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

Youlong Wu, Michèle Wigger

Abstract—We propose two coding schemes for the two-receiver discrete memoryless broadcast channel (BC) with rate-limited feedback from one or both receivers. They improve over the nofeedback capacity region for a large class of channels, including the class of *strictly essentially less-noisy BCs* that we introduce in this article. Examples of strictly essentially less-noisy BCs are the binary symmetric BC (BSBC) or the binary erasure BC (BEBC) with unequal cross-over or erasure probabilities at the two receivers. When the feedback rates are sufficiently large, our schemes recover all previously known capacity results for discrete memoryless BCs with feedback.

In both our schemes, we let the receivers feed back quantization messages about their receive signals. In the first scheme, the transmitter simply *relays* the quantization information obtained from Receiver 1 to Receiver 2, and vice versa. This provides each receiver with a second observation of the input signal and can thus improve its decoding performance unless the BC is physically degraded. Moreover, each receiver uses its knowledge of the quantization message describing its own outputs so as to attain the same performance as if this message had not been transmitted at all.

In our second scheme the transmitter first *reconstructs and processes* the quantized output signals, and then sends the outcome as a common update information to both receivers. A special case of our second scheme applies also to memoryless BCs without feedback but with strictly-causal state-information at the transmitter and causal state-information at the receivers. It recovers all previous achievable regions also for this setup with state-information.

Index Terms—Broadcast channel, channel capacity, rate-limited feedback

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 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.

Other achievable regions for general DMBCs with perfect or noisy feedback have been proposed by Kramer [4] and by Venkataramanan and Pradhan [5]. Kramer's achievable region can be used to show that feedback improves capacity for some specific *binary symmetric BCs* (BSBC). Comparing the general achievable regions in [3], [4], [5] 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) [6], [7], [8], [9], [10], [11], [12], [13], [14], [15]. The best achievable region for perfect feedback and when the noises at the two receivers are independent is given in [10] and is based on a MAC-BC duality approach. In [12], the asymptotic high-SNR sum-capacity for arbitrary noise correlation is derived.

In this paper, we consider DMBCs with *rate-limited feedback*, where the feedback links from the receivers to the transmitter are assumed to be instantaneous and noiseless but rate-limited. We present two types of coding schemes that use Marton coding [16] in a block-Markov framework (similar to [3]), and where in both types the receivers feed back compression information about their channel outputs of the previous block.

In our type-I schemes, (Schemes IA–IC), the encoder simply *relays* the obtained compression informations as part of the cloud center of the Marton code employed in the next-following block. Each receiver reconstructs the compressed version of the other receiver's outputs, and decodes its intended data and compression information based on this compressed signal and its own channel outputs. The key novelty of our scheme is that in this decoding each receiver cleverly uses its knowledge of the compression messages describing its own previous outputs in a way as to attain the same performance as if this message had not been transmitted at all.

In our type-II scheme, (Scheme II), the encoder uses the feedback messages to *reconstruct* compressed versions of both receivers' channel outputs, and then *processes* these compressed signals together with the previously sent codewords to generate update information for the two receivers. This update information is sent as part of the cloud center of the Marton code employed in the next-following block. Each receiver

uses backward decoding to simultaneously reconstruct the encoder's compressed signal and decode its intended messages sent in the cloud center and satellite.

Our coding schemes exhibit the following features:

- Unlike previous schemes, our new schemes strictly improve over the nofeedback capacity region for a large class of memoryelss BCs, which includes:
  1) *Strictly essentially less-noisy DMBCs*, which we define in this paper and which represent a subclass of Nair's essentially less-noisy DMBCs [17]. They include as special cases the BSBC and the *binary erasure BC* (BEBC) with unequal cross-over probabilities or unequal erasure probabilities at the two receivers, and the *binary symmetric/binary erasure channel BC* (BSC/BEC-BC) for a large range of parameters.
  2) The BSC/BEC-BC for all parameters where this DMBC is more capable [18] and the BSC and the BEC have different capacities.
  3) The memoryless Gaussian BC with unequal noise variances at the two receivers.[1]
  4) An instance of the semideterministic BC as is proved in [19].

- When the feedback-rates are sufficiently large, our new schemes recover all previously known capacity results for DMBCs with perfect feedback. In particular, they improve over the Shayevitz-Wigger scheme for perfect feedback [3] when this latter is restricted to send all the update information in the cloud center. This represents a prominent special case of the Shayevitz-Wigger scheme that subsumes the capacity-achieving scheme by Wang [20] or by Georgiadis & Tassiulas [21] for the two-user BEBC where both receivers know all erasure events.

- Subject to a slight modification, our coding schemes apply also to a setup with noisy feedback links when the receivers can code over them. All our achievable regions remain valid also in this modified setup.

- A special case of our type-II coding scheme applies also to state-dependent DMBCs without feedback but where the receivers learn the state causally and the transmitter learns it strictly causally. Our new achievable region for this state-dependent setup recovers all previous achievable regions. In particular the Degrees of Freedom (DoF) result by Maddah-Ali & Tse [22], and the achievable regions and capacity regions presented by Kim, Chia, and El Gamal in [23].

The idea of our type-I schemes extends to more general networks. In [24], such extended coding schemes are proposed for the discrete memoryless multicast network (DMMN), where one transmitter wishes to communicate a message to multiple receivers over a relay network. The result shows that with feedback, one can strictly improve over noisy network
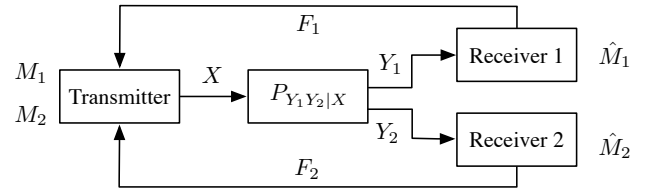

Fig. 1. Broadcast channel with rate-limited feedback

coding [25] and distributed decode-forward coding [26] for some DMMNs.

The rest of this paper is organized as follows: In Sections II and III, we describe the channel model and present some previous results on BCs with and without feedback. In Section IV, we define the class of strictly essentially less-noisy BCs. In Section V, we propose a simple coding scheme that motivates our work. Sections VI and VII present our main results: new achievable regions with rate-limited feedback and conditions under which they improve over the nofeedback capacity region. Section IX describes our new coding schemes achieving the rate regions in Section VI. In Section X, we compare our new achievable regions with previous results and discuss extensions of our results to related setups. Finally, in Section VIII, we numerically evaluate one of our achievable regions for several examples.

A. Notation

Let $\mathbb{R}$ denote the set of real numbers and $\mathbb{Z}^+$ the set of positive integers. 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 j-tuples $X^j := (X_1, \ldots, X_j)$ and $x^j := (x_1, \ldots, x_j)$. Sets are usually denoted by caligraphic letters, e.g., $\mathcal{S}$. For a finite set $\mathcal{S}$, we use $|\mathcal{S}|$ for its cardinality and $\mathcal{S}^j$ for its j-fold Cartesian product $\mathcal{S}^j := \mathcal{S} \times \cdots \times \mathcal{S}$, for $j \in \mathbb{Z}^+$. For a subset $\mathcal{S} \subset \mathbb{R}^2$ we use $\mathrm{bd}(\mathcal{S})$ to denote its boundary and $\mathrm{int}(\mathcal{S})$ for its interior. We also use caligraphic letters for events, mostly $\mathcal{E}$. Moreover, we denote the complement of event $\mathcal{E}$ by $\mathcal{E}^c$.

Given a distribution P_A over some alphabet $\mathcal{A}$, a positive real number $\varepsilon > 0$, and a positive integer n, $\mathcal{T}_\varepsilon^{(n)}(P_A)$ is the typical set in [27]. Given a positive integer n, let $\mathbf{1}_{[n]}$ denote the all-one tuple of length n, e.g., $\mathbf{1}_{[3]} = (1,1,1)$.

We use definitions $\bar{a} := (1-a)$ and $a * b := \bar{a}b + a\bar{b}$, for $a, b \in [0,1]$. Moreover, $Z \sim \mathrm{Bern}(p)$ denotes that Z is a binary random variable taking values 0 and 1 with probabilities $1-p$ and p. We use $H_\mathrm{b}(\cdot)$ for the binary entropy function; thus the entropy of random variable $Z \sim \mathrm{Bern}(p)$ is given by $H_\mathrm{b}(p)$.

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 $P_{Y_1 Y_2 | X}$, and nonnegative feedback rates $R_{\mathrm{Fb},1}$ and $R_{\mathrm{Fb},2}$. 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 $(Y_{1,t}, Y_{2,t}) \sim P_{Y_1 Y_2 | X}(\cdot, \cdot | x_t)$. 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

---

[1]Interestingly, our result hinges on the fact that the receivers are allowed to code over the feedback links: A recent result by Pillai and Prabhakaran [14] shows that when the feedback links are additive Gaussian noise channels of noise variance exceeding a certain threshold, then one cannot improve over the nofeedback capacity if the receivers simply feed back their channel outputs.

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 $R_{\text{Fb},1} \geq 0$, $R_{\text{Fb},2} \geq 0$ is called achievable if for every blocklength N, there exists a set of encoding functions $\{f_t^{(N)}\}_{t=1}^{N}$ and for $i \in \{1,2\}$ 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

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

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 $\mathcal{C}_{\text{NoFB}}$.

When $R_{\text{Fb},1} \geq \log_2 |\mathcal{Y}_1|$ and $R_{\text{Fb},2} \geq \log_2 |\mathcal{Y}_2|$, our setup is equivalent to a perfect-feedback setup where after each channel use the receivers feed back their channel outputs.

III. PREVIOUS INNER AND OUTER BOUNDS

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

A. DMBC without feedback

1) Marton's coding: The capacity region of DMBCs without feedback is in general unknown. The best known inner bound without feedback is Marton's region [16], $\mathcal{R}_{\text{Marton}}$, which is 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) $P_{U_0 U_1 U_2}$ and a function $f : \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$ such that $X = f(U_0, U_1, U_2)$.

To evaluate Marton's region, it suffices to consider distributions $P_{U_0 U_1 U_2 X}$ for which one of the following conditions holds [28], [29], [30]:

- $I(U_0; Y_1) = I(U_0; Y_2)$;
- $I(U_0; Y_1) < I(U_0; Y_2)$ and $U_1 = \text{const.}$;
- $I(U_0; Y_1) > I(U_0; Y_2)$ and $U_2 = \text{const.}$.

2) Superposition coding region: An important subset of Marton's region is the *superposition coding region*, $\mathcal{R}_{\text{SuperPos}}^{(1)}$, which results when Marton's constraints (6) are specialized to $U_1 = \text{const.}$ and $X = U_2$. That means, $\mathcal{R}_{\text{SuperPos}}^{(1)}$ 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 $\mathcal{R}_{\text{SuperPos}}^{(2)}$ is defined in the same way as $\mathcal{R}_{\text{SuperPos}}^{(1)}$ but with exchanged indices 1 and 2.

3) Nair-El Gamal outer bound: In [31] Nair-El Gamal proposed an outer bound on the capacity region of DMBCs without feedback. It is 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 [17]*
- *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.

The various classes of BCs satisfy the relationships [17], [34]:

$$R_1 \leq I(U_0, U_1; Y_1, V_1 | Q) - I(U_0, U_1, U_2, Y_2; V_0, V_1 | Y_1, Q) \tag{9a}$$

$$R_2 \leq I(U_0, U_2; Y_2, V_2 | Q) - I(U_0, U_1, U_2, Y_1; V_0, V_2 | Y_2, Q) \tag{9b}$$

$$R_1 + R_2 \leq I(U_1; Y_1, V_1 | U_0, Q) + I(U_2; Y_2, V_2 | U_0, Q) + \min_{i \in \{1,2\}} I(U_0; Y_i, V_i | Q) - \max_{i \in \{1,2\}} I(U_0, U_1, U_2, Y_1, Y_2; V_0 | Y_i, Q)$$
$$- I(U_0, U_1, U_2, Y_2; V_1 | V_0, Y_1, Q) - I(U_0, U_1, U_2, Y_1, Y_2; V_2 | V_0, Y_2, Q) - I(U_1; U_2 | U_0, Q) \tag{9c}$$

$$R_1 + R_2 \leq I(U_0, U_1; Y_1, V_1 | Q) + I(U_0, U_2; Y_2, V_2 | Q) - I(U_1; U_2 | U_0, Q)$$
$$- I(U_0, U_1, U_2, Y_2; V_0, V_1 | Y_1, Q) - I(U_0, U_1, U_2, Y_1; V_0, V_2 | Y_2, Q) \tag{9d}$$

- 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.

B. DMBC with feedback

Previous results on the DMBC with feedback mostly focus on perfect feedback, which in our setup corresponds to $R_{\text{Fb},1} \geq \log_2 |\mathcal{Y}_1|$ and $R_{\text{Fb},2} \geq \log_2 |\mathcal{Y}_2|$. The previous results that are most closely related to our work are:

1) Shayevitz-Wigger achievable region: The achievable region with feedback that is most closely related to our paper is the Shayevitz-Wigger region [3]. It is the set of all nonnegative rate pairs (R_1, R_2) that satisfy (9) for some pmf $P_Q P_{U_0 U_1 U_2 | Q} P_{V_0 V_1 V_2 | U_0 U_1 U_2 Y_1 Y_2 Q}$ and some function $f: \mathcal{Q} \times \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$, where $X = f(U_0, U_1, U_2, Q)$.

2) Ozarow-Leung outer bound: A simple outer bound on the capacity region with output feedback is given in [6]. It equals the capacity region $\mathcal{C}_{\text{Enh}}^{(1)}$ 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

$$R_1 \leq I(U; Y_1) \tag{10a}$$

$$R_2 \leq I(X; Y_1, Y_2 | U) \tag{10b}$$

for some pmf P_{UX}.

Exchanging everywhere in the previous paragraph indices 1 and 2, we can define a similar enhanced capacity region $\mathcal{C}_{\text{Enh}}^{(2)}$, which is also an outer bound to $\mathcal{C}_{\text{Fb}}$. The intersection $\mathcal{C}_{\text{Enh}}^{(1)} \cap \mathcal{C}_{\text{Enh}}^{(2)}$ yields an even tighter outerbound [20], [21].

IV. DEFINITIONS

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

Definition 1 (From [17]). *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*

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

$$I_P(V; Y_2) + I_P(X; Y_1 | V) \leq I_{P'}(V; Y_2) + I_{P'}(X; Y_1 | V)$$

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 [17]). *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 $P_{U|X}$,*

$$I(U; Y_1) \leq I(U; Y_2). \tag{11}$$

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.

For essentially less-noisy DMBCs, Marton's coding (or superposition coding) is known to achieve capacity [17]. To evaluate the superposition coding region $\mathcal{R}_{\text{SuperPos}}^{(1)}$ of an essentially less-noisy DMBC, it suffices to evaluate the region given by constraints (7) for pmfs P_{UX} that satisfy $I(U; Y_1) \leq I(U; Y_2)$.

In this paper, we introduce the new term *strictly essentially less-noisy BC*, a subclass of essentially less-noisy DMBCs.

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 (11) 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.

V. MOTIVATION: A SIMPLE SCHEME

We sketch a simple scheme that motivates our work. Consider first the coding scheme in Subsection V-A without feedback, on which we build our coding scheme with feedback in Subsection V-B.

A. A coding scheme without feedback

Assume each message is split into B submessages, $M_1 = (M_{1,1}, \ldots, M_{1,B})$ and $M_2 = (M_{2,1}, \ldots, M_{2,B})$. We apply block-Markov coding with $B+1$ blocks of length $n = \lfloor \frac{N}{B+1} \rfloor$, and in each block $b \in \{1, \ldots, B+1\}$ we use superposition coding 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 Message $M_{2,b}$ in the satellite codeword X_b^n. Thus, the scheme is expected to perform well when the following gap is nonnegative:

$$\Gamma := I(U; Y_2) - I(U; Y_1) \geq 0. \tag{12}$$

After each block, both Receivers 1 and 2 decode the cloud center codeword $U_b^n(M_{1,b})$ by means of joint typicality decoding. By the Packing Lemma [27], Receiver 1 will be successful with high probability whenever

$$R_1 < I(U; Y_1) \tag{13}$$

and Receiver 2 will be successful whenever

$$R_1 < I(U; Y_2). \tag{14}$$

Receiver 2 also decodes the satellite codeword $X_b^n(M_{2,b}|M_{1,b})$, which is possible with very high probability whenever $R_2 < I(X; Y_2|U)$.

We notice that when

$$\Gamma > 0, \tag{15}$$

Constraint (14) is not active in view of Constraint (13). In this case, Receiver 2 would be able to decode the cloud center even if it contained a second additional message of rate Γ. The problem is that adding an arbitrary additional message to the cloud center, might make it impossible for Receiver 1 to decode since then (13) could be violated.

We now show that when there is feedback from Receiver 1, the transmitter can identify a suitable additional message that it can add to the cloud center and that will improve the performance of the scheme.

B. Our coding scheme with feedback

Assume there is feedback from Receiver 1, i.e., $R_{\text{Fb},1} > 0$; we ignore feedback from Receiver 2. As in the previous

codebook of satellite codewords $X_b^n(M_{2,b}|M_{1,b}, M_{\text{Fb},1,b-1})$

codebook of cloud centers $U_b^n(M_{1,b}, M_{\text{Fb},1,b-1})$

$M_{\text{Fb},1,b-1}$

$M_{2,b}$

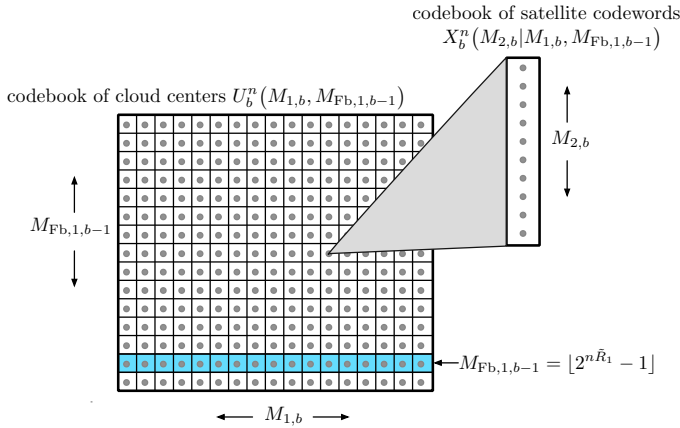$M_{\text{Fb},1,b-1} = \lfloor 2^{n\tilde{R}_1} \rfloor - 1$

$M_{1,b}$

Fig. 2. Code construction in the simple motivating scheme for each block b. Each dot represents a codeword.

subsection, we employ a block-Markov strategy. In each block we use superposition coding as shown in Figure 2: the block-b cloud center U_b^n encodes the two messages $M_{1,b}$ and $M_{\text{Fb},1,b-1}$ and the only satellite encodes message $M_{2,b}$. Here, $M_{\text{Fb},1,b-1}$ is a feedback message that Receiver 1 sent back at the end of block $b-1$. Specifically, Receiver 1 generates $M_{\text{Fb},1,b-1}$ as a Wyner-Ziv message that compresses Receiver 1's block-$(b-1)$ outputs $Y_{1,b-1}^n$ so that a decoder that has side-information $Y_{2,b-1}^n$ can reconstruct the compressed

outputs $\tilde{Y}_{1,b-1}^n$ of $Y_{1,b-1}^n$. Message $M_{\text{Fb},1,b-1}$'s rate $\tilde{R}_1$ thus has to satisfy

$$\tilde{R}_1 > I(\tilde{Y}_1; Y_1|Y_2). \tag{16a}$$

Also, since it is sent over the feedback link, it has to satisfy

$$\tilde{R}_1 < R_{\text{Fb},1}, \tag{16b}$$

and for ease of exposition, we further restrict

$$\tilde{R}_1 < \Gamma. \tag{16c}$$

Decoding is performed as follows. After each block b, Receiver 1 decodes the cloud center. Since it is already aware of message $M_{\text{Fb},1,b-1}$ (it created it itself), in the decoding it can restrict attention to all cloud-center codewords that correspond to the correct value of $M_{\text{Fb},1,b-1}$. (In the code construction in Figure 2, it restricts to a specific row of the cloud center codebook. For example, when $M_{\text{Fb},1,b-1} = \lfloor 2^{n\tilde{R}_1} \rfloor - 1$ it restricts to the light blue row.) Thus, for Receiver 1 the situation is as if message $M_{\text{Fb},1,b-1}$ was not present and had not been sent at all.

