CCIT Report #504 September 2004

Coding for the Degraded Broadcast Channel with Random Parameters, with Causal and Non-Causal Side Information

Yossef Steinberg

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel
ysteinbe@ee.technion.ac.il

September 12, 2004

Abstract

In this work coding for the degrarded broadcast channel controlled by random parameters is studied. Two main paradigms are considered: where side information on the random parameters is provided to the transmitter in a noncausal manner (termed here *non-causal coding*), and where side information is provided in a causal manner (termed *causal coding*). Inner and outer bounds are derived on the capacity region with non-causal coding. For the special case where the non-degrarded user is informed about the channel parameters, we show that the inner bound is tight, thus deriving the capacity region for that case. For causal coding, a single-letter characterization of the capacity region is derived. This characterization is expressed via auxiliary random variables, and can also be interpreted by means of *Shannon strategies*, as the formula for the capacity of the single-user channel with causal coding derived by Shannon. The capacity region of a class of binary broadcast channels with causal coding is computed, as an example. Applications to watermarking are suggested. In particular, our results on non-causal coding can be used to derive the capacity region of a watermarking system where the channel (attacker) is fixed, and the encoder is required to encode watermarks for both, private and public users.

Index terms — Broadcast channel, causal coding, degraded broadcast channel, information hiding, non-causal coding, side information, Shannon strategies, watermarking.

1 Introduction

Channels that depend on random parameters have been extensively studied, due to the wide range of applications in which such models appear. In many cases, it is reasonable to assume that the random parameters controlling the channel are known either to the receiver or to the transmitter. When such knowledge is present at the receiver, the known parameters can be regarded as part of the channel output, and hence, at least from theoretical point of view, this model does not differ from the classical (state-independent) model. Similarly, when the parameters are known at both sides, the problem can be handled with classical (state-independent) methods, by splitting the channel into many parallel channels, each with one (deterministic) state, constructing an optimal codebook for that state, and combining the codebooks again according to the relative frequencies of the corresponding states. A digress from state-independent methods appears when the parameters are known only at the transmitter. A key factor in the study of coding for, and the capacity of, such channels, is whether the parameters controlling the channel are known causally or non-causally at the encoder. The capacity of the single-user channel with random parameters, known causally at the encoder, was derived by Shannon [18]. Let $P_{Y|X,S}$ stand for the channel transition probability, where Y is the channel output, X the channel input, and S the random parameters, distributed according to P_S. Shannon showed that the capacity of this channel can be expressed as

$$C = \max_{P_T(t)} I(T;Y) \tag{1}$$

where $t(\cdot)$ stands for *strategy*, i.e., a mapping from the state alphabet $\mathcal{S}$ to the channel input alphabet $\mathcal{X}$, and the maximization is over all distributions $P_T(t)$ on the space of strategies. It should be noted that evaluation of (1) for specific models is usually hard, as it involves maximization with respect to distributions over the space of strategies. There are few cases, however, where (1) admits a simpler form. These include cases where the state sequence known at the encoder is a subset of the channel output [1], and the discrete memoryless modulo-additive channels, treated by Erez and Zamir in [10]. We will refer to some of the derivations in [10] in our examples section.

Coding for state-dependent channels, where the state is known non-causally at the encoder, was initiated by the work of Kusnetsov and Tsybakov [14], who considered this problem in the context of coding for a storage unit, say a computer memory, with defective cells, where the location of the defective cells is known a priori to the encoder. This application initiated a series of works that dealt with specific coding schemes for computer memories – see, e.g., [22] and [15], and references therein. A complete treatment, with a capacity formula for the general case (i.e., not restricted to a computer memory model), was given by Gal'fand and Pinsker [11]. In their setting, a single user channel $P_{Y|X,S}$ is controlled by random state S, whose realization s^n is known non-causally to the encoder, whereas the decoder is kept ignorant of it. Gel'fand and Pinsker showed that the capacity of this channel is given by

$$C = \max \left[I(U;Y) - I(U;S) \right] \tag{2}$$

where U is an external random variable, and the maximum is over all random variables (U, S, X, Y), where Y is connected to (U, S, X) via the channel $P_{Y|X,S}$. In general, the evaluation of (2) for specific models is a difficult task. For the special case of Gaussian channel with additive Gaussian interference S (that is, the state plays the role of an interferer), Costa [5] have shown that the capacity equals that of the same Gaussian channel, without the additive interference S. Costa's result was extended to vector Gaussian channels with additive interference in [27]. Although non-causal coding might seem to be of limited applicability, it turned out that the Gel'fand and Pinsker result, and the interesting observations made by Costa, has found numerous applications in various fields. The non-causal coding model was recently extensively applied to the problem of watermarking and information hiding ([4], [16], [17], [21] is only a partial list). In these models, the channel does not explicitely depend on the state S. Instead, a host data, or a covertext into which the watermark is to be encoded, is represented by S^n. The channel input words X^n satisfy a distortion constraint with respect to the host data S^n. Thus the capacity is given by the maximum in (2), with an additional distortion constraint between X and S. Costa's result [5], with its extension to the vector case [27], has found applications in coding for the broadcast channel [2]. In particular, it turned out that dirty paper coding for the vector broadcast channel, a technique based on Costa's construction, is optimal in the sense that it exhausts the whole capacity region of the general MIMO broadcast channel [23], [24].

While for single user channels capacity formulas were derived for both, causal and non-causal side information, much less is known for multiuser models. Das and Narayan [8] studied causal coding for the state-dependent multiple access channel (MAC), with various degrees of side information available at the encoders and decoder. They considered the most general model - i.e., general channel statistics and general state process (need not be stationary and ergodic). Therefore, most of the formulas obtained are expressed as unions of regions characterized by limits of information quantities – that is, not a single-letter formula. The only case where these expressions collapse into a single-letter formula is when the following conditions are fulfilled: (a) the side information available to the encoders is a subset of the channel output, (b) the channel is memoryless and time invariant, (c) the state is stationary and ergodic, and (d) the coding starategies are restricted to depend only on a finite window of the respective side information. Condition (d) can be relaxed if we restrict the class of state processes: if the state is memoryless, then a single letter formula can be obtained even if the coding strategies have infinite memory with respect to the encoders side information.

Coding for multiuser models, with non-causal side information, was studied recently in two aspects: capacity regions for specific channel models, and random coding error exponents. Inner and outer bounds on the capacity region of the degraded broadcast channel were presented in [20]. The inner bound presented is tight for a special class of broadcast Gaussian channels, treated later in [13]. In addition, the inner bound of [20] is tight for the scenario where the stronger user has full side information, as the encoder. In [13], Kim, Sutivong, and Sigurjónsson derived the capacity region for three Gaussian channels with additive interference known non-causally at the encoders - the Gaussian multiple access channel, the Gaussian broadcast channel, and the physically degraded Gaussian relay channel. For the broadcast channel, it was assumed that the same interference appears at the two marginal channels, so the channel remains a degraded one. Similarly for the physically degraded relay channel: an interference signal is added to both, the direct and the rely path. For these three models, it was demonstrated that once the interference in known non-causally to the encoders, no loss in capacity is incurred relative to the case of no interference. The authors termed it as the *WDP property*. It should be noted that due to the WDP property, no converse has to be proved, that is, once a coding scheme is devised, that achieves that capacity of the channel without interference, it is clear that this scheme is optimal. In [12], a model is suggested, where one transmitter wishes to convey the same message to many different users via binary channels, each of which suffers from its own additive interference. All the interfering signals are known to the encoder. Under certain conditions on the joint statistics of the interfering signals, inner and outer bounds are derived on the capacity, which are tight for the case of two users. Random coding error exponents for the single-user and multiple access channels were derived in [19].

In this work we study the degraded broadcast channel with random parameters, with non-causal and causal coding. We provide the full proofs, with extensions, of the results presented in [20]. A general description of such channel is given in Figure 1. Inner and outer bounds are derived, on the capacity region of the degraded broadcast channel with random parameters known non-causally at the encoder. As a result of the bounds presented, a full characterization of the capacity region is given for the special case that the stronger user (the non-degraded component) has full knowledge about the state S. For the causal case, a full characterization of the capacity region is derived. This characterization is given in terms of auxiliary random variables, and can also be interpreted via Shannon strategies, as (1) for the single-user case.

In analogy to the single-user case, our results on non-causal coding can be applied to problems of watermaking, when one encoder is required to encode watermarks for more than one user, or when the encoded information is supposed to pass several stages of attacks, thus resulting in a degraded channel model. Another situation in which our results can be applied is where the encoder is supposed to convey information without a priori knowledge whether the decoder has an access to the host data S (the *private* version in the terminology of waternarking) or is ignorant of it (the *public* version). In this case a tradeoff exists between information that can be sent to public and that for private users. This is well represented by the capacity region of our degraded broadcast model.

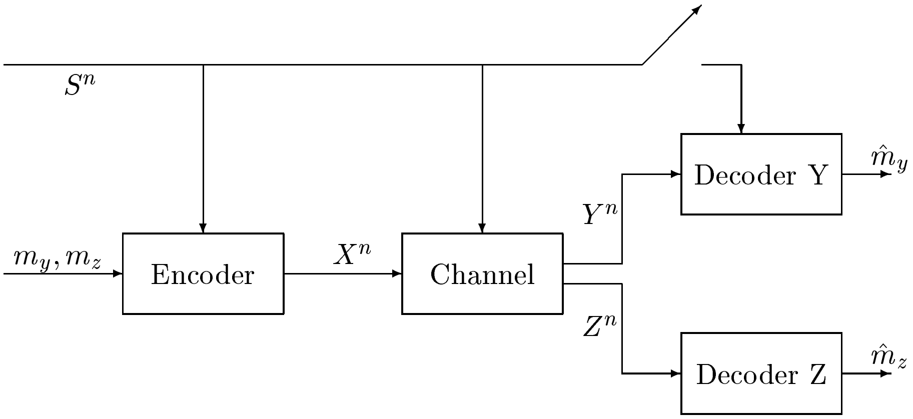

Figure 1: The broadcast channel with random parameters

This work is organized as follows. In Section 2 the basic definitions and notation are presented. All our main results, for causal and non-causal side information, are presented in Section 3. Section 3.1 is devoted to non-causal side information, and Section 3.2 to coding with causal side information. In Section 4 examples are given. In particular, the results of Section 3.2 are used to derive the capacity region of a certain class of binary broadcast channels. Finally, the proofs of our main results are given in Section 5.

