# An Achievable Rate Region for the Broadcast Channel with Feedback

Ramji Venkataramanan
Yale University
ramji.venkataramanan@yale.edu

S. Sandeep Pradhan *
University of Michigan, Ann Arbor
pradhanv@eecs.umich.edu

May 13, 2011

**Abstract**

A single-letter achievable rate region is proposed for the two-receiver discrete memoryless broadcast channel with noiseless or noisy feedback. The coding strategy involves block-Markov superposition coding using Marton's coding scheme for the broadcast channel without feedback as the starting point. If the message rates in the Marton scheme are too high to be decoded at the end of a block, each receiver is left with a list of messages compatible with its output. Resolution information is sent in the following block to enable each receiver to resolve its list. The key observation is that the resolution information of the first receiver is correlated with that of the second. This correlated information is efficiently transmitted via joint source-channel coding, using ideas similar to the Han-Costa coding scheme. The proposed rate region is computed for two examples, including the degraded AWGN broadcast channel, which show that the region can be strictly larger than the capacity region in the absence of feedback. Finally, the proposed rate region is shown to contain the achievable region proposed independently by Shayevitz and Wigger under certain mild conditions.

## 1 Introduction

The two-receiver discrete memoryless broadcast channel is shown in Figure 1(a). The channel has one transmitter which generates a channel input $X$, and two receivers which receive $Y$ and $Z$, respectively. The channel is characterized by a conditional law $P_{YZ|X}$. The transmitter wishes to communicate information simultaneously to the receivers at rates $(R_0, R_1, R_2)$, where $R_0$ is the rate of the common message, and $R_1, R_2$ are the rates of the private messages of the two receivers. This channel has been studied extensively. The largest known set of achievable rates for this channel without feedback is due to Marton [1]. Marton's rate region is equal to the capacity region in all cases where it is known. (See [2], for example, for a list of such channels.)

Figure 1(b) shows a broadcast channel with noiseless feedback where the channel outputs at both receivers are available at the transmitter with a finite delay. El Gamal showed in [3] that feedback does not enlarge the capacity region of a physically degraded broadcast channel. Later, through a simple example, Dueck [4] demonstrated that feedback can strictly improve the capacity region of a general broadcast channel. For

*This work was supported by NSF grant CCF-0915619. It was presented in part at the IEEE International Symposium on Information Theory (ISIT) 2010, held in Austin, TX.

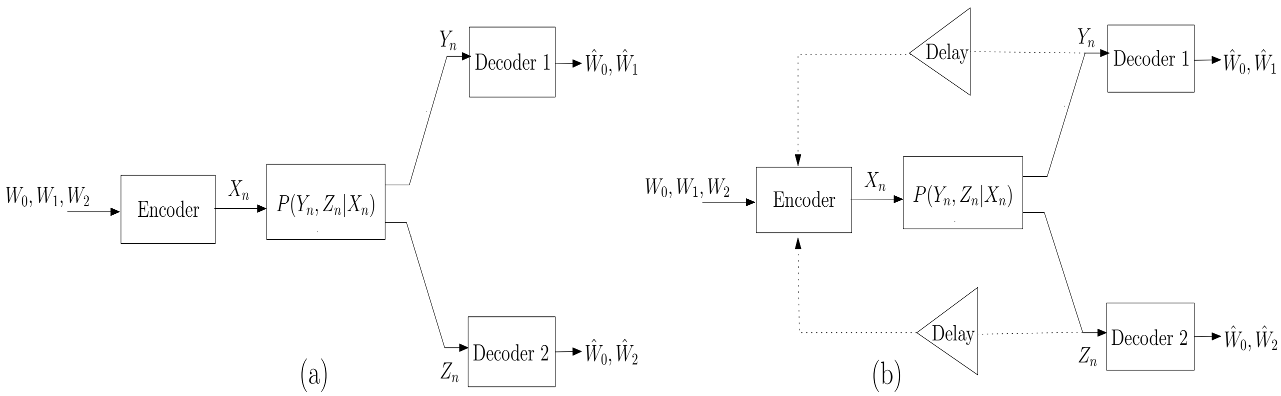


Figure 1: The discrete memoryless broadcast channel with a) no feedback b) noiseless feedback.

the degraded AWGN broadcast channel with feedback, an achievable rate region larger than the no-feedback capacity region was established in [5], and more recently, in [6]. In this paper, we establish a single-letter achievable rate region for the discrete memoryless broadcast channel with noiseless or noisy feedback.

Before describing our coding strategy, let us revisit the example from [4]. Consider the broadcast channel in Figure 2. The channel input is a binary triple $(X_0, X_1, X_2)$. $X_0$ is transmitted cleanly to both receivers. In addition, receiver 1 receives $X_1 \oplus N$ and receiver 2 receives $X_2 \oplus N$, where $N$ is an independent binary Bernoulli($\frac{1}{2}$) noise variable. Here, the operation $\oplus$ denotes the modulo-two sum. Without feedback, the maximum sum rate for this channel is 1 bit/channel use, achieved by using the clean input $X_0$ alone. In other words, no information can be reliably transmitted through inputs $X_1$ and $X_2$.

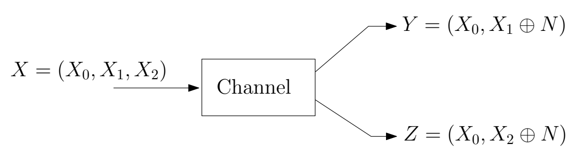


Figure 2: The channel input is a binary triple $(X_0, X_1, X_2)$. $N \sim$ Bernoulli($\frac{1}{2}$) is an independent noise variable.

Dueck described a simple scheme to achieve a greater sum rate using feedback. In the first channel use, transmit one bit to each receiver $i$ through $X_i$, $i = 1, 2$. Receivers 1 and 2 then receive $Y = X_1 \oplus N$ and $Z = X_2 \oplus N$, respectively, and cannot recover $X_i$. The transmitter learns $Y, Z$ through feedback and can compute $N = Y \oplus X_1 = Z \oplus X_2$. For the next channel use, the transmitter sets $X_0 = N$. Since $X_0$ is received noiselessly by both receivers, receiver 1 can now recover $X_1$ as $Y \oplus N$. Similarly, receiver 2 reconstructs $X_2$ as $Z \oplus N$. We can repeat this idea over several transmissions: in each channel use, transmit a fresh pair of bits (through $X_1, X_2$) as well as the noise realization of the previous channel use (through $X_0$). This yields a sum rate of 2 bits/channel use. This is, in fact, the sum-capacity of the channel since it equals the cut-set bound $\max_{P_X} I(X; YZ)$.

The example suggests a natural way to exploit feedback in a broadcast channel. If we transmit a block of information at rates outside the no-feedback capacity region, the receivers cannot uniquely decode their messages at the end of the block. Each receiver now has a list of its codewords that are jointly typical with its channel output. In the next block, we attempt to resolve these lists at the two receivers. The key observation is that the resolution information needed by receiver 1 is in general *correlated* with the resolution information needed by receiver 2. The above example is an extreme case of this: the resolution information needed by the two receivers is identical, i.e., the correlation is perfect!

It is known that correlated information can be transmitted over the broadcast channel at higher rates than independent information [7–11]. At the heart of the proposed coding scheme is a way to represent the resolution information of the two receivers as a pair of correlated sources, which is then transmitted efficiently in the next block using joint source-channel coding, along the lines of [7]. We repeat this idea over several blocks of transmission, with each block containing independent fresh information superimposed over correlated resolution information for the previous block.

The following are the main contributions of this paper:

- We obtain a single-letter achievable rate region for the discrete memoryless broadcast channel with noiseless or noisy feedback. The proposed region contains three extra random variables in addition to those in Marton's rate region.
- Using a simpler form of the rate region with only one extra random variable, we compute achievable rates for two examples including the degraded AWGN broadcast channel. These show that the achievable region is strictly larger than the capacity region in the absence of feedback.
- At the conference where our result was first presented [12], another rate region for the broadcast channel with feedback was proposed independently by Shayevitz and Wigger [13]. We show that under some (mild) conditions, the Shayevitz-Wigger region is contained in our rate region.

*Notation*: We use uppercase letters to denote random variables, lower-case for their realizations and calligraphic notation for their alphabets. Bold-face notation is used for random vectors. Unless otherwise stated, all vectors have length $n$. Thus $\mathbf{A} \triangleq A^n \triangleq (A_1, \ldots, A_n)$. The $\epsilon$-strongly typical set of block-length $n$ of a random variable with distribution $P$ is denoted $A_\epsilon^{(n)}(P)$. Logarithms are with base 2, and entropy and mutual information are measured in bits. For $\alpha \in (0, 1)$, $\bar{\alpha} \triangleq 1 - \alpha$.

In the following, we give an intuitive description of a two-phase coding scheme for communicating over a broadcast channel with noiseless feedback. We will use the notation $\sim$ to indicate the random variables used in the first phase. Thus $(\tilde{Y}, \tilde{Z})$ denote the channel output pair for the first phase, and $(Y, Z)$ the channel output pair for the second phase. We start with Marton's coding strategy for the discrete memoryless broadcast channel without feedback. The rates of the messages of the two receivers are assumed to lie outside Marton's achievable rate region. Let $\tilde{U}, \tilde{V}$, and $\tilde{W}$ denote the auxiliary random variables used to encode the information. $\tilde{W}$ carries the information meant to be decoded at both receivers. $\tilde{U}$ and $\tilde{V}$ carry the rest of the information meant for the first and the second receiver, respectively. The $\tilde{U}$- and $\tilde{V}$-codebooks are formed by randomly sampling the $\tilde{U}$- and $\tilde{V}$-typical sets, respectively. Let $\tilde{\mathbf{U}}$, $\tilde{\mathbf{V}}$ and $\tilde{\mathbf{W}}$ denote the three random codewords chosen by the transmitter. The channel input vector $\tilde{\mathbf{X}}$ is obtained by 'fusing' the triple $(\tilde{\mathbf{U}}, \tilde{\mathbf{V}}, \tilde{\mathbf{W}})$.

Since the rates lie outside Marton's region, the receivers are not able to decode the information contained in $\tilde{U}, \tilde{V}$, and $\tilde{W}$. Instead, they obtain just a list of highly likely codewords given the respective channel output vectors. The first phase can be thought of as transmission of independent messages over a broadcast channel with list decoding. At the first decoder, this list is formed by collecting all $(\tilde{U}, \tilde{W})$-codeword pairs that are jointly typical with the channel output. A similar list of $(\tilde{V}, \tilde{W})$-codeword pairs is formed at the second receiver. Note that even with feedback, the total transmission rate of the broadcast channel cannot exceed the capacity of the point-to-point channel with input $\tilde{X}$ and outputs $(\tilde{Y}, \tilde{Z})$. (This is because the channel is memoryless.) Hence, given *both* channel output vectors $(\tilde{\mathbf{Y}}, \tilde{\mathbf{Z}})$, the posterior probability of the codewords will be concentrated on the transmitted codeword triple.

The channel output vectors $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{Z}}$ are fed back to the encoder. For the second phase, we treat $(\tilde{\mathbf{U}}, \tilde{\mathbf{W}})$ as the source of information to be transmitted to the first decoder, and $(\tilde{\mathbf{V}}, \tilde{\mathbf{W}})$ as the source of information to be transmitted to the second decoder. The objective in the second phase is to communicate these two correlated pairs to the decoders over the broadcast channel, while treating $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{Z}}$ as source state information. This is accomplished using a joint-source-channel coding strategy. Transmission of correlated information over a broadcast channel has been addressed in [7,10]. The former addresses the case when the correlated information is modeled as a pair of memoryless sources characterized by a fixed single-letter distribution. In the latter, the correlated information is modeled as a random edge in an exponentially large nearly semi-regular bipartite graph. A bipartite graph is called semi-regular if the degrees of all the left vertices are the same, and the degrees of all the right vertices are the same.

In the current setup, the correlated information given by $(\tilde{\mathbf{U}}, \tilde{\mathbf{W}})$ and $(\tilde{\mathbf{V}}, \tilde{\mathbf{W}})$ does not exhibit a memoryless-source-like behavior. This is because the vectors $\tilde{\mathbf{U}}$, $\tilde{\mathbf{V}}$ and $\tilde{\mathbf{W}}$ come from codebooks. Instead, the correlated information can be modeled as a random edge in a nearly semi-regular bipartite graph. This follows from the law of large numbers since the codewords are sufficiently long and are chosen randomly. Transmission of such correlated information is addressed in [10]. We will use a combination of the techniques in [7] and [10] to develop a scheme for the second phase. At the end of the second phase, the decoders are able to decode the respective messages.

We will superimpose these two phases using a block-Markov strategy. Therefore, the overall transmission scheme has several blocks, with fresh information entering in each block being decoded in the subsequent block. The fresh information gets encoded in the first phase, and is superimposed on the second phase which corresponds to information that entered in the previous block.

It turns out that the performance of such a scheme cannot be directly captured by single-letter information quantities. This is because the state information, given by the channel outputs of all the previous blocks, keeps accumulating, leading to a different joint distribution of the random variables in each block. We address this issue by constraining the distributions used in the second phase so that in every block, all the sequences follow a stationary joint distribution. This results in a first-order stationary Markov process of the sequences across blocks.

The rest of paper is organized as follows. In Section 2, we give the formal problem statement and the first main result of the paper, an achievable rate region for broadcast channel with noiseless feedback. We give an outline of the proof of the coding theorem in Section 3. In Section 4, we give the second main result of the paper, an achievable rate region for broadcast channels with noisy feedback. In Section 5, the rate region is computed for two examples, including the degraded AWGN broadcast channel. In Section 6, we compare our rate region with the one proposed by Shayevitz and Wigger. The formal proof of the coding theorem is given in Section 7, and Section 8 concludes the paper.

# 2 Problem Statement and Main Result

A two-user general discrete memoryless broadcast channel is a quadruple $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, P_{YZ|X})$ of input alphabet $\mathcal{X}$, two output alphabets $\mathcal{Y}$, $\mathcal{Z}$ and a set of probability distributions $P_{YZ|X}(\cdot|x)$ on $\mathcal{Y} \times \mathcal{Z}$ for every $x \in \mathcal{X}$.

The channel satisfies the following conditions for all $n = 1, 2, \ldots$

$$Pr(Y_n = y_n, Z_n = z_n | X^n = \mathbf{x}, Y^{n-1} = \mathbf{y}, Z^{n-1} = \mathbf{z}) = P_{YZ|X}(y_n, z_n | x_n), \tag{1}$$

for all $y_n \in \mathcal{Y}$, $z_n \in \mathcal{Z}$, $\mathbf{x} \in \mathcal{X}^n$, $\mathbf{y} \in \mathcal{Y}^{n-1}$ and $\mathbf{z} \in \mathcal{Z}^{n-1}$. The outputs are fed back noiselessly to the encoder as shown in Figure 1 (b).

**Definition 2.1.** *An $(n, M_0, M_1, M_2)$ transmission system for a given broadcast channel with noiseless feedback consists of*

- *A sequence of mappings for the encoder:*

$$e_m : \{1, 2, \ldots, M_0\} \times \{1, 2, \ldots, M_1\} \times \{1, 2, \ldots, M_2\} \times \mathcal{Y}^{m-1} \times \mathcal{Z}^{m-1} \to \mathcal{X}, \quad m = 1, 2, \ldots, n, \tag{2}$$

- *A pair of decoder mappings:*

$$g_1 : \mathcal{Y}^n \to \{1, 2, \ldots, M_0\} \times \{1, 2, \ldots, M_1\}, \quad g_2 : \mathcal{Z}^n \to \{1, 2, \ldots, M_0\} \times \{1, 2, \ldots, M_2\}. \tag{3}$$

***Remark***: Though we have defined the transmission system above for feedback delay 1, all the results in this paper hold for feedback with any finite delay $k$.

We use $W_0$ to denote the common message, and $W_1, W_2$ to denote the private messages of decoders 1 and 2, respectively. The messages $(W_0, W_1, W_2)$ are uniformly distributed over the set $\{1, 2, \ldots, M_0\} \times \{1, 2, \ldots, M_1\} \times \{1, 2, \ldots, M_2\}$. The channel input at time $n$ is given by $X_n = e_n(W_0, W_1, W_2, Y^{n-1}, Z^{n-1})$. The average error probability of the above transmission system is given by

$$\tau = \frac{1}{M_0 M_1 M_2} \sum_{k=1}^{M_0} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} Pr\left( (g_1(Y^n), g_2(Z^n)) \neq ((k,i),(k,j)) | (W_0, W_1, W_2) = (k,i,j) \right). \tag{4}$$

**Definition 2.2.** *A triple of non-negative real numbers $(R_0, R_1, R_2)$ is said to be achievable for a given broadcast channel with feedback if $\forall \epsilon > 0$, there exists an $N(\epsilon) > 0$ such that for all $n > N(\epsilon)$, there exists an $(n, M_0, M_1, M_2)$ transmission system satisfying the following constraints:*

$$\frac{1}{n} \log M_0 \geq R_0 - \epsilon, \quad \frac{1}{n} \log M_1 \geq R_1 - \epsilon, \quad \frac{1}{n} \log M_2 \geq R_2 - \epsilon, \quad \tau \leq \epsilon. \tag{5}$$

*The set of all achievable rate pairs is the capacity region of the channel.*

Before stating the main result of the paper, we define the structure for the joint distribution of variables in each block of our coding scheme, and also the joint distribution of variables across successive blocks.