After each block b, Receiver 2 performs the following three decoding steps:

- Based on $Y_{2,b}^n$ it decodes both messages $M_{1,b}$ and $M_{\text{Fb},1,b-1}$ in the cloud center.
- It uses the Wyner-Ziv message $M_{\text{Fb},1,b-1}$ and its block-$(b-1)$ outputs $Y_{2,b-1}^n$ to reconstruct $\tilde{Y}_{1,b-1}^n$, the compressed version of $Y_{1,b-1}^n$.
- Based on the triple $(\tilde{Y}_{1,b-1}^n, Y_{2,b-1}^n, U_{b-1}^n)$ it decodes its intended message $M_{2,b-1}$ sent in the satellite $X_{b-1}^n(M_{2,b-1}|M_{1,b-1}, M_{\text{Fb},1,b-2})$ of the previous block.

Receiver 1 errs with vanishingly small probability of error if

$$R_1 < I(U; Y_1). \tag{17a}$$

Receiver 2 errs with vanishingly small probability of error if

$$R_2 < I(X; \tilde{Y}_1, Y_2|U) = I(X; Y_2|U) + I(X; \tilde{Y}_1|U, Y_2), \tag{17b}$$

and if

$$R_1 + \tilde{R}_1 < I(U; Y_2) = I(U; Y_1) + \Gamma,$$

which is already implied by (16c) and (17a).

We conclude by (16) and (17) that the error probability of our scheme tends to 0 as the blocklength n and the number of blocks B tend to infinity, whenever the rate pair (R_1, R_2) satisfies (17) for some pmfs P_{UX} and $P_{\tilde{Y}_1|Y_1}$ that satisfy

$$I(\tilde{Y}_1; Y_1|Y_2) < \min\{\Gamma, R_{\text{Fb},1}\}. \tag{18}$$

Our new constraints (17) differ from the original superposition coding constraints (7) mainly in that the output Y_2 can be replaced by the pair $(Y_2, \tilde{Y}_1)$. This is because in our scheme the compressed output $\tilde{Y}_1$ is conveyed to Receiver 2. Remarkably, there is no cost in conveying this compressed output $\tilde{Y}_1$ to Receiver 2: the compression information for Receiver 2 can be freely *piggybacked* on the data sent to Receiver 1. Our new scheme thus improves over the standard superposition coding scheme whenever $\tilde{Y}_1$ is useful at Receiver 2, i.e., whenever $I(X; \tilde{Y}_1|U, Y_2) > 0$.

In the presented scheme, the transmitter 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 allow to establish a communication link from Receiver 1 to Receiver 2, where the link is of rate

$$\min\{\Gamma, R_{\text{Fb},1}\}. \qquad (19)$$

We use this link to describe the compressed output $\tilde{Y}_1$ to Receiver 2.

C. Extensions

For ease of exposition, we kept above coding scheme as simple as possible. It is easily extended in the following directions:

- The block-b Wyner-Ziv code that compresses $Y_{1,b}^n$ can be superposed on the cloud center U_b^n, since Receiver 1 has already decoded this cloud center before generating (or using) the Wyner-Ziv message $M_{\text{FB},1,b}$.
- When $R_{\text{Fb},2} > 0$, also Receiver 2 can send a Wyner-Ziv compression messages over the feedback link; now *two* additional messages $M_{\text{Fb},1,b-1}$ and $M_{\text{Fb},2,b-1}$ have to be included in the block-b cloud center.
- Superposition coding can be replaced by full Marton coding.
- The receivers can decode the cloud center and their satellites *jointly* based on their own outputs and the compressed version of the other receiver's outputs.
- Sliding-window decoding at the receivers can be replaced by backward decoding.

Each of these modifications can only improve our scheme. However, there is a tension between the first modification and the last two modifications. If a receiver uses backward decoding instead of sliding-window decoding, it can't superpose its Wyner-Ziv code on the cloud center. This is simply because at the time it has to generate its Wyner-Ziv mesage, it hasn't yet decoded the cloud center. The same applies under sliding-window decoding if the receiver decodes the cloud center and the satellite *jointly*.

For each receiver, one has thus to decide on the following two options:

1) use successive sliding-window-decoding of the cloud center and the satellite of the Marton code, and superpose the Wyner-Ziv code on the Marton cloud center; or
2) use joint backward-decoding of the cloud center and the satellite of the Marton code, and *do not* superpose the Wyner-Ziv code on the cloud center.

It is unclear which of the two options performs better. In Section IX we present all three possible combinations: both receivers apply Option 1 (Scheme IA); both receivers apply Option 2 (Scheme IB); one receiver applies Option 1 and the other Option 2 (Scheme IC). The corresponding three achievable regions are given in Theorems 1–3 in the next-following Section VI.

VI. New Achievable regions

The following achievable regions are based on the coding schemes in Section IX.

Our first region is achieved by our scheme IA described in Section IX-A, where both receivers apply successive sliding-window-decoding of the cloud center and the satellite of the Marton code, and they superpose the Wyner-Ziv code on the Marton cloud center.

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*

$$R_1 \leq I(U_0, U_1; Y_1, \tilde{Y}_2 | Q) - I(\tilde{Y}_2; U_0, Y_2 | Y_1, Q) \quad (20\text{a})$$
$$R_2 \leq I(U_0, U_2; Y_2, \tilde{Y}_1 | Q) - I(\tilde{Y}_1; U_0, Y_1 | Y_2, Q) \quad (20\text{b})$$
$$R_1 \leq I(U_0; Y_2 | Q) + \Delta_2 - I(\tilde{Y}_1; Y_1 | U_0, U_2, Y_2, Q) \quad (20\text{c})$$
$$R_2 \leq I(U_0; Y_1 | Q) + \Delta_1 - I(\tilde{Y}_2; Y_2 | U_0, U_1, Y_1, Q) \quad (20\text{d})$$
$$R_1 + R_2 \leq I(U_0, U_1; Y_1, \tilde{Y}_2 | Q) + \Delta_1$$
$$\qquad -I(\tilde{Y}_2; U_0, Y_2 | Y_1, Q) - I(U_1; U_2 | U_0, Q) \quad (20\text{e})$$
$$R_1 + R_2 \leq I(U_0, U_2; Y_2, \tilde{Y}_1 | Q) + \Delta_2$$
$$\qquad -I(\tilde{Y}_1; U_0, Y_1 | Y_2, Q) - I(U_1; U_2 | U_0, Q) \quad (20\text{f})$$
$$R_1 + R_2 \leq I(U_0, U_1; Y_1, \tilde{Y}_2 | Q) + I(U_0, U_2; Y_2, \tilde{Y}_1 | Q)$$
$$\qquad -I(\tilde{Y}_2; U_0, Y_2 | Y_1, Q) - I(\tilde{Y}_1; U_0, Y_1 | Y_2, Q)$$
$$\qquad -I(U_1; U_2 | U_0, Q) \quad (20\text{g})$$

where

$$\Delta_1 := \min\{I(U_2; Y_2, \tilde{Y}_1 | U_0, Q),$$
$$\qquad I(U_2; Y_2, \tilde{Y}_1 | U_0, Q) - I(\tilde{Y}_1; Y_1 | U_0, Y_2, Q) + R_{\text{Fb},1}\}$$
$$\Delta_2 := \min\{I(U_1; Y_1, \tilde{Y}_2 | U_0, Q),$$
$$\qquad I(U_1; Y_1, \tilde{Y}_2 | U_0, Q) - I(\tilde{Y}_2; Y_2 | U_0, Y_1, Q) + R_{\text{Fb},2}\}$$

for some pmf $P_Q P_{U_0 U_1 U_2 | Q} P_{\tilde{Y}_1 | Y_1 U_0 Q} P_{\tilde{Y}_2 | Y_2 U_0 Q}$ *and some function* $f: \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Q} \to \mathcal{X}$ *such that* $\Delta_1, \Delta_2 \geq 0$ *and*

$$I(\tilde{Y}_1; Y_1 | U_0, U_2, Y_2, Q) \leq \min\{I(U_0; Y_2 | Q), R_{\text{Fb},1}\} \quad (21\text{a})$$
$$I(\tilde{Y}_2; Y_2 | U_0, U_1, Y_1, Q) \leq \min\{I(U_0; Y_1 | Q), R_{\text{Fb},2}\} \quad (21\text{b})$$

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

Proof: See Section IX-A. ∎

For sufficiently large feedback rates $R_{\text{Fb},1}$ and $R_{\text{Fb},2}$ (in particular for $R_{\text{Fb},1} \geq \log_2 |\mathcal{Y}_1|$ and $R_{\text{Fb},2} \geq \log_2 |\mathcal{Y}_2|$), we have

$$\Delta_1 = I(U_2; Y_2, \tilde{Y}_1 | U_0, Q) \quad (22\text{a})$$
$$\Delta_2 = I(U_1; Y_1, \tilde{Y}_2 | U_0, Q) \quad (22\text{b})$$

The second region is based on our Scheme IB described in Section IX-B, where the receivers apply joint backward-decoding of the cloud center and the satellite of the Marton code, but they *do not* superpose the Wyner-Ziv code on the cloud center.

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*

$$R_1 \leq I(U_0, U_1; Y_1, \tilde{Y}_2 | Q) - I(\tilde{Y}_2; Y_2 | Y_1, Q) \quad (23\text{a})$$

---

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

[3]The subscript "bw" stands for backward decoding.

$$R_2 \leq I(U_0, U_2; Y_2, \tilde{Y}_1 | Q) - I(\tilde{Y}_1; Y_1 | Y_2, Q) \quad (23b)$$

$$R_1 + R_2 \leq I(U_0, U_1; Y_1, \tilde{Y}_2 | Q) + \Delta_1$$
$$- I(\tilde{Y}_2; Y_2 | Y_1, Q) - I(U_1; U_2 | U_0, Q) \quad (23c)$$

$$R_1 + R_2 \leq I(U_0, U_2; Y_2, \tilde{Y}_1 | Q) + \Delta_2$$
$$- I(\tilde{Y}_1; Y_1 | Y_2, Q) - I(U_1; U_2 | U_0, Q) \quad (23d)$$

$$R_1 + R_2 \leq I(U_0, U_1; Y_1, \tilde{Y}_2 | Q) - I(\tilde{Y}_2; Y_2 | Y_1, Q)$$
$$+ I(U_0, U_2; Y_2, \tilde{Y}_1 | Q) - I(\tilde{Y}_1; Y_1 | Y_2, Q)$$
$$- I(U_1; U_2 | U_0, Q) \quad (23e)$$

for some pmf $P_Q P_{U_0 U_1 U_2 | Q} P_{\tilde{Y}_1 | Y_1 Q} P_{\tilde{Y}_2 | Y_2 Q}$ *and some function* $f : \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Q} \to \mathcal{X}$ *such that*

$$I(\tilde{Y}_1; Y_1 | U_0, U_2, Y_2, Q) \leq R_{\mathrm{Fb},1} \quad (24a)$$

$$I(\tilde{Y}_2; Y_2 | U_0, U_1, Y_1, Q) \leq R_{\mathrm{Fb},2} \quad (24b)$$

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

Proof: See Section IX-B. ∎

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

Remark 1. *Constraints (23) and (24) are looser than Constraints (20) and (21), 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.*

The third region is based on our Scheme IC (Section IX-C), where Receiver 1 applies successive sliding-window-decoding of the cloud center and the satellite of the Marton code, and superposes the Wyner-Ziv code on the Marton cloud center, and Receiver 2 applies joint backward-decoding of the cloud center and the satellite of the Marton code, but *does not* superpose the Wyner-Ziv code.

The scheme is particularly interesting when there is no feedback from Receiver 2, $R_{\mathrm{Fb},2} = 0$, and when Marton's scheme specializes to superposition coding with no satellite codeword for Receiver 1. The rate region corresponding to this special case is presented in Corollary 1 ahead.

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

$$R_1 \leq I(U_0, U_1; Y_1 | Q) \quad (25a)$$

$$R_2 \leq I(U_0, U_2; \tilde{Y}_1, Y_2 | Q)$$
$$- I(\tilde{Y}_1; U_0, U_1, U_2, Y_1 | Y_2, Q) \quad (25b)$$

$$R_1 + R_2 \leq I(U_0, U_1; Y_1 | Q) + \Delta_1$$
$$- I(U_1; U_2 | U_0, Q) \quad (25c)$$

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

for some pmf $P_Q P_{U_0 U_1 U_2 | Q} P_{\tilde{Y}_1 | Y_1 U_0 Q}$ *and some function* $f : \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Q} \to \mathcal{X}$ *such that*

$$I(\tilde{Y}_1; U_1, Y_1 | U_0, U_2, Y_2, Q) \leq R_{\mathrm{Fb},1}. \quad (26)$$

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

Proof: See Section IX-C. ∎

If superposition coding is used instead of Marton coding, Theorem 3 reduces to the following corollary.

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

$$R_1 \leq I(U; Y_1 | Q) \quad (27a)$$

$$R_1 + R_2 \leq I(U; Y_1 | Q) + I(X; Y_2, \tilde{Y}_1 | U, Q) \quad (27b)$$

$$R_1 + R_2 \leq I(X; Y_2 | Q) - I(\tilde{Y}_1; Y_1 | U, Y_2, Q) \quad (27c)$$

for some pmf $P_Q P_{UX|Q} P_{\tilde{Y}_1 | Y_1 U Q}$ *such that*

$$I(\tilde{Y}_1; Y_1 | U, Y_2, Q) \leq R_{\mathrm{Fb},1}. \quad (28)$$

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

Proof: Let $\tilde{Y}_2 = U_1 = $ const., $U = U_0$ and $X = U_2$. Constraint (25a) then specializes to (27a) and Constraint (25b) is redundant compared to Constraint (25d). Observe that Constraints (25d) and (26) are looser than Constraints (27c) and (28), respectively. Also, by (28), Constraint (25c) reduces to (27b). Thus the capacity region $\mathcal{C}_{\mathrm{Fb}}$ includes the region $\mathcal{R}_{\mathrm{relay,sp}}^{(1)}$. Similar arguments hold for $\mathcal{R}_{\mathrm{relay,sp}}^{(2)}$. ∎

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

In our Schemes IA–IC 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 V.

Alternatively, the transmitter can 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 rate region achieved by this Scheme II.

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

$$R_1 \leq I(U_0, U_1; Y_1, \tilde{Y}_1, V | Q)$$
$$- I(V; U_0, U_1, U_2, \tilde{Y}_2 | \tilde{Y}_1, Y_1, Q) \quad (29a)$$

$$R_2 \leq I(U_0, U_2; Y_2, \tilde{Y}_2, V | Q)$$
$$- I(V; U_0, U_1, U_2, \tilde{Y}_1 | \tilde{Y}_2, Y_2, Q) \quad (29b)$$

---

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

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

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

$$R_1 + R_2 \leq I(U_0, U_1; Y_1, \tilde{Y}_1, V|Q) - I(U_1; U_2|U_0, Q)$$
$$- I(V; U_0, U_1, U_2, \tilde{Y}_2|\tilde{Y}_1, Y_1, Q)$$
$$+ I(U_2; Y_2, \tilde{Y}_2, V|U_0, Q) \quad (29c)$$

$$R_1 + R_2 \leq I(U_0, U_2; Y_2, \tilde{Y}_2, V|Q) - I(U_1; U_2|U_0, Q)$$
$$- I(V; U_0, U_1, U_2, \tilde{Y}_1|\tilde{Y}_2, Y_2, Q)$$
$$+ I(U_1; Y_1, \tilde{Y}_1, V|U_0, Q) \quad (29d)$$

$$R_1 + R_2 \leq I(U_0, U_1; Y_1, \tilde{Y}_1, V|Q) - I(U_1; U_2|U_0, Q)$$
$$- I(V; U_0, U_1, U_2, \tilde{Y}_2|\tilde{Y}_1, Y_1, Q)$$
$$- I(V; U_0, U_1, U_2, \tilde{Y}_1|\tilde{Y}_2, Y_2, Q)$$
$$+ I(U_0, U_2; Y_2, \tilde{Y}_2, V|Q) \quad (29e)$$

for some pmf $P_Q P_{U_0 U_1 U_2|Q} P_{\tilde{Y}_1|Y_1 Q} P_{\tilde{Y}_2|Y_2 Q} P_{V|U_0 U_1 U_2 \tilde{Y}_1 \tilde{Y}_2 Q}$ and some function $f: \mathcal{X} \to \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Q}$ where the feedback-rates have to satisfy

$$I(Y_1; \tilde{Y}_1|U_0, U_1, U_2, \tilde{Y}_2, Q) \leq R_{\text{Fb},1} \quad (30a)$$
$$I(Y_2; \tilde{Y}_2|U_0, U_1, U_2, \tilde{Y}_1, Q) \leq R_{\text{Fb},2} \quad (30b)$$
$$I(Y_1, Y_2; \tilde{Y}_1, \tilde{Y}_2|U_0, U_1, U_2, Q) \leq R_{\text{Fb},1} + R_{\text{Fb},2} \quad (30c)$$

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

Proof: See Section IX-D. ∎

When $R_{\text{Fb},1}, R_{\text{Fb},2}$ are sufficiently large, so that we can choose $\tilde{Y}_1 = Y_1$ and $\tilde{Y}_2 = Y_2$, we have the following corollary to Theorem 4.

Corollary 2. *Let $\mathcal{R}_{\text{proc.}}^{\infty}$ be the set of all nonnegative rate pairs (R_1, R_2) that satisfy*

$$R_1 \leq I(U_0, U_1; Y_1, V|Q) - I(V; U_0, U_1, U_2, Y_2|Y_1, Q) \quad (31a)$$

$$R_2 \leq I(U_0, U_2; Y_2, V|Q) - I(V; U_0, U_1, U_2, Y_1|Y_2, Q) \quad (31b)$$

$$R_1 + R_2 \leq I(U_1; Y_1, V|U_0, Q) + I(U_2; Y_2, V|U_0, Q)$$
$$- I(U_1; U_2|U_0, Q) + \min_{i \in \{1,2\}} \{I(U_0; Y_i, V|Q)$$
$$- I(V; U_0, U_1, U_2, Y_1, Y_2|Y_i, Q)\} \quad (31c)$$

$$R_1 + R_2 \leq I(U_0, U_1; Y_1, V|Q) + I(U_0, U_2; Y_2, V|Q)$$
$$- I(V; U_0, U_1, U_2, Y_1|Y_2, Q)$$
$$- I(V; U_0, U_1, U_2, Y_2|Y_1, Q)$$
$$- I(U_1; U_2|U_0, Q) \quad (31d)$$

for some pmf $P_Q P_{U_0 U_1 U_2|Q} P_{V|U_0 U_1 U_2 Y_1 Y_2}$ and some function $f: \mathcal{X} \to \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Q}$, where $X = f(U_0, U_1, U_2, Q)$.