2 Definitions

Let $\mathcal{X}$, $\mathcal{S}$, $\mathcal{Y}$, $\mathcal{Z}$ be finite sets, and let $P_S(\cdot)$ be a probability mass function (PMF) on $\mathcal{S}$. A broadcast channel with random parameters $(\mathcal{S}, P_S(s), \mathcal{X}, W(y,z|s,x), \mathcal{Y}, \mathcal{Z})$ is a channel with input alphabet $\mathcal{X}$, state space $\mathcal{S}$, output alphabet $\mathcal{Y} \times \mathcal{Z}$, and transition probability matrix $W(y,z|s,x)$, where the states s are random, taking values in $\mathcal{S}$ according to the PMF P_S. When the choice of the alphabets and of P_S is clear, we will refere to the broadcast channel just by $W(y,z|s,x)$. Throughout, we let y_m^n denote the vector $(y_m, y_{m+1}, \ldots, y_n)$. When the choice of n is clear, we use boldface letters to denote n-vectors, i.e., $\boldsymbol{y} = y_1^n = y^n$. We assume memoryless, time invariant channel and state parameters. I.e.,

$$W^n(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{s}, \boldsymbol{x}) = \prod_{i=1}^{n} W(y_i, z_i|s_i, x_i),$$

and

$$P_S^n(\boldsymbol{s}) = \prod_{i=1}^{n} P_S(s_i).$$

We denote the marginals by $W_{Y|S,X}(y|s,x)$ and $W_{Z|S,X}(z|s,x)$. A broadcast channel $W(y,z|s,x)$ is said to be *physically degraded* if we can write

$$W(y,z|s,x) = W_{Y|S,X}(y|s,x) W_{Z|Y}(z|y) \tag{3}$$

for some transition probability matrix $W_{Z|Y}$, in which case Z is said to be the degraded component. Note that according to this definition, the state parameter s controls only the non-degraded channel, whereas $W_{Z|Y}$, the channel from Y to Z, is independent of the states. A broadcast channel $W(y,z|s,x)$ is said to be *stochastically degraded* if there exists some transition probability matrix $W'_{Z|Y}$ such that

$$W_{Z|S,X}(z|s,x) = \sum_y W_{Y|S,X}(y|s,x) W'_{Z|Y}(z|y). \tag{4}$$

Definition 1 An (n, M_Y, M_Z, λ) noncausal code for the broadcast channel with random parameters $W(y,z|s,x)$ is an encoder map

$$f: \quad \{1,2,\ldots,M_Y\} \times \{1,2,\ldots,M_Z\} \times \mathcal{S}^n \longrightarrow \mathcal{X}^n, \tag{5}$$

and a pair of decoding maps

$$\begin{aligned} g_y: &\quad \mathcal{Y}^n \longrightarrow \{1,2,\ldots,M_Y\}, \\ g_z: &\quad \mathcal{Z}^n \longrightarrow \{1,2,\ldots,M_Z\}, \end{aligned}$$

such that the probability of error in decoding the messages is not larger than λ, i.e.,

$$\frac{1}{M_Y M_Z} \sum_{m_y=1}^{M_Y} \sum_{m_z=1}^{M_Z} \sum_{s^n \in \mathcal{S}^n} P_S^n(\boldsymbol{s}) W^n \left([g_Y^{-1}(m_y) \times g_Z^{-1}(m_z)]^c \mid \boldsymbol{s}, f(m_y, m_z, \boldsymbol{s}) \right) \leq \lambda. \tag{6}$$

The rate pair (R_Y, R_Z) of the code is defined as $(R_Y, R_Z) = 1/n(\log M_Y, \log M_Z)$. A rate pair (R_Y, R_Z) is said to be λ-achievable if for any $\gamma > 0$ and sufficiently large n there exists an $(n, 2^{n(R_Y - \gamma)}, 2^{n(R_Z-\gamma)}, \lambda)$ code for $W(y, z|s, x)$. The closure of all λ-achievable rate pairs is the λ-capacity region $\mathcal{C}(\lambda)$. The capacity region $\mathcal{C}$ is the closure of all rate-pairs (R_Y, R_Z) that are λ-achievable for every $\lambda > 0$.

The definition of an (n, M_Y, M_Z, λ) *causal* code for the channel $W(y, z|s, x)$ is similar to that of the noncausal code in Definition 1, except that the encoder consists of a *sequence* of maps $\{f_i\}_{i=1}^n$, where f_i is the mapping at time i. Thus, the encoder mapping f in (5) is replaced by

$$f_i: \quad \{1, 2, \ldots, M_Y\} \times \{1, 2, \ldots, M_Z\} \times \mathcal{S}^i \longrightarrow \mathcal{X}, \quad i = 1, 2, \ldots, n, \tag{7}$$

and, accordingly, the probability of error (6) is replaced by

$$\frac{1}{M_Y M_Z} \sum_{m_y=1}^{M_Y} \sum_{m_z=1}^{M_Z} \sum_{s^n \in \mathcal{S}^n} P_S^n(\boldsymbol{s}) W^n \left([g_Y^{-1}(m_y) \times g_Z^{-1}(m_z)]^c \mid \boldsymbol{s}, \boldsymbol{f}(m_y, m_z, \boldsymbol{s}) \right) \leq \lambda, \tag{8}$$

where $\boldsymbol{f}$ is the sequence of encoder outputs, i.e.,

$$\boldsymbol{f}(m_y, m_z, \boldsymbol{s}) = \boldsymbol{x} = \left(f_1(m_y, m_z, s_1), f_2(m_y, m_z, s^2), \ldots, f_n(m_y, m_z, s^n) \right).$$

The definitions of achievable rates and capacity region remain as in the noncausal case. We denote the capacity region of the channel W with causal coding by $\mathcal{C}_c$, where the subscript c stands for causal.

In this work we confine attention to the degraded broadcast channel with random parameters. Similar to the situation in the classical broadcast channel, the capacity regions (causal and noncausal) depend on $W(y, z|s, x)$ only via the marginal channels $W_Y(y|s, x)$ and $W_Z(z|s, x)$. Therefore, no distinction has to be made between stochastically degraded and physically degraded broadcast channels, and hereafter they will be termed as degraded broadcast channels with random parameters.

3 Main Results

3.1 Non-causal Side Information

Let $\mathcal{P}$ stand for the collection of all random variables (RVs) $(\tilde{K}, S, X, Y, Z)$ such that $\tilde{K}$ and X take values in the finite alphabets $\tilde{\mathcal{K}}$ and $\mathcal{X}$, respectrively, $\mathcal{X}$ is the input alphabet of the channel W, and

$$\begin{aligned} P_{\tilde{K},S,X,Y,Z}(\tilde{k}, s, x, y, z) &= P_{\tilde{K},X,S}(\tilde{k}, x, s) W(y, z|x, s) && (9) \\ \sum_{\tilde{k},x} P_{\tilde{K},X,S}(\tilde{k}, x, s) &= P_S(s). && (10) \end{aligned}$$

By (9), the following Markov relations hold

$$\tilde{K} \leftrightarrow (S, X) \leftrightarrow (Y, Z). \tag{11}$$

Define $\mathcal{R}_i$ to be the set of all rate pairs (R_Y, R_Z) such that

$$\begin{aligned} R_Z &\leq I(K; Z) - I(K; S) \\ R_Y &\leq I(U; Y|K) - I(U; S|K) \quad \text{for some } ((K, U), X, S, Y, Z) \in \mathcal{P}. && (12) \end{aligned}$$

We have the following properties of $\mathcal{R}_i$.

Proposition 1

1. *The set $\mathcal{R}_i$ is convex.*

2. *To exhaust $\mathcal{R}_i$, it is enough to take X to be a deterministic function of the triple (K, U, S).*

3. *To exhaust $\mathcal{R}_i$, it is enough to restrict $\mathcal{K}$ and $\mathcal{U}$ to satisfy*

$$\begin{aligned} |\mathcal{K}| &\leq |\mathcal{S}||\mathcal{X}| + 1, && (13) \\ |\mathcal{U}| &\leq |\mathcal{S}||\mathcal{X}| \left(|\mathcal{S}||\mathcal{X}| + 1\right). && (14) \end{aligned}$$

The proof is given in Section 5.1.

We have the following inner bound.

Theorem 1 *For any discrete memoryless degraded broadcast channel with random parameters,*

$$\mathcal{R}_i \subseteq \mathcal{C}.$$

The proof is based on a random code consisting of a combination of the code construction for the degraded broadcast channel [6] and that of Gel'fand and Pinsker [11]. It is given in Section 5.2

We state now the outer bound. Define $\mathcal{R}_o$ to be the set of all rate pairs (R_Y, R_Z) such that

$$\begin{aligned} R_Z &\leq I(K;Z) - I(K;S), \\ R_Y &\leq I(U;Y|K,V) - I(U;S|K,V), \\ R_Y + R_Z &\leq I(K,V,U;Y) - I(K,V,U;S), \quad \text{for some } ((K,V,U),S,X,Y,Z) \in \mathcal{P}. \end{aligned} \tag{15}$$

Proposition 2 *The set $\mathcal{R}_o$ is convex. Moreover, in order to exhaust the whole set $\mathcal{R}_o$, it is enough to restrict the alphabets of K, U, and V to satisfy*

$$\begin{aligned} |\mathcal{K}| &\leq |\mathcal{X}||\mathcal{S}| + 2 \\ |\mathcal{V}| &\leq |\mathcal{X}||\mathcal{S}|(|\mathcal{X}||\mathcal{S}| + 2) + 1 \\ |\mathcal{U}| &\leq (|\mathcal{X}||\mathcal{S}|(|\mathcal{X}||\mathcal{S}| + 2) + 1)\,(|\mathcal{X}||\mathcal{S}| + 2)|\mathcal{X}||\mathcal{S}| + 1 \end{aligned}$$

The proof is similar to the proof of Proposition 1, and is omitted.

Theorem 2 *For any discrete memoryless degraded broadcast channel with random parameters,*

$$\mathcal{C} \subseteq \mathcal{R}_o.$$

The proof is deferred to Section 5.3.

The cases where the state S is available to decoder Y, or to decoder Z, or to both, are special cases of definition 1. This can be viewed by incorporating the state as part of the corresponding channel output, Y, or Z, or both (see, for example, [3]). If decoder Z is informed, while decoder Y is kept ignorant, then (3) is not satisfied, and the channel is not a degraded one.