**Definition 2.3.** *Given a broadcast channel $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, P_{YZ|X})$, define $\mathcal{P}$ as the set of all distributions $P$ on $\mathcal{U} \times \mathcal{V} \times \mathcal{A} \times \mathcal{B} \times \mathcal{C} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ of the form*

$$P_{ABC} \; P_{UV|C} \; P_{X|ABCUV} \; P_{YZ|X},$$

*where $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{U}$, and $\mathcal{V}$ are arbitrary sets. Consider two sets of random variables $(U, V, A, B, C, X, Y, Z)$ and $(\tilde{U}, \tilde{V}, \tilde{A}, \tilde{B}, \tilde{C}, \tilde{X}, \tilde{Y}, \tilde{Z})$ each having the same distribution $P$. For brevity, we often refer to the collection*

$(A,B,Y,Z)$ *as* $K$*, to* $(\tilde{A},\tilde{B},\tilde{Y},\tilde{Z})$ *as* $\tilde{K}$*, and to* $\mathcal{A}\times\mathcal{B}\times\mathcal{Y}\times\mathcal{Z}$ *as* $\mathcal{K}$*. Hence*

$$P_{\tilde{U}\tilde{V}\tilde{C}\tilde{K}\tilde{X}} = P_{UVCKX} = P.$$

*For a given* $P\in\mathcal{P}$*, define* $\mathcal{Q}(P)$ *as the set of conditional distributions* $Q$ *that satisfy the following consistency condition*

$$P_{ABC}(a,b,c) = \sum_{\tilde{u},\tilde{v},,\tilde{k},\tilde{c}\in\mathcal{U}\times\mathcal{V}\times\mathcal{K}\times\mathcal{C}} Q_{ABC|\tilde{U}\tilde{V}\tilde{K}\tilde{C}}(a,b,c|\tilde{u},\tilde{v},\tilde{k},\tilde{c})P_{UVKC}(\tilde{u},\tilde{v},\tilde{k},\tilde{c}), \quad \forall(a,b,c). \tag{6}$$

*Then for any* $P\in\mathcal{P}$ *and* $Q\in\mathcal{Q}(P)$*, the* ***joint*** *distribution of the two sets* $(U,V,K,C,X)$ *and* $(\tilde{U},\tilde{V},\tilde{K},\tilde{C},\tilde{X})$ *is*

$$P_{\tilde{U}\tilde{V}\tilde{K}\tilde{C}\tilde{X}}\; Q_{ABC|\tilde{U}\tilde{V}\tilde{C}\tilde{K}}\; P_{UVKX|ABC}. \tag{7}$$

With the above definitions, we have the following theorem.

**Theorem 1.** *Given a broadcast channel* $(\mathcal{X},\mathcal{Y},\mathcal{Z},P_{YZ|X})$*, for any distribution* $P\in\mathcal{P}$ *and* $Q\in\mathcal{Q}(P)$*, the following rate region is achievable with noiseless feedback.*

$$R_0 < \min\{T_1,T_2,T_3,T_4,T_5\} \tag{8}$$
$$R_0+R_1 < I(\tilde{U}AC;Y\tilde{Y}\tilde{A}|\tilde{C}) - I(\tilde{V}\tilde{K};AC|\tilde{U}\tilde{C}) \tag{9}$$
$$R_0+R_2 < I(\tilde{V}BC;Z\tilde{Z}\tilde{B}|\tilde{C}) - I(\tilde{U}\tilde{K};BC|\tilde{V}\tilde{C}) \tag{10}$$
$$\begin{aligned}R_0+R_1+R_2 < {} & I(\tilde{U}AC;Y\tilde{Y}\tilde{A}|\tilde{C}) - I(\tilde{V}\tilde{K};AC|\tilde{U}\tilde{C}) - I(U;V|C)\\ & + I(\tilde{V};C|\tilde{C}) + I(\tilde{V}B;Z\tilde{Z}\tilde{B}|C\tilde{C}) - I(\tilde{U}\tilde{K}A;B|C\tilde{V}\tilde{C})\end{aligned} \tag{11}$$
$$\begin{aligned}R_0+R_1+R_2 < {} & I(\tilde{V}BC;Z\tilde{Z}\tilde{B}|\tilde{C}) - I(\tilde{U}\tilde{K};BC|\tilde{V}\tilde{C}) - I(U;V|C)\\ & + I(\tilde{U};C|\tilde{C}) + I(\tilde{U}A;Y\tilde{Y}\tilde{A}|C\tilde{C}) - I(\tilde{V}\tilde{K}B;A|C\tilde{U}\tilde{C})\end{aligned} \tag{12}$$
$$\begin{aligned}2R_0+R_1+R_2 < {} & I(\tilde{U}AC;Y\tilde{Y}\tilde{A}|\tilde{C}) - I(\tilde{V}\tilde{K};AC|\tilde{U}\tilde{C}) - I(U;V|C)\\ & + I(\tilde{V}BC;Z\tilde{Z}\tilde{B}|\tilde{C}) - I(\tilde{U}\tilde{K};BC|\tilde{V}\tilde{C}) - I(A;B|C\tilde{C}\tilde{U}\tilde{V}\tilde{K})\end{aligned} \tag{13}$$

*where*

$$\begin{aligned}
T_1 &\triangleq I(AC;Y\tilde{Y}\tilde{A}|\tilde{C}\tilde{U}) - I(\tilde{V}\tilde{K};AC|\tilde{C}\tilde{U})\\
T_2 &\triangleq I(BC;Z\tilde{Z}\tilde{B}|\tilde{C}\tilde{V}) - I(\tilde{U}\tilde{K};BC|\tilde{C}\tilde{V})\\
T_3 &\triangleq I(AC;Y\tilde{Y}\tilde{A}|\tilde{C}\tilde{U}) + I(B;Z\tilde{Z}\tilde{B}|\tilde{C}\tilde{V}C) - I(\tilde{V}\tilde{K};AC|\tilde{C}\tilde{U}) - I(\tilde{U}\tilde{K}A;B|C\tilde{C}\tilde{V})\\
T_4 &\triangleq I(A;Y\tilde{Y}\tilde{A}|\tilde{C}\tilde{U}C) + I(BC;Z\tilde{Z}\tilde{B}|\tilde{C}\tilde{V}) - I(\tilde{V}\tilde{K}B;A|C\tilde{C}\tilde{U}) - I(\tilde{U}\tilde{K};BC|\tilde{C}\tilde{V})\\
T_5 &\triangleq \frac{1}{2}\Big[I(AC;Y\tilde{Y}\tilde{A}|\tilde{C}\tilde{U}) - I(\tilde{V}\tilde{K};AC|\tilde{C}\tilde{U}) + I(BC;Z\tilde{Z}\tilde{B}|\tilde{C}\tilde{V}) - I(\tilde{U}\tilde{K};BC|\tilde{C}\tilde{V}) - I(A;B|C\tilde{C}\tilde{U}\tilde{V}\tilde{K})\Big]
\end{aligned}$$

*Proof.* This theorem is proved in Section 7.

***Remark***: We can recover Marton's achievable rate region for the broadcast channel without feedback by setting $A=B=\phi$, and $C=W$ with $Q_{C|\tilde{U}\tilde{V}\tilde{K}\tilde{C}} = P_W$.

## 3 Coding scheme

In this section, we give an informal outline of the proof of the theorem. The formal proof is given in Section 7. Let us first consider the case when the rate $R_0$ of the common message equals 0. Let the message rate pair $(R_1, R_2)$ lie outside Marton's achievable region [1]. The coding scheme uses a block-Markov superposition strategy, with the communication taking place over $L$ blocks, each of length $n$.

In each block, a fresh pair of messages is encoded using the Marton coding strategy (for the broadcast channel without feedback). In block $l$, random variables $U$ and $V$ carry the fresh information for receivers 1 and 2, respectively. At the end of this block, the receivers are not able to decode the information in $(U, V)$ completely, so we send 'resolution' information in block $(l + 1)$ using random variables $(A, B, C)$. The pair $(A, C)$ is meant to be decoded by the first receiver, and the pair $(B, C)$ by the second receiver. Thus in each block, we obtain the channel output by superimposing fresh information on the resolution information for the previous block. At the end of the block, the first receiver decodes $(A, C)$, the second receiver decodes $(B, C)$, thereby resolving the uncertainty about their messages of the previous block.

*Codebooks*: The $A$-, $B$-, and $C$-codebooks are constructed on the alphabets $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ respectively. The exact procedure for this construction, and the method for selecting codewords from these codebooks will be described in the sequel. Since $C$ is decoded first by both the receivers, conditioned on each codeword in the $C$-codebook, we construct $\mathbf{U}$- and $\mathbf{V}$-codebooks of sizes $2^{nR'_1}$ and $2^{nR'_2}$ by generating codewords according to $P_{U|C}$ and $P_{V|C}$, respectively. Each $\mathbf{U}$-codebook is divided into $2^{nR_1}$ bins, and each $\mathbf{V}$-codebook into $2^{nR_2}$ bins.

*Encoding*: In each block $l$, the encoder chooses a tuple of five codewords $(\mathbf{A}_l, \mathbf{B}_l, \mathbf{C}_l, \mathbf{U}_l, \mathbf{V}_l)$ as follows. The resolution information for block $(l-1)$ is used to select $(\mathbf{A}_l, \mathbf{B}_l, \mathbf{C}_l)$ from the $A$-, $B$- and $C$-codebooks. $\mathbf{C}_l$ determines the $\mathbf{U}$- and $\mathbf{V}$-codebooks to be used to encode the message pair of block $l$. Denoting the message pair by $(m_{1l}, m_{2l})$, the encoder choose a $U$-codeword from bin $m_{1l}$ of the $U$-codebook and a $V$-codeword from bin $m_{2l}$ of the $V$-codebook that are jointly typical according to $P_{UV|C}$. This pair of jointly typical codewords is set to be $(\mathbf{U}_l, \mathbf{V}_l)$.

This step is successful if the product of the sizes of $U$-bin and $V$-bin is nearly equal to $2^{nI(U;V|C)}$ [14]. Therefore, we have

$$R'_1 + R'_2 - R_1 - R_2 > I(U; V|C). \tag{14}$$

These five codewords are combined using the transformation $P_{X|ABCUV}$ (applied componentwise) to generate the channel input $\mathbf{X}_l$.

*Decoding*: After receiving the channel output of block $l$, receiver 1 first decodes $(\mathbf{A}_l, \mathbf{C}_l)$, and receiver 2 decodes $(\mathbf{B}_l, \mathbf{C}_l)$ However, the rates $R'_1, R'_2$ of the $U$- and $V$-codebooks are too large for receivers 1 and 2 to uniquely decode $\mathbf{U}_l$ and $\mathbf{V}_l$, respectively. Hence receiver 1 is left with a list of $U$-codewords that are jointly typical with its channel output $\mathbf{Y}_l$ and the just-decoded resolution information $(\mathbf{A}_l, \mathbf{C}_l)$; receiver 2 has a similar list of $V$-codewords that are jointly typical with its channel output $\mathbf{Z}_l$, and the just-decoded resolution information $(\mathbf{B}_l, \mathbf{C}_l)$. The sizes of the lists are nearly equal to $2^{n(R'_1 - I(U;Y|AC))}$ and $2^{n(R'_2 - I(V;Z|BC))}$, respectively. The transmitter knows both these lists due to feedback, and resolves them in the next block as follows.

In block $(l + 1)$, the random variables of block $l$ are represented using the notation $\sim$. Thus we have

$$\tilde{\mathbf{U}}_{l+1} = \mathbf{U}_l,\ \tilde{\mathbf{V}}_{l+1} = \mathbf{V}_l,\ \tilde{\mathbf{C}}_{l+1} = \mathbf{C}_l,\ \tilde{\mathbf{A}}_{l+1} = \mathbf{A}_l,\ \tilde{\mathbf{B}}_{l+1} = \mathbf{B}_l,$$

For brevity, we denote the collection of random variables $(A, B, Y, Z)$ as $K$, and $(\mathbf{A}_l, \mathbf{B}_l, \mathbf{Y}_l, \mathbf{Z}_l)$ as $\mathbf{K}_l = \tilde{\mathbf{K}}_{l+1}$. The random variables $(U, V, K, C)$ in block $l$ are jointly distributed via $P_{ABC}P_{UV|C}P_{YZ|ABCUV}$ chosen from $\mathcal{P}$ as given in the statement of the theorem.

For block $l+1$, $(\tilde{\mathbf{U}}_{l+1}, \tilde{\mathbf{V}}_{l+1}) = (\tilde{\mathbf{U}}_l, \tilde{\mathbf{V}}_l)$ can be considered to be a realization of a pair of correlated 'sources' ($\tilde{U}$ and $\tilde{V}$), jointly distributed according to $P_{\tilde{U}\tilde{V}|\tilde{Y}\tilde{Z}\tilde{A}\tilde{B}\tilde{C}}$ along with the transmitter side information given by $(\tilde{\mathbf{A}}_{l+1}, \tilde{\mathbf{B}}_{l+1}, \tilde{\mathbf{Y}}_{l+1}, \tilde{\mathbf{Z}}_{l+1})$, and the common side-information $\tilde{\mathbf{C}}_{l+1}$. The goal in block $(l+1)$ is to transmit this pair of correlated sources over the broadcast channel, with

- Receiver 1 needing to decode $\tilde{\mathbf{U}}_{l+1}$, treating $(\tilde{\mathbf{A}}_{l+1}, \tilde{\mathbf{Y}}_{l+1}, \tilde{\mathbf{C}}_{l+1})$ as receiver side-information,

- Receiver 2 needing to decode $\tilde{\mathbf{V}}_{l+1}$, treating $(\tilde{\mathbf{B}}_{l+1}, \tilde{\mathbf{Z}}_{l+1}, \tilde{\mathbf{C}}_{l+1})$ as receiver side-information.

We use the ideas of Han and Costa [7] to transmit this pair of correlated sources over the broadcast channel (with appropriate extensions to take into account the different side-information available at the transmitter and the receivers). This is shown in Figure 3. The triple of correlated random variables $(A, B, C)$ is used to cover the sources. This triple carries the resolution information intended to disambiguate the lists of the two receivers. The random variables of block $(l+1)$, given by $(A, B, C)$ are related to the random variables in block $l$ via $Q_{ABC|\tilde{U}\tilde{V}\tilde{K}\tilde{C}}$, chosen from $\mathcal{Q}$ given in the statement of the theorem. We now describe the construction of the $A$-, $B$-, and $C$- codebooks.

*Covering the Sources*: For each $\tilde{\mathbf{c}} \in \mathcal{C}^n$, a $C$-codebook $\mathcal{C}_C(\tilde{\mathbf{c}})$ of rate $\rho_0$ is constructed randomly from $P_{C|\tilde{C}}$. For every realization of $\tilde{\mathbf{u}} \in \mathcal{U}^n$, $\tilde{\mathbf{c}} \in \mathcal{C}^n$, and $\mathbf{c} \in \mathcal{C}^n$, an $A$-codebook $\mathcal{C}_A(\tilde{\mathbf{u}}, \tilde{\mathbf{c}}, \mathbf{c})$ of rate $\rho_1$ is constructed with codewords picked randomly according to $P_{A|\tilde{U},\tilde{C},C}$. Similarly, for every realization of $\tilde{\mathbf{v}} \in \mathcal{V}^n$, $\tilde{\mathbf{c}} \in \mathcal{C}^n$, and $\mathbf{c} \in \mathcal{C}^n$, a $B$-codebook $\mathcal{C}_B(\tilde{\mathbf{v}}, \tilde{\mathbf{c}}, \mathbf{c})$ of rate $\rho_2$ is constructed with codewords picked randomly according to $P_{B|\tilde{V},\tilde{C},C}$.

At the beginning of block $(l+1)$, for a given realization $(\tilde{\mathbf{U}}_{l+1}, \tilde{\mathbf{V}}_{l+1}, \tilde{\mathbf{K}}_{l+1}, \tilde{\mathbf{C}}_{l+1})$, of correlated 'sources', and side information, the encoder chooses a triple of codewords $(\mathbf{A}_{l+1}, \mathbf{B}_{l+1}, \mathbf{C}_{l+1})$ from the appropriate $A$-, $B$- and $C$-codebooks such that the two tuples are jointly typical according to $P_{\tilde{U}\tilde{V}\tilde{K}\tilde{C}}Q_{ABC|\tilde{U}\tilde{K}\tilde{V}\tilde{C}}$. The channel input $\mathbf{X}_{l+1}$ is generated by fusing this $(\mathbf{A}_{l+1}, \mathbf{B}_{l+1}, \mathbf{C}_{l+1})$ with the pair of codewords $(\mathbf{U}_{l+1}, \mathbf{V}_{l+1})$, which carry fresh information in block $(l+1)$.

Now consider the general case when $R_0 > 0$. We can use the variable $C$ to encode common information to be decoded by both receivers. Hence $C$ serves two purposes - it is used to (a) cover the correlated sources and transmitter side-information, thus being part of the resolution information, and (b) to carry common fresh information. Note that in every block, two communication tasks are being accomplished simultaneously. The first is channel coding over the broadcast channel with list decoding, accomplished via $(U, V, C)$. The second is joint-source-channel coding of correlated sources over the broadcast channel, accomplished via $(A, B, C)$. Recall that in Marton's achievable region for the broadcast channel without feedback, there is a random variable $W$ meant to be decoded by both receivers. In the present case, it turns out that $C$ can be made to assume the dual role of the common random variable associated with both the tasks.

