Coding Schemes with Rate-Limited Feedback that Improve over the Nofeedback Capacity for a Large Class of Broadcast Channels

Youlong Wu, Michèle Wigger

arXiv:1401.6219v2 [cs.IT] 29 Jun 2014

Abstract

We propose new coding schemes for the two-receiver discrete memoryless broadcast channel (BC) with rate-limited feedback from one or both receivers. Our schemes apply a block-Markov strategy and Marton coding in each block. Also, the receivers use the feedback links to send Wyner-Ziv compression messages about their previously observed outputs to the transmitter.

We have two types of schemes. In the first type the transmitter simply relays the compression messages obtained over the feedback links by encoding them into the Marton cloud center of the next-following block. Each receiver uses these compression messages to reconstruct a quantized version of the other receiver's outputs in order to improve the decoding of its desired data message. In our scheme, each receiver can decode the Marton cloud center with the same performance as if the compression messages that it had sent itself over the feedback link in the previous block was not present. This implies in particular that for asymmetric setups where one receiver is stronger than the other (e.g., less noisy), the feedback allows to improve the stronger receiver's performance—by conveying it the information about the weaker receiver's outputs—without degrading the weaker receiver's performance.

The described coding scheme is analyzed with sliding-window decoding, backward decoding, or a mixture thereof. These schemes strictly improve over the nofeedback capacity region for the class of *strictly essentially less-noisy BCs*, which we introduce in this paper. This holds even when there is only feedback from the weaker receiver and no matter how small (but positive) the feedback rate. Examples of

This paper was in part presented at the *IEEE International Symposium on Information Theory*, in Honolulu, HI, USA, June–July 2014.

Y. Wu and M. Wigger are with the Department of Communications and Electronics, Telecom ParisTech, Paris, France. (e-mail: youlong.wu@telecom-paristech; michele.wigger@telecom-paristech.fr).

M. Wigger was partially supported by the city of Paris under program "Emergences".

strictly essentially less-noisy BCs are the binary symmetric BC (BSBC) or the binary erasure BC (BEBC) with unequal cross-over probabilities or unequal erasure probabilities to the two receivers. Our schemes also improve over the nofeedback capacity region of the binary symmetric channel/binary erasure channel BC and also for parameter ranges where the BC is not essentially less-noisy but more capable. Previous to our work, feedback was known to increase capacity only for a few very specific memoryless BCs with feedback.

In our second type of scheme, the encoder decodes all the feedback information and processes it with some local information before sending the result to the receivers. When the feedback-rates are sufficiently large, then our scheme can recover all previously known capacity and degrees of freedom results for memoryless BCs with feedback. This includes in particular the result by Wang and by Georgiadis and Tassiulas for the binary erasure BC when all erasures are known to both receivers, the results by Shayevitz and Wigger and by Chia, Kim, and El Gamal on variations of the Blackwell DMBC, and the result by Maddah-Ali and Tse on memoryless fading BCs with completely stale state information. In fact, as the feedback-rates tend to infinity our scheme improves over a special case of the Shayevitz-Wigger scheme which is known to recover the mentioned results.

With appropriate modification, our schemes can achieve the same rates also for BCs with noisy feedback assuming that the receivers can code over the feedback links.

I. Introduction

For most discrete memoryless broadcast channels (DMBC), it is not known whether feedback can increase the capacity region, even when the feedback links are noise-free and of infinite rate. There are some exceptions. For example, for all physically degraded DMBCs the capacity regions with and without feedback coincide [1]. The first simple example DMBC where (even rate-limited) feedback increases capacity was presented by Dueck [2]. His example and coding scheme were generalized by Shayevitz and Wigger [3] who proposed a general scheme and achievable region for DMBCs with generalized feedback. In the generalized feedback model, the feedback to the transmitter is modeled as an additional output of the DMBC that can depend on the input and the receivers' outputs in an arbitrary manner. It has recently been shown [4] that the Shayevitz-Wigger scheme for generalized feedback includes as special cases the two-user schemes by Wang [5], by Georgiadis and Tassiulas [6], and by Maddah-Ali and Tse [7], which achieve the capacity region and the degrees of freedom region of their respective channels. Further, the Shayevitz-Wigger scheme also includes the schemes in [8], [9], [10] and [11] for setups with imperfect (rate-limited) delayed channel state-information (CSI) at the transmitter. (A more detailed discussion of these connections is provided at the end of Section III-B.)

Other achievable regions for general DMBCs with perfect or noisy feedback have been proposed by Kramer [12] and by Venkataramanan and Pradhan [13]. Kramer's achievable region can be used to show that feedback improves capacity for some specific *binary symmetric BC* (BSBC). Comparing the general achievable regions in [3], [12], [13] to each other is hard because of their complex form which involves several auxiliary random variables.

A different line of works has concentrated on the memoryless Gaussian broadcast channel (BC) [14], [15], [16], [17], [18], [19], [20], [21]. The best achievable region when the noises at the two receivers are independent is given in [18] and is based on a MAC-BC duality approach. In [20], the asymptotic high-SNR sum-capacity for arbitrary noise correlation is derived.

In this paper, we present two types of coding schemes for DMBCs with *rate-limited feedback*. Our schemes use a block-Markov strategy where in each block they apply Marton coding [22], which to date is the best known coding scheme without feedback. The messages sent over the feedback links are simply compression information that describe the channel outputs that the receivers observed during a block.

In our first type of scheme, (Schemes 1A–1C), the encoder transmits exactly these compression informations as part of the cloud center of the Marton code employed in the next block. Thus, here, the encoder only *relays* the feedback messages from one receiver to the other. Each receiver can hence reconstruct a compressed version of the other receiver's outputs and apply a modified Marton decoding to these compressed outputs and its own observed outputs. The Marton decoding is modified to account for the fact that each receiver already knows a part of the message sent in the cloud center—namely the compression information it had generated itself after the previous block. As we will see, the decoding can be performed as well as if the part of the cloud-center message known at a receiver was not present. In this sense, in the cloud center we are sending information that is useful to one of the two receivers without disturbing the other receiver, or in other words, without occupying the other receiver's resources. For asymmetric setups where one of the two receivers is stronger than the other, e.g., less noisy, this implies that we can send the compression message, and thus the information about the other receiver's outputs, to the stronger receiver without harming the performance of the weaker receiver. This allows in particular to improve over Marton's original nofeedback scheme.

We discuss the described coding strategy when the two receivers apply backward decoding (Scheme 1A), when they apply sliding-window decoding (Scheme 1B), and when one receiver applies backward decoding and the other sliding-window decoding (Scheme 1C).

Our coding strategy is reminiscent of the compress-and-forward relay strategy [23] or the noisy network coding for general networks [24], [25] in the sense that the two receivers compress their channel outputs and send these compression indices over the feedback links. However, in our schemes, we use Marton coding since our transmitter has to send two independent private messages to the two receivers (we could treat them as a big common message, but this would perform poorly). Moreover, whereas in noisy network coding the transmitter where to generate new compression indices that describe its observed feedback outputs, in our schemes the transmitter *decodes and forwards (or relays)* the compression messages that were sent over the feedback links. Thus, in our schemes the transmitter sends compression indices that describe the outputs observed at the two receivers.

Our schemes are particularly beneficial for the class of *strictly essentially less-noisy* DMBCs, which we define in this paper and which represents a subclass of Nair's essentially less-noisy DMBCs [26]. Our class includes the BSBC and the *binary erasure BCs* (BEBC) with unequal cross-over probabilities or unequal erasure probabilities at the two receivers, and the *binary symmetric channel/binary erasure channel BC* (BSC/BEC-BC) for a large range of parameters. For strictly essentially less-noisy DMBCs Marton coding is known to achieve capacity [26]. For this class of DMBCs, our schemes improve strictly over the nofeedback capacity region no matter how small but positive the feedback rates are and even when there is feedback only from the weaker receiver. In fact, for most of these channels our scheme can improve over all boundary points $(R_1 > 0, R_2 > 0)$ of the nofeedback capacity region. The described schemes also improve over the nofeedback capacity region of the BSC/BEC-BC when the DMBC is more capable [27], unless the BSC and BEC have same capacities.

Thus, unlike for previous schemes, with our new schemes we can easily show that feedback increases the capacity region for a large set of DMBCs.

We present a fourth scheme, Scheme 2, where the encoder uses the feedback messages to *reconstruct* compressed versions of the channel outputs, and then *processes* these compressed signals together with the previously sent codewords to generate update (compression) information intended to both receivers. This update information is sent as part of the cloud center of the Marton code employed in the next-following block. This scheme is reminiscent of the Shayevitz-Wigger scheme [3] but for rate-limited feedback. Moreover, in our Scheme 2 here, the update information is sent only in the cloud center and using a joint source-channel code, whereas in the Shayevitz-Wigger scheme parts of it are also sent in the satellite codewords but using only a separate source-channel code.

Since here the update information is sent using a joint source-channel code, in the limit as the feedback rates increase, the region achieved with our Scheme 2 improves over the region achieved by the Shayevitz-Wigger scheme when this latter is restricted to send all the update information in the cloud center. Notice that this represents a prominent special case of the Shayevitz-Wigger scheme which subsumes the schemes

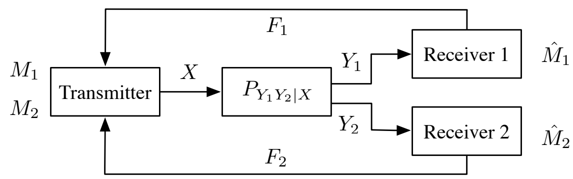

Fig. 1. Broadcast channel with rate-limited feedback

by Wang [5], by Georgiadis and Tassiulas [6], by Maddah-Ali and Tse [7], and also the schemes in [8], [9], [10] and [11] when these are specialized to memoryless BCs and to delayed state-information only.

All our results hold also with noisy feedback when the receivers can code over the feedback links.

A. Notation

Let $\mathbb{R}$ denote the set of reals and $\mathbb{Z}^+$ the set of positive integers. For a finite set $\mathcal{A}$, we denote by $|\mathcal{A}|$ its cardinality and by $\mathcal{A}^j$, for $j \in \mathbb{Z}^+$, its j-fold Cartesian product, $\mathcal{A}^j := \mathcal{A}_1 \times \ldots \times \mathcal{A}_j$.

We use capital letters to denote random variables and small letters for their realizations, e.g. X and x. For $j \in \mathbb{Z}^+$, we use the short hand notations X^j and x^j for the tuples $X^j := (X_1, \ldots, X_j)$ and $x^j := (x_1, \ldots, x_j)$. Given a set $\mathcal{S} \in \mathbb{R}^2$, we denote by $\text{bd}(\mathcal{S})$ and $\text{int}(\mathcal{S})$ the boundary and the interior of $\mathcal{S}$.

$Z \sim \text{Bern}(p)$ denotes that Z is a binary random variable taking values 0 and 1 with probabilities $1-p$ and p. Also, we use the definitions $\bar{a} := (1-a)$ and $a * b := \bar{a}b + a\bar{b}$, for $a, b \in [0,1]$. Given a positive integer n, let $\mathbf{1}_{[n]}$ denote the all-one tuple of length n, e.g., $\mathbf{1}_{[3]} = (1,1,1)$. The abbreviation i.i.d. stands for *independent and identically distributed*.

Given a distribution P_A over some alphabet $\mathcal{A}$, a positive real number $\varepsilon > 0$, and a positive integer n, let $\mathcal{T}_\varepsilon^n(P_A)$ denote the typical set in [28]. $H_b(\cdot)$ denotes the binary entropy function. Finally, given an event ϵ, we denote its complement by ϵ^c.

II. Channel model

Communication takes place over a DMBC with rate-limited feedback, see Figure 1. The setup is characterized by the finite input alphabet $\mathcal{X}$, the finite output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, the channel law $P_{Y_1Y_2|X}$, and nonnegative feedback rates $R_{\text{Fb},1}$ and $R_{\text{Fb},2}$. If at discrete-time t the transmitter sends the channel input $x_t \in \mathcal{X}$, then Receiver $i \in \{1,2\}$ observes the output $Y_{i,t} \in \mathcal{Y}_i$, where the pair $(Y_{1,t}, Y_{2,t}) \sim P_{Y_1Y_2|X}(\cdot,\cdot|x_t)$. Also, after observing $Y_{i,t}$, Receiver i can send a feedback signal $F_{i,t} \in \mathcal{F}_{i,t}$ to the transmitter, where $\mathcal{F}_{i,t}$ denotes the finite alphabet of $F_{i,t}$ and is a design parameter of a scheme.

The feedback link between the transmitter and Receiver i is assumed to be instantaneous and noiseless but *rate-limited* to $R_{\text{Fb},i}$ bits on average. Thus, if the transmission takes place over a total blocklength N, then

$$|\mathcal{F}_{i,1}| \times \cdots \times |\mathcal{F}_{i,N}| \leq 2^{NR_{\text{Fb},i}}, \quad i \in \{1,2\}. \tag{1a}$$

The goal of the communication is that the transmitter conveys two independent private messages $M_1 \in \{1,\ldots,\lfloor 2^{NR_1}\rfloor\}$ and $M_2 \in \{1,\ldots,\lfloor 2^{NR_2}\rfloor\}$, to Receiver 1 and 2, respectively. Each M_i, $i \in \{1,2\}$, is uniformly distributed over the set $\mathcal{M}_i := \{1,\ldots,\lfloor 2^{NR_i}\rfloor\}$, where R_i denotes the private rate of transmission of Receiver i.

The transmitter is comprised of a sequence of encoding functions $\{f_t^{(N)}\}_{t=1}^N$ of the form $f_t^{(N)}$: $\mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{F}_{1,1} \times \cdots \times \mathcal{F}_{1,t-1} \times \mathcal{F}_{2,1} \times \cdots \times \mathcal{F}_{2,t-1} \to \mathcal{X}$ that is used to produce the channel inputs as

$$X_t = f_t^{(N)}\big(M_1, M_2, F_1^{t-1}, F_2^{t-1}\big), \qquad t \in \{1,\ldots,N\}. \tag{2}$$

Receiver $i \in \{1,2\}$ is comprised of a sequence of feedback-encoding functions $\{\psi_{i,t}^{(N)}\}_{t=1}^N$ of the form $\psi_{i,t}^{(N)} : \mathcal{Y}_i^t \to \mathcal{F}_{i,t}$ that is used to produce the symbols

$$F_{i,t} = \psi_{i,t}^{(N)}(Y_{i,1},\ldots,Y_{i,t}), \qquad t \in \{1,\ldots,N\}, \tag{3}$$

sent over the feedback link, and of a decoding function $\Phi_i^{(N)} : \mathcal{Y}_i^N \to \mathcal{M}_i$ used to produce a guess of Message M_i:

$$\hat{M}_i = \Phi_i^{(N)}(Y_i^N). \tag{4}$$

A rate region (R_1, R_2) with averaged feedback rates $R_{\text{Fb},1}$, $R_{\text{Fb},2}$ is called achievable if for every blocklength N, there exists a set encoding functions $\{f_t^{(N)}\}_{t=1}^N$ and for $i = \{1,2\}$ there exists a set of decoding functions $\Phi_i^{(N)}$, feedback alphabets $\{\mathcal{F}_{i,t}\}_{t=1}^N$ satisfying (1), and feedback-encoding functions $\{\psi_{i,t}^{(N)}\}_{t=1}^N$ such that the error probability

$$P_e^{(N)} := \Pr(M_1 \neq \hat{M}_1 \text{ or } M_2 \neq \hat{M}_2) \tag{5}$$

tends to zero as the blocklength N tends to infinity. The closure of the set of achievable rate pairs (R_1, R_2) is called the *feedback capacity region* and is denoted by $\mathcal{C}_{\text{Fb}}$.

In the special case $R_{\text{Fb},1} = R_{\text{Fb},2} = 0$ the feedback signals are constant and the setup is equivalent to a setup without feedback. We denote the capacity region for this setup $\mathcal{C}_{\text{NoFB}}$.

III. Preliminaries

We recall some previous results on the capacity region of DMBCs without and with feedback.

A. Previous Results Without Feedback

The capacity region of DMBCs without feedback is in general unknown. The best known inner bound without feedback is Marton's region [22], $\mathcal{R}_{\text{Marton}}$, which is defined as the set of all nonnegative rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U_0, U_1; Y_1) \tag{6a}$$

$$R_2 \leq I(U_0, U_2; Y_2) \tag{6b}$$

$$R_1+R_2 \leq I(U_0, U_1;Y_1)+I(U_2;Y_2|U_0)-I(U_1;U_2|U_0) \tag{6c}$$

$$R_1+R_2 \leq I(U_0, U_2;Y_2)+I(U_1;Y_1|U_0)-I(U_1;U_2|U_0) \tag{6d}$$

for some probability mass function (pmf) $P_{U_0U_1U_2}$ and a function $f\colon \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$ such that $X = f(U_0, U_1, U_2)$.

An important subset of Marton's region is the *superposition coding region*, $\mathcal{R}^{(1)}_{\text{SuperPos}}$, which results when Marton's constraints (6) are specialized to U_1 =const. and $X = U_2$. That means, $\mathcal{R}^{(1)}_{\text{SuperPos}}$ is defined as the set of all nonnegative rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U; Y_1) \tag{7a}$$

$$R_2 \leq I(X; Y_2|U) \tag{7b}$$

$$R_1 + R_2 \leq I(X; Y_2) \tag{7c}$$

for some pmf P_{UX}. The superposition coding region $\mathcal{R}^{(2)}_{\text{SuperPos}}$ is defined similarly to $\mathcal{R}^{(1)}_{\text{SuperPos}}$ but where indices 1 and 2 need to be exchanged.

Remark 1. *To evaluate Marton's region, it suffices to consider distributions $P_{U_0U_1U_2X}$ for which one of the following conditions holds:*

- $I(U_0; Y_1) = I(U_0; Y_2)$*;*
- $I(U_0; Y_1) < I(U_0; Y_2)$ *and* $U_1 =$ const.*;*
- $I(U_0; Y_1) > I(U_0; Y_2)$ *and* $U_2 =$ const..

The proof is given in [29], [30]. For convenience to the reader, we reprove the statement in Appendix E.

The best known outer bound without feedback is the Nair-El Gamal outer bound [31] which is defined as the set of all nonnegative rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U; Y_1) \tag{8a}$$

$$R_2 \leq I(V;Y_2) \tag{8b}$$

$$R_1 + R_2 \leq I(U;Y_1) + I(X;Y_2|U) \tag{8c}$$

$$R_1 + R_2 \leq I(V;Y_2) + I(X;Y_1|V) \tag{8d}$$

for some pmf P_{UVX}.

The Nair-El Gamal outer bound is known to coincide with Marton's region for the following classes of DMBCs, which also play a role in the present paper:

- *stochastically or physically degraded DMBCs [32]*
- *less noisy DMBCs [33]*
- *essentially less noisy DMBCs [26]*
- *more capable DMBCs [33].*

In all these classes of DMBCs one of the two receivers is stronger than the other receiver in some sense. This makes that superposition coding is as good as the more general Marton coding and achieves capacity.

Remark 2. *The relationship among these various classes of BCs is established in [26], [34]:*

- *degraded $\subsetneq$ less-noisy $\subsetneq$ more capable,*
- *less noisy $\subsetneq$ essentially less noisy,*
- *essentially less-noisy $\nsubseteq$ more capable,*
- *more capable $\nsubseteq$ essentially less-noisy.*

We recall the definition of essentially less-noisy DMBCs as they are most important for this paper.

Definition 1 (From [26])**.** *A subset $\mathcal{P}_{\mathcal{X}}$ of all pmfs on the input alphabet $\mathcal{X}$ is said to be a* sufficient class *of pmfs for a DMBC if the following holds: Given any joint pmf P_{UVX} there exists a joint pmf P'_{UVX} that satisfies*

$$\begin{aligned} P'_X(x) &\in \mathcal{P}_{\mathcal{X}} \\ I_P(U;Y_1) &\leq I_{P'}(U;Y_1) \\ I_P(V;Y_2) &\leq I_{P'}(V;Y_2) \\ I_P(U;Y_1) + I_P(X;Y_2|U) &\leq I_{P'}(U;Y_1) + I_{P'}(X;Y_2|U) \\ I_P(V;Y_2) + I_P(X;Y_1|V) &\leq I_{P'}(V;Y_2) + I_{P'}(X;Y_1|V) \end{aligned} \tag{9}$$

where the notations I_P and $I_{P'}$ indicate that the mutual informations are computed assuming that $(U,V,X) \sim P_{UVX}$ and $(U,V,X) \sim P'_{UVX}$ and $P'_X(x)$ is the marginal obtained from P'_{UVX}.

Definition 2 (From [26])**.** *A DMBC is called* essentially less-noisy *if there exists a* sufficient class *of pmfs $\mathcal{P}_{\mathcal{X}}$ such that whenever $P_X \in \mathcal{P}_{\mathcal{X}}$, then for all conditional pmfs $P_{U|X}$,*

$$I(U;Y_1) \leq I(U;Y_2). \tag{10}$$

The class of essentially less-noisy DMBCs contains as special cases the BSBC and the BEBC. Also the memoryless Gaussian BC is essentially less noisy.

To evaluate the superposition coding region $\mathcal{R}^{(1)}_{\text{SuperPos}}$ of an essentially less-noisy DMBC, it suffices to evaluate the region given by constraints (7) for pmfs P_{UX} that satisfy $I(U;Y_1) \leq I(U;Y_2)$.

B. Previous Results with Feedback

Previous results on the DMBC with feedback mostly focus on perfect output feedback, which in our setup corresponds to infinite feedback rates $R_{\text{Fb},1}, R_{\text{Fb},2} \to \infty$.

A simple outer bound on the capacity with output feedback is given in [35]. It equals the capacity region $\mathcal{C}^{(1)}_{\text{Enh}}$ of an enhanced DMBC where the outputs Y_1^n are also revealed to Receiver 2. Notice that this enhanced DMBC is physically degraded and thus, with and without feedback, its capacity region is given by the set of all nonnegative rate pairs (R_1, R_2) that satisfy

$$R_1 \leq I(U;Y_1) \tag{11a}$$

$$R_2 \leq I(X;Y_1,Y_2|U) \tag{11b}$$

for some pmf P_{UX}.

Exchanging everywhere in the previous paragraph indices 1 and 2, we can define a similar enhanced capacity region $\mathcal{C}^{(2)}_{\text{Enh}}$, which is also an outer bound to $\mathcal{C}_{\text{Fb}}$. The intersection $\mathcal{C}^{(1)}_{\text{Enh}} \cap \mathcal{C}^{(2)}_{\text{Enh}}$ yields an even tighter outerbound [5], [6].

The achievable region with feedback that is most closely related to our paper is the Shayevitz-Wigger region [3], which is defined as the set of all nonnegative rate pairs (R_1, R_2) satisfying (12) for some pmfs $P_Q P_{U_0U_1U_2|Q}$, $P_{V_0V_1V_2|U_0U_1U_2Y_1Y_2Q}$ and some function $f\colon \mathcal{Q} \times \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$, where $X = f(U_0, U_1, U_2, Q)$.