In what follows we restrict attention to a model where the state S ia available to Y but not to Z. It turns out that in such a case, the inner bound $\mathcal{R}_i$, given in (12), is tight, and admits a simpler form. Define the set $\mathcal{R}$ to be the collection of all pairs (R_Y, R_Z) such that

$$\begin{aligned} R_Z &\leq I(K;Z) - I(K;S), \\ R_Y &\leq I(X;Y|K,S), \quad (K,S,X,Y,Z) \in \mathcal{P}. \end{aligned} \tag{16}$$

Similarly to the statements made in Proposition 1 and Proposition 2, the following can be shown

Proposition 3 *1. The set $\mathcal{R}$ is convex*

2. To exhaust $\mathcal{R}$, it is enough to take X to be a deterministic function of the pair (K,S)

3. To exhaust $\mathcal{R}$, it is enough to limit the alphabet of K to

$$|\mathcal{K}| \leq |\mathcal{S}||\mathcal{X}| + 1. \tag{17}$$

The proof of Proposition 3 follows exactly the lines of the proof of Proposition 1, and is therefore omitted. We have the following result for the capacity region in case that the non-degraded user is informed

Theorem 3 *For any discrete memoryless degraded broadcast channel with random parameters and informed Y decoder*

$$\mathcal{C} = \mathcal{R}.$$

The proof is given in Section 5.4. A possible application of this model is a watermarking system [16] where the encoder is required to encode watermarks without knowing a priori whether the decoder has an access to the covertext S. I.e., two kinds of users are expected to decode the watermarks: private and public users. The covertext S is available to the Y decoder, and the channel from Y to Z, $W_{Z|Y}$, is an identity channel. This corresponds to the case where the switch in Figure 1 is closed.

3.2 Causal Side Information

For the case that the SI is provided to the encoder in a causal manner, we can characterize the transmission capacity region when both decoders are not informed. Let $\mathcal{P}_c$ be the collection of RVs

$(\tilde{K}, S, X, Y, Z)$ such that

$$\begin{aligned} (\tilde{K}, S, X, Y, Z) &\in \mathcal{P} && (18) \\ P_{\tilde{K},S} &= P_{\tilde{K}} P_S, && (19) \end{aligned}$$

that is, the subset of $\mathcal{P}$ where $\tilde{K}$ is independent of S. We define the region $\mathcal{R}_c$ as the collection of all pairs (R_Y, R_Z) satisfying

$$\begin{aligned} R_Z &\leq I(K; Z) \\ R_Y &\leq I(U; Y|K) \quad \text{for some } ((K,U), S, X, Y, Z) \in \mathcal{P}_c. && (20) \end{aligned}$$

Similarly to Propositions 1 and 3, the following properties of $\mathcal{R}_c$ hold.

Proposition 4 *1. The set $\mathcal{R}_c$ is convex*

2. To exhaust $\mathcal{R}_c$, it is enough to take X to be a deterministic function of the triplet (K, U, S)

3. The alphabets of K and U can be bounded as

$$\begin{aligned} \mathcal{K} &\leq |\mathcal{S}||\mathcal{X}| + 1 && (21) \\ \mathcal{U} &\leq |\mathcal{S}||\mathcal{X}|(|\mathcal{S}||\mathcal{X}| + 1) && (22) \end{aligned}$$

The proof follows the lines of the proof of Proposition 1 and is omitted. We have the following result for causal transmission.

Theorem 4 *For any discrete memoryless degraded broadcast channel with random parameters*

$$\mathcal{C}_c = \mathcal{R}_c.$$

The proof is given in Section 5.5.

Due to Proposition 4, it is possible to express the region $\mathcal{R}_c$ (and hence the capacity region $\mathcal{C}_c$) in terms of strategies, as in Shannon's formula for causal transmission via the single user channel. To that end, we first view briefly the result for the single user channel. The capacity of the single user channel with causal transmission is give by

$$C = \max I(U; Y) \tag{23}$$

where the maximization is over all random variables U independent of S, and where X is a deterministic function of the pair (U, S), i.e.,

$$\begin{aligned} P_{U,S} &= P_U P_S && (24) \\ x &= f(u,s), \quad \text{for some deterministic function } f. && (25) \end{aligned}$$

(The direct part follows easily from (20), with $R_Z = 0$. For the converse, substitute again $R_Z = 0$, and use $I(U;Y|K) \leq I(U,K;Y)$, defining a new auxiliary RV $\tilde{U} = (U,K)$.) Denote by $\mathcal{T}$ the familiy of all *Shannon strategies*, i.e., mappings from $\mathcal{S}$ to $\mathcal{X}$:

$$\mathcal{T} = \{t | t : \mathcal{S} \to \mathcal{X}\}\,. \tag{26}$$

Note that for a fixed u, $f(u,\cdot) \in \mathcal{T}$. Hence $\{f(u,\cdot)\}_u$ is a family of strategies, indexed by a parameter $u \in \mathcal{U}$. A distribution P_U on $\mathcal{U}$ induces a distribution $P_T(t)$ on the family of all strategies $\mathcal{T}$. Moreover, since U is independent of S, so is the ditribution of strategies, i.e., $P_{T|S}(t|s) = P_T(t)$. By the channel structure, we have

$$U -\!\!\circ\!\!- (T,S) -\!\!\circ\!\!- Y \tag{27}$$

and since the pair (T,U) is independent of S, we also have

$$U -\!\!\circ\!\!- T -\!\!\circ\!\!- Y. \tag{28}$$

To see (28), observe that

$$P_{U,T,S,Y} \overset{(a)}{=} P_{Y|T,S} P_{U|T,S} P_T P_S = P_{Y|T,S} P_{U,T} P_S = P_{Y,S|T} P_{U,T} \tag{29}$$

where (a) follows from (27), and (28) follows by summing (29) over s. For fixed f, a letter u defines an element $t \in \mathcal{T}$. A simple application of the data processing inequality and the chain rule for entropy, shows that if T is a deterministic function of U, and (28) holds, then necessarily $I(U;Y) = I(T;Y)$. Since T is independent of S, we conclude that (23)-(25) is equivalent to Shannon's formula.

We now give a similar interpretation of the capacity region $\mathcal{C}_c$. By Proposition 4, we can write $X = f(K,U,S)$, for some deterministic function f. For fixed k and u, we have $f(k,u,\cdot) \in \mathcal{T}$. Let us fix k, and put a distribution $P_{U|K}(\cdot|k)$ on $\mathcal{U}$. This induces a conditional distribution $P_{T|K}(t|k)$ on $\mathcal{T}$, conditioned on k. By the channel structure, we have

$$(U,K) -\!\!\circ\!\!- (T,S) -\!\!\circ\!\!- Y \tag{30}$$

from which we also have, as (U,K) is independent of S,

$$(U,K) \ominus T \ominus Y \tag{31}$$

For fixed f, a pair of letters (u,k) defines a unique $t \in \mathcal{T}$. Again, by (31), the fact that T is a deterministic function of (U,K), and the chain rule, we deduce that

$$I(U;Y|K) = I(U,K;Y|K) = I(T;Y|K). \tag{32}$$

Therefore the region $\mathcal{R}_c$ an be written as

$$\begin{aligned} \mathcal{R}_c = \bigcup_{P_K P_{T|K}} \{(R_Y,R_Z): \quad R_Z &\leq I(K;Z) \\ R_Y &\leq I(T;Y|K)\} \end{aligned} \tag{33}$$

where K is an auxiliary random variable, $P_{T|K}(\cdot|k)$ is a conditional distribution on the set of Shannon strategies $\mathcal{T}$, conditioned on $K=k$, and the pair (T,K) is independent of S.

4 Examples

The Gaussian broadcast channel, with additive interference, is modeled as

$$\begin{aligned} Y &= X+S+V_Y \\ Z &= X+S+V_Z \end{aligned} \tag{34}$$

where the state S is Gaussian, known non-causally at the transmitter, and V_Y, V_Z are Gaussian noise, unknown to all parties. Since the state S is common to the two channels, this model falls within the category of the channels treated in this paper. It was shown in [13] that the capacity region of the channel (34) with S known noncausally at the transmitter equals the capacity region of the Gaussian broadcast channel without interference. This can be shown by following closely the arguments of Costa [5] - either directly by suggesting a specific coding scheme, or by a proper choice of the random variable U and K that appear in $\mathcal{R}_i$ of Theorem 1, following the choice of the auxiliary random variable in [5]. Note that once it is proved that the capcity region without interference is achievable also in the model (34), there is no need to prove a converse. As this is already a known result, the details are omitted.

4.1 The Symmetric Broadcast Channel

Let $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathcal{W}_1 = \mathcal{W}_2 = \{0, 1, \ldots, |\mathcal{X}| - 1\}$. The symmetric, state dependent, degraded broadcast channel can be described as

$$\begin{aligned} Y &= X \oplus W_1 \qquad (35) \\ Z &= Y \oplus W_2 = X \oplus W_1 \oplus W_2 = X \oplus W \qquad (36) \end{aligned}$$

where W_1 is a state dependent noise, with conditional distribution $P_{W_1|S}(w_1|s)$, W_2 is an additive noise independent of W_1 and of the state S, and $\oplus$ is the modulo-$|\mathcal{X}|$ addition. We denote by $P_{W|S}$ the distribution of $W = W_1 \oplus W_2$ conditioned on S. The channel is assumed memoryless, that is,

$$P^n_{S,W_1,W_2}(s^n, w_1^n, w_2^n) = \prod_{i=1}^{n} P_{W_1|S}(w_{1,i}|s_i) P_S(s_i) P_{W_2}(w_{2,i}) \qquad (37)$$

We now use (33), to characterize the capacity region of this channel with causal SI. Following the approach employed in [10] for the single user channel, we can write

$$\begin{aligned} P_{Z|K}(z|k) &= \sum_{s,t} P_S(s) P_{T|K}(t|k) P_{Z|K,T,S}(z|k, t, s) \\ &= \sum_{s,t} P_S(s) P_{T|K}(t|k) P_{W|S}(z \ominus t(s)|s) \\ &= \sum_{t} P_{T|K}(t|k) P_r(W \oplus t(S) = Z) \qquad (38) \end{aligned}$$

where $\ominus$ stands for modulo-$|\mathcal{X}|$ subtraction. In a similar manner,

$$P_{Y|K}(y|k) = \sum_{t} P_{T|K}(t|k) P_r(W_1 \oplus t(S) = y) \qquad (39)$$

and

$$\begin{aligned} P_{Y|K,T}(y|k, t) &= \sum_{s} P_S(s) P_{Y|K,T,S}(y|k, t, s) = \sum_{s} P_S(s) P_{W_1|S}(y \ominus t(s)|s) \\ &= P_r(W_1 \oplus t(S) = y). \qquad (40) \end{aligned}$$

Using (38)-(40) in (33), we arrive at the following characterization

$$\begin{aligned} \mathcal{R}_c = \bigcup_{P_K P_{T|K}} \{(R_Y, R_Z) : \quad R_Z &\leq H(Z) - H(W \oplus T(S)|K) \\ R_Y &\leq H(W_1 \oplus T(S)|K) - H(W_1 \oplus T(S)|T)\}. \qquad (41) \end{aligned}$$

Note that the (random) strategy $T(S)$ appears in both, the equation for R_z and the equation for R_Y. Following [10], we can interpret t as a noise predictor, that strives to minimize the noise entropy (see also [9]). In contrast to the single-user case, we run here into some difficulties. First, the derivation of the capacity region of the broadcast channel composed of two cascaded symmetric DMCs is not an easy task, even for the case of no random parameters. This is due to the fact that there is no "Mrs. Gerber's Lemma" for the general alphabet case [25], [26]. Thus, although an achievability result can be suggested, proving the converse, i.e., that a certain input distribution exhausts the whole capacity region, is quite a hard problem. Returning to the case of random parameters, note that in general, the predictor that minimizes the entropy of $W_1 \oplus t(S)$ need not coincide with the one that minimizes $W \oplus t(S)$. Moreover, the random predictor T depends on the random variable K, which controls the tradeoff between the rates R_Z and R_Y. Nevertheless, for the binary case, we can further obtain from (41) an explicit formula for the capacity region.