*Analysis*: For this encoding to be successful, we need the following covering conditions. These are similar

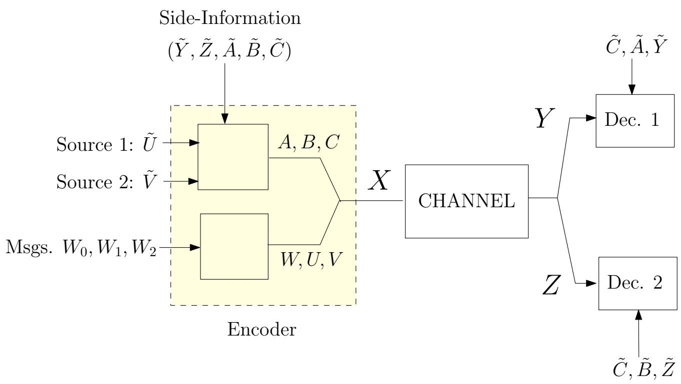


Figure 3: Transmitting correlated sources with side-information at the receivers through $(A, B, C)$, and fresh information through $U, V$.

to equations (3.1)-(3.5) in [7].

$$\rho_0 > I(\tilde{U}\tilde{K}\tilde{V}; C|\tilde{C}) + R_0 \quad (15)$$

$$\rho_0 + \rho_1 > I(\tilde{V}\tilde{K}; A|C\tilde{C}\tilde{U}) + I(\tilde{U}\tilde{K}\tilde{V}; C|\tilde{C}) + R_0 \quad (16)$$

$$\rho_0 + \rho_2 > I(\tilde{U}\tilde{K}; B|C\tilde{C}\tilde{V}) + I(\tilde{U}\tilde{K}\tilde{V}; C|\tilde{C}) + R_0 \quad (17)$$

$$\rho_0 + \rho_1 + \rho_2 > I(\tilde{V}\tilde{K}; A|C\tilde{C}\tilde{U}) + I(\tilde{U}\tilde{K}; B|C\tilde{C}\tilde{V}) + I(A; B|\tilde{U}\tilde{K}\tilde{V}C\tilde{C}) + I(\tilde{U}\tilde{K}\tilde{V}; C|\tilde{C}) + R_0 \quad (18)$$

At the end of block $(l+1)$, receiver 1 determines $\mathbf{U}_l = \tilde{\mathbf{U}}_{l+1}$ by finding the pair $(\tilde{\mathbf{U}}_{l+1}, \mathbf{A}_{l+1}, \mathbf{C}_{l+1})$ using joint typical decoding in the composite $U$-, $A$-, and $C$-codebooks. A similar procedure is followed at the second reciever. For the decoding to be successful, we need the following packing conditions.

$$R_1' + \rho_0 + \rho_1 < I(\tilde{U}; Y\tilde{Y}\tilde{A}|\tilde{C}) + I(C; Y\tilde{A}\tilde{Y}\tilde{U}|\tilde{C}) + I(A; Y\tilde{A}\tilde{Y}|\tilde{U}C\tilde{C}) \quad (19)$$

$$R_1' + \rho_1 < I(\tilde{U}; Y\tilde{A}\tilde{Y}C|\tilde{C}) + I(A; Y\tilde{A}\tilde{Y}|\tilde{U}C\tilde{C}) \quad (20)$$

$$R_2' + \rho_0 + \rho_2 < I(\tilde{V}; Z\tilde{Z}\tilde{B}|\tilde{C}) + I(C; Z\tilde{B}\tilde{Z}\tilde{V}|\tilde{C}) + I(B; Z\tilde{B}\tilde{Z}|\tilde{V}C\tilde{C}) \quad (21)$$

$$R_2' + \rho_2 < I(\tilde{V}; Z\tilde{B}\tilde{Z}C|\tilde{C}) + I(B; Z\tilde{B}\tilde{Z}|\tilde{V}C\tilde{C}) \quad (22)$$

$$\rho_0 + \rho_1 < I(C; Y\tilde{A}\tilde{Y}\tilde{U}|\tilde{C}) + I(A; Y\tilde{A}\tilde{Y}|\tilde{U}\tilde{C}C) \quad (23)$$

$$\rho_0 + \rho_2 < I(C; Z\tilde{B}\tilde{Z}\tilde{V}|\tilde{C}) + I(B; Z\tilde{B}\tilde{Z}|\tilde{V}\tilde{C}C) \quad (24)$$

$$\rho_1 < I(A; Y\tilde{A}\tilde{Y}|\tilde{U}\tilde{C}C) \quad (25)$$

$$\rho_2 < I(B; Z\tilde{B}\tilde{Z}|\tilde{V}\tilde{C}C) \quad (26)$$

Performing Fourier-Motzkin elimination on equations (14), (15-18) and (19-26), we obtain the statement of the theorem.

To get a single-letter characterization of achievable rates, we need to ensure that the random variables in each block follow a stationary joint distribution. We now describe how we ensure that the sequences in each block are jointly distributed according to

$$P_{ABC} \cdot P_{UV|C} \cdot P_{X|ABCUV} \cdot P_{YZ|X} \quad (27)$$

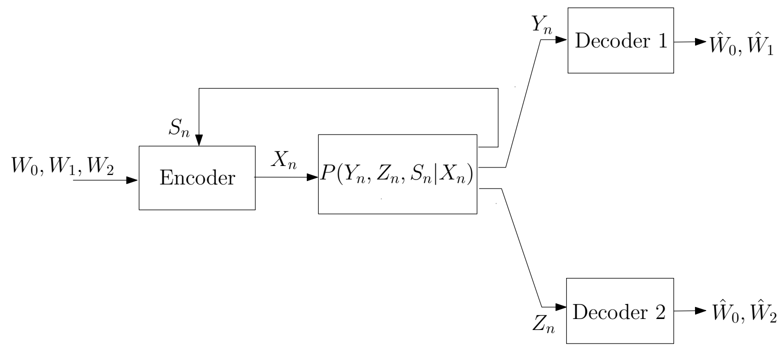


Figure 4: The discrete memoryless broadcast channel with noisy feedback.

for some chosen $P_{ABC}, P_{UV}$, and $P_{X|ABCUV}$.

Suppose that the sequences in a given block are jointly distributed according to (27). These sequences become the source pair $(\tilde{U}, \tilde{V})$, and the side-information $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{Y}, \tilde{Z})$ in the next block. To cover the source pair with $(A, B, C)$, we pick a conditional distribution $Q_{ABC|\tilde{A}\tilde{B}\tilde{C}\tilde{U}\tilde{V}\tilde{Y}\tilde{Z}}$ such that the covering sequences are distributed according to $P_{ABC}$. This holds when the consistency condition given by (6) is satisfied. We thereby ensure that the sequences in each block are jointly distributed according to (27). Our technique of exploiting the correlation induced by feedback is similar in spirit to the coding scheme of Han for two-way channels [15].

Note that the transmitter side information $\tilde{K} = (\tilde{A}\tilde{B}\tilde{Y}\tilde{Z})$ is exploited at the encoder in the covering operation implicitly, without using codebooks conditioned on $\tilde{K}$. This is because this side information is only partially available at the receivers, with receiver 1 having only $(\tilde{A}, \tilde{Y})$, and receiver 2 having only $(\tilde{B}, \tilde{Z})$. Hence this approach can be extended to the case of noisy feedback in a straightforward way[1]. This is done is the next section.

# 4 Noisy Feedback

A two-user discrete memoryless broadcast channel with noisy feedback is a quintuple $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{S}, P_{YZS|X})$ of input alphabet $\mathcal{X}$, two output alphabets $\mathcal{Y}$, $\mathcal{Z}$, noisy feedback alphabet $\mathcal{S}$ and a set of probability distributions $P_{YZS|X}(\cdot|x)$ on $\mathcal{Y} \times \mathcal{Z} \times S$ for every $x \in \mathcal{X}$. The channel satisfies the following conditions for all $n = 1, 2, \ldots,$

$$Pr(Y_n = y_n, Z_n = z_n, S = s_n | X^n = \mathbf{x}, Y^{n-1} = \mathbf{y}, Z^{n-1} = \mathbf{z}, S^{n-1} = \mathbf{s}) = P_{YZS|X}(y_n, z_n, s_n | x_n), \qquad (28)$$

for all $y_n \in \mathcal{Y}$, $z_n \in \mathcal{Z}$, $s_n \in \mathcal{S}$, $\mathbf{x} \in \mathcal{X}^n$, $\mathbf{y} \in \mathcal{Y}^{n-1}$, $\mathbf{s} \in \mathcal{S}^{n-1}$ and $\mathbf{z} \in \mathcal{Z}^{n-1}$. The schematic is shown in Figure 4. We note that we can obtain the broadcast channel with noiseless feedback as a special case by setting $\mathcal{S} = \mathcal{Y} \times \mathcal{Z}$, and $S_n = (Y_n, Z_n)$.

**Definition 4.1.** *An $(n, M_0, M_1, M_2)$ transmission system for a given broadcast channel with noisy feedback consists of*

- *A sequence of mappings for the encoder:*

$$e_m : \{1, 2, \ldots, M_0\} \times \{1, 2, \ldots, M_1\} \times \{1, 2, \ldots, M_2\} \times \mathcal{S}^{m-1} \to \mathcal{X}, \quad m = 1, 2, \ldots, n, \qquad (29)$$

[1]This is in contrast to communication over a multiple-access channel with feedback, where there is a significant difference between the noiseless feedback and noisy feedback [16].

- *A pair of decoder mappings:*

$$g_1 : \mathcal{Y}^n \to \{1,2,\dots,M_0\}\times\{1,2,\dots,M_1\}, \quad g_2 : \mathcal{Z}^n \to \{1,2,\dots,M_0\}\times\{1,2,\dots,M_2\}. \tag{30}$$

The messages $(W_0, W_1, W_2)$ are uniformly distributed over the set $\{1,2,\dots,M_0\}\times\{1,2,\dots,M_1\}\times\{1,2,\dots,M_2\}$. The channel input at time $n$ is given by $X_n = e_n(W_0, W_1, W_2, S^{n-1})$. The average error probability of the above transmission system is given by

$$\tau = \frac{1}{M_0 M_1 M_2}\sum_{k=1}^{M_0}\sum_{i=1}^{M_1}\sum_{j=1}^{M_2} Pr\left((g_1(Y^n), g_2(Z^n)) \neq ((k,i),(k,j)) | (W_0,W_1,W_2) = (k,i,j)\right). \tag{31}$$

The definition of achievable rates and the capacity region for a broadcast channel with noisy feedback is identical to Definition 2.2. We now state the second result of the paper, an achievable rate region for the discrete memoryless broadcast channel with noisy feedback.

**Definition 4.2.** *Given a broadcast channel with noisy feedback $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{S}, P_{YZS|X})$, define $\mathcal{P}$ as the set of all distributions $P$ on $\mathcal{U}\times\mathcal{V}\times\mathcal{A}\times\mathcal{B}\times\mathcal{C}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$ of the form*

$$P_{ABC}\ P_{UV|C}\ P_{X|ABCUV}\ P_{YZ|X},$$

*where $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{U}$, and $\mathcal{V}$ are arbitrary sets. Consider two sets of random variables $(U,V,A,B,C,X,Y,Z,S)$ and $(\tilde{U},\tilde{V},\tilde{A},\tilde{B},\tilde{C},\tilde{X},\tilde{Y},\tilde{Z},\tilde{S})$ each having the same distribution $P$. For brevity, we often refer to the collection $(A,B,S)$ as $K$, to $(\tilde{A},\tilde{B},\tilde{S})$ as $\tilde{K}$, and to $\mathcal{A}\times\mathcal{B}\times\mathcal{S}$ as $\mathcal{K}$. Hence*

$$P_{\tilde{U}\tilde{V}\tilde{C}\tilde{K}\tilde{Y}\tilde{Z}} = P_{UVCKYZ} = P.$$

*For a given $P\in\mathcal{P}$, define $\mathcal{Q}(P)$ as the set of conditional distributions $Q$ that satisfy the following consistency condition*

$$P_{ABC}(a,b,c) = \sum_{\tilde{u},\tilde{v},,\tilde{k},\tilde{c}\in\mathcal{U}\times\mathcal{V}\times\mathcal{K}\times\mathcal{C}} Q_{ABC|\tilde{U}\tilde{V}\tilde{K}\tilde{C}}(a,b,c|\tilde{u},\tilde{v},\tilde{k},\tilde{c})P_{UVKC}(\tilde{u},\tilde{v},\tilde{k},\tilde{c}), \quad \forall (a,b,c). \tag{32}$$

*Then for any $P\in\mathcal{P}$ and $Q\in\mathcal{Q}(P)$, the joint distribution of the two sets $(U,V,K,C,X,Y,Z)$ and $(\tilde{U},\tilde{V},\tilde{K},\tilde{C},\tilde{X},\tilde{Y},\tilde{Z})$ is*

$$P_{\tilde{U}\tilde{V}\tilde{K}\tilde{C}\tilde{X}\tilde{Y}\tilde{Z}}\ Q_{ABC|\tilde{U}\tilde{V}\tilde{C}\tilde{K}}\ P_{UVKXYZ|ABC}. \tag{33}$$

With the above definitions, we have the following theorem:

**Theorem 2.** *Given a broadcast channel with noisy feedback $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{S}, P_{YZS|X})$, for any distribution $P\in\mathcal{P}$*

*and $Q \in \mathcal{Q}(P)$, the following rate region is achievable.*

$$
\begin{aligned}
R_0 &< \min\{T_1, T_2, T_3, T_4, T_5\} && (34)\\
R_0 + R_1 &< I(\tilde{U}AC; Y\tilde{Y}\tilde{A}|\tilde{C}) - I(\tilde{V}\tilde{K}; AC|\tilde{U}\tilde{C}) && (35)\\
R_0 + R_2 &< I(\tilde{V}BC; Z\tilde{Z}\tilde{B}|\tilde{C}) - I(\tilde{U}\tilde{K}; BC|\tilde{V}\tilde{C}) && (36)\\
R_0 + R_1 + R_2 &< I(\tilde{U}AC; Y\tilde{Y}\tilde{A}|\tilde{C}) - I(\tilde{V}\tilde{K}; AC|\tilde{U}\tilde{C}) - I(U;V|C) && (37)\\
&\quad + I(\tilde{V}; C|\tilde{C}) + I(\tilde{V}B; Z\tilde{Z}\tilde{B}|C\tilde{C}) - I(\tilde{U}\tilde{K}A; B|C\tilde{V}\tilde{C}) &&\\
R_0 + R_1 + R_2 &< I(\tilde{V}BC; Z\tilde{Z}\tilde{B}|\tilde{C}) - I(\tilde{U}\tilde{K}; BC|\tilde{V}\tilde{C}) - I(U;V|C) && (38)\\
&\quad + I(\tilde{U}; C|\tilde{C}) + I(\tilde{U}A; Y\tilde{Y}\tilde{A}|C\tilde{C}) - I(\tilde{V}\tilde{K}B; A|C\tilde{U}\tilde{C}) &&\\
2R_0 + R_1 + R_2 &< I(\tilde{U}AC; Y\tilde{Y}\tilde{A}|\tilde{C}) - I(\tilde{V}\tilde{K}; AC|\tilde{U}\tilde{C}) - I(U;V|C) && (39)\\
&\quad + I(\tilde{V}BC; Z\tilde{Z}\tilde{B}|\tilde{C}) - I(\tilde{U}\tilde{K}; BC|\tilde{V}\tilde{C}) - I(A;B|C\tilde{C}\tilde{U}\tilde{V}\tilde{K}) &&
\end{aligned}
$$

*where*

$$
\begin{aligned}
T_1 &\triangleq I(AC; Y\tilde{Y}\tilde{A}|\tilde{C}\tilde{U}) - I(\tilde{V}\tilde{K}; AC|\tilde{C}\tilde{U})\\
T_2 &\triangleq I(BC; Z\tilde{Z}\tilde{B}|\tilde{C}\tilde{V}) - I(\tilde{U}\tilde{K}; BC|\tilde{C}\tilde{V})\\
T_3 &\triangleq I(AC; Y\tilde{Y}\tilde{A}|\tilde{C}\tilde{U}) + I(B; Z\tilde{Z}\tilde{B}|\tilde{C}\tilde{V}C) - I(\tilde{V}\tilde{K}; AC|\tilde{C}\tilde{U}) - I(\tilde{U}\tilde{K}A; B|C\tilde{C}\tilde{V})\\
T_4 &\triangleq I(A; Y\tilde{Y}\tilde{A}|\tilde{C}\tilde{U}C) + I(BC; Z\tilde{Z}\tilde{B}|\tilde{C}\tilde{V}) - I(\tilde{V}\tilde{K}B; A|C\tilde{C}\tilde{U}) - I(\tilde{U}\tilde{K}; BC|\tilde{C}\tilde{V})\\
T_5 &\triangleq \frac{1}{2}\Big[I(AC; Y\tilde{Y}\tilde{A}|\tilde{C}\tilde{U}) - I(\tilde{V}\tilde{K}; AC|\tilde{C}\tilde{U}) + I(BC; Z\tilde{Z}\tilde{B}|\tilde{C}\tilde{V}) - I(\tilde{U}\tilde{K}; BC|\tilde{C}\tilde{V}) - I(A;B|C\tilde{C}\tilde{U}\tilde{V}\tilde{K})\Big]
\end{aligned}
$$

*Proof.* This theorem is obtained by replacing $K = (A, B, Y, Z)$ in Theoem 1 by $K = (A, B, S)$. The proof of this theorem is identical to that of Theorem 1, and is omitted.

# 5 Special Cases and Examples

In this section, we obtain a simpler version of the rate region of Theorem 1, and use it to compute achievable rates for two examples.

