}^{(i)}) &\leq \Pr(\mathcal{E}_0 \cup \mathcal{E}_1 \cup \mathcal{E}_{2,i} \cup \mathcal{E}_{3,i} \cup \mathcal{E}_{4,i} \cup \mathcal{E}_{5,i}) \\ &\leq \Pr(\mathcal{E}_0) + \Pr(\mathcal{E}_1|\mathcal{E}_0^c) \\ &\quad + \Pr(\mathcal{E}_{2,i}|\mathcal{E}_1^c) + \Pr(\mathcal{E}_{3,i}|\mathcal{E}_{2,i}^c) \\ &\quad + \Pr(\mathcal{E}_{4,i}) + \Pr(\mathcal{E}_{5,i}). \end{aligned} \tag{22}$$

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:
$$\lim_{n\to\infty} \Pr(\mathcal{E}_0) = 0 \tag{23}$$
- By the code construction and the covering lemma (Lemma 2):
$$\lim_{n\to\infty} \Pr(\mathcal{E}_1|\mathcal{E}_0^c) = 0 \tag{24}$$
whenever
$$R_0' + R_0 > I(X;V_0) + \delta(\epsilon) \tag{25}$$
- Again, by the code construction and the covering lemma:
$$\lim_{n\to\infty} \Pr(\mathcal{E}_{2,i}|\mathcal{E}_1^c) = 0 \tag{26}$$
whenever
$$R_i' + R_{i,i} > I(V_i;X,V_0) + \delta(\epsilon) \tag{27}$$
- The pair $\left(V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),V_i^n(K_{0,i}^*,K_{i,i}^*,L_i^*)\right)$ depends on Y_i^n only through X^n, i.e., the Markov chain
$$V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),V_i^n(K_{0,i}^*,K_{i,i}^*,L_i^*) -\!\!\circ\!\!- X^n -\!\!\circ\!\!- Y_i^n$$
holds. Therefore, Y_i^n is $P_{Y_i|XV_0V_i} = P_{Y_i|X}$-independent given $(X^n,V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),V_i^n(K_{0,i}^*,K_{i,i}^*,L_i^*))$ and by the conditional typicality lemma (Lemma 1):
$$\lim_{n\to\infty} \Pr\left(\mathcal{E}_{3,i}|\mathcal{E}_{2,i}^c\right) = 0. \tag{28}$$
- Notice that the codewords $\{V_i^n(K_{0,i}^*,K_{i,i}^*,\ell_i)\}$ for $\ell_i \in [2^{nR_i'}]\backslash\{L_i^*\}$ are not independent and P_{V_i}-i.i.d.[3] In Appendix A we prove Inequality (29) on top of the next page, where on the right-hand side we have the probability that one of the $\lfloor 2^{nR_i'} \rfloor$ independent and P_{V_i}-i.i.d. codewords $\{V_i^n(1,1,\ell_i)\}_{\ell_i=1}^{\lfloor 2^{nR_i'}\rfloor}$ is jointly ϵ-typical with the pair $(V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),Y_i^n)$. By the packing lemma (Lemma 3) this probability tends to 0 as n tends to ∞ whenever
$$R_i' < I(V_i;V_0,Y_i) - \delta(\epsilon). \tag{31}$$
We thus conclude that
$$\lim_{n\to\infty} \Pr(\mathcal{E}_{4,i}) = 0 \tag{32}$$
whenever (31) holds.
- Following similar steps as in Appendix A, upper bound (30) can be proved. Then, by the multivariate packing lemma (Lemma 4):
$$\lim_{n\to\infty} \Pr(\mathcal{E}_{5,i}) = 0, \tag{33}$$
whenever
$$R_0' - R_{i,0} + R_i' < I(V_0;Y_i) + I(V_i;V_0,Y_i) - \delta(\epsilon) \tag{34}$$

Combining (22) with (25), (27), (31), and (34) we obtain that $\Pr\left(\mathcal{E}^{(i)}\right) \to 0$ as $n \to \infty$ whenever:

$$R_0' + R_{0,0} > I(X;V_0) + \delta(\epsilon) \tag{35a}$$
$$R_1' + R_{0,1} + R_{1,1} > I(V_1;X,V_0) + \delta(\epsilon) \tag{35b}$$
$$R_2' + R_{0,2} + R_{2,2} > I(V_2;X,V_0) + \delta(\epsilon) \tag{35c}$$
$$R_0' - R_{1,0} + R_1' < I(V_0;Y_1) + I(V_1;V_0,Y_1) - \delta(\epsilon) \tag{35d}$$
$$R_0' - R_{2,0} + R_2' < I(V_0;Y_2) + I(V_2;V_0,Y_2) - \delta(\epsilon) \tag{35e}$$
$$R_1' < I(V_1;V_0,Y_1) - \delta(\epsilon) \tag{35f}$$
$$R_2' < I(V_2;V_0,Y_2) - \delta(\epsilon). \tag{35g}$$

We now argue that with an appropriate choice of the auxiliary rates R_0', R_1', R_2', $R_{0,0}$, $R_{0,1}$, $R_{0,2}$, $R_{1,0}$, $R_{1,1}$, $R_{2,0}$, $R_{2,2} > 0$ our scheme achieves the region $\mathcal{R}_{\text{LGW}}^{\text{inner}}$. We first replace $R_{i,i}$ by $R_i - R_{i,0}$, for $i \in \{1,2\}$ and $R_{0,0}$ by $R_0 - R_{0,1} - R_{0,2}$ to obtain

$$R_0' + R_0 - R_{0,1} - R_{0,2} > I(X;V_0) + \delta(\epsilon) \tag{36a}$$
$$R_1' + R_{0,1} + R_1 - R_{1,0} > I(V_1;X,V_0) + \delta(\epsilon) \tag{36b}$$
$$R_2' + R_{0,2} + R_2 - R_{2,0} > I(V_2;X,V_0) + \delta(\epsilon) \tag{36c}$$
$$R_0' - R_{1,0} + R_1' < I(V_0;Y_1) + I(V_1;V_0,Y_1) - \delta(\epsilon) \tag{36d}$$
$$R_0' - R_{2,0} + R_2' < I(V_0;Y_2) + I(V_2;V_0,Y_2) - \delta(\epsilon) \tag{36e}$$
$$R_1' < I(V_1;V_0,Y_1) - \delta(\epsilon) \tag{36f}$$
$$R_2' < I(V_2;V_0,Y_2) - \delta(\epsilon). \tag{36g}$$

[3] This 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 Bern($\frac{1}{2}$). Let $\bar{C}$ be the index in $\{0,1\}$ not equal to C. Notice that $\Pr(A_0 \geq 1.89) = \frac{1}{20}$, whereas $\Pr(A_{\bar{C}} \geq 1.89) = \Pr(A_0 \geq 1.89 \text{ and } A_1 \geq 1.89) = \frac{1}{400}$. Thus, $A_{\bar{C}}$ is not uniform over $[1.7, 1.9]$.