Binary channel ($|\mathcal{X}| = 2$), with $\alpha = P_{W_2}(w_2 = 1) < 0.5$. Let $\mathcal{S} = \{1, 2\}$, and let the state-dependent noise distribution be $P_{W_1|S}(W_1 = 1|s) = \theta_s$, with $\theta_1 < 0.5 < \theta_2$. It is easy to verify, by direct calculation, that the strategy (state predictor) that minimizes the entropy of $W_1 \oplus t(S)$, minimizes also the entropy of $W \oplus t(S)$. Moreover, the optimal strategy t^* is given by

$$t^*(s) = \begin{cases} 0 & s = 1 \\ 1 & s = 2, \end{cases} \tag{42}$$

that is, t^* flips the input if $s = 2$, i.e., if the probability of having $w_1 = 1$ is greater than 0.5. Accordingly, we have

$$\begin{aligned} H(W_1 \oplus t^*(S)) &= h([\theta_1 P_S(1) + (1-\theta_2)P_S(2)]) & (43) \\ H(W \oplus t^*(S)) &= h([\theta_1 P_S(1) + (1-\theta_2)P_S(2)] \star \alpha) & (44) \end{aligned}$$

where $a \star b = a(1-b) + (1-a)b$, and h is the binary entropy fuction

$$h(\alpha) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha).$$

Let K be a binary random variable, with $P_K(0) = P_K(1) = 0.5$, fix $\beta \in (0,1)$, and define

$$P_{T|K}(t|k) = \begin{cases} \beta & \text{for} \quad t = t^* \oplus k \\ 1-\beta & \text{for} \quad t = t^* \oplus \bar{k}, \quad \text{where} \quad \bar{k} = (1-k). \end{cases} \tag{45}$$

We have for the conditional entropies

$$\begin{aligned} H(W_1 \oplus T(S)|K) &= H(W_1 \oplus T(S)|K=0)P_K(0) + H(W_1 \oplus T(S)|K=1)P_K(1) \\ &\quad = h([\theta_1 P_S(1) + (1-\theta_2)P_S(2)] \star \beta) & (46) \\ H(W \oplus T(S)|K) &= h([\theta_1 P_S(1) + (1-\theta_2)P_S(2)] \star \alpha \star \beta) & (47) \\ H(W \oplus T(S)|T) &= h([\theta_1 P_S(1) + (1-\theta_2)P_S(2)]) & (48) \end{aligned}$$

where (48) is due to the fact that all the realizations of the strategy are entropy minimizing, as they differ from t^* by a constant shift, k or $\bar{k}$ (see (45)). Next, since K is unifrmly distributed, the entropy of Z is maximized, regardless of the value of β. Substituting $H(Z) = 1$ and (46)-(48) in (41), we arrive at the following achievable region, parametrized by β:

$$R_Z \leq 1 - h(\tilde{\theta} \star \alpha \star \beta) \tag{49}$$

$$R_Y \leq h(\tilde{\theta} \star \beta) - h(\tilde{\theta}) \tag{50}$$

with

$$\beta \in (0, 1), \tag{51}$$

where

$$\tilde{\theta} = \theta_1 P_S(1) + (1 - \theta_2) P_S(2). \tag{52}$$

We claim that this achievable region is actually the capacity region. To prove this, we have to show that if a pair (R_Y, R_Z) is achievable, and

$$R_Y > h(\tilde{\theta} \star \beta) - h(\tilde{\theta}) \tag{53}$$

for some $\beta \in (0, 1)$, then necessarily

$$R_Z < 1 - h(\tilde{\theta} \star \alpha \star \beta) \tag{54}$$

Indeed, since $\min_t H(W_1 \oplus t(S)) = H(W_1 \oplus t^*(S)) = h(\tilde{\theta})$, the bound on R_Y in (41), with (53), imply

$$H(W_1 \oplus T(S)|K) > h(\tilde{\theta} \star \beta). \tag{55}$$

To proceed, we have to invoke a monotonicity argument. As $W = W_1 \oplus W_2$ and $P_{W_2}(1) = \alpha$, (55) in turn implies that

$$H(W \oplus T(S)|K) > h(\tilde{\theta} \star \alpha \star \beta), \tag{56}$$

where we have used [25, Corollary 4]. Since $H(Z)$ is upper bounded by 1, the bound on R_Z in (41) together with (56) yield (54).

The capacity region given by (49)-(51) coincides with the capacity region of the broadcast channel composed of two cascaded BSCs, with crossover probabilities $\tilde{\theta}$ for the first channel (from X to Y) and α for the second channel (from Y to Z) (see [26], or [6, Section 14.6]). This suggests an optimal coding scheme that resembles the one suggested by Erez and Zamir in [10]: design a code for a broadcast channel without random parameters, with crossover probabilities $\tilde{\theta}$ and α. Use this code for the current channel, followed by a state dependent noise predictor $t(s)$, that minimizes $H(W_1 \oplus t(S))$. This completes the binary example.

5 Proofs

5.1 Proof of Proposition 1

We start with the proof of convexity, that is, with the proof of Part 1 of the proposition. Introduce a time-sharing random variable Q, and define the joint distribution

$$\begin{aligned} P_{Q,K,U,X,S,Y,Z}(q,k,u,x,s,y,z,) &= P_{Q,K,U,X,S}(q,k,u,x,s)W(y,z|x,s) && (57) \\ \sum_{k,u,x} P_{Q,K,U,X,S}(q,k,u,x,s) &= P_Q(q)P_S(s), && (58) \end{aligned}$$

that is, S is independent of Q (since the state distribution is fixed and cannot be controlled by time-sharing scheme). Denote by (R_Y^Q, R_Z^Q) the rates resulting from time sharing. Then

$$\begin{aligned} R_Z^Q &= I(K;Z|Q) - I(K;S|Q) = I(K;Z|Q) - I(K,Q;S) \leq I(K,Q;Z) - I(K,Q;S) \\ &= I(\tilde{K};Z) - I(\tilde{K};S), \\ R_Y^Q &= I(U;Y|K,Q) - I(U;S|K,Q) = I(U;Y|\tilde{K}) - I(U;S|\tilde{K}), && (59) \end{aligned}$$

where $\tilde{K} = (K,Q)$, that is, we have incorporated the time sharing random variable Q into the auxiliary random variable. Therefore, we conclude that time sharing cannot yield rates that are not included in $\mathcal{R}_i$ defined above, and hence $\mathcal{R}_i$ is convex.

We proceed to prove Part 2 of Proposition 1. For that end, define

$$J(\lambda) = I(K;Z) - I(K;S) + \lambda \left[I(U;Y|K) - I(U;S|K) \right]. \tag{60}$$

To prove Part 2 of Proposition 1, it is enough to show that for every fixed $\lambda > 0$ and $P_{K,U,S}$, $J(\lambda)$ is maximized when X is a deterministic function of (K,U,S). In turn, to show this, it is enough to prove that $J(\lambda)$ is a convex $\cup$ function of $P_{X|K,U,S}$ for fixed $P_{K,U,S}$ and $\lambda > 0$. Thus, observe that

$$J(\lambda) = H(K|S) + \lambda H(U|K,S) + \psi_1(P_{X|K,U,S}) + \lambda \psi_2(P_{X|K,U,S}), \tag{61}$$

where

$$\begin{aligned} \psi_1(P_{X|K,U,S}) &= -H(K|Z) \\ \psi_2(P_{X|K,U,S}) &= -H(U|K,Y), \end{aligned}$$

and where we have used the special notation ψ_1 and ψ_2 instead of the entropies, to stress the dependence on $P_{X|K,U,S}$. To complete the proof we have to show that ψ_1 and ψ_2 are convex $\cup$ functions of $P_{X|K,U,S}$ for fixed $P_{K,U,S}$. But this follows from the concavity of the entropy function. Indeed, let

$$P_{X|K,U,S} = \alpha P^{(1)}_{X|K,U,S} + (1-\alpha) P^{(2)}_{X|K,U,S} \tag{62}$$

and let

$$\begin{aligned} P^{(i)}_{K,Z}(k,z) &= \sum_{u,s,x,y} P_{K,U,S}(k,u,s) P^{(i)}_{X|K,U,S}(x|k,u,s) P_{Y,Z|X,S}(y,z|x,s), \quad i=1,2 & (63) \\ P^{(i)}_{U,Y,K}(u,y,k) &= \sum_{s,x,z} P_{K,U,S}(k,u,s) P^{(i)}_{X|K,U,S}(x|k,u,s) P_{Y,Z|X,S}(y,z|x,s), \quad i=1,2. & (64) \end{aligned}$$

Now note that

$$\begin{aligned} \psi_1(P_{X|K,U,S}) &= \sum_{k,z} \left[\alpha P^{(1)}_{K,Z}(k,z) + (1-\alpha) P^{(2)}_{K,Z}(k,z)\right] \log \frac{\alpha P^{(1)}_{K,Z}(k,z) + (1-\alpha) P^{(2)}_{K,Z}(k,z)}{\alpha P^{(1)}_{Z}(z) + (1-\alpha) P^{(2)}_{Z}(z)} \\ &\overset{(a)}{\leq} \alpha \psi_1(P^{(1)}_{X|K,U,S}) + (1-\alpha)\psi_1(P^{(2)}_{X|K,U,S}) & (65) \end{aligned}$$

where (a) is due to the log-sum inequality. Similar argument holds for ψ_2. This completes the proof of Part 2 of Proposition 1.