When $R_{\text{Fb},1} \geq \log_2 |\mathcal{Y}_1|$ and $R_{\text{Fb},2} \geq \log_2 |\mathcal{Y}_2|$,[7]

$$\mathcal{R}_{\text{proc.}}^{\infty} \in \mathcal{C}_{\text{Fb}}. \quad (32)$$

As we show in Subsection X-A, our new region $\mathcal{R}_{\text{proc.}}^{\infty}$ includes an important special case of the Shayevitz-Wigger region [3].

---

[7]Smaller feedback rates suffice in general; for simplicity we use these conditions on the feedback rates.

VII. USEFULNESS OF FEEDBACK

Our Scheme IC (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*

$$\Gamma^{(\text{M})} := I(U_0^{(\text{M})}; Y_2^{(\text{M})}) - I(U_0^{(\text{M})}; Y_1^{(\text{M})}) > 0. \quad (33)$$

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),

$$R_1 = (1 - \gamma) R_1^{(\text{M})} + \gamma R_1^{(\text{Enh})} \quad (34a)$$
$$R_2 = (1 - \gamma) R_2^{(\text{M})} + \gamma R_2^{(\text{Enh})} \quad (34b)$$

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

$$(R_1, R_2) \in \mathcal{R}_{\text{relay,hb}}^{(1)}, \quad (35)$$

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. It will be used in the proof of Corollary 4.

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

$$\mathcal{R}_{\text{Marton}} \subsetneq \mathcal{C}_{\text{Fb}}. \quad (36)$$

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

$$\mathcal{C}_{\text{NoFB}} \subsetneq \mathcal{C}_{\text{Fb}}. \quad (37)$$

Proof: Inclusion (37) follows from (36). We show (36). Since $(R_1^{(\text{M})}, R_2^{(\text{M})})$ is in the interior of $\mathcal{C}_{\text{Enh}}^{(1)}$, there exists a rate pair $(R_1^{(\text{Enh})}, R_2^{(\text{Enh})}) \in \mathcal{C}_{\text{Enh}}^{(1)}$ with $R_1^{(\text{Enh})} > R_1^{(\text{M})}$ and $R_2^{(\text{Enh})} > R_2^{(\text{M})}$. Now, since $(R_1^{(\text{M})}, R_2^{(\text{M})})$ lies on the boundary of $\mathcal{R}_{\text{Marton}}$, the rate pair in (34) must lie outside $\mathcal{R}_{\text{Marton}}$ for any $\gamma \in (0, 1)$. By Theorem 5, Equation (35), 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, it is easily verified that the conditions of Corollary 3 hold whenever the DMBCs are not physically degraded. Thus, our corollary immediately shows that for these DMBCs rate-limited feedback strictly increases capacity. (See also Examples 1 and 2 in the next Section.)

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

Remark 3. *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^{(\text{M})} = $ const. and $U_2^{(\text{M})} = X^{(\text{M})}$ and when Condition (33) 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 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.

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 (33). Thus, by Remark 3, 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 $\mathcal{C}_{\text{Enh}}^{(1)}$. ∎

As mentioned, all BSBCs and BEBCs with unequal crossover 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 $\mathcal{C}_{\text{NoFB}} = \mathcal{C}_{\text{Enh}}^{(1)}$. In this case, (even infinite-rate) feedback does not increase the capacity of physically degraded DMBCs [1].

Theorem 5 exhibits sufficient conditions that our coding schemes improve over the nofeedback capacity. These conditions however are not necessary. In fact, by our discussions in Section X and the examples in [21], [20], [23], our schemes can improve over the nofeedback capacity even for DMBCs where for any choice of the auxiliaries $I(U_0; Y_2) = I(U_0; Y_1)$ holds.

On a related note, Bracher and Wigger [19] showed that our type-I schemes can improve over the nofeedback capacity even for semideterministic[8] DMBCs. This might be surprising since without feedback the capacity region of semideterministic DMBCs is achieved by degenerate Marton coding *without cloud center* [35], i.e., $U_0 = $ const..

VIII. EXAMPLES

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

$$Y_1 = X \oplus Z_1, \quad Y_2 = X \oplus Z_2, \tag{38a}$$

for $Z_1 \sim Bern(p_1)$ and $Z_2 \sim Bern(p_2)$ independent noises. Let $Q = $ const., $U \sim Bern(1/2)$, $W_1 \sim Bern(\beta_1)$ and $W_2 \sim Bern(\beta_2)$, for $\beta_1, \beta_2 \in [0, 1/2]$, where U, W_1, W_2 are 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:

$$R_1 \leq 1 - H_b(\beta_1 * p_1) \tag{39a}$$
$$R_1 + R_2 \leq 1 - H_b(\beta_1 * p_1) + H(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$$
$$- H_b(p_2) - H_b(\beta_2 * p_1) \tag{39b}$$
$$R_1 + R_2 \leq 1 - H_b(p_2) - H(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$$
$$+ H_b(\beta_1 * p_2) + H_b(\beta_2) \tag{39c}$$

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

$$H(\alpha_1, \alpha_2, \alpha_3, \alpha_4) - H_b(\beta_1 * p_2) - H_b(\beta_2) \leq R_{\text{Fb},1} \tag{40}$$

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 3 against the no-feedback capacity region $\mathcal{C}_{\text{NoFB}}$.

---

[8]Semideterministic means that the outputs at one of the two receivers are deterministic functions of the inputs.
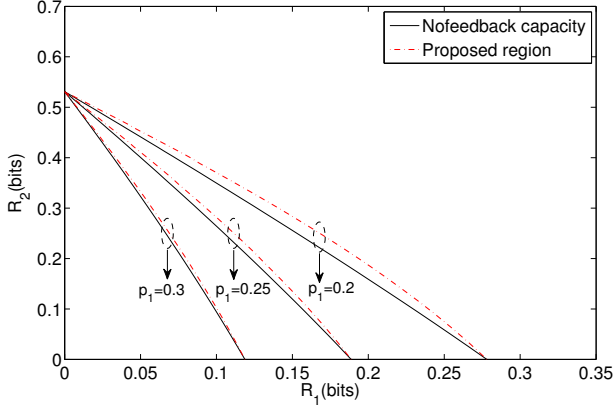
Fig. 3. $\mathcal{C}_{\text{NoFB}}$ and the achievable region in (39) are plotted for BSBCs with parameters $p_2 = 0.1$ and $p_1 \in \{0.2, 0.25, 0.3\}$ and for feedback rate $R_{\text{Fb},1} = 0.8$.

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 $\mathcal{R}_{\text{relay,sp}}^{(1)}$ and $\mathcal{R}_{\text{relay,sp}}^{(2)}$ 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 $R_{\text{Fb},1}, R_{\text{Fb},2} > 0$.*

We distinguish different parameter ranges of our channel.

- $0 < e < H_b(p)$*: In this case, the nofeedback capacity region $\mathcal{C}_{\text{NoFB}}$ [17] is formed by the set of rate pairs (R_1, R_2) that for some $s \in [0, 1/2]$ satisfy*

$$R_1 \leq 1 - H_b(s * p), \tag{41a}$$
$$R_2 \leq (1 - e)H_b(s), \tag{41b}$$
$$R_1 + R_2 \leq 1 - e. \tag{41c}$$

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

$$\gamma \leq \frac{R_{\text{Fb},1}}{(1 - e)H_b(p) + eH_b(s * p)}. \tag{42}$$

Condition (42) assures that the described choice satisfies (28). Then,

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

and

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

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

$$R_1 \leq 1 - H_b(s * p) \tag{43a}$$
$$R_1 + R_2 \leq 1 - H_b(s * p) + (1 - e)H_b(s)$$
$$+ \gamma e(H_b(s * p) - H_b(p)) \tag{43b}$$
$$R_1 + R_2 \leq 1 - e - \gamma(H_b(p)(1 - e) + eH_b(s * p)) \tag{43c}$$

are achievable for any $\gamma \in (0, 1)$ satisfying (42).
As shown in [17], the points (R_1, R_2) of the form

$$(1 - H_b(s * p), (1 - e)H_b(s)), \qquad s \in (0, s_0), \tag{44}$$

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

$$1 - H_b(s_0 * p) + (1 - e)H_b(s_0) = 1 - e. \tag{45}$$

For these boundary points, only the single-rate constraints (41a) and (41b) are active, but not (41c). Thus, comparing (44) to our feedback region (43), we can conclude that by choosing γ sufficiently small, all boundary points (44) lie strictly in the interior of our feedback region $\mathcal{R}_{\text{relay,sp}}^{(1)}$ 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*

$$R_1 \leq \alpha(1 - H_b(p)) \tag{46a}$$
$$R_2 \leq (1 - \alpha)(1 - e). \tag{46b}$$

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

$$\gamma \leq \frac{R_{\text{Fb},2}}{\alpha((1 - e)H_b(p) + H_b(e))}. \tag{47}$$

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

$$R_1 \leq \alpha(1 - H_b(p)) + \alpha(1 - e)\gamma H_b(p) \tag{48a}$$
$$R_1 + R_2 \leq (1 - \alpha)(1 - e) + \alpha(1 - H_b(p))$$
$$+ \alpha(1 - e)\gamma H_b(p) \tag{48b}$$
$$R_1 + R_2 \leq (1 - H_b(p)) - (1 - \alpha)\gamma H_b(e). \tag{48c}$$

are achievable for any $\gamma \in (0, 1)$ satisfying (47).
Since here $1 - H_b(p) > 1 - e$, for small $\gamma > 0$ the feedback region in (48) improves over $\mathcal{C}_{\text{NoFB}}$ given in (46). In fact, (48) improves over all boundary points $(R_1 > 0, R_2 > 0)$ of $\mathcal{C}_{\text{NoFB}}$.

Remark 4. *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 [17]. We conclude that when Receiver 1 is "stronger" than Receiver 2, $\mathcal{R}_{\text{Relay,sp}}^{(2)}$ improves over the no-feedback capacity, for all these classes of BCs even with only one feedback link that is of arbitrary small, but positive rate. Similar arguments hold for $\mathcal{R}_{\text{relay,sp}}^{(1)}$ when Receiver 2 is "stronger" than Receiver 1.*

We plotted our regions (43) and (48) versus the nofeedback capacity region in Figure 4 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.
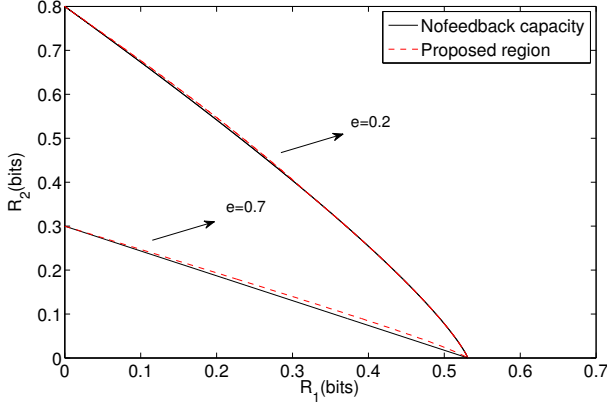
Fig. 4. $\mathcal{C}_{\text{NoFB}}$ and the achievable regions in (43) and (48) are plotted for a BSC/BEC-BC when the BSC has parameter $p = 0.1$ and the BEC has parameter $e \in \{0.2, 0.7\}$. Notice that $0.2 < H_b(p) < 0.7$. The feedback rates $R_{\text{Fb},1} = R_{\text{Fb},2} = 0.8$.
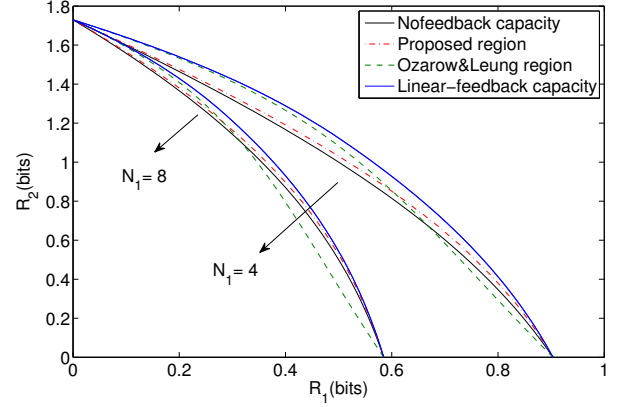
Fig. 5. $\mathcal{C}_{\text{NoFB}}$, linear-feedback capacity [10], Ozarow and Leung's achievable region [6], and the achievable region in (50) are plotted for Gaussian BCs with parameters $P = 10$, $N_2 = 1$, $N_1 \in \{4, 8\}$ and feedback rate $R_{\text{Fb},1} = 0.8$.

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

$$Y_1 = X + Z_1 \tag{49a}$$
$$Y_2 = X + Z_2 \tag{49b}$$

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 independent. We use $C(x) := \frac{1}{2}\log(1+x)$, for $x \geq 0$. Setting $X = U + W_1$, $\tilde{Y}_1 = Y_1 + W_2$, we have

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

and

$$I(X; Y_2, \tilde{Y}_1 | U) = C\left(\frac{\alpha P}{N_2}\right) + C\left(\frac{\alpha P N_2}{(\alpha P + N_2)(N_1 + \beta)}\right)$$
$$I(\tilde{Y}_1; Y_1 | Y_2, U) = C\left(\frac{\alpha P(N_1 + N_2) + N_1 N_2}{\beta(N_2 + \alpha P)}\right).$$

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

$$R_1 \leq C\left(\frac{\bar{\alpha}P}{\alpha P + N_1}\right) \tag{50a}$$

$$R_1 + R_2 \leq C\left(\frac{\bar{\alpha}P}{\alpha P + N_1}\right) + C\left(\frac{\alpha P}{N_2}\right)$$
$$+ C\left(\frac{\alpha P N_2}{(\alpha P + N_2)(N_1 + \beta)}\right) \tag{50b}$$

$$R_1 + R_2 \leq C\left(\frac{P}{N_2}\right) - C\left(\frac{\alpha P(N_1 + N_2) + N_1 N_2}{\beta(N_2 + \alpha P)}\right) \tag{50c}$$

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

$$C\left(\frac{\alpha P(N_1 + N_2) + N_1 N_2}{\beta(N_2 + \alpha P)}\right) \leq R_{\text{Fb},1}. \tag{51}$$

The region is plotted in Figure 5 against the nofeedback capacity region $\mathcal{C}_{\text{NoFB}}$, linear-feedback capacity [10] and the Ozarow-Leung's achievable region [6] with perfect feedback.

From the plots and from expressions (50) and (51) above, it is evident that the capacity region is increased for any positive feedback-rate $R_{\text{Fb},1}, R_{\text{Fb},2} > 0$. By Proposition 1, the same holds also when the feedback links are additive Gaussian noise channels of capacities $R_{\text{Fb},1}$ and $R_{\text{Fb},2}$. This result very much hinges upon the fact that we allow the receivers to code over the feedback link. In fact, Pillai and Prabhakaran showed in a recent work [14] that when the receivers cannot code over the feedback links but are obliged to send back their received channel outputs, then the capacities without feedback and with noisy feedback coincide whenever the noise variances on the two feedback links exceed a certain threshold.

IX. CODING SCHEMES

A. Coding Scheme IA: Sliding-window decoding (Theorem 1)

For simplicity, we only describe the scheme for $Q = \text{const.}$ A general Q can be introduced by coded time-sharing [27, Section 4.5.3]. That means all the codebooks need to be superposed 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 $f: \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$, and pmfs $P_{U_0 U_1 U_2}$, $P_{\tilde{Y}_1 | U_0 Y_1}$, $P_{\tilde{Y}_2 | U_0 Y_2}$. 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 $x_b^n, y_{1,b}^n$ 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 $R_i = R_{p,i} + R_{c,i}$[9]. The feedback message $M_{\mathrm{Fb},i,b-1}$ is uniformly distributed over the set $\mathcal{L}_i$. The common message $M_{c,b} := (M_{c,1,b}, M_{c,2,b})$ is uniformly distributed over the set $\mathcal{M}_c := \mathcal{M}_{c,1} \times \mathcal{M}_{c,2}$.

The coding scheme is illustrated in Table I. In each block b, the transmitter uses Marton's coding to send both feedback messages $(M_{\mathrm{Fb},1,b-1}, M_{\mathrm{Fb},2,b-1})$ and the common message $M_{c,b}$ in the cloud center $U_{0,b}^n(M_{c,b}, M_{\mathrm{Fb},1,b-1}, M_{\mathrm{Fb},2,b-1})$, and the private message $M_{p,i,b}$ in the satellite $U_{i,b}^n(M_{p,i,b}, K_{i,b} | M_{c,b}, M_{\mathrm{Fb},1,b-1}, M_{\mathrm{Fb},2,b-1})$, for $K_{i,b} \in \mathcal{K}_i$, $i \in \{1,2\}$. Receiver 1 first decodes the messages sent in the cloud center $U_{0,b}^n$, and then simultaneously reconstructs Receiver 2's compression outputs $\tilde{Y}_{2,b-1}^n$ and decodes its intended messages sent in the satellite $U_{1,b-1}^n$. Finally, it compresses its channel outputs $Y_{1,b}^n$ by means of Wyner-Ziv coding and sends the compression index $M_{\mathrm{Fb},1,b}$ over the feedback link. Receiver 2 behaves in an analogous way.

1) Codebook generation: For each block $b \in \{1,\ldots,B+1\}$, randomly and independently generate $\lfloor 2^{n(R_{c,1}+R_{c,2}+\tilde{R}_1+\tilde{R}_2)} \rfloor$ sequences $u_{0,b}^n(m_{c,b}, m_{\mathrm{Fb},1,b-1}, m_{\mathrm{Fb},2,b-1})$, for $m_{c,b} \in \mathcal{M}_c$ and $m_{\mathrm{Fb},i,b-1} \in \mathcal{L}_i$, for $i \in \{1,2\}$. Each sequence $u_{0,b}^n(m_{c,b}, m_{\mathrm{Fb},1,b-1}, m_{\mathrm{Fb},2,b-1})$ 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 $u_{0,b}^n(m_{c,b}, m_{\mathrm{Fb},1,b-1}, m_{\mathrm{Fb},2,b-1})$.

For $i \in \{1,2\}$ and each $u_{0,b}^n(m_{c,b}, m_{\mathrm{Fb},1,b-1}, m_{\mathrm{Fb},2,b-1})$ randomly and conditionally independently generate $\lfloor 2^{n(R_{p,i}+R_i')} \rfloor$ sequences $u_{i,b}^n(m_{p,i,b}, k_{i,b} | m_{c,b}, m_{\mathrm{Fb},1,b-1}, m_{\mathrm{Fb},2,b-1})$, for $m_{p,i,b} \in \mathcal{M}_{p,i}$ and $k_{i,b} \in \mathcal{K}_i$, where each $u_{i,b}^n(m_{p,i,b}, k_{i,b} | m_{c,b}, m_{\mathrm{Fb},1,b-1}, m_{\mathrm{Fb},2,b-1})$ 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 $u_{i,b}^n(m_{p,i,b}, k_{i,b} | m_{c,b}, m_{\mathrm{Fb},1,b-1}, m_{\mathrm{Fb},2,b-1})$.

Similarly, for $i \in \{1,2\}$ and each tuple $(m_{c,b}, m_{\mathrm{Fb},1,b-1}, m_{\mathrm{Fb},2,b-1}) \in \mathcal{M}_c \times \mathcal{L}_1 \times \mathcal{L}_2$ randomly generate $\lfloor 2^{n(\tilde{R}_i+\hat{R}_i)} \rfloor$ sequences $\tilde{y}_{i,b}^n(m_{\mathrm{Fb},i,b}, j_{i,b} | m_{c,b}, m_{\mathrm{Fb},1,b-1}, m_{\mathrm{Fb},2,b-1})$, for $m_{\mathrm{Fb},i,b} \in \mathcal{L}_i$ and $j_{i,b} \in \mathcal{J}_i$, by drawing each $\tilde{y}_{i,b}^n(m_{\mathrm{Fb},i,b}, j_{i,b} | m_{c,b}, m_{\mathrm{Fb},1,b-1}, m_{\mathrm{Fb},2,b-1})$ 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 $\tilde{y}_{i,b}^n(m_{\mathrm{Fb},i,b}, j_{i,b} | m_{c,b}, m_{\mathrm{Fb},1,b-1}, m_{\mathrm{Fb},2,b-1})$.