$$\Pr\left(\mathcal{E}_{4,i}\right) \leq \Pr\left(\bigcup_{\ell_i=1}^{\lfloor 2^{nR_i'} \rfloor} (V_0^n(i; K_{0,0}^*, K_{i,0}^*, L_{i,0}^*), V_i^n(1,1,\ell_i), Y_i^n) \in T_\epsilon^n(P_{V_0 V_i Y_i})\right) \tag{29}$$

$$\Pr\left(\mathcal{E}_{5,i}\right) \leq \Pr\left(\bigcup_{\substack{\ell_{i,0}\in\left[2^{n(R_0'-R_{i,0})}\right],\\ \ell_i\in[2^{nR_i'}]}} (V_0^n(i;1,1,\ell_{i,0}), V_i^n(1,1,\ell_i), Y_i^n) \in T_\epsilon^n(P_{V_0 V_i Y_i})\right) \tag{30}$$

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

$$\begin{aligned} R_0 + R_1 &> I(X;V_0) + I(V_1;X,V_0) - I(V_0;Y_1) \\ &\quad -I(V_1;V_0,Y_1) + \delta(\epsilon) && \text{(37a)} \\ R_0 + R_2 &> I(X;V_0) + I(V_2;X,V_0) - I(V_0;Y_2) \\ &\quad -I(V_2;V_0,Y_2) + \delta(\epsilon) && \text{(37b)} \\ R_0 + R_1 + R_2 &> I(X;V_0) + I(V_1;X,V_0) + I(V_2;X,V_0) \\ &\quad -I(V_1;V_0,Y_1) - I(V_2;V_0,Y_2) \\ &\quad -\min_i I(V_0;Y_i) + \delta(\epsilon) && \text{(37c)} \end{aligned}$$

or equivalently (due to the Markov chain (V_0,V_1,V_2)—∘—X—∘—(Y_1,Y_2)), if

$$\begin{aligned} R_0 + R_1 &> I(X;V_0,V_1|Y_1) + \delta(\epsilon) && \text{(38a)} \\ R_0 + R_2 &> I(X;V_0,V_2|Y_2) + \delta(\epsilon) && \text{(38b)} \\ R_0 + R_1 + R_2 &> I(X;V_1|V_0,Y_1) + I(X;V_2|V_0,Y_2) \\ &\quad +\max_i I(X;V_0|Y_i) + \delta(\epsilon) && \text{(38c)} \end{aligned}$$

then there exists a choice of nonnegative rates R_0', R_1', R_2', $R_{0,1}$, $R_{0,2}$, $R_{1,0}$, $R_{2,0}$ that satisfies (36) and

$$\begin{aligned} R_1 - R_{1,0} &\geq 0 \\ R_2 - R_{2,2} &\geq 0 \\ R_0' - R_{1,0} &\geq 0 \\ R_0' - R_{2,0} &\geq 0 \\ R_0 - R_{0,1} - R_{0,2} &\geq 0. \end{aligned}$$

Thus, we conclude that the region (38) is ϵ-achievable for all choices of the auxiliary random variable V_0 satisfying the Markov chain (V_0,V_1,V_2)—∘—X—∘—(Y_1,Y_2). Letting $\epsilon \to 0$, the achievability of $\mathcal{R}_{\text{LGW}}^{\text{inner}}$ is established. The existence of a deterministic coding scheme achieving the same region follows from standard arguments.

The following remark is found useful in the sequel.

Remark 3. *In our error analysis, only Limits* (23) *and* (28) *rely on the assumption that (X^n, Y_1^n, Y_2^n) are $P_{XY_1Y_2}$-i.i.d. It is easy to check that replacing this assumption with the more general assumptions*

(i) $\Pr(X^n \in \mathcal{T}_{\epsilon/8}^n(P_X)) \to 1$ *as* $n \to \infty$.
(ii) (Y_1^n, Y_2^n) *is* $P_{Y_1Y_2|X}$*-independent given* X^n.

still guarantees the existence of associated parameters such that the scheme above ϵ-achieves the region (38). In particular,

$$\Pr\left((X^n, V_i^n(K_{0,i}^*, K_{i,i}^*, L_i^*)) \notin T_\epsilon^n(P_{XV_i})\right) \to 0$$

and

$$\Pr\left(\hat{V}_i^n \neq V_i^n(K_{0,i}^*, K_{i,i}^*, L_i^*)\right) \to 0,$$

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

VI. Main Result

A. Achievable Region

Consider a DMBC with generalized feedback given by $\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2, \tilde{\mathcal{Y}}, P_{Y_1Y_2\tilde{Y}|X}$. Let $\mathcal{R}_{\text{inner}}$ be the closed convex hull of the set of all nonnegative triplets (R_0, R_1, R_2) that satisfy Inequalities (41) shown on top of the next page, for some choice of auxiliary random variables $(U_0,U_1,U_2,V_0,V_1,V_2)$ and function f such that $X = f(U_0,U_1,U_2)$,

$$(V_0,V_1,V_2)—∘—(U_0,U_1,U_2,\widetilde{Y})—∘—(Y_1,Y_2) \tag{39}$$

and

$$(U_0,U_1,U_2)—∘—X—∘—(Y_1,Y_2,\widetilde{Y}) \tag{40}$$

form Markov chains, and $(Y_1,Y_2,\tilde{Y}) \sim P_{Y_1Y_2\tilde{Y}|X}$.

Notice that for noise-free feedback where $\tilde{Y} = (Y_1,Y_2)$ the Markov chain (39) is satisfied for any choice of the auxiliary random variables $(U_0,U_1,U_2,V_0,V_1,V_2)$.

Theorem 3. $\mathcal{R}_{\text{inner}} \subseteq \mathcal{C}_{\text{GenFB}}$.

The proof of the theorem is given in Subsection VI-B. A few remarks are in order:

Remark 4. *The region* $\mathcal{R}_{\text{inner}}$ *includes* $\mathcal{R}_{\text{Marton}}$*, because when for a given choice of* (U_0,U_1,U_2)*, constraints* (41) *are specialized to* $(V_0,V_1,V_2) = \text{const}$*, then it results in the Marton region* (4)*. The inclusion is also clear from the construction of our scheme in Subsection VI-B ahead.*

Remark 5. *In our coding scheme we can allow f to be a* randomized *function. In this case, the scheme achieves the region* $\mathcal{R}_{\text{inner}}$ *but where the input X can be an arbitrary random variable satisfying the Markov chain* (40)*. It is not clear whether this results in an improved region compared to* $\mathcal{R}_{\text{inner}}$.