Next, we prove Part 3 of the proposition. To prove that the region $\mathcal{R}_i$ is not altered if we bound the alphabet sizes as in (13), (14), we use the support lemma, which is a result of Carathéodory's theorem (see [7, p. 310]). Thus, let P_{UXS} stand for the (U,S,X) marginal of some distribution $P \in \mathcal{P}$. Without loss of generality, let us denote the product set $\mathcal{X} \times \mathcal{S} = \{1,2,\ldots,m\}$, $m = |\mathcal{X} \times \mathcal{S}|$. Define the following functionals

$$\begin{aligned} r_i(P_{U,S,X}) &= \sum_u P_{U,S,X}(u,s,x) = P_{S,X}(s,x), \quad i = 1,2,\ldots,m-1, \\ r_m(P_{U,S,X}) &= H(Z) - H(S) - \sum_{u,s,x} P_{U,S,X}(u,s,x) \log\left(\sum_{u'x'} P_{U,S,X}(u',s,x')\right) \\ &\quad + \sum_{u,x,s,y,z} P_{U,S,X}(u,s,x) W(y,z|x,s) \log\left(\sum_{u',x',y',s'} P_{U,S,X}(u',s',x') W(y',z|x',s')\right) \\ r_{m+1}(P_{U,S,X}) &= \sum_{u,x,s} P_{U,S,X}(u,s,x) \log\left(\sum_{u',x'} P_{U,S,X}(u',s,x')\right) \\ &\quad + \sum_{u,x,s} P_{U,S,X}(u,s,x) \log \frac{\sum_{s',x'} P_{U,S,X}(u,s',x')}{\sum_{x'} P_{U,S,X}(u,s,x')} \end{aligned}$$

$$
\begin{aligned}
&-\sum_{u,x,s,y,z} P_{U,S,X}(u,s,x)W(y,z|x,s)\log\left(\sum_{u',x',s',z'} P_{U,S,X}(u',s',x')W(y,z'|x',s')\right)\\
&-\sum_{u,x,s,y,z} P_{U,S,X}(u,s,x)W(y,z|x,s)\log\frac{\sum_{s',x'} P_{U,S,X}(u,s',x')}{\sum_{x',s',z'} P_{U,S,X}(u,s',x')W(y,z'|x',s')},
\end{aligned}
\tag{66}
$$

and observe that

$$
\begin{aligned}
\sum_k P_K(k)r_i(P_{U,S,X}(\cdot|k)) &= P_{S,X}(s,x), \quad i=1,2,\ldots,m-1,\\
\sum_k P_K(k)r_m(P_{U,S,X}(\cdot|k)) &= I(K;Z)-I(K;S),\\
\sum_k P_K(k)r_{m+1}(P_{U,S,X}(\cdot|k)) &= I(U;Y|K)-I(U;S|K)
\end{aligned}
$$

Therefore, by the support lemma [7, p. 310], the alphabet of the random variable K can be restricted as indicated in (13). Once the alphabet of K is fixed, we apply similar arguments to bound the alphabet of U, where this time we have $|\mathcal{S}||\mathcal{X}|\left(|\mathcal{S}||\mathcal{X}|+1\right)-1$ equations to preserve the joint distribution of S, X, and K, and one more equation to preserve $I(U;Y|K)-I(U;S|K)$, yielding the bound indicated in (14). □

5.2 Proof of Theorem 1

Before proving the direct part, we need some additional notation, and an elementary result that will be used in the random coding argument. Given a distribution P_{KU}, we denote by T_K^δ the set of all n-tuples $\boldsymbol{k}$ that are δ-typical according to P_K, i.e.,

$$
\begin{aligned}
T_K^\delta = \Big\{\boldsymbol{k}\in\mathcal{K}^n: &\left|\frac{1}{n}N(k|\boldsymbol{k})-P_K(k)\right|\le\delta \ \ \forall\, k\in\mathcal{K},\\
&\text{and } N(k|\boldsymbol{k})=0 \text{ whenever } P_K(k)=0\Big\},
\end{aligned}
$$

where $N(k|\boldsymbol{k})$ is the number of occurances of the letter k in the n-tuple $\boldsymbol{k}$. Similarly, we denote by $T_{U|K}^\delta(\boldsymbol{k})$ the set of all $\boldsymbol{u}$ that are $P_{U|K}$ δ-typical conditioned on a given $\boldsymbol{k}$, that is

$$
\begin{aligned}
T_{U|K}^\delta(\boldsymbol{k}) = \Big\{\boldsymbol{u}\in\mathcal{U}^n: &\left|\frac{1}{n}N(k,u|\boldsymbol{k},\boldsymbol{u})-\frac{1}{n}N(k|\boldsymbol{k})P_{U|K}(u|k)\right|\le\delta \ \ \forall\, k\in\mathcal{K}, u\in\mathcal{U},\\
&\text{and } N(k,u|\boldsymbol{k},\boldsymbol{u})=0 \text{ whenever } P_{U|K}(u|k)=0\Big\}.
\end{aligned}
$$

We will need the following known auxiliary results [7]. For any $\boldsymbol{s}\in T_S^\delta$ and any $\delta'>\delta$

$$
\exp\left(-nI(K;S)-n\epsilon_k\right)\le\sum_{\boldsymbol{k}:\ (\boldsymbol{k},\boldsymbol{s})\in T_{KS}^{\delta'}} P_K(\boldsymbol{k})\le\exp\left(-nI(K;S)+n\epsilon\right) \tag{67}
$$

where $\epsilon \to 0$ as $\delta, \delta' \to 0$, and $\epsilon_k \to 0$ as $\delta, \delta' \to 0$, but ϵ_k depends on P_K, thus the lower bound in (67) in not uniform over all P_K. Similarly, for any pair $(\boldsymbol{k}, \boldsymbol{s}) \in T^{\delta'}_{KS}$ and any $\delta'' > \delta'$

$$\exp\left(-nI(U;S|K) - n\eta_{u|k}\right) \leq \sum_{\boldsymbol{u}\colon\, (\boldsymbol{u},\boldsymbol{s}) \in T^{\delta''}_{US|K}(\boldsymbol{k})} P_{U|K}(\boldsymbol{u}|\boldsymbol{k}) \leq \exp\left(-nI(U;S|K) + n\eta\right) \qquad (68)$$

where $\eta, \eta_{u|k} \to 0$ as $\delta', \delta'' \to 0$, but $\eta_{u|k}$ depends on $P_{U|K}$, and thus the lower bound in (68) is not uniform over all conditional distributions $P_{U|K}$.

We use a random code consisting of a combination of the code construction for the degraded broadcast channel [6] and that of Gel'fand and Pinsker [11]. Fix a joint distribution $P \in \mathcal{P}$. Fix an arbitrary $\gamma > 0$, define $J_Y \overset{\triangle}{=} \exp(nI(U;S|K) + n\gamma)$, $J_Z \overset{\triangle}{=} \exp(nI(K;S) + n\gamma)$, $M_Y \overset{\triangle}{=} \exp(nI(U;Y|K) - nI(U;S|K) - 2n\gamma)$, and $M_Z \overset{\triangle}{=} \exp(nI(K;Z) - nI(K;S) - 2n\gamma)$. Generate an auxiliary collection a of $\boldsymbol{k}$-vectors

$$a = \{\boldsymbol{k}_{j_z,m_z},\ \ j_z \in \{1,2,\dots J_Z\},\ \ m_z \in \{1,2,\dots M_Z\}\} \qquad (69)$$

iid, independently of each other, according to P_K. For each vector $\boldsymbol{k} \in a$, generate a collection b of $\boldsymbol{u}$-vectors

$$b(\boldsymbol{k}_{j_z,m_z}) = \{\boldsymbol{u}_{j_y,m_y}(\boldsymbol{k}_{j_z,m_z}),\ \ j_y \in \{1,2,\dots J_Y\},\ \ m_y \in \{1,2,\dots M_Y\}\} \qquad (70)$$

independently of each other, according to $P_{U|K}(u_i|(\boldsymbol{k}_{j_z,m_z})_i)$. Reveal the collection a and the collections b to the encoder and decoder.

Encoding: Fix arbitrary parameters $0 < \delta < \delta_1$. Given a state vector $\boldsymbol{s}$ and a pair of messages indices m_y, m_z, let $j_z(\boldsymbol{s}, m_z)$ be the smallest integer j_z such that $(\boldsymbol{k}_{j_z,m_z}, \boldsymbol{s}) \in T^{\delta_1}_{KS}$. If such j_z does not exist, set $j_z(\boldsymbol{s}, m_z) = J_Z$. Let $j_y(\boldsymbol{s}, m_y, m_z)$ be the smallest j_y such that

$$(\boldsymbol{u}_{j_y,m_y}(\boldsymbol{k}_{j_z(\boldsymbol{s},m_z),m_z}), \boldsymbol{s}) \in T^{2\delta_1}_{US|K}(\boldsymbol{k}_{j_z(\boldsymbol{s},m_z),m_z}). \qquad (71)$$

If such j_y does not exist, set $j_y = J_Y$. For convenience, denote the pair $(\boldsymbol{k}, \boldsymbol{u})$ satisfying (71) by $\boldsymbol{ku}(\boldsymbol{s}, m_y, m_z)$. Finally, generate a vector of input letters $\boldsymbol{x} \in \mathcal{X}^n$ according to the memoryless distribution defined by the n-product of $P_{X|KUS}$

$$P^n_{X|KUS}(\cdot|\boldsymbol{ku}(\boldsymbol{s}, m_y, m_z), \boldsymbol{s}).$$

Decoding: We start with receiver Z. For convenience, define the parameters

$$\begin{aligned}
\delta_2 &= 2\delta_1(|\mathcal{S}| + 2)|\mathcal{US}| \\
\delta_3 &= \delta_1(|\mathcal{S}| + 2) \\
\delta_4 &= 4\delta_1(|\mathcal{S}| + 2)|\mathcal{S}|.
\end{aligned}$$

Given a vector $\boldsymbol{z} \in \mathcal{Z}^n$, the receiver looks for the set of all $\boldsymbol{k}_{j_z,m_z} \in a$ such that $(\boldsymbol{k}_{j_z,m_z}, \boldsymbol{z}) \in T^{\delta_2}_{KZ}$. Denote this set by $E_Z(\boldsymbol{z})$. If this set is nonempty, and all its elements have the same index message, say m'_z, then $g_z(\boldsymbol{z}) = m'_z$. Otherwize, i.e., if $E_Z(\boldsymbol{z})$ is empty, or has multiple elements with different message indices, set $g_z(\boldsymbol{z}) = M_Z$ and the decoder declares an error.