This region is generally difficult to evaluate due to the presence of six auxiliary random variables in the rate constraints. Recently, Kim, Chia, and El Gamal [4] studied the more general Shayevitz-Wigger region for generalized feedback [3], which differs from the above region only in that in some places the outputs Y_1 or Y_2 have to be replaced by the generalized feedback output $\tilde{Y}$. In particular, they proposed two choices for the auxiliary random variables for the Shayevitz-Wigger region with generalized feedback

$$R_1 \leq I(U_0, U_1; Y_1, V_1|Q) - I(U_0, U_1, U_2, Y_2; V_0, V_1|Q, Y_1) \tag{12a}$$

$$R_2 \leq I(U_0, U_2; Y_2, V_2|Q) - I(U_0, U_1, U_2, Y_1; V_0, V_2|Q, Y_2) \tag{12b}$$

$$\begin{aligned} R_1 + R_2 \leq\ & I(U_1; Y_1, V_1|Q, U_0) + I(U_2; Y_2, V_2|Q, U_0) + \min_{i \in \{1,2\}} I(U_0; Y_i, V_i Q) - \max_{i \in \{1,2\}} I(U_0, U_1, U_2, Y_1, Y_2; V_0|Q, Y_i) \\ & - I(U_0, U_1, U_2, Y_2; V_1|Q, V_0, Y_1) - I(U_0, U_1, U_2, Y_1, Y_2; V_2|Q, V_0, Y_2) - I(U_1; U_2|Q, U_0) \end{aligned} \tag{12c}$$

$$\begin{aligned} R_1 + R_2 \leq\ & I(U_0, U_1; Y_1, V_1|Q) + I(U_0, U_2; Y_2, V_2|Q) - I(U_1; U_2|Q, U_0) \\ & - I(U_0, U_1, U_2, Y_2; V_0, V_1|Q, Y_1) - I(U_0, U_1, U_2, Y_1; V_0, V_2|Q, Y_2) \end{aligned} \tag{12d}$$

---

and presented simplified expressions for the maximum sum-rates that these choices achieve for symmetric state-dependent DMBCs with state known at both receivers and where the generalized feedback equals the delayed state sequence. Their first choice is given by

$$Q = \begin{cases} 0 & \text{w. p. } 1 - 2p \\ 1 & \text{w. p. } p \\ 2 & \text{w. p. } p \end{cases}, \tag{13a}$$

$$V_0 = V_1 = V_2 = \begin{cases} \emptyset & \text{if } Q = 0 \\ Y_1 & \text{if } Q = 1 \\ Y_2 & \text{if } Q = 2 \end{cases}, \tag{13b}$$

and

$$X = \begin{cases} U_0 & \text{if } Q = 0 \\ U_1 & \text{if } Q = 1 \\ U_2 & \text{if } Q = 2 \end{cases} \tag{13c}$$

for joint pmf $P_{U_0 U_1 U_2} = P_{U_0} P_{U_1} P_{U_2}$. This choice essentially results in a coded time-sharing scheme. Their second choice is

$$Q = \begin{cases} 1 & \text{w. p. } 1/2 \\ 2 & \text{w. p. } 1/2 \end{cases}, \tag{14a}$$

$$V_0 = V_1 = V_2 = \begin{cases} Y_1 & \text{if } Q = 1 \\ Y_2 & \text{if } Q = 2 \end{cases}, \tag{14b}$$

and

$$X = \begin{cases} U_1 & \text{if } Q = 1 \\ U_2 & \text{if } Q = 2 \end{cases} \tag{14c}$$

for some joint pmf $P_{U_0U_1U_2} = P_{U_0}P_{U_1|U_0}P_{U_2|U_0}$. This choice results in a randomized superposition coding scheme.

Notice that the two-user i.i.d. fading BC with delayed state information at the transmitter considered by Maddah-Ali and Tse [7], and the BEBCs where *all* erasure events are known at all receivers considered by Wang [5] and by Georgiadis and Tassiulas [6], belong to this class of DMBCs. In fact, it is shown in [4] that the degrees-of-freedom optimal scheme in [7] and the capacity-achieving schemes in [5], [6] are special cases of the general Shayevitz-Wigger scheme specialized to the choice (13). The Maddah-Ali&Tse scheme [7] has been modified and extended to setups with imperfect (rate-limited) current and delayed channel state-information (CSI) at the transmitter, e.g., in [8], [9], [10] and [11]. In the special case of only delayed CSI, these schemes are also special cases[1] of the Shayevitz-Wigger scheme for generalized feedback, more precisely for feedback signals that are noisy versions of the channel state. For example, the scheme in [9] corresponds to the Shayevitz-Wigger scheme with the following auxiliaries: $Q = \{1, 2\}$; $X = U_0 + U_1 + U_2$; $U_0 - U_1 - U_2$ form a Markov chain; $U_0 =$ const. when $Q = 1$ and U_0 arbitrary when $Q = 2$; $V_0 = V_1 = V_2 = (\hat{\eta}_1, \hat{\eta}_2)$ where $(\hat{\eta}_1, \hat{\eta}_2)$ are defined in [9]. The scheme in [8] corresponds to the Shayevitz-Wigger scheme with the following auxiliaries: $X = U_0 + U_1 + U_2$; $U_0 - U_1 - U_2$ form a Markov chain; and $V_0 = V_1 = V_2 = (\bar{\bar{l}}^{(1)}, \bar{\bar{l}}^{(2)})$ where $(\bar{\bar{l}}^{(1)}, \bar{\bar{l}}^{(2)})$ are defined in (54) in [8].

IV. Motivation: A Simple Scheme

We sketch a simple scheme that motivates our schemes in Section VII. We assume there is only feedback from Receiver 1, i.e., $R_{\text{Fb},1} > 0$ and $R_{\text{Fb},2} = 0$.

We apply block-Markov coding with $B+1$ blocks of length n, where in each block we use superposition coding (without feedback) to send fresh messages $M_{1,b}$ and $M_{2,b}$. Message $M_{1,b}$ is sent in the cloud center U_b^n and $M_{2,b}$ in the satellite codeword X_b^n. Thus, the scheme is expected to perform well when

[1]The schemes in [9] and [8], however, use successive decoding instead of the more performant joint decoding applied in [3].

the following gap is nonnegative:

$$\Gamma := I(U;Y_2) - I(U;Y_1) \geq 0. \tag{15}$$

(This is for example the case in a BSBC when the cross-over probability to Receiver 2 is no larger than the cross-over probability to Receiver 1.)

After each block, both Receivers 1 and 2 decode the cloud center codeword U_b^n by means of joint typicality decoding. By the Packing Lemma, this is possible whenever

$$R_1 \leq I(U;Y_1) \tag{16}$$

$$R_1 \leq I(U;Y_2), \tag{17}$$

where here, by (15), the second constraint is not active. We notice that when

$$\Gamma > 0 \tag{18}$$

Receiver 2 would be able to decode the cloud center even if—besides $M_{1,b}$—it also encoded an extra message of rate not exceeding Γ. Of course, we cannot just add an arbitrary rate-Γ message to the cloud center, because this would make it impossible for Receiver 1 to decode this larger cloud center. Instead, we shall add a rate-Γ message that is known to Receiver 1. If in the typicality check Receiver 1 only considers the candidate codewords for the cloud center that correspond to the correct value of this extra message, then the decoding at Receiver 1 performs as well as if the additional message was not present. Thus, if the additional message is known at Receiver 1, it does not disturb its decoding.

With rate-limited feedback, we can identify a suitable additional message to send in the cloud center of block b: the feedback message $M_{\text{Fb},1,b-1}$ that Receiver 1 had fed back after the previous block $b-1$. In fact, as we describe shortly, in our scheme Receiver 1 only feeds back a message at the end of each block.

The transmitter thus simply relays the information it received over the feedback link to the other receiver. In this sense, the feedback link and part of the cloud center can be seen as an independent communication pipe from Receiver 1 to Receiver 2, where the pipe is rate-limited to

$$\min\{\Gamma, R_{\text{Fb},1}\}. \tag{19}$$

In our scheme, we use this pipe to send a compressed version of the channel outputs observed at Receiver 1 to Receiver 2. Specifically, the feedback message $M_{\text{Fb},1,b-1}$ sent after block $b-1$ is a Wyner-Ziv message that compresses outputs $Y_{1,b-1}^n$ while taking into account that the reconstructor has side-information

$Y_{2,b-1}^n, U_{b-1}^n$. The rate-required for this Wyner-Ziv message is

$$\tilde{R}_1 > I(\tilde{Y}_1; Y_1 | Y_2, U). \tag{20}$$

and, in order to satisfy the feedback-rate constraint, it also has to satisfy

$$\tilde{R}_1 < R_{\text{Fb},1}. \tag{21}$$

After decoding the additional message $M_{\text{Fb},1,b-1}$, which is transmitted in the cloud center of block b, Receiver 2 first reconstructs a compressed version of Receiver 1's outputs $\tilde{Y}_{1,b-1}^n$. It then uses this reconstruction to decode its intended Message $M_{2,b-1}$ based on the tuple $(\tilde{Y}_{1,b-1}^n, Y_{2,b-1}^n, U_{b-1}^n)$. This is possible, with arbitrary small probability of error, if

$$R_2 \leq I(X; \tilde{Y}_1, Y_2 | U). \tag{22}$$

Combining now constraints (16), (20), (21), and (22), we conclude that our scheme achieves all rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U; Y_1) \tag{23a}$$

$$R_2 \leq I(X; \tilde{Y}_1, Y_2 | U) = I(X; Y_2 | U) + I(X; \tilde{Y}_1 | U, Y_2) \tag{23b}$$

for some pmfs P_{UX} and $P_{\tilde{Y}_1|U,Y_1}$ that satisfy

$$I(\tilde{Y}_1; Y_1 | Y_2, U) \leq \min\{\Gamma, R_{\text{Fb},1}\}. \tag{24}$$

The left-hand side of (24) gives the minimum rate required for a Wyner-Ziv code that compresses $Y_{1,b-1}^n$ given that the reconstructor has side-information $Y_{2,b-1}^n$ and U_{b-1}^n.

Comparing constraints (23) to the superposition coding constraints in (7), we see that the constraints here are strictly looser whenever $I(X; \tilde{Y}_1 | U, Y_2) > 0$. Or in other words, whenever observing a compressed version of Receiver 1's outputs improves the decoding at Receiver 2.

What is remarkable about this scheme is that when $\Gamma > 0$, there is no cost in conveying the compressed version of Receiver 1's outputs to Receiver 2. It is as if there were free resources in the communication from the transmitter to Receiver 2, which the feedback allows to exploit. Without feedback, the resources cannot be exploited because the transmitter cannot identify a messages that is known at Receiver 1 and useful at Receiver 2.

V. New Achievable regions and Usefulness of Feedback

A. Achievable Regions

The following achievable regions are based on the coding schemes in Section VII. These coding schemes are motivated by the scheme sketched in the previous section, but use the more general Marton coding instead of superposition coding and exploit the feedback from both receivers.

In our first scheme 1A (Section VII-A), the receivers apply sliding-window decoding. The scheme achieves the region in the following Theorem 1.

Theorem 1 (Sliding-Window Decoding)**.** *The capacity region $\mathcal{C}_{\text{Fb}}$ includes the set $\mathcal{R}_{\text{relay,sw}}$[2] of all nonnegative rate pairs (R_1, R_2) that satisfy*

$$R_1 \leq I(U_0, U_1; Y_1, \tilde{Y}_2|Q) - I(\tilde{Y}_2; U_0, Y_2|Y_1, Q) \tag{25a}$$

$$R_1 \leq I(U_0; Y_2|Q) + I(U_1; Y_1, \tilde{Y}_2|U_0, Q) - \Delta_2 - I(\tilde{Y}_1; Y_1|U_0, U_2, Y_2, Q) \tag{25b}$$

$$R_2 \leq I(U_0, U_2; Y_2, \tilde{Y}_1|Q) - I(\tilde{Y}_1; U_0, Y_1|Y_2, Q) \tag{25c}$$

$$R_2 \leq I(U_0; Y_1|Q) + I(U_2; Y_2, \tilde{Y}_1|U_0, Q) - \Delta_1 - I(\tilde{Y}_2; Y_2|U_0, U_1, Y_1, Q) \tag{25d}$$

$$R_1+R_2 \leq I(U_0, U_1; Y_1, \tilde{Y}_2|Q) - I(\tilde{Y}_2; U_0, Y_2|Y_1, Q) + I(U_2; Y_2,\tilde{Y}_1|U_0,Q) - \Delta_1 - I(U_1; U_2|U_0,Q) \tag{25e}$$

$$R_1+R_2 \leq I(U_0, U_2; Y_2, \tilde{Y}_1|Q) - I(\tilde{Y}_1; U_0, Y_1|Y_2, Q) + I(U_1; Y_1,\tilde{Y}_2|U_0,Q) - \Delta_2 - I(U_1; U_2|U_0,Q) \tag{25f}$$

$$R_1+R_2 \leq I(U_0, U_1; Y_1, \tilde{Y}_2|Q) + I(U_0, U_2; Y_2, \tilde{Y}_1|Q) - I(\tilde{Y}_2; U_0, Y_2|Y_1, Q) - I(\tilde{Y}_1; U_0, Y_1|Y_2, Q) - I(U_1; U_2|U_0, Q) \tag{25g}$$

where

$$\Delta_1 := \max\{0, I(\tilde{Y}_1; Y_1|U_0, Y_2, Q) - R_{\text{Fb},1}\}$$

[2]The subscript "relay" indicates that the transmitter simply relays the feedback information and the subscript "sw" indicates that sliding-window decoding is applied.

$$\Delta_2 := \max\{0, I(\tilde{Y}_2; Y_2|U_0, Y_1, Q) - R_{\text{Fb},2}\}$$

for some pmfs $P_Q, P_{U_0U_1U_2|Q}, P_{\tilde{Y}_1|Y_1U_0Q}, P_{\tilde{Y}_2|Y_2U_0Q}$ *and some function* $f\colon \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Q} \to \mathcal{X}$ *such that*

$$I(U_1; Y_1, \tilde{Y}_2|U_0, Q) - \Delta_2 \geq 0 \tag{27a}$$

$$I(U_2; Y_2, \tilde{Y}_1|U_0, Q) - \Delta_1 \geq 0. \tag{27b}$$

$$I(\tilde{Y}_1; Y_1|U_0, U_2, Y_2, Q) \leq \min\{I(U_0; Y_2|Q), R_{\text{Fb},1}\} \tag{27c}$$

$$I(\tilde{Y}_2; Y_2|U_0, U_1, Y_1, Q) \leq \min\{I(U_0; Y_1|Q), R_{\text{Fb},2}\} \tag{27d}$$

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

Proof: See Section VII-A. ■

For sufficiently large feedback rates $R_{\text{Fb},1}$ and $R_{\text{Fb},2}$ (in particular for $R_{\text{Fb},1} \geq |\mathcal{Y}_1|$ and $R_{\text{Fb},2} \geq |\mathcal{Y}_2|$), the terms Δ_1 and Δ_2 as defined in (26) are 0.

In our second scheme 1B (Section VII-B), the receivers apply backward decoding. This way, for each block, they can jointly decode the cloud center and their intended satellite codewords. In this scheme, the Wyner-Ziv compression cannot be superpositioned on the cloud center because the receivers have not yet decoded this latter when compressing their channel outputs at the end of each block. The following Theorem 2 presents the achievable region for this second scheme.

Theorem 2 (Backward Decoding)**.** *The capacity region* $\mathcal{C}_{\text{Fb}}$ *includes the set* $\mathcal{R}_{\text{relay,bw}}$[3] *of all nonnegative rate pairs* (R_1, R_2) *that satisfy*

$$R_1 \leq I(U_0, U_1; Y_1, \tilde{Y}_2|Q) - I(\tilde{Y}_2; Y_2|Y_1, Q) \tag{28a}$$

$$R_2 \leq I(U_0, U_2; Y_2, \tilde{Y}_1|Q) - I(\tilde{Y}_1; Y_1|Y_2, Q) \tag{28b}$$

$$\begin{aligned} R_1+R_2 \leq{}& I(U_0, U_1; Y_1, \tilde{Y}_2|Q) - I(\tilde{Y}_2; Y_2|Y_1, Q) \\ &+I(U_2; Y_2, \tilde{Y}_1|U_0,Q)-\Delta_1-I(U_1; U_2|U_0,Q) \end{aligned} \tag{28c}$$

$$\begin{aligned} R_1+R_2 \leq{}& I(U_0, U_2; Y_2, \tilde{Y}_1|Q) - I(\tilde{Y}_1; Y_1|Y_2, Q) \\ &+I(U_1; Y_1, \tilde{Y}_2|U_0,Q)-\Delta_2-I(U_1; U_2|U_0,Q) \end{aligned} \tag{28d}$$

$$R_1+R_2 \leq I(U_0, U_1; Y_1, \tilde{Y}_2|Q) - I(\tilde{Y}_2; Y_2|Y_1, Q)$$

[3]The subscript “bw” stands for backward decoding.

$$+I(U_0,U_2;Y_2,\tilde{Y}_1|Q)-I(\tilde{Y}_1;Y_1|Y_2,Q)$$
$$-I(U_1;U_2|U_0,Q) \tag{28e}$$

for some pmfs $P_Q, P_{U_0U_1U_2|Q}, P_{\tilde{Y}_1|Y_1Q}, P_{\tilde{Y}_2|Y_2Q}$ *and some function* $f\colon \mathcal{U}_0\times\mathcal{U}_1\times\mathcal{U}_2\times\mathcal{Q}\to\mathcal{X}$ *such that*

$$I(\tilde{Y}_1;Y_1|U_0,U_2,Y_2,Q)\leq R_{\text{Fb},1} \tag{29a}$$

$$I(\tilde{Y}_2;Y_2|U_0,U_1,Y_1,Q)\leq R_{\text{Fb},2} \tag{29b}$$

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

Proof: See Section VII-B. ■

Setting $\tilde{Y}_1=\tilde{Y}_2=\text{const.}$, i.e., both receivers do not send any feedback, the region $\mathcal{R}_{\text{relay,bw}}$ specializes to $\mathcal{R}_{\text{Marton}}$.

Remark 3. *Constraints* (28) *and* (29) *are looser than Constraints* (25) *and* (27)*, respectively. But in Theorem 2 we have the conditional pmfs* $P_{\tilde{Y}_1|Y_1}$ *and* $P_{\tilde{Y}_2|Y_2}$ *whereas in Theorem 1 we allow for more general pmfs* $P_{\tilde{Y}_1|Y_1,U_0}$ *and* $P_{\tilde{Y}_2|Y_2,U_0}$*. It is thus not clear in general which of the achievable regions in Theorems 1 or 2 is larger.*

Remark 4. *Consider the Shayevitz-Wigger region* (12) *restricted to the choice of auxiliaries*

$$V_1=V_2=V_0=(f_1(Y_1,Q),f_2(Y_2,Q)) \tag{30}$$

for two deterministic functions f_1 *and* f_2*. (Notice that Kim, Chia, and El Gamal's choice of auxiliaries* (13) *or* (14) *is of this form.) Our new achievable region* $\mathcal{R}_{\text{relay,bw}}$ *improves over this restricted Shayevitz-Wigger region whenever the feedback rates* $R_{\text{Fb},1}$*,* $R_{\text{Fb},2}$ *are sufficiently large so that in our new region we can choose*

$$\tilde{Y}_1=f_1(Y_1,Q)\quad\text{and}\quad\tilde{Y}_2=f_2(Y_2,Q) \tag{31}$$

and so that $\Delta_1=\Delta_2=0$.

In fact, for the choices (30) *and* (31) *the rate constraints in* (28a), (28b), *and* (28e) *characterizing our new region coincide with the rate constraints* (12a)–(12b) *which characterize the Shayevitz-Wigger region. Moreover, the combination of the two sum-rate constraints* (28c) *and* (28d) *is looser than the sum-rate constraint* (12c)*, because the former involves a* "$\min_{i=\{1,2\}}\{a_i-b_i\}$*-term" whereas the latter involves the smaller* "$\min_{i\in\{1,2\}}a_i-\max_{i\in\{1,2\}}b_i$*-term", for* $a_i,b_i\geq 0$.

Our third scheme 1C (Section VII-C) is a mixture of the first two: Receiver 1 behaves as in the first scheme and Receiver 2 as in the second scheme. This is particularly interesting when there is no feedback from Receiver 2, $R_{\text{Fb},2} = 0$, and when Marton's scheme specializes to superposition coding with no satellite codeword for Receiver 1. Theorem 3 presents the region achieved by this third scheme with Marton coding and Corollary 1 with superposition coding.

Theorem 3 (Hybrid Sliding-Window Decoding and Backward Decoding)**.** *Even for* $R_{\text{Fb},2} = 0$*, the capacity region* $\mathcal{C}_{\text{Fb}}$ *includes the set* $\mathcal{R}^{(1)}_{\text{relay,hb}}$[4] *of all nonnegative rate pairs* (R_1, R_2) *that satisfy*

$$R_1 \leq I(U_0, U_1; Y_1|Q) \tag{32a}$$

$$\begin{aligned} R_2 \leq\ & I(U_0, U_2; \tilde{Y}_1, Y_2|Q) \\ & - I(\tilde{Y}_1; U_0, U_1, U_2, Y_1|Y_2, Q) \end{aligned} \tag{32b}$$

$$\begin{aligned} R_1 + R_2 \leq\ & I(U_0, U_1; Y_1|Q) + I(U_2; Y_2, \tilde{Y}_1|U_0, Q) \\ & - \Delta_1 - I(U_1; U_2|U_0, Q) \end{aligned} \tag{32c}$$

$$\begin{aligned} R_1 + R_2 \leq\ & I(U_1; Y_1|U_0, Q) + I(U_0, U_2; \tilde{Y}_1, Y_2|Q) \\ & - I(\tilde{Y}_1; U_0, U_1, U_2, Y_1|Y_2, Q) - I(U_1; U_2|U_0, Q) \end{aligned} \tag{32d}$$

for some pmfs $P_Q, P_{U_0U_1U_2|Q}, P_{\tilde{Y}_1|Y_1U_0Q}$ *and some function* $f: \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Q} \to \mathcal{X}$ *such that*

$$I(\tilde{Y}_1; U_1, Y_1|U_0, U_2, Y_2, Q) \leq R_{\text{Fb},1}. \tag{33}$$

The capacity region $\mathcal{C}_{\text{Fb}}$ *also includes the region* $\mathcal{R}^{(2)}_{\text{relay,hb}}$ *which is obtained by exchanging indices* 1 *and* 2 *in the above definition of* $\mathcal{R}^{(1)}_{\text{relay,hb}}$.

Proof: See Section VII-C. ■

If superposition coding is used instead of Marton coding and only one of the two receivers sends feedback, Theorem 3 reduces to the following corollary.

Corollary 1. *The capacity region* $\mathcal{C}_{\text{Fb}}$ *includes the set* $\mathcal{R}^{(1)}_{\text{relay,sp}}$[5] *of all nonnegative rate pairs* (R_1, R_2) *that satisfy*

$$R_1 \leq I(U; Y_1|Q) \tag{34a}$$

$$R_1 + R_2 \leq I(U; Y_1|Q) + I(X; Y_2, \tilde{Y}_1|U, Q) \tag{34b}$$

$$R_1 + R_2 \leq I(X; Y_2|Q) - I(\tilde{Y}_1; Y_1|U, Y_2, Q) \tag{34c}$$

[4]The subscript "hb" stands for hybrid decoding.

[5]The subscript "sp" stands for superposition coding.

for some pmfs $P_Q, P_{UX|Q}, P_{\tilde{Y}_1|Y_1UQ}$ *such that*

$$I(\tilde{Y}_1; Y_1|U, Y_2, Q) \leq R_{\text{Fb},1}. \tag{35}$$

The capacity region $\mathcal{C}_{\text{Fb}}$ *also includes the region* $\mathcal{R}^{(2)}_{\text{relay,sp}}$ *which is obtained by exchanging indices* 1 *and* 2 *in the above definition of* $\mathcal{R}^{(1)}_{\text{relay,sp}}$.

Proof: Let $\tilde{Y}_2 = U_1 = \text{const.}$, $U = U_0$ and $X = U_2$. Constraint (32a) then specializes to (34a) and Constraint (32b) is redundant compared to Constraint (32d). Observe that Constraints (32d) and (33) are looser than Constraints (34c) and (35), respectively. Also, by (35), Constraint (32c) reduces to (34b). Thus the capacity region $\mathcal{C}_{\text{Fb}}$ includes the region $\mathcal{R}^{(1)}_{\text{relay,sp}}$. Similar arguments hold for $\mathcal{R}^{(2)}_{\text{relay,sp}}$. ■

Remark 5. *The region* $\mathcal{R}^{(1)}_{\text{relay,hb}}$ *contains the regions in Theorems 1 and 2 when these latter are specialized to* U_1 =*const.,* $U_2 = X$*, and* $R_{\text{Fb},2} = 0$.

In our first three schemes 1A–1C the transmitter simply relays the compression information it received over each of the feedback links to the other receiver, as is the case also for our motivating scheme in the previous section IV. Alternatively, the transmitter can also use this feedback information to first reconstruct the compressed versions of the channel outputs and then compress them jointly with the Marton codewords. The indices resulting from this latter compression are then sent to the two receivers. The following Theorem 4 presents the region achieved by this fourth scheme 2.