Remark 6. *Recall that for fixed finite alphabets, the Shannon information measures are continuous (say w.r.t. Euclidean*

$$R_0 + R_1 \leq I(U_0, U_1; Y_1, V_1) - I(U_0, U_1, U_2, \widetilde{Y}; V_0, V_1 | Y_1) \tag{41a}$$
$$R_0 + R_2 \leq I(U_0, U_2; Y_2, V_2) - I(U_0, U_1, U_2, \widetilde{Y}; V_0, V_2 | Y_2) \tag{41b}$$
$$\begin{aligned} R_0 + R_1 + R_2 \leq\ & I(U_1; Y_1, V_1 | U_0) + I(U_2; Y_2, V_2 | U_0) + \min_{i \in \{1,2\}} I(U_0; Y_i, V_i) - I(U_1; U_2 | U_0) \\ & - I(U_0, U_1, U_2, \widetilde{Y}; V_1 | V_0, Y_1) - I(U_0, U_1, U_2, \widetilde{Y}; V_2 | V_0, Y_2) - \max_{i \in \{1,2\}} I(U_0, U_1, U_2, \widetilde{Y}; V_0 | Y_i) \end{aligned} \tag{41c}$$
$$\begin{aligned} 2R_0 + R_1 + R_2 \leq\ & I(U_1 U_0; Y_1, V_1) + I(U_2, U_0; Y_2, V_2) - I(U_1; U_2 | U_0) \\ & - I(U_0, U_1, U_2, \widetilde{Y}; V_0, V_1 | Y_1) - I(U_0, U_1, U_2, \widetilde{Y}; V_0, V_2 | Y_2) \end{aligned} \tag{41d}$$

---

distance) in the joint distribution [23]. Fix the channel's input, output, and feedback alphabets. Then for any fixed choice of $(P_{U_0 U_1 U_2}, f, P_{V_0 V_1 V_2 | U_0 U_1 U_2 \widetilde{Y}})$, the quantities on the right-hand side of Inequalities (41) are continuous in $P_{Y_1 Y_2 \widetilde{Y} | X}$.

Remark 7. *By the previous remark, the following conclusion holds for any DMBC $P_{Y_1 Y_2 | X}$ with feedback alphabet $\widetilde{\mathcal{Y}} = \mathcal{Y}_1 \times \mathcal{Y}_2$. Assume that the region $\mathcal{R}_{\text{inner}}$ associated with noiseless feedback (i.e., $\tilde{Y} = (Y_1, Y_2)$) strictly contains $\mathcal{C}_{\text{NoFB}}$. Now, if we consider a* noisy *feedback channel $P_{\widetilde{Y} | X Y_1 Y_2}$ that is close enough to the noiseless feedback (i.e., $\tilde{Y}$ close to (Y_1, Y_2)), then also the region $\mathcal{R}_{\text{inner}}$ associated with this noisy feedback strictly contains $\mathcal{C}_{\text{NoFB}}$.*

B. Scheme achieving $\mathcal{R}_{\text{innner}}$

Our scheme combines Marton's no-feedback scheme of Section IV-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 5. 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 $n' = \gamma n$ for $\gamma > 1$. We denote the input/output/feedback sequences in Block $b \in [B]$ by $X^n_{(b)}, Y^n_{i,(b)}, \widetilde{Y}^n_{(b)}$, respectively, and the input/output sequences in Block $B+1$ by $X^{n'}_{(B+1)}, Y^{n'}_{i,(B+1)}$. 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 $[2^{nR_i}]$. The effective rates of transmission are thus

$$\left(\frac{B}{B+\gamma} R_0, \frac{B}{B+\gamma} R_1, \frac{B}{B+\gamma} R_2 \right) \tag{42}$$

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 $Y^n_{i,(b)}$. This facilitates the 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-informations" consist 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 $Y^{n'}_{i,(B+1)}$. Denote its guess by $\hat{K}_{0,i,(B)}, \hat{K}_{i,(B)}$. Then, for each block $b \in [B]$, starting from block B and going backwards, it performs the following steps:

1) Using $(\hat{K}_{0,b}, \hat{K}_{i,b})$, it "improves" its block-b outputs $Y^n_{i,(b)}$.
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 $(\hat{M}_{0,(b)}, \hat{M}_{i,(b)})$ and $(\hat{K}_{0,i,(b-1)}, \hat{K}_{i,(b-1)})$.

We now describe the coding scheme in more detail. Our scheme has parameters ($\mathcal{U}_0, \mathcal{U}_1, \mathcal{U}_2, \mathcal{V}_0, \mathcal{V}_1, \mathcal{V}_2, P_{U_0 U_1 U_2}, f, P_{V_0 V_1 V_2 | U_0 U_1 U_2 \widetilde{Y}}, R_0, R_1, R_2, \bar{R}'_1, \bar{R}'_2, \tilde{R}_0, \tilde{R}_1, \tilde{R}_2, \tilde{R}'_0, \tilde{R}'_1, \tilde{R}'_2, \epsilon, \gamma, n, B$), where:

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

Our scheme is of rates R_0, R_1, R_2 and of blocklength $N = Bn + \gamma n$. Before the transmission starts, we divide the blocklength N into $B+1$ blocks: the first B blocks are of length n and the last block is of length $n' \overset{\text{def}}{=} \gamma n$. We also construct the following codes.

For each block $b \in [B]$ we construct a Marton code for a DMBC with parameters $(\mathcal{X}, \mathcal{Y}_1 \times \mathcal{V}_1, \mathcal{Y}_2 \times \mathcal{V}_2, P_{(Y_1 V_1)(Y_2 V_2) | X})$ using the code construction in Subsection IV-B1. As parameters of this construction we choose:

- the auxiliary alphabets $\mathcal{U}_0, \mathcal{U}_1, \mathcal{U}_2$;
- the joint law P_{U_0, U_1, U_2} over these alphabets;
- the function $f : \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$;
- the nonnegative communication rates $\bar{R}_0$, $\bar{R}_{1,p}$, $\bar{R}_{2,p}$, $\bar{R}_{1,c}$, $\bar{R}_{2,c}$ where we require that $\bar{R}_0 = R_0 + \tilde{R}_0$, $\bar{R}_{1,p} + \bar{R}_{1,c} = R_1 + \tilde{R}_1$, and $\bar{R}_{2,p} + \bar{R}_{2,c} = R_2 + \tilde{R}_2$;

Fig. 5. Block-Markov strategy of our feedback-scheme.

- the nonnegative binning rates $\bar{R}'_1, \bar{R}'_2$;
- the small number ϵ; and
- the blocklength n.

For block $B+1$, we use a Marton scheme for the DMBC with parameters $(\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2, P_{Y_1Y_2|X})$ of block length n' where the scheme is chosen as to achieve the rate triplet $(\gamma^{-1}\widetilde{R}_0, \gamma^{-1}\widetilde{R}_1, \gamma^{-1}\widetilde{R}_2)$. To make sure that such a scheme exists, we assume throughout the proof that the single-user channels $P_{Y_1|X}$ and $P_{Y_2|X}$ both have positive capacities.[4] 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 $X^n_{(b)}, Y^n_{1,(b)}, Y^n_{2,(b)}$ and $\widetilde{Y}^n_{(b)}$ denote the corresponding blocks of channel inputs, channel outputs, and feedback outputs, respectively.