Receiver Y: Given a channel output vector $\boldsymbol{y}$, the decoding of the Y message is done in two steps. First, the decoder looks for a vector $\boldsymbol{k}_{j_z,m_z} \in a$ such that $(\boldsymbol{k}_{j_z,m_z}, \boldsymbol{y}) \in T^{\delta_2}_{KY}$. If such a vector does not exist, or there are more than one such vector, an error is declared. Denote the corresponding vector by $\hat{\boldsymbol{k}}$. Note that we have performed a full decoding of the index message m_z and also the index j_z. In the second step, the Y decoder looks for the set of all vectors $\boldsymbol{u}_{j_y,m_y}(\hat{\boldsymbol{k}}) \in b(\hat{\boldsymbol{k}})$ such that $(\boldsymbol{u}_{j_y,m_y}(\hat{\boldsymbol{k}}), \boldsymbol{y}) \in T^{\delta}_{UY|K}(\hat{k})$. Denote this set by $E_Y(\boldsymbol{y})$. If $E_Y(\boldsymbol{y})$ is nonempty and all its elements have the same message index, say m'_y, set $g_y(\boldsymbol{y}) = m'_y$. Otherwize, the decoder declares an error. Note that the Y receiver performs decoding of the full vector $\boldsymbol{k}_{j_z,m_z}$, and not only the message index m_z. The fact that Z can actually decode the full vector $\boldsymbol{k}$ can be proved following [11], with minor modifications. This fact was extensively used, for example, in [21]. Since Z is the degraded user, this can be done also by Y.

Probability of error: We analyse first the probability of error of receiver Z. For any $\boldsymbol{s} \in \mathcal{S}^n$, $\boldsymbol{k} \in a$, and any pair of message indices m_y, m_z, define the sets

$$\begin{aligned}
A_1(\boldsymbol{s}, m_z) &= \left\{\text{there is no } j_z \in \{1,2,\ldots,J_Z\} \text{ s.t. } (\boldsymbol{k}_{j_z,m_z}, \boldsymbol{s}) \in T^{\delta_1}_{KS}\right\} \\
A_2(\boldsymbol{s}, m_y|\boldsymbol{k}) &= \left\{\text{there is no } j_y \in \{1,2,\ldots,J_Y\} \text{ s.t. } (\boldsymbol{u}_{j_y,m_y}(\boldsymbol{k}), \boldsymbol{s}) \in T^{2\delta_1}_{US|K}(\boldsymbol{k})\right\} \\
A_3(\boldsymbol{s}, m_z) &= \left\{(\boldsymbol{k}_{j_z(\boldsymbol{s},m_z),m_z} \boldsymbol{z}) \notin T^{\delta_2}_{KZ}\right\} \\
A_4(\boldsymbol{s}, m_z) &= \left\{\exists\ \boldsymbol{k}_{j'_z,m'_z} \in a \text{ s.t. } m'_z \neq m_z \text{ and } (\boldsymbol{k}_{j'_z,m'_z}, \boldsymbol{z}) \in T^{\delta_2}_{KZ}\right\}.
\end{aligned}$$

The probability of error of the Z decoder can be upper bounded as

$$\begin{aligned}
P_{\mathrm{e},Z} \leq \frac{1}{M_Y M_Z} \sum_{m_y,m_z} \sum_{\boldsymbol{s} \in T^{\delta}_S} P_S(\boldsymbol{s}) \Big[& P(A_1) + P(A_2(\boldsymbol{s}, m_y|\boldsymbol{k}_{j_z(\boldsymbol{s},m_z),m_z})|A_1^c) \\
& + P(A_3|A_1^c, A_2^c(\boldsymbol{s}, m_y|\boldsymbol{k}_{j_z(\boldsymbol{s},m_z),m_z})) + P(A_4|A_1^c, A_2^c(\boldsymbol{s}, m_y|\boldsymbol{k}_{j_z(\boldsymbol{s},m_z),m_z}), A_3^c)\Big] \\
& + P_S({T^{\delta}_S}^c). \qquad (72)
\end{aligned}$$

For any message index m_z, any $\boldsymbol{s} \in T^{\delta}_S$ and $\delta < \delta_1$:

$$P(A_1(\boldsymbol{s}, m_z)) = P\left(\bigcap_{j_z=1}^{J_Z} \left\{(\boldsymbol{k}_{j_z,m_z}, \boldsymbol{s}) \notin T^{\delta_1}_{KS}\right\}\right)$$

$$\begin{aligned}
&= \left[P\left((\boldsymbol{k}_{1,m_z}, \boldsymbol{s}) \notin T_{KS}^{\delta_1}\right)\right]^{J_Z} \\
&= \left[1 - P\left((\boldsymbol{k}_{1,m_z}, \boldsymbol{s}) \in T_{KS}^{\delta_1}\right)\right]^{J_Z} \\
&\leq [1 - \exp(-nI(K;S) - n\epsilon_k)]^{\exp(nI(K;S)+n\gamma)} \\
&\leq \exp\left(-2^{n(\gamma-\epsilon_k)}\right) \qquad (73)
\end{aligned}$$

where

$$\epsilon_k \to 0 \;\; \text{as } \delta, \delta_1 \to 0, \qquad (74)$$

and where the first inequality in (73) is due to the lower bound in (67).

Conditioned on $A_1^c(\boldsymbol{s}, m_z)$, the vector $\boldsymbol{k}_{j_z(\boldsymbol{s},m_z),m_z}$ is in $T_{KS}^{\delta_1}$, and since $J_Y = \exp(nI(U;S|K) + n\gamma)$, we have

$$\begin{aligned}
&P\left(A_2(\boldsymbol{s}, m_z | \boldsymbol{k}_{j_z(\boldsymbol{s},m_z),m_z}) \mid A_1(\boldsymbol{s}, m_z)\right) = \\
&\quad = P\left(\bigcap_{j_y=1}^{J_Y} \left\{\left(\boldsymbol{u}_{j_y,m_y}(\boldsymbol{k}_{j_z(\boldsymbol{s},m_z),m_z}), \boldsymbol{s}\right) \notin T_{US|K}^{2\delta_1}\left(\boldsymbol{k}_{j_z(\boldsymbol{s},m_z),m_z}\right)\right\}\right) \\
&\quad = \left[1 - P\left(\left(\boldsymbol{u}_{1,m_y}(\boldsymbol{k}_{j_z(\boldsymbol{s},m_z)}), \boldsymbol{s}\right) \in T_{US|K}^{2\delta_1}(\boldsymbol{k}_{j_z(\boldsymbol{s},m_z),m_z})\right)\right]^{J_Y} \\
&\quad \leq \left[1 - \exp\left(-nI(U;S|K) - n\eta_{u|k}\right)\right]^{\exp(nI(U;S|K)+n\gamma)} \\
&\quad \leq \exp(-2^{n(\gamma-\eta_{u|k})}) \qquad (75)
\end{aligned}$$

where

$$\eta_{u|k} \to 0 \;\; \text{as } \delta, \delta_1 \to 0, \qquad (76)$$

and the first inequality in (75) is due to the lower bound in (68).

Conditioning on A_1^c, A_2^c, we have $\boldsymbol{kus} \in T_{KUS}^{\delta_3}$. Therefore,

$$P\left(\left\{\boldsymbol{kusz} \in T_{KUSZ}^{2\delta_3}\right\} \mid A_1^c, A_2^c\right) \longrightarrow 1 \quad \text{as } \; n \to \infty \qquad (77)$$

Moreover, the event $\boldsymbol{kusz} \in T_{KUSZ}^{2\delta_3}$ implies $\boldsymbol{kz} \in T_{KZ}^{\delta_2}$. Using this fact and (77), we have

$$P\left(A_3(\boldsymbol{s}, m_z) \mid A_1^c, A_2^c\right) \longrightarrow 0 \quad \text{as } \; n \to \infty. \qquad (78)$$

To evaluate $P(A_4|A_1^c, A_2^c, A_3^c)$, observe that conditioned on A_1^c, A_2^c, A_3^c, we have $\boldsymbol{z} \in T_Z^{\delta_2|\mathcal{K}|}$. In addition, for any $m_z' \neq m_z$, the vector $\boldsymbol{k}_{j_z', m_z'}$ was generated independently of the output vector $\boldsymbol{z}$.

Therefore

$$
\begin{aligned}
& P(A_4(\boldsymbol{s}, m_z)|A_1^c, A_2^c, A_3^c) = \\
&\quad = P\left(\bigcup_{j'_z m'_z :\ m'_z \neq m_z} \left\{ (\boldsymbol{k}_{j'_z m'_z}, \boldsymbol{z}) \in T_{KZ}^{\delta_2} \right\} \right) \\
&\quad \leq J_Z M_Z P\left((\boldsymbol{k}_{j'_z m'_z}, \boldsymbol{z}) \in T_{KZ}^{\delta_2} \right) \\
&\quad = J_Z M_Z \sum_{\boldsymbol{k}:\ (\boldsymbol{k}, \boldsymbol{z}) \in T_{KZ}^{\delta_2}} P_K(\boldsymbol{k}) \\
&\quad \leq J_Z M_Z \exp\left(-nI(K;Z) + n\epsilon\right) \\
&\quad = \exp(-n(\gamma - \epsilon))
\end{aligned}
\tag{79}
$$

where the second inequality is due to the upper bound in (67), and

$$
\epsilon \longrightarrow 0 \quad \text{as } \delta, \delta_1 \to 0. \tag{80}
$$

Clearly, for any $\delta > 0$

$$
P_S\left((T_S^\delta)^c \right) \longrightarrow 0 \qquad \text{as } n \to \infty. \tag{81}
$$

Collecting (73)–(81), we conclude that for any $\gamma > 0$ and sufficiently small δ, δ_1 $(\delta < \delta_1)$, the right hand side of (72) tends to 0 as $n \to \infty$.

As mentioned above, the decoding process in reciever Y is composed of two steps: in the first, the vector $\boldsymbol{k}$ is fully decoded. Then, based on the knowledge of $\boldsymbol{k}$, the message index m_y is decoded. Note that events A_1 and A_2 deal with the encoding process, and thus are relevant for both, Z and Y decoders. We define now the following events, analogous to events A_3, A_4:

$$
\begin{aligned}
B_1(\boldsymbol{s}, m_y) &= \left\{ (\boldsymbol{k}_{j_z(\boldsymbol{s}, m_z), m_z}, \boldsymbol{y}) \notin T_{KY}^{\delta_2} \right\} \\
B_2(\boldsymbol{s}, m_y) &= \left\{ \exists\ \boldsymbol{k}_{j'_z, m'_z} \in a \text{ s.t. } m'_z \neq m_z \text{ and } (\boldsymbol{k}_{j'_z, m'_z}, \boldsymbol{y}) \in T_{KY}^{\delta_2} \right\} \\
B_3(\boldsymbol{s}, m_y, m_z|\boldsymbol{k}) &= \left\{ (\boldsymbol{u}_{j_y(\boldsymbol{s}, m_y, m_z), m_y}(\boldsymbol{k}), \boldsymbol{y}) \notin T_{UY|K}^{\delta_5}(\boldsymbol{k}) \right\} \\
B_4(\boldsymbol{s}, m_y, m_z|\boldsymbol{k}) &= \left\{ \exists\ \boldsymbol{u}_{j'_y, m'_y}(\boldsymbol{k}) \in b(\boldsymbol{k}) \ \text{ s.t. } m'_y \neq m_y \text{ and } (\boldsymbol{u}_{j'_y, m'_y}, \boldsymbol{y}) \in T_{UY|K}^{\delta_5}(\boldsymbol{k}) \right\}.
\end{aligned}
$$

The probability of error of decoder Y can be upper bounded as

$$
\begin{aligned}
P_{\mathrm{e},Y} \leq\ & \frac{1}{M_Y M_Z} \sum_{m_y, m_z} \sum_{\boldsymbol{s} \in T_S^\delta} P_S(\boldsymbol{s}) \Big[P(A_1) + P(A_2(\boldsymbol{s}, m_y | \boldsymbol{k}_{j_z(\boldsymbol{s}, m_z), m_z}) | A_1^c) \\
& + \sum_{i=1}^{4} P(B_i | A_1^c, A_2^c, \cap_{l=1}^{i-1} B_l^c) \Big] \\
& + P_S({T_S^\delta}^c).
\end{aligned}
\tag{82}
$$

The first and second terms in the right hand side of (82) are treated in (73) and (75), and vanish as $n \to \infty$. As for the third and fourth terms, note that Z is the degraded user, therefore the claims

$$\begin{aligned}
\lim_{n\to\infty} P(B_1|A_1^c, A_2^c) &= 0, && (83)\\
\lim_{n\to\infty} P(B_2|A_1^c, A_2^c, B_1^c) &= 0 && (84)
\end{aligned}$$

are proved similar to (78), (79), and the details are omitted.