Theorem 4. *The capacity region* $\mathcal{C}_{\text{Fb}}$ *includes the set* $\mathcal{R}_{\text{proc.}}$[6] *of all nonnegative rate pairs* (R_1, R_2) *that satisfy*

$$\begin{aligned}
R_1 &\leq I(U_0, U_1; Y_1, \tilde{Y}_1, V|Q) \\
&\quad - I(V; U_0, U_1, U_2, \tilde{Y}_2|\tilde{Y}_1, Y_1, Q) \\
R_2 &\leq I(U_0, U_2; Y_2, \tilde{Y}_2, V|Q) \\
&\quad - I(V; U_0, U_1, U_2, \tilde{Y}_1|\tilde{Y}_2, Y_2, Q) \\
R_1 + R_2 &\leq I(U_0, U_1; Y_1, \tilde{Y}_1, V|Q) + I(U_2; Y_2, \tilde{Y}_2, V|U_0, Q) \\
&\quad - I(V; U_0, U_1, U_2, \tilde{Y}_2|\tilde{Y}_1, Y_1, Q) - I(U_1; U_2|U_0, Q) \\
R_1 + R_2 &\leq I(U_0, U_2; Y_2, \tilde{Y}_2, V|Q) + I(U_1; Y_1, \tilde{Y}_1, V|U_0, Q)
\end{aligned}$$

[6]The subscript "proc." indicates that the transmitter processes the feedback information it receives.

$$-I(V;U_0,U_1,U_2,\tilde{Y}_1|\tilde{Y}_2,Y_2,Q)-I(U_1;U_2|U_0,Q)$$

$$R_1+R_2 \leq I(U_0,U_1;Y_1,\tilde{Y}_1,V|Q)+I(U_0,U_2;Y_2,\tilde{Y}_2,V|Q)$$

$$-I(V;U_0,U_1,U_2,\tilde{Y}_2|\tilde{Y}_1,Y_1,Q)$$

$$-I(V;U_0,U_1,U_2,\tilde{Y}_1|\tilde{Y}_2,Y_2,Q)-I(U_1;U_2|U_0,Q)$$

for some pmf $P_Q, P_{U_0U_1U_2|Q}, P_{\tilde{Y}_1|Y_1Q}, P_{\tilde{Y}_2|Y_2Q}, P_{V|U_0U_1U_2\tilde{Y}_1\tilde{Y}_2}$ *and some function* $f\colon \mathcal{X} \to \mathcal{U}_0\times\mathcal{U}_1\times\mathcal{U}_2\times\mathcal{Q}$ *where the feedback-rates have to satisfy*

$$I(Y_1;\tilde{Y}_1|U_0,U_1,U_2,\tilde{Y}_2,Q) \leq R_{\text{Fb},1} \tag{37a}$$

$$I(Y_2;\tilde{Y}_2|U_0,U_1,U_2,\tilde{Y}_1,Q) \leq R_{\text{Fb},2} \tag{37b}$$

$$I(Y_1,Y_2;\tilde{Y}_1,\tilde{Y}_2|U_0,U_1,U_2,Q) \leq R_{\text{Fb},1}+R_{\text{Fb},2}. \tag{37c}$$

and where $X=f(U_0,U_1,U_2,Q)$.

Proof: See Section VII-D. ∎

When the feedback rates $R_{\text{Fb},1}, R_{\text{Fb},2}$ are sufficiently large, we can choose $\tilde{Y}_i = Y_i$ for $i\in\{1,2\}$.

Corollary 2. *In the limit* $R_{\text{Fb},1}, R_{\text{Fb},2} \to \infty$, $\mathcal{C}_{\text{Fb}}$ *includes the set* $\mathcal{R}^{\infty}_{\text{proc.}}$ *of all nonnegative rate pairs* (R_1,R_2) *that satisfy*

$$R_1 \leq I(U_0,U_1;Y_1,V|Q)-I(V;U_0,U_1,U_2,Y_2|Y_1,Q) \tag{38a}$$

$$R_2 \leq I(U_0,U_2;Y_2,V|Q)-I(V;U_0,U_1,U_2,Y_1|Y_2,Q) \tag{38b}$$

$$R_1+R_2 \leq I(U_1;Y_1,V|U_0,Q)+I(U_2;Y_2,V|U_0,Q) \tag{38c}$$

$$-I(U_1;U_2|U_0,Q)+\min_{i\in\{1,2\}}\{I(U_0;Y_i,V|Q)$$

$$-I(V;U_0,U_1,U_2,Y_1,Y_2|Y_i,Q)\} \tag{38d}$$

$$R_1+R_2 \leq I(U_0,U_1;Y_1,V|Q)-I(V;U_0,U_1,U_2,Y_2|Y_1,Q)$$

$$+I(U_0,U_2;Y_2,V|Q)-I(V;U_0,U_1,U_2,Y_1|Y_2,Q)$$

$$-I(U_1;U_2|U_0,Q) \tag{38e}$$

for some pmf $P_Q, P_{U_0U_1U_2|Q}P_{V|U_0U_1U_2Y_1Y_2}$ *and some function* $f\colon \mathcal{X}\to\mathcal{U}_0\times\mathcal{U}_1\times\mathcal{U}_2\times\mathcal{Q}$, *where* $X = f(U_0,U_1,U_2,Q)$.

Remark 6. *The region* $\mathcal{R}^{\infty}_{\text{proc.}}$ *improves over the Shayevitz-Wigger region for output feedback when this latter is specialized to the choice* $V_1=V_2=V_0$. *Observe that except for the sum-rate constraints* (38d) *and* (12c)*, all other rate constraints defining* $\mathcal{R}_{\text{proc.}}$ *and the Shayevitz-Wigger region coincide when the latter are specialized to* $V_1 = V_2 = V_0$*. Since* $\min_{i=\{1,2\}}\{a_i - b_i\} \geq \min_{i\in\{1,2\}} a_i - \max_{i\in\{1,2\}} b_i$ *holds for any nonnegative* $\{a_i, b_i\}_{i=1}^2$*, we conclude that the rate region* $\mathcal{R}_{\text{proc.}}$ *contains the Shayevitz-Wigger region specialized to the choice* $V_1 = V_2 = V_0$*. As proved in [4], our region* $\mathcal{R}_{\text{proc.}}^{\infty}$ *thus also recovers the two-user capacity result in [5], [6] and the degrees of freedom achievability result in [7].*

B. Usefulness of Feedback

Our third scheme 1C (which leads to Theorem 3) can be used to prove the following result on the usefulness of rate-limited feedback for DMBCs. (Similar results can be shown based on our other proposed schemes.)

Theorem 5. *Fix a DMBC. Consider random variables* $(U_0^{(\text{M})}, U_1^{(\text{M})}, U_2^{(\text{M})}, X^{(\text{M})})$ *such that*

$$\Gamma^{(\text{M})} := I(U_0^{(\text{M})}; Y_2^{(\text{M})}) - I(U_0^{(\text{M})}; Y_1^{(\text{M})}) > 0. \tag{39}$$

Let the rate pair $(R_1^{(\text{M})}, R_2^{(\text{M})})$ *satisfy Marton's constraints* (6) *when evaluated for* $(U_0^{(\text{M})}, U_1^{(\text{M})}, U_2^{(\text{M})}, X^{(\text{M})})$ *where Constraint* (6b) *has to hold with strict inequality.*

Also, let $(R_1^{(\text{Enh})}, R_2^{(\text{Enh})})$ *be a rate pair in the capacity region* $\mathcal{C}_{\text{Enh}}^{(1)}$ *of the enhanced DMBC.*

If the feedback-rate from Receiver 1 is positive, $R_{\text{Fb},1} > 0$*, then for all sufficiently small* $\gamma \in (0,1)$*, the rate pair* (R_1, R_2)*,*

$$R_1 = (1-\gamma)R_1^{(\text{M})} + \gamma R_1^{(\text{Enh})} \tag{40a}$$

$$R_2 = (1-\gamma)R_2^{(\text{M})} + \gamma R_2^{(\text{Enh})} \tag{40b}$$

lies in $\mathcal{R}_{\text{relay,hb}}^{(1)}$*,*

$$(R_1, R_2) \in \mathcal{R}_{\text{relay,hb}}^{(1)}, \tag{41}$$

and is thus achievable.

An analogous statement holds when indices 1 *and* 2 *are exchanged.*

Proof: See Appendix D. ■

The following remark elaborates on the condition of the theorem that a rate pair satisfies Constraint (6b) with strict inequality.

Remark 7. *For given random variables* $U_0^{(\text{M})}, U_1^{(\text{M})}, U_2^{(\text{M})}, X^{(\text{M})}$ *Marton's region, i.e., the rate region defined by Constraints* (6)*, is either a pentagon (both single-rate constraints as well as at least one of* *the sum-rates are active), a quadrilateral (only the two single-rate constraints are active), or a triangle (only one single-rate constraint and at least one of the sum-rate constraints are active).*

In the case of superposition coding with $U_1^{(\mathrm{M})}$ *=const. and* $U_2^{(\mathrm{M})} = X^{(\mathrm{M})}$ *and when Condition* (39) *holds, then the region is a quadrilateral and the only active constraints are* (6a) *and* (6c)*. Thus, in this case, constraint* (6b) *holds with strict inequality for all rate pairs in this region.*

Whenever the region defined by Marton's constraints (6) *is a pentagon, then the only rate pair in this pentagon that satisfies Constraint* (6b) *with equality is the dominant corner point of maximum* R_2*-rate.*

Corollary 3. *Assume* $R_{\mathrm{Fb},1} > 0$*. If there exists a rate pair* $(R_1^{(\mathrm{M})}, R_2^{(\mathrm{M})})$ *that satisfies the conditions in Theorem 5 and that lies on the boundary of* $\mathcal{R}_{\mathrm{Marton}}$ *but strictly in the interior of* $\mathcal{C}_{\mathrm{Enh}}^{(1)}$*, then*

$$\mathcal{R}_{\mathrm{Marton}} \subsetneq \mathcal{C}_{\mathrm{Fb}}. \tag{42}$$

If for the considered DMBC moreover $\mathcal{R}_{\mathrm{Marton}} = \mathcal{C}_{\mathrm{NoFB}}$*,*

$$\mathcal{C}_{\mathrm{NoFB}} \subsetneq \mathcal{C}_{\mathrm{Fb}}. \tag{43}$$

Proof: Inclusion (43) follows from (42). We show (42). Since $(R_1^{(\mathrm{M})}, R_2^{(\mathrm{M})})$ is in the interior of $\mathcal{C}_{\mathrm{Enh}}^{(1)}$, there exists a rate pair $(R_1^{(\mathrm{Enh})}, R_2^{(\mathrm{Enh})}) \in \mathcal{C}_{\mathrm{Enh}}^{(1)}$ with $R_1^{(\mathrm{Enh})} > R_1^{(\mathrm{M})}$ and $R_2^{(\mathrm{Enh})} > R_2^{(\mathrm{M})}$. Now, since $(R_1^{(\mathrm{M})}, R_2^{(\mathrm{M})})$ lies on the boundary of $\mathcal{R}_{\mathrm{Marton}}$, the rate pair in (40) must lie outside $\mathcal{R}_{\mathrm{Marton}}$ for any $\gamma \in (0,1)$. By Theorem 5, Equation (41), this rate pair is achievable with rate-limited feedback for all $\gamma \in (0,1)$ that are sufficiently close to 0. ■

For many DMBCs such as the BSBC or the BEBC with unequal cross-over probabilities or unequal erasure probabilities to the two receivers, or the BSC/BEC-BC where the two channels have different capacities, the conditions of Corollary 3 can easily be checked. Thus, our corollary immediately shows that for these DMBCs rate-limited feedback strictly increases capacity. (See also Examples 1 and 2

For the BSBC and the BEBC, Theorem 5 can even be used to show that all the boundary points $(R_1 > 0, R_2 > 0)$ of $\mathcal{C}_{\mathrm{NoFB}}$ can be improved with rate-limited feedback, see the following Corollary 4, the paragraph thereafter, and Example 1 in the next Section.

More generally speaking, Corollary 3 is particularly interesting in view of the following class of BCs. We introduce the new term *strictly essentially less-noisy*.

Definition 3 (Strictly Essentially Less-Noisy)**.** *The definition of a* strictly essentially less-noisy *DMBC coincides with the definition of an essentially less-noisy DMBC except that Inequality* (10) *needs to be strict whenever* $I(U;Y_1) > 0$*.*

The BSBC and the BEBC with different cross-over probabilities or different erasure probabilities at the two receivers are strictly essentially less-noisy.

Corollary 4. *Consider a DMBC where Y_2 is strictly essentially less-noisy than Y_1. Assume $R_{\text{Fb},1} > 0$. We have:*

1) *If a rate pair (R_1, R_2) lies on the boundary of $\mathcal{C}_{\text{NoFB}}$ but in the interior of $\mathcal{C}_{\text{Enh}}^{(1)}$, then (R_1, R_2) lies in the interior of $\mathcal{C}_{\text{Fb}}$, i.e., with rate-limited feedback one can improve over this rate pair.*
2) *If $\mathcal{C}_{\text{NoFB}}$ does not coincide with $\mathcal{C}_{\text{Enh}}^{(1)}$, then $\mathcal{C}_{\text{NoFB}}$ is also a strict subset of $\mathcal{C}_{\text{Fb}}$, i.e., feedback strictly improves capacity.*

Analogous statements hold if indices 1 and 2 are exchanged.

As mentioned, all BSBCs and BEBCs with unequal cross-over probabilities or unequal erasure probabilities to the two receivers are strictly essentially less-noisy. Also, for these BCs $\mathcal{C}_{\text{NoFB}}$ has no common boundary points $(R_1 > 0, R_2 > 0)$ with the sets $\mathcal{C}_{\text{Enh}}^{(1)}$ or $\mathcal{C}_{\text{Enh}}^{(2)}$ unless the BC is physically degraded. Thus, for these BCs the corollary implies that, unless the BC is physically degraded, rate-limited feedback improves all boundary points $(R_1 > 0, R_2 > 0)$ of $\mathcal{C}_{\text{NoFB}}$ whenever $R_{\text{Fb},1}, R_{\text{Fb},2} > 0$.

Notice that when a DMBC is physically degraded in the sense that output Y_1 is a degraded version of Y_2, then $\mathcal{C}_{\text{NoFB}} = \mathcal{C}_{\text{Enh}}^{(1)}$. Of course (even infinite-rate) feedback does not increase the capacity of physically degraded DMBCs [1].

Proof of Corollary 4: 2.) follows from 1.) We prove 1.) For strictly essentially less-noisy DMBCs, $\mathcal{C}_{\text{NoFB}}$ is achieved by superposition coding. Thus, $\mathcal{R}_{\text{Marton}} = \mathcal{C}_{\text{NoFB}}$ and in the evaluation of Marton's region one can restrict attention to auxiliaries of the form $U_1 =$const. and $U_2 = X$. By the definition of strictly essentially-less noisy, when evaluating Marton's region we can further restrict attention to auxiliary random variables that satisfy (39). Thus, by Remark 7, any boundary point of $\mathcal{R}_{\text{Marton}}$ satisfies the conditions of Theorem 5. Repeating the proof steps for Corollary 3, we can prove that these boundary points cannot be boundary points of $\mathcal{C}_{\text{Fb}}$ whenever they lie in the interior of $\mathcal{C}_{\text{Enh}}^{(1)}$. ∎

VI. EXAMPLES

Example 1. *Consider the BSBC with input X and outputs Y_1 and Y_2 described by:*

$$Y_1 = X \oplus Z_1, \quad Y_2 = X \oplus Z_2, \tag{44a}$$

for $Z_1 \sim$ Bern(p_1) and $Z_2 \sim$ Bern(p_2) independent noises. Let $Q = const.$, $U \sim$ Bern$(1/2)$, $W_1 \sim$ Bern(β_1) and $W_2 \sim$ Bern(β_2), for $\beta_1, \beta_2 \in [0, 1/2]$, where U, W_1, W_2 are mutually independent. Also *set* $X = U \oplus W_1$*, and* $\tilde{Y}_1 = Y_1 \oplus W_2$*. Then*

$$I(U;Y_1) = 1 - H_b(\beta_1 * p_1), \quad I(X;Y_2) = 1 - H_b(p_2),$$

and

$$I(X;\tilde{Y}_1,Y_2|U) = H(\alpha_1,\alpha_2,\alpha_3,\alpha_4) - H_b(p_2) - H_b(\beta_2 * p_1)$$

$$I(\tilde{Y}_1;Y_1|Y_2,U) = H(\alpha_1,\alpha_2,\alpha_3,\alpha_4) - H_b(\beta_1 * p_2) - H_b(\beta_2)$$

where

$$\alpha_1 = (p_1 * \beta_2)p_2\beta_1 + (1 - p_1 * \beta_2)\bar{p_2}\bar{\beta}_1$$

$$\alpha_2 = (p_1 * \beta_2)\bar{p_2}\beta_1 + (1 - p_1 * \beta_2)p_2\bar{\beta}_1$$

$$\alpha_3 = (p_1 * \beta_2)\bar{p_2}\bar{\beta}_1 + (1 - p_1 * \beta_2)p_2\beta_1$$

$$\alpha_4 = (p_1 * \beta_2)p_2\bar{\beta}_1 + (1 - p_1 * \beta_2)\bar{p_2}\beta_1.$$

For this choice, the region defined by the constraints in Corollary 1 evaluates to:

$$R_1 \leq 1 - H_b(\beta_1 * p_1) \tag{45a}$$

$$\begin{aligned} R_1 + R_2 \leq 1 &- H_b(\beta_1 * p_1) + H(\alpha_1,\alpha_2,\alpha_3,\alpha_4) \\ &- H_b(p_2) - H_b(\beta_2 * p_1) \end{aligned} \tag{45b}$$

$$\begin{aligned} R_1 + R_2 \leq 1 &- H_b(p_2) - H(\alpha_1,\alpha_2,\alpha_3,\alpha_4) \\ &+ H_b(\beta_1 * p_2) + H_b(\beta_2) \end{aligned} \tag{45c}$$

for some $\beta_1, \beta_2 \in [0, 1/2]$ *satisfying*

$$H(\alpha_1,\alpha_2,\alpha_3,\alpha_4) - H_b(\beta_1 * p_2) - H_b(\beta_2) \leq R_{\text{Fb},1} \tag{46}$$

and where $H(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ *denotes the entropy of a quaternary random variable with probability masses* $(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$.

The region is plotted in Figure 2 against the no-feedback capacity region $\mathcal{C}_{\text{NoFB}}$.

Example 2. *Consider a DMBC where the channel from* X *to* Y_1 *is a BSC with cross-over probability* $p \in (0, 1/2)$*, and the channel from* X *to* Y_2 *is an independent BEC with erasure probability* $e \in (0, 1)$*. We show that our feedback regions* $\mathcal{R}^{(1)}_{\text{relay,sp}}$ *and* $\mathcal{R}^{(2)}_{\text{relay,sp}}$ *improve over a large part of the boundary points of* $\mathcal{C}_{\text{NoFB}}$ *for all values of* e, p *unless* $H_b(p) = e$*, no matter how small* $R_{\text{Fb},1}, R_{\text{Fb},2} > 0$.

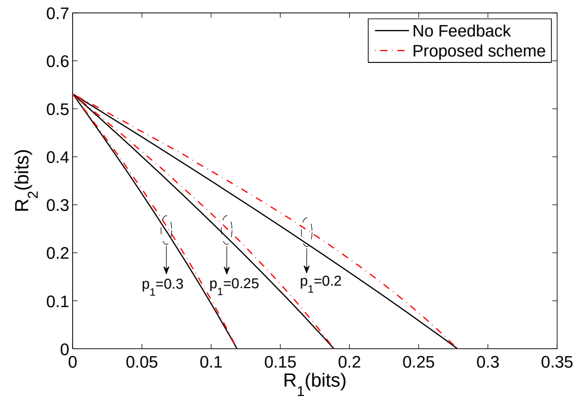

Fig. 2. $\mathcal{C}_{\text{NoFB}}$ and the achievable region in (45) are plotted for BSBCs with parameters $p_2 = 0.1$ and $p_1 \in \{0.2, 0.25, 0.3\}$ and for feedback rate $R_{\text{Fb},1} = 0.8$.

We distinguish different parameter ranges of our channel.

- $0 < e < H_b(p)$*: In this case, the nofeedback capacity region* $\mathcal{C}_{\text{NoFB}}$ *[26] is formed by the set of rate pairs* (R_1, R_2) *that for some* $s \in [0, 1/2]$ *satisfy*

$$R_1 \leq 1 - H_b(s * p), \tag{47a}$$

$$R_2 \leq (1-e)H_b(s), \tag{47b}$$

$$R_1 + R_2 \leq 1 - e. \tag{47c}$$

We specialize the region $\mathcal{R}^{(1)}_{\text{relay,sp}}$ *to the following choices. Let* $Q = const.$, $U \sim \text{Bern}(1/2)$, $X = U \oplus V$*, where* $V \sim \text{Bern}(s)$ *independent of* U*, and* $\tilde{Y}_1 = Y_1$ *with probability* $\gamma \in (0,1)$ *and* $\tilde{Y}_1 = ?$ *with probability* $1-\gamma$*, where*

$$\gamma \leq \frac{R_{\text{Fb},1}}{(1-e)H_b(p) + eH_b(s*p)}. \tag{48}$$

Condition (48) *assures that the described choice satisfies* (35)*. Then,*

$$I(U;Y_1) = 1 - H_b(s*p), \quad I(X;Y_2) = 1-e,$$

and

$$I(X;\tilde{Y}_1, Y_2|U) = \gamma e(H_b(s*p) - H_b(p)) + (1-e)H_b(s)$$

$$I(\tilde{Y}_1; Y_1|Y_2, U) = \gamma(H_b(p)(1-e) + eH_b(s*p)).$$

When $R_{\text{Fb},1} > 0$*, by Corollary 1, all rate pairs* (R_1, R_2) *satisfying*

$$R_1 \leq 1 - H_b(s*p) \tag{49a}$$

$$R_1+R_2 \leq 1 - H_b(s*p) + (1-e)H_b(s)$$
$$+\gamma e(H_b(s*p) - H_b(p)) \tag{49b}$$
$$R_1+R_2 \leq 1-e-\gamma(H_b(p)(1-e)+eH_b(s*p)) \tag{49c}$$

are achievable for any $\gamma \in (0,1)$ *satisfying* (48).

As shown in [26], the points (R_1, R_2) *of the form*

$$(1 - H_b(s*p), (1-e)H_b(s)), \qquad s \in (0, s_0), \tag{50}$$

are all on the dominant boundary of $\mathcal{C}_{\text{NoFb}}$*, where* $s_0 \in (0, 1/2)$ *is the unique solution to*

$$1 - H_b(s_0 * p) + (1-e)H_b(s_0) = 1 - e. \tag{51}$$

For these boundary points, only the single-rate constraints (47a) *and* (47b) *are active, but not* (47c). *Thus, comparing* (50) *to our feedback region* (49)*, we can conclude that by choosing* γ *sufficiently small, all boundary points* (50) *lie strictly in the interior of our feedback region* $\mathcal{R}^{(1)}_{\text{relay,sp}}$ *when* $R_{\text{Fb},1} > 0$.

- $0 < H_b(p) < e < 1$*: The nofeedback capacity region* $\mathcal{C}_{\text{NoFb}}$ *equals the time-sharing region given by the union of all rate pairs* (R_1, R_2) *that for some* $\alpha \in [0,1]$ *satisfy*

$$R_1 \leq \alpha(1 - H_b(p)) \tag{52a}$$
$$R_2 \leq (1-\alpha)(1-e). \tag{52b}$$

We specialize the region $\mathcal{R}^{(2)}_{\text{relay,sp}}$ *to the following choices:* $Q \sim \text{Bern}(\alpha)$*; if* $Q = 0$ *then* $U \sim \text{Bern}(1/2)$*,* $X = U$*, and* $\tilde{Y}_2 = const.$*; if* $Q = 1$ *then* U =*const.,* $X \sim \text{Bern}(1/2)$*, and* $\tilde{Y}_2 = Y_2$ *with probability* $\gamma \in (0,1)$ *and* $\tilde{Y}_2 = ?$ *with probability* $1-\gamma$*, where in order to satisfy the average feedback rate constraint,*

$$\gamma \leq \frac{R_{\text{Fb},2}}{\alpha((1-e)H_b(p) + H_b(e))}. \tag{53}$$

When $R_{\text{Fb},2} > 0$*, by Theorem 3, all rate pairs* (R_1, R_2) *satisfying*

$$R_1 \leq \alpha(1 - H_b(p)) + \alpha(1-e)\gamma H_b(p) \tag{54a}$$
$$R_1 + R_2 \leq (1-\alpha)(1-e) + \alpha(1 - H_b(p))$$
$$+ \alpha(1-e)\gamma H_b(p) \tag{54b}$$
$$R_1 + R_2 \leq (1 - H_b(p)) - (1-\alpha)\gamma H_b(e). \tag{54c}$$

are achievable for any $\gamma \in (0,1)$ *satisfying* (53).