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

- the source alphabet $(\mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \tilde{\mathcal{Y}})$;
- the decoder side-information alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$;
- the reconstruction alphabets $\mathcal{V}_1$ and $\mathcal{V}_2$;
- the source-side-information law $P_{(U_0U_1U_2\tilde{Y})Y_1Y_2}$; and
- the reconstruction laws $P_{V_1|U_0U_1U_2\tilde{Y}}(v_1|u_0,u_1,u_2,\tilde{y}) = \sum_{v_0,v_2} P_{V_0V_1V_2|U_0U_1U_2\tilde{Y}}(v_0,v_1,v_2|u_0,u_1,u_2,\tilde{y})$ and $P_{V_2|U_0U_1U_2\tilde{Y}}(v_2|u_0,u_1,u_2,\tilde{y}) = \sum_{v_0,v_1} P_{V_0V_1V_2|U_0U_1U_2\tilde{Y}}(v_0,v_1,v_2|u_0,u_1,u_2,\tilde{y})$.

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 $\mathcal{V}_0$;
- the conditional law $P_{V_0V_1V_2|U_0U_1U_2\tilde{Y}}$;
- the nonnegative rates $\tilde{R}_{0,0}$, $\tilde{R}_{0,1}$, $\tilde{R}_{0,2}$, $\tilde{R}_{1,0}$, $\tilde{R}_{1,1}$, $\tilde{R}_{2,0}$, $\tilde{R}_{2,2}$, $\tilde{R}'_0$, $\tilde{R}'_1$, $\tilde{R}'_2$;
- the binning rates $\tilde{R}'_0, \tilde{R}'_1, \tilde{R}'_2 \geq 0$, where $\tilde{R}'_0$ cannot be smaller than $\max\{\tilde{R}_{1,0}, \tilde{R}_{2,0}\}$;
- the small number $\epsilon/2$; and
- the blocklength n.

In what follows, let $\lambda_{(b)}$, $\Lambda_{1,(b)}$, and $\Lambda_{2,(b)}$ denote the LGW-SI encoding and decoding rules corresponding to these codes.

We are now ready to describe the scheme in detail.

1) Encoding: In the first block $b=1$, the transmitter forms the product messages $J_{0,(1)} \overset{\text{def}}{=} (M_{0,(1)}, 1)$, $J_{1,(1)} \overset{\text{def}}{=} (M_{1,(1)}, 1)$, and $J_{2,(1)} \overset{\text{def}}{=} (M_{2,(1)}, 1)$, and applies the Marton encoding rule $\varphi_{(1)}$ to this triplet $J_{0,(1)}, J_{1,(1)}, J_{2,(1)}$.

In blocks $b \in 2,\ldots,B$ the transmitter first applies the LGW-SI encoding function $\lambda_{(b-1)}$ to its "source" sequence $(U^n_{0,(b-1)}, U^n_{1,(b-1)}, U^n_{2,(b-1)}, \tilde{Y}^n_{(b-1)})$ to generate the update messages $(K_{0,(b-1)}, K_{1,(b-1)}, K_{2,(b-1)})$. (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 $J_{i,(b)} \overset{\text{def}}{=} (M_{i,(b)}, K_{i,(b-1)})$, and encodes them via the Marton encoding rule $\varphi^{(n)}_{(b)}$. 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 $\lambda_{(B)}$ to the sequences $(U^n_{0,(B)}, U^n_{1,(B)}, U^n_{2,(B)}, \tilde{Y}^n_{(B)})$ to generate the update messages $K_{i,(B)}$, for $i \in \{0,1,2\}$. It then forms the tuple $J_{0,(B+1)} \overset{\text{def}}{=} (1, K_{0,(B)})$, $J_{1,(B+1)} \overset{\text{def}}{=} (1, K_{1,(B)})$, and $J_{2,(B+1)} \overset{\text{def}}{=} (1, K_{2,(B)})$ and encodes them via the Marton encoding rule $\varphi_{(B+1)}$.

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

Now, for every block $b \in \{2,\ldots,B\}$, starting with block B and going backwards, the receiver performs the following steps. It applies the LGW-SI decoder $\Lambda_{i,(b)}$ to its guess of the update messages $(\hat{K}_{0,i,(b)}, \hat{K}_{i,(b)})$ obtained in block $b+1$, and to its "side-information" $Y^n_{i,(b)}$. It then applies Marton's decoding rule $\Phi_{i,(b)}$ to the pair $(Y^n_{i,(b)}, \hat{V}^n_{i,(b)})$, where $\hat{V}^n_{i,(b)}$ denotes the reconstruction sequence produced by the LGW-SI decoder $\Lambda_{i,(b)}$. Finally, it parses the guess produced by Marton's decoding rule $(\hat{J}_{0,i,(b)}, \hat{J}_{i,(b)})$ as $\hat{J}_{0,i,(b)} = (\hat{M}_{0,i,(b)}, \hat{K}_{0,i,(b)})$ and $\hat{J}_{i,(b)} = (\hat{M}_{i,(b)}, \hat{K}_{i,(b)})$.

In the last processed block $b=1$, the receiver applies the LGW-SI decoder $\Lambda_{i,(1)}$ to its guess of the update messages $(\hat{K}_{0,i,(1)}, \hat{K}_{1,(1)})$ produced in block 2 and to its "side-information" $Y^n_{i,(1)}$. It then applies Marton's decoding rule $\Phi_{i,(b)}$ to the pair $(Y^n_{i,(1)}, \hat{V}^n_{i,(1)})$ to decode the messages $(J_{0,(1)}, J_{i,(1)})$, and finally parses the produced guesses as $(\hat{J}_{0,(1)}, \hat{J}_{i,(1)})$ as $\hat{J}_{0,i,(1)} = (\hat{M}_{0,i,(1)}, 1)$ and $\hat{J}_{i,(1)} = (\hat{M}_{i,(1)}, 1)$.