We treat now the fifth term, $P(B_3|A_1^c, A_2^c, B_1^c, B_2^c)$. Conditioned on $A_1^c A_2^c B_1^c B_2^c$, we have $\boldsymbol{kus} \in T_{KUS}^{\delta_3}$, and $\boldsymbol{k}$ was decoded correctly. Moreover, $\boldsymbol{k} \in T_K^{\delta_1|\mathcal{S}|}$, and $\boldsymbol{uy} \in T_{UY|K}(\boldsymbol{k})^{\delta_4}$. Therefore, similarly to (77) and (78), we have

$$\lim_{n\to\infty} P(B_3(\boldsymbol{s}, m_y, m_z|\boldsymbol{k})|A_1^c, A_2^c, B_1^c, B_2^c) = 0. \tag{85}$$

It remains to evaluate the probability of the event B_4. Conditioned on $A_1^c A_2^c B_1^c B_2^c B_3^c$, the vector $\boldsymbol{y}$ is typical. More precisely, $\boldsymbol{y} \in T_Y^{\delta_2|\mathcal{K}|}$. Moreover, for any $m_y' \neq m_y$, the vector $\boldsymbol{u}_{j_y', m_y'}(\boldsymbol{k})$ was generated independently of the output vector $\boldsymbol{y}$. Therefore

$$\begin{aligned}
P\left(B_4(\boldsymbol{s}, m_y, m_z \mid \boldsymbol{k})|A_1^c A_2^c \bigcap_{l=1}^{3} B_l^c\right) &= \\
&= P\left(\bigcup_{j_y' m_y' :\ m_y' \neq m_y} \left\{(\boldsymbol{u}_{j_y' m_y'}(\boldsymbol{k}), \boldsymbol{y}) \in T_{UY|K}^{\delta_4}(\boldsymbol{k})\right\}\right)\\
&\leq J_Y M_Y P\left((\boldsymbol{u}_{j_y' m_y'}(\boldsymbol{k}), \boldsymbol{y}) \in T_{UY|K}^{\delta_4}(\boldsymbol{k})\right)\\
&= J_Y M_Y \sum_{\boldsymbol{u}:\ (\boldsymbol{u}, \boldsymbol{y}) \in T_{UY|K}^{\delta_4}(\boldsymbol{k})} P_{U|K}(\boldsymbol{u}|\boldsymbol{k})\\
&\leq J_Y M_Y \exp(-nI(U;Y|K) + n\eta)\\
&= \exp(-n(\gamma - \eta)) && (86)
\end{aligned}$$

where the second inequality is due to the upper bound in (68), and

$$\eta \longrightarrow 0 \ \text{ as } n \to \infty. \tag{87}$$

Collecting (81) and (83)-(87), we conclude that for any$\gamma > 0$ and sufficiently small δ, δ_1 ($\delta < \delta_1$), the right hand side of (82) tends to 0 as $n \to \infty$. □

5.3 Proof of Theorem 2

For the proof of Theorem 2, we need an auxiliary result. Let $(n, \exp(nR_Y), \exp(nR_Z), \lambda)$ be a code for the broadcast channel with random parameters, and denote by m_y, m_z the random messages indices. Define the random variables

$$\begin{aligned} K_i &= m_z Z^{i-1} S_{i+1}^n \\ V_i &= Y^{i-1} \\ U_i &= m_y. \end{aligned}$$

Observe that the random variables so defined satisfy

$$((K_i, V_i, U_i), S_i, X_i, Y_i, Z_i) \in \mathcal{P}, \quad \forall i \in \{1, 2, \ldots, n\}. \tag{88}$$

We have

Lemma 1 The following inequalities hold

$$\begin{aligned} I(m_z; Z^n) - I(m_z; S^n) &\leq \sum_{i=1}^n I(K_i; Z_i) - I(K_i; S_i) \qquad (89) \\ I(m_y; Y^n | m_z) - I(m_y; S^n | m_z) &\leq \sum_{i=1}^n I(U_i; Y_i | K_i V_i) - I(U_i; S_i | K_i V_i) \qquad (90) \\ I(m_y m_z; Y^n) - I(m_y m_z; S^n) &\leq \sum_{i=1}^n I(K_i V_i U_i; Y_i) - I(K_i V_i U_i; S_i). \qquad (91) \end{aligned}$$

Proof The proof of (89) and (91) follows exactly the lines of the proof of Lemma 4 of [11], and is omitted. To show (90), we decompose the terms in the left hand side as

$$\begin{aligned} I(m_y; Y^n | m_z) \leq I(m_y; Y^n Z^n | m_z) &= \sum_{i=1}^n I(m_y; Y_i Z_i | Y^{i-1} Z^{i-1} m_z), \qquad (92) \\ I(m_y; S^n | m_z) &= \sum_{i=1}^n I(m_y; S_i | S_{i+1}^n m_z). \qquad (93) \end{aligned}$$

Each of the terms in these sums can be further written as

$$\begin{aligned} I(m_y; Y_i Z_i | Y^{i-1} Z^{i-1} m_z) &= I(m_y S_{i+1}^n; Y_i Z_i | Y^{i-1} Z^{i-1} m_z) - I(S_{i+1}^n; Y_i Z_i | Y^{i-1} Z^{i-1} m_y m_z) \\ &= I(S_{i+1}^n; Y_i Z_i | Y^{i-1} Z^{i-1} m_z) + I(m_y; Y_i Z_i | S_{i+1}^n Y^{i-1} Z^{i-1} m_z) \\ &\quad - I(S_{i+1}^n; Y_i Z_i | Y^{i-1} Z^{i-1} m_y m_z) \qquad (94) \\ I(m_y; S_i | S_{i+1}^n m_z) &= I(m_y Y^{i-1} Z^{i-1}; S_i | S_{i+1}^n m_z) - I(Y^{i-1} Z^{i-1}; S_i | S_{i+1}^n m_y m_z) \\ &= I(Y^{i-1} Z^{i-1}; S_i | S_{i+1}^n m_z) + I(m_y; S_i | S_{i+1}^n Y^{i-1} Z^{i-1} m_z) \\ &\quad - I(Y^{i-1} Z^{i-1}; S_i | S_{i+1}^n m_y m_z) \qquad (95) \end{aligned}$$

Thus

$$\begin{aligned} I(m_y;Y^n|m_z) &\le \sum_{i=1}^n I(m_y;Y_iZ_i|S_{i+1}^nY^{i-1}Z^{i-1}m_z)+\Delta_1-\Delta_2 \\ &= \sum_{i=1}^n I(m_y;Y_i|S_{i+1}^nY^{i-1}Z^{i-1}m_z)+\Delta_1-\Delta_2 \qquad (96) \\ I(m_y;S^n|m_z) &= \sum_{i=1}^n I(m_y;S_i|S_{i+1}^nY^{i-1}Z^{i-1}m_z)+\Delta_1^*-\Delta_2^* \qquad (97) \end{aligned}$$

where in the equality in (96) we have used the Markov structure of the channel, and where

$$\begin{aligned} \Delta_1 &= \sum_{i=1}^n I(S_{i+1}^n;Y_iZ_i|Y^{i-1}Z^{i-1}m_z) \\ \Delta_2 &= \sum_{i=1}^n I(S_{i+1}^n;Y_iZ_i|Y^{i-1}Z^{i-1}m_ym_z) \\ \Delta_1^* &= \sum_{i=1}^n I(Y^{i-1}Z^{i-1};S_i|S_{i+1}^nm_z) \\ \Delta_2^* &= \sum_{i=1}^n I(Y^{i-1}Z^{i-1};S_i|S_{i+1}^nm_ym_z) \end{aligned}$$

We claim that

$$\begin{aligned} \Delta_1 &= \Delta_1^* \qquad (98) \\ \Delta_2 &= \Delta_2^* \qquad (99) \end{aligned}$$

To see (98), observe that

$$\begin{aligned} I(S_{i+1}^n;Y_iZ_i|Y^{i-1}Z^{i-1}m_z) &= \sum_{j=i+1}^n I(S_j;Y_iZ_i|Y^{i-1}Z^{i-1}S_{j+1}^nm_z) \qquad (100) \\ I(Y^{i-1}Z^{i-1};S_i|S_{i+1}^nm_z) &= \sum_{j=1}^{i-1} I(Y_jZ_j;S_i|Y^{j-1}Z^{j-1}S_{i+1}^nm_z) \qquad (101) \end{aligned}$$

and therefore

$$\begin{aligned} \Delta_1 &= \sum_{i=1}^n\sum_{j=i+1}^n I(S_j;Y_iZ_i|Y^{i-1}Z^{i-1}S_{j+1}^nm_z) \qquad (102) \\ \Delta_1^* &= \sum_{i=1}^n\sum_{j=1}^{i-1} I(S_i;Y_jZ_j|Y^{j-1}Z^{j-1}S_{i+1}^nm_z)=\Delta_1. \qquad (103) \end{aligned}$$

The equality (99) is shown similarly. Equations (96), (97), (98), and (99) imply (90). □