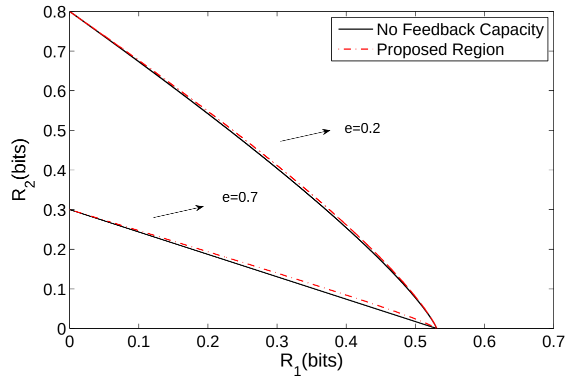

Fig. 3. $\mathcal{C}_{\text{NoFB}}$ and the achievable regions in (49) and (54) are plotted for a BSC/BEC-BC when the BSC has parameter $p = 0.1$ and the BEC has parameter $e \in \{0.2, 0.7\}$. Notice that $0.2 < H_b(p) < 0.7$. The feedback rates $R_{\text{Fb},1} = R_{\text{Fb},2} = 0.8$.

Since here $1 - H_b(p) > 1 - e$, for small $\gamma > 0$ the feedback region in (54) *improves over $\mathcal{C}_{\text{NoFB}}$ given in* (52)*. In fact,* (54) *improves over all boundary points $(R_1 > 0, R_2 > 0)$ of $\mathcal{C}_{\text{NoFB}}$.*

Remark 8. *The BSC/BEC-BC in Example 2, is particularly interesting, because depending on the values of the parameters e and p, the BC is either degraded, less noisy, more capable, or essentially less-noisy [26]. We conclude that our feedback regions $\mathcal{R}^{(1)}_{\text{relay,sp}}$ and $\mathcal{R}^{(2)}_{\text{relay,sp}}$ can improve over the nofeedback capacity regions for all these classes of BCs even with only one feedback link that is of arbitrary small, but positive rate.*

We plotted our regions (49) *and* (54) *versus the nofeedback capacity region in Figure 3 for $p = 0.1$ and $e = 0.2$ or $e = 0.7$. In the first case the DMBC is more capable and in the second case it is essentially less-noisy.*

In the next example we consider the Gaussian BC with independent noises. We evaluate the region defined by the constraints of Corollary 1 for a set of jointly Gaussian distributions on the input and the auxiliary random variables. A rigorous proof that our achievability result in Corollary 1 holds also for the Gaussian BC and Gaussian random variables is omitted for brevity.

Example 3. *Consider the* Gaussian broadcast channel

$$Y_1 = X + Z_1 \tag{55a}$$

$$Y_2 = X + Z_2 \tag{55b}$$

where $Z_1 \sim \mathcal{N}(0, N_1)$ and $Z_2 \sim \mathcal{N}(0, N_2)$ are independent noises. Assume an average transmission *power* P*, and* $0 < N_2 < N_1 < P$*.*

Let $Q = \text{const.}$*,* $U \sim \mathcal{N}(0, \bar{\alpha}P)$*,* $W_1 \sim \mathcal{N}(0, \alpha P)$ *and* $W_2 \sim \mathcal{N}(0, \beta)$*, for* $\alpha \in [0,1], \beta > 0$*, where* U, W_1, W_2 *are mutually independent. Set* $X = U + W_1$*,* $\tilde{Y}_1 = Y_1 + W_2$*, then*

$$I(U;Y_1) = C\Big(\frac{\bar{\alpha}P}{\alpha P + N_1}\Big), \quad I(X;Y_2) = C\Big(\frac{P}{N_2}\Big),$$

and

$$\begin{aligned} I(X;Y_2,\tilde{Y}_1|U) &= C\Big(\frac{\alpha P}{N_2}\Big) + C\Big(\frac{\alpha P N_2}{(\alpha P + N_2)(N_1+\beta)}\Big) \\ I(\tilde{Y}_1;Y_1|Y_2,U) &= C\Big(\frac{\alpha P(N_1+N_2)+N_1N_2}{\beta(N_2+\alpha P)}\Big). \end{aligned}$$

For these choices, the region defined by the constraints in Corollary 1 evaluates to:

$$R_1 \leq C\Big(\frac{\bar{\alpha}P}{\alpha P + N_1}\Big) \tag{56a}$$

$$\begin{aligned} R_1 + R_2 \leq\; & C\Big(\frac{\bar{\alpha}P}{\alpha P + N_1}\Big) + C\Big(\frac{\alpha P}{N_2}\Big) \\ & + C\Big(\frac{\alpha P N_2}{(\alpha P + N_2)(N_1+\beta)}\Big) \end{aligned} \tag{56b}$$

$$R_1 + R_2 \leq C\Big(\frac{P}{N_2}\Big) - C\Big(\frac{\alpha P(N_1+N_2)+N_1N_2}{\beta(N_2+\alpha P)}\Big) \tag{56c}$$

for some $\alpha \in [0,1]$ *and* $\beta > 0$ *satisfying*

$$C\Big(\frac{\alpha P(N_1+N_2)+N_1N_2}{\beta(N_2+\alpha P)}\Big) \leq R_{\text{Fb},1}. \tag{57}$$

Here, we use $C(x) := \frac{1}{2}\log(1+x)$*, for any* $x \geq 0$*.*

The region is plotted in Figure 4 against the no-feedback capacity region $\mathcal{C}_{\text{NoFB}}$ *and the region achieved by Ozarow&Leung's coding scheme [35].*

Example 4. *(Blackwell Channel with State [4]) We consider the Blackwell DMBC with random state. The state is described by a random variable* $S \sim \text{Bern}(1/2)$*, which is also part of the outputs. That means Receiver 1's output is* $Y_1 = (Y_1^*, S)$ *and Receiver 2's output is* $Y_2 = (Y_2^*, S)$*. If* $S = 0$ *then the BC to* Y_1^* *and* Y_2^* *is a reversed Blackwell channel:*

$$Y_1^* = \begin{cases} 0 & X = 0 \\ 1 & X = 1,2 \end{cases} \qquad Y_2^* = \begin{cases} 0 & X = 0,2 \\ 1 & X = 1. \end{cases} \tag{58}$$

If $S = 1$*, then the BC to* Y_1^* *and* Y_2^* *is a standard Blackwell channel:*

$$Y_1^* = \begin{cases} 0 & X = 0,2 \\ 1 & X = 1 \end{cases} \qquad Y_2^* = \begin{cases} 0 & X = 0 \\ 1 & X = 1,2. \end{cases} \tag{59}$$

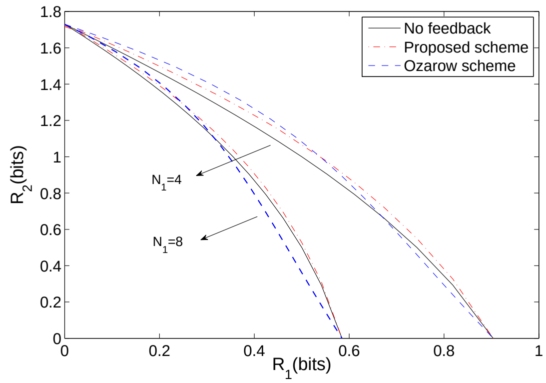

Fig. 4. $\mathcal{C}_{\text{NoFB}}$ and the achievable region in (56) are plotted for Gaussian BCs with parameters $P = 10$, $N_2 = 1$, $N_1 \in \{4, 8\}$ and feedback rate $R_{\text{Fb},1} = 0.8$.

For this BC, the nofeedback capacity region is achieved by time-sharing and the maximum sum-rate is 1. In [4] it was shown that the Shayevitz-Wigger scheme with choices of auxiliary random variables as in (13) *and* (14) *achieves the rate pairs* $(0.5958, 0.5958)$ *and* $(0.6103, 0.6103)$*, respectively. By Remark 4, we obtain that the proposed scheme pertaining to Theorem 2 can enlarge the nofeedback-capacity of this BC. Notice that for this setup,* $I(U;Y_2) - I(U;Y_1) = 0$ *holds for all* P_{UX}*, which means the statement above holds even when one of the receivers is not "stronger" than the other.*

VII. Coding Schemes

A. Coding Scheme 1A: Sliding-Window Decoding (Theorem 1)

For simplicity, we only describe the scheme for Q =const. A general Q can be introduced by coded time-sharing [28, Section 4.5.3]. That means all the codebooks need to be superpositioned on a P_Q-i.i.d. random vector Q^n that is revealed to transmitter and receivers, and this Q^n sequence needs to be included in all the joint-typicality checks.

Choose nonnegative rates $R_1', R_2', \tilde{R}_1, \tilde{R}_2, \hat{R}_1, \hat{R}_2$, auxiliary finite alphabets $\mathcal{U}_0, \mathcal{U}_1, \mathcal{U}_2, \tilde{\mathcal{Y}}_1, \tilde{\mathcal{Y}}_2$, a function f of the form $f\colon \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$, and pmfs $P_{U_0U_1U_2}$, $P_{\tilde{Y}_1|U_0Y_1}$, $P_{\tilde{Y}_2|U_0Y_2}$. Transmission takes place over $B+1$ consecutive blocks, with length n for each block. We denote the n-length blocks of inputs and outputs in block b by x_b^n, $y_{1,b}^n$ and $y_{2,b}^n$.

Define $\mathcal{J}_i := \{1, \ldots, \lfloor 2^{n\hat{R}_i} \rfloor\}$, $\mathcal{K}_i := \{1, \ldots, \lfloor 2^{nR_i'} \rfloor\}$, and $\mathcal{L}_i := \{1, \ldots, \lfloor 2^{n\tilde{R}_i} \rfloor\}$, for $i \in \{1, 2\}$. The messages are in product form: $M_i = (M_{i,1}, \ldots, M_{i,B})$, $i \in \{1, 2\}$, with $M_{i,b} = (M_{c,i,b}, M_{p,i,b})$ for $b \in \{1, \ldots, B\}$. The submessages $M_{c,i,b}$, and $M_{p,i,b}$ are uniformly distributed over the sets $\mathcal{M}_{c,i} :=$

$\{1,\ldots,\lfloor 2^{nR_{c,i}}\rfloor\}$ and $\mathcal{M}_{p,i} := \{1,\ldots,\lfloor 2^{nR_{p,i}}\rfloor\}$, respectively, where $R_{p,i}, R_{c,i} > 0$ and so that $R_i = R_{p,i} + R_{c,i}$[7]. Let $R_c := (R_{c,1} + R_{c,2} + \tilde{R}_1 + \tilde{R}_2)$.

1) Codebook generation: For each block $b \in \{1,\ldots,B+1\}$, randomly and independently generate 2^{nR_c} sequences $u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$, for $\mathbf{m}_{c,b} \in \mathcal{M}_c := \mathcal{M}_{c,1} \times \mathcal{M}_{c,2}$ and $l_{i,b-1} \in \mathcal{L}_i$, for $i \in \{1,2\}$. (We use vector notation for $\mathbf{m}_{c,b}$ to emphasize that it represents a pair of indices.) Each sequence $u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$ is drawn according to the product distribution $\prod_{t=1}^n P_{U_0}(u_{0,b,t})$, where $u_{0,b,t}$ denotes the t-th entry of $u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$.

For $i \in \{1,2\}$ and each $u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$ randomly and conditionally independently generate [28] $2^{n(R_{p,i}+R_i')}$ sequences $u_{i,b}^n(m_{p,i,b}, k_{i,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$, for $m_{p,i,b} \in \mathcal{M}_{p,i}$ and $k_{i,b} \in \mathcal{K}_i$, where each $u_{i,b}^n(m_{p,i,b}, k_{i,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$ is drawn according to the product distribution $\prod_{t=1}^n P_{U_i|U_0}(u_{i,b,t}|u_{0,b,t})$, where $u_{i,b,t}$ denotes the t-th entry of $u_{i,b}^n\big(m_{p,i,b}, k_{i,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}\big)$.

Similarly, for $i \in \{1,2\}$ and each tuple $(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}) \in \mathcal{M}_c \times \mathcal{L}_1 \times \mathcal{L}_2$ randomly generate $2^{n(\tilde{R}_i+\hat{R}_i)}$ sequences $\tilde{y}_{i,b}^n(l_{i,b}, j_{i,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}\big)$, for $l_{i,b} \in \mathcal{L}_i$ and $j_{i,b} \in \mathcal{J}_i$, by drawing each $\tilde{y}_{i,b}^n(l_{i,b}, j_{i,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}\big)$ according to the product distribution $\prod_{t=1}^n P_{\tilde{Y}_i|U_0,Y_i}(\tilde{y}_{i,b,t}|u_{0,b,t})$ where $\tilde{y}_{i,b,t}$ denotes the t-th entry of $\tilde{y}_{i,b}^n(l_{i,b}, j_{i,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}\big)$.

All codebooks are revealed to transmitter and receivers.

2) Encoding: We describe the encoding for a fixed block $b \in \{1,\ldots,B+1\}$. Assume that $M_{c,i,b} = m_{c,i,b}$, $M_{p,i,b} = m_{p,i,b}$, for $i \in \{1,2\}$ and that the feedback messages sent after block $b-1$ are $L_{1,b-1} = l_{1,b-1}$ and $L_{2,b-1} = l_{2,b-1}$. Define $\mathbf{m}_{c,b} := (m_{c,1,b}, m_{c,2,b})$. To simplify notation, let $l_{i,0} = m_{c,i,B+1} = m_{p,i,B+1} = 1$, for $i \in \{1,2\}$ and $\mathbf{m}_{c,B+1} = (1,1)$.

The transmitter looks for a pair $(k_{1,b}, k_{2,b}) \in \mathcal{K}_1 \times \mathcal{K}_2$ that satisfies

$$\begin{aligned}&\big(u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}),\\ &\quad u_{1,b}^n(m_{p,1,b}, k_{1,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}),\\ &\qquad u_{2,b}^n(m_{p,2,b}, k_{2,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})\big) \in \mathcal{T}_{\varepsilon/16}^n(P_{U_0U_1U_2}).\end{aligned} \tag{60}$$

If there is exactly one pair $(k_{1,b}, k_{2,b})$ that satisfies the above condition, the transmitter chooses this pair.

[7]Due to the floor operations and since transmission takes place over $B+1$ blocks whereas the messages M_1 and M_2 are split into only B submessages, R_1 and R_2 here do not exactly represent the transmission rates of messages M_1 and M_2. In the limit $n \to \infty$ and $B \to \infty$, which is our case of interest, R_1 and R_2 however approach these transmission rates. Therefore, we neglect this technicality in the following.

If there are multiple such pairs, it chooses one of them uniformly at random. Otherwise it chooses a pair $(k_{1,b}, k_{2,b})$ uniformly at random over the entire set $\mathcal{K}_1 \times \mathcal{K}_2$. In block b the transmitter then sends the inputs $x_b^n = (x_{b,1}, \ldots, x_{b,n})$, where

$$x_{b,t} = f(u_{0,b,t}, u_{1,b,t}, u_{2,b,t}), \qquad t \in \{1, \ldots, n\}, \tag{61}$$

and $u_{0,b,t}$, $u_{1,b,t}$, $u_{2,b,t}$ denote the t-th symbols of the chosen Marton codewords $u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$, $u_{1,b}^n(m_{p,1,b}, k_{1,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$, and
$u_{2,b}^n(m_{p,2,b}, k_{2,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$.

3) Decoding and Generation of Feedback Messages at Receivers: We describe the operations performed at Receiver 1. Receiver 2 behaves in an analogous way.

After each block $b \in \{1, \ldots, B+1\}$, and after observing the outputs $y_{1,b}^n$, Receiver 1 looks for a pair of indices $(\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{l}_{2,b-1}) \in \mathcal{M}_c \times \mathcal{L}_2$ that satisfies

$$\big(u_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(1)}, l_{1,b-1}, \hat{l}_{2,b-1}), y_{1,b}^n\big) \in \mathcal{T}_{\varepsilon/8}^n(P_{U_0Y_1}).$$

Notice that Receiver 1 already knows $l_{1,b-1}$ because it has created it itself after the previous block $b-1$. If there are multiple such pairs, the receiver chooses one of them at random. If there is no such pair, then it chooses $(\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{l}_{2,b-1})$ randomly over the set $\mathcal{M}_c \times \mathcal{L}_2$.

After decoding the cloud center in block b, Receiver 1 then looks for a tuple $(\hat{j}_{2,b-1}, \hat{m}_{p,1,b-1}, \hat{k}_{1,b-1}) \in \mathcal{J}_2 \times \mathcal{M}_{p,1} \times \mathcal{K}_1$ that satisfies

$$\begin{aligned}\big(&u_{0,b-1}^n(\hat{\mathbf{m}}_{c,b-1}^{(1)}, l_{1,b-2}, \hat{l}_{2,b-2}), y_{1,b-1}^n,\\ &u_{1,b-1}^n(\hat{m}_{p,1,b-1}, \hat{k}_{1,b-1}|\hat{\mathbf{m}}_{c,b-1}^{(1)}, l_{1,b-2}, \hat{l}_{2,b-2}),\\ &\tilde{y}_{2,b-1}^n(\hat{l}_{2,b-1}, \hat{j}_{2,b-1}|\hat{\mathbf{m}}_{c,b-1}^{(1)}, l_{1,b-2}, \hat{l}_{2,b-2})\big) \in \mathcal{T}_\epsilon^n(P_{U_0U_1Y_1\tilde{Y}_2}).\end{aligned}$$

It further looks for a pair $(l_{1,b}, j_{1,b}) \in \mathcal{L}_1 \times \mathcal{J}_1$ that satisfies

$$\begin{aligned}(&\tilde{y}_{1,b}^n(l_{1,b}, j_{1,b}|\hat{\mathbf{m}}_{c,b}^{(1)}, l_{1,b-1}, \hat{l}_{2,b-1}),\\ &\qquad u_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(1)}, l_{1,b-1}, \hat{l}_{2,b-1}), y_{1,b}^n) \in \mathcal{T}_{\varepsilon/4}^n(P_{Y_1U_0\tilde{Y}_1})\end{aligned}$$

and sends the index $l_{1,b}$ over the feedback link. If there is more than one such pair $(l_{1,b}, j_{1,b})$ the encoder chooses one of them at random. If there is none, it chooses the index $l_{1,b}$ that it sends over the feedback link uniformly at random over $\mathcal{L}_1$. The receivers thus only send a feedback message at the end of each block $1, \ldots, B$.

After decoding Block $B+1$, Receiver 1 produces the product message $\hat{m}_1 = (\hat{m}_{1,1}, \ldots, \hat{m}_{1,B})$ as its guess, where $\hat{m}_{1,b} = (\hat{m}^{(1)}_{c,1,b}, \hat{m}_{p,1,b})$, for $b \in \{1, \ldots, B\}$, and $\hat{m}^{(1)}_{c,1,b}$ denotes the first component of $\hat{\mathbf{m}}^{(1)}_{c,b}$.

5) Analysis: See Appendix A.

B. Coding Scheme 1B: Backward Decoding (Theorem 2)

For simplicity, we describe the scheme without the coded time-sharing random variable Q, i.e., for $Q =$const.

Choose nonnegative rates $R'_1, R'_2, \tilde{R}_1, \tilde{R}_2, \hat{R}_1, \hat{R}_2$, auxiliary finite alphabets $\mathcal{U}_0, \mathcal{U}_1, \mathcal{U}_2, \tilde{\mathcal{Y}}_1, \tilde{\mathcal{Y}}_2$, a function f of the form $f\colon \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$, and pmfs $P_{U_0U_1U_2}$, $P_{\tilde{Y}_1|Y_1}$, $P_{\tilde{Y}_2|Y_2}$. Transmission takes place over $B+1$ consecutive blocks, with length n for each block. We denote the n-length blocks of inputs and outputs in block b by x^n_b, $y^n_{1,b}$ and $y^n_{2,b}$.

Define $\mathcal{J}_i := \{1, \ldots, \lfloor 2^{n\hat{R}_i} \rfloor\}$, $\mathcal{K}_i := \{1, \ldots, \lfloor 2^{nR'_i} \rfloor\}$, and $\mathcal{L}_i := \{1, \ldots, \lfloor 2^{n\tilde{R}_i} \rfloor\}$, for $i \in \{1, 2\}$. The messages are in product form: $M_i = (M_{i,1}, \ldots, M_{i,B})$, $i \in \{1, 2\}$, with $M_{i,b} = (M_{c,i,b}, M_{p,i,b})$ for $b \in \{1, \ldots, B\}$. The submessages $M_{c,i,b}$, and $M_{p,i,b}$ are uniformly distributed over the sets $\mathcal{M}_{c,i} := \{1, \ldots, \lfloor 2^{nR_{c,i}} \rfloor\}$ and $\mathcal{M}_{p,i} := \{1, \ldots, \lfloor 2^{nR_{p,i}} \rfloor\}$, respectively, where $R_{p,i}, R_{c,i} > 0$ and so that $R_i = R_{p,i} + R_{c,i}$. Let $R_c := (R_{c,1} + R_{c,2} + \tilde{R}_1 + \tilde{R}_2)$.

1) Codebook generation: For each block $b \in \{1, \ldots, B+1\}$, randomly and independently generate 2^{nR_c} sequences $u^n_{0,b}(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$, for $\mathbf{m}_{c,b} \in \mathcal{M}_c := \mathcal{M}_{c,1} \times \mathcal{M}_{c,2}$ and $l_{i,b-1} \in \mathcal{L}_i$, for $i \in \{1, 2\}$. Each sequence $u^n_{0,b}(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$ is drawn according to the product distribution $\prod_{t=1}^n P_{U_0}(u_{0,b,t})$, where $u_{0,b,t}$ denotes the t-th entry of $u^n_{0,b}(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$.

For $i \in \{1, 2\}$ and each tuple $(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$ randomly generate $2^{n(R_{p,i}+R'_i)}$ sequences $u^n_{i,b}(m_{p,i,b}, k_{i,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$, for $m_{p,i,b} \in \mathcal{M}_{p,i}$ and $k_{i,b} \in \mathcal{K}_i$ by randomly drawing each codeword $u^n_{i,b}(m_{p,i,b}, k_{i,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$ according to the product distribution $\prod_{t=1}^n P_{U_i|U_0}(u_{i,b,t}|u_{0,b,t})$, where $u_{i,b,t}$ denotes the t-th entry of $u^n_{i,b}\big(m_{p,i,b}, k_{i,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}\big)$.

Also, for $i \in \{1, 2\}$, generate $2^{n(\tilde{R}_i+\hat{R}_i)}$ sequences $\tilde{y}^n_{i,b}\big(l_{i,b}, j_{i,b}\big)$, for $l_{i,b} \in \mathcal{L}_i$ and $j_{i,b} \in \mathcal{J}_i$, by drawing all the entries independently according to the same distribution $P_{\tilde{Y}_i}$.

All codebooks are revealed to transmitter and receivers.

2) Encoding: We describe the encoding for a fixed block $b \in \{1, \ldots, B+1\}$. Assume that $M_{c,i,b} = m_{c,i,b}$, $M_{p,i,b} = m_{p,i,b}$, for $i \in \{1, 2\}$, and that the feedback messages sent after block $b-1$ are $L_{1,b-1} = l_{1,b-1}$ and $L_{2,b-1} = l_{2,b-1}$. Define $\mathbf{m}_{c,b} := (m_{c,1,b}, m_{c,2,b})$. To simplify notation, let $l_{i,0} = m_{c,i,B+1} = m_{p,i,B+1} = 1$, for $i \in \{1, 2\}$ and $m_{c,B+1} = (1, 1)$.

The transmitter looks for a pair $(k_{1,b}, k_{2,b}) \in \mathcal{K}_1 \times \mathcal{K}_2$ that satisfies

$$\begin{aligned}&\big(u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}),\\&\quad u_{1,b}^n(m_{p,1,b}, k_{1,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1}),\\&\qquad u_{2,b}^n(m_{p,2,b}, k_{2,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})\big) \in \mathcal{T}^n_{\varepsilon/16}(P_{U_0U_1U_2}).\end{aligned} \tag{62}$$

If there is exactly one pair $(k_{1,b}, k_{2,b})$ that satisfies the above condition, the transmitter chooses this pair. If there are multiple such pairs, it chooses one of them uniformly at random. Otherwise it chooses a pair $(k_{1,b}, k_{2,b})$ uniformly at random over the entire set $\mathcal{K}_1 \times \mathcal{K}_2$. In block b the transmitter then sends the inputs $x_b^n = (x_{b,1}, \ldots, x_{b,n})$, where

$$x_{b,t} = f(u_{0,b,t}, u_{1,b,t}, u_{2,b,t}), \qquad t \in \{1, \ldots, n\}, \tag{63}$$

and $u_{0,b,t}$, $u_{1,b,t}$, $u_{2,b,t}$ denote the t-th symbols of the chosen Marton codewords $u_{0,b}^n(\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$, $u_{1,b}^n(m_{p,1,b}, k_{1,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$, and
$u_{2,b}^n(m_{p,2,b}, k_{2,b}|\mathbf{m}_{c,b}, l_{1,b-1}, l_{2,b-1})$.

