The State-Dependent Semideterministic Broadcast Channel

Amos Lapidoth and Ligong Wang

Abstract—We derive the capacity region of the state-dependent semideterministic broadcast channel with noncausal state-information at the transmitter. One of the two outputs of this channel is a deterministic function of the channel input and the channel state, and the state is assumed to be known noncausally to the transmitter but not to the receivers. We show that appending the state to the deterministic output does not increase capacity.

We also derive an outer bound on the capacity of general (not necessarily semideterministic) state-dependent broadcast channels.

Index Terms—Broadcast channel, capacity region, channel-state information, Gel’fand-Pinsker problem, semideterministic.

I. INTRODUCTION

We characterize the capacity region of the discrete, memoryless, state-dependent, semideterministic broadcast channel. This channel has a single transmitting node, two receiving nodes, and an internal state, all of which are assumed to take value in finite sets. One of the receiving nodes—the “deterministic receiver”—observes a symbol $Y$ that is a deterministic function of the transmitted symbol $x$ and the (random) state $S$

$$Y = f(x, S) \quad \text{with probability one,}$$

and the other receiving node—the “nondeterministic receiver”—observes a symbol $Z$, which is random: conditional on the input being $x$ and the state being $s$, the probability that it equals $z$ is $W(z|x, s)$:

$$\Pr[Z = z | X = x, S = s] = W(z|x, s).$$

(1b)

The state sequence $S$ is assumed to be independent and identically distributed (IID) according to some law $P_S(\cdot)$

$$\Pr[S = s] = P_S(s)$$

(1c)

and to be revealed to the encoder in a noncausal way: all future values of the state are revealed to the transmitter before transmission begins.

We consider a scenario where the encoder wishes to convey two private messages: $M_y \in \{1, \ldots, 2^{nR_y}\}$ to the deterministic receiver, and $M_z \in \{1, \ldots, 2^{nR_z}\}$ to the nondeterministic receiver, where $R_y$ and $R_z$ denote the rates (in bits per channel use) of data transmission to the deterministic and nondeterministic receivers. The messages $M_y$ and $M_z$ are assumed to be independent and uniformly distributed. As for the broadcast channel without a state [1], [2], we define the capacity region of this channel as the closure of all rate-pairs that are achievable in the sense that the probability that at least one of the receivers decodes its message incorrectly can be made arbitrarily close to zero.

The main result of this paper is a single-letter characterization of the capacity region:

**Theorem 1:** The capacity region of the channel (1) when the states are known noncausally to the transmitter is the convex closure of the union of rate-pairs $(R_y, R_z)$ satisfying

$$R_y < H(Y|S)$$

(2a)

$$R_z < I(U; Z) - I(U; S)$$

(2b)

$$R_y + R_z < H(Y|S) + I(U; Z) - I(U; S, Y)$$

(2c)

over all joint distribution on $(X, Y, Z, S, U)$ whose marginal on $S$ is the given state distribution $P_S$ and under which, conditional on $X$ and $S$, the channel outputs $Y$ and $Z$ are drawn according to the channel law (1) independently of $U$:

$$P_{XYSU}(x, y, z, s, u) = P_S(s) P_{X|S}(x | s) 1\{y = f(x, s)\} W(z|x, s).$$

(3)

Here $1\{\cdot\}$ denotes the indicator function. Moreover, this is also the capacity region when the state sequence is also revealed to the deterministic receiver, i.e., when the mapping $f(\cdot, \cdot)$ is replaced by the mapping $(x, s) \mapsto (f(x, s), s)$.

**Proof:** See Sections II and III.

As to the cardinality of the auxiliary random variable $U$:

**Proposition 1:** To exhaust the capacity region of the channel (1), we may restrict the auxiliary random variable $U$ in (2) to take value in a set $\mathcal{U}$ whose cardinality $|\mathcal{U}|$ is bounded by

$$|\mathcal{U}| \leq |\mathcal{X}| \cdot |S| + 1,$$

(4)

where $\mathcal{X}$ and $S$ denote the input and state alphabets.

**Proof:** See Appendix A.

Broadcast channels without states have been studied extensively [3]. Our work can be considered as an extension to broadcast channels with states of prior work by Gel’fand, Marton, and Pinsker on deterministic and semideterministic broadcast channels without states [2], [4]–[8]. State-dependent

\footnote{To be precise, we should replace $2^{nR_y}$ and $2^{nR_z}$ with their integer parts, but, for typographical reasons, we shall not.}

\footnote{The value of $1\{\text{statement}\}$ is 1 if the statement is true and is 0 otherwise.
broadcast channels were also considered before [9]–[11], but capacity regions of most such channels are still unknown.