Proof of Theorem 2 Let $(n, \exp(nR_Y), \exp(nR_Z), \lambda)$ be a code for the broadcast channel with random parameters, and denote by m_y, m_z the random messages indices. We can write the following chains of inequalities

$$\begin{aligned} nR_Z - n\epsilon_n &\overset{(a)}{\leq} I(m_z; Z^n) - I(m_z; S^n) \\ &\overset{(b)}{\leq} \sum_{i=1}^{n} I(K_i; Z_i) - I(K_i; S_i) \qquad (104) \\ nR_Y - n\epsilon_n &\overset{(c)}{\leq} I(m_y; Y^n|m_z) - I(m_y; S^n|m_z) \\ &\overset{(d)}{\leq} \sum_{i=1}^{n} I(U_i; Y_i|K_iV_i) - I(U_i; S_i|K_iV_i) \qquad (105) \\ n(R_Y + R_Z) - n\epsilon_n &\overset{(e)}{\leq} I(m_ym_z; Y^n) - I(m_ym_z; S^n) \\ &\overset{(f)}{\leq} \sum_{i=1}^{n} I(K_iV_iU_i; Y_i) - I(K_iV_iU_i; S_i) \qquad (106) \end{aligned}$$

where $\epsilon_n \to 0$ as $\lambda \to 0$, (a) (c) and (e) follow from Fano inequality and the fact that the messages are independent of each other and of the state sequence, (b) (d) and (f) result from Lemma 1. The statement of Theorem 2 follows now by applying to (104), (105) and (106) the standard time-sharing argument and taking the limits of large n and small probability of error λ. We show it here briefly. Let Q be a random variable independent of S, and uniformly distributed over $\{1, 2, \ldots, n\}$. We can rewrite (104), (105), and (106) as

$$\begin{aligned} R_Z - \epsilon_n &\leq I(K_Q; Z_Q|Q) - I(K_Q; S|Q) \qquad (107) \\ R_Y - \epsilon_n &\leq I(U_Q; Y_Q|K_Q, V_Q, Q) - I(U_Q; S|K_Q, V_Q, Q) \qquad (108) \\ R_Y + R_Z - \epsilon_n &\leq I(K_Q, V_Q, U_Q; Y_Q|Q) - I(K_Q, V_Q, U_Q; S|Q). \qquad (109) \end{aligned}$$

For every realization of Q, $Q = i$, $i \in \{1, 2, \ldots, n\}$, the relation (88) holds. Therefore, by the convexity of the set $\mathcal{R}_o$ (Proposition 2), we can replace the convex combination in (107)-(109) by the union over all random variables $((K, V, U), S, X, Y, Z) \in \mathcal{P}$. This completes the proof of Theorem 2. □

5.4 Proof of Theorem 3

Since the state S is available to the Y decoder, we can incorporate it into the output sequence Y^n. Set

$$\tilde{Y} \triangleq YS. \qquad (110)$$

We start with the converse part. Each of the terms in the right hand side of (105) can be written as

$$\begin{aligned} I(U_i;\tilde{Y}_i|K_i\tilde{Y}^{i-1}) - I(U_i;S_i|K_i\tilde{Y}^{i-1}) &= I(U_i\tilde{Y}^{i-1};\tilde{Y}_i|K_i) - I(U_i\tilde{Y}^{i-1};S_i|K_i) \\ &\quad - \Big(I(\tilde{Y}^{i-1};\tilde{Y}_i|K_i) - I(\tilde{Y}^{i-1};S_i|K_i)\Big) \\ &= I(U_i;\tilde{Y}_i|K_i) - I(U_i;S_i|K_i) \\ &\quad - \Big(H(\tilde{Y}^{i-1}|S_iK_i) - H(\tilde{Y}^{i-1}|\tilde{Y}_iK_i)\Big) \qquad (111) \end{aligned}$$

Therefore we get from (105) and (111)

$$\begin{aligned} nR_Y - n\epsilon_n &\leq \sum_{i=1}^{n} I(U_i\tilde{Y}^{i-1};Y_iS_i|K_i) - I(U_i\tilde{Y}^{i-1};S_i|K_i) \\ &\quad - \sum_{i=1}^{n}\Big(H(Y^{i-1}S^{i-1}|S_iK_i) - H(Y^{i-1}S^{i-1}|Y_iS_iK_i)\Big) \\ &\leq \sum_{i=1}^{n} I(U_i\tilde{Y}^{i-1};Y_iS_i|K_i) - I(U_i\tilde{Y}^{i-1};S_i|K_i) \\ &= \sum_{i=1}^{n} I(U_i\tilde{Y}^{i-1};Y_i|K_iS_i) \\ &\leq \sum_{i=1}^{n} I(X_i;Y_i|K_iS_i) \qquad (112) \end{aligned}$$

The upper bound on R_Z remains as in (104). The converse of Theorem 3 follows from (104) and (112), and the standard time sharing principle.

The direct part of Theorem 3 results from Theorem 1 by substituting $U \equiv X$ in $\mathcal{R}_i$, and using (110). □

5.5 Proof of Theorem 4

As in the single-user channel, the coding theorem for the causal case can be proved along the lines of the proof for the case of noncausal coding. Therefore, we do not give here the full proof, as it parallels many of the arguments used in the proofs of Theorems 1 and 2. Instead, we only highlight the points where the arguments differ from those used in the noncausal case.

We start with the direct part. Recall that in the proof of Theorem 1, the collections a and b of (69), (70) are generated independently of the realization of the state vector $\boldsymbol{s}$. Given a state vector $\boldsymbol{s}$ and a pair of message indices (m_y, m_z), the encoder seeks vectors $\boldsymbol{k}_{j_z,m_z} \in a$ and $\boldsymbol{u}_{j_y,m_y}(\boldsymbol{k}_{j_z,m_z}) \in$

$b(\boldsymbol{k}_{j_z,m_z})$, so that $\boldsymbol{k}_{j_z,m_z}$ and $\boldsymbol{u}_{j_y,m_y}(\boldsymbol{k}_{j_z,m_z})$ are jointly typical with the state vector $\boldsymbol{s}$. This is the point where the noncausality comes into account. If the pair (U, K) is independent of S, with high probability any pair of vectors $\boldsymbol{k}_{j_z,m_z}$ and $\boldsymbol{u}_{j_y,m_y}(\boldsymbol{k}_{j_z,m_z})$ are jointly typical with the realization of the state vector $\boldsymbol{s}$, so the encoding can proceed without reference to the specific realization of the states. From this point on, the encoding and decoding proceed as in the noncausal case.

As for the converse part, observe that if the encoder is causal, the random variables Y^{i-1}, K_i, V_i, and U_i in the proof of Theorem 2, are independent of S_i. Therefore the terms $I(K_i;S_i)$ and $I(U_i;S_i|K_iY^{i-1})$ in equations (104) and (105), respectively, can be dropped. We obtain the inequalities

$$nR_Z - n\epsilon_n \leq \sum_{i=1}^{n} I(K_i;Z_i) \tag{113}$$

$$nR_Y - n\epsilon_n \leq \sum_{i=1}^{n} I(U_i;Y_i|K_iY^{i-1}) \leq \sum_{i=1}^{n} I(U_iY^{i-1};Y_i|K_i), \tag{114}$$

Define now a new auxiliary random variable $\tilde{U}_i = (U_i, Y^{i-1})$. From this point on, we proceed as in the proof of the outer bound for the noncausal case (Theorem 2), with $\tilde{U}_i$ replacing U_i there. The details are omitted. □

References

[1] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 2007–2019, Sept. 1999.

[2] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[3] M. Chiang and T. M. Cover, "Unified duality between channel capacity and rate distortion with side information," *Proceedings of 2001 IEEE International Symposium on Information Theory*, Washington, D.C., June 24–29, 2001.

[4] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1639–1667, June 2002.

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

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley Inc., N.Y. 1992.

[7] I. Csiszár and J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Systems*. Academic Press, London, 1981.

[8] A. Das and P. Narayan, "Capacities of time-varying multiple-access channels with side information," *IEEE Trans. Inform. Theory*, vol. 48, mo. 1, pp. 4–25, January 2002.

[9] P. Elias, "Predictive coding," *IRE Trans. Inform. Theory*, vol. 1,pp. 16–33, Mar. 1955.

[10] U. Erez and R. Zamir, "Noise prediction for channels with side information at the transmitter," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1610–1617, July 2000.

[11] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters." *Probl. Inform. & Control*, vol. 9, no. 1,pp. 19–31, 1980.

[12] A. Khisti, U. Erez, and G. Wornell, "Writing on many pieces of dirty paper at once: the binary case," in *Proc. IEEE Int. Symp. Information Theory*. Chicago, June 27 – July 2, 2004.

[13] Y. H. Kim, A. Sutivong, and S. Sigurjónsson, "Multiple user writing on dirty paper," in *Proc. IEEE Int. Symp. Information Theory*. Chicago, June 27–July 2, 2004.

[14] A. V. Kusnetsov and B. S. Tsybakov, "Coding in a memory with defective cells," translated from *Prob. Peredach. Inform.*, vol. 10, no. 2, pp. 52–60, April–June 1974.

[15] A. V. Kusnetsov, T. Kasami, and S. Tamamura, "An error correcting scheme for defective memory," *IEEE Trans. Inform. Theory*,vol. 24, no. 6, pp. 712–718, Nov. 1978.

[16] P. Moulin and J. A. O'Sullivan, "Information theoretic analysis of information hiding," *IEEE Trans. Inform. Theory*, vol. 49, no. 3, pp. 563–593, March 2003.

[17] J. A. O'Sullivan, P. Moulin, and J. M. Ettinger, "Information theoretic analysis of steganography," in *Proc. Int. Symp. Information Theory*, Cambridge, MA, Aug. 1998, p. 297.

[18] C. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Devel.*, vol. 2, pp. 289–293, 1958.

[19] A. Somekh-Baruch and N. Merhav, "On the random coding error exponents for the single-user and the multiple-access Gel'fand-Pinsker channels," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, USA, June 27 – July 2, 2004.

[20] Y. Steinberg, "On the broadcast channel with random parameters," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, June 30 – July 5, 2002.

[21] Y. Steinberg and N. Merhav, "Identification in the presence of side information with application to watermarking." *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1410–1422, May 2001.

[22] B. S. Tsybakov, "Additive group codes for defect correction," translated from *Prob. Peredach. Inform.*, vol. 2, no. 1, pp. 111–113, January–March 1975.

[23] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," in *Proc. CISS 2004*, Princeton, NJ, March 2004.

[24] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," submitted.

[25] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part I," *IEEE Trans. Inform. Theory*, vol. 19, no. 6, pp. 769–772, Nov. 1973.

[26] A. D. Wyner, "A theorem on the entropy of certain binary sequences and applications: Part II," *IEEE Trans. Inform. Theory*, vol. 19, no. 6, pp. 772–777, Nov. 1973.

[27] W. Yu, A. Sutivong, D. Julian, T. M. Cover, and M. Chiang, "Writing on colored paper," in *Proc. IEEE Int. Symp. Information Theory*, Washington D.C., June 24–29, 2001.