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 $M_{\mathrm{Fb},1,b-1} = m_{\mathrm{Fb},1,b-1}$ and $M_{\mathrm{Fb},2,b-1} = m_{\mathrm{Fb},2,b-1}$. Define $m_{c,b} := (m_{c,1,b}, m_{c,2,b})$. To simplify notation, let $m_{\mathrm{Fb},i,0} = m_{c,i,B+1} = m_{p,i,B+1} = 1$, for $i \in \{1,2\}$ and $m_{c,B+1} = (1,1)$.

---

[9] 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.

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

$$\left(u_{1,b}^n(m_{p,1,b}, k_{1,b} | m_{c,b}, m_{\mathrm{Fb},1,b-1}, m_{\mathrm{Fb},2,b-1}), \right.$$
$$u_{2,b}^n(m_{p,2,b}, k_{2,b} | m_{c,b}, m_{\mathrm{Fb},1,b-1}, m_{\mathrm{Fb},2,b-1}),$$
$$\left. u_{0,b}^n(m_{c,b}, m_{\mathrm{Fb},1,b-1}, m_{\mathrm{Fb},2,b-1}) \right) \in \mathcal{T}_{\varepsilon/16}^{(n)}(P_{U_0U_1U_2}).$$

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

$$x_{b,t} = f(u_{0,b,t}, u_{1,b,t}, u_{2,b,t}), \qquad t \in \{1,\ldots,n\}, \quad (52)$$

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

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 $(\hat{m}_{c,b}^{(1)}, \hat{m}_{\mathrm{Fb},2,b-1}) \in \mathcal{M}_c \times \mathcal{L}_2$ that satisfies

$$\left(u_{0,b}^n(\hat{m}_{c,b}^{(1)}, m_{\mathrm{Fb},1,b-1}, \hat{m}_{\mathrm{Fb},2,b-1}), y_{1,b}^n\right) \in \mathcal{T}_{\varepsilon/8}^{(n)}(P_{U_0Y_1}).$$

Notice that Receiver 1 already knows $m_{\mathrm{Fb},1,b-1}$ 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 $(\hat{m}_{c,b}^{(1)}, \hat{m}_{\mathrm{Fb},2,b-1})$ randomly over the set $\mathcal{M}_c \times \mathcal{L}_2$.

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

$$\left(\tilde{y}_{2,b-1}^n(\hat{m}_{\mathrm{Fb},2,b-1}, \hat{\hat{j}}_{2,b-1} | \hat{m}_{c,b-1}^{(1)}, m_{\mathrm{Fb},1,b-2}, \hat{m}_{\mathrm{Fb},2,b-2}), y_{1,b-1}^n, \right.$$
$$u_{1,b-1}^n(\hat{m}_{p,1,b-1}, \hat{k}_{1,b-1} | \hat{m}_{c,b-1}^{(1)}, m_{\mathrm{Fb},1,b-2}, \hat{m}_{\mathrm{Fb},2,b-2}),$$
$$\left. u_{0,b-1}^n(\hat{m}_{c,b-1}^{(1)}, m_{\mathrm{Fb},1,b-2}, \hat{m}_{\mathrm{Fb},2,b-2}) \right) \in \mathcal{T}_{\varepsilon}^{(n)}(P_{U_0U_1Y_1\tilde{Y}_2}).$$

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

$$\left(\tilde{y}_{1,b}^n(m_{\mathrm{Fb},1,b}, j_{1,b} | \hat{m}_{c,b}^{(1)}, m_{\mathrm{Fb},1,b-1}, \hat{m}_{\mathrm{Fb},2,b-1}), \right.$$
$$\left. u_{0,b}^n(\hat{m}_{c,b}^{(1)}, m_{\mathrm{Fb},1,b-1}, \hat{m}_{\mathrm{Fb},2,b-1}), y_{1,b}^n\right) \in \mathcal{T}_{\varepsilon/4}^{(n)}(P_{Y_1U_0\tilde{Y}_1})$$

and sends the index $m_{\mathrm{Fb},1,b}$ over the feedback link. If there is more than one such pair $(m_{\mathrm{Fb},1,b}, j_{1,b})$ the encoder chooses one of them at random. If there is none, it chooses the index $m_{\mathrm{Fb},1,b}$ 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 $\hat{m}_{1,b} = (\hat{m}_{c,1,b}^{(1)}, \hat{m}_{p,1,b})$, for $b \in \{1,\ldots,B\}$, and $\hat{m}_{c,1,b}^{(1)}$ denotes the first component of $\hat{m}_{c,b}^{(1)}$.

4) Analysis: See Appendix A.

TABLE I
Coding scheme IA: The first four rows in the table depict the encoding process at the transmitter; the following two rows the compression mechanism at the receivers; and the last two rows the messages decoded at the two receivers. The left-to-right arrows in the last two lines indicate that the receivers apply forward decoding.

Block	1	2	$\ldots$	b	$\ldots$
X	$u_{0,1}^n(m_{c,1},1,1)$ $u_{1,1}^n(m_{p,1,1},k_{1,1}\mid m_{c,1},1,1)$ $u_{2,1}^n(m_{p,2,1},k_{2,1}\mid m_{c,1},1,1)$ $x_1^n(u_{0,1}^n,u_{1,1}^n,u_{2,1}^n)$	$u_{0,2}^n(m_{c,2},m_{\text{Fb},1,1},m_{\text{Fb},2,1})$ $u_{1,2}^n(m_{p,1,2},k_{1,2}\mid m_{c,2},m_{\text{Fb},1,1},m_{\text{Fb},2,1})$ $u_{2,2}^n(m_{p,2,2},k_{2,2}\mid m_{c,2},m_{\text{Fb},1,1},m_{\text{Fb},2,1})$ $x_2^n(u_{0,2}^n,u_{1,2}^n,u_{2,2}^n)$	$\ldots$ $\ldots$ $\ldots$ $\ldots$	$u_{0,b}^n(m_{c,b},m_{\text{Fb},1,b-1},m_{\text{Fb},2,b-1})$ $u_{1,b}^n(m_{p,1,b},k_{1,b}\mid m_{c,b},m_{\text{Fb},1,b-1},m_{\text{Fb},2,b-1})$ $u_{2,b}^n(m_{p,2,b},k_{2,b}\mid m_{c,b},m_{\text{Fb},1,b-1},m_{\text{Fb},2,b-1})$ $x_b^n(u_{0,b}^n,u_{1,b}^n,u_{2,b}^n)$	$\ldots$ $\ldots$ $\ldots$ $\ldots$
$\tilde{Y}_1$ $\tilde{Y}_2$	$\tilde{y}_{1,1}^n(m_{\text{Fb},1,1},j_{1,1}\mid m_{c,1},1,1)$ $\tilde{y}_{2,1}^n(m_{\text{Fb},2,1},j_{2,1}\mid m_{c,1},1,1)$	$\tilde{y}_{1,2}^n(m_{\text{Fb},1,2},j_{1,2}\mid m_{c,2},m_{\text{Fb},1,1},m_{\text{Fb},2,1})$ $\tilde{y}_{2,2}^n(m_{\text{Fb},2,2},j_{2,2}\mid m_{c,2},m_{\text{Fb},1,1},m_{\text{Fb},2,1})$	$\ldots$ $\ldots$	$\tilde{y}_{1,b}^n(m_{\text{Fb},1,b},j_{1,b}\mid m_{c,b},m_{\text{Fb},1,b-1},m_{\text{Fb},2,b-1})$ $\tilde{y}_{2,b}^n(m_{\text{Fb},2,b},j_{2,b}\mid m_{c,b},m_{\text{Fb},1,b-1},m_{\text{Fb},2,b-1})$	$\ldots$ $\ldots$
Y_1 Y_2	$\hat{m}_{c,1}^{(1)}$ $\hat{m}_{c,1}^{(2)}$	$\rightarrow(\hat{m}_{c,2}^{(1)},\hat{m}_{\text{Fb},2,1}),(\hat{j}_{2,1},\hat{m}_{p,1,1},\hat{k}_{1,1})$ $\rightarrow(\hat{m}_{c,2}^{(2)},\hat{m}_{\text{Fb},1,1}),(\hat{j}_{1,1},\hat{m}_{p,2,1},\hat{k}_{2,1})$	$\ldots$ $\ldots$	$\rightarrow(\hat{m}_{c,b}^{(1)},\hat{m}_{\text{Fb},2,b-1}),(\hat{j}_{2,b-1},\hat{m}_{p,1,b-1},\hat{k}_{1,b-1})$ $\rightarrow(\hat{m}_{c,b}^{(2)},\hat{m}_{\text{Fb},1,b-1}),(\hat{j}_{1,b-1},\hat{m}_{p,2,b-1},\hat{k}_{2,b-1})$	$\ldots$ $\ldots$

B. Coding Scheme IB: Backward decoding (Theorem 2)

The coding scheme is similar to Scheme IA, except that here the two receivers jointly decode the cloud center and their intended satellite using backward coding, and the Wyner-Ziv codes are not superposed on the Marton cloud center. The scheme is illustrated in Table II.

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

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 $P_{U_0 U_1 U_2}, P_{\tilde{Y}_1\mid Y_1}, P_{\tilde{Y}_2\mid Y_2}$. 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_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 $R_i = R_{p,i} + R_{c,i}$.

1) *Codebook generation*: For each block $b \in \{1,\ldots,B+1\}$, randomly and independently generate $\lfloor 2^{n(R_{c,1}+R_{c,2}+\tilde{R}_1+\tilde{R}_2)}\rfloor$ sequences $u_{0,b}^n(m_{c,b},m_{\text{Fb},1,b-1},m_{\text{Fb},2,b-1})$, for $m_{c,b} \in \mathcal{M}_c := \mathcal{M}_{c,1} \times \mathcal{M}_{c,2}$ and $m_{\text{Fb},i,b-1} \in \mathcal{L}_i$, for $i \in \{1,2\}$. Each sequence $u_{0,b}^n(m_{c,b},m_{\text{Fb},1,b-1},m_{\text{Fb},2,b-1})$ 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 $u_{0,b}^n(m_{c,b},m_{\text{Fb},1,b-1},m_{\text{Fb},2,b-1})$.

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

Also, for $i \in \{1,2\}$, generate $\lfloor 2^{n(\tilde{R}_i+\hat{R}_i)}\rfloor$ sequences $\tilde{y}_{i,b}^n(m_{\text{Fb},i,b},j_{i,b})$, for $m_{\text{Fb},i,b} \in \mathcal{L}_i$ 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 $M_{\text{Fb},1,b-1} = m_{\text{Fb},1,b-1}$ and $M_{\text{Fb},2,b-1} = m_{\text{Fb},2,b-1}$. Define $m_{c,b} := (m_{c,1,b},m_{c,2,b})$. To simplify notation, let $m_{\text{Fb},i,0} = m_{c,i,B+1} = 1$, 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

$$\big(u_{1,b}^n(m_{p,1,b},k_{1,b}\mid m_{c,b},m_{\text{Fb},1,b-1},m_{\text{Fb},2,b-1}),$$
$$u_{2,b}^n(m_{p,2,b},k_{2,b}\mid m_{c,b},m_{\text{Fb},1,b-1},m_{\text{Fb},2,b-1}),$$
$$u_{0,b}^n(m_{c,b},m_{\text{Fb},1,b-1},m_{\text{Fb},2,b-1})\big) \in \mathcal{T}_{\varepsilon/16}^{(n)}(P_{U_0 U_1 U_2}).$$

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

$$x_{b,t} = f(u_{0,b,t},u_{1,b,t},u_{2,b,t}), \qquad t \in \{1,\ldots,n\}, \qquad (53)$$

and $u_{0,b,t}$, $u_{1,b,t}$, $u_{2,b,t}$ denote the t-th symbols of the chosen Marton codewords $u_{0,b}^n(m_{c,b},m_{\text{Fb},1,b-1},m_{\text{Fb},2,b-1})$, $u_{1,b}^n(m_{p,1,b},k_{1,b}\mid m_{c,b},m_{\text{Fb},1,b-1},m_{\text{Fb},2,b-1})$, and $u_{2,b}^n(m_{p,2,b},k_{2,b}\mid m_{c,b},m_{\text{Fb},1,b-1},m_{\text{Fb},2,b-1})$.

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 $(m_{\text{Fb},1,b},j_{1,b}) \in \mathcal{L}_1 \times \mathcal{J}_1$ that satisfies

$$(\tilde{y}_{1,b}^n(m_{\text{Fb},1,b},j_{1,b}),y_{1,b}^n) \in \mathcal{T}_{\varepsilon/4}^{(n)}(P_{\tilde{Y}_1 Y_1}) \qquad (54)$$

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

TABLE II
Coding scheme IB: The first four rows in the table depict the encoding process at the transmitter; the following two rows the compression mechanism at the receivers; and the last two rows the messages decoded at the two receivers. The right-to-left arrows in the last two rows indicate that the receivers use backward decoding.

Block	1	2	$\ldots$	b	$\ldots$
X	$u_{0,1}^n(m_{c,1},1,1)$ $u_{1,1}^n(m_{p,1,1},k_{1,1}\|m_{c,1},1,1)$ $u_{2,1}^n(m_{p,2,1},k_{2,1}\|m_{c,1},1,1)$ $x_1^n(u_{0,1}^n,u_{1,1}^n,u_{2,1}^n)$	$u_{0,1}^n(m_{c,2},m_{\mathrm{Fb},1,1},m_{\mathrm{Fb},2,1})$ $u_{1,2}^n(m_{p,1,2},k_{1,2}\|m_{c,2},m_{\mathrm{Fb},1,1},m_{\mathrm{Fb},2,1})$ $u_{2,2}^n(m_{p,2,2},k_{2,2}\|m_{c,2},m_{\mathrm{Fb},1,1},m_{\mathrm{Fb},2,1})$ $x_2^n(u_{0,2}^n,u_{1,2}^n,u_{2,2}^n)$	$\ldots$	$u_{0,b}^n(m_{c,b},m_{\mathrm{Fb},1,b-1},m_{\mathrm{Fb},2,b-1})$ $u_{1,b}^n(m_{p,1,b},k_{1,b}\|m_{c,b},m_{\mathrm{Fb},1,b-1},m_{\mathrm{Fb},2,b-1})$ $u_{2,b}^n(m_{p,2,b},k_{2,b}\|m_{c,b},m_{\mathrm{Fb},1,b-1},m_{\mathrm{Fb},2,b-1})$ $x_b^n(u_{0,b}^n,u_{1,b}^n,u_{2,b}^n)$	$\ldots$
$\tilde{Y}_1$ $\tilde{Y}_2$	$\tilde{y}_{1,1}^n(m_{\mathrm{Fb},1,1},j_{1,1})$ $\tilde{y}_{2,1}^n(m_{\mathrm{Fb},2,1},j_{2,1})$	$\tilde{y}_{1,2}^n(m_{\mathrm{Fb},1,2},j_{1,2})$ $\tilde{y}_{2,2}^n(m_{\mathrm{Fb},2,2},j_{2,2})$	$\ldots$	$\tilde{y}_{1,b}^n(m_{\mathrm{Fb},1,b},j_{1,b})$ $\tilde{y}_{2,b}^n(m_{\mathrm{Fb},2,b},j_{2,b})$	$\ldots$
Y_1 Y_2	$(\hat{j}_{2,1},\hat{m}_{c,1}^{(1)},\hat{m}_{p,1,1},\hat{k}_{1,1})$ $(\hat{j}_{1,1},\hat{m}_{c,1}^{(2)},\hat{m}_{p,2,1},\hat{k}_{2,1})$	$\leftarrow(\hat{j}_{2,2},\hat{m}_{c,2}^{(1)},\hat{m}_{\mathrm{Fb},2,1},\hat{m}_{p,1,2},\hat{k}_{1,2})$ $\leftarrow(\hat{j}_{1,2},\hat{m}_{c,2}^{(2)},\hat{m}_{\mathrm{Fb},1,1},\hat{m}_{p,2,2},\hat{k}_{2,2})$	$\ldots$	$\leftarrow(\hat{j}_{2,b},\hat{m}_{c,b}^{(1)},\hat{m}_{\mathrm{Fb},2,b-1},\hat{m}_{p,1,b},\hat{k}_{1,b})$ $\leftarrow(\hat{j}_{1,b},\hat{m}_{c,b}^{(2)},\hat{m}_{\mathrm{Fb},1,b-1},\hat{m}_{p,2,b},\hat{k}_{2,b})$	$\ldots$

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 $m_{\mathrm{Fb},2,b}$. Moreover, it also knows its own feedback messages $m_{\mathrm{Fb},1,b-1}$ and $m_{\mathrm{Fb},1,b}$ because it has created them itself, see point 3). Now, when observing $y_{1,b}^n$, Receiver 1 looks for a tuple $(\hat{\hat{j}}_{2,b},\hat{m}_{c,b}^{(1)},\hat{m}_{\mathrm{Fb},2,b-1},\hat{m}_{p,1,b},\hat{k}_{1,b}) \in \mathcal{J}_2 \times \mathcal{M}_c \times \mathcal{L}_2 \times \mathcal{M}_{p,1} \times \mathcal{K}_1$ that satisfies

$$\Big(u_{1,b}^n(\hat{m}_{p,1,b},\hat{k}_{1,b}|\hat{m}_{c,b}^{(1)},m_{\mathrm{Fb},1,b-1},\hat{m}_{\mathrm{Fb},2,b-1}),\tilde{y}_{2,b}^n(m_{\mathrm{Fb},2,b},\hat{\hat{j}}_{2,b}),$$
$$u_{0,b}^n(\hat{m}_{c,b}^{(1)},m_{\mathrm{Fb},1,b-1},\hat{m}_{\mathrm{Fb},2,b-1}),y_{1,b}^n\Big) \in \mathcal{T}_\varepsilon^{(n)}(P_{U_0U_1Y_1\tilde{Y}_2}).$$

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 $\hat{m}_{c,b}^{(1)}$.

5) Analysis: See Appendix B.

C. Coding Scheme IC: Hybrid sliding-window decoding and backward decoding (Theorem 3)

Our third scheme IC is a mixture of the Scheme IA and IB: Receiver 1 behaves as in Scheme IA and Receiver 2 as in Scheme IB. It achieves region $\mathcal{R}_{\mathrm{relay,hb}}^{(1)}$. A similar scheme achieves region $\mathcal{R}_{\mathrm{relay,hb}}^{(2)}$.

For simplicity we consider only $Q = \mathrm{const}$.

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

2) Encoding: The transmitter performs the same encoding procedure as in Section IX-A, but where $m_{\mathrm{Fb},2,b-1} = 1$ is constant for all blocks $b \in \{1,\ldots,B+1\}$.