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

3) Error Analysis: We bound the average probability of error (where the average is over the random messages, codes,

[4]When one of the two single-user channels has capacity 0, then the broadcast problem is not very interesting. In fact, in this case both the capacity regions with noiseless feedback and with no-feedback are degenerate.

and channel realizations). Let $\mathcal{E}$ be the error event:

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

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

$$\mathcal{F}_b \stackrel{\text{def}}{=} \bigcup_{i=1}^{2} \left\{ (\widehat{J}_{0,i,(b)}, \widehat{J}_{i,(b)}) \neq \left(J_{0,(b)}, J_{i,(b)}\right) \right\}.$$

Then,

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

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

$$I(U_1; Y_1, V_1 | U_0) + I(U_2; Y_2, V_2 | U_0) \geq I(U_1; U_2 | U_0); \quad (43)$$

and

$$\bar{R}_0 + \bar{R}_1 < I(U_0, U_1; Y_1, V_1) - \delta(\epsilon) \quad (44a)$$

$$\bar{R}_0 + \bar{R}_2 < I(U_0, U_2; Y_2, V_2) - \delta(\epsilon) \quad (44b)$$

$$\begin{aligned} \bar{R}_0 + \bar{R}_1 + \bar{R}_2 < {} & I(U_1; Y_1, V_1 | U_0) + I(U_2; Y_2, V_2 | U_0) \\ & + \min_i I(U_0; Y_i, V_i) - I(U_1; U_2 | U_0) - \delta(\epsilon); \end{aligned} \quad (44c)$$

$$\begin{aligned} 2\bar{R}_0 + \bar{R}_1 + \bar{R}_2 < {} & I(U_0, U_1; Y_1, V_1) + I(U_0, U_2; Y_2, V_2) \\ & - I(U_1; U_2 | U_0) - \delta(\epsilon); \end{aligned} \quad (44d)$$

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

$$\Pr((U_{0,(b)}^n, U_{i,(b)}^n, Y_{i,(b)}^n, \hat{V}_{i,(b)}^n) \notin \mathcal{T}_\epsilon^n(P_{U_0 U_i Y_i, V_i})) \to 0 \quad (45)$$

as $n \to \infty$, then there exists a choice of the parameters such that $\Pr\left(F_b | F_{b+1}^{\mathrm{c}}\right) \to 0$ as $n \to \infty$. From this point forward we assume that conditions (43) and (44) hold. In the following we wish to prove that if additionally

$$\widetilde{R}_0 + \widetilde{R}_1 > I(U_0, U_1, U_2, \widetilde{Y}; V_0, V_1 | Y_1) + \delta(\epsilon) \quad (46a)$$

$$\widetilde{R}_0 + \widetilde{R}_2 > I(U_0, U_1, U_2, \widetilde{Y}; V_0, V_2 | Y_2) + \delta(\epsilon) \quad (46b)$$

$$\begin{aligned} \widetilde{R}_0 + \widetilde{R}_1 + \widetilde{R}_2 > {} & I(U_0, U_1, U_2, \widetilde{Y}; V_1 | V_0, Y_1) \\ & + I(U_0, U_1, U_2, \widetilde{Y}; V_2 | V_0, Y_2) \\ & + \max_i I(U_0, U_1, U_2, \widetilde{Y}; V_0 | Y_i) + \delta(\epsilon) \end{aligned} \quad (46c)$$

then the limit (45) holds. We notice that

$$\begin{aligned} & \Pr\Big((U_{0,(b)}^n, U_{i,(b)}^n, Y_{i,(b)}^n, \hat{V}_{i,(b)}^n) \notin \mathcal{T}_\epsilon^n(P_{U_0 U_i Y_i V_i})\Big) \\ & \quad \leq \Pr\Big((U_{0,(b)}^n, U_{i,(b)}^n, Y_{i,(b)}^n, V_{i,(b)}^n) \notin \mathcal{T}_\epsilon^n(P_{U_0 U_i Y_i V_i})\Big) \\ & \qquad + \Pr\Big(\widehat{V}_{i,(b)}^n \neq V_{i,(b)}^n\Big), \end{aligned} \quad (47)$$

where $V_{1,(b)}^n$ and $V_{2,(b)}^n$ denote the codewords chosen by the LGW-SI encoding rule $\lambda_{(b)}$. We shall now verify that under conditions (46), both terms on the right-hand side of (47) vanish as $n \to \infty$.

Since the input $X_{(b)}^n$ is a component-wise function of $(U_{0,(b)}^n, U_{1,(b)}^n, U_{2,(b)}^n)$, since the Markov condition $(\widetilde{Y}_{(b)}^n) \multimap X_{(b)}^n \multimap (U_{0,(b)}^n, U_{1,(b)}^n, U_{2,(b)}^n)$ holds, and since the channel law is memoryless, $(\widetilde{Y}_{(b)}^n)$ is $P_{\widetilde{Y}|U_0 U_1 U_2}$-independent given $(U_{0,(b)}^n, U_{1,(b)}^n, U_{2,(b)}^n)$. Furthermore, from Marton's code construction and in light of Remark 1, we have that under conditions (43) and (44)

$$\Pr\left((U_{0,(b)}^n, U_{1,(b)}^n, U_{2,(b)}^n) \notin \mathcal{T}_{\epsilon/32}(P_{U_0 U_1 U_2})\right) \to 0.$$

Therefore, by the conditional typicality Lemma, also

$$\Pr\left((U_{0,(b)}^n, U_{1,(b)}^n, U_{2,(b)}^n, \widetilde{Y}_{(b)}^n) \notin \mathcal{T}_{\epsilon/16}(P_{U_0 U_1 U_2 \widetilde{Y}})\right) \to 0$$

as $n \to \infty$. Moreover, by similar arguments, $(Y_{1,(b)}^n, Y_{2,(b)}^n)$ is $P_{Y_1 Y_2 | U_0 U_1 U_2 \widetilde{Y}}$-independent given $U_{0,(b)}^n, U_{1,(b)}^n, U_{2,(b)}^n, \widetilde{Y}_{(b)}^n$. We conclude that the conditions of the modified LGW-SI setup of Remark 3 are satisfied (recall we have used the parameter $\epsilon/2$ for the LGW-SI code). Therefore, by that same remark and under conditions (46) it holds that

$$\Pr\Big(\widehat{V}_{i,(b)}^n \neq V_{i,(b)}^n\Big) \to 0 \quad (48)$$

$$\Pr\Big((U_{0,(b)}^n, U_{1,(b)}^n, U_{2,(b)}^n, \widetilde{Y}_{(b)}^n, V_{i,(b)}^n) \notin \mathcal{T}_{\epsilon/2}^n(P_{U_0 U_1 U_2 \widetilde{Y} V_i})\Big) \to 0 \quad (49)$$

as $n \to \infty$. We note that $Y_{i,(b)}^n$ is $P_{Y_i | U_0 U_1 U_2 \widetilde{Y}}$-independent given $(U_{0,(b)}^n, U_{1,(b)}^n, U_{2,(b)}^n, \widetilde{Y}_{(b)}^n)$, and the Markov condition

$$V_{i,(b)}^n \multimap (U_{0,(b)}^n, U_{1,(b)}^n, U_{2,(b)}^n, \widetilde{Y}_{(b)}^n) \multimap Y_{i,(b)}^n$$

holds. Together with (49) and the conditional typicality Lemma, this implies that

$$\Pr\Big((U_{0,(b)}^n, U_{i,(b)}^n, Y_{i,(b)}^n, V_{i,(b)}^n) \notin \mathcal{T}_\epsilon^n(P_{U_0 U_i Y_i V_i})\Big) \to 0 \quad (50)$$

Combining (47), (48) and (50), the limit in (45) is established.