3) Generation of Feedback Messages at Receivers: We describe the operations performed at Receiver 1. Receiver 2 behaves in an analogous way.

After each block $b \in \{1, \ldots, B\}$, and after observing the outputs $y_{1,b}^n$, Receiver 1 looks for a pair $(l_{1,b}, j_{1,b}) \in \mathcal{L}_1 \times \mathcal{J}_1$ that satisfies

$$(\tilde{y}_{1,b}^n(l_{1,b}, j_{1,b}), y_{1,b}^n) \in \mathcal{T}^n_{\varepsilon/4}(P_{Y_1\tilde{Y}_1}) \tag{64}$$

and sends the index $l_{1,b}$ over the feedback link. If there is more than one such pair $(l_{1,b}, j_{1,b})$ the encoder chooses one of them at random. If there is none, it chooses the index $l_{1,b}$ that it sends over the feedback link uniformly at random over $\mathcal{L}_1$.

In our scheme the receivers thus only send a feedback message at the end of each block $1, \ldots, B$.

4) Decoding at Receivers: We describe the operations performed at Receiver 1. Receiver 2 behaves in an analogous way.

The receivers apply backward decoding and thus start decoding only after the transmission terminates. Then, for each block $b \in \{1, \ldots, B+1\}$, starting with the last block $B+1$, Receiver 1 performs the following operations. From the previous decoding step in block $b+1$, it already knows the feedback message $l_{2,b}$. Moreover, it also knows its own feedback messages $l_{1,b-1}$ and $l_{1,b}$ because it has created them itself, see point *3)*. Now, when observing $y_{1,b}^n$, Receiver 1 looks for a tuple $(\hat{j}_{2,b}, \hat{\mathbf{m}}_{c,b}^{(1)}, \hat{l}_{2,b-1}, \hat{m}_{p,1,b}, \hat{k}_{1,b}) \in \mathcal{J}_2 \times \mathcal{M}_c \times \mathcal{L}_2 \times \mathcal{M}_{p,1} \times \mathcal{K}_1$ that satisfies

$$\big(u_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(1)}, l_{1,b-1}, \hat{l}_{2,b-1}), u_{1,b}^n(\hat{m}_{p,1,b}, \hat{k}_{1,b} | \hat{\mathbf{m}}_{c,b}^{(1)}, l_{1,b-1}, \hat{l}_{2,b-1}),$$
$$\tilde{y}_{2,b}^n(l_{2,b}, \hat{j}_{2,b}), y_{1,b}^n\big) \in \mathcal{T}_\epsilon^n(P_{U_0 U_1 Y_1 \tilde{Y}_2}).$$

After decoding Block 1, Receiver 1 produces the product message $\hat{m}_1 = (\hat{m}_{1,1}, \ldots, \hat{m}_{1,B})$ as its guess, where $\hat{m}_{1,b} = (\hat{m}_{c,1,b}^{(1)}, \hat{m}_{p,1,b})$, for $b \in \{1, \ldots, B\}$, and $\hat{m}_{c,1,b}^{(1)}$ denotes the first component of $\hat{\mathbf{m}}_{c,b}^{(1)}$.

5) Analysis: See Appendix B.

C. Coding Scheme 1C: Hybrid Sliding-Window Decoding and Backward Decoding (Theorem 3)

For simplicity, we only describe the scheme achieving region $\mathcal{R}_{\text{relay,hb}}^{(1)}$ for Q =const. A scheme achieving region $\mathcal{R}_{\text{relay,hb}}^{(2)}$ is obtained if in the following description indices 1 and 2 are exchanged.

1) Codebook generation: The codebooks are generated as in Scheme 1A, described in point 1) in Section VII-A, but where $\tilde{R}_2 = \hat{R}_2 = 0$.

2) Encoding: The transmitter performs the same encoding procedure as in Section VII-A, but where $l_{2,b-1} = 1$ is constant for each block $b \in \{1, \ldots, B+1\}$.

3) Receiver 1: In each block $b \in \{1, \ldots, B+1\}$, Receiver 1 first simultaneously decodes the cloud center and its satellite. Specifically, Receiver 1 looks for a tuple $(\hat{\mathbf{m}}_{c,b-1}, \hat{m}_{p,1,b-1}, \hat{k}_{1,b-1}) \in \mathcal{M}_c \times \mathcal{M}_{p,1} \times \mathcal{K}_1$ that satisfies

$$\big(u_{0,b-1}^n(\hat{\mathbf{m}}_{c,b-1}, l_{1,b-2}, 1), y_{1,b-1}^n,$$
$$u_{1,b-1}^n(\hat{m}_{p,1,b-1}, \hat{k}_{1,b-1} | \hat{\mathbf{m}}_{c,b-1}, l_{1,b-2}, 1)\big) \in \mathcal{T}_\epsilon^n(P_{U_0 U_1 Y_1}).$$

It further compresses the outputs $y_{1,b}^n$ and sends the feedback message $l_{1,b}$ over the feedback link as in Scheme 1A, see point 3) in Section VII-A.

4) Receiver 2: Receiver 2 performs backward decoding as in Scheme 1B, see point 4) in Section VII-B.

5) Analysis: Similar to the analysis of the schemes 1A and 1B in presented in appendices A and B. Details are omitted.

D. Coding Scheme 2: Encoder Processes Feedback-Info

The scheme described in this subsection differs from the previous scheme in that in each block b, after receiving the feedback messages $M_{\text{Fb},1,b}, M_{\text{Fb},2,b}$, the encoder first reconstructs the compressed versions of the channel outputs, $\tilde{Y}_{1,b}^n$ and $\tilde{Y}_{2,b}^n$, and then newly compresses the quintuple consisting of $\tilde{Y}_{1,b}^n$ and

$\tilde{Y}^n_{2,b}$ and the Marton codewords $U^n_{0,b}$, $U^n_{1,b}$, $U^n_{2,b}$ that it had sent during block b. This new compression information is then sent to the two receivers in the next-following block $b+1$ as part of the cloud center of Marton's code.

Decoding at the receivers is based on backward decoding. For each block b, each receiver $i \in \{1,2\}$ uses its observed outputs $Y^n_{i,b}$ to simultaneously reconstruct the encoder's compressed signal and decode its intended messages sent in block b.

For simplicity, we only describe the scheme for $Q =$const.

Choose nonnegative rates $R'_1, R'_2, \tilde{R}_1, \tilde{R}_2, \hat{R}_1, \hat{R}_2, \tilde{R}_v$, auxiliary finite alphabets $\mathcal{U}_0, \mathcal{U}_1, \mathcal{U}_2, \tilde{\mathcal{Y}}_1, \tilde{\mathcal{Y}}_2$, $\mathcal{V}$, a function f of the form $f\colon \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$, and pmfs $P_{U_0U_1U_2}$, $P_{\tilde{Y}_1|Y_1}$, $P_{\tilde{Y}_2|Y_2}$, and $P_{V|U_0U_1U_2\tilde{Y}_1\tilde{Y}_2}$. Transmission takes place over $B+1$ consecutive blocks, with length n for each block. We denote the n-length blocks of channel inputs and outputs in block b by x^n_b, $y^n_{1,b}$ and $y^n_{2,b}$.

Define $\mathcal{J}_i := \{1, \ldots, \lfloor 2^{n\hat{R}_i} \rfloor\}$, $\mathcal{K}_i := \{1, \ldots, \lfloor 2^{nR'_i} \rfloor\}$, and $\mathcal{L}_i := \{1, \ldots, \lfloor 2^{n\tilde{R}_i} \rfloor\}$, for $i \in \{1,2\}$, and $\mathcal{N} := \{1, \ldots, \lfloor 2^{n\tilde{R}_v} \rfloor\}$ The messages are in product form: $M_i = (M_{i,1}, \ldots, M_{i,B})$, $i \in \{1,2\}$, with $M_{i,b} = (M_{c,i,b}, M_{p,i,b})$ for $b \in \{1, \ldots, B\}$. The submessages $M_{c,i,b}$, and $M_{p,i,b}$ are uniformly distributed over the sets $\mathcal{M}_{c,i} := \{1, \ldots, \lfloor 2^{nR_{c,i}} \rfloor\}$ and $\mathcal{M}_{p,i} := \{1, \ldots, \lfloor 2^{nR_{p,i}} \rfloor\}$, respectively, where $R_{p,i}, R_{c,i} > 0$ and so that $R_i = R_{p,i} + R_{c,i}$. Let $R_c := (R_{c,1} + R_{c,2} + \tilde{R}_v)$.

1) Codebook generation: For each block $b \in \{1, \ldots, B+1\}$, randomly and independently generate 2^{nR_c} sequences $u^n_{0,b}(\mathbf{m}_{c,b}, n_{b-1})$, for $\mathbf{m}_{c,b} \in \mathcal{M}_c := \mathcal{M}_{c,1} \times \mathcal{M}_{c,2}$ and $n_{b-1} \in \mathcal{N}$. Each sequence $u^n_{0,b}(\mathbf{m}_{c,b}, n_{b-1})$ is drawn according to the product distribution $\prod_{t=1}^n P_{U_0}(u_{0,b,t})$, where $u_{0,b,t}$ denotes the t-th entry of $u^n_{0,b}(\mathbf{m}_{c,b}, n_{b-1})$.

For $i \in \{1,2\}$ and each pair $(\mathbf{m}_{c,b}, n_{b-1})$ randomly generate $2^{n(R_{p,i}+R'_i)}$ sequences $u^n_{i,b}(m_{p,i,b}, k_{i,b}|\mathbf{m}_{c,b}, n_{b-1})$, for $m_{p,i,b} \in \mathcal{M}_{p,i}$ and $k_{i,b} \in \mathcal{K}_i$, by drawing each codeword $u^n_{i,b}(m_{p,i,b}, k_{i,b}|\mathbf{m}_{c,b}, n_{b-1})$ according to the product distribution $\prod_{t=1}^n P_{U_i|U_0}(u_{i,b,t}|u_{0,b,t})$, where $u_{i,b,t}$ denotes the t-th entry of $u^n_{i,b}\big(m_{p,i,b}, k_{i,b}|\mathbf{m}_{c,b}, n_{b-1}\big)$.

Also, for $i \in \{1,2\}$, generate $2^{n(\tilde{R}_i + \hat{R}_i)}$ sequences $\tilde{y}^n_{i,b}\big(l_{i,b}, j_{i,b}\big)$, for $l_{i,b} \in \mathcal{L}_i$ and $j_{i,b} \in \mathcal{J}_i$ by drawing all the entries independently according to the same distribution $P_{\tilde{Y}_i}$;

Finally, for each $n_{b-1} \in \mathcal{N}$, generate 2^{nR_v} sequences $v^n_b(n_b|n_{b-1})$, for $n_b \in \mathcal{N}$ by drawing all entries independently according to the same distribution P_V.

All codebooks are revealed to transmitter and receivers.

2) Encoding: We describe the encoding for a fixed block $b \in \{1, \ldots, B+1\}$. Assume that in this block we wish to send messages $M_{c,i,b} = m_{c,i,b}$, $M_{p,i,b} = m_{p,i,b}$, for $i \in \{1,2\}$, and define $\mathbf{m}_{c,b} :=$

$(m_{c,1,b}, m_{c,2,b})$. To simplify notation, let $l_{i,0} = m_{c,i,B+1} = m_{p,i,B+1} = 1$, for $i \in \{1,2\}$, and also $n_{-1} = n_0 = 1$.

The first step in the encoding is to reconstruct the compressed outputs pertaining to the previous block $\tilde{Y}^n_{1,b-1}$ and $\tilde{Y}^n_{2,b-1}$. Assume that after block $b-1$ the transmitter received the feedback messages $L_{1,b-1} = l_{1,b-1}$ and $L_{2,b-1} = l_{2,b-1}$, and that in this previous block it had produced the Marton codewords $u^n_{0,b-1} := u^n_{0,b-1}(\mathbf{m}_{c,b-1}, n_{b-2})$, $u^n_{1,b-1} := u^n_{1,b-1}(m_{p,1,b-1}, k_{1,b-1}|\mathbf{m}_{c,b-1}, n_{b-2})$, and $u^n_{2,b-1} := u^n_{2,b-1}(m_{p,2,b-1}, k_{2,b-1}|\mathbf{m}_{c,b-1}, n_{b-2})$. The transmitter then looks for a pair $(\hat{j}_{1,b-1}, \hat{j}_{2,b-1}) \in \mathcal{J}_1 \times \mathcal{J}_2$ that satisfies

$$\begin{aligned}\big(u^n_{0,b-1}, u^n_{1,b-1}, u^n_{2,b-1}, \tilde{y}^n_{1,b-1}(l_{1,b-1}, \hat{j}_{1,b-1}),\\ \tilde{y}^n_{2,b-1}(l_{2,b-1}, \hat{j}_{2,b-1})\big) \in \mathcal{T}_{\varepsilon/4}(P_{U_0U_1U_2\tilde{Y}_1,\tilde{Y}_2}).\end{aligned}$$

In a second step the encoder produces the new compression information pertaining to block $b-1$, which it then sends to the receivers during block b. To this end, it looks for an index $\hat{n}_{b-1} \in \mathcal{N}$ that satisfies

$$\begin{aligned}\big(u^n_{0,b-1}, u^n_{1,b-1}, u^n_{2,b-1}, \tilde{y}^n_{1,b-1}(l_{1,b-1}, \hat{j}_{1,b-1}),\\ \tilde{y}^n_{2,b-1}(l_{2,b-1}, \hat{j}_{2,b-1}), v^n_{b-1}(\hat{n}_{b-1}|n_{b-2})\big)\\ \in \mathcal{T}_{\varepsilon/2}(P_{U_0U_1U_2\tilde{Y}_1,\tilde{Y}_2V}).\end{aligned}$$

The transmitter now sends the fresh data and the compression message $\hat{n}_{b-1}$ over the channel: It thus looks for a pair $(k_{1,b}, k_{2,b}) \in \mathcal{K}_1 \times \mathcal{K}_2$ that satisfies

$$\begin{aligned}\big(u^n_{0,b}(\mathbf{m}_{c,b}, \hat{n}_{b-1}),\\ u^n_{1,b}(m_{p,1,b}, k_{1,b}|\mathbf{m}_{c,b}, \hat{n}_{b-1}),\\ u^n_{2,b}(m_{p,2,b}, k_{2,b}|\mathbf{m}_{c,b}, \hat{n}_{b-1})\big) \in \mathcal{T}^n_{\epsilon/64}(P_{U_0U_1U_2}).\end{aligned}$$

If there is exactly one pair $(k_{1,b}, k_{2,b})$ that satisfies the above condition, the transmitter chooses this pair. If there are multiple such pairs, it chooses one of them uniformly at random. Otherwise it chooses a pair $(k_{1,b}, k_{2,b})$ uniformly at random over the entire set $\mathcal{K}_1 \times \mathcal{K}_2$. In block b the transmitter then sends the inputs $x^n_b = (x_{b,1}, \ldots, x_{b,n})$, where

$$x_{b,t} = f(u_{0,b,t}, u_{1,b,t}, u_{2,b,t}), \qquad t \in \{1, \ldots, n\}. \tag{65}$$

and $u_{0,b,t}$, $u_{1,b,t}$, $u_{2,b,t}$ denote the t-th symbols of the chosen Marton codewords $u^n_{0,b}(\mathbf{m}_{c,b}, \hat{n}_{b-1})$, $u^n_{1,b}(m_{p,1,b}, k_{1,b}|\mathbf{m}_{c,b}, \hat{n}_{b-1})$, and $u^n_{2,b}(m_{p,2,b}, k_{2,b}|\mathbf{m}_{c,b}, \hat{n}_{b-1})$.

3) Generation of Feedback Messages at Receivers: We describe the operations performed at Receiver 1. Receiver 2 behaves in an analogous way.

After each block $b \in \{1, \dots, B\}$, and after observing the outputs $y_{1,b}^n$, Receiver 1 looks for a pair of indices $(l_{1,b}, j_{1,b}) \in \mathcal{L}_1 \times \mathcal{J}_1$ that satisfies

$$(\tilde{y}_{1,b}^n(l_{1,b}, j_{1,b}), y_{1,b}^n) \in \mathcal{T}_{\varepsilon/16}^n(P_{Y_1\tilde{Y}_1}) \tag{66}$$

and sends the index $l_{1,b}$ over the feedback link. If there is more than one such pair $(l_{1,b}, j_{1,b})$ the encoder chooses one of them at random. If there is none, it chooses the index $l_{1,b}$ sent over the feedback link uniformly at random over $\mathcal{L}_1$.

In our scheme the receivers thus only send a feedback message at the end of each block.

4) Decoding at Receivers: We describe the operations performed at Receiver 1. Receiver 2 behaves in an analogous way.

The receivers apply backward decoding, so they wait until the end of the transmission. Then, for each block $b \in \{1, \dots, B+1\}$, starting with the last block $B+1$, Receiver 1 performs the following operations. From the previous decoding step in block $b+1$, it already knows the compression index n_b. Now, when observing $y_{1,b}^n$, Receiver 1 looks for a tuple $(\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{m}_{p,1,b}, \hat{k}_{1,b}, \hat{n}_{b-1}) \in \mathcal{M}_c \times \mathcal{M}_{p,1} \times \mathcal{K}_1 \times \mathcal{N}$ that satisfies

$$\begin{aligned}\big(u_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{n}_{b-1}), u_{1,b}^n(\hat{m}_{p,1,b}, \hat{k}_{1,b}|\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{n}_{b-1}),&\\ v_b^n(n_b|\hat{n}_{b-1}), y_{1,b}^n, \tilde{y}_{1,b}^n(l_{1,b}, j_{1,b})\big)&\\ \in \mathcal{T}_\epsilon^n(P_{U_0U_1VY_1\tilde{Y}_1}),&\end{aligned}$$

where recall that Receiver 1 knows the indices $l_{1,b}$ and $j_{1,b}$ because it has constructed them itself under *3)*.

After the decoding Block 1, Receiver 1 produces the product message $\hat{m}_1 = (\hat{m}_{1,1}, \dots, \hat{m}_{1,B})$ as its guess, where $\hat{m}_{1,b} = (\hat{m}_{c,1,b}^{(1)}, \hat{m}_{p,1,b})$, for $b \in \{1, \dots, B\}$, and $\hat{m}_{c,1,b}^{(1)}$ denotes the first component of $\hat{m}_{c,1,b}^{(1)}$.

5) Analysis: See Appendix C.

VIII. Extension: Noisy Feedback

Our results also apply to the related setup where the two feedback links are noisy channels of capacities $R_{\text{Fb},1}$ and $R_{\text{Fb},2}$ and where *the decoders can code over their feedback links*. The following three modifications to our coding schemes suffice to ensure that our achievable regions remain valid:

- We time-share two instances of our coding schemes: one scheme operates during the odd blocks of the BC and occupies the even blocks on the feedback links; the other scheme operates during the even blocks of the BC and occupies the odd blocks on the feedback links.
- Instead of sending after each block an uncoded feedback message over the feedback links, the receivers encode them using a capacity-achieving code for their feedback links and send these codewords during the next block.
- After each block, the transmitter first decodes the messages sent over the feedback links during this block, and then uses the decoded feedback-messages in the same way as it used them in the original scheme.

Let $\varepsilon_{\text{Fb},i,b}$, for $i = 1,2$, denote the event that during Block b there is an error in the feedback communication from Receiver i to the transmitter, and let ε denote the event that $\hat{M}_1 \neq M_1$ or $\hat{M}_2 \neq M_2$. Then,

$$\begin{aligned}
&\Pr[\hat{M}_1 \neq M_1 \text{ or } \hat{M}_2 \neq M_2] \\
&\quad \leq \Pr\left[\varepsilon \cup \left(\bigcup_{b=1}^{B} \varepsilon_{\text{Fb},1,b}\right) \cup \left(\bigcup_{b=1}^{B} \varepsilon_{\text{Fb},2,b}\right)\right] \\
&\quad \leq \Pr\left[\varepsilon \middle| \left(\bigcup_{b=1}^{B} \varepsilon_{\text{Fb},1,b}\right)^c \cap \left(\bigcup_{b=1}^{B} \varepsilon_{\text{Fb},2,b}\right)^c\right] \\
&\quad\quad + \Pr\left[\bigcup_{b=1}^{B} \varepsilon_{\text{Fb},1,b}\right] + \Pr\left[\bigcup_{b=1}^{B} \varepsilon_{\text{Fb},2,b}\right] \\
&\quad \leq \Pr\left[\varepsilon \middle| \left(\bigcup_{b=1}^{B} \varepsilon_{\text{Fb},1,b}\right)^c \cap \left(\bigcup_{b=1}^{B} \varepsilon_{\text{Fb},2,b}\right)^c\right] \\
&\quad\quad + \sum_{b=1}^{B} \Pr[\varepsilon_{\text{Fb},1,b}] + \Pr[\varepsilon_{\text{Fb},2,b}] \,.
\end{aligned} \tag{67}$$

Since we use capacity-achieving codes on the feedback links, the probabilities $\Pr[\varepsilon_{\text{Fb},1,b}]$ and $\Pr[\varepsilon_{\text{Fb},2,b}]$ vanish as the blocklength increases. When the feedback communications in all the blocks are error-free, then the probability of error in the setup with noisy feedback is no larger than that in the setup with noise-free feedback. Thus, under the corresponding rate constraints, also the probability $\Pr\left[\varepsilon \middle| \left(\bigcup_{b=1}^{B} \varepsilon_{\text{Fb},1,b}\right)^c \cap \left(\bigcup_{b=1}^{B} \varepsilon_{\text{Fb},2,b}\right)^c\right]$ vanishes as the blocklength increases. Combining all these observations proves that the rate regions in Theorems 1–4 are achievable also in a setup with noisy feedback if the receivers can code over the feedback links.

ACKNOWLEDGMENT

The authors would like to thank R. Timo for helpful discussions and the city of Paris who supported this work under the "Emergences" programm.

APPENDIX A
ANALYSIS OF SCHEME 1A (THEOREM 1)

By the symmetry of our code construction, the probability of error does not depend on the realizations of $M_{c,i,b}$, $M_{p,i,b}$, $K_{i,b}$, $J_{i,b}$, $L_{i,b}$, for $i \in \{1,2\}$ and $b \in \{1,\ldots,B\}$. To simplify exposition we therefore assume that $M_{c,i,b} = M_{p,i,b} = K_{i,b} = J_{i,b} = L_{i,b} = 1$ for all $i \in \{1,2\}$ and $b \in \{1,\ldots,B\}$. Under this assumption, an error occurs if, and only if, for some $b \in \{1,\ldots,B\}$,

$$(\hat{M}_{p,1,b}, \hat{M}_{p,2,b}, \hat{M}^{(1)}_{c,1,b}, \hat{M}^{(2)}_{c,2,b}) \neq (1,1,1,1).$$

For each $b \in \{1,\ldots,B\}$, let ϵ_b denote the event that in our coding scheme at least one of the following holds for $i \in \{1,2\}$:

- $\hat{J}_{i,b-1} \neq 1$;
- $\hat{K}_{i,b-1} \neq 1$;
- $\hat{L}_{i,b-1} \neq 1$;
- $\hat{M}_{p,i,b-1} \neq 1$;
- $\hat{\mathbf{M}}^{(i)}_{c,b} \neq (1,1)$;
- There is no pair $(k_{1,b}, k_{2,b}) \in \mathcal{K}_1 \times \mathcal{K}_2$ that satisfies
$$\big(U^n_{0,b}(\mathbf{1}_{[4]}), U^n_{1,b}(1, k_{1,b}|\mathbf{1}_{[4]}), U^n_{2,b}(1, k_{2,b}|\mathbf{1}_{[4]})\big) \in \mathcal{T}^n_{\varepsilon/16}(P_{U_0U_1U_2})$$
- $\big(U^n_{0,b-1}(\mathbf{1}_{[4]}), U^n_{1,b-1}(1,1|\mathbf{1}_{[4]}), U^n_{2,b-1}(1,1,|\mathbf{1}_{[4]}),$ $Y^n_{1,b-1}, Y^n_{2,b-1}\big) \notin \mathcal{T}^n_{\varepsilon/12}(P_{U_0U_1U_2Y_1Y_2})$
- There is no pair $(l_{i,b}, j_{i,b}) \in \mathcal{L}_i \times \mathcal{J}_i$ that satisfies
$$\big(\tilde{Y}^n_{i,b}(l_{i,b}, j_{i,b}|\mathbf{1}_{[4]}), U^n_{0,b}(\mathbf{1}_{[4]}), Y^n_{i,b}\big) \in \mathcal{T}^n_{\varepsilon/4}(P_{\tilde{Y}_iU_0Y_i}).$$

Then,

$$P_e^{(N)} \leq \Pr\left[\bigcup_{b=1}^{B+1} \epsilon_b\right] \leq \sum_{b=2}^{B+1} \Pr\big[\epsilon_b | \epsilon^c_{b-1}\big] + \Pr[\epsilon_1]. \tag{68}$$

In the following we analyze the probabilities of these events averaged over the random code construction. In particular, we shall identify conditions such that for each $b \in \{2,\ldots,B+1\}$, the probability $\Pr\big[\epsilon_b|\epsilon^c_{b-1}\big]$

tends to 0 as $n \to \infty$. Similar arguments can be used to show that under the same conditions also $\Pr[\epsilon_1] \to 0$ as $n \to \infty$. Using standard arguments one can then conclude that there must exist a deterministic code for which the probability of error $P_e^{(N)}$ tends to 0 as $N \to \infty$ when the mentioned conditions are satisfied.