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 $(\hat{m}_{c,b-1},\hat{m}_{p,1,b-1},\hat{k}_{1,b-1}) \in \mathcal{M}_c \times \mathcal{M}_{p,1} \times \mathcal{K}_1$ that satisfies

$$\Big(u_{0,b-1}^n(\hat{m}_{c,b-1},m_{\mathrm{Fb},1,b-2},1),y_{1,b-1}^n,$$
$$u_{1,b-1}^n(\hat{m}_{p,1,b-1},\hat{k}_{1,b-1}|\hat{m}_{c,b-1},m_{\mathrm{Fb},1,b-2},1)\Big) \in \mathcal{T}_\varepsilon^{(n)}(P_{U_0U_1Y_1}).$$

It further compresses the outputs $y_{1,b}^n$ and sends the feedback message $m_{\mathrm{Fb},1,b}$ over the feedback link as in Scheme IA, see point 3) in Section IX-A.

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

5) Analysis: Similar to the analysis of the schemes IA and IB 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 $M_{\mathrm{Fb},1,b},M_{\mathrm{Fb},2,b}$, the encoder first reconstructs the compressed versions of the channel outputs, $\tilde{Y}_{1,b}^n$ and $\tilde{Y}_{2,b}^n$. It then newly compresses the quintuple consisting of $\tilde{Y}_{1,b}^n$ and $\tilde{Y}_{2,b}^n$ and the Marton codewords $U_{0,b}^n, U_{1,b}^n, U_{2,b}^n$ that it had sent during block b. This new compression information is 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 $Y_{i,b}^n$ to simultaneously reconstruct the encoder's compressed signal and decode its intended messages sent in block b. The scheme is illustrated in Table III.

For simplicity, we only describe the scheme for $Q = \mathrm{const}$.

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 $f: \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$, 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 $x_b^n, y_{1,b}^n$ 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 $i \in \{1,2\}$, 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

TABLE III

Coding scheme II: The first row depicts the compression mechanism at the transmitter; the following four rows the encoding mechanism at the transmitter; the subsequent two rows the compression mechanism at the receivers; and the last two rows the messages decoded at the two receivers. The right-to-left arrows in the last two rows indicates that backward-decoding is used.

Block	1	2	...	b	...
	$v_1^n(n_1\vert 1)$	$v_2^n(n_2\vert n_1)$	...	$v_b^n(n_b\vert n_{b-1})$	...
X	$u_{0,b}^n(m_{c,1},1)$ $u_{1,1}^n(m_{p,1,1},k_{1,1}\vert m_{c,1},1)$ $u_{2,1}^n(m_{p,2,1},k_{2,1}\vert m_{c,1},1)$ $x_1^n(u_{0,1}^n,u_{1,1}^n,u_{2,1}^n)$	$u_{0,2}^n(m_{c,2},n_1)$ $u_{1,2}^n(m_{p,1,2},k_{1,2}\vert m_{c,2},n_1)$ $u_{2,2}^n(m_{p,2,2},k_{2,2}\vert m_{c,2},n_1)$ $x_2^n(u_{0,2}^n,u_{1,2}^n,u_{2,2}^n)$	...	$u_{0,b}^n(m_{c,b},n_{b-1})$ $u_{1,b}^n(m_{p,1,b},k_{1,b}\vert m_{c,b},n_{b-1})$ $u_{2,b}^n(m_{p,2,b},k_{2,b}\vert m_{c,b},n_{b-1})$ $x_b^n(u_{0,b}^n,u_{1,b}^n,u_{2,b}^n)$	...
$\tilde Y_1$	$\tilde y_{1,1}^n(m_{\mathrm{Fb},1,1},j_{1,1})$	$\tilde y_{1,2}^n(m_{\mathrm{Fb},1,2},j_{1,2})$	...	$\tilde y_{1,b}^n(m_{\mathrm{Fb},1,b},j_{1,b})$	...
$\tilde Y_2$	$\tilde y_{2,1}^n(m_{\mathrm{Fb},2,1},j_{2,1})$	$\tilde y_{2,2}^n(m_{\mathrm{Fb},2,2},j_{2,2})$	...	$\tilde y_{2,b}^n(m_{\mathrm{Fb},2,b},j_{2,b})$	...
Y_1	$(\hat m_{c,1}^{(1)},\hat m_{p,1,1},\hat k_{1,1})$	$\leftarrow (\hat m_{c,2}^{(1)},\hat m_{p,1,2},\hat k_{1,2},\hat n_1^{(1)})$	...	$\leftarrow (\hat m_{c,b}^{(1)},\hat m_{p,1,b},\hat k_{1,b},\hat n_{b-1}^{(1)})$	...
Y_2	$(\hat m_{c,1}^{(2)},\hat m_{p,2,1},\hat k_{2,1})$	$\leftarrow (\hat m_{c,2}^{(2)},\hat m_{p,2,2},\hat k_{2,2},\hat n_1^{(2)})$	...	$\leftarrow (\hat m_{c,b}^{(2)},\hat m_{p,2,b},\hat k_{2,b},\hat n_{b-1}^{(2)})$	...

$b \in \{1,\dots,B\}$. The submessages $M_{c,i,b}$, and $M_{p,i,b}$ are uniformly distributed over the sets $\mathcal{M}_{c,i} := \{1,\dots,\lfloor 2^{nR_{c,i}}\rfloor\}$ and $\mathcal{M}_{p,i} := \{1,\dots,\lfloor 2^{nR_{p,i}}\rfloor\}$, respectively, where $R_{p,i}, R_{c,i} > 0$ and so that $R_i = R_{p,i} + R_{c,i}$.

1) Codebook generation: For each block $b \in \{1,\dots,B+1\}$, randomly and independently generate $\lfloor 2^{n(R_{c,1}+R_{c,2}+\tilde R_1+\tilde R_2)}\rfloor$ sequences $u_{0,b}^n(m_{c,b},n_{b-1})$, for $m_{c,b} \in \mathcal{M}_c := \mathcal{M}_{c,1} \times \mathcal{M}_{c,2}$ and $n_{b-1} \in \mathcal{N}$. Each sequence $u_{0,b}^n(m_{c,b},n_{b-1})$ 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 $u_{0,b}^n(m_{c,b},n_{b-1})$.

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

Also, for $i \in \{1,2\}$, generate $\lfloor 2^{n(\tilde R_i+\hat R_i)}\rfloor$ sequences $\tilde y_{i,b}^n(m_{\mathrm{Fb},i,b},j_{i,b})$, for $m_{\mathrm{Fb},i,b} \in \mathcal{L}_i$ 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 $\lfloor 2^{nR_v}\rfloor$ sequences $v_b^n(n_b\vert n_{b-1})$, 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 $b \in \{1,\dots,B+1\}$. 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 $m_{c,b} := (m_{c,1,b},m_{c,2,b})$. To simplify notation, let $m_{\mathrm{Fb},i,0} = m_{c,i,B+1} = m_{p,i,B+1} = 1$, 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 $\tilde Y_{1,b-1}^n$ and $\tilde Y_{2,b-1}^n$. Assume that after block $b-1$ the transmitter received the feedback messages $M_{\mathrm{Fb},1,b-1} = m_{\mathrm{Fb},1,b-1}$ and $M_{\mathrm{Fb},2,b-1} = m_{\mathrm{Fb},2,b-1}$, and that in this previous block it had produced the Marton codewords $u_{0,b-1}^n := u_{0,b-1}^n(m_{c,b-1},n_{b-2})$, $u_{1,b-1}^n := u_{1,b-1}^n(m_{p,1,b-1},k_{1,b-1}\vert m_{c,b-1},n_{b-2})$, and $u_{2,b-1}^n := u_{2,b-1}^n(m_{p,2,b-1},k_{2,b-1}\vert m_{c,b-1},n_{b-2})$. 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

$$\big(u_{0,b-1}^n,u_{1,b-1}^n,u_{2,b-1}^n,\tilde y_{1,b-1}^n(m_{\mathrm{Fb},1,b-1},\hat j_{1,b-1}),$$
$$\tilde y_{2,b-1}^n(m_{\mathrm{Fb},2,b-1},\hat j_{2,b-1})\big) \in \mathcal{T}_{\varepsilon/4}^{(n)}(P_{U_0U_1U_2\tilde Y_1\tilde Y_2}).$$

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 $n_{b-1} \in \mathcal{N}$ that satisfies

$$\big(u_{0,b-1}^n,u_{1,b-1}^n,u_{2,b-1}^n,\tilde y_{1,b-1}^n(m_{\mathrm{Fb},1,b-1},\hat j_{1,b-1}),$$
$$\tilde y_{2,b-1}^n(m_{\mathrm{Fb},2,b-1},\hat j_{2,b-1}),v_{b-1}^n(n_{b-1}\vert n_{b-2})\big)$$
$$\in \mathcal{T}_{\varepsilon/2}^{(n)}(P_{U_0U_1U_2\tilde Y_1\tilde Y_2V}).$$

The transmitter now sends the fresh data and the compression message 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

$$\big(u_{0,b}^n(m_{c,b},n_{b-1}),$$
$$u_{1,b}^n(m_{p,1,b},k_{1,b}\vert m_{c,b},n_{b-1}),$$
$$u_{2,b}^n(m_{p,2,b},k_{2,b}\vert m_{c,b},n_{b-1})\big) \in \mathcal{T}_{\varepsilon/64}^{(n)}(P_{U_0U_1U_2}).$$

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 $x_b^n = (x_{b,1},\dots,x_{b,n})$, where

$$x_{b,t} = f(u_{0,b,t},u_{1,b,t},u_{2,b,t}), \qquad t \in \{1,\dots,n\}. \quad (55)$$

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

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 $(m_{\mathrm{Fb},1,b},j_{1,b}) \in \mathcal{L}_1 \times \mathcal{J}_1$ that satisfies

$$\big(\tilde y_{1,b}^n(m_{\mathrm{Fb},1,b},j_{1,b}),y_{1,b}^n\big) \in \mathcal{T}_{\varepsilon/16}^{(n)}(P_{\tilde Y_1Y_1}) \quad (56)$$

and sends the index $m_{\text{Fb},1,b}$ over the feedback link. If there is more than one such pair $(m_{\text{Fb},1,b}, j_{1,b})$ the encoder chooses one of them at random. If there is none, it chooses the index $m_{\text{Fb},1,b}$ 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 $b \in \{1, \ldots, B+1\}$, 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 compression index n_b. Now, when observing $y_{1,b}^n$, Receiver 1 looks for a tuple $(\hat{m}_{c,b}^{(1)}, \hat{m}_{p,1,b}, \hat{k}_{1,b}, \hat{n}_{b-1}) \in \mathcal{M}_c \times \mathcal{M}_{p,1} \times \mathcal{K}_1 \times \mathcal{N}$ that satisfies

$$\big(u_{0,b}^n(\hat{m}_{c,b}^{(1)}, \hat{n}_{b-1}), u_{1,b}^n(\hat{m}_{p,1,b}, \hat{k}_{1,b} | \hat{m}_{c,b}^{(1)}, \hat{n}_{b-1}),$$
$$v_b^n(n_b | \hat{n}_{b-1}), y_{1,b}^n, \tilde{y}_{1,b}^n(m_{\text{Fb},1,b}, j_{1,b}) \big)$$
$$\in \mathcal{T}_\varepsilon^{(n)}(P_{U_0 U_1 V Y_1 \tilde{Y}_1}),$$

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

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 $\hat{m}_{c,1,b}^{(1)}$.

5) Analysis: See Appendix C.

X. RELATED SETUPS AND PREVIOUS WORKS

In this section we compare our results to previous works, and we discuss the related setups of DMBCs with noisy feedback channels and state-dependent DMBCs with strictly-causal state-information at the transmitter and causal state-information at the receivers.

A. The Shayevitz-Wigger region and our region $\mathcal{R}_{\text{proc.}}^\infty$

Consider a restricted form of the Shayevitz-Wigger region for output feedback where in the evaluation of Constraints (9) we limit the choices of the auxiliaries to

$$V_1 = V_2 = V_0 = V. \tag{57}$$

This restricted Shayevitz-Wigger region for output feedback is included in our new achievable region $\mathcal{R}_{\text{proc.}}^\infty$ when the feedback rates satisfy

$$R_{\text{Fb},1} \geq \log_2 |\mathcal{Y}_1| \quad \text{and} \quad R_{\text{Fb},2} \geq \log_2 |\mathcal{Y}_2|, \tag{58}$$

which is seen as follows. Recall that our achievable region $\mathcal{R}_{\text{proc.}}^\infty$ is characterized by Constraints (31). When the constraints in (9) are specialized to (57), then (9a), (9b), and (9d) coincide with Constraints (31a), (31b), and (31d), and the sum-rate constraint (9c) is tighter than the sum-rate constraint in (31c). The latter holds because $\min_{i=\{1,2\}}\{a_i - b_i\} \geq \min_{i \in \{1,2\}} a_i - \max_{i \in \{1,2\}} b_i$ for any nonnegative $\{a_i, b_i\}_{i=1}^2$.

As explained at the end of the next subsection, choices as in (57) are particularly interesting, see also (59) ahead.

B. The Shayevitz-Wigger region and our region $\mathcal{R}_{\text{relay,bw}}$

Consider a restricted Shayevitz-Wigger region where in the evaluation of Constraints (9) we limit the choices of the auxiliaries to

$$V_1 = V_2 = V_0 = V = (\psi_1(Y_1, Q), \psi_2(Y_2, Q)) \tag{59}$$

for some functions ψ_1 and ψ_2 on appropriate domains.

For choices satisfying (59), the Shayevitz-Wigger rate-constraints in (9) reduce to

$$R_1 \leq I(U_0, U_1; Y_1, V | Q) - I(Y_2; V | Q, Y_1) \tag{60a}$$
$$R_2 \leq I(U_0, U_2; Y_2, V | Q) - I(Y_1; V | Q, Y_2) \tag{60b}$$
$$R_1 + R_2 \leq I(U_1; Y_1, V | Q, U_0) + I(U_2; Y_2, V | Q, U_0)$$
$$+ \min_{i \in \{1,2\}} I(U_0; Y_i, V | Q)$$
$$- \max_{i \in \{1,2\}} I(Y_1, Y_2; V | Q, Y_i) - I(U_1; U_2 | Q, U_0) \tag{60c}$$

$$R_1 + R_2 \leq I(U_0, U_1; Y_1, V | Q) + I(U_0, U_2; Y_2, V | Q)$$
$$- I(U_1; U_2 | Q, U_0) - I(Y_2; V | Q, Y_1)$$
$$- I(Y_1; V | Q, Y_2) \tag{60d}$$

Our new achievable region $\mathcal{R}_{\text{relay,bw}}$, which is characterized by Constraints (23), 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

$$\tilde{Y}_1 = \psi_1(Y_1, Q) \quad \text{and} \quad \tilde{Y}_2 = \psi_2(Y_2, Q) \tag{61}$$

and so that Δ_1, Δ_2 satisfy (22).

In fact, for choices as in (59), the rate constraints in (60a), (60b), and (60d) coincide with constraints (23a), (23b), and (23e) when in the latter we choose $\tilde{Y}_1$ and $\tilde{Y}_2$ as described by (61). Moreover, the combination of the two sum-rate constraints (23c) and (23d) is looser than the sum-rate constraint (60c), 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.

Choices as in (57) or (59) describe an important special case of the Shayevitz-Wigger scheme, as all evaluations of the Shayevitz-Wigger region known to date are based on them. In particular, the capacity-achieving scheme for the two-user BEC-BC when both receivers are informed of all erasure events [20], [21] is a special case of the Shayevitz-Wigger scheme with a choice of auxiliaries as in (59). See also the following Subsections X-E and X-F for further usages of this choice.

C. Noisy feedback channels

Consider the related scenario where the feedback links are replaced by noisy channels of capacities $R_{\text{Fb},1}$ and $R_{\text{Fb},2}$ and where *the decoders can code over these feedback links*. The following three modifications to our coding schemes suffice to ensure that our achievable regions remain valid:

- We interleave 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 capacity-achieving codes 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.

Proposition 1. *Consider a DMBC with noisy feedback channels of capacities $R_{\text{Fb},1}$ and $R_{\text{Fb},2}$ where the receivers can code over the feedback channels. When modified as described above, our coding schemes in Section IX achieve the rate regions in Theorems 1–4.*

Proof: See Appendix E. ∎

D. State-dependent DMBCs with state-information at the transmitter and the receivers

Consider a state-dependent DMBC of channel law $P_{Y_1 Y_2 | X S}$, where the state S is observed causally at both receivers and strictly-causally at the transmitter. That means, the transmitter learns the time-t state S_t right after producing the time-t input X_t, and the receivers learn S_t at the same time as their outputs $Y_{1,t}$ or $Y_{2,t}$. There is no feedback.

Our coding schemes in the previous section also apply to this related scenario when the following modifications are applied: Let the "feedback messages" $\{M_{1,\text{Fb},b}\}$ and $\{M_{2,\text{Fb},b}\}$, for $b \in \{1, \ldots, B\}$, be computed in function of the state symbols S_b^n only. In this case, the receivers do not have to feed back anything, since the transmitter observes S_b^n and can generate $\{M_{1,\text{Fb},b}\}$ and $\{M_{2,\text{Fb},b}\}$ locally.

This means that our new achievable regions in Theorems 1–4 are also valid for this related scenario, if we replace

- the channel law $P_{Y_1 Y_2 | X}$ by a state-dependent law $P_{Y_1 Y_2 | X S}$;
- the compression variables $\tilde{Y}_1$ and $\tilde{Y}_2$ by the state S;
- the outputs Y_1 and Y_2 by the state-augmented outputs (Y_1, S) and (Y_2, S);

and we set $R_{\text{Fb},1}, R_{\text{Fb},2} > H(S)$. (This last condition implies that Constraints (30) are always satisfied and thus can be dropped.)

In particular, we have the following corollary to Theorem 4:

Corollary 5. *Consider a state-dependent DMBC $P_{Y_1 Y_2 | X S}$ where the transmitter observes the state-sequence strictly-causally, and the receivers observe it causally.*

For this scenario the region $\mathcal{R}_{\text{proc.}}^{\text{state}}$ is achievable, where $\mathcal{R}_{\text{proc.}}^{\text{state}}$ is the set of all nonnegative rate pairs (R_1, R_2) that satisfy

$$R_1 \leq I(U_0, U_1; Y_1, V | S, Q)$$
$$- I(V; U_0, U_1, U_2, Y_2 | Y_1, S, Q) \quad \text{(62a)}$$
$$R_2 \leq I(U_0, U_2; Y_2, V | S, Q)$$
$$- I(V; U_0, U_1, U_2, Y_1 | Y_2, S, Q) \quad \text{(62b)}$$
$$R_1 + R_2 \leq I(U_1; Y_1, V | U_0, S, Q) + I(U_2; Y_2, V | U_0, S, Q)$$

$$- I(U_1; U_2 | U_0, S, Q)$$
$$+ \min_{i \in \{1,2\}} \{ I(U_0; Y_i, V | S, Q)$$
$$- I(V; U_0, U_1, U_2, Y_1, Y_2 | Y_i, S, Q) \} \quad \text{(62c)}$$
$$R_1 + R_2 \leq I(U_0, U_1; Y_1, V | S, Q) + I(U_0, U_2; Y_2, V | S, Q)$$
$$- I(V; U_0, U_1, U_2, Y_1 | Y_2, S, Q)$$
$$- I(V; U_0, U_1, U_2, Y_2 | Y_1, S, Q)$$
$$- I(U_1; U_2 | U_0, Q) \quad \text{(62d)}$$

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

The state-dependent setup considered in this subsection can also be modelled as a DMBC with generalized feedback [3]. Consequently, the Shayevitz-Wigger achievable region for generalized feedback [3][10] (see in particular also [23, Corollary 1]) applies.

By similar arguments as used in the previous two subsections, it can be shown that for state-dependent DMBCs with causal and strictly-causal state-information at the receivers and the transmitter our new achievable region $\mathcal{R}_{\text{proc.}}^{\text{state}}$ improves over a restricted version of the Shayevitz-Wigger region for generalized feedback when in the latter the choice of auxiliaries is limited to (57).