Combining Constraints (44) and (46), replacing $\bar{R}_i$ by $R_i + \tilde{R}_i$, and employing the Fourier-Motzkin elimination algorithm, we obtain that under the set of constraints (41) and when (43) holds, then there exists a choice of the parameters such that the probability of error of our scheme tends to 0 as $n \to \infty$, for any ϵ small enough.

Whenever (43) is not satisfied, then the rate region (41) is strictly enlarged if the random triple (U_0, U_1, U_2) is replaced by the triple (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') moreover satisfies (43) because both sides are 0. It also satisfies the Markov chain (39) and X can be expressed as a function of the new auxiliaries U_0', U_1', U_2'.

Thus, since by (42) the effective rates of transmission tend to (R_0, R_1, R_2) as $B \to \infty$, any rate triplet satisfying the constraints (41) is achievable by our scheme.

The existence of a deterministic coding scheme achieving the same region 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

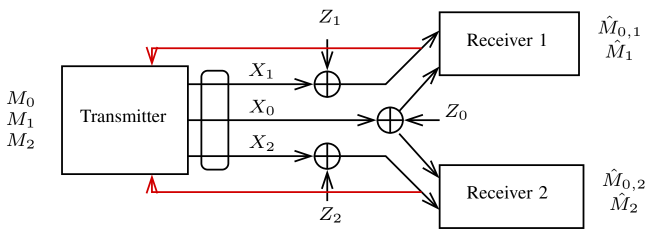

Fig. 6. Generalization of Dueck's DMBC with feedback example.

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

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

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 noiseless 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*

$$R_0 + R_1 \leq 1 \qquad \text{and} \qquad R_0 + R_2 \leq 1. \tag{51}$$

Proof. The converse follows from the cutset bound. The achievability 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 $\{X_{1,t}\}_{t=1}^n$ and $\{X_{2,t}\}_{t=1}^n$, respectively. Additionally, for $t = 2, \ldots, (n+1)$, it repeats the previous noise 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 $\{Z_t\}_{t=1}^n$ from its sequence of outputs $\{Y_{i,0,t}\}_{t=2}^{n+1}$. Receiver i can thus compute the channel inputs $X_{i,t} = Y_{i,i,t} - Z_t$, for $t = 1, \ldots, n$, and recover the desired pair of messages (M_0, 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 6. We assume that all three binary channels are (possibly) noisy, and the first and third channels are corrupted by (possibly) 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, $Y_1 = (Y_{1,1}, Y_{1,0})$ and $Y_2 = (Y_{2,0}, Y_{2,2})$. However, now,

$$\begin{aligned} Y_{1,0} = Y_{2,0} &= X_0 \oplus Z_0, \\ Y_{1,1} &= X_1 \oplus Z_1, \\ Y_{2,2} &= X_2 \oplus Z_2, \end{aligned}$$

where Z_0, Z_1, Z_2 are binary random variables of a given joint law $P_{Z_0 Z_1 Z_2}$. We restrict attention to laws $P_{Z_0 Z_1 Z_2}$ such that

$$H(Z_0, Z_1) \leq 1 \qquad \text{and} \qquad H(Z_0, Z_2) \leq 1. \tag{52}$$

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

$$R_0 + R_1 \leq 2 - H(Z_0, Z_1), \tag{53a}$$

$$R_0 + R_2 \leq 2 - H(Z_0, Z_2), \tag{53b}$$

$$R_0 + R_1 + R_2 \leq 3 - H(Z_0, Z_1, Z_2) - I(Z_1; Z_2|Z_0). \tag{53c}$$

Proof. The no-feedback capacity of a DMBC depends on the channel law $P_{Y_1Y_2|X}(y_1, y_2|x)$ only through the marginal laws $P_{Y_1|X}(y_1|x)$ and $P_{Y_2|X}(y_2|x)$ (see e.g., [22]). We therefore assume in the following that $Z_2 -\!\!\circ\!\!- Z_0 -\!\!\circ\!\!- Z_1$. 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 (18) (where we let $\epsilon \to 0$) we choose U_0, U_1, U_2 to be i.i.d. Bern$(1/2)$ and $X_i = U_i$, for $i \in \{0,1,2\}$, then it evaluates to our region in (53). (Notice that since we choose U_0, U_1, U_2 independent, constraint (18d) on $2R_0 + R_1 + R_2$ is not active.) ∎

Our scheme in Section VI-B allows us to obtain the following result for the Generalized Dueck DMBC with noiseless feedback.

Theorem 5. *Under condition* (52) *and when no common message is sent, i.e.,* $R_0 = 0$*, the noiseless-feedback capacity of the Generalized Dueck DMBC is the set of all nonnegative rate pairs* (R_1, R_2) *satisfying*

$$R_1 \leq 2 - H(Z_0, Z_1), \tag{54a}$$

$$R_2 \leq 2 - H(Z_0, Z_2), \tag{54b}$$

$$R_1 + R_2 \leq 3 - H(Z_0, Z_1, Z_2). \tag{54c}$$

Proof. The converse follows from the cutset bound. The direct part follows from Theorem 3 by taking the convex hull of the achievable regions that result when (41) 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 (41d) is subsumed by Constraints (41a) and (41b).) ∎

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

Corollary 1. *If the triplet* (Z_0, Z_1, Z_2) *satisfies* (52) *and does not form the Markov chain* $Z_1 - Z_0 - Z_2$*, then noiseless feedback strictly increases the capacity of our Generalized Dueck DMBC.*

Let's briefly consider the case of noisy feedback $\widetilde{Y} = (Y_{1,1} \oplus W_1, Y_{1,0} \oplus W_0, Y_{2,2} \oplus W_2)$ 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 7, we can conclude the following.

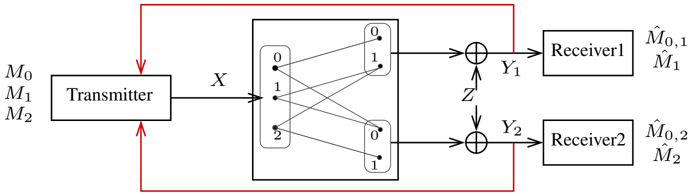

Fig. 7. A noisy version of Blackwell's DMBC with noiseless feedback.

Remark 8. *If the noise triplet* (Z_0, Z_1, Z_2) *satisfies* (52) *and does not form the Markov chain* $Z_1 - Z_0 - Z_2$*, 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 [7] in Figure 7. 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 $Z \sim \text{Bern}(p)$, with $p < \frac{1}{2}$, be independent of X. The channel law $P_{Y_1Y_2|X}$ is described as follows.

$$Y_1 = \begin{cases} Z & X = 0 \\ 1-Z & X = 1,2 \end{cases} \qquad Y_2 = \begin{cases} Z & X = 0,1 \\ 1-Z & X = 2 \end{cases} \tag{55}$$

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 $\mathcal{R}_{\text{Inner}}$ (41) in Theorem 3.