## 5.1 A Simpler Rate Region

**Corollary 5.1.** *Given a broadcast channel with noisy feedback $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{S}, P_{YZS|X})$, define any joint distribution $P$ of the form*

$$P_{C_0} P_{WUV} P_{X|WUVC_0} P_{YZS|X}. \tag{40}$$

*for some discrete random variables $W, U, V, C_0$. Let $(C_0, W, U, V, X, Y, Z, S)$ and $(\tilde{C}_0, \tilde{W}, \tilde{U}, \tilde{V}, \tilde{X}, \tilde{Y}, \tilde{Z}, \tilde{S})$ be two sets of variables each distributed according to $P$ and jointly distributed as*

$$P_{\tilde{C}_0\tilde{W}\tilde{U}\tilde{V}\tilde{X}\tilde{Y}\tilde{Z}\tilde{S}}\ Q_{C_0|\tilde{C}_0\tilde{W}\tilde{U}\tilde{V}\tilde{S}}\ P_{WUVXYZS|C_0}. \tag{41}$$

*where* $Q_{C_0|\tilde{C}_0\tilde{W}\tilde{U}\tilde{V}\tilde{S}}$ *is a distribution such that*

$$P_{C_0}(c_0) = \sum_{\tilde{c}_0,\tilde{w},\tilde{u},\tilde{v},\tilde{s}} Q_{C_0|\tilde{C}_0\tilde{W}\tilde{U}\tilde{V}\tilde{S}}(c_0|\tilde{c}_0,\tilde{w},\tilde{u},\tilde{v},\tilde{s})P(\tilde{c}_0,\tilde{w},\tilde{u},\tilde{v},\tilde{s}), \quad \forall c_0 \in \mathcal{C}_0. \tag{42}$$

*Then the following region is achievable.*

$$R_0 < \min\{T_1, T_2\} \tag{43}$$

$$R_0 + R_1 < I(UW;Y|C_0) + I(C_0;Y|\tilde{Y}\tilde{C}_0\tilde{W}) + I(C_0;\tilde{Y}|\tilde{C}_0\tilde{W}\tilde{U}) - I(\tilde{V}\tilde{S};C_0|\tilde{C}_0\tilde{W}\tilde{U}) \tag{44}$$

$$R_0 + R_2 < I(VW;Z|C_0) + I(C_0;Z|\tilde{Z}\tilde{C}_0\tilde{W}) + I(C_0;\tilde{Z}|\tilde{C}_0\tilde{W}\tilde{V}) - I(\tilde{U}\tilde{S};C_0|\tilde{C}_0\tilde{W}\tilde{V}) \tag{45}$$

$$\begin{aligned} R_0 + R_1 + R_2 < {} & I(UW;Y|C_0) + I(C_0;Y|\tilde{Y}\tilde{C}_0\tilde{W}) + I(C_0;\tilde{Y}|\tilde{C}_0\tilde{W}\tilde{U}) - I(\tilde{V}\tilde{S};C_0|\tilde{C}_0\tilde{W}\tilde{U}) \\ & + I(C_0;\tilde{V}|\tilde{C}_0\tilde{W}) + I(\tilde{V};\tilde{Z}|C_0\tilde{C}_0\tilde{W}) - I(U;V|W) \end{aligned} \tag{46}$$

$$\begin{aligned} R_0 + R_1 + R_2 < {} & I(VW;Z|C_0) + I(C_0;Z|\tilde{Z}\tilde{C}_0\tilde{W}) + I(C_0;\tilde{Z}|\tilde{C}_0\tilde{W}\tilde{V}) - I(\tilde{U}\tilde{S};C_0|\tilde{C}_0\tilde{W}\tilde{V}) \\ & + I(C_0;\tilde{U}|\tilde{C}_0\tilde{W}) + I(\tilde{U};\tilde{Y}|C_0\tilde{C}_0\tilde{W}) - I(U;V|W) \end{aligned} \tag{47}$$

$$\begin{aligned} 2R_0 + R_1 + R_2 < {} & I(UW;Y|C_0) + I(C_0;Y|\tilde{Y}\tilde{C}_0\tilde{W}) + I(C_0;\tilde{Y}|\tilde{C}_0\tilde{W}\tilde{U}) - I(\tilde{V}\tilde{S};C_0|\tilde{C}_0\tilde{W}\tilde{U}) \\ & + I(VW;Z|C_0) + I(C_0;Z|\tilde{Z}\tilde{C}_0\tilde{W}) + I(C_0;\tilde{Z}|\tilde{C}_0\tilde{W}\tilde{V}) - I(\tilde{U}\tilde{S};C_0|\tilde{C}_0\tilde{W}\tilde{V}) - I(U;V|W) \end{aligned} \tag{48}$$

*where*

$$T_1 \triangleq I(C_0;\tilde{Y}|\tilde{C}_0\tilde{W}\tilde{U}) + I(C_0W;Y|\tilde{Y}\tilde{C}_0\tilde{W}\tilde{U}) - I(\tilde{V}\tilde{S};C_0|\tilde{C}_0\tilde{W}\tilde{U})$$
$$T_2 \triangleq I(C_0;\tilde{Z}|\tilde{C}_0\tilde{W}\tilde{V}) + I(C_0W;Z|\tilde{Z}\tilde{C}_0\tilde{W}\tilde{V}) - I(\tilde{U}\tilde{S};C_0|\tilde{C}_0\tilde{W}\tilde{V})$$

*Proof.* In Theorem 2, set $A = B = \phi$, and $C = (C_0, W)$, with

$$Q_{C|\tilde{C}\tilde{U}\tilde{V}\tilde{S}} = Q_{C_0W|\tilde{C}_0\tilde{W}\tilde{U}\tilde{V}\tilde{S}} = P_W Q_{C_0|\tilde{C}_0\tilde{W}\tilde{U}\tilde{V}\tilde{S}}.$$

This choice of $Q_{C|\tilde{C}\tilde{U}\tilde{V}\tilde{S}} \in \mathcal{Q}(P)$ if (42) is satisfied. □

**Dueck's feedback example**: The rate region of Corollary 5.1 yields the optimal rates for the example described in Section 1. To see this, set

$$\begin{gathered} W = \phi, \quad (U,V) \sim P_{UV} = P_U P_V, \text{ with } P_U(0) = P_U(1) = P_V(0) = P_V(1) = \frac{1}{2} \\ P_{C_0}(0) = P_{C_0}(1) = \frac{1}{2} \\ X : (X_0 = C_0,\ X_1 = U,\ X_2 = V) \end{gathered} \tag{49}$$

We define the distribution $Q$ that generates $C_0$ for each block from the variables of the previous block as

$$Q : C_0 = \tilde{Y} \oplus \tilde{U} = \tilde{Z} \oplus \tilde{V} \tag{50}$$

Since $Y \oplus U = Z \oplus V = N$, the noise variable which is Bernoulli($\frac{1}{2}$), the above choice satisfies (42). Finally, substituting (49) in Corollary 5.1, the mutual information quantities are

$$\begin{aligned} I(V;Z|C_0) = I(U;Y|C_0) &= 0, & I(C_0;Y|\tilde{Y}\tilde{C}_0) &= I(C_0;Z|\tilde{Z}\tilde{C}_0) = 1 \\ I(C_0;\tilde{Y}|\tilde{C}_0\tilde{U}) = I(C_0;\tilde{Z}|\tilde{C}_0\tilde{V}) &= 1, & I(\tilde{V}\tilde{Y}\tilde{Z};C_0|\tilde{C}_0\tilde{U}) &= I(\tilde{U}\tilde{Y}\tilde{Z};C_0|\tilde{C}_0\tilde{V}) = 1 \\ I(C_0;\tilde{V}|\tilde{C}_0) = I(C_0;\tilde{U}|\tilde{C}_0) &= 1 & I(\tilde{U};\tilde{Y}|C_0\tilde{C}_0) &= I(\tilde{V};\tilde{Z}|C_0\tilde{C}_0) = 1 \\ I(C_0;\tilde{Y}|\tilde{C}_0\tilde{U}) = I(C_0;\tilde{Z}|\tilde{C}_0\tilde{U}) &= 1, & I(C_0;Y|\tilde{Y}\tilde{C}_0\tilde{U}) &= I(C_0;Z|\tilde{Z}\tilde{C}_0\tilde{V}) = 0. \end{aligned}$$

Using these, we see that a rate of $R_1 = R_2 = 1$ is achievable.

## 5.2 The AWGN Broadcast Channel with Feedback

In this subsection, we compute the rate region of Corollary 5.1 for the scalar AWGN broadcast channel with noiseless feedback and average power constraint $P$. We compare the obtained sum rate with: a) the maximum sum rate in the absence of feedback, and b) the achievable region for the AWGN broadcast channel with noiseless feedback obtained by Ozarow and Leung in [5] using a generalization of the Schalkwijk-Kailath coding scheme [17].

The channel, with $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathbb{R}$, is described by

$$Y = X + N_1, \quad Z = X + N_2, \tag{51}$$

where $N_1, N_2$ are Gaussian noise variables with zero mean and unit variance. $N_1$ and $N_2$ are independent of one another as well as the channel input $X$. The input sequence $\mathbf{x}$ for each block satisfies $\sum_{i=1}^{n} x_i^2 \leq P$.

In the absence of feedback, the capacity region of the AWGN broadcast channel is known [18, 19] and can be obtained from Marton's inner bound using the following choice of random variables.

$$V = \sqrt{\bar{\alpha}P} Q_2, \quad U = \sqrt{\alpha P} Q_1 + \frac{\alpha P}{\alpha P + \sigma^2} V$$

where $\alpha \in (0, 1)$, and $Q_1, Q_2$ are independent Gaussian variables with zero mean and unit variance. The Marton sum rate is then given by

$$R_{\text{no-FB}} = I(V;Z) + I(U;Y) - I(U;V) = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma^2}\right). \tag{52}$$

This is essentially the 'writing on dirty paper' coding strategy [20, 21]: for the channel from $U$ to $Y$, $V$ can be considered as channel state information known at the encoder. We note that an alternate way of achieving the no-feedback capacity region of the AWGN broadcast channel is through superposition coding [2].

Using Corollary 5.1, we now compute an achievable region for the channel (51) with noiseless feedback.[2] The joint distribution $P_{C_0} P_{UV} P_{X|C_0UV}$ is chosen as

$$V = \sqrt{\bar{\alpha}P_1} Q_2, \quad U = \sqrt{\alpha P_1} Q_1 + \beta V \tag{53}$$

$$X = \sqrt{P - P_1} C_0 + \sqrt{\bar{\alpha}P_1} Q_2 + \sqrt{\alpha P_1} Q_1 \tag{54}$$

[2] Theorems 1 and 2 were established for a discrete memoryless broadcast channel with feedback. These theorems can be extended to the AWGN broadcast channel using a similar proof, recognizing that in the Gaussian case superposition is equivalent to addition.

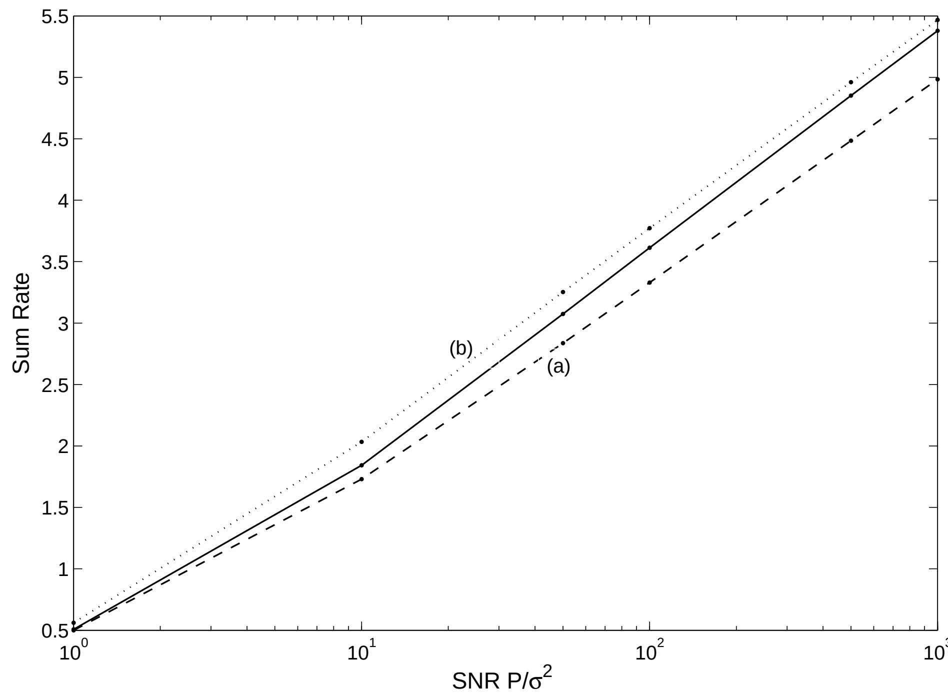


Figure 5: Achievable sum rates for the AWGN broadcast channel. The solid line shows the sum rate computed using Corollary 5.1. The dashed line (a) is the no-feedback sum rate, and the dotted line (b) is the sum rate of the Ozarow-Leung scheme.

where $C_0, Q_1, Q_2$ are independent Gaussians with zero mean and unit variance, and $\alpha, \beta \in (0, 1)$, $P_1 \in (0, P)$ are parameters to optimized later.

Next we define a conditional distribution $Q_{C_0|\tilde{C}_0\tilde{U}\tilde{V}\tilde{Y}\tilde{Z}}$ that satisfies (42). Let

$$\tilde{S}_1 = \frac{\tilde{U} - E[\tilde{U}|\tilde{Y}\tilde{C}_0]}{\sqrt{E[(\tilde{U} - E[\tilde{U}|\tilde{Y}\tilde{C}_0])^2]}}. \tag{55}$$

Then define $Q_{C_0|\tilde{C}_0\tilde{U}\tilde{V}\tilde{Y}\tilde{Z}}$ by the relation

$$C_0 = \sqrt{1 - D}\tilde{S}_1 + \eta \tag{56}$$

where $\eta$ is a Gaussian random variable with zero mean and variance $D$ independent of $S_1$.

In words, $\tilde{S}_1$ is the normalized error in the estimate of $\tilde{U}$ at receiver 1. This estimation error is quantized at distortion level $D$ and suitably scaled to obtain $C_0$. Thus, in each block, $C_0$ represents a quantized version of the estimation error at receiver 1 in the previous block. If we similarly denote by $\tilde{S}_2$ the error in the estimate of $\tilde{V}$ at receiver 2, then $\tilde{S}_2$ is correlated with $\tilde{S}_1$. Therefore, $C_0$ simultaneously plays the role of conveying information about $\tilde{S}_2$ to receiver 2. With the above choice of joint distribution, the information quantities in Corollary 5.1 are computed and listed in Appendix A.

For different values of the signal-to-noise ratio $P/\sigma^2$, we then numerically compute the maximum sum rate by optimizing over the parameters $(\alpha, \beta, D, P_1)$. This is plotted in Figure 5. The figure also shows the sum rate in the absence of feedback and the maximum sum rate of the Ozarow-Leung scheme with noiseless feedback. We observe that the obtained sum rate is higher than the sum rate without feedback, but lower than that of the Ozarow-Leung scheme. However, we emphasize that the Ozarow-Leung coding scheme is specific to the AWGN broadcast channel and does not extend to other discrete memoryless broadcast channels, unlike the rate region proposed in this paper.

For the AWGN broadcast channel with noisy feedback, a similar joint distribution can be used to obtain achievable rates outside the no-feedback capacity region. In this case, the variable $\tilde{S}_1$ should be defined as the encoder's *estimate* of the error at receiver 1. This is then quantized to $C_0$, and sent to both receivers in the subsequent block.

# 6 Comparison with the Shayevitz-Wigger (S-W) Rate Region

In [13], the following achievable rate region was proposed for a broadcast channel with noisy feedback.

**Fact 1.** ***S-W Region***. *Consider a broadcast channel with noisy feedback defined by $P_{YZS|X}$ , where $S$ denotes the noisy feedback signal available to the transmitter. Let $U_0, U_1, U_2, V_0, V_1, V_2$ be discrete auxiliary random variables jointly distributed according to*

$$P_{U_0U_1U_2}\ P_{X|U_0U_1U_2}\ P_{YZS|X}\ P_{V_0V_1V_2|U_0U_1U_2S}. \tag{57}$$

*Then the following rate region is achievable.*

$$R_0 + \Theta_1 < \Theta_2 \tag{58}$$

$$R_0 + R_1 < I(U_0U_1; YV_1) - I(U_0U_1U_2S; V_1|V_0Y) - \Theta_1 \tag{59}$$

$$R_0 + R_2 < I(U_0U_2; ZV_2) - I(U_0U_1U_2S; V_2|V_0Z) - \Theta_1 \tag{60}$$

$$\begin{aligned} R_0 + R_1 + R_2 < I(U_1; YV_1|U_0) + I(U_2; ZV_2|U_0) - I(U_1; U_2|U_0) + \Theta_2 - \Theta_1 \\ - I(U_0U_1U_2S; V_1|V_0Y) - I(U_0U_1U_2S; V_2|V_0Z) \end{aligned} \tag{61}$$

*where*

$$\Theta_1 \triangleq \max\{I(U_0U_1U_2S; V_0|Y), I(U_0U_1U_2S; V_0|Z)\}, \qquad \Theta_2 \triangleq \min\{I(U_0; YV_1), I(U_0; ZV_2)\}. \tag{62}$$

***Remark*** 1: There appears to be an error in the derivation of the rate region in [13]. In particular, there is an inequality missing in the set of constraints obtained by performing Fourier-Motzkin elimination on equation (35) in [13]. We have communicated this to the authors. The correct version of the S-W rate region is given above.

***Remark*** 2: The joint distribution of $(V_0, V_1, V_2)$ conditioned on all the other variables can be written as

$$P_{V_0V_1V_2|U_0U_1U_2S} = P_{V_0|U_0U_1U_2S} \cdot P_{V_1V_2|V_0U_0U_1U_2S}. \tag{63}$$

Now define another conditional distribution $P'_{V_0V_1V_2|U_0U_1U_2S}$ as

$$P'_{V_0V_1V_2|U_0U_1U_2S} = P_{V_0|U_0U_1U_2S} \cdot P_{V_1|V_0U_0U_1U_2S} \cdot P_{V_2|V_0U_0U_1U_2S} \tag{64}$$

where we have used the marginals from (63) to define $P'$. We observe that if we replace $P_{V_0V_1V_2|U_0U_1U_2S}$ by $P'_{V_0V_1V_2|U_0U_1U_2S}$ in (57), the Shayevitz-Wigger region remains the same because at most one of $V_1$ and $V_2$ appears in each of the mutual information terms. Hence, without loss of generality, we can assume that the joint distribution in (57) is of the form

$$P_{U_0U_1U_2}\ P_{X|U_0U_1U_2}\ P_{YZS|X}\ P_{V_0|U_0U_1U_2S}\ P_{V_1|V_0U_0U_1U_2S}\ P_{V_2|V_0U_0U_1U_2S}. \tag{65}$$

The S-W rate region is obtained using a block-Markov superposition coding scheme with several blocks of transmission. The rates of the message are too high for the receivers to decode them directly at the end of each block. At the end of the block, the encoder is therefore left with the task of transmitting a triple of correlated sequences $(\mathbf{U_0}, \mathbf{U_1}, \mathbf{U_2})$ with side-information $\mathbf{Y}$ and $\mathbf{Z}$ available at decoders 1 and 2, respectively. These correlated 'sources' at the encoder are covered using three random variables $(V_0, V_1, V_2)$, which are transmitted in the following block using *separate* source-channel coding. Receiver 1 first recovers $(V_0, V_1)$ and then uses backward decoding with $V_1$ as an additional output to decode its message from the previous block. Receiver 2 decodes in a similar fashion using $V_2$ as an additional output.