E. The Kim-Chia-El Gamal region for state-dependent DM-BCs and our regions $\mathcal{R}_{\text{proc.}}^{\text{state}}$ and $\mathcal{R}_{\text{relay,bw}}$

Consider the class of *deterministic* state-dependent DMBCs where the two outputs Y_1 and Y_2 are deterministic functions of the input X and the random state S. As in the previous Subsection we assume that the transmitter has strictly-causal state-information and both receivers have causal state-information.

Kim, Chia, and El Gamal [23] evaluated the maximum sum-rate achieved by the Shayevitz-Wigger scheme with generalized feedback [3] for deterministic state-dependent DMBCs when the auxiliary random variables of the Shayevitz-Wigger scheme are restricted to one of the following two choices:

- Choice 1 (coded time-sharing):

$$Q = \begin{cases} 0 & \text{w. p.} \quad 1 - 2p \\ 1 & \text{w. p.} \quad p \\ 2 & \text{w. p.} \quad p \end{cases}, \quad \text{(63a)}$$

$$V_0 = V_1 = V_2 = \begin{cases} \emptyset & \text{if } Q = 0 \\ Y_2 & \text{if } Q = 1 \\ Y_1 & \text{if } Q = 2 \end{cases}, \quad \text{(63b)}$$

and

$$X = \begin{cases} U_0 & \text{if } Q = 0 \\ U_1 & \text{if } Q = 1 \\ U_2 & \text{if } Q = 2 \end{cases} \quad \text{(63c)}$$

for $0 \leq p \leq 0.5$ and some pmf $P_{U_0 U_1 U_2} = P_{U_0} P_{U_1} P_{U_2}$.

---

[10]The Shayevitz-Wigger region for generalized feedback [3] is characterized by constraints as shown in (9) but where in some places the outputs Y_1 and Y_2 have to be replaced by the generalized feedback output $\tilde{Y}$.

- Choice 2 (randomized superposition coding):

$$Q = \begin{cases} 1 & \text{w. p. } 1/2 \\ 2 & \text{w. p. } 1/2 \end{cases}, \qquad (64a)$$

$$V_0 = V_1 = V_2 = \begin{cases} Y_2 & \text{if } Q = 1 \\ Y_1 & \text{if } Q = 2 \end{cases}, \qquad (64b)$$

and

$$X = \begin{cases} U_1 & \text{if } Q = 1 \\ U_2 & \text{if } Q = 2 \end{cases} \qquad (64c)$$

for some pmf $P_{U_0 U_1 U_2} = P_{U_0} P_{U_1 | U_0} P_{U_2 | U_0}$.

Choice 1 essentially results in a coded time-sharing scheme, and was shown [23] to be optimal for some deterministic state-dependent DMBCs, e.g., the state-dependent broadcast erasure channels [21], [22] and the finite field deterministic channel [22]. Choice 2 results in a randomized superposition coding scheme and can achieve larger rate regions than Choice 1. For no state-dependent DMBC a better choice for the auxiliaries for the Shayevitz-Wigger region is known.

Since Kim, Chia, and El Gamal's choices of auxiliaries in (63) and (64) satisfy Condition (57), by the discussion in the last paragraph of the previous subsection X-D, our new region $\mathcal{R}_{\text{proc.}}^{\text{state}}$ recovers the sum-rates obtained by Kim-Chia-El Gamal.

F. The Maddah-Ali&Tse DoF region and our regions $\mathcal{R}_{\text{proc.}}^{\text{state}}$ and $\mathcal{R}_{\text{relay,bw}}$

As pointed out in [23], the celebrated Maddah-Ali & Tse scheme [22] for the two-user i.i.d. fading BC with strictly causal state-information is a special case of the Shayevitz-Wigger scheme for generalized feedback, when the scheme is extended to real-valued alphabets and specialized to the choice of auxiliaries in (63). Since their choice also satisfies (61), by the discussion at the end of Section X-D, their DoF result for the two-user i.i.d. fading BC is also included in our new achievable region $\mathcal{R}_{\text{proc.}}^{\text{state}}$. The same holds for the capacity of Maddah-Ali & Tse's deterministic approximation of the i.i.d. fading BC, which is a deterministic state-dependent DMBC.

Various works have improved and extended Maddah-Ali & Tse's DoF result. For example, Yang, Kobayashi, Gesbert, and Yi [37] modified and improved Maddah-Ali & Tse's scheme to apply to the more general setup where the transmitters also obtain imperfect (rate-limited) *causal* state-information. They showed that, under some mild assumptions, their improved scheme achieves the optimal DoF region for arbitrary stationary and ergodic fading processes. In this sense, they could bridge the gap between Maddah-Ali & Tse's setup and a setup where the transmitter learns the state causally (and not only strictly causally).

Chen and Elia [36] proposed a coding scheme for an even more general setup where the transmitter accumulates state-information about each fading sample over time. They were able to determine the optimal DoF region under some mild assumptions.

Both the Yang et al. scheme [37] and the Chen and Elia scheme [36] are related to the Shayevitz-Wigger scheme with a choice of auxiliaries as in (59), and thus also to our new region $\mathcal{R}_{\text{proc.}}^{\text{state}}$. See [30, Chapter 4] for a more detailed discussion.

G. Comparison with noisy network coding

Our coding schemes are reminiscent of the compress-and-forward relay strategy [38] or the noisy network coding for general networks [25], [39] in the sense that the two receivers compress their channel outputs and send the compression indices over the feedback links. The operations at the transmitter are however very different from noisy network coding. On one hand, we use Marton coding to send independent private and common messages to the two receivers. On the other hand, the transmitter either reconstructs the quantized version of the receivers' outputs (our type-II scheme) or it simply relays the compression messages (our type-I schemes). In particular, as explained in Section X-C, when the feedback links are noisy, the transmitter first decodes the compression messages sent over the feedback links and then operates on these compression messages. In noisy network coding the transmitter would compress and forward the observed noisy feedback signals, without attempting to recover the transmitted compression messages.

APPENDIX A
ANALYSIS OF SCHEME IA (THEOREM 1)

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

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

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

- $\hat{J}_{i,b-1} \neq 1$;
- $\hat{K}_{i,b-1} \neq 1$;
- $\hat{M}_{\text{Fb},i,b-1} \neq 1$;
- $\hat{M}_{p,i,b-1} \neq 1$;
- $\hat{M}_{c,b}^{(i)} \neq (1,1)$;
- There is no pair $(k_{1,b}, k_{2,b}) \in \mathcal{K}_1 \times \mathcal{K}_2$ that satisfies

$$\left(U_{0,b}^n(\mathbf{1}_{[4]}), U_{1,b}^n(1, k_{1,b}|\mathbf{1}_{[4]}), U_{2,b}^n(1, k_{2,b}|\mathbf{1}_{[4]}) \right)$$
$$\in \mathcal{T}_{\varepsilon/16}^{(n)}(P_{U_0 U_1 U_2})$$

- $\big(U_{0,b-1}^n(\mathbf{1}_{[4]}), U_{1,b-1}^n(1,1|\mathbf{1}_{[4]}), U_{2,b-1}^n(1,1,|\mathbf{1}_{[4]}), Y_{1,b-1}^n, Y_{2,b-1}^n \big) \notin \mathcal{T}_{\varepsilon/12}^{(n)}(P_{U_0 U_1 U_2 Y_1 Y_2})$

- There is no pair $(m_{\text{Fb},i,b}, j_{i,b}) \in \mathcal{L}_i \times \mathcal{J}_i$ that satisfies

$$\left(\tilde{Y}_{i,b}^n(m_{\text{Fb},i,b}, j_{i,b}|\mathbf{1}_{[4]}), U_{0,b}^n(\mathbf{1}_{[4]}), Y_{i,b}^n \right) \in \mathcal{T}_{\varepsilon/4}^{(n)}(P_{\tilde{Y}_i U_0 Y_i}).$$

Then,

$$P_e^{(N)} \leq \Pr\left[\bigcup_{b=1}^{B+1} \mathcal{E}_b\right] \leq \sum_{b=2}^{B+1} \Pr\left[\mathcal{E}_b|\mathcal{E}_{b-1}^c\right] + \Pr[\mathcal{E}_1]. \quad (65)$$

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 $\Pr[\mathcal{E}_b|\mathcal{E}_{b-1}^c]$ tends to 0 as $n \to \infty$. Similar arguments can be used to show that under the same conditions also $\Pr[\mathcal{E}_1] \to 0$ 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 $\mathcal{E}_{0,b}$ be the event that there is no pair $(k_{1,b}, k_{2,b}) \in \mathcal{K}_1 \times \mathcal{K}_2$ that satisfies

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

By the Covering Lemma [27], $\Pr(\mathcal{E}_{0,b})$ tends to 0 as $n \to \infty$ if

$$R_1' + R_2' \geq I(U_1; U_2|U_0) + \delta(\varepsilon), \quad (66)$$

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

- Let $\mathcal{E}_{1,b}$ be the event that

$$\big(U_{0,b}^n(\mathbf{1}_{[4]}), U_{1,b}^n(1, 1|\mathbf{1}_{[4]}), U_{2,b}^n(1, 1, |\mathbf{1}_{[4]}), Y_{1,b}^n, Y_{2,b}^n\big)$$
$$\notin \mathcal{T}_{\varepsilon/12}^{(n)}(P_{U_0 U_1 U_2 Y_1 Y_2}).$$

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

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

$$\big(U_{0,b}^n(\hat{m}_{c,b}^{(1)}, 1, \hat{m}_{\text{Fb},2,b-1}), Y_{1,b}^n\big) \in \mathcal{T}_{\varepsilon/8}^{(n)}(P_{U_0 Y_1}).$$

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

$$\tilde{R}_2 + R_{c,1} + R_{c,2} \leq I(U_0; Y_1) + \delta(\varepsilon). \quad (67)$$

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

$$\big(U_{0,b}^n(\hat{m}_{c,b}^{(2)}, \hat{m}_{\text{Fb},1,b-1}, 1), Y_{2,b}^n\big) \in \mathcal{T}_{\varepsilon/8}^{(n)}(P_{U_0 Y_2}).$$

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

$$\tilde{R}_1 + R_{c,1} + R_{c,2} \leq I(U_0; Y_2) + \delta(\varepsilon). \quad (68)$$

- Let $\mathcal{E}_{3,1,b}$ be the event that

$$\big(U_{0,b-1}^n(\mathbf{1}_{[4]}), U_{1,b-1}^n(1, 1|\mathbf{1}_{[4]}),$$
$$\tilde{Y}_{2,b-1}^n(1, 1), Y_{1,b-1}^n\big) \notin \mathcal{T}_{\varepsilon/2}^{(n)}(P_{U_0 U_1 \tilde{Y}_2 Y_1}).$$

By the Markov Lemma [27], $\Pr(\mathcal{E}_{3,1,b}|\mathcal{E}_{b-1}^c)$ tends to 0 as $n \to \infty$.

- Let $\mathcal{E}_{3,2,b}$ be the event that

$$\big(U_{0,b-1}^n(\mathbf{1}_{[4]}), U_{2,b-1}^n(1, 1|\mathbf{1}_{[4]}),$$
$$\tilde{Y}_{1,b-1}^n(1, 1), Y_{2,b-1}^n\big) \notin \mathcal{T}_{\varepsilon/2}^{(n)}(P_{U_0 U_2 \tilde{Y}_1 Y_2}).$$

By the Markov Lemma, $\Pr(\mathcal{E}_{3,2,b}|\mathcal{E}_{b-1}^c)$ tends to 0 as $n \to \infty$.

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

$$\big(U_{0,b-1}^n(\mathbf{1}_{[4]}), U_{1,b-1}^n(\hat{m}_{p,1,b-1}, \hat{k}_{1,b-1}|\mathbf{1}_{[4]}),$$
$$\tilde{Y}_{2,b-1}^n(1, \hat{j}_{2,b-1}|\mathbf{1}_{[4]}), Y_{1,b-1}^n\big) \in \mathcal{T}_{\varepsilon}^{(n)}(P_{U_0 U_1 \tilde{Y}_2 Y_1}).$$

By the Packing Lemma, $\Pr(\mathcal{E}_{4,1,b}|\mathcal{E}_{3,1,b}^c)$ tends to zero as $n \to \infty$, if

$$\hat{R}_2 \leq I(\tilde{Y}_2; U_1, Y_1|U_0) - \delta(\varepsilon) \quad (69)$$
$$R_{p,1} + R_1' \leq I(U_1; Y_1, \tilde{Y}_2|U_0) - \delta(\varepsilon) \quad (70)$$
$$R_{p,1} + R_1' + \hat{R}_2 \leq I(U_1; Y_1, \tilde{Y}_2|U_0)$$
$$+ I(\tilde{Y}_2; Y_1|U_0) - \delta(\varepsilon). \quad (71)$$

- Let $\mathcal{E}_{4,2,b}$ 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

$$\big(U_{0,b-1}^n(\mathbf{1}_{[4]}), U_{2,b-1}^n(\hat{m}_{p,2,b-1}, \hat{k}_{2,b-1}|\mathbf{1}_{[4]}),$$
$$\tilde{Y}_{1,b-1}^n(1, \hat{j}_{1,b-1}|\mathbf{1}_{[4]}), Y_{2,b-1}^n\big) \in \mathcal{T}_{\varepsilon}^{(n)}(P_{U_0 U_2 \tilde{Y}_1 Y_2}).$$

By the Packing Lemma, $\Pr(\mathcal{E}_{4,2,b}|\mathcal{E}_{3,2,b}^c)$ tends to zero as $n \to \infty$, if

$$\hat{R}_1 \leq I(\tilde{Y}_1; U_2, Y_2|U_0) - \delta(\varepsilon) \quad (72)$$
$$R_{p,2} + R_2' \leq I(U_2; Y_2, \tilde{Y}_1|U_0) - \delta(\varepsilon) \quad (73)$$
$$R_{p,2} + R_2' + \hat{R}_1 \leq I(U_2; Y_2, \tilde{Y}_1|U_0)$$
$$+ I(\tilde{Y}_1; Y_2|U_0) - \delta(\varepsilon). \quad (74)$$

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

$$\big(\tilde{Y}_{i,b}^n(m_{\text{Fb},i,b}, j_{i,b}|\mathbf{1}_{[4]}), U_{0,b}^n(\mathbf{1}_{[4]}), Y_{i,b}^n\big) \in \mathcal{T}_{\varepsilon/4}^{(n)}(P_{\tilde{Y}_i U_0 Y_i}).$$

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

$$\tilde{R}_i + \hat{R}_i \geq I(\tilde{Y}_i; Y_i|U_0) + \delta(\varepsilon). \quad (75)$$

Whenever the event $\mathcal{E}_{b-1}^c$ occurs but none of the events $\{\mathcal{E}_{0,b}, \mathcal{E}_{1,b}, \mathcal{E}_{2,i,b}, \mathcal{E}_{3,i,b}, \mathcal{E}_{3,i,b}, \mathcal{E}_{4,i,b}, \mathcal{E}_{5,i,b}\}$ above, for $i = 1, 2$, then $\mathcal{E}_b^c$. Therefore,

$$\Pr\big[\mathcal{E}_b|\mathcal{E}_{b-1}^c\big]$$
$$\leq \Pr\Big[\mathcal{E}_{0,b} \cup \mathcal{E}_{1,b} \cup \bigcup_{i=1}^{2}\Big(\mathcal{E}_{2,i,b} \cup \mathcal{E}_{3,i,b} \cup \mathcal{E}_{4,i,b} \cup \mathcal{E}_{5,i,b}\Big)\Big|\mathcal{E}_{b-1}^c\Big]$$
$$\leq \Pr\big[\mathcal{E}_{0,b}|\mathcal{E}_{b-1}^c\big] + \Pr\big[\mathcal{E}_{1,b}|\mathcal{E}_{0,b}^c, \mathcal{E}_{b-1}^c\big]$$
$$+ \sum_{i=1}^{2}\Big(\Pr\big[\mathcal{E}_{2,i,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{b-1}^c\big] + \Pr\big[\mathcal{E}_{3,i,b}|\mathcal{E}_{b-1}^c\big]$$

$$+ \Pr\big[\mathcal{E}_{4,i,b}|\mathcal{E}_{3,i,b}^c,\mathcal{E}_{b-1}^c\big] + \Pr\big[\mathcal{E}_{5,i,b}|\mathcal{E}_{1,b}^c,\mathcal{E}_{b-1}^c\big]\Big)$$

$$= \Pr[\mathcal{E}_{0,b}] + \Pr\big[\mathcal{E}_{1,b}|\mathcal{E}_{0,b}^c\big]$$

$$+ \sum_{i=1}^{2}\Big(\Pr\big[\mathcal{E}_{2,i,b}|\mathcal{E}_{1,b}^c\big] + \Pr\big[\mathcal{E}_{3,i,b}|\mathcal{E}_{b-1}^c\big]$$

$$+ \Pr\big[\mathcal{E}_{4,i,b}|\mathcal{E}_{3,i,b}^c\big] + \Pr\big[\mathcal{E}_{5,i,b}|\mathcal{E}_{1,b}^c\big]\Big).$$

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 (66)–(75) are satisfied. Thus, by (65) and (76) we conclude that the probability of error $P_e^{(N)}$ (averaged over all code constructions) vanishes as $n \to \infty$ if Constraints (66)–(75) hold. Letting $\varepsilon \to 0$, we obtain that the probability of error can be made to tend to 0 as $n \to \infty$ whenever

$$R_1' + R_2' > I(U_1; U_2|U_0) \tag{76a}$$

$$\tilde{R}_2 + R_{c,1} + R_{c,2} < I(U_0; Y_1) \tag{76b}$$

$$\tilde{R}_1 + R_{c,1} + R_{c,2} < I(U_0; Y_2) \tag{76c}$$

$$\hat{R}_1 < I(\tilde{Y}_1; U_2, Y_2|U_0) \tag{76d}$$

$$\hat{R}_2 < I(\tilde{Y}_2; U_1, Y_1|U_0) \tag{76e}$$

$$R_{p,1} + R_1' < I(U_1; Y_1, \tilde{Y}_2|U_0) \tag{76f}$$

$$R_{p,2} + R_2' > I(U_2; Y_2, \tilde{Y}_1|U_0) \tag{76g}$$

$$R_{p,1} + R_1' + \hat{R}_2 < I(U_1; Y_1, \tilde{Y}_2|U_0) + I(\tilde{Y}_2; Y_1|U_0) \tag{76h}$$

$$R_{p,2} + R_2' + \hat{R}_1 < I(U_2; Y_2, \tilde{Y}_1|U_0) + I(\tilde{Y}_1; Y_2|U_0) \tag{76i}$$

$$\hat{R}_1 + \tilde{R}_1 > I(\tilde{Y}_1; Y_1|U_0) \tag{76j}$$

$$\hat{R}_2 + \tilde{R}_2 > I(\tilde{Y}_2; Y_2|U_0). \tag{76k}$$

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

$$\tilde{R}_1 \le R_{\mathrm{Fb},1} \tag{76l}$$

$$\tilde{R}_2 \le R_{\mathrm{Fb},2}. \tag{76m}$$

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

$$\Delta_1 + \Delta_2 - I(U_1; U_2|U_0) \ge 0 \tag{77}$$

Notice that we can ignore Constraint (77) because for any tuple $(U_0, U_1, U_2, X, Y_1, Y_2, \tilde{Y}_1, \tilde{Y}_2)$ that violates (77), the region defined by the constraints in Theorem 1 is contained in the time-sharing region.