Fix $b \in \{2, \ldots, B+1\}$ and $\varepsilon > 0$, and define the following events.

- Let $\epsilon_{0,b}$ be the event that there is no pair $(k_{1,b}, k_{2,b}) \in \mathcal{K}_1 \times \mathcal{K}_2$ that satisfies

$$\begin{aligned}\big(U_{0,b}^n(\mathbf{1}_{[4]}), U_{1,b}^n(1, k_{1,b}|\mathbf{1}_{[4]}), U_{2,b}^n(1, k_{2,b}|\mathbf{1}_{[4]})\big) \\ \in \mathcal{T}_{\varepsilon/16}^n(P_{U_0U_1U_2}).\end{aligned}$$

  By the Covering Lemma [28], $\Pr(\epsilon_{0,b})$ tends to 0 as $n \to \infty$ if

$$R_1' + R_2' \geq I(U_1; U_2|U_0) + \delta(\varepsilon), \tag{69}$$

  where throughout this section $\delta(\varepsilon)$ stands for some function that tends to 0 as $\varepsilon \to 0$.

- Let $\epsilon_{1,b}$ be the event that

$$\begin{aligned}\big(U_{0,b}^n(\mathbf{1}_{[4]}), U_{1,b}^n(1, 1|\mathbf{1}_{[4]}), U_{2,b}^n(1, 1, |\mathbf{1}_{[4]}), Y_{1,b}^n, Y_{2,b}^n\big) \\ \notin \mathcal{T}_{\varepsilon/12}^n(P_{U_0U_1U_2Y_1Y_2}).\end{aligned}$$

  Since the channel is memoryless, by the law of large numbers, $\Pr(\epsilon_{1,b}|\epsilon_{0,b}^c)$ tends to 0 as $n \to \infty$.

- Let $\epsilon_{2,1,b}$ be the event that there is no tuple $(\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{l}_{2,b-1}) \in \mathcal{M}_c \times \mathcal{L}_2$ that is not equal to $(\mathbf{1}_{[2]}, 1)$ and that satisfies

$$\big(U_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(1)}, 1, \hat{l}_{2,b-1}), Y_{1,b}^n\big) \in \mathcal{T}_{\varepsilon/8}^n(P_{U_0Y_1}).$$

  By the Packing Lemma [28], $\Pr(\epsilon_{2,1,b}|\epsilon_{1,b}^c)$ tends to 0 as $n \to \infty$, if

$$\tilde{R}_2 + R_{c,1} + R_{c,2} \leq I(U_0; Y_1) + \delta(\varepsilon). \tag{70}$$

- Let $\epsilon_{2,2,b}$ be the event that there is no tuple $(\hat{\mathbf{m}}_{c,b}^{(2)}, \hat{l}_{1,b-1}) \in \mathcal{M}_c \times \mathcal{L}_1$ with $(\hat{\mathbf{m}}_{c,b}^{(2)}, \hat{l}_{1,b-1})$ not equal to $(\mathbf{1}_{[2]}, 1)$ that satisfies

$$\big(U_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(2)}, \hat{l}_{1,b-1}, 1), Y_{2,b}^n\big) \in \mathcal{T}_{\varepsilon/8}^n(P_{U_0Y_2}).$$

  By the Packing Lemma, $\Pr(\epsilon_{2,2,b}|\epsilon_{1,b}^c)$ tends to 0 as $n \to \infty$, if

$$\tilde{R}_1 + R_{c,1} + R_{c,2} \leq I(U_0; Y_2) + \delta(\varepsilon). \tag{71}$$

- Let $\epsilon_{3,1,b}$ be the event that

$$\big(U_{0,b-1}^n(\mathbf{1}_{[4]}),\ U_{1,b-1}^n(1, 1|\mathbf{1}_{[4]}),$$

$$\tilde{Y}^n_{2,b-1}(1,1), Y^n_{1,b-1}\big) \notin \mathcal{T}^n_{\varepsilon/2}(P_{U_0U_1\tilde{Y}_2Y_1}).$$

By the Markov Lemma [28], $\Pr(\epsilon_{3,1,b}|\epsilon^c_{b-1})$ tends to 0 as $n \to \infty$.

- Let $\epsilon_{3,2,b}$ be the event that

$$\big(U^n_{0,b-1}(\mathbf{1}_{[4]}),\, U^n_{2,b-1}(1,1|\mathbf{1}_{[4]}),$$
$$\tilde{Y}^n_{1,b-1}(1,1), Y^n_{2,b-1}\big) \notin \mathcal{T}^n_{\varepsilon/2}(P_{U_0U_2\tilde{Y}_1Y_2}).$$

By the Markov Lemma, $\Pr(\epsilon_{3,2,b}|\epsilon^c_{b-1})$ tends to 0 as $n \to \infty$.

- Let $\epsilon_{4,1,b}$ be the event that there exists a tuple $(\hat{m}_{p,1,b-1}, \hat{k}_{1,b-1}, \hat{j}_{2,b-1}) \in \mathcal{M}_{p,1} \times \mathcal{K}_1 \times \mathcal{J}_2$ not equal to the all-one tuple and that satisfies

$$\big(U^n_{0,b-1}(\mathbf{1}_{[4]}), U^n_{1,b-1}(\hat{m}_{p,1,b-1}, \hat{k}_{1,b-1}|\mathbf{1}_{[4]}),$$
$$\tilde{Y}^n_{2,b-1}(1, \hat{j}_{2,b-1}|\mathbf{1}_{[4]}), Y^n_{1,b-1}\big) \in \mathcal{T}^n_{\varepsilon}(P_{U_0U_1\tilde{Y}_2Y_1}).$$

By the Packing Lemma, $\Pr(\epsilon_{4,1,b}|\epsilon^c_{3,1,b})$ tends to zero as $n \to \infty$, if

$$\hat{R}_2 \le I(\tilde{Y}_2; U_1, Y_1|U_0) - \delta(\varepsilon) \tag{72}$$

$$R_{p,1} + R'_1 \le I(U_1; Y_1, \tilde{Y}_2|U_0) - \delta(\varepsilon) \tag{73}$$

$$R_{p,1} + R'_1 + \hat{R}_2 \le I(U_1; Y_1, \tilde{Y}_2|U_0) + I(\tilde{Y}_2; Y_1|U_0) - \delta(\varepsilon). \tag{74}$$

- Let $\epsilon_{4,2,b}$ be the event that there exists a tuple $(\hat{m}_{p,2,b-1}, \hat{k}_{2,b-1}, \hat{j}_{1,b-1}) \in \mathcal{M}_{p,2} \times \mathcal{K}_2 \times \mathcal{J}_1$ not equal to the all-one tuple and that satisfies

$$\big(U^n_{0,b-1}(\mathbf{1}_{[4]}), U^n_{2,b-1}(\hat{m}_{p,2,b-1}, \hat{k}_{2,b-1}|\mathbf{1}_{[4]}),$$
$$\tilde{Y}^n_{1,b-1}(1, \hat{j}_{1,b-1}|\mathbf{1}_{[4]}), Y^n_{2,b-1}\big) \in \mathcal{T}^n_{\varepsilon}(P_{U_0U_2\tilde{Y}_1Y_2}).$$

By the Packing Lemma, $\Pr(\epsilon_{4,2,b}|\epsilon^c_{3,2,b})$ tends to zero as $n \to \infty$, if

$$\hat{R}_1 \le I(\tilde{Y}_1; U_2, Y_2|U_0) - \delta(\varepsilon) \tag{75}$$

$$R_{p,2} + R'_2 \le I(U_2; Y_2, \tilde{Y}_1|U_0) - \delta(\varepsilon) \tag{76}$$

$$R_{p,2} + R'_2 + \hat{R}_1 \le I(U_2; Y_2, \tilde{Y}_1|U_0) + I(\tilde{Y}_1; Y_2|U_0) - \delta(\varepsilon). \tag{77}$$

- For $i \in \{1,2\}$, let $\epsilon_{5,i,b}$ be the event that there is no pair $(l_{i,b}, j_{i,b}) \in \mathcal{L}_i \times \mathcal{J}_i$ that satisfies

$$\big(\tilde{Y}^n_{i,b}(l_{i,b}, j_{i,b}|\mathbf{1}_{[4]}), U^n_{0,b}(\mathbf{1}_{[4]}), Y^n_{i,b}\big) \in \mathcal{T}^n_{\varepsilon/4}(P_{\tilde{Y}_iU_0Y_i}).$$

By the Covering Lemma, $\Pr(\epsilon_{5,i,b}|\epsilon_{1,b}^c)$ tends to 0 as $n \to \infty$, if

$$\tilde{R}_i + \hat{R}_i \geq I(\tilde{Y}_i; Y_i|U_0) + \delta(\varepsilon). \tag{78}$$

Whenever the event ϵ_{b-1}^c occurs but none of the events $\{\epsilon_{0,b}, \epsilon_{1,b}, \epsilon_{2,1,b}, \epsilon_{2,2,b}, \epsilon_{3,1,b}, \epsilon_{3,2,b}, \epsilon_{4,1,b}, \epsilon_{4,2,b}, \epsilon_{5,1,b}, \epsilon_{5,2,b}\}$ above, then ϵ_b^c. Therefore,

$$\begin{aligned}
&\Pr\Big[\epsilon_b|\epsilon_{b-1}^c\Big] \\
&\leq \Pr\Big[\epsilon_{0,b} \cup \epsilon_{1,b} \cup \bigcup_{i=1}^{2} \Big(\epsilon_{2,i,b} \cup \epsilon_{3,i,b} \cup \epsilon_{4,i,b} \cup \epsilon_{5,i,b}\Big)\Big|\epsilon_{b-1}^c\Big] \\
&\leq \Pr\Big[\epsilon_{0,b}|\epsilon_{b-1}^c\Big] + \Pr\Big[\epsilon_{1,b}|\epsilon_{0,b}^c, \epsilon_{b-1}^c\Big] \\
&\qquad + \sum_{i=1}^{2} \Big(\Pr\Big[\epsilon_{2,i,b}|\epsilon_{1,b}^c, \epsilon_{b-1}^c\Big] + \Pr\Big[\epsilon_{3,i,b}|\epsilon_{b-1}^c\Big] \\
&\qquad + \Pr\Big[\epsilon_{4,i,b}|\epsilon_{3,i,b}^c, \epsilon_{b-1}^c\Big] + \Pr\Big[\epsilon_{5,i,b}|\epsilon_{1,b}^c, \epsilon_{b-1}^c\Big] \Big) \\
&= \Pr[\epsilon_{0,b}] + \Pr\Big[\epsilon_{1,b}|\epsilon_{0,b}^c\Big] \\
&\qquad + \sum_{i=1}^{2} \Big(\Pr\Big[\epsilon_{2,i,b}|\epsilon_{1,b}^c\Big] + \Pr\Big[\epsilon_{3,i,b}|\epsilon_{b-1}^c\Big] \\
&\qquad\qquad + \Pr\Big[\epsilon_{4,i,b}|\epsilon_{3,i,b}^c\Big] + \Pr\Big[\epsilon_{5,i,b}|\epsilon_{1,b}^c\Big] \Big).
\end{aligned}$$

The last equality holds because the channel is memoryless and the codebooks employed in blocks $b-1$ and b are drawn independently. As explained in the previous paragraphs, the remaining terms in the last three lines tend to 0 as $n \to \infty$, if Constraints (69)–(78) are satisfied. Thus, by (68) and (79) we conclude that the probability of error $P_e^{(N)}$ (averaged over all code constructions) vanishes as $n \to \infty$ if Constraints (69)–(78) hold. Letting $\varepsilon \to 0$, we obtain that the probability of error can be made to tend to 0 as $n \to \infty$ whenever

$$R_1' + R_2' > I(U_1; U_2|U_0) \tag{79a}$$

$$\tilde{R}_2 + R_{c,1} + R_{c,2} < I(U_0; Y_1) \tag{79b}$$

$$\tilde{R}_1 + R_{c,1} + R_{c,2} < I(U_0; Y_2) \tag{79c}$$

$$\hat{R}_1 < I(\tilde{Y}_1; U_2, Y_2|U_0) \tag{79d}$$

$$\hat{R}_2 < I(\tilde{Y}_2; U_1, Y_1|U_0) \tag{79e}$$

$$R_{p,1} + R_1' < I(U_1; Y_1, \tilde{Y}_2|U_0) \tag{79f}$$

$$R_{p,2} + R_2' > I(U_2; Y_2, \tilde{Y}_1|U_0) \tag{79g}$$

$$R_{p,1}+R'_1+\hat{R}_2 < I(U_1;Y_1,\tilde{Y}_2|U_0)+I(\tilde{Y}_2;Y_1|U_0) \tag{79h}$$

$$R_{p,2}+R'_2+\hat{R}_1 < I(U_2;Y_2,\tilde{Y}_1|U_0)+I(\tilde{Y}_1;Y_2|U_0) \tag{79i}$$

$$\hat{R}_1+\tilde{R}_1 > I(\tilde{Y}_1;Y_1|U_0) \tag{79j}$$

$$\hat{R}_2+\tilde{R}_2 > I(\tilde{Y}_2;Y_2|U_0). \tag{79k}$$

Moreover, the feedback-rate constraints (1) impose that:

$$\tilde{R}_1 \leq R_{\text{Fb},1} \tag{79l}$$

$$\tilde{R}_2 \leq R_{\text{Fb},2}. \tag{79m}$$

Applying the Fourier-Motzkin elimination algorithm to these constraints, we obtain the desired result in Theorem 1 with the additional constraint that

$$\begin{aligned} &I(U_1;Y_1,\tilde{Y}_2|U_0)+I(U_2;Y_2,\tilde{Y}_1|U_0) \\ &\qquad -\Delta_1-\Delta_2-I(U_1;U_2|U_0) \geq 0 \end{aligned} \tag{80}$$

Notice that we can ignore Constraint (80) because for any tuple $(U_0,U_1,U_2,X,Y_1,Y_2,\tilde{Y}_1,\tilde{Y}_2)$ that violates (80), the region defined by the constraints in Theorem 1 is contained in the time-sharing region.

APPENDIX B
ANALYSIS OF THE SCHEME 1B (THEOREM 2)

An error occurs whenever

$$\hat{M}_{1,b} \neq M_{1,b} \text{ or } \hat{M}_{2,b} \neq M_{2,b}, \text{ for some } b \in \{1,\ldots,B\}.$$

For each $b \in \{1,\ldots,B+1\}$, let ϵ_b denote the event that in our coding scheme at least one of the following holds for $i \in \{1,2\}$:

$$\hat{J}_{i,b} \neq J_{i,b} \tag{81}$$

$$\hat{K}_{i,b} \neq K_{i,b} \tag{82}$$

$$\hat{L}_{i,b-1} \neq L_{i,b-1} \tag{83}$$

$$\hat{M}_{p,i,b} \neq M_{p,i,b} \tag{84}$$

$$\hat{\mathbf{M}}_{c,b}^{(i)} \neq \mathbf{M}_{c,b}^{(i)} \tag{85}$$

Then,

$$P_e^{(N)} \leq \Pr\left[\bigcup_{b=1}^{B+1}\epsilon_b\right] \leq \sum_{b=1}^{B} \Pr\left[\epsilon_b|\epsilon_{b+1}^c\right] + \Pr[\epsilon_{B+1}]. \tag{86}$$

In the following we analyze the probabilities of these events averaged over the random code construction. In particular, we shall identify conditions such that for each $b \in \{1, \dots, B\}$, the probability $\Pr\left[\epsilon_b | \epsilon_{b+1}^c\right]$ tends to 0 as $n \to \infty$. Similar arguments can be used to show that under the same conditions also $\Pr[\epsilon_{B+1}] \to 0$ as $n \to \infty$. Using standard arguments one can then conclude that there must exist a deterministic code for which the probability of error $P_e^{(N)}$ tends to 0 as $N \to \infty$ when the mentioned conditions are satisfied.

Fix $b \in \{1, \dots, B\}$ and $\varepsilon > 0$. By the symmetry of our code construction, the probability $\Pr\left[\epsilon_b | \epsilon_{b+1}^c\right]$ does not depend on the realization of $M_{c,i,b}$, $M_{p,i,b}$, $K_{i,b}$, $J_{i,b}$, $L_{i,b}$, $L_{i,b-1}$, for $i \in \{1, 2\}$. To simplify exposition we therefore assume that $M_{c,i,b} = M_{p,i,b} = K_{i,b} = J_{i,b} = L_{i,b} = L_{i,b-1} = 1$.

Define the following events.

- Let $\epsilon_{0,b}$ be the event that there is no pair $(k_{1,b}, k_{2,b}) \in \mathcal{K}_1 \times \mathcal{K}_2$ that satisfies

$$\begin{aligned}\big(U_{0,b}(\mathbf{1}_{[4]}), U_{1,b}^n(1, k_{1,b}|\mathbf{1}_{[4]}), U_{2,b}^n(1, k_{2,b}|\mathbf{1}_{[4]})\big)& \\ \in \mathcal{T}_{\varepsilon/16}^n(P_{U_0 U_1 U_2}).&\end{aligned}$$

  By the Covering Lemma, $\Pr(\epsilon_{0,b})$ tends to 0 as $n \to \infty$, if

$$R_1' + R_2' \geq I(U_1; U_2|U_0) + \delta(\varepsilon), \tag{87}$$

  where throughout this section $\delta(\varepsilon)$ stands for some function that tends to 0 as $\varepsilon \to 0$.

- Let $\varepsilon_{1,b}$ be the event that

$$\begin{aligned}\big(U_{0,b}^n(\mathbf{1}_{[4]}), U_{1,b}^n(1, 1|\mathbf{1}_{[4]}, U_{2,b}^n(1, 1, |\mathbf{1}_{[4]}), Y_{1,b}^n, Y_{2,b}^n\big)& \\ \notin \mathcal{T}_{\varepsilon/8}^n(P_{U_0 U_1 Y_2 Y_1 Y_2}).&\end{aligned}$$

  Since the channel is memoryless, according to the law of large numbers, $\Pr(\epsilon_{1,b}|\epsilon_{0,b}^c)$ tends to 0 as $n \to \infty$.

- For $i \in \{1, 2\}$, let $\epsilon_{2,i,b}$ be the event that there is no pair $(l_{i,b}, j_{i,b}) \in \mathcal{L}_i \times \mathcal{J}_i$ that satisfies

$$\big(\tilde{Y}_{i,b}^n(l_{i,b}, j_{i,b}), Y_{i,b}^n\big) \in \mathcal{T}_{\varepsilon/4}^n(P_{\tilde{Y}_i Y_i}).$$

  By the Covering Lemma, $\Pr(\epsilon_{2,i,b}|\epsilon_{1,b}^c)$ tends to 0 as $n \to \infty$ if

$$\tilde{R}_i + \hat{R}_i \geq I(\tilde{Y}_i; Y_i) + \delta(\varepsilon). \tag{88}$$

- Let $\epsilon_{3,1,b}$ be the event that

$$\big(U_{0,b}^n(\mathbf{1}_{[4]}), U_{1,b}^n(1, 1|\mathbf{1}_{[4]}),$$

$$\tilde{Y}^n_{2,b}(1,1), Y^n_{1,b}\big) \notin \mathcal{T}^n_{3\varepsilon/4}(P_{U_0U_1\tilde{Y}_2Y_1}).$$

By the Markov Lemma, $\Pr(\epsilon_{3,1,b}|\epsilon^c_{2,2,b}, \epsilon^c_{1,b})$ tends to 0 as $n \to \infty$.

- Let $\epsilon_{3,2,b}$ be the event that

$$\big(U^n_{0,b}(\mathbf{1}_{[4]}), U^n_{2,b}(1,1|\mathbf{1}_{[4]}),$$
$$\tilde{Y}^n_{1,b}(1,1), Y^n_{2,b}\big) \notin \mathcal{T}^n_{3\varepsilon/4}(P_{U_0U_2\tilde{Y}_1Y_2}).$$

By the Markov Lemma, $\Pr(\epsilon_{3,2,b}|\epsilon^c_{2,1,b}, \epsilon^c_{1,b})$ tends to 0 as $n \to \infty$.