We now evaluate the rate region of Theorem 2 with a choice of random variables that results in the joint distribution (65). We show that the resulting rate constraints on $R_0, R_0 + R_1, R_0 + R_2$, and $2R_0 + R_1 + R_2$ are implied by the S-W rate constraints. Further, the $R_0 + R_1 + R_2$ constraints are also implied by the S-W constraints as long as a couple of conditions on the mutual information terms hold.

**Proposition 6.1.** *Consider any joint distribution of the form* (65) *for which the following inequalities hold.*

$$\begin{aligned} &I(V_0; U_0U_1|V_1Y) + I(V_2; U_2U_0|V_0Z) > I(V_2; U_2|U_0Z), \\ &I(V_0; U_0U_2|V_2Z) + I(V_1; U_1U_0|V_0Y) > I(V_1; U_1|U_0Y) \end{aligned} \tag{66}$$

*The S-W region evaluated with such a joint distribution is contained in the rate region of Theorem 2.*

*Proof.* For a given broadcast channel with noisy feedback $P_{YZS|X}$, pick any joint distribution of the form (65). We will evaluate Theorem 2 by identifying distributions $P$ and $Q \in \mathcal{Q}(P)$ such that the joint distribution coincides with the one in (65). Consider the following choice of random variables. Pick $A = (\tilde{V}_1, U_1)$, $B = (\tilde{V}_2, U_2)$, $C = (\tilde{V}_0, U_0)$, and $U = V = \phi$, where $(U_0, U_1, U_2)$ are jointly distributed according to $P_{U_0U_1U_2}$, and $(\tilde{V}_0, \tilde{V}_1, \tilde{V}_2)$ are jointly distributed according to $P_{V_0V_1V_2}$, the marginals from (65). Further, $(\tilde{V}_0, \tilde{V}_1, \tilde{V}_2)$ is independent of $(U_0, U_1, U_2)$. The input $X$ is generated according to $P_{X|U_0U_1U_2}$.

We choose a conditional distribution in $\mathcal{Q}(P)$ to generate $(\mathbf{A}, \mathbf{B}, \mathbf{C})_{l+1}$ from the information $(\mathbf{U}, \mathbf{V}, \mathbf{C}, \mathbf{K})_l$ as follows.

$$Q_{CAB|\tilde{U}\tilde{V}\tilde{C}\tilde{K}} = Q_{CAB|\tilde{A}\tilde{B}\tilde{C}\tilde{S}} = Q_{\tilde{V}_0U_0\tilde{V}_1U_1\tilde{V}_2U_2|\tilde{\tilde{V}}_1\tilde{U}_1\tilde{\tilde{V}}_2\tilde{U}_2\tilde{\tilde{V}}_0\tilde{U}_0\tilde{S}} \triangleq P_{U_0U_1U_2}P_{\tilde{V}_0\tilde{V}_1\tilde{V}_2|\tilde{U}_0\tilde{U}_1\tilde{U}_2\tilde{S}}.$$

where $P_{\tilde{V}_0\tilde{V}_1\tilde{V}_2|\tilde{U}_0\tilde{U}_1\tilde{U}_2\tilde{S}} = P_{V_0V_1V_2|U_0U_1U_2S}$, the marginal from (65).

To summarize, the joint distribution over two successive blocks with $C = (\tilde{V}_0, U_0), A = (\tilde{V}_1, U_1), B =$

$(\tilde{V}_2, U_2)$, and $U = V = \phi$ is

$$P_{\tilde{\tilde{V}}_0\tilde{\tilde{V}}_1\tilde{\tilde{V}}_2} P_{\tilde{U}_0\tilde{U}_1\tilde{U}_2} P_{\tilde{X}|\tilde{U}_0\tilde{U}_1\tilde{U}_2} P_{\tilde{Y}\tilde{Z}|\tilde{X}} \; P_{\tilde{V}_0\tilde{V}_1\tilde{V}_2|\tilde{U}_0\tilde{U}_1\tilde{U}_2\tilde{S}} \; P_{U_0U_1U_2} P_{X|U_0U_1U_2} P_{YZ|X} P_{V_0V_1V_2|U_0U_1U_2S}. \tag{67}$$

Therefore, we have ensured that the joint distribution of the sequences $(\mathbf{U}_0, \mathbf{U_1}, \mathbf{U}_2, \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{S}, \mathbf{V_0}, \mathbf{V}_1, \mathbf{V}_2)$ is the same as that used for the Shayevitz-Wigger region in (65). Substituting $C = (\tilde{V}_0, U_0), A = (\tilde{V}_1, U_1), B = (\tilde{V}_2, U_2)$, and $U = V = \phi$, the first constraint of Theorem 2 becomes

$$\begin{aligned} R_0 + R_1 &< I(U_0U_1\tilde{V}_1\tilde{V}_0; Y\tilde{Y}\tilde{U}_1|\tilde{U}_0) - I(\tilde{U}_1\tilde{U}_2\tilde{S}; \tilde{V}_1\tilde{V}_0|\tilde{U}_0) \\ &\overset{(a)}{=} I(U_0U_1; Y) + I(\tilde{V}_0\tilde{V}_1; \tilde{Y}\tilde{U}_1|\tilde{U}_0) - I(\tilde{U}_1\tilde{U}_2\tilde{S}; \tilde{V}_1\tilde{V}_0|\tilde{U}_0) \\ &\overset{(b)}{=} I(U_0U_1; Y) + I(V_0V_1; YU_1|U_0) - I(U_0U_1U_2S; V_1V_0) + I(V_1V_0; U_0) \\ &\overset{(c)}{=} I(U_0U_1; Y) + I(V_0V_1; YU_1U_0) - I(U_0U_1U_2SY; V_1V_0) \\ &= I(U_0U_1; Y) + I(V_0V_1; U_1U_0|Y) - I(U_0U_1U_2S; V_1V_0|Y). \end{aligned} \tag{68}$$

In the above, $(a)$ is due to the fact that $Y$ is independent of $(\tilde{U}_0, \tilde{U}_1, \tilde{Y}, \tilde{V}_0, \tilde{V}_1)$, as can be seen from (67). $(b)$ is true because $(\tilde{U_0}, \tilde{U_1}, \tilde{U_2}, \tilde{Y}, \tilde{Z}, \tilde{S}, \tilde{V}_0, \tilde{V}_1, \tilde{V}_2)$ has the same joint distribution as $(U_0, U_1, U_2, Y, Z, S, V_0, V_1, V_2)$. $(c)$ holds because $I(Y; V_1V_0|U_0U_1U_2S) = 0$, since $Y - (U_0, U_1, U_2, S) - (V_1, V_0)$ form a Markov chain. The other rate constraints of Theorem 2 can be similarly evaluated, and are given below.

$$R_0 + R_1 < I(U_0U_1; Y) + I(V_0V_1; U_1U_0|Y) - I(U_0U_1U_2S; V_1V_0|Y) \tag{69}$$

$$R_0 + R_2 < I(U_0U_2; Z) + I(V_0V_2; U_2U_0|Z) - I(U_0U_1U_2S; V_2V_0|Z) \tag{70}$$

$$\begin{aligned} R_0 + R_1 + R_2 &< I(U_0U_1; Y) + I(U_2; Z|U_0) + I(V_0V_1; U_1U_0|Y) + I(V_2; U_2U_0|V_0Z) \\ &\quad - I(U_0U_1U_2S; V_1V_0|Y) - I(U_0U_1U_2S; V_2|V_0Z) - I(U_1; U_2|U_0) \end{aligned} \tag{71}$$

$$\begin{aligned} R_0 + R_1 + R_2 &< I(U_1; Y|U_0) + I(U_0U_2; Z) + I(V_1; U_1U_0|V_0Y) + I(V_0V_2; U_2U_0|Z) \\ &\quad - I(U_0U_1U_2S; V_1|V_0Y) - I(U_0U_1U_2S; V_2V_0|Z) - I(U_1; U_2|U_0) \end{aligned} \tag{72}$$

$$\begin{aligned} 2R_0 + R_1 + R_2 &< I(U_0U_1; Y) + I(U_0U_2; Z) + I(V_0V_1; U_1U_0|Y) + I(V_0V_2; U_2U_0|Z) \\ &\quad - I(U_0U_1U_2S; V_1V_0|Y) - I(U_0U_1U_2S; V_2V_0|Z) - I(U_1; U_2|U_0) \end{aligned} \tag{73}$$

(The $R_0$ constraint of Theorem 2 does not appear because it subsumed by the above constraints.) We now have the following lemma.

**Lemma 6.1.** *Given a joint distribution $P_{WUV} P_{X|WUV} P_{YZS|X} P_{V_0|WUVS} P_{V_1|V_0WUVS} P_{V_2|V_0WUVS}$, the following statements hold.*

1. (69) *is implied by* (59).
2. (70) *is implied by* (60).
3. (73) *is implied by* (61) *and* (58).
4. (71) *and* (72) *are implied by* (61) *if the inequalities in* (66) *hold.*

The proof of the lemma is given in Appendix B. This completes the proof that the rate region of Theorem 2 contains the S-W achievable region for the broadcast channel with noisy/noiseless feedback. $\square$

# 7 Proof of Theorem 1

## 7.1 Preliminaries

We shall use the notion of strong typicality as defined in [22]. Consider three finite sets $\mathcal{V}, \mathcal{Z}_1$ and $\mathcal{Z}_2$, and an arbitrary distribution $P_{VZ_1Z_2}$ on them.

**Definition 7.1.** *For any distribution $P_V$ on $\mathcal{V}$, a sequence $v^n \in \mathcal{V}^n$ is said to be $\epsilon$-typical with respect to $P_V$, if*

$$\left| \frac{1}{n} \#(a|v^n) - P_V(a) \right| \leq \frac{\epsilon}{|\mathcal{V}|},$$

*for all $a \in \mathcal{V}$, and no $a \in \mathcal{V}$ with $P_V(a) = 0$ occurs in $v^n$, where $\#(a|v^n)$ denotes the number of occurrences of $a$ in $v^n$. Let $A_\epsilon^{(n)}(P_V)$ denote the set of all sequences that are $\epsilon$-typical with respect to $P_V$.*

The following are some of the properties of typical sequences that will be used in the proof.

**Property 0:** For all $\epsilon > 0$, and for all sufficiently large $n$, we have $P_V^n[A_\epsilon^{(n)}(P_V)] > 1 - \epsilon$.

**Property 1:** Let $v^n \in A_\epsilon^{(n)}(P_V)$ for some fixed $\epsilon > 0$. If a random vector $Z_1^n$ is generated from the product distribution $\prod_{i=1}^n P_{Z_1|V}(\cdot|v_i)$, then for all sufficiently large $n$, we have $Pr[(v^n, Z_1^n) \notin A_{\tilde{\epsilon}}^{(n)}(P_{VZ_1})] < \epsilon$, where $\tilde{\epsilon} = \epsilon(|\mathcal{V}| + |\mathcal{Z}_1|)$.

**Property 2:** Let $v^n \in A_\epsilon^{(n)}(P_V)$ for some fixed $\epsilon > 0$. If a random vector $Z_1^n$ is generated from the product distribution $\prod_{i=1}^n P_{Z_1|V}(\cdot|v_i)$ and $Z_2^n$ is generated from the product distribution $\prod_{i=1}^n P_{Z_2|V}(\cdot|v_i)$, then for all sufficiently large $n$, we have

$$\frac{2^{-n\delta(\epsilon)}\ 2^{nH(Z_1Z_2|V)}}{2^{nH(Z_1|V)}\ 2^{nH(Z_2|V)}} < Pr[(v^n, Z_1^n, Z_2^n) \in A_{\tilde{\epsilon}}^{(n)}(P_{VZ_1Z_2})] < \frac{2^{n\delta(\epsilon)}\ 2^{nH(Z_1Z_2|V)}}{2^{nH(Z_1|V)}\ 2^{nH(Z_2|V)}},$$

where $\tilde{\epsilon} = \epsilon(|\mathcal{V}| + |\mathcal{Z}_1||\mathcal{Z}_2|)$, and $\delta(\epsilon)$ is a continuous positive function of $\epsilon$ that goes to 0 as $\epsilon \to 0$.

## 7.2 Random Codebook Generation

Fix a distribution $P_{UVABCXYZ}$ from $\mathcal{P}$ and a conditional distribution $Q_{ABC|\tilde{U}\tilde{V}\tilde{K}\tilde{C}}$ satisfying (6), as required by the statement of the theorem. Fix a positive integer $L$. There are $L$ blocks in encoding and decoding. Fix positive real numbers $R_1', R_2'$, $R_0, R_1, R_2$, $\rho_0$, $\rho_1$ and $\rho_2$ such that $R_1' > R_1$ and $R_2' > R_2$, where these numbers denote the rates of codebooks to be constructed as described below. Fix $\epsilon > 0$ and a positive integer $n$. $n$ denotes the block length. Recall that $K$ denotes the collection $(A, B, Y, Z)$, and $\mathcal{K}$ denotes the set $\mathcal{A} \times \mathcal{B} \times \mathcal{Y} \times \mathcal{Z}$. Let $\epsilon[l] = \epsilon(4|\mathcal{K}|^2|\mathcal{U}|^2|\mathcal{V}|^2|\mathcal{C}|^2)^l$ for $l = 0, 1, 2, \ldots, L$.
For $l = 1, 2, 3, \ldots, L$ independently perform the following random experiments.

- For each sequence $\mathbf{c} \in \mathcal{C}^n$, generate $2^{n(R_1' - R_1)}$ sequences $\mathbf{U}_{[l,i,\mathbf{c}]}$, $i = 1, 2, \ldots, 2^{n(R_1' - R_1)}$, independently where each sequence is generated from the product distribution $\prod_{i=1}^n P_{U|C}(\cdot|c_i)$. Call this the first $U$-bin. Independently repeat this experiment $2^{nR_1}$ times to generate $2^{nR_1}$ $U$-bins, and a total of $2^{nR_1'}$ sequences. The $i$th sequence in the $j$th bin is $\mathbf{U}_{[l,(j-1)2^{nR_1}+i,\mathbf{c}]}$.

- For each sequence $\mathbf{c} \in \mathcal{C}^n$, similarly generate $2^{nR_2}$ $V$-bins each containing $2^{n(R_1'-R_1)}$ sequences with each sequence being generated from the product distribution $\prod_{i=1}^n P_{V|C}(\cdot|c_i)$. The $i$th sequence in the $j$th bin is $\mathbf{V}_{[l,(j-1)2^{nR_2}+i,\mathbf{c}]}$.
- For each sequence $\tilde{\mathbf{c}} \in \mathcal{C}^n$, generate $2^{n\rho_0}$ sequences $\mathbf{C}_{[l,i,\tilde{\mathbf{c}}]}$, $i = 1, 2, \ldots, 2^{n\rho_0}$, independently where each sequence is generated from $\prod_{i=1}^n P_{C|\tilde{C}}(\cdot|\tilde{c}_i)$.
- For each $(\tilde{\mathbf{u}}, \tilde{\mathbf{c}}, \mathbf{c}) \in \mathcal{U}^n \times \mathcal{C}^{2n}$ generate independently $2^{n\rho_1}$ sequences $\mathbf{A}_{[l,i,\tilde{\mathbf{u}},\tilde{\mathbf{c}},\mathbf{c}]}$, for $i = 1, 2, \ldots, 2^{n\rho_1}$, where each sequence is generated from $\prod_{j=1}^n P_{A|\tilde{U}\tilde{C}C}(\cdot|\tilde{u}_j, \tilde{c}_j, c_j)$.
- For each $(\tilde{\mathbf{v}}, \tilde{\mathbf{c}}, \mathbf{c}) \in \mathcal{V}^n \times \mathcal{C}^{2n}$ generate independently $2^{n\rho_2}$ sequences $\mathbf{B}_{[l,i,\tilde{\mathbf{v}},\tilde{\mathbf{c}},\mathbf{c}]}$, for $i = 1, 2, \ldots, 2^{n\rho_2}$, where each sequence is generated from $\prod_{j=1}^n P_{B|\tilde{V}\tilde{C}C}(\cdot|\tilde{v}_j, \tilde{c}_j, c_j)$.
- For each $(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{u}, \mathbf{v}) \in \mathcal{A}^n \times \mathcal{B}^n \times \mathcal{C}^n \times \mathcal{U}^n \times \mathcal{V}^n$ generate one sequence $\mathbf{X}_{[l,\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{u},\mathbf{v}]}$ using $\prod_{i=1}^n P_{X|ABCUV}(\cdot|a_i, b_i, c_i, u_i, v_i)$.

Generate independently a tuple of sequences $(\mathbf{U}[0], \mathbf{V}[0], \mathbf{K}[0], \mathbf{C}[0])$ from the product distribution $P^n_{U,V,K,C}$. These sequences are known to all terminals before transmission begins.