APPENDIX B
ANALYSIS OF THE SCHEME IB (THEOREM 2)

An error occurs whenever

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

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

$$\hat{J}_{i,b} \ne J_{i,b} \tag{78}$$

$$\hat{K}_{i,b} \ne K_{i,b} \tag{79}$$

$$\hat{M}_{\mathrm{Fb},i,b-1} \ne M_{\mathrm{Fb},i,b-1} \tag{80}$$

$$\hat{M}_{p,i,b} \ne M_{p,i,b} \tag{81}$$

$$\hat{M}_{c,b}^{(i)} \ne M_{c,b}^{(i)}. \tag{82}$$

Then,

$$P_e^{(N)} \le \Pr\left[\bigcup_{b=1}^{B+1} \mathcal{E}_b\right] \le \sum_{b=1}^{B} \Pr\big[\mathcal{E}_b|\mathcal{E}_{b+1}^c\big] + \Pr[\mathcal{E}_{B+1}]. \tag{83}$$

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 $b \in \{1,\ldots,B\}$, the probability $\Pr\big[\mathcal{E}_b|\mathcal{E}_{b+1}^c\big]$ tends to 0 as $n \to \infty$. Similar arguments can be used to show that under the same conditions also $\Pr[\mathcal{E}_{B+1}] \to 0$ 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 $b \in \{1,\ldots,B\}$ and $\varepsilon > 0$. By the symmetry of our code construction, the probability $\Pr\big[\mathcal{E}_b|\mathcal{E}_{b+1}^c\big]$ does not depend on the realization of $M_{c,i,b}, M_{p,i,b}, K_{i,b}, J_{i,b}, M_{\mathrm{Fb},i,b}, M_{\mathrm{Fb},i,b-1}$, for $i \in \{1,2\}$. To simplify exposition we therefore assume that $M_{c,i,b} = M_{p,i,b} = K_{i,b} = J_{i,b} = M_{\mathrm{Fb},i,b} = M_{\mathrm{Fb},i,b-1} = 1$.

Define the following events.

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

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

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

$$R_1' + R_2' \ge I(U_1; U_2|U_0) + \delta(\varepsilon), \tag{84}$$

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

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

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

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

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

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

$$\tilde{R}_i + \hat{R}_i \ge I(\tilde{Y}_i; Y_i) + \delta(\varepsilon). \tag{85}$$

- Let $\mathcal{E}_{3,1,b}$ be the event that

$$\big(U_{0,b}^n(\mathbf{1}_{[4]}), U_{1,b}^n(1,1|\mathbf{1}_{[4]}),$$
$$\tilde{Y}_{2,b}^n(1,1), Y_{1,b}^n\big) \notin \mathcal{T}_{3\varepsilon/4}^{(n)}(P_{U_0 U_1 \tilde{Y}_2 Y_1}).$$

By the Markov Lemma, $\Pr(\mathcal{E}_{3,1,b}|\mathcal{E}_{2,2,b}^c, \mathcal{E}_{1,b}^c)$ tends to 0 as $n \to \infty$.

- Let $\mathcal{E}_{3,2,b}$ be the event that

$$\big(U_{0,b}^n(\mathbf{1}_{[4]}), U_{2,b}^n(1,1|\mathbf{1}_{[4]}),$$

$$\tilde{Y}_{1,b}^n(1,1), Y_{2,b}^n) \notin \mathcal{T}_{3\varepsilon/4}^{(n)}(P_{U_0 U_2 \tilde{Y}_1 Y_2}).$$

By the Markov Lemma, $\Pr(\mathcal{E}_{3,2,b}|\mathcal{E}_{2,1,b}^c, \mathcal{E}_{1,b}^c)$ tends to 0 as $n \to \infty$.

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

$$\Big(U_{0,b}^n(\hat{m}_{c,b}^{(1)}, 1, \hat{m}_{\text{Fb},2,b-1}),$$
$$U_{1,b}^n(\hat{m}_{p,1,b}, \hat{k}_{1,b}|\hat{m}_{c,b}^{(1)}, 1, \hat{m}_{\text{Fb},2,b-1}),$$
$$\tilde{Y}_{2,b}^n(1, \hat{j}_{2,b}), Y_{1,b}^n \Big) \in \mathcal{T}_\varepsilon^{(n)}(P_{U_0 U_1 \tilde{Y}_2 Y_1}).$$

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

$$\hat{R}_2 \le I(U_0, U_1, Y_1; \tilde{Y}_2|U_0) - \delta(\varepsilon)$$
$$R_{p,1} + R_1' \le I(U_1; Y_1, \tilde{Y}_2|U_0) - \delta(\varepsilon)$$
$$R_1 + R_{c,2} + \tilde{R}_2 + R_1' \le I(U_0, U_1; Y_1, \tilde{Y}_2) - \delta(\varepsilon)$$
$$R_1 + R_{c,2} + \tilde{R}_2 + R_1' + \hat{R}_2 \le I(U_0, U_1; Y_1, \tilde{Y}_2)$$
$$+ I(Y_1; \tilde{Y}_2) - \delta(\varepsilon)$$
$$R_{p,1} + R_1' + \hat{R}_2 \le I(U_1; Y_1, \tilde{Y}_2|U_0)$$
$$+ I(\tilde{Y}_2; Y_1, U_0) - \delta(\varepsilon). \quad (86)$$

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

$$\Big(U_{0,b}^n(\hat{m}_{c,b}^{(2)}, \hat{m}_{\text{Fb},1,b-1}, 1),$$
$$U_{1,b}^n(\hat{m}_{p,2,b}, \hat{k}_{2,b}|\hat{m}_{c,b}^{(2)}, \hat{m}_{\text{Fb},1,b-1}, 1),$$
$$\tilde{Y}_{1,b}^n(1, \hat{j}_{1,b}), Y_{2,b}^n \Big) \in \mathcal{T}_\varepsilon^{(n)}(P_{U_0 U_2 \tilde{Y}_1 Y_2}).$$

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

$$\hat{R}_1 \le I(U_0, U_2, Y_2; \tilde{Y}_1|U_0) - \delta(\varepsilon)$$
$$R_{p,2} + R_2' \le I(U_2; Y_2, \tilde{Y}_1|U_0) - \delta(\varepsilon)$$
$$R_2 + R_{c,1} + \tilde{R}_1 + R_2' \le I(U_0, U_2; Y_2, \tilde{Y}_1) - \delta(\varepsilon)$$
$$R_2 + R_{c,1} + \tilde{R}_1 + R_2' + \hat{R}_1 \le I(U_0, U_2; Y_2, \tilde{Y}_1)$$
$$+ I(Y_2; \tilde{Y}_1) - \delta(\varepsilon)$$
$$R_{p,2} + R_2' + \hat{R}_1 \le I(U_2; Y_2, \tilde{Y}_1|U_0)$$
$$+ I(\tilde{Y}_1; Y_2, U_0) - \delta(\varepsilon). \quad (87)$$

Whenever the event $\mathcal{E}_{b+1}^c$ occurs but none of the events above, then $\mathcal{E}_b^c$. Therefore,

$$\Pr\big[\mathcal{E}_b|\mathcal{E}_{b+1}^c\big]$$
$$\le \Pr\Big[\mathcal{E}_{0,b} \cup \mathcal{E}_{1,b} \cup \bigcup_{i=1}^2 \Big(\mathcal{E}_{2,i,b} \cup \mathcal{E}_{3,i,b} \cup \mathcal{E}_{4,i,b}\Big)\Big|\mathcal{E}_{b+1}^c\Big]$$
$$\le \Pr\big[\mathcal{E}_{0,b}|\mathcal{E}_{b+1}^c\big] + \Pr\big[\mathcal{E}_{3,1,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{2,2,b}^c, \mathcal{E}_{b+1}^c\big]$$
$$+ \Pr\big[\mathcal{E}_{1,b}|\mathcal{E}_{0,b}^c, \mathcal{E}_{b+1}^c\big] + \Pr\big[\mathcal{E}_{3,2,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{2,1,b}^c, \mathcal{E}_{b+1}^c\big]$$
$$+ \sum_{i=1}^2 \Big(\Pr\big[\mathcal{E}_{2,i,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{b+1}^c\big] + \Pr\big[\mathcal{E}_{4,i,b}|\mathcal{E}_{3,i,b}^c, \mathcal{E}_{b+1}^c\big] \Big)$$
$$= \Pr\big[\mathcal{E}_{0,b}\big] + \Pr\big[\mathcal{E}_{1,b}|\mathcal{E}_{0,b}^c\big]$$

$$+ \Pr\big[\mathcal{E}_{3,1,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{2,2,b}^c\big] + \Pr\big[\mathcal{E}_{3,2,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{2,1,b}^c\big]$$
$$+ \sum_{i=1}^2 \Big(\Pr\big[\mathcal{E}_{2,i,b}|\mathcal{E}_{1,b}^c\big] + \Pr\big[\mathcal{E}_{4,i,b}|\mathcal{E}_{3,i,b}^c\big] \Big), \quad (88)$$

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 (84)–(87) are satisfied. Thus, by (83) and (88) we conclude that the probability of error $P_e^{(N)}$ (averaged over all code constructions) vanishes as $n \to \infty$ if constraints (84)–(87) hold. Letting $\varepsilon \to 0$, we obtain that the probability of error can be made to tend to 0 as $n \to \infty$ whenever

$$R_1' + R_2' > I(U_1; U_2|U_0) \quad (89\text{a})$$
$$\hat{R}_1 + \tilde{R}_1 > I(\tilde{Y}_1; Y_1) \quad (89\text{b})$$
$$\hat{R}_2 + \tilde{R}_2 > I(\tilde{Y}_2; Y_2) \quad (89\text{c})$$
$$\hat{R}_1 < I(U_0, U_2, Y_2; \tilde{Y}_1|U_0) \quad (89\text{d})$$
$$\hat{R}_2 < I(U_0, U_1, Y_1; \tilde{Y}_2|U_0) \quad (89\text{e})$$
$$R_{p,1} + R_1' < I(U_1; Y_1, \tilde{Y}_2|U_0) \quad (89\text{f})$$
$$R_{p,2} + R_2' < I(U_2; Y_2, \tilde{Y}_1|U_0) \quad (89\text{g})$$
$$R_1 + R_{c,2} + \tilde{R}_2 + R_1' < I(U_0, U_1; Y_1, \tilde{Y}_2) \quad (89\text{h})$$
$$R_2 + R_{c,1} + \tilde{R}_1 + R_2' < I(U_0, U_2; Y_2, \tilde{Y}_1) \quad (89\text{i})$$
$$R_1 + R_{c,2} + \tilde{R}_2 + R_1' + \hat{R}_2 < I(U_0, U_1; Y_1, \tilde{Y}_2)$$
$$+ I(Y_1; \tilde{Y}_2) \quad (89\text{j})$$
$$R_2 + R_{c,1} + \tilde{R}_1 + R_2' + \hat{R}_1 < I(U_0, U_2; Y_2, \tilde{Y}_1)$$
$$+ I(Y_2; \tilde{Y}_1) \quad (89\text{k})$$
$$R_{p,1} + R_1' + \hat{R}_2 < I(U_1; Y_1, \tilde{Y}_2|U_0)$$
$$+ I(\tilde{Y}_2; Y_1, U_0) \quad (89\text{l})$$
$$R_{p,2} + R_2' + \hat{R}_1 < I(U_2; Y_2, \tilde{Y}_1|U_0)$$
$$+ I(\tilde{Y}_1; Y_2, U_0). \quad (89\text{m})$$

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

$$\tilde{R}_1 \le R_{\text{Fb},1} \quad (89\text{n})$$
$$\tilde{R}_2 \le R_{\text{Fb},2}. \quad (89\text{o})$$

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

$$\Delta_1 \ge 0 \quad (90\text{a})$$
$$\Delta_2 \ge 0 \quad (90\text{b})$$
$$\Delta_1 + \Delta_2 \ge I(U_1; U_2|U_0). \quad (90\text{c})$$

We can ignore Constraint (90c) because for any tuple $(U_0, U_1, U_2, X, Y_1, Y_2, \tilde{Y}_1, \tilde{Y}_2)$ that violates (90c), the region defined by the constraints in Theorem 2 is contained in the time-sharing region. Constraint (90a) 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 (90a), the region defined by the constraints in Theorem 2 is contained in the region in Theorem 2 for the choice $\tilde{Y}_2 = \text{const.}$, for which (90a) is always satisfied. Constraint (90b) can be ignored by analogous arguments.

APPENDIX C
ANALYSIS OF SCHEME II (THEOREM 4)

An error occurs whenever

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

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

$$\hat{J}_{i,b} \neq J_{i,b} \tag{91}$$
$$\hat{K}_{i,b} \neq K_{i,b} \tag{92}$$
$$\hat{M}_{\text{Fb},i,b} \neq M_{\text{Fb},i,b} \tag{93}$$
$$\hat{M}_{p,i,b} \neq M_{p,i,b} \tag{94}$$
$$\hat{M}_{c,b}^{(i)} \neq M_{c,b}^{(i)} \tag{95}$$

or when

$$\hat{N}_{b-1} \neq N_{b-1}. \tag{96}$$

Then,

$$P_e^{(n)} \leq \Pr\left[\bigcup_{b=1}^{B+1} \mathcal{E}_b\right] \leq \sum_{b=1}^{B} \Pr\left[\mathcal{E}_b|\mathcal{E}_{b+1}^c\right] + \Pr[\mathcal{E}_{B+1}]. \tag{97}$$

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 $b \in \{1,\ldots,B\}$, the probability $\Pr[\mathcal{E}_b|\mathcal{E}_{b+1}^c]$ tends to 0 as $n \to \infty$. Similar arguments can be used to show that under the same conditions also $\Pr[\mathcal{E}_{B+1}] \to 0$ 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 $b \in \{1,\ldots,B\}$ and $\varepsilon > 0$. By the symmetry of our code construction, the probability $\Pr[\mathcal{E}_b|\mathcal{E}_{b+1}^c]$ does not depend on the realizations of N_{b-1}, N_b, or $M_{c,i,b}$, $M_{p,i,b}$, $K_{i,b}$, $J_{i,b}$, $M_{\text{Fb},i,b}$, for $i \in \{1,2\}$. To simplify exposition we therefore assume that for $i \in \{1,2\}$, $M_{c,i,b} = M_{p,i,b} = K_{i,b} = J_{i,b} = M_{\text{Fb},i,b} = 1$, and $N_b = N_{b-1} = 1$.

Define the following events.

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

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

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

$$R_1' + R_2' \geq I(U_1; U_2|U_0) + \delta(\varepsilon), \tag{98}$$

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

- Let $\mathcal{E}_{1,b}$ be the event that

$$\left(U_{0,b}^n(\mathbf{1}_{[3]}), U_{1,b}^n(1,1|\mathbf{1}_{[3]}), U_{2,b}^n(1,1,|\mathbf{1}_{[3]}), Y_{1,b}^n, Y_{2,b}^n\right)$$
$$\notin \mathcal{T}_{\varepsilon/32}^{(n)}(P_{U_0U_1U_2Y_1Y_2}).$$

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

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

$$\left(\tilde{Y}_{i,b}^n(m_{\text{Fb},i,b}, j_{i,b}), Y_{i,b}^n\right) \in \mathcal{T}_{\varepsilon/16}^{(n)}(P_{\tilde{Y}_iY_i}).$$

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

$$\tilde{R}_i + \hat{R}_i \geq I(\tilde{Y}_i; Y_i) + \delta(\mathcal{E}). \tag{99}$$

- Let $\mathcal{E}_{3,b}$ be the event that

$$\left(U_{0,b}^n(\mathbf{1}_{[3]}), U_{1,b}^n(1,1|\mathbf{1}_{[3]}), U_{2,b}^n(1,1|\mathbf{1}_{[3]}),\right.$$
$$\tilde{Y}_{1,b}^n(1,1), \tilde{Y}_{2,b}^n(1,1), Y_{1,b}^n, Y_{2,b}^n\Big)$$
$$\notin \mathcal{T}_{\varepsilon/6}^{(n)}(P_{U_0U_1U_2\tilde{Y}_1\tilde{Y}_2Y_1Y_2}).$$

By the Markov Lemma, $\Pr(\mathcal{E}_{3,b}|\mathcal{E}_{2,1,b}^c, \mathcal{E}_{2,2,b}^c, \mathcal{E}_{1,b}^c)$ tends to 0 as $n \to \infty$.

- Let $\mathcal{E}_{4,b}$ 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

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

By the Packing Lemma, $\Pr(\mathcal{E}_{4,b}|\mathcal{E}_{3,b}^c)$ tends to 0 as $n \to \infty$, if

$$\hat{R}_1 \leq I(U_0, U_1, U_2, \tilde{Y}_2; \tilde{Y}_1) - \delta(\varepsilon) \tag{100}$$
$$\hat{R}_2 \leq I(U_0, U_1, U_2, \tilde{Y}_1; \tilde{Y}_2) - \delta(\varepsilon) \tag{101}$$
$$\hat{R}_1 + \hat{R}_2 \leq I(U_0, U_1, U_2; \tilde{Y}_1, \tilde{Y}_2) + I(\tilde{Y}_1; \tilde{Y}_2) - \delta(\varepsilon). \tag{102}$$

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

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

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

$$\tilde{R}_v \geq I(U_0, U_1, U_2, \tilde{Y}_1, \tilde{Y}_2; V) + \delta(\varepsilon). \tag{103}$$

- Let $\mathcal{E}_{6,1,b}$ be the event that

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

By the Markov Lemma $\Pr(\mathcal{E}_{6,1,b}|\mathcal{E}_{3,b}^c, \mathcal{E}_{5,b}^c)$ tends to zero as $n \to \infty$.

- Let $\mathcal{E}_{6,2,b}$ be the event that

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

By the Markov Lemma $\Pr(\mathcal{E}_{6,2,b}|\mathcal{E}_{3,b}^c, \mathcal{E}_{5,b}^c)$ tends to zero as $n \to \infty$.