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

$P_{U_1U_2|U_0=0}$:

	$U_2 = 0$	$U_2 = 1$
$U_1 = 0$	α	0
$U_1 = 1$	$1-\alpha-\beta$	β

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

$P_{U_1U_2|U_0=1}$:

	$U_2 = 0$	$U_2 = 1$
$U_1 = 0$	β	0
$U_1 = 1$	$1-\alpha-\beta$	α

Set $X \stackrel{\text{def}}{=} U_1 + U_2$ (real addition), and let $V_1 \stackrel{\text{def}}{=} U_1$, $V_2 \stackrel{\text{def}}{=} U_2$, and $V_0 \stackrel{\text{def}}{=} V_1 \oplus Y_1 = Z$. Evaluating the region in (41) for this choice of random variables, we obtain the following theorem.

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

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

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

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* (56) *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 $P_{Y_1|X}, P_{Y_2|X}$ [22], the capacity region for the Noisy Blackwell channel remains the same if in the definitions of Y_1 and Y_2 (see (55)) we replace Z by independent $\text{Bern}(p)$ 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

$$\begin{aligned}
&R_1 + R_2 \\
&\leq \sup_{\alpha \in (0, \frac{1}{2})} \Big\{ H\big([\alpha(p-\bar{p})^2 + p\bar{p}, \bar{p}^2 + 2\alpha\bar{p}(p-\bar{p}), \\
&\qquad p^2 + 2\alpha p(\bar{p}-p), \alpha(p-\bar{p})^2 + p\bar{p}]\big) \Big\} - 2h_b(p),
\end{aligned} \tag{57}$$

where $\bar{p} \stackrel{\text{def}}{=} 1-p$. Figure 8 depicts the bounds (56) and (57) together with a cut-set upper bound on the sum-rates $R_1 + R_2$ that are achievable with noiseless feedback. By this Figure 8:

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

Remark 9. *Let* $\widetilde{Y} = (Y_1 \oplus W_1, Y_2 \oplus W_2)$*, 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 7, 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.*

$$R_1 + R_2 \geq \sup_{\substack{\alpha,\beta\geq 0:\\ \alpha+\beta\leq 1}} \min\Bigg\{ h_b\left(\left(\frac{\alpha+\beta}{2}\right)\star p\right) + \frac{1-\beta}{2}h_b\left(\frac{\alpha}{1-\beta}\right) + \frac{1-\alpha}{2}h_b\left(\frac{\beta}{1-\alpha}\right) - h_b(p),$$
$$2h_b\left(\left(\frac{\alpha+\beta}{2}\right)\star p\right) + H\left([\alpha,\beta,1-\alpha-\beta]\right) - h_b(\alpha) - h_b(\beta) - 2h_b(p)\Bigg\} \quad (56)$$

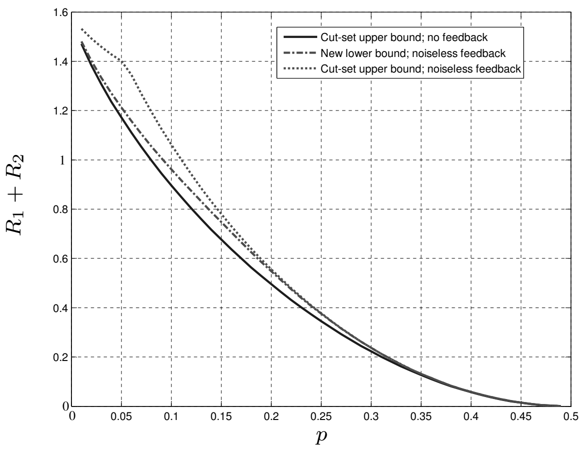

Fig. 8. Bounds on the maximum sum-rates $R_1 + R_2$ that are achievable over the Noisy Blackwell DMBC with no feedback and noiseless feedback.

ACKNOWLEDGEMENT

The authors thank the Associate Editor and the reviewers for their valuable comments. In particular for pointing them to [18].

APPENDIX A
PROOF OF INEQUALITY (29)

Our proof of Inequality (29) is similar to a proof in [14, Appendix 12A].

By the symmetry of the code construction, the probability that the codewords $\{V_i^n(K_{0,i}^*, K_{i,i}^*, \ell_i)\}_{\ell_i\neq L_i^*}$ are jointly typical with the pair $\left(V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*), Y_i^n\right)$ does not depend on the specific value of $K_{0,i}^*, K_{i,i}^*, L_i^*$. Therefore, Equality (58) on top of the next page holds. In the following we show Inequality (59), which combined with (58) establishes the desired inequality (29).

To this end we shall prove that Inequality (60) holds for all triplets $(k_{0,i}^*, k_{i,i}^*, \ell_i^*) \in [2^{nR_{0,i}}] \times [2^{nR_{i,i}}] \times [2^{nR_i'}]$. Since by the symmetry of our construction the triplet $\left(K_{0,i}^*, K_{i,i}^*, L_i^*\right)$ is independent and uniformly distributed, this proves (59).

We notice that by the symmetry of the code construction, the right-hand side of (60) does not depend on the value of ℓ_i^*. Moreover, the right-hand side of (60) is the same for all triplets $(k_{0,i}^*, k_{i,i}^*, \ell_i^*)$ where $(k_{0,i}^*, k_{i,i}^*) \neq (1,1)$. Thus, it suffices to prove (60) for the tuples $(k_{0,i}^*, k_{i,i}^*, \ell_i^*) = (1,1,1)$ and $(k_{0,i}^*, k_{i,i}^*, \ell_i^*) = (2,2,1)$. That (60) holds for $(k_{0,i}^*, k_{i,i}^*, \ell_i^*) = (1,1,1)$ follows because in this case the event on the right-hand side contains the event on the left-hand side. To prove (60) for $(k_{0,i}^*, k_{1,1}^*, \ell_i^*) = (2,2,1)$, we again use the symmetry of the code construction to conclude Equalities (61) and (62). In fact, (61) holds because conditioned on $K_{0,i}^* = K_{i,i}^* = L_i^* = 1$ every set of $\lfloor 2^{nR_i'}\rfloor - 1$ codewords $\{V_i^n(k_{0,i}, k_{i,i}, \ell_i)\}$ for $(k_{0,i}, k_{i,i}, \ell_i) \neq (1,1,1)$ has the same joint distribution. Inequality (63) finally follows because the event on the right-hand side contains the event on the left-hand side. This proves (60) and concludes our proof.