## 7.3 Encoding Operation

Let $W_0[l]$ denote the common message, and $W_1[l], W_2[l]$, the private messages for block $l$. These are independent random variables distributed uniformly over $\{1, 2, \ldots, 2^{nR_0}\}$, $\{1, 2, \ldots, 2^{nR_1}\}$, and $\{1, 2, \ldots, 2^{nR_2}\}$, respectively. We set $W_0[0] = W_1[0] = W_2[0] = W_0[L] = W_1[L] = W_2[L] = 1$. For each block $l$, the encoder chooses a quintuple of sequences $(\mathbf{A}[l], \mathbf{B}[l], \mathbf{C}[l], \mathbf{U}[l], \mathbf{V}[l])$ from the five codebooks generated above, according to the encoding rule described below. The channel input, and channel output sequences in block $l$ are denoted $\mathbf{X}[l]$, $\mathbf{Y}[l]$ and $\mathbf{Z}[l]$, respectively.
**Blocks** $l = 1, 2, 3, \ldots, L$ : The encoder performs the following sequence of operations.

- Step 1: The encoder chooses a triple of indices $(G_0[l], G_1[l], G_2[l])$ such that

$$G_0[l] \mod 2^{nR_0} = W_0[l],$$

and the tuple $(\mathbf{U}[l-1], \mathbf{V}[l-1], \mathbf{K}[l-1], \mathbf{C}[l-1])$ is jointly $\epsilon[l]$-typical with the triple of sequences

$$(\mathbf{C}_{[l,G_0[l],\mathbf{C}[l-1]]}, \mathbf{A}_{[l,G_1[l],\mathbf{U}[l-1],\mathbf{C}[l-1],\mathbf{C}_{[l,G_0[l],\mathbf{C}[l-1]]}]}, \mathbf{B}_{[l,G_2[l],\mathbf{V}[l-1],\mathbf{C}[l-1],\mathbf{C}_{[l,G_0[l],\mathbf{C}[l-1]]}]}),$$

with respect to $P_{\tilde{U},\tilde{V},\tilde{K},\tilde{C},C,A,B}$. If no such index triple is found, it declares error and sets $(G_0[l], G_1[l], G_2[l]) = (1, 1, 1)$.

The encoder then sets

$$\mathbf{C}[l] = \mathbf{C}_{[l,G_0[l],\mathbf{C}[l-1]]}, \quad \mathbf{A}[l] = \mathbf{A}_{[l,G_1[l],\mathbf{U}[l-1],\mathbf{C}[l-1],\mathbf{C}_{[l,G_0[l],\mathbf{C}[l-1]]}]}, \quad \mathbf{B}[l] = \mathbf{B}_{[l,G_2[l],\mathbf{V}[l-1],\mathbf{C}[l-1],\mathbf{C}_{[l,G_0[l],\mathbf{C}[l-1]]}]}.$$

- Step 2: The encoder chooses a pair of indices $(G_3[l], G_4[l])$ such that the triple of sequences

$$(\mathbf{U}_{[l,G_3[l],\mathbf{C}[l]]}, \mathbf{V}_{[l,G_4[l],\mathbf{C}[l]]}, \mathbf{C}[l])$$

is $\epsilon$-typical with respect to $P_{UVC}$, and $\mathbf{U}_{[l,G_3[l],\mathbf{C}[l]]}$ belongs to the $U$-bin with index $W_1[l]$, and $\mathbf{V}_{[l,G_4[l],\mathbf{C}[l]]}$ belongs to the $V$-bin with index $W_2[l]$. If no such index pair is found, it declares error and sets $(G_3[l], G_4[l]) = (1,1)$.

The encoder then sets $\mathbf{U}[l] = \mathbf{U}_{[l,G_3[l],\mathbf{C}[l]]}$, $\mathbf{V}[l] = \mathbf{V}_{[l,G_4[l],\mathbf{C}[l]]}$, and $\mathbf{X}[l] = \mathbf{X}_{[l,\mathbf{A}[l],\mathbf{B}[l],\mathbf{C}[l],\mathbf{U}[l],\mathbf{V}[l]]}$. The encoder sends $\mathbf{X}[l]$ as the channel input sequence for block $l$.

- Step 3: The broadcast channel produces $(\mathbf{Y}[l], \mathbf{Z}[l])$.
- Step 4: After receiving $(\mathbf{Y}[l], \mathbf{Z}[l])$ via the feedback link, the encoder sets $\mathbf{K}[l] = (\mathbf{A}[l], \mathbf{B}[l], \mathbf{Y}[l], \mathbf{Z}[l])$.

## 7.4 Decoding Operation

**Block 1:**

- The first decoder receives $\mathbf{Y}[1]$, and the second decoder receives $\mathbf{Z}[1]$.
- The first decoder determines the unique index pair $(\hat{G}_{01}[1], \hat{G}_1[1])$ such that the tuples

$$(\mathbf{C}[0], \mathbf{A}[0], \mathbf{U}[0], \mathbf{Y}[0]) \text{ and } (\bar{\mathbf{C}}_1[1], \mathbf{A}_{[1,\hat{G}_1[1],\mathbf{U}[0],\mathbf{C}[0],\bar{\mathbf{C}}_1[1]]}, \mathbf{Y}[1])$$

are jointly $\epsilon[l]$-typical with respect to $P_{\tilde{C}\tilde{A}\tilde{U}\tilde{Y}CAY}$, where $\bar{\mathbf{C}}_1[1] \triangleq \mathbf{C}_{[1,\hat{G}_{01}[1],\mathbf{C}[0]]}$. Note that $\bar{\mathbf{C}}_1[1]$ is the estimate of $\mathbf{C}[1]$ at the first decoder.

If not successful in this operation, the first decoder declares an error and sets $(\hat{G}_{01}[1], \hat{G}_1[1]) = (1,1)$, and $\bar{\mathbf{C}}_1[1] \triangleq \mathbf{C}_{[1,\hat{G}_{01}[1],\mathbf{C}[0]]}$.

- The first decoder outputs $\hat{W}_0[1] = \hat{G}_{01}[1] \mod 2^{nR_0}$, and sets

$$\bar{\mathbf{A}}[1] = \mathbf{A}_{[1,\hat{G}_1[1],\mathbf{U}[0],\mathbf{C}[0],\bar{\mathbf{C}}_1[1]]}.$$

$\bar{\mathbf{A}}[1]$ is the first decoder's estimate of $\mathbf{A}[1]$.

- The second decoder determines the unique index pair $(\hat{G}_{02}[1], \hat{G}_2[1])$ such that the tuples

$$(\mathbf{C}[0], \mathbf{B}[0], \mathbf{V}[0], \mathbf{Z}[0]) \text{ and } (\bar{\mathbf{C}}_2[1], \mathbf{B}_{[1,\hat{G}_2[1],\mathbf{V}[0],\mathbf{C}[0],\bar{\mathbf{C}}_2[1]]}), \mathbf{Z}[1])$$

are jointly $\epsilon[l]$-typical with respect to $P_{\tilde{C}\tilde{B}\tilde{V}\tilde{Z}CBZ}$, where $\bar{\mathbf{C}}_2[1] \triangleq \mathbf{C}_{[1,\hat{G}_{02}[1],\mathbf{C}[0]]}$. Note that $\bar{\mathbf{C}}_2[1]$ is the estimate of $\mathbf{C}[1]$, at the second decoder.

If not successful in this operation, the second decoder declares an error and sets $(\hat{G}_{02}[1], \hat{G}_2[1]) = (1,1)$, and $\bar{\mathbf{C}}_2[1] \triangleq \mathbf{C}_{[1,\hat{G}_{02}[1],\mathbf{C}[0]]}$.

- The second decoder outputs $\bar{W}_0[1] = \hat{G}_{02}[1] \mod 2^{nR_0}$, and sets

$$\bar{\mathbf{B}}[1] = \mathbf{B}_{[1,\hat{G}_2[1],\mathbf{V}[0],\mathbf{C}[0],\bar{\mathbf{C}}_2[1]]}.$$

$\bar{\mathbf{B}}[1]$ is the second decoder's estimate of $\mathbf{B}[1]$.

**Block** $l, l = 2, 3, \ldots, L$:

- The first decoder receives $\mathbf{Y}[l]$ and the second decoder receives $\mathbf{Z}[l]$.
- The first decoder determines the unique index triple $(\hat{G}_{01}[l], \hat{G}_1[l], \hat{G}_3[l-1])$ such that the tuples

$$(\bar{\mathbf{C}}_1[l-1], \bar{\mathbf{A}}[l-1], \bar{\mathbf{U}}[l-1], \mathbf{Y}[l-1]) \text{ and } (\bar{\mathbf{C}}_1[l], \mathbf{A}_{[l,\hat{G}_1[l],\bar{\mathbf{U}}[l-1],\bar{\mathbf{C}}_1[l-1],\bar{\mathbf{C}}_1[l]]}, \mathbf{Y}[l])$$

are jointly $\epsilon[l]$-typical with respect to $P_{\tilde{C}\tilde{A}\tilde{U}\tilde{Y}CAY}$, where

$$\bar{\mathbf{U}}[l-1] \triangleq \mathbf{U}_{[(l-1),\hat{G}_3[l-1],\bar{\mathbf{C}}_1[l-1]]}, \quad \bar{\mathbf{C}}_1[l] \triangleq \mathbf{C}_{[l,\hat{G}_{01}[l],\bar{\mathbf{C}}_1[l-1]]}.$$

If not successful in this operation, the first decoder declares an error and sets $(\hat{G}_{01}[l], \hat{G}_1[l], \hat{G}_3[l-1]) = (1,1,1)$, and

$$\bar{\mathbf{U}}[l-1] = \mathbf{U}_{[(l-1),1,\bar{\mathbf{C}}_1[l-1]]}, \quad \bar{\mathbf{C}}_1[l] \triangleq \mathbf{C}_{[l,1,\bar{\mathbf{C}}_1[l-1]]}.$$

Note that $\bar{\mathbf{U}}[l-1]$ and $\bar{\mathbf{C}}_1[l]$ are the estimates of $\mathbf{U}[l-1]$ and $\mathbf{C}[l]$, respectively, at the first decoder.
- The first decoder then outputs $\hat{W}_0[l] = \hat{G}_{01}[l] \mod 2^{nR_0}$, and $\hat{W}_1[l-1]$ as the index of $U$-bin that contains the sequence $\mathbf{U}_{[(l-1),\hat{G}_3[l-1],\bar{\mathbf{C}}_1[l-1]]}$. The decoder sets

$$\bar{\mathbf{A}}[l] = \mathbf{A}_{[l,\hat{G}_1[l],\bar{\mathbf{U}}[l-1],\bar{\mathbf{C}}_1[l-1],\bar{\mathbf{C}}_1[l]]}.$$

$\bar{\mathbf{A}}[l]$ is the first decoder's estimate of $\mathbf{A}[l]$.
- The second decoder determines the unique index triple $(\hat{G}_{02}[l], \hat{G}_2[l], \hat{G}_4[l-1])$ such that the tuples

$$(\bar{\mathbf{C}}_2[l-1], \bar{\mathbf{B}}[l-1], \bar{\mathbf{V}}[l-1], \mathbf{Z}[l-1]) \text{ and } (\bar{\mathbf{C}}_2[l], \mathbf{B}_{[l,\hat{G}_2[l],\bar{\mathbf{V}}[l-1],\bar{\mathbf{C}}_2[l-1],\bar{\mathbf{C}}_2[l]]}, \mathbf{Z}[l])$$

are jointly $\epsilon[l]$-typical with respect to $P_{\tilde{C}\tilde{B}\tilde{V}\tilde{Z}CBZ}$, where, where

$$\bar{\mathbf{V}}[l-1] \triangleq \mathbf{V}_{[(l-1),\hat{G}_4[l-1],\bar{\mathbf{C}}_2[l-1]]}, \quad \bar{\mathbf{C}}_2[l] \triangleq \mathbf{C}_{[l,\hat{G}_{02}[l],\bar{\mathbf{C}}_2[l-1]]}.$$

If not successful in this operation, the second decoder declares an error and sets $(\hat{G}_{02}[l], \hat{G}_2[l], \hat{G}_4[l-1]) = (1,1,1)$, and

$$\bar{\mathbf{V}}[l-1] \triangleq \mathbf{V}_{[(l-1),1,\bar{\mathbf{C}}_2[l-1]]}, \quad \bar{\mathbf{C}}_2[l] \triangleq \mathbf{C}_{[l,1,\bar{\mathbf{C}}_2[l-1]]};$$

Note that $\bar{\mathbf{V}}[l-1]$ and $\bar{\mathbf{C}}_2[l]$ are the estimates of $\mathbf{V}[l-1]$ and $\mathbf{C}[l]$, respectively, at the second decoder.
- The second decoder then outputs $\bar{W}_0[l] = \hat{G}_{02}[l] \mod 2^{nR_0}$, and $\hat{W}_2[l-1]$ as the index of $V$-bin that contains the sequence $\mathbf{V}_{[(l-1),\hat{G}_4[l-1],\bar{\mathbf{C}}_2[l-1]]}$. The decoder sets

$$\bar{\mathbf{B}}[l] = \mathbf{B}_{[l,\hat{G}_2[l],\bar{\mathbf{V}}[l-1],\bar{\mathbf{C}}_2[l-1],\bar{\mathbf{C}}_2[l]]}.$$

$\bar{\mathbf{B}}[l]$ is the second decoder's estimate of $\mathbf{B}[l]$.

### 7.5 Error Analysis

Let $E[0]$ denote the event that $(\mathbf{U}[0], \mathbf{K}[0], \mathbf{V}[0], \mathbf{C}[0])$ is not $\epsilon[0]$-typical with respect to $P_{UKVC}$. By Property 0, we have $Pr[E[0]] \leq \epsilon$ for all sufficiently large $n$.

**Block 1:** Let $E_1[1]$ denote the event that the encoder declares error in step 1 of encoding (described in Section 7.3). Let $E_2[1]$ denote the event that the encoder declares error in the second step. Let $E_3[1]$ denote the event that the tuples

$$(\mathbf{U}[0], \mathbf{V}[0], \mathbf{K}[0], \mathbf{C}[0]) \quad \text{and} \quad (\mathbf{U}[1], \mathbf{V}[1], \mathbf{K}[1], \mathbf{C}[1])$$

are not jointly $\epsilon[1]$-typical with respect to $P_{\tilde{U}\tilde{V}\tilde{K}\tilde{C}UVKC}$. Let $E_4[1]$ denote the event that $(\hat{G}_{01}[1], \hat{G}_1[1]) \neq (G_0[1], G_1[1])$ and $E_5[1]$ denote the event that $(\hat{G}_{02}[1], \hat{G}_2[1]) \neq (G_0[1], G_2[1])$. The error event in Block 1 is given by $E[1] = E_1[1] \cup E_2[1] \cup E_3[1] \cup E_4[1] \cup E_5[1]$.

**Lemma 7.1.** *$Pr[E_1[1]|E[0]^c] \leq \epsilon$ for all sufficiently large $n$ if $R_0$, $\rho_0, \rho_1$, and $\rho_2$ satisfy*

$$\rho_0 > I(\tilde{U}\tilde{V}\tilde{K}; C|\tilde{C}) + R_0 + 2\delta(\epsilon[1]) \tag{74}$$

$$\rho_0 + \rho_1 > I(\tilde{V}\tilde{K}; A|C\tilde{C}\tilde{U}) + I(\tilde{U}\tilde{V}\tilde{K}; C|\tilde{C}) + R_0 + 2\delta(\epsilon[1]) \tag{75}$$

$$\rho_0 + \rho_2 > I(\tilde{U}\tilde{K}; B|C\tilde{C}\tilde{V}) + I(\tilde{U}\tilde{V}\tilde{K}; C|\tilde{C}) + R_0 + 2\delta(\epsilon[1]) \tag{76}$$

$$\rho_0 + \rho_1 + \rho_2 > I(\tilde{V}\tilde{K}; A|C\tilde{C}\tilde{U}) + I(\tilde{U}\tilde{K}; B|C\tilde{C}\tilde{V}) + I(A; B|\tilde{U}\tilde{V}\tilde{K}C\tilde{C}) + I(\tilde{U}\tilde{V}\tilde{K}; C|\tilde{C}) + R_0 + 2\delta(\epsilon[1]) \tag{77}$$

*Proof.* See Appendix C. □

**Lemma 7.2.** *$Pr[E_2[1]|E[0]^c] \leq \epsilon$ for all sufficiently large $n$ if $R'_1, R'_2$, and $R_1, R_2$ satisfy*

$$R'_1 + R'_2 - R_1 - R_2 > I(U; V|C) + \delta(\epsilon[1]) \tag{78}$$

*Proof.* Follows along the lines of the proof of Lemma 7.1. It is omitted for conciseness. □

From Property 1, it follows that $Pr[E_3[1]|E_1[1]^c, E_2[1]^c, E[0]^c] \leq \epsilon$ for all sufficiently large $n$.

**Lemma 7.3.** *$Pr[E_4[1] \cup E_5[1] \mid E_3[1]^c, E_2[1]^c, E_1[1]^c, E[0]^c] \leq 2\epsilon$, if*

$$\rho_0 + \rho_1 < I(C; Y\tilde{A}\tilde{Y}\tilde{U}|\tilde{C}) + I(A; Y\tilde{A}\tilde{Y}|\tilde{U}\tilde{C}C) - \delta(\epsilon[1]) \tag{79}$$

$$\rho_0 + \rho_2 < I(C; Z\tilde{B}\tilde{Z}\tilde{V}|\tilde{C}) + I(B; Z\tilde{B}\tilde{Z}|\tilde{V}\tilde{C}C) - \delta(\epsilon[1]) \tag{80}$$

$$\rho_1 < I(A; Y\tilde{A}\tilde{Y}|\tilde{U}\tilde{C}C) - \delta(\epsilon[1]) \tag{81}$$

$$\rho_2 < I(B; Z\tilde{B}\tilde{Z}|\tilde{V}\tilde{C}C) - \delta(\epsilon[1]) \tag{82}$$

*Proof.* The proof is very similar to that of Lemma 7.4 given below, and is omitted for conciseness. □

Hence $P[E[1]|E[0]^c] < 5\epsilon$ if the conditions given in Lemmas 7.1, 7.2, and 7.3 are satisfied. This implies that $\bar{\mathbf{A}}[1] = \mathbf{A}[1]$, $\bar{\mathbf{C}}_1[1] = \bar{\mathbf{C}}_2[1] = \mathbf{C}[1]$, and similarly $\bar{\mathbf{B}}[1] = \mathbf{B}[1]$ with high probability.