- Let $\epsilon_{4,1,b}$ be the event that there exists a tuple $(\hat{j}_{2,b}, \hat{\mathbf{m}}^{(1)}_{c,b}, \hat{l}_{2,b-1}, \hat{m}_{p,1,b}, \hat{k}_{1,b}) \in \mathcal{J}_2 \times \mathcal{M}_c \times \mathcal{L}_2 \times \mathcal{M}_{p,1} \times \mathcal{K}_1$ not equal to the all-one tuple $(1, \mathbf{1}_{[2]}, 1, 1, 1)$ and that satisfies

$$\Big(U^n_{0,b}(\hat{\mathbf{m}}^{(1)}_{c,b}, 1, \hat{l}_{2,b-1}),$$
$$U^n_{1,b}(\hat{m}_{p,1,b}, \hat{k}_{1,b}|\hat{\mathbf{m}}^{(1)}_{c,b}, 1, \hat{l}_{2,b-1}),$$
$$\tilde{Y}^n_{2,b}(1, \hat{j}_{2,b}), Y^n_{1,b}\Big) \in \mathcal{T}^n_{\varepsilon}(P_{U_0U_1\tilde{Y}_2Y_1}).$$

By the Packing Lemma, we conclude that $\Pr(\epsilon_{4,1,b}|\epsilon^c_{3,1,b})$ tends to zero as $n \to \infty$ if

$$\begin{aligned}
\hat{R}_2 &\le I(U_0, U_1, Y_1; \tilde{Y}_2|U_0) - \delta(\varepsilon)\\
R_{p,1} + R'_1 &\le I(U_1; Y_1, \tilde{Y}_2|U_0) - \delta(\varepsilon)\\
R_1 + R_{c,2} + \tilde{R}_2 + R'_1 &\le I(U_0, U_1; Y_1, \tilde{Y}_2) - \delta(\varepsilon)\\
R_1 + R_{c,2} + \tilde{R}_2 + R'_1 + \hat{R}_2 &\le I(U_0, U_1; Y_1, \tilde{Y}_2)\\
&\quad + I(Y_1; \tilde{Y}_2) - \delta(\varepsilon)\\
R_{p,1} + R'_1 + \hat{R}_2 &\le I(U_1; Y_1, \tilde{Y}_2|U_0)\\
&\quad + I(\tilde{Y}_2; Y_1, U_0) - \delta(\varepsilon).
\end{aligned} \tag{89}$$

- Let $\epsilon_{4,2,b}$ be the event that there exists a tuple $(\hat{j}_{1,b}, \hat{\mathbf{m}}^{(2)}_{c,b}, \hat{l}_{1,b-1}, \hat{m}_{p,2,b}, \hat{k}_{2,b}) \in \mathcal{J}_1 \times \mathcal{M}_c \times \mathcal{L}_1 \times \mathcal{M}_{p,2} \times \mathcal{K}_2$ not equal to the all-one tuple and that satisfies

$$\big(U^n_{0,b}(\hat{\mathbf{m}}^{(2)}_{c,b}, \hat{l}_{1,b-1}, 1),$$
$$U^n_{1,b}(\hat{m}_{p,2,b}, \hat{k}_{2,b}|\hat{\mathbf{m}}^{(2)}_{c,b}, \hat{l}_{1,b-1}, 1),$$
$$\tilde{Y}^n_{1,b}(1, \hat{j}_{1,b}), Y^n_{2,b}\big) \in \mathcal{T}^n_{\varepsilon}(P_{U_0U_2\tilde{Y}_1Y_2}).$$

By the Packing Lemma, we conclude that $\Pr(\epsilon_{4,2,b}|\epsilon^c_{3,2,b})$ tends to zero as $n \to \infty$ if

$$\hat{R}_1 \le I(U_0, U_2, Y_2; \tilde{Y}_1|U_0) - \delta(\varepsilon)$$

$$\begin{aligned}
R_{p,2} + R_2' &\le I(U_2; Y_2, \tilde{Y}_1|U_0) - \delta(\varepsilon) \\
R_2 + R_{c,1} + \tilde{R}_1 + R_2' &\le I(U_0, U_2; Y_2, \tilde{Y}_1) - \delta(\varepsilon) \\
R_2 + R_{c,1} + \tilde{R}_1 + R_2' + \hat{R}_1 &\le I(U_0, U_2; Y_2, \tilde{Y}_1) \\
&\quad + I(Y_2; \tilde{Y}_1) - \delta(\varepsilon) \\
R_{p,2} + R_2' + \hat{R}_1 &\le I(U_2; Y_2, \tilde{Y}_1|U_0) \\
&\quad + I(\tilde{Y}_1; Y_2, U_0) - \delta(\varepsilon).
\end{aligned} \tag{90}$$

Whenever the event ϵ_{b+1}^c occurs but none of the events above, then ϵ_b^c. Therefore,

$$\begin{aligned}
&\Pr\big[\epsilon_b|\epsilon_{b+1}^c\big] \\
&\le \Pr\left[\epsilon_{0,b} \cup \epsilon_{1,b} \cup \bigcup_{i=1}^{2} \Big(\epsilon_{2,i,b} \cup \epsilon_{3,i,b} \cup \epsilon_{4,i,b}\Big)\Big|\epsilon_{b+1}^c\right] \\
&\le \Pr\big[\epsilon_{0,b}|\epsilon_{b+1}^c\big] + \Pr\big[\epsilon_{1,b}|\epsilon_{0,b}^c, \epsilon_{b+1}^c\big] \\
&\qquad + \Pr\big[\epsilon_{3,1,b}|\epsilon_{1,b}^c, \epsilon_{2,2,b}^c, \epsilon_{b+1}^c\big] + \Pr\big[\epsilon_{3,2,b}|\epsilon_{1,b}^c, \epsilon_{2,1,b}^c, \epsilon_{b+1}^c\big] \\
&\qquad + \sum_{i=1}^{2} \Big(\Pr\big[\epsilon_{2,i,b}|\epsilon_{1,b}^c, \epsilon_{b+1}^c\big] + \Pr\big[\epsilon_{4,i,b}|\epsilon_{3,i,b}^c, \epsilon_{b+1}^c\big] \Big) \\
&= \Pr[\epsilon_{0,b}] + \Pr\big[\epsilon_{1,b}|\epsilon_{0,b}^c\big] \\
&\qquad + \Pr\big[\epsilon_{3,1,b}|\epsilon_{1,b}^c, \epsilon_{2,2,b}^c\big] + \Pr\big[\epsilon_{3,2,b}|\epsilon_{1,b}^c, \epsilon_{2,1,b}^c\big] \\
&\qquad + \sum_{i=1}^{2} \Big(\Pr\big[\epsilon_{2,i,b}|\epsilon_{1,b}^c\big] + \Pr\big[\epsilon_{4,i,b}|\epsilon_{3,i,b}^c\big] \Big),
\end{aligned} \tag{91}$$

where the last equality follows because the channel is memoryless and the codebooks for blocks b and $b+1$ have been generated independently. As explained in the previous paragraphs, each of the terms in the last three lines tends to 0 as $n \to \infty$, if Constraints (87)–(90) are satisfied. Thus, by (86) and (91) we conclude that the probability of error $P_e^{(N)}$ (averaged over all code constructions) vanishes as $n \to \infty$ if constraints (87)–(90) hold. Letting $\varepsilon \to 0$, we obtain that the probability of error can be made to tend to 0 as $n \to \infty$ whenever

$$R_1' + R_2' > I(U_1; U_2|U_0) \tag{92a}$$

$$\hat{R}_1 + \tilde{R}_1 > I(\tilde{Y}_1; Y_1) \tag{92b}$$

$$\hat{R}_2 + \tilde{R}_2 > I(\tilde{Y}_2; Y_2) \tag{92c}$$

$$\hat{R}_1 < I(U_0, U_2, Y_2; \tilde{Y}_1|U_0) \tag{92d}$$

$$\hat{R}_2 < I(U_0, U_1, Y_1; \tilde{Y}_2|U_0) \tag{92e}$$

$$R_{p,1} + R'_1 < I(U_1; Y_1, \tilde{Y}_2|U_0) \tag{92f}$$

$$R_{p,2} + R'_2 < I(U_2; Y_2, \tilde{Y}_1|U_0) \tag{92g}$$

$$R_1 + R_{c,2} + \tilde{R}_2 + R'_1 < I(U_0, U_1; Y_1, \tilde{Y}_2) \tag{92h}$$

$$R_2 + R_{c,1} + \tilde{R}_1 + R'_2 < I(U_0, U_2; Y_2, \tilde{Y}_1) \tag{92i}$$

$$\begin{aligned} R_1+R_{c,2}+\tilde{R}_2+R'_1+\hat{R}_2 < I(U_0, U_1; Y_1, \tilde{Y}_2) \\ + I(Y_1; \tilde{Y}_2) \end{aligned} \tag{92j}$$

$$\begin{aligned} R_2+R_{c,1}+\tilde{R}_1+R'_2+\hat{R}_1 < I(U_0, U_2; Y_2, \tilde{Y}_1) \\ + I(Y_2; \tilde{Y}_1) \end{aligned} \tag{92k}$$

$$\begin{aligned} R_{p,1} + R'_1 + \hat{R}_2 < I(U_1; Y_1, \tilde{Y}_2|U_0) \\ + I(\tilde{Y}_2; Y_1, U_0) \end{aligned} \tag{92l}$$

$$\begin{aligned} R_{p,2} + R'_2 + \hat{R}_1 < I(U_2; Y_2, \tilde{Y}_1|U_0) \\ + I(\tilde{Y}_1; Y_2, U_0). \end{aligned} \tag{92m}$$

Moreover, the feedback-rate constraints (1) impose that:

$$\tilde{R}_1 \leq R_{\text{Fb},1} \tag{92n}$$

$$\tilde{R}_2 \leq R_{\text{Fb},2}. \tag{92o}$$

Applying the Fourier-Motzkin elimination algorithm to these constraints, we obtain the desired result in Theorem 2 with the additional constraint that

$$\begin{aligned} I(U_1; Y_1, \tilde{Y}_2|U_0) + I(U_2; Y_2, \tilde{Y}_1|U_0) \\ -\Delta_1 - \Delta_2 - I(U_1; U_2|U_0) \geq 0 \end{aligned} \tag{93a}$$

$$I(U_1; Y_1, \tilde{Y}_2|U_0) - \Delta_2 \geq 0 \tag{93b}$$

$$I(U_2; Y_2, \tilde{Y}_1|U_0) - \Delta_1 \geq 0. \tag{93c}$$

We can ignore Constraint (93a) because for any tuple $(U_0, U_1, U_2, X, Y_1, Y_2, \tilde{Y}_1, \tilde{Y}_2)$ that violates (93a), the region defined by the constraints in Theorem 2 is contained in the time-sharing region. Constraint (93b) can also be ignored because for any tuple $(U_0, U_1, U_2, X, Y_1, Y_2, \tilde{Y}_1, \tilde{Y}_2)$ that violates (93b), the region defined by the constraints in Theorem 2 is contained in the region in Theorem 2 for the choice $\tilde{Y}_2 =$ const., for which (93b) is always satisfied. Constraint (93c) can be ignored by analogous arguments.

APPENDIX C
ANALYSIS OF SCHEME 2 (THEOREM 4)

An error occurs whenever

$$\hat{M}_{1,b} \neq M_{1,b} \text{ or } \hat{M}_{2,b} \neq M_{2,b}, \text{ for some } b \in \{1, \dots, B\}.$$

For each $b \in \{1, \dots, B+1\}$, let ϵ_b denote the event that in our coding scheme at least one of the following holds for $i \in \{1, 2\}$:

$$\hat{J}_{i,b} \neq J_{i,b} \tag{94}$$

$$\hat{K}_{i,b} \neq K_{i,b} \tag{95}$$

$$\hat{L}_{i,b} \neq L_{i,b} \tag{96}$$

$$\hat{M}_{p,i,b} \neq M_{p,i,b} \tag{97}$$

$$\hat{\mathbf{M}}_{c,b}^{(i)} \neq \mathbf{M}_{c,b}^{(i)} \tag{98}$$

or when

$$\hat{N}_{b-1} \neq N_{b-1}. \tag{99}$$

Then,

$$P_e^{(n)} \leq \Pr\left[\bigcup_{b=1}^{B+1} \epsilon_b\right] \leq \sum_{b=1}^{B} \Pr\left[\epsilon_b | \epsilon_{b+1}^c\right] + \Pr[\epsilon_{B+1}]\,. \tag{100}$$

In the following we analyze the probabilities of these events averaged over the random code construction. In particular, we shall identify conditions such that for each $b \in \{1, \dots, B\}$, the probability $\Pr\left[\epsilon_b | \epsilon_{b+1}^c\right]$ tends to 0 as $n \to \infty$. Similar arguments can be used to show that under the same conditions also $\Pr[\epsilon_{B+1}] \to 0$ as $n \to \infty$. Using standard arguments one can then conclude that there must exist a deterministic code for which the probability of error $P_e^{(N)}$ tends to 0 as $N \to \infty$ when the mentioned conditions are satisfied.

Fix $b \in \{1, \dots, B\}$ and $\varepsilon > 0$. By the symmetry of our code construction, the probability $\Pr\left[\epsilon_b | \epsilon_{b+1}^c\right]$ does not depend on the realizations of N_{b-1}, N_b, or $M_{c,i,b}$, $M_{p,i,b}$, $K_{i,b}$, $J_{i,b}$, $L_{i,b}$, for $i \in \{1, 2\}$. To simplify exposition we therefore assume that for $i \in \{1, 2\}$, $M_{c,i,b} = M_{p,i,b} = K_{i,b} = J_{i,b} = L_{i,b} = 1$, and $N_b = N_{b-1} = 1$.

Define the following events.

- Let $\epsilon_{0,b}$ be the event that there is no pair $(k_{1,b}, k_{2,b}) \in \mathcal{K}_1 \times \mathcal{K}_2$ that satisfies

$$\big(U_{0,b}(\mathbf{1}_{[3]}), U^n_{1,b}(1, k_{1,b}|\mathbf{1}_{[2]}), U^n_{2,b}(1, k_{2,b}|\mathbf{1}_{[3]})\big) \\ \in \mathcal{T}^n_{\varepsilon/64}(P_{U_0U_1U_2}).$$

  By the Covering Lemma, $\Pr(\epsilon_{0,b})$ tends to 0 as $n \to \infty$ if

$$R'_1 + R'_2 \geq I(U_1; U_2|U_0) + \delta(\varepsilon), \tag{101}$$

  where throughout this section $\delta(\varepsilon)$ stands for some function that tends to 0 as $\varepsilon \to 0$.

- Let $\epsilon_{1,b}$ be the event that

$$\big(U^n_{0,b}(\mathbf{1}_{[3]}), U^n_{1,b}(1, 1|\mathbf{1}_{[3]}), U^n_{2,b}(1, 1, |\mathbf{1}_{[3]}), Y^n_{1,b}, Y^n_{2,b}\big) \\ \notin \mathcal{T}^n_{\varepsilon/32}(P_{U_0U_1U_2Y_1Y_2}).$$

  Since the channel is memoryless, according to the law of large numbers, $\Pr(\epsilon_{1,b}|\epsilon^c_{0,b})$ tends to 0 as $n \to \infty$.

- For $i \in \{1, 2\}$, let $\epsilon_{2,i,b}$ be the event that there is no pair $(l_{i,b}, j_{i,b}) \in \mathcal{L}_i \times \mathcal{J}_i$ that satisfies

$$\big(\tilde{Y}^n_{i,b}(l_{i,b}, j_{i,b}), Y^n_{i,b}\big) \in \mathcal{T}^n_{\varepsilon/16}(P_{\tilde{Y}_iY_i}).$$

  By the Covering Lemma, $\Pr(\epsilon_{2,i,b}|\epsilon^c_{1,b})$ tends to 0 as $n \to \infty$ if

$$\tilde{R}_i + \hat{R}_i \geq I(\tilde{Y}_i; Y_i) + \delta(\epsilon). \tag{102}$$

- Let $\epsilon_{3,b}$ be the event that

$$\big(U^n_{0,b}(\mathbf{1}_{[3]}), U^n_{1,b}(1, 1|\mathbf{1}_{[3]}), U^n_{2,b}(1, 1|\mathbf{1}_{[3]}), \\ \tilde{Y}^n_{1,b}(1, 1), \tilde{Y}^n_{2,b}(1, 1), Y^n_{1,b}, Y^n_{2,b}\big) \\ \notin \mathcal{T}^n_{\varepsilon/6}(P_{U_0U_1U_2\tilde{Y}_1\tilde{Y}_2Y_1Y_2}).$$

  By the Markov Lemma, $\Pr(\epsilon_{3,b}|\epsilon^c_{2,1,b}, \epsilon^c_{2,2,b}, \epsilon^c_{1,b})$ tends to 0 as $n \to \infty$.

- Let $\epsilon_{4,b}$ be the event that there is a pair of indices $\hat{j}_{1,b} \in \mathcal{J}_1$ and $\hat{j}_{2,b} \in \mathcal{J}_2$ not equal to the all-one pair $(1, 1)$ and that satisfies

$$\big(U^n_{0,b}(\mathbf{1}_{[3]}), U^n_{1,b}(1, 1|\mathbf{1}_{[3]}), U^n_{2,b}(1, 1|\mathbf{1}_{[3]}), \\ \tilde{Y}^n_{1,b}(1, \hat{j}_{1,b}), \tilde{Y}^n_{2,b}(1, \hat{j}_{2,b})\big) \in \mathcal{T}^n_{\varepsilon/4}(P_{U_0U_1U_2\tilde{Y}_1\tilde{Y}_2}).$$

  By the Packing Lemma, $\Pr(\epsilon_{4,b}|\epsilon^c_{3,b})$ tends to 0 as $n \to \infty$, if

$$\hat{R}_1 \leq I(U_0, U_1, U_2, \tilde{Y}_2; \tilde{Y}_1) - \delta(\varepsilon) \tag{103}$$

$$\hat{R}_2 \le I(U_0, U_1, U_2, \tilde{Y}_1; \tilde{Y}_2) - \delta(\varepsilon) \tag{104}$$

$$\hat{R}_1 + \hat{R}_2 \le I(U_0, U_1, U_2; \tilde{Y}_1, \tilde{Y}_2) + I(\tilde{Y}_1; \tilde{Y}_2) - \delta(\varepsilon). \tag{105}$$

- Let $\epsilon_{5,b}$ be the event that there is no index $n_b \in \mathcal{N}$ that satisfies

$$\big(U_{0,b}^n(\mathbf{1}_{[3]}), U_{1,b}^n(1,1|\mathbf{1}_{[3]}), U_{2,b}^n(1,1|\mathbf{1}_{[3]}),$$
$$\tilde{Y}_{1,b}^n(1,1), \tilde{Y}_{2,b}^n(1,1), V_b^n(n_b|1)\big)$$
$$\in \mathcal{T}_{\varepsilon/2}^n(P_{U_0U_1U_2\tilde{Y}_1\tilde{Y}_2V}).$$

  By the Covering Lemma, $\Pr(\epsilon_{5,b}|\epsilon_{3,b}^c)$ tends to 0 as $n \to \infty$, if

$$\tilde{R}_v \ge I(U_0, U_1, U_2, \tilde{Y}_1, \tilde{Y}_2; V) + \delta(\varepsilon). \tag{106}$$

- Let $\epsilon_{6,1,b}$ be the event that

$$\Big(U_{0,b}^n(\mathbf{1}_{[3]}, 1), U_{1,b}^n(1,1|\mathbf{1}_{[3]}, 1),$$
$$V_b^n(1|1), Y_{1,b}^n, \tilde{Y}_{1,b}^n(1,1)\Big)$$
$$\in \mathcal{T}_{\varepsilon}^n(P_{U_0U_1VY_1\tilde{Y}_1}).$$

  By the Markov Lemma $\Pr(\epsilon_{6,1,b}|\epsilon_{3,b}^c, \epsilon_{5,b}^c)$ tends to zero as $n \to \infty$.
- Let $\epsilon_{6,2,b}$ be the event that

$$\Big(U_{0,b}^n(\mathbf{1}_{[3]}, 1), U_{2,b}^n(1,1|\mathbf{1}_{[3]}, 1),$$
$$V_b^n(1|1), Y_{2,b}^n, \tilde{Y}_{2,b}^n(1,1)\Big)$$
$$\in \mathcal{T}_{\varepsilon}^n(P_{U_0U_2VY_2\tilde{Y}_2}).$$

  By the Markov Lemma $\Pr(\epsilon_{6,2,b}|\epsilon_{3,b}^c, \epsilon_{5,b}^c)$ tends to zero as $n \to \infty$.
- Let $\epsilon_{7,1,b}$ be the event that there is a tuple $(\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{n}_{b-1}, \hat{m}_{p,1,b}, \hat{k}_{1,b}) \in \mathcal{M}_c \times \mathcal{N} \times \mathcal{M}_{p,1} \times \mathcal{K}_1$ that is not equal to the all-one tuple $(\mathbf{1}_{[3]}, 1, 1, 1)$ and that satisfies

$$\Big(U_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{n}_{b-1}), U_{1,b}^n(\hat{m}_{p,1,b}, \hat{k}_{1,b}|\hat{\mathbf{m}}_{c,b}^{(1)}, \hat{n}_{b-1}),$$
$$V_b^n(1|\hat{n}_{b-1}), Y_{1,b}^n, \tilde{Y}_{1,b}^n(1,1)\Big)$$
$$\in \mathcal{T}_{\varepsilon}^n(P_{U_0U_1VY_1\tilde{Y}_1}).$$

  By the Packing Lemma, we conclude that $\Pr(\epsilon_{7,1,b}|\epsilon_{6,1,b}^c)$ tends to zero as $n \to \infty$ if

$$R_1 + R_{c,2} + R_1' \le I(U_0, U_1; Y_1, \tilde{Y}_1, V) - \delta(\varepsilon) \tag{107}$$

$$R_1+R_{c,2}+\tilde{R}_v+R_1' \leq I(U_0,U_1;Y_1,\tilde{Y}_1,V) \\ +I(V;\tilde{Y}_1,Y_1)-\delta(\varepsilon) \tag{108}$$

$$R_{p,1}+R_1' \leq I(U_1;Y_1,\tilde{Y}_1,V|U_0)-\delta(\varepsilon). \tag{109}$$

- Let $\epsilon_{7,2,b}$ be the event that there is a tuple $(\hat{\mathbf{m}}_{c,b}^{(2)},\hat{n}_{b-1},\hat{m}_{p,2,b},\hat{k}_{2,b}) \in \mathcal{M}_c \times \mathcal{N} \times \mathcal{M}_{p,2} \times \mathcal{K}_2$ that is not equal to the all-one tuple $(\mathbf{1}_{[3]},1,1,1)$ and that satisfies

$$\Big(U_{0,b}^n(\hat{\mathbf{m}}_{c,b}^{(2)},\hat{n}_{b-1}),U_{2,b}^n(\hat{m}_{p,2,b},\hat{k}_{2,b}|\hat{\mathbf{m}}_{c,b}^{(2)},\hat{n}_{b-1}), \\ V_b^n(1|\hat{n}_{b-1}),Y_{2,b}^n,\tilde{Y}_{2,b}^n(1,1)\Big) \\ \in \mathcal{T}_\varepsilon^n(P_{U_0U_2VY_2\tilde{Y}_2}).$$

By the Markov Lemma and the Packing Lemma, we conclude that $\Pr(\epsilon_{7,2,b}|\epsilon_{6,2,b}^c)$ tends to zero as $n\to\infty$, if

$$R_2+R_{c,1}+R_2' \leq I(U_0,U_2;Y_2,\tilde{Y}_2,V)-\delta(\varepsilon) \tag{110}$$

$$R_2+R_{c,1}+\tilde{R}_v+R_2' \leq I(U_0,U_2;Y_2,\tilde{Y}_2,V) \\ +I(V;\tilde{Y}_2,Y_2)-\delta(\varepsilon) \tag{111}$$

$$R_{p,2}+R_2' \leq I(U_2;Y_2,\tilde{Y}_2,V|U_0)-\delta(\varepsilon). \tag{112}$$