APPENDIX B
CONVEXITY IN THEOREM 2

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

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

Let $Q \sim \text{Bern}(\alpha)$ be independent of the union of the two sets, and define $\bar{V}_0 \overset{\text{def}}{=} V_{0,Q}, \bar{V}_i \overset{\text{def}}{=} V_{i,Q}, \bar{X} \overset{\text{def}}{=} X_Q, \bar{Y}_i \overset{\text{def}}{=} Y_{i,Q}$, 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 $(\bar{X}, \bar{Y}_1, \bar{Y}_2)$. Therefore, and since by assumption

$$(\bar{V}_0, \bar{V}_1, \bar{V}_2) \multimap (\bar{X}, Q) \multimap (\bar{Y}_1, \bar{Y}_2),$$

we conclude that defining $\tilde{V}_0 \overset{\text{def}}{=} (Q, \bar{V}_0)$ we have the Markov chain

$$(\tilde{V}_0, \bar{V}_1, \bar{V}_2) \multimap \bar{X} \multimap (\bar{Y}_1, \bar{Y}_2). \quad (64)$$

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

$$I(\bar{X}; \bar{V}_i | \bar{V}_0, \bar{Y}_i, Q) = I(\bar{X}; \bar{V}_i | \tilde{V}_0, \bar{Y}_i) \quad (65)$$

and

$$I(\bar{X}; \bar{V}_0 | \bar{Y}_i, Q) = I(\bar{X}; \bar{V}_0, Q | \bar{Y}_i) = I(\bar{X}; \tilde{V}_0 | \bar{Y}_i), \quad (66)$$

where the first equality holds because of the independence of Q and $(\bar{X}, \bar{Y}_i)$. Moreover, by (65) and (66)

$$I(\bar{X}; \bar{V}_0, \bar{V}_i | \bar{Y}_i, Q) = I(\bar{X}; \bar{V}_i, \tilde{V}_0 | \bar{Y}_i) \quad (67)$$

Combining these inequalities with the Markov condition, we conclude that the region $\mathcal{R}_{\text{LGW}}^{\text{inner}}$ is convex.

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," *in Proc. of 21st Annual Allerton Conference on Communication, Control, and Computing,* Monticello (IL), USA, October 5–7, 1983.

[3] D. Tuninetti, On interference channels with generalized feedback, *in Proc. of ISIT 2007,* Nice, France, June 2007.

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

$$\Pr\left(\bigcup_{\substack{\ell_i\in[2^{nR'_i}]\\ \ell_i\neq L_i^*}} (V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),V_i^n(K_{0,i}^*,K_{i,i}^*,\ell_i),Y_i^n)\in T_\epsilon^n(P_{V_0V_iY_i})\right)$$
$$= \Pr\left(\bigcup_{\ell_i=2}^{\lfloor 2^{nR'_i}\rfloor} (V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),V_i^n(1,1,\ell_i),Y_i^n)\in T_\epsilon^n(P_{V_0V_iY_i})\middle| K_{0,i}^*=K_{i,i}^*=L_i^*=1\right) \tag{58}$$

$$\Pr\left(\bigcup_{\ell_i=2}^{\lfloor 2^{nR'_i}\rfloor} (V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),V_i^n(1,1,\ell_i),Y_i^n)\in T_\epsilon^n(P_{V_0V_iY_i})\middle| K_{0,i}^*=K_{i,i}^*=L_i^*=1\right)$$
$$\leq \Pr\left(\bigcup_{\ell_i=1}^{\lfloor 2^{nR'_i}\rfloor} (V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),V_i^n(1,1,\ell_i),Y_i^n)\in T_\epsilon^n(P_{V_0V_iY_i})\right) \tag{59}$$

$$\Pr\left(\bigcup_{\ell_i=2}^{\lfloor 2^{nR'_i}\rfloor} (V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),V_i^n(1,1,\ell_i),Y_i^n)\in T_\epsilon^n(P_{V_0V_iY_i})\middle| K_{0,i}^*=K_{i,i}^*=L_i^*=1\right)$$
$$\leq \Pr\left(\bigcup_{\ell_i=1}^{\lfloor 2^{nR'_i}\rfloor} (V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),V_i^n(1,1,\ell_i),Y_i^n)\in T_\epsilon^n(P_{V_0V_iY_i})\middle| K_{0,i}^*=k_{0,i}^*,K_{i,i}^*=k_{i,i}^*,L_i^*=\ell_i^*\right) \tag{60}$$

$$\Pr\left(\bigcup_{\ell_i=2}^{\lfloor 2^{nR'_i}\rfloor} V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),V_i^n(1,1,\ell_i),Y_i^n)\in T_\epsilon^n(P_{V_0V_iY_i})\middle| K_{0,i}^*=K_{i,i}^*=L_i^*=1\right)$$
$$= \Pr\left(\bigcup_{\ell_i=2}^{\lfloor 2^{nR'_i}\rfloor} (V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),V_i^n(2,2,\ell_i),Y_i^n)\in T_\epsilon^n(P_{V_0V_iY_i})\middle| K_{0,i}^*=K_{i,i}^*=L_i^*=1\right) \tag{61}$$
$$= \Pr\left(\bigcup_{\ell_i=2}^{\lfloor 2^{nR'_i}\rfloor} (V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),V_i^n(1,1,\ell_i),Y_i^n)\in T_\epsilon^n(P_{V_0V_iY_i})\middle| K_{0,i}^*=K_{i,i}^*=2,L_i^*=1\right) \tag{62}$$
$$\leq \Pr\left(\bigcup_{\ell_i=1}^{\lfloor 2^{nR'_i}\rfloor} (V_0^n(i;K_{0,0}^*,K_{i,0}^*,L_{i,0}^*),V_i^n(1,1,\ell_i),Y_i^n)\in T_\epsilon^n(P_{V_0V_iY_i})\middle| K_{0,i}^*=K_{i,i}^*=2,L_i^*=1\right) \tag{63}$$

---

[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, 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, 1977.

[8] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. on Inf. Theory,* vol. 25, pp. 60–64, no. 3, 1979.

[9] S. I. Gelfand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," *Probl. Inf. Transm.,* vol. 16, no. 1, pp. 24–34, 1980.

[10] A. Lapidoth and Y. Steinberg, "The multiple access channel with two independent states each known causally to one encoder," *in Proc. ISIT 2010*, Austin, Texas, June 13–18, 2010.

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

[12] R. Venkataramanan and S.S. Pradhan, "Achievable rates for the broadcast channel with feedback,", *in Proc. ISIT 2010*, Austin, Texas, June 13–18, 2010.

[13] R. Venkataramanan 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`.

[14] 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

[15] Y. Liang, G. Kramer, and H. V. Poor, "Equivalence of two inner bounds on the capacity region of the broadcast channel," *in Proc. 46th Annual Allerton Conference on Communication, Control, and Computing,* Monticello (IL), USA, Sept. 2008.

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

[17] 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.

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

[19] Y. Steinberg and N. Merhav, “On successive refinement for the Wyner-Ziv problem,” IT-50, no. 8, pp. 1636–1654, Aug. 2004.

[20] R. Timo 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.

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

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