**Block** 2: Let $E_1[2]$ denote the event that the encoder declares error in step 1 of encoding, and $E_2[2]$ the event that the encoder declares error in step 2 of encoding. Let $E_3[2]$ denote the event that the tuples

$$(\mathbf{U}[1], \mathbf{V}[1], \mathbf{K}[1], \mathbf{C}[1]) \quad \text{and} \quad (\mathbf{U}[2], \mathbf{V}[2], \mathbf{K}[2], \mathbf{C}[2])$$

are not jointly $\epsilon[2]$-typical with respect to $P_{\tilde{U}\tilde{V}\tilde{K}\tilde{C}UVKC}$. Let $E_4[2]$ denote the event that

$$\left\{(\hat{G}_{01}[2], \hat{G}_1[2], \hat{G}_3[1]) \neq (G_0[2], G_1[2], G_3[1])\right\}$$

Similarly, let $E_5[2]$ denote the event that

$$\left\{(\hat{G}_{02}[2], \hat{G}_2[2], \hat{G}_4[1]) \neq (G_0[2], G_2[2], G_4[1])\right\}$$

Hence the error event in this block is given by $E[2] = \cup_{i=1}^{5} E_i[2]$. In the following we show that $P[E[2]|E[1]^c E[0]^c]$ is small under certain conditions.

Using arguments similar to those used in Block 1, one can show that if $\rho_0, \rho_1, \rho_2$, $R_1', R_2'$, $R_1$ and $R_2$ satisfy the conditions given in (78) and (74-77) with $\epsilon[1]$ replaced with $\epsilon[2]$, then for all sufficiently large $n$,

$$Pr[E_1[2] \cup E_2[2] \cup E_3[2]|E[1]^c, E[0]^c] \leq 3\epsilon.$$

**Lemma 7.4.** *$Pr[E_4[2] \cup E_5[2] \mid E_3[2]^c, E_2[2]^c, E_1[2]^c, E[1]^c, E[0]^c] \leq 2\epsilon$, if*

$$R_1' + \rho_0 + \rho_1 < I(\tilde{U}; Y\tilde{Y}\tilde{A}|\tilde{C}) + I(C; Y\tilde{A}\tilde{Y}\tilde{U}|\tilde{C}) + I(A; Y\tilde{A}\tilde{Y}|\tilde{U}\tilde{C}C) - \delta(\epsilon[2]) \quad (83)$$

$$R_1' + \rho_1 < I(\tilde{U}; Y\tilde{A}\tilde{Y}C|\tilde{C}) + I(A; Y\tilde{A}\tilde{Y}|\tilde{U}\tilde{C}C) - \delta(\epsilon[2]) \quad (84)$$

$$R_2' + \rho_0 + \rho_2 < I(\tilde{V}; Z\tilde{Z}\tilde{B}|\tilde{C}) + I(C; Z\tilde{B}\tilde{Z}\tilde{V}|\tilde{C}) + I(B; Z\tilde{B}\tilde{Z}|\tilde{V}\tilde{C}C) - \delta(\epsilon[2]) \quad (85)$$

$$R_2' + \rho_2 < I(\tilde{V}; Z\tilde{B}\tilde{Z}C|\tilde{C}) + I(B; Z\tilde{B}\tilde{Z}|\tilde{V}\tilde{C}C) - \delta(\epsilon[2]) \quad (86)$$

$$\rho_0 + \rho_1 < I(C; Y\tilde{A}\tilde{Y}\tilde{U}|\tilde{C}) + I(A; Y\tilde{A}\tilde{Y}|\tilde{U}\tilde{C}C) - \delta(\epsilon[2]) \quad (87)$$

$$\rho_0 + \rho_2 < I(C; Z\tilde{B}\tilde{Z}\tilde{V}|\tilde{C}) + I(B; Z\tilde{B}\tilde{Z}|\tilde{V}\tilde{C}C) - \delta(\epsilon[2]) \quad (88)$$

$$\rho_1 < I(A; Y\tilde{A}\tilde{Y}|\tilde{U}\tilde{C}C) - \delta(\epsilon[2]) \quad (89)$$

$$\rho_2 < I(B; Z\tilde{B}\tilde{Z}|\tilde{V}\tilde{C}C) - \delta(\epsilon[2]) \quad (90)$$

*Proof.* See Appendix D. □

Hence $Pr[E[2]|E[1]^c E[0]^c] < 5\epsilon$. Under the event $(E[2]^c \cap E[1]^c \cap E[0]^c)$, we have $\bar{\mathbf{A}}[2] = \mathbf{A}[2]$, $\bar{\mathbf{C}}_1[2] = \bar{\mathbf{C}}_2[2] = \mathbf{C}[2]$, and $\bar{\mathbf{B}}[2] = \mathbf{B}[2]$.

**Block** $l$, $l = 3, 4, \ldots, L$:

Let $E_1[l]$ denote the event that the encoder declares error in the step 1 of encoding, and $E_2[l]$ the event that it declares error in step 2 of decoding. Let $E_3[l]$ denote the event that the tuples

$$(\mathbf{U}[l-1], \mathbf{V}[l-1], \mathbf{K}[l-1], \mathbf{C}[l-1]) \quad \text{and} \quad (\mathbf{U}[l], \mathbf{V}[l], \mathbf{K}[l], \mathbf{C}[l])$$

are not jointly $\epsilon[l]$-typical with respect to $P_{\tilde{U}\tilde{V}\tilde{K}\tilde{C}UVKC}$. Let $E_4[l]$ denote the event that

$$\left\{(\hat{G}_{01}[l], \hat{G}_1[l], \hat{G}_3[l-1]) \neq (G_0[l], G_1[l], G_3[l-1])\right\}$$

Similarly, let $E_5[l]$ denote the event that

$$\left\{(\hat{G}_{02}[l], \hat{G}_2[l], \hat{G}_4[l-1]) \neq (G_0[l], G_2[l], G_4[l-1])\right\}$$

Hence the error event in this block is given by $E[l] = \cup_{i=1}^{5} E_i[l]$. Using the arguments similar to those used in block 2, it can be shown that $Pr[E[l]| \cap_{i=0}^{l-1} E[i]^c] \leq 5\epsilon$ for all sufficiently large $n$ if the conditions given in (78), (74)-(77) and (83)-(90) are satisfied with $\epsilon[1]$ and $\epsilon[2]$ are replaced with $\epsilon[l]$.

**Overall Probability of Decoding Error** : Hence the probability of decoding error over $L$ blocks satisfies

$$Pr[E] = Pr\left[\cup_{l=0}^{L} E[l]\right] \leq 5\epsilon L$$

if the conditions given in (78), (74)-(77) and (83)-(90) are satisfied with $\delta(\epsilon[1])$ and $\delta(\epsilon[2])$ are replaced with $\theta$, where $\theta = \sum_{i=1}^{L} \delta(\epsilon[l])$. This implies that the rate region given by (14), (15)-(18), (19)-(26) is achievable. By applying Fourier-Motzkin elimination to these equations, we obtain that the rate region given in the statement of the theorem is achievable. The details of this elimination are omitted since they are elementary, but somewhat tedious.

# 8 Conclusion

We have derived a single-letter rate region for the two-user broadcast channel with feedback. Using the Marton coding scheme as the starting point, our scheme uses three additional random variables $(A, B, C)$ to cover the correlated information generated at the end of each block. The key to obtaining a single-letter characterization was to impose a constraint on the distribution used to generate these covering variables. A similar idea was used in [16] for multiple-access channels with feedback. This approach to harnessing correlated information is quite general, and it is likely that it can be used to obtain improved rate regions for other multi-user channels with feedback such as interference and relay channels.

## APPENDIX

# A Mutual Information terms for the AWGN example

With the joint distribution described in Section 5.2, we first compute the following quantities.

$$M_u \triangleq E[(\tilde{U} - E[\tilde{U}|\tilde{Y}\tilde{C}_0])^2] = \alpha P_1 \left(\frac{\sigma^2 + \bar{\beta}\bar{\alpha}P_1}{P_1 + \sigma^2}\right)^2 + \bar{\alpha}P_1 \left(\frac{\beta\sigma^2 - \alpha P_1 \bar{\beta}}{P_1 + \sigma^2}\right)^2 + \sigma^2 \left(\frac{\alpha P_1 + \beta\bar{\alpha}P_1}{P_1 + \sigma^2}\right)^2 \tag{91}$$

$$\mathrm{var}(\tilde{S}_1|\tilde{C}_0\tilde{Z}) \triangleq E[(\tilde{S}_1 - E[\tilde{S}_1|\tilde{C}_0\tilde{Z}])^2] = 1 - \frac{P_1^2\sigma^4(\alpha + \bar{\alpha}\beta)^2}{M_u(P_1 + \sigma^2)^3} \tag{92}$$

$$\mathrm{var}(\tilde{S}_1|\tilde{C}_0\tilde{V}\tilde{Z}) \triangleq E[(\tilde{S}_1 - E[\tilde{S}_1|\tilde{C}_0\tilde{V}\tilde{Z}])^2] = 1 - \frac{(\alpha P_1\sigma^2 + \alpha\bar{\alpha}\bar{\beta}P_1^2)^2}{M_u(P_1 + \sigma^2)^2(\alpha P_1 + \sigma^2)} - \frac{\bar{\alpha}P_1(\beta\sigma^2 - \alpha P_1\bar{\beta})^2}{M_u(P_1 + \sigma^2)^2} \tag{93}$$

In terms of the above quantities, the mutual information terms are

$$\begin{aligned}
I(V;Z|C_0) &= \frac{1}{2}\log_2\left(\frac{P_1 + \sigma^2}{\alpha P_1 + \sigma^2}\right) \\
I(U;Y|C_0) &= \frac{1}{2}\log_2\left(\frac{(P_1 + \sigma^2)(\alpha + \beta^2\bar{\alpha})}{(P_1 + \sigma^2)(\alpha + \beta^2\bar{\alpha}) - P_1(\alpha + \beta\bar{\alpha})^2}\right) \\
I(U;V) &= \frac{1}{2}\log_2\left(1 + \frac{\beta^2\bar{\alpha}}{\alpha}\right) \\
I(C_0;\tilde{U}|\tilde{C}_0\tilde{Y}) &= \frac{1}{2}\log_2\left(\frac{1}{D}\right) \\
I(C_0;\tilde{V}|\tilde{C}_0\tilde{Z}) &= \frac{1}{2}\log_2\left(\frac{(1-D)\mathrm{var}(\tilde{S}_1|\tilde{C}_0\tilde{Z}) + D}{(1-D)\mathrm{var}(\tilde{S}_1|\tilde{C}_0\tilde{V}\tilde{Z}) + D}\right) \\
I(C_0;Z|\tilde{C}_0\tilde{V}\tilde{Z}) &= \frac{1}{2}\log_2\left(1 + \frac{(P - P_1)((1-D)\mathrm{var}(\tilde{S}_1|\tilde{C}_0\tilde{V}\tilde{Z}) + D)}{P_1 + \sigma^2}\right) \\
I(C_0;Z|\tilde{C}_0\tilde{Z}) &= \frac{1}{2}\log_2\left(1 + \frac{(P - P_1)((1-D)\mathrm{var}(\tilde{S}_1|\tilde{C}_0\tilde{Z}) + D)}{P_1 + \sigma^2}\right) \\
I(C_0;Y|\tilde{Y}\tilde{C}_0) &= \frac{1}{2}\log_2\left(\frac{P + \sigma^2}{P_1 + \sigma^2}\right) \\
I(\tilde{U}\tilde{Y};C_0|\tilde{C}_0\tilde{V}\tilde{Z}) &= \frac{1}{2}\log_2\left(1 + \frac{1-D}{D}\mathrm{var}(\tilde{S}_1|\tilde{C}_0\tilde{V}\tilde{Z})\right) \\
I(\tilde{V}\tilde{Z};C_0|\tilde{C}_0\tilde{U}\tilde{Y}) &= 0
\end{aligned}$$

# B Proof of Lemma 6.1

We start with the first claim of the lemma. (59) implies the following constraint for the S-W region.

$$\begin{aligned}
R_0 + R_1 &< I(U_0U_1;YV_1) - I(U_0U_1U_2S;V_0V_1|Y) \\
&= I(U_0U_1;Y) + I(U_0U_1;V_1|Y) - I(U_0U_1U_2S;V_0V_1|Y) \\
&= \mathbf{I(U_0U_1;Y)} + \mathbf{I(U_0U_1;V_0V_1|Y)} - \mathbf{I(U_0U_1U_2S;V_0V_1|Y)} - I(U_0U_1;V_0|V_1Y)
\end{aligned} \tag{94}$$

where we observe that the part in boldface is the right hand side of (69). Therefore, (69) is implied by (59). In a similar manner, one can show that (70) is implied by (60).

Adding (58) and (61), we get the following S-W constraint.

$$\begin{aligned} 2R_0 + R_1 + R_2 < I(U_1; YV_1|U_0) + I(U_2; ZV_2|U_0) - I(U_1; U_2|U_0) + 2\Theta_2 - 2\Theta_1 \\ - I(U_0U_1U_2S; V_1|V_0Y) - I(U_0U_1U_2S; V_2|V_0Z) \end{aligned} \tag{95}$$

(95) implies the following constraint for the S-W region.

$$\begin{aligned} 2R_0 + R_1 + R_2 &< I(U_1; YV_1|U_0) + I(U_2; ZV_2|U_0) - I(U_1; U_2|U_0) + I(U_0; ZV_2) + I(U_0; YV_1) \\ &\quad - I(U_0U_1U_2S; V_1V_0|Y) - I(U_0U_1U_2S; V_2V_0|Z) \\ &= I(U_0U_1; YV_1) + I(U_0U_2; ZV_2) - I(U_1; U_2|U_0) - I(U_0U_1U_2S; V_1V_0|Y) - I(U_0U_1U_2S; V_2V_0|Z) \\ &= I(U_0U_1; Y) + I(U_0U_2; Z) + I(U_0U_1; V_1|Y) + I(U_0U_2; V_2|Z) - I(U_1; U_2|U_0) \\ &\quad - I(U_0U_1U_2S; V_1V_0|Y) - I(U_0U_1U_2S; V_2V_0|Z) \\ &= \mathbf{I(U_0U_1; Y) + I(U_0U_2; Z) + I(U_0U_1; V_0V_1|Y) + I(U_0U_2; V_0V_2|Z) - I(U_1; U_2|U_0)} \\ &\quad \mathbf{- I(U_0U_1U_2S; V_1V_0|Y) - I(U_0U_1U_2S; V_2V_0|Z)} - I(U_0U_1; V_0|V_1Y) - I(U_0U_2; V_0|V_2Z) \end{aligned} \tag{96}$$

where the part in boldface is the right hand side of (73). Hence (73) is implied by (58) and (61)

Finally, (61) implies the following constraint for the S-W region.

$$\begin{aligned} R_0 + R_1 + R_2 &< I(U_0U_1; YV_1) - I(U_0U_1U_2S; V_1V_0|Y) - I(U_1; U_2|U_0) + I(U_2; ZV_2|U_0) - I(U_0U_1U_2S; V_2|V_0Z) \\ &= I(U_0U_1; Y) + I(U_0U_1; V_1|Y) - I(U_0U_1U_2S; V_1V_0|Y) - I(U_1; U_2|U_0) \\ &\quad + I(U_2; Z|U_0) + I(U_2; V_2|U_0Z) - I(U_0U_1U_2S; V_2|V_0Z) \\ &= I(U_0U_1; Y) + I(U_0U_1; V_0V_1|Y) - I(U_0U_1; V_0|V_1Y) - I(U_0U_1U_2S; V_1V_0|Y) - I(U_1; U_2|U_0) \\ &\quad + I(U_2; Z|U_0) + I(U_2; V_2|U_0Z) - I(U_0U_1U_2S; V_2|V_0Z) + I(U_0U_2; V_2|V_0Z) - I(U_0U_2; V_2|V_0Z) \\ &= \mathbf{I(U_0U_1; Y) + I(U_2; Z|U_0) - I(U_0U_1U_2S; V_1V_0|Y) - I(U_0U_1U_2S; V_2|V_0Z) - I(U_1; U_2|U_0)} \\ &\quad \mathbf{+ I(U_0U_1; V_0V_1|Y) + I(U_0U_2; V_2|V_0Z)} + I(U_2; V_2|U_0Z) - I(U_0U_2; V_2|V_0Z) - I(U_0U_1; V_0|V_1Y) \end{aligned} \tag{97}$$

where the part in bold-face is the right hand side of (71). Thus, (71) is implied by (61) if

$$I(U_2; V_2|U_0Z) - I(U_0U_2; V_2|V_0Z) - I(U_0U_1; V_0|V_1Y) < 0,$$

which is the first condition in (66). Similarly, one can show that (72) is implied by (61) if the second condition in (66) holds.