Whenever the event ϵ_{b+1}^c occurs but none of the events above, then ϵ_b^c. Therefore,

$$\begin{aligned}
&\Pr\big[\epsilon_b|\epsilon_{b+1}^c\big] \\
&\leq \Pr\Big[\epsilon_{0,b}\cup\epsilon_{1,b}\cup\epsilon_{2,1,b}\cup\epsilon_{2,2,b}\cup\epsilon_{3,b} \\
&\qquad \cup\epsilon_{4,b}\cup\epsilon_{5,b}\cup\epsilon_{6,1,b}\cup\epsilon_{6,2,b}\Big|\epsilon_{b+1}^c\Big] \\
&\leq \Pr\big[\epsilon_{0,b}|\epsilon_{b+1}^c\big]+\Pr\big[\epsilon_{1,b}|\epsilon_{0,b}^c,\epsilon_{b+1}^c\big] \\
&\quad +\sum_{i=1}^{2}\Pr\big[\epsilon_{2,i,b}|\epsilon_{1,b}^c,\epsilon_{b+1}^c\big] \\
&\quad +\Pr\big[\epsilon_{3,b}|\epsilon_{1,b}^c,\epsilon_{2,1,b}^c,\epsilon_{2,2,b}^c,\epsilon_{b+1}^c\big]+\Pr\big[\epsilon_{4,b}|\epsilon_{3,b}^c,\epsilon_{b+1}^c\big] \\
&\quad +\Pr\big[\epsilon_{5,b}|\epsilon_{3,b}^c,\epsilon_{b+1}^c\big]+\sum_{i=1}^{2}\Pr\big[\epsilon_{6,i,b}|\epsilon_{3,b}^c,\epsilon_{b+1}^c\big] \\
&= \Pr[\epsilon_{0,b}]+\Pr\big[\epsilon_{1,b}|\epsilon_{0,b}^c\big]+\sum_{i=1}^{2}\Pr\big[\epsilon_{2,i,b}|\epsilon_{1,b}^c\big] \\
&\quad +\Pr\big[\epsilon_{3,b}|\epsilon_{1,b}^c,\epsilon_{2,1,b}^c,\epsilon_{2,2,b}^c\big]+\Pr\big[\epsilon_{4,b}|\epsilon_{3,b}^c\big]+\Pr\big[\epsilon_{5,b}|\epsilon_{3,b}^c\big]
\end{aligned}$$

$$+\sum_{i=1}^{2} \Pr\left[\epsilon_{6,i,b}|\epsilon_{3,b}^{c}\right], \tag{113}$$

where the last equality follows because the channel is memoryless and the codebooks in blocks b and $b+1$ have been chosen independently. As explained in the previous paragraphs, each of the terms in the last five lines tends to 0 as $n \to \infty$, if Constraints (101)–(112) are satisfied. Thus, by (100) and (113) we conclude that the probability of error $P_e^{(N)}$ (averaged over all code constructions) vanishes as $n \to \infty$ if Constraints (101)–(112) hold. Letting $\varepsilon \to 0$, we obtain that the probability of error can be made to tend to 0 as $n \to \infty$ whenever

$$R_1' + R_2' > I(U_1; U_2|U_0) \tag{114a}$$

$$\hat{R}_1 + \tilde{R}_1 > I(\tilde{Y}_1; Y_1) \tag{114b}$$

$$\hat{R}_2 + \tilde{R}_2 > I(\tilde{Y}_2; Y_2) \tag{114c}$$

$$\hat{R}_1 < I(U_0, U_1, U_2, \tilde{Y}_2; \tilde{Y}_1) \tag{114d}$$

$$\hat{R}_2 < I(U_0, U_1, U_2, \tilde{Y}_1; \tilde{Y}_2) \tag{114e}$$

$$\hat{R}_1 + \hat{R}_2 < I(U_0, U_1, U_2; \tilde{Y}_1, \tilde{Y}_2) + I(\tilde{Y}_1; \tilde{Y}_2) \tag{114f}$$

$$\tilde{R}_v > I(U_0, U_1, U_2, \tilde{Y}_1, \tilde{Y}_2; V) \tag{114g}$$

$$R_1 + R_{c,2} + \tilde{R}_v + R_1' < I(U_0, U_1; Y_1, \tilde{Y}_1, V) + I(V; \tilde{Y}_1, Y_1) \tag{114h}$$

$$R_1 + R_{c,2} + R_1' < I(U_0, U_1; Y_1, \tilde{Y}_1, V) \tag{114i}$$

$$R_{c,1} + R_2 + R_2' < I(U_0, U_2; Y_2, \tilde{Y}_2, V) \tag{114j}$$

$$R_{c,1} + R_2 + \tilde{R}_v + R_2' < I(U_0, U_2; Y_2, \tilde{Y}_2, V) + I(V; \tilde{Y}_2, Y_2) \tag{114k}$$

$$R_{p,1} + R_1' < I(U_1; Y_1, \tilde{Y}_1, V|U_0) \tag{114l}$$

$$R_{p,2} + R_2' < I(U_2; Y_2, \tilde{Y}_2, V|U_0). \tag{114m}$$

Moreover, the feedback-rate constraints (1) impose that:

$$\tilde{R}_1 \leq R_{\text{Fb},1} \tag{114n}$$

$$\tilde{R}_2 \leq R_{\text{Fb},2}. \tag{114o}$$

Eliminating the auxiliaries $\tilde{R}_1, \tilde{R}_2, \hat{R}_1, \hat{R}_2, \tilde{R}_v$ from the above (using the Fourier-Motzkin algorithm), we obtain:

$$R_1' + R_2' > I(U_1; U_2|U_0) \tag{115a}$$

$$R_1 + R_{c,2} + R_1' < I(U_0, U_1; Y_1, \tilde{Y}_1, V) - I(V; U_0, U_1, U_2, \tilde{Y}_2 | \tilde{Y}_1, Y_1) \tag{115b}$$

$$R_{c,1} + R_2 + R_2' < I(U_0, U_2; Y_2, \tilde{Y}_2, V) - I(V; U_0, U_1, U_2, \tilde{Y}_1 | \tilde{Y}_2, Y_2) \tag{115c}$$

$$R_{p,1} + R_1' < I(U_1; Y_1, \tilde{Y}_1, V|U_0) \tag{115d}$$

$$R_{p,2} + R_2' < (U_2; Y_2, \tilde{Y}_2, V|U_0) \tag{115e}$$

where the feedback-rate constraints have to satisfy

$$I(Y_1; \tilde{Y}_1 | U_0, U_1, U_2, \tilde{Y}_2) \leq R_{\text{Fb},1} \tag{116a}$$

$$I(Y_2; \tilde{Y}_2 | U_0, U_1, U_2, \tilde{Y}_1) \leq R_{\text{Fb},2} \tag{116b}$$

$$I(Y_1, Y_2; \tilde{Y}_1, \tilde{Y}_2 | U_0, U_1, U_2) \leq R_{\text{Fb},1} + R_{\text{Fb},2}. \tag{116c}$$

Applying again the Fourier-Motzkin elimination algorithm to Constraints (115) and keeping Constraints (116), we obtain the desired result in Theorem 4 with the additional constraint that

$$I(U_1; U_2|U_0) \leq I(U_1; Y_1, \tilde{Y}_1, V|U_0) + (U_2; Y_2, \tilde{Y}_2, V|U_0). \tag{117}$$

Finally, this last constraint can be ignored because for any tuple $(U_0, U_1, U_2, X, Y_1, Y_2, \tilde{Y}_1, \tilde{Y}_2)$ that violates (117), the region defined by the constraints in Theorem 4 is contained in the time-sharing region.

Appendix D
Proof of Theorem 5

Let $R_{\text{Fb},1} > 0$. Fix a tuple $(U_0^{(\text{M})}, U_1^{(\text{M})}, U_2^{(\text{M})}, X^{(\text{M})})$ and rate pairs $(R_1^{(\text{M})}, R_2^{(\text{M})})$ and $(R_1^{(\text{Enh})}, R_2^{(\text{Enh})}) \in \mathcal{C}_{\text{Enh}}^{(1)}$ as stated in the theorem. Then, by the assumptions in the theorem,

$$R_1^{(\text{M})} \leq I(U_0^{(\text{M})}, U_1^{(\text{M})}; Y_1^{(\text{M})}) \tag{118a}$$

$$R_2^{(\text{M})} < I(U_0^{(\text{M})}, U_2^{(\text{M})}; Y_2^{(\text{M})}) \tag{118b}$$

$$R_1^{(\text{M})} + R_2^{(\text{M})} \leq I(U_0^{(\text{M})}, U_1^{(\text{M})}; Y_1^{(\text{M})}) + I(U_2^{(\text{M})}; Y_2^{(\text{M})} | U_0^{(\text{M})})$$

$$-I(U_1^{(\mathrm{M})};U_2^{(\mathrm{M})}|U_0^{(\mathrm{M})}), \tag{118c}$$

where $Y_1^{(\mathrm{M})}$ and $Y_2^{(\mathrm{M})}$ denote the outputs of the considered DMBC corresponding to input $X^{(\mathrm{M})}$. (Notice the strict inequality of the second constraint.)

By the definition of $\mathcal{C}_{\text{Enh}}^{(1)}$ we can identify random variables $U_0^{(\text{Enh})}$ and $X^{(\text{Enh})}$ such that

$$R_1^{(\text{Enh})} \leq I(U_0^{(\text{Enh})};Y_1^{(\text{Enh})}) \tag{119a}$$

$$R_2^{(\text{Enh})} \leq I(X^{(\text{Enh})};Y_1^{(\text{Enh})},Y_2^{(\text{Enh})}|U_0^{(\text{Enh})}), \tag{119b}$$

where $Y_1^{(\text{Enh})}$ and $Y_2^{(\text{Enh})}$ denote the outputs of the considered DMBC corresponding to input $X^{(\text{Enh})}$.

Define further $U_1^{(\text{Enh})}$ =const., $U_2^{(\text{Enh})} = X^{(\text{Enh})}$, $\tilde{Y}_1^{(\text{Enh})} = Y_1^{(\text{Enh})}$, $\tilde{Y}_1^{M}$ =const, and a binary random variable Q independent of all previously defined random variables and of pmf

$$P_Q(q) = \begin{cases} \gamma, & q = \text{Enh} \\ 1-\gamma, & q = \text{M}. \end{cases} \tag{120}$$

We show that when γ is sufficiently small, then the random variables

$$U_0 := U_0^{(Q)},\ U_1 := U_1^{(Q)},\ U_2 := U_2^{(Q)}$$
$$X := X^{(Q)},\ \text{and } \tilde{Y}_1 := \tilde{Y}_1^{(Q)} \tag{121}$$

satisfy the feedback rate constraints (33) and the rate pair (R_1', R_2'),

$$R_1' := (1-\gamma)R_1^{(\mathrm{M})} + \gamma R_1^{(\text{Enh})} \tag{122a}$$

$$R_2' := (1-\gamma)R_2^{(\mathrm{M})} + \gamma R_2^{(\text{Enh})}, \tag{122b}$$

satisfies the constraints in (32) for the choice in (121). The two imply that the rate pair (R_1', R_2') lies in $\mathcal{R}_{\text{relay,hb}}^{(1)}$ and concludes our proof.

Notice that the pmf of the tuple $U_0, U_1, U_2, X, Y_1, Y_2, \tilde{Y}_1$ has the desired form

$$P_Q P_{U_0U_1U_2|Q} P_{X|U_0U_1U_2Q} P_{Y_1Y_2|X} P_{\tilde{Y}_1|Y_1Q}. \tag{123}$$

where $P_{Y_1Y_2|X}$ denotes the channel law.

For the described choice of random variables (121), the feedback-rate constraint (33) specializes to

$$\gamma H(Y_1^{(\text{Enh})}|Y_2^{(\text{Enh})}, X^{(\text{Enh})}) \leq R_{\text{Fb},1}, \tag{124}$$

which is satisfied for all sufficiently small $\gamma \in (0,1)$. Moreover, for this choice the constraints in (32) specialize to

$$R_1 \leq (1-\gamma)I(U_0^{(\mathrm{M})}, U_1^{(\mathrm{M})}; Y_1^{(\mathrm{M})})$$

$$
\begin{aligned}
&+\gamma I(U_0^{(\text{Enh})};Y_1^{(\text{Enh})}) && \text{(125a)}\\
R_2 \leq\ &(1-\gamma)I(U_0^{(\text{M})},U_2^{(\text{M})};Y_2^{(\text{M})})\\
&+\gamma\big(I(X^{(\text{Enh})};Y_1^{(\text{Enh})},Y_2^{(\text{Enh})})\\
&\qquad -H(Y_1^{(\text{Enh})}|Y_2^{(\text{Enh})})\big) && \text{(125b)}\\
R_1+R_2 \leq\ &(1-\gamma)\big(I(U_0^{(\text{M})},U_1^{(\text{M})};Y_1^{(\text{M})})\\
&\qquad +I(U_2^{(\text{M})};Y_2^{(\text{M})}|U_0^{(\text{M})})\\
&\qquad\quad -I(U_1^{(\text{M})};U_2^{(\text{M})}|U_0^{(\text{M})})\big)\\
&+\gamma\big(I(U_0^{(\text{Enh})};Y_1^{(\text{Enh})})\\
&\quad +I(X^{(\text{Enh})};Y_1^{(\text{Enh})},Y_2^{(\text{Enh})}|U_0^{(\text{Enh})})\big) && \text{(125c)}\\
R_1+R_2 \leq\ &(1-\gamma)\big(I(U_1^{(\text{M})};Y_1^{(\text{M})}|U_0^{(\text{M})})\\
&\qquad +I(U_0^{(\text{M})},U_2^{(\text{M})};Y_2^{(\text{M})})\\
&\qquad\quad -I(U_1^{(\text{M})};U_2^{(\text{M})}|U_0^{(\text{M})})\big)\\
&+\gamma\big(I(X^{(\text{Enh})};Y_1^{(\text{Enh})},Y_2^{(\text{Enh})})\\
&\qquad -H(Y_1^{(\text{Enh})}|Y_2^{(\text{Enh})})\big). && \text{(125d)}
\end{aligned}
$$

We argue in the following that the rate pair $(R_1 = R_1', R_2 = R_2')$ defined in (122) satisfies these constraints for all sufficiently small $\gamma > 0$. Comparing (118a), (119a), and (122a), we see that the first constraint (125a) is satisfied for any choice of $\gamma \in [0,1]$. Similarly, comparing (118c), (119a), (119b), and (122a) and (122b), we note that also the third constraint (125c) is satisfied for any $\gamma \in [0,1]$. The second constraint (125b) is satisfied when γ is sufficiently small. This can be seen by comparing (118b), (119b), and (122b), and because Constraint (118b) holds with strict inequality. The last constraint (125d) is not active in view of Constraint (125c) whenever

$$\gamma H(Y_1^{(\text{Enh})}|Y_2^{(\text{Enh})}) \leq (1-\gamma)\Gamma^{(\text{M})}, \tag{126}$$

where $\Gamma^{(\text{M})}$ is defined in (39). Thus, also this last constraint is satisfied when γ is sufficiently small. This concludes our proof.

Appendix E
Proof of Remark 1

Fix a distribution $P_{U_0U_1U_2X}$. We prove that there exists a distribution $P_{U_0'U_1'U_2'X'}$ that satisfies one of the three conditions in Remark 1 and so that the rate region defined by Marton's constraints (6) and distribution $P_{U_0'U_1'U_2'X'}$ contains the rate region defined by Marton's constraints (6) and distribution $P_{U_0U_1U_2X}$.

We assume without loss of generality that $I(U_0;Y_1) \leq I(U_0;Y_2)$, and we separately treat the two cases

- $I(U_0,U_1;Y_1) \leq I(U_0,U_1;Y_2)$
- $I(U_0,U_1;Y_1) > I(U_0,U_1;Y_2)$.

For the first case, $I(U_0,U_1;Y_1) \leq I(U_0,U_1;Y_2)$, let $U_0' = (U_0,U_1)$, $U_1' =$ const., $U_2' = U_2$ and $X' = X$. Evaluating Marton's constraints (6) for the auxiliaries (U_0',U_1',U_2',X') results in

$$R_1 \leq I(U_0,U_1;Y_1) \tag{127a}$$

$$R_2 \leq I(U_0,U_1,U_2;Y_2) \tag{127b}$$

$$R_1+R_2 \leq I(U_0,U_1;Y_1)+I(U_2;Y_2|U_0,U_1) \tag{127c}$$

$$R_1+R_2 \leq I(U_0,U_1,U_2;Y_2) \tag{127d}$$

Note that the fourth constraint is redundant in view of the second.

We show that the first three constraints are no tighter than Marton's constraints in (6), which proves the desired result for the first case. In fact, the R_1-constraint in (127a) coincides with Marton's R_1-constraint (6a). The R_2-constraint in (127b) is looser than Marton's R_2-constraint (6b):

$$I(U_0,U_1,U_2;Y_2) \geq I(U_0,U_2;Y_2).$$

The sum-rate constraint in (127c) is looser than Marton's sum-rate constraint (6c),

$$\begin{aligned}
&I(U_0,U_1;Y_1) + I(U_2;Y_2|U_0,U_1) \\
&\quad = I(U_0,U_1;Y_1) + H(U_2|U_0,U_1) - H(U_2|U_0,U_1,Y_2) \\
&\quad \geq I(U_0,U_1;Y_1) + H(U_2|U_0,U_1) - H(U_2|U_0,Y_2) \\
&\quad = I(U_0,U_1;Y_1) + H(U_2|U_0,U_1) - H(U_2|U_0,Y_2) \\
&\qquad + H(U_2|U_0) - H(U_2|U_0) \\
&\quad = I(U_0,U_1;Y_1) + I(U_2;Y_2|U_0) - I(U_1;U_2|U_0).
\end{aligned}$$

We now treat the second case $I(U_0,U_1;Y_1) > I(U_0,U_1;Y_2)$. Since $I(U_0;Y_1) < I(U_0;Y_2)$ by assumption and by the continuity of mutual information, there exists a deterministic function f such that

$$I(U_0,f(U_1);Y_1) = I(U_0,f(U_1);Y_2) \tag{128}$$

Let now $U_0' = (U_0, f(U_1))$, $U_1' = U_1$, $U_2' = U_2$ and $X' = X$. For this choice of auxiliaries, Marton's constraints (6) result in:

$$R_1 \leq I(U_0, f(U_1), U_1; Y_1) \tag{129a}$$

$$R_2 \leq I(U_0, f(U_1), U_2; Y_2) \tag{129b}$$

$$\begin{aligned} R_1+R_2 \leq\ & I(U_0, f(U_1), U_1; Y_1)+I(U_2; Y_2|U_0, f(U_1)) \\ & -I(U_1; U_2|U_0, f(U_1)) \end{aligned} \tag{129c}$$

$$\begin{aligned} R_1+R_2 \leq\ & I(U_0, f(U_1), U_2; Y_2) + I(U_1; Y_1|U_0, f(U_1)) \\ & -I(U_1; U_2|U_0, f(U_1)) \end{aligned} \tag{129d}$$

Note that the two sum-rate constraints (129c) and (129d) coincide because $I(U_0, f(U_1); Y_1) = I(U_0, f(U_1); Y_2)$.

We again show that these constraints are no tighter than Marton's constraints in (6), which proves the desired result also for this second case and concludes the proof. The R_1-constraint in (129a) coincides with Marton's R_1-constraint (6a):

$$I(U_0, f(U_1), U_1; Y_1) = I(U_0, U_1; Y_1).$$

The R_2-constraint in (129b) is looser than Marton's R_2-constraint (6b):

$$I(U_0, f(U_1), U_2; Y_2) \geq I(U_0, U_2; Y_2).$$

The sum-rate constraints in (129c) and (129d) are looser than Marton's sum-rate constraint (6c):

$$\begin{aligned} & I(U_0, f(U_1), U_1; Y_1) - I(U_1; U_2|U_0, f(U_1)) + I(U_2; Y_2|U_0, f(U_1)) \\ & = I(U_0, U_1; Y_1) - H(U_2|U_0, f(U_1)) + H(U_2|U_0, U_1) \\ & \quad + H(U_2|U_0, f(U_1)) - H(U_2|U_0, f(U_1), Y_2) \\ & = I(U_0, U_1; Y_1) + H(U_2|U_0, U_1) - H(U_2|U_0, f(U_1), Y_2) \\ & \quad + H(U_2|U_0) - H(U_2|U_0) \\ & \geq I(U_0, U_1; Y_1) + H(U_2|U_0, U_1) - H(U_2|U_0, Y_2) \\ & \quad + H(U_2|U_0) - H(U_2|U_0) \\ & = I(U_0, U_1; Y_1) + I(U_2; Y_2|U_0) - I(U_1; U_2|U_0). \end{aligned}$$

REFERENCES

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

[2] G. Dueck, "Partial feedback for two-way and broadcast channels," *Inform. and Control*, vol. 46, pp. 1–15, July 1980.

[3] O. Shayevitz and M. Wigger, "On the capacity of the discrete memoryless broadcast channel with feedback," *IEEE Trans. on Inf. Th.*, vol. 59, no. 3, pp. 1329–1345, Mar. 2013.

[4] H. Kim, Y.-K. Chia, and Abbas El Gamal, "A note on broadcast channels with stale state information at the transmitter," submitted to *IEEE Trans. on Inf. Th.*, Sept. 2013, available at `http://arxiv.org/abs/1309.7437`.

[5] C. C. Wang, "Capacity of 1-to-k broadcast erasure channel with output feedback," *in Proc. of Allerton Conf.*, Monticello, IL, Sept. 2010.

[6] L. Georgiadis and L. Tassiulas, "Broadcast erasure channel with feedback capacity and algorithms," *in Proc. 5th Workshop on Network Coding, Theory, & Applications*, Lausanne, Switzerland, June 2009.

[7] M. A. Maddah-Ali and D. Tse, "Completely Stale Transmitter Channel State Information is Still Very Useful," *IEEE Trans. on Inf. Th.*, vol. 58, no. 7, pp. 4418–4431, July 2012.

[8] J. Chen and P. Elia, "Toward the performance versus feedback tradeoff for the two-user MISO broadcast channel," *IEEE Trans. on Inf. Th.*, vol. 59, no. 12, pp. 8336–8356, Dec. 2013.

[9] S. Yang, M, Kobayashi, D. Gesbert and X. Yi, "Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT," *IEEE Trans. on Inf. Th.*, vol. 59, no. 1, pp. 315–328, Jan. 2013.

[10] R. Tandon, S. A. Jafar, S. Shamai, and H. V. Poor, "On the synergistic benefits of alternating CSIT for the MISO broadcast channel," *IEEE Trans. on Inf. Th.*, vol. 59, no. 7, pp. 4106–4128, Jul. 2013.

[11] T. Gou and S. Jafar, "Optimal use of current and outdated channel state information: Degrees of freedom of the MISO BC with mixed CSIT," *IEEE Communications Letters*, vol. 16, no. 7, pp. 1084–1087, July 2012.

[12] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. on Inf. Th.*, vol. 49, no. 1, pp. 4–20, Jan. 2003.

[13] R. Venkataramanan and S. S. Pradhan, "An achievable rate region for the broadcast channel with feedback," *IEEE Trans. on Inf. Th.*, vol. 59, no. 10, pp. 6175–6191, Oct. 2013.

[14] L. Ozarow and S. Leung-Yan-Cheong, "An achievable region and outer bound for the Gaussian broadcast channel with feedback," *IEEE Trans. on Inf. Th.*, vol. 30, no. 4, pp. 667–671, July 1984.

[15] S. Vishwanath, W. Wu, and A. Arapostathis, "Gaussian interference networks with feedback: duality, sum capacity and dynamic team problems," in *Proc. 44th Ann. Allerton Conf.* 2005.

[16] N. Elia, "When Bode meets Shannon: control-oriented feedback communication schemes," *IEEE Trans. Automat. Contr.*, vol. 49, no. 9, pp, 1477–1488, Sep. 2004.

[17] E. Ardestanizadeh, P. Minero, and M. Franceschetti, "LQG control approach to Gaussian broadcast channels with feedback," *IEEE Trans. on Inf. Th.*, vol. 58, no. 8, pp. 5267–5278, Aug. 2012.

[18] S. Belhadj Amor, Y. Steinberg, and M. Wigger, "MAC-BC duality with linear-feedback schemes," submitted to *IEEE Trans. on Inf. Th.*, Apr. 2014, available at `http://arxiv.org/abs/1404.2584`.

[19] S. R. Bhaskaran, "Gaussian broadcast channel with feedback," *IEEE Trans. on Inf. Th.* vol. 54, no. 11, pp. 5252–5257, Nov. 2008.

[20] M. Gastpar, A. Lapidoth, Y. Steinberg, and M. Wigger, "Coding schemes and asymptotic capacity for the Gaussian broadcast and interference channels with feedback," *IEEE Trans. on Inf. Th.*, vol. 60, no. 1, pp. 54–71, Jan. 2014.

[21] Y. Wu, P. Minero and M. Wigger, "Insufficiency of linear-feedback schemes in Gaussian broadcast channels with common message," accepted by *IEEE Trans. on Inf. Th.*, available at `http://arxiv.org/abs/1307.5549`.

[22] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. on Inf. Th.*, vol. 25, pp. 306–311, May 1979.

[23] A. El Gamal and T.M Cover, "Capacity theorems for the relay channel," *IEEE Trans. on Inf. Th.*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[24] S. Lim, Y-H Kim, A. El Gamal and S-Y Chung, "Noisy network coding," *IEEE Trans. on Inf. Th.*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[25] J. Hou and G. Kramer, "Short message noisy network coding with a decode-forward option," submitted to *IEEE Trans. on Inf. Th.*, Apr. 2013, available at `http://arxiv.org/abs/1304.1692`.

[26] C. Nair, "Capacity regions of two new classes of two-receiver broadcast channels," *IEEE Trans. on Inf. Th.*, vol. 56, no. 9, pp. 4207–4214, Sept. 2010.

[27] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. on Inf. Th.*, vol. 25, no. 2, pp. 166–169, Mar. 1979.

[28] A. El Gamal and Y.-H. Kim, *Network information theory.* Cambridge, U.K.: Cambridge Univ. Press, 2011.

[29] S. I. Gelfand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," *Probl. Inform. Transm.*, vol. 16, no. 1, pp. 17–25, Jan.-Mar. 1980.

[30] Y. Liang, G. Kramer, and H. V. Poor, "Equivalence of two inner bounds on the capacity region of the broadcast channel," in *Proc. 46th Annu. Allerton Conf. on Commun., Control and Comp.*, pp. 1417–1421, 2008.

[31] C. Nair and A. El Gamal, "An outer bound to the capacity region of the broadcast channel," *IEEE Trans. on Inf. Th.*, vol. 53, no. 1, pp. 350–355, Jan. 2007.

[32] T. M. Cover, "Comments on broadcast channels," *IEEE Trans. on Inf. Th.*, vol. 44, no. 6, pp. 2524–2530, Oct. 1998.

[33] J. Köner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. on Inf. Th.*, vol. 23, no. 1, pp. 60–64, Jan. 1977.

[34] J. Körner and K. Marton, "Comparison of two noisy channels," *Topics in Inf. Th. (ed. by I. Csiszar and P. Elias), Keszthely, Hungary*, pp. 411–423, Aug. 1975.

[35] L. H. Ozarow, and S. K. Leung-Yan-Cheong, "An achievable region and outer bound for the Gaussian broadcast channel with feedback," *IEEE Trans. on Inf. Th.*, vol. 30, no. 4, pp. 667–671, July 1984.