- Let $\mathcal{E}_{7,1,b}$ be the event that there is a tuple $(\hat{m}_{c,b}^{(1)}, \hat{n}_{b-1}, \hat{m}_{p,1,b}, \hat{k}_{1,b}) \in \mathcal{M}_c \times \mathcal{N} \times \mathcal{M}_{p,1} \times \mathcal{K}_1$ that

is not equal to the all-one tuple $(\mathbf{1}_{[3]}, 1, 1, 1)$ and that satisfies

$$\Big(U_{0,b}^n(\hat{m}_{c,b}^{(1)}, \hat{n}_{b-1}), U_{1,b}^n(\hat{m}_{p,1,b}, \hat{k}_{1,b}|\hat{m}_{c,b}^{(1)}, \hat{n}_{b-1}),$$

$$V_b^n(1|\hat{n}_{b-1}), Y_{1,b}^n, \tilde{Y}_{1,b}^n(1,1)\Big)$$

$$\in \mathcal{T}_\varepsilon^{(n)}(P_{U_0 U_1 V Y_1 \tilde{Y}_1}).$$

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

$$R_1 + R_{c,2} + R_1' \le I(U_0, U_1; Y_1, \tilde{Y}_1, V) - \delta(\varepsilon) \quad (104)$$

$$R_1 + R_{c,2} + \tilde{R}_v + R_1' \le I(U_0, U_1; Y_1, \tilde{Y}_1, V)$$
$$+ I(V; \tilde{Y}_1, Y_1) - \delta(\varepsilon) \quad (105)$$

$$R_{p,1} + R_1' \le I(U_1; Y_1, \tilde{Y}_1, V|U_0) - \delta(\varepsilon). \quad (106)$$

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

$$\Big(U_{0,b}^n(\hat{m}_{c,b}^{(2)}, \hat{n}_{b-1}), U_{2,b}^n(\hat{m}_{p,2,b}, \hat{k}_{2,b}|\hat{m}_{c,b}^{(2)}, \hat{n}_{b-1}),$$

$$V_b^n(1|\hat{n}_{b-1}), Y_{2,b}^n, \tilde{Y}_{2,b}^n(1,1)\Big)$$

$$\in \mathcal{T}_\varepsilon^{(n)}(P_{U_0 U_2 V Y_2 \tilde{Y}_2}).$$

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

$$R_2 + R_{c,1} + R_2' \le I(U_0, U_2; Y_2, \tilde{Y}_2, V) - \delta(\varepsilon) \quad (107)$$

$$R_2 + R_{c,1} + \tilde{R}_v + R_2' \le I(U_0, U_2; Y_2, \tilde{Y}_2, V)$$
$$+ I(V; \tilde{Y}_2, Y_2) - \delta(\varepsilon) \quad (108)$$

$$R_{p,2} + R_2' \le I(U_2; Y_2, \tilde{Y}_2, V|U_0) - \delta(\varepsilon). \quad (109)$$

Whenever the event $\mathcal{E}_{b+1}^c$ occurs but none of the events above, then $\mathcal{E}_b^c$. Therefore,

$$\Pr\big[\mathcal{E}_b|\mathcal{E}_{b+1}^c\big]$$

$$\le \Pr\Big[\mathcal{E}_{0,b} \cup \mathcal{E}_{1,b} \cup \mathcal{E}_{2,1,b} \cup \mathcal{E}_{2,2,b} \cup \mathcal{E}_{3,b}$$

$$\cup \mathcal{E}_{4,b} \cup \mathcal{E}_{5,b} \cup \mathcal{E}_{6,1,b} \cup \mathcal{E}_{6,2,b}\big|\mathcal{E}_{b+1}^c\Big]$$

$$\le \Pr\big[\mathcal{E}_{0,b}\big|\mathcal{E}_{b+1}^c\big] + \Pr\big[\mathcal{E}_{1,b}|\mathcal{E}_{0,b}^c, \mathcal{E}_{b+1}^c\big]$$

$$+ \sum_{i=1}^2 \Pr\big[\mathcal{E}_{2,i,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{b+1}^c\big]$$

$$+ \Pr\big[\mathcal{E}_{3,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{2,1,b}^c, \mathcal{E}_{2,2,b}^c, \mathcal{E}_{b+1}^c\big] + \Pr\big[\mathcal{E}_{4,b}|\mathcal{E}_{3,b}^c, \mathcal{E}_{b+1}^c\big]$$

$$+ \Pr\big[\mathcal{E}_{5,b}|\mathcal{E}_{3,b}^c, \mathcal{E}_{b+1}^c\big] + \sum_{i=1}^2 \Pr\big[\mathcal{E}_{6,i,b}|\mathcal{E}_{3,b}^c, \mathcal{E}_{b+1}^c\big]$$

$$= \Pr\big[\mathcal{E}_{0,b}\big] + \Pr\big[\mathcal{E}_{1,b}|\mathcal{E}_{0,b}^c\big] + \sum_{i=1}^2 \Pr\big[\mathcal{E}_{2,i,b}|\mathcal{E}_{1,b}^c\big]$$

$$+ \Pr\big[\mathcal{E}_{3,b}|\mathcal{E}_{1,b}^c, \mathcal{E}_{2,1,b}^c, \mathcal{E}_{2,2,b}^c\big] + \Pr\big[\mathcal{E}_{4,b}|\mathcal{E}_{3,b}^c\big]$$

$$+ \Pr\big[\mathcal{E}_{5,b}|\mathcal{E}_{3,b}^c\big] + \sum_{i=1}^2 \Pr\big[\mathcal{E}_{6,i,b}|\mathcal{E}_{3,b}^c\big], \quad (110)$$

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 (98)–(109) are satisfied. Thus, by (97) and (110) we conclude that the probability of error $P_e^{(N)}$ (averaged over all code constructions) vanishes as $n \to \infty$ if Constraints (98)–(109) hold. Letting $\varepsilon \to 0$, we obtain that the probability of error can be made to tend to 0 as $n \to \infty$ whenever

$$R_1' + R_2' > I(U_1; U_2|U_0) \quad (111a)$$

$$\hat{R}_1 + \tilde{R}_1 > I(\tilde{Y}_1; Y_1) \quad (111b)$$

$$\hat{R}_2 + \tilde{R}_2 > I(\tilde{Y}_2; Y_2) \quad (111c)$$

$$\hat{R}_1 < I(U_0, U_1, U_2, \tilde{Y}_2; \tilde{Y}_1) \quad (111d)$$

$$\hat{R}_2 < I(U_0, U_1, U_2, \tilde{Y}_1; \tilde{Y}_2) \quad (111e)$$

$$\hat{R}_1 + \hat{R}_2 < I(U_0, U_1, U_2; \tilde{Y}_1, \tilde{Y}_2)$$
$$+ I(\tilde{Y}_1; \tilde{Y}_2) \quad (111f)$$

$$\tilde{R}_v > I(U_0, U_1, U_2, \tilde{Y}_1, \tilde{Y}_2; V) \quad (111g)$$

$$R_1 + R_{c,2} + \tilde{R}_v + R_1' < I(U_0, U_1; Y_1, \tilde{Y}_1, V)$$
$$+ I(V; \tilde{Y}_1, Y_1) \quad (111h)$$

$$R_1 + R_{c,2} + R_1' < I(U_0, U_1; Y_1, \tilde{Y}_1, V) \quad (111i)$$

$$R_{c,1} + R_2 + R_2' < I(U_0, U_2; Y_2, \tilde{Y}_2, V) \quad (111j)$$

$$R_{c,1} + R_2 + \tilde{R}_v + R_2' < I(U_0, U_2; Y_2, \tilde{Y}_2, V)$$
$$+ I(V; \tilde{Y}_2, Y_2) \quad (111k)$$

$$R_{p,1} + R_1' < I(U_1; Y_1, \tilde{Y}_1, V|U_0) \quad (111l)$$

$$R_{p,2} + R_2' < I(U_2; Y_2, \tilde{Y}_2, V|U_0). \quad (111m)$$

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

$$\tilde{R}_1 \le R_{\text{Fb},1} \quad (111n)$$

$$\tilde{R}_2 \le R_{\text{Fb},2}. \quad (111o)$$

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:

$$R_1' + R_2' > I(U_1; U_2|U_0) \quad (112a)$$

$$R_1 + R_{c,2} + R_1' < I(U_0, U_1; Y_1, \tilde{Y}_1, V)$$
$$- I(V; U_0, U_1, U_2, \tilde{Y}_2|\tilde{Y}_1, Y_1) \quad (112b)$$

$$R_{c,1} + R_2 + R_2' < I(U_0, U_2; Y_2, \tilde{Y}_2, V)$$
$$- I(V; U_0, U_1, U_2, \tilde{Y}_1|\tilde{Y}_2, Y_2) \quad (112c)$$

$$R_{p,1} + R_1' < I(U_1; Y_1, \tilde{Y}_1, V|U_0) \quad (112d)$$

$$R_{p,2} + R_2' < (U_2; Y_2, \tilde{Y}_2, V|U_0) \quad (112e)$$

where the feedback-rate constraints have to satisfy

$$I(Y_1; \tilde{Y}_1|U_0, U_1, U_2, \tilde{Y}_2) \le R_{\text{Fb},1} \quad (113a)$$

$$I(Y_2; \tilde{Y}_2|U_0, U_1, U_2, \tilde{Y}_1) \le R_{\text{Fb},2} \quad (113b)$$

$$I(Y_1, Y_2; \tilde{Y}_1, \tilde{Y}_2|U_0, U_1, U_2) \le R_{\text{Fb},1} + R_{\text{Fb},2}. \quad (113c)$$

Applying again the Fourier-Motzkin elimination algorithm to Constraints (112) and keeping Constraints (113), we obtain the desired result in Theorem 4 with the additional constraint that

$$I(U_1; U_2|U_0) \le I(U_1; Y_1, \tilde{Y}_1, V|U_0) + I(U_2; Y_2, \tilde{Y}_2, V|U_0). \quad (114)$$

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 (114), 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,

$$R_1^{(\text{M})} \leq I(U_0^{(\text{M})}, U_1^{(\text{M})}; Y_1^{(\text{M})}) \tag{115a}$$

$$R_2^{(\text{M})} < I(U_0^{(\text{M})}, U_2^{(\text{M})}; Y_2^{(\text{M})}) \tag{115b}$$

$$R_1^{(\text{M})} + R_2^{(\text{M})} \leq I(U_0^{(\text{M})}, U_1^{(\text{M})}; Y_1^{(\text{M})}) + I(U_2^{(\text{M})}; Y_2^{(\text{M})} | U_0^{(\text{M})})$$
$$- I(U_1^{(\text{M})}; U_2^{(\text{M})} | U_0^{(\text{M})}), \tag{115c}$$

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

$$R_1^{(\text{Enh})} \leq I(U_0^{(\text{Enh})}; Y_1^{(\text{Enh})}) \tag{116a}$$

$$R_2^{(\text{Enh})} \leq I(X^{(\text{Enh})}; Y_1^{(\text{Enh})}, Y_2^{(\text{Enh})} | U_0^{(\text{Enh})}), \tag{116b}$$

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

$$P_Q(q) = \begin{cases} \gamma, & q = \text{Enh} \\ 1 - \gamma, & q = \text{M}. \end{cases} \tag{117}$$

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

$$U_0 := U_0^{(Q)},\ U_1 := U_1^{(Q)},\ U_2 := U_2^{(Q)}$$
$$X := X^{(Q)},\ \text{and}\ \tilde{Y}_1 := \tilde{Y}_1^{(Q)} \tag{118}$$

satisfy the feedback rate constraints (26) and the rate pair (R_1', R_2'),

$$R_1' := (1 - \gamma) R_1^{(\text{M})} + \gamma R_1^{(\text{Enh})} \tag{119a}$$

$$R_2' := (1 - \gamma) R_2^{(\text{M})} + \gamma R_2^{(\text{Enh})}, \tag{119b}$$

satisfies the constraints in (25) for the choice in (118). 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

$$P_Q P_{U_0 U_1 U_2 | Q} P_{X | U_0 U_1 U_2 Q} P_{Y_1 Y_2 | X} P_{\tilde{Y}_1 | Y_1 Q}. \tag{120}$$

where $P_{Y_1 Y_2 | X}$ denotes the channel law.

For the described choice of random variables (118), the feedback-rate constraint (26) specializes to

$$\gamma H(Y_1^{(\text{Enh})} | Y_2^{(\text{Enh})}, X^{(\text{Enh})}) \leq R_{\text{Fb},1}, \tag{121}$$

which is satisfied for all sufficiently small $\gamma \in (0, 1)$. Moreover, for this choice the constraints in (25) specialize to

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

$$+ \gamma I(U_0^{(\text{Enh})}; Y_1^{(\text{Enh})}) \tag{122a}$$

$$R_2 \leq (1 - \gamma) I(U_0^{(\text{M})}, U_2^{(\text{M})}; Y_2^{(\text{M})})$$
$$+ \gamma \big(I(X^{(\text{Enh})}; Y_1^{(\text{Enh})}, Y_2^{(\text{Enh})})$$
$$- H(Y_1^{(\text{Enh})} | Y_2^{(\text{Enh})}) \big) \tag{122b}$$

$$R_1 + R_2 \leq (1 - \gamma) \big(I(U_0^{(\text{M})}, U_1^{(\text{M})}; Y_1^{(\text{M})})$$
$$+ I(U_2^{(\text{M})}; Y_2^{(\text{M})} | U_0^{(\text{M})})$$
$$- I(U_1^{(\text{M})}; U_2^{(\text{M})} | U_0^{(\text{M})}) \big)$$
$$+ \gamma \big(I(X^{(\text{Enh})}; Y_1^{(\text{Enh})}, Y_2^{(\text{Enh})} | U_0^{(\text{Enh})})$$
$$+ I(U_0^{(\text{Enh})}; Y_1^{(\text{Enh})}) \big) \tag{122c}$$

$$R_1 + R_2 \leq (1 - \gamma) \big(I(U_1^{(\text{M})}; Y_1^{(\text{M})} | U_0^{(\text{M})})$$
$$+ I(U_0^{(\text{M})}, U_2^{(\text{M})}; Y_2^{(\text{M})})$$
$$- I(U_1^{(\text{M})}; U_2^{(\text{M})} | U_0^{(\text{M})}) \big)$$
$$+ \gamma \big(I(X^{(\text{Enh})}; Y_1^{(\text{Enh})}, Y_2^{(\text{Enh})})$$
$$- H(Y_1^{(\text{Enh})} | Y_2^{(\text{Enh})}) \big). \tag{122d}$$

We argue in the following that the rate pair $(R_1 = R_1', R_2 = R_2')$ defined in (119) satisfies these constraints for all sufficiently small $\gamma > 0$. Comparing (115a), (116a), and (119a), we see that the first constraint (122a) is satisfied for any choice of $\gamma \in [0, 1]$. Similarly, comparing (115c), (116a), (116b), and (119a) and (119b), we note that also the third constraint (122c) is satisfied for any $\gamma \in [0, 1]$. The second constraint (122b) is satisfied when γ is sufficiently small. This can be seen by comparing (115b), (116b), and (119b), and because Constraint (115b) holds with strict inequality. The last constraint (122d) is not active in view of Constraint (122c) whenever

$$\gamma H(Y_1^{(\text{Enh})} | Y_2^{(\text{Enh})}) \leq (1 - \gamma) \Gamma^{(\text{M})}, \tag{123}$$

where $\Gamma^{(\text{M})}$ is defined in (33). Thus, also this last constraint is satisfied when γ is sufficiently small. This concludes our proof.

APPENDIX E
PROOF OF PROPOSITION 1

Let $\mathcal{E}_{\text{Fb},i,b}$, 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,

$$\Pr[\hat{M}_1 \neq M_1 \text{ or } \hat{M}_2 \neq M_2]$$
$$\leq \Pr\left[\varepsilon \cup \left(\bigcup_{b=1}^{B} \mathcal{E}_{\text{Fb},1,b} \right) \cup \left(\bigcup_{b=1}^{B} \mathcal{E}_{\text{Fb},2,b} \right) \right]$$
$$\leq \Pr\left[\varepsilon \Big| \left(\bigcup_{b=1}^{B} \mathcal{E}_{\text{Fb},1,b} \right)^c \cap \left(\bigcup_{b=1}^{B} \mathcal{E}_{\text{Fb},2,b} \right)^c \right]$$
$$+ \Pr\left[\bigcup_{b=1}^{B} \mathcal{E}_{\text{Fb},1,b} \right] + \Pr\left[\bigcup_{b=1}^{B} \mathcal{E}_{\text{Fb},2,b} \right]$$
$$\leq \Pr\left[\varepsilon \Big| \left(\bigcup_{b=1}^{B} \mathcal{E}_{\text{Fb},1,b} \right)^c \cap \left(\bigcup_{b=1}^{B} \mathcal{E}_{\text{Fb},2,b} \right)^c \right]$$

$$+\sum_{b=1}^{B}\Pr[\mathcal{E}_{\text{Fb},1,b}]+\Pr[\mathcal{E}_{\text{Fb},2,b}]. \tag{124}$$

Since we use capacity-achieving codes on the feedback links, the probabilities $\Pr[\mathcal{E}_{\text{Fb},1,b}]$ and $\Pr[\mathcal{E}_{\text{Fb},2,b}]$ 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 $\Pr\left[\mathcal{E}\Big|\left(\bigcup_{b=1}^{B}\mathcal{E}_{\text{Fb},1,b}\right)^{c}\cap\left(\bigcup_{b=1}^{B}\mathcal{E}_{\text{Fb},2,b}\right)^{c}\right]$ vanishes as the blocklength increases.

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

[5] 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.

[6] 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.

[7] 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.* , Monticello, IL, Sept. 2005.

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

[9] 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.

[10] S. Belhadj Amor, Y. Steinberg, and M. Wigger, "MAC-BC duality with linear-feedback schemes," *IEEE Trans. on Inf. Th.,* vol. 61, no. 11, pp. 5976–5998, Nov. 2015.

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

[12] 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.

[13] A. Ben-Yishai and O. Shayevitz, "The AWGN BC with MAC feedback: a reduction to noiseless feedback via interaction," in *Proc. of IEEE ITW 2015,* April 26–May 1 2015, Jerusalem, Israel, April 26–May 1 2015.

[14] S. Pillai Bhaskaran and V. Prabhakaran, "On the noisy feedback capacity of Gaussian broadcast channels," in *Proc. of IEEE ITW 2015,* Jerusalem, Israel, April 26–May 1, 2015.

[15] Y. Wu, P. Minero, and M. Wigger, "Insufficiency of linear-feedback schemes in Gaussian broadcast channels with common message," *IEEE Trans. on Inf. Th.,* vol. 60, no. 8, pp. 4553–4566, Aug. 2014.

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

[17] 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, Sep. 2010.

[18] 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.

[19] A. Bracher and M. Wigger, "Feedback and partial message side-Information on the semideterministic broadcast channel," submitted to *IEEE Trans. on Inf. Th.,* Aug. 2015. Available: http://arxiv.org/abs/1508.01880

[20] C. C. Wang, "Capacity of 1-to-k broadcast erasure channel with output feedback," in *Proc. 48th Ann. Allerton Conf.* , Monticello, IL, Sep. 2010.

[21] 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.

[22] 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.

[23] H. Kim, Y.-K. Chia, and A. El Gamal, "A note on broadcast channels with stale state information at the transmitter," *IEEE Trans. on Inf. Th.,* vol. 61, no. 7, pp. 3622–3631, July 2015.

[24] Y. Wu, "Coding schemes for discrete memoryless multicast networks with and without Feedback," in *Proc. 53th Ann. Allerton Conf.* , Monticello, IL, Sept. 2015.

[25] 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.

[26] S. Lim, K. T. Kim, and Y-H Kim, "Distributed decode-forward for multicast, *Proc. of IEEE ISIT, 2014,* Honolulu, HI, June 29–July 4, Nov. 2014.

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

[28] 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.

[29] Y. Liang, G. Kramer, and H. V. Poor, "Equivalence of two inner bounds on the capacity region of the broadcast channel," in *Proc. 46th Ann. Allerton Conf.* , Monticello, IL, Sept. 2008.

[30] Y. Wu, *New ways of exploiting feedback in memoryless broadcast channels,* PhD thesis, Telecom Paristech, 2014.

[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] S. I. Gel'fand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," *Probl. Inf. Transmiss.* vol. 16, no. 1, pp. 17–25, Jan.–Mar. 1980.

[36] 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.

[37] 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.

[38] 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.

[39] J. Hou and G. Kramer, "Short message noisy network coding with a decode-forward option," *IEEE Trans. on Inf. Th.,* vol. 62, no. 1, pp. 89–107, Jan. 2016

Youlong WU (S'13-M'15) Youlong Wu obtained his B.S. degree in electrical engineering from Wuhan University, Wuhan, China, in 2007. He received the M.S. degree in electrical engineering from Shanghai Jiaotong University, Shanghai, China, in 2011. In 2014, he received the Ph.D. degree at Telecom ParisTech, in Paris, France. Since December 2014, he has been a postdoc at the Institute for Communication Engineering, Technische Universität München (TUM), Munich, Germany. His research interests include information theory and wireless communication.

Michèle Wigger (S'05-M'09) received the M.Sc. degree in electrical engineering (with distinction) and the Ph.D. degree in electrical engineering both from ETH Zurich in 2003 and 2008, respectively. In 2009 she was a postdoctoral researcher at the ITA center at the University of California, San Diego. Since December 2009 she is an Assistant Professor at Telecom ParisTech, in Paris, France. Her research interests are in information and communications theory.