## C Proof of Lemma 7.1

For a given sequence quadruple $(\tilde{u}^n, \tilde{v}^n, \tilde{k}^n, \tilde{c}^n) \in A^{(n)}_{\epsilon[0]}(\tilde{U}, \tilde{V}, \tilde{K}, \tilde{C})$, let $G(\tilde{u}^n, \tilde{v}^n, \tilde{k}^n, \tilde{c}^n)$ denote the function given by

$$G = \sum_{i,j,k} \phi(i,j,k)$$

where $\phi(i,j,k) = 1$ if

$$i \bmod 2^{nR_0} = W_0[1] \text{ AND } (\tilde{u}^n, \tilde{v}^n, \tilde{k}^n, \tilde{c}^n, \mathbf{C}_{[1,i,\tilde{c}^n]}, \mathbf{A}_{[1,j,\tilde{u}^n,\tilde{c}^n,\mathbf{C}_{[1,i,\tilde{c}^n]}]}\mathbf{B}_{[1,k,\tilde{v}^n,\tilde{c}^n,\mathbf{C}_{[1,i,\tilde{c}^n]}]}) \in A^{(n)}_{\epsilon[1]}(\tilde{U}, \tilde{V}, \tilde{K}, \tilde{C}, A, B, C)$$

and equals 0 otherwise. We will show that $\Pr[G = 0] \leq \epsilon$ for sufficiently large $n$ if $\rho_0, \rho_1$, and $\rho_2$ satisfy the conditions given in the statement of the lemma. Using Chebyshev's inequality, we have

$$\Pr[G = 0] \leq \frac{\text{var}(G)}{(E[G])^2} = \frac{E[G^2] - (E[G])^2}{(E[G])^2}. \tag{98}$$

We will now compute $E[G]$ and $E[G^2]$.

Using Property 2 of typical sequences in Section 7.1, we have

$$\begin{aligned} E[G] &= \sum_{i,j,k} Pr[\phi(i,j,k) = 1] \\ &\geq \frac{2^{n(\rho_0+\rho_1+\rho_2-R_0-\delta(\epsilon[1]))} 2^{nH(ABC|\tilde{U}\tilde{V}\tilde{K}\tilde{C})}}{2^{nH(C|\tilde{C})} 2^{nH(A|\tilde{U}\tilde{C}C)} 2^{nH(B|\tilde{V}\tilde{C}C)}} \\ &= \frac{2^{n(\rho_0+\rho_1+\rho_2-R_0-\delta(\epsilon[1]))}}{2^{nI(\tilde{U}\tilde{V}\tilde{K};C|\tilde{C})} 2^{nI(\tilde{U}K;B|C\tilde{C}\tilde{V})} 2^{nI(\tilde{V}K;A|C\tilde{C}\tilde{U})} 2^{nI(A;B|C\tilde{C}\tilde{U}\tilde{V}\tilde{K})}} \\ &= 2^{n(\kappa-\delta(\epsilon[1]))} \end{aligned} \tag{99}$$

where

$$\kappa \triangleq \rho_0 + \rho_1 + \rho_2 - R_0 - I(\tilde{U}\tilde{V}\tilde{K}; C|\tilde{C}) - I(\tilde{U}K; B|C\tilde{C}\tilde{V}) - I(\tilde{V}K; A|C\tilde{C}\tilde{U}) - I(A; B|C\tilde{C}\tilde{U}\tilde{V}\tilde{K}). \tag{100}$$

$E[G^2]$ can be written as

$$E[G^2] = \Phi_1 + \Phi_2 + \Phi_3 + \Phi_4 + \Phi_5 \tag{101}$$

where

$$\begin{aligned} \Phi_1 &\triangleq \sum_{i,j,k} Pr[\phi(i,j,k) = 1] = E[G], \\ \Phi_2 &\triangleq \sum_{i,j,k} \sum_{k' \neq k} P[\phi(i,j,k)\phi(i,j,k') = 1], \\ \Phi_3 &\triangleq \sum_{i,j,k} \sum_{j' \neq j} P[\phi(i,j,k)\phi(i,j',k) = 1], \\ \Phi_4 &\triangleq \sum_{i,j,k} \sum_{k' \neq k} \sum_{j \neq j'} P[\phi(i,j,k)\phi(i,j',k') = 1], \\ \Phi_5 &\triangleq \sum_{i,j,k} \sum_{\substack{i',j',k' \\ i \neq i'}} P[\phi(i,j,k)\phi(i',j',k') = 1]. \end{aligned} \tag{102}$$

The second term $\Phi_2$ can be bounded as

$$\Phi_2 \leq \frac{2^{n\kappa}2^{n\rho_2}2^{2n\delta(\epsilon[1])}2^{nH(B|\tilde{U}\tilde{K}\tilde{V}\tilde{C}AC)}}{2^{nH(B|\tilde{V}\tilde{C}C)}} = \frac{2^{n\kappa}2^{n\rho_2}2^{2n\delta(\epsilon[1])}}{2^{nI(B;\tilde{U}\tilde{K}A|\tilde{C}\tilde{V}C)}}. \tag{103}$$

Similarly, the third term $\Phi_3$ is bounded as

$$\Phi_3 \leq \frac{2^{n\kappa}2^{n\rho_1}2^{2n\delta(\epsilon[1])}}{2^{nI(A;\tilde{U}\tilde{K}B|\tilde{C}\tilde{U}C)}}. \tag{104}$$

The fourth term $\Phi_4$ can be bounded as

$$\Phi_4 \leq \frac{2^{n\kappa}2^{n(\rho_1+\rho_2)}2^{2n\delta(\epsilon[1])}2^{nH(AB|\tilde{U}\tilde{K}\tilde{V}\tilde{C}C)}}{2^{nH(A|\tilde{U}\tilde{C}C)}2^{nH(B|\tilde{V}\tilde{C}C)}} = \frac{2^{n\kappa}2^{n(\rho_1+\rho_2)}2^{2n\delta(\epsilon[1])}}{2^{n(I(\tilde{V}\tilde{K};A|\tilde{U}\tilde{C}C)+I(\tilde{U}\tilde{K};B|\tilde{V}\tilde{C}C)+I(A;B|\tilde{U}\tilde{K}\tilde{V}\tilde{C}C))}}. \tag{105}$$

Finally, the fifth term satisfies

$$\Phi_5 \leq \left(\sum_{i,j,k}\Pr[\phi(i,j,k)=1]\right)^2 = (E[G])^2. \tag{106}$$

Using the above in (101), we can get an upper bound on $E[G^2]$. Substituting in (98), we obtain

$$\begin{aligned}\Pr[G=0] &\leq 2^{2n\delta(\epsilon[1])}\Big[2^{-n\kappa} + 2^{-n(\kappa-\rho_1+I(A;\tilde{V}\tilde{K}B|\tilde{U}\tilde{C}C))} + 2^{-n(\kappa-\rho_2+I(B;\tilde{U}\tilde{K}A|\tilde{C}\tilde{V}C))} \\ &\qquad + 2^{-n(\kappa-\rho_1-\rho_2+I(A;\tilde{K}\tilde{V}|C\tilde{C}\tilde{U})+I(B;\tilde{K}\tilde{U}|C\tilde{C}\tilde{V})+I(A;B|\tilde{U}\tilde{V}\tilde{K}C\tilde{C}))}\Big] \\ &\leq \epsilon\end{aligned} \tag{107}$$

if the conditions given in the statement of the lemma are satisfied. Hence

$$\begin{aligned}\Pr[E_1[1]|E[0]^c] &= \sum_{\tilde{u}^n,\tilde{v}^n,\tilde{k}^n,\tilde{c}^n \in A^{(n)}_{\epsilon[1]}(\tilde{U},\tilde{V},\tilde{K},\tilde{C})} \Pr[(\mathbf{U}[0],\mathbf{V}[0],\mathbf{K}[0],\mathbf{C}[0]) = (\tilde{u}^n,\tilde{v}^n,\tilde{k}^n,\tilde{c}^n)]\ \Pr[G(\tilde{u}^n,\tilde{v}^n,\tilde{k}^n,\tilde{c}^n)=0]\frac{1}{\Pr[E[0]^c]} \\ &\leq \epsilon.\end{aligned} \tag{108}$$

# D Proof of Lemma 7.4

For this section alone, let $\hat{E}$ denote the event $(E_3[2]^c \cap E_2[2]^c \cap E_1[2]^c \cap E[1]^c \cap E[0]^c)$. Note that $\hat{E}$ is the conditioning event in the statement of Lemma 7.4, and hence $\bar{\mathbf{A}}[1] = \mathbf{A}[1]$, $\bar{\mathbf{C}}_1[1] = \bar{\mathbf{C}}_2[1] = \mathbf{C}[1]$, and similarly $\bar{\mathbf{B}}[1] = \mathbf{B}[1]$. For conciseness, let $\tilde{E}(\tilde{y}^n, y^n, \tilde{a}^n, \tilde{c}^n, \alpha, \beta, \gamma)$ denote the event

$$(\mathbf{Y}[1], \mathbf{Y}[2], \bar{\mathbf{A}}[1], \bar{\mathbf{C}}_1[1], G_0[2], G_1[2], G_3[1]) = (\tilde{y}^n, y^n, \tilde{a}^n, \tilde{c}^n, \alpha, \beta, \gamma)$$

for any $(\tilde{y}^n, y^n, \tilde{a}^n, \tilde{c}^n) \in A^{(n)}_{\epsilon[2]}(\tilde{Y}, Y, \tilde{A}, \tilde{C})$ and any index triple $(\alpha, \beta, \gamma)$.

Now consider the following argument.

$$P[E_4|\hat{E}] = \frac{P[E_4 \cap \hat{E}]}{P[\hat{E}]} \tag{109}$$

$$= \frac{1}{P[\hat{E}]} \sum_{(\tilde{y}^n, y^n, \tilde{a}^n, \tilde{c}^n) \in A^{(n)}_{\epsilon[2]}} \sum_{\alpha,\beta,\gamma} P[E_4 \cap \hat{E} \cap \tilde{E}(\tilde{y}^n, y^n, \tilde{a}^n, \tilde{c}^n, \alpha, \beta, \gamma)]$$

$$= \frac{1}{P[\hat{E}]} \sum_{(\tilde{y}^n, y^n, \tilde{a}^n, \tilde{c}^n) \in A^{(n)}_{\epsilon[2]}} \sum_{\alpha,\beta,\gamma} P[\hat{E} \cap \tilde{E}(\tilde{y}^n, y^n, \tilde{a}^n, \tilde{c}^n, \alpha, \beta, \gamma] P[E_4|\tilde{E}\hat{E}] \tag{110}$$

Let us now define the following indicator random variable: $\psi(i,j,k) = 1$ if the tuples

$$(\mathbf{U}_{[1,k,\mathbf{C}_1[1]]}, \mathbf{A}[1], \mathbf{Y}[1], \mathbf{C}_1[1]) \quad \text{and} \quad (\mathbf{C}_{[2,i,\mathbf{C}_1[1]]}, \mathbf{Y}[2], \mathbf{A}_{[2,j,\mathbf{U}_{[1,k,\mathbf{C}_1[1]]}, \mathbf{C}_1[1], \mathbf{C}_{[2,i,\mathbf{C}_1[1]]}]})$$

are jointly $\epsilon[2]$-typical with respect to $P_{\tilde{U}\tilde{A}\tilde{Y}\tilde{C}CAY}$ and 0 otherwise.

Now we have

$$P[E_4|\tilde{E}\hat{E}] \leq \sum_{(i,j,k) \neq (\alpha,\beta,\gamma)} P[\psi(i,j,k) = 1|\tilde{E}\hat{E}] \tag{111}$$

$$= \Phi_1 + \Phi_2 + \Phi_3 + \Phi_4, \tag{112}$$

where

$$\Phi_1 = \sum_{j \neq \beta} P[\psi(\alpha, j, \gamma) = 1|\tilde{E}\hat{E}], \tag{113}$$

$$\Phi_2 = \sum_{i \neq \alpha} \sum_{j} P[\psi(i, j, \gamma) = 1|\tilde{E}\hat{E}], \tag{114}$$

$$\Phi_3 = \sum_{k \neq \gamma} \sum_{j} P[\psi(\alpha, j, k) = 1|\tilde{E}\hat{E}], \tag{115}$$

$$\Phi_4 = \sum_{i \neq \alpha} \sum_{k \neq \gamma} \sum_{j} P[\psi(i, j, k) = 1|\tilde{E}\hat{E}]. \tag{116}$$

Now, using Property 2 of typical sequences in Section 7.1, we have the following bounds.

$$\Phi_1 \leq \frac{2^{n\rho_1} 2^{n\delta_1(\epsilon[2])} 2^{nH(A|\tilde{U}\tilde{C}\tilde{A}\tilde{Y}CY)}}{2^{nH(A|\tilde{U}\tilde{C}C)}}$$
$$= \exp_2\left[n\left(\rho_1 + \delta_1(\epsilon[2]) - I(A; Y\tilde{A}\tilde{Y}|\tilde{U}C\tilde{C})\right)\right], \tag{117}$$
$$\Phi_2 \leq \frac{2^{n(\rho_0+\rho_1)} 2^{n\delta_1(\epsilon[2])} 2^{nH(CA|\tilde{U}\tilde{C}\tilde{A}\tilde{Y}Y)}}{2^{n(H(C|\tilde{C})+H(A|\tilde{U}\tilde{C}C))}}$$
$$= \exp_2\left[n\left(\rho_0 + \rho_1 + \delta_1(\epsilon[2]) - I(C; \tilde{U}\tilde{A}\tilde{Y}Y|\tilde{C}) - I(A; Y\tilde{A}\tilde{Y}|\tilde{U}C\tilde{C})\right)\right], \tag{118}$$
$$\Phi_3 \leq \frac{2^{n(R_1'+\rho_1)} 2^{n\delta_1(\epsilon[2])} 2^{nH(\tilde{U}A|\tilde{C}\tilde{A}\tilde{Y}CY)}}{2^{n(H(\tilde{U}|\tilde{C})+H(A|\tilde{U}\tilde{C}C))}}$$
$$= \exp_2\left[n\left(\rho_1 + R_1' + \delta_1(\epsilon[2]) - I(\tilde{U}; \tilde{A}\tilde{Y}CY|\tilde{C}) - I(A; Y\tilde{A}\tilde{Y}|\tilde{U}C\tilde{C})\right)\right] \tag{119}$$
$$\Phi_4 \leq \frac{2^{n(\rho_0+R_1'+\rho_1)} 2^{n\delta_1(\epsilon[2])} 2^{nH(\tilde{U}CA|\tilde{C}\tilde{A}\tilde{Y}Y)}}{2^{n(H(\tilde{U}|\tilde{C})+H(C|\tilde{C})+H(A|\tilde{U}\tilde{C}C))}}$$
$$= \exp_2\left[n\left(\rho_0 + \rho_1 + R_1' + \delta_1(\epsilon[2]) - I(\tilde{U}; Y\tilde{A}\tilde{Y}|\tilde{C}) + I(C; Y\tilde{Y}\tilde{U}\tilde{A}|\tilde{C}) + I(A; Y\tilde{A}\tilde{Y}|\tilde{U}C\tilde{C})\right)\right]. \tag{120}$$

Therefore we have

$$P[E_4[2]|\hat{E}\tilde{E}] \leq \frac{\epsilon}{2}$$

for all sufficiently large $n$ if the conditions in the statement of the lemma are satisfied. Substituting back in (110), we obtain $P[E_4[2]|\hat{E}] \leq \epsilon$ for all sufficiently large $n$. Similarly, one can show that $P[E_5|\hat{E}] \leq \epsilon$ if the conditions in the statement of lemma are satisfied.

# References


[1] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform Theory*, vol. 25, no. 3, pp. 306–311, 1979.

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

[3] A. El Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. 24, pp. 379–381, May 1980.

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

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

[6] S. R. Bhaskaran, "Gaussian broadcast channel with feedback," *IEEE Trans. on Inform. Theory*, vol. 54, no. 11, pp. 5252–5257, 2008.

[7] T. S. Han and M. H. M. Costa, "Broadcast channels with arbitrarily correlated sources," *IEEE Trans. Inform, Theory*, vol. IT-33, no. 5, pp. 641–650, 1987.

[8] G. Kramer and C. Nair, "Comments on 'Broadcast channels with arbitrarily correlated sources'," in *Proc. IEEE Int. Symp. on Inf. Theory*, June 2009.

[9] P. Minero and Y.-H. Kim, "Correlated sources over broadcast channels," in *Proc. IEEE Int. Symp. Inf. Theory*, 2009.

[10] S. Choi and S. S. Pradhan, "A graph-based framework for transmission of correlated sources over broadcast channels," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 2841–2856, 2008.

[11] W. Kang and G. Kramer, "Broadcast channel with degraded source random variables and receiver side information," in *Proc. IEEE Int. Symp. on Inf. Theory*, June 2008.

[12] R. Venkataramanan and S. S. Pradhan, "Achievable rates for the broadcast channel with feedback," *Proc. IEEE Int. Symp. on Inf. Theory*, 2010.

[13] O. Shayevitz and M. A. Wigger, "On the capacity of the discrete memoryless broadcast channel with feedback," *Proc. IEEE Int. Symp. on Inf. Theory*, 2010. http://arxiv.org/abs/1012.6012.

[14] A. El Gamal and E. C. van der Muelen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 120–122, 1981.

[15] T. S. Han, "A general coding scheme for the two-way channel," *IEEE Trans. Inform. Theory*, vol. 30, no. 1, pp. 35–43, 1984.

[16] R. Venkataramanan and S. S. Pradhan, "A new achievable rate region for the discrete memoryless multiple-access channel with feedback," in *Proc. IEEE Int. Symp. on Inf. Theory*, June 2009. http://arxiv.org/abs/0901.1892.

[17] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback- part II: Band-limited signals," *IEEE Trans. on Information Theory*, vol. IT-12, pp. 183–189, April 1966.

[18] T. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. 18, pp. 2 – 14, Jan 1972.

[19] P. Bergmans, "A simple converse for broadcast channels with additive white gaussian noise (corresp.)," *IEEE Trans. Inform. Theory*, vol. 20, pp. 279 – 280, Mar 1974.

[20] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, pp. 439 – 441, May 1983.

[21] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19 – 31, 1980.

[22] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic Press,, 1981.