Steinberg [9] studied the degraded state-dependent broadcast channel with causal and with noncausal state-information at the transmitter. He derived the capacity region for the causal case, but for the noncausal case his outer and inner bounds do not coincide. Steinberg and Shamai [10] then derived an inner bound for general (not necessarily degraded) state-dependent broadcast channels with noncausal state-information. This inner bound is based on Marton’s inner bound for broadcast channels without states [7] and on Gel’fand-Pinsker coding [12]. In fact, the direct part of our Theorem 1 can be deduced from [10] with a proper choice of the auxiliary random variables (see Section II-A).

Our proof of the converse part of Theorem 1 borrows from the Gel’fand-Pinsker converse for single-user channels with states [12] as well as from the Körner-Marton [7] and the Nair-El Gamal [13] approaches to outer-bounding the capacity region of broadcast channels without states. But it also has a new element: the choice/definition of the auxiliary random variable depends on the codebook. As we demonstrate in Section V, our proof can be extended to general (not necessarily semideterministic) state-dependent broadcast channels.

Some special cases of Theorem 1 were solved by Khosravi-Farsani and Marvasti [11]: the semideterministic case, the case where the states are known to the nondeterministic transmitter contains the convex closure of the union of rate-pairs \( (R_y, R_z) \) satisfying

\[
R_y \leq I(U_0, U_y; Y) - I(U_0, U_y; S) \tag{5a}
\]

\[
R_z \leq I(U_0, U_z; Z) - I(U_0, U_z; S) \tag{5b}
\]

\[
R_y + R_z \leq [\max\{I(U_0; Y), I(U_0; Z)\} - I(U_0; S)]^+ + I(U_0, U_y; Y) - I(U_0, U_y; S) + I(U_0, U_z; Z) - I(U_0, U_z; S) - I(U_0, U_z; S), \tag{5c}
\]

where the union is over all joint distribution on \((X, Y, Z, S, U_0, U_y, U_z)\) whose marginal is \(P_S\); that satisfies the Markov condition

\[
(U_0, U_y, U_z) \rightarrow (X, S) \rightarrow (Y, Z); \tag{6}
\]

and under which the conditional law of \((Y, Z)\) given \((X, S)\) is that of the given channel.

For the semideterministic channel, we choose the auxiliary random variables in (5) as follows:

\[
U_0 = 0 \quad \text{(deterministic)} \tag{7a}
\]

\[
U_y = Y \tag{7b}
\]

\[
U_z = U. \tag{7c}
\]

Note that the Markov condition (6) is satisfied because \(Y\) is a deterministic function of \((X, S)\) and because in Theorem 1 we restrict \(U\) to be such that \(U \rightarrow (X, S) \rightarrow (Y, Z)\). With this choice of \(U_0, U_y, U_z, \) (5) reduces to (2).

B. Self-contained proof

We next provide a self-contained proof of the direct part of Theorem 1. As in [10, Theorem 1], our proof is based on Marton’s inner bound for general broadcast channels [7], [14] and on Gel’fand-Pinsker coding [12].

First note that the joint distribution (3) can also be written as

\[
P_{XYSU}(x, y, z, s, u) = P_S(s) P_{Y|S}(y, u|s) P_{X|YSU}(x|y, s, u) W(z|x, s) \tag{8}
\]

with the additional requirement that

\[
y = f(x, s). \tag{9}
\]

Further note that, when \(P_{YSU}\) is fixed, all the terms on the right-hand side (RHS) of (2) are fixed except for \(I(U; Z)\), which is convex in \(P_{X|YSU}\). Since \(I(U; Z)\) only appears with a positive sign on the RHS of (2), it follows that the union over all joint distributions of the form (2) can be replaced by a union only over those where \(x\) is a deterministic function of \((y, u, s)\), i.e., of the form

\[
P_{XYSU}(x, y, z, s, u) = P_S(s) P_{Y|S}(y, u|s) \mathbf{1}\{x = g(y, u, s)\} W(z|x, s) \tag{10}
\]

for some \(g: (y, u, s) \rightarrow x\) (and subject to (9)). We shall thus only establish the achievability of rate pairs that satisfy (2) for some distribution of the form (10).

Choose a stochastic kernel \(P_{YSU}\) and a mapping \(g: (y, u, s) \rightarrow x\) which, combined with \(P_S\) and the channel law, determines the joint distribution (10) for which (9) is satisfied. For a given block-length \(n\), we construct a random code as follows:

**Codebook:** Generate \(2^{nR_y}\) \(y\)-bins, each containing \(2^{nR_y}\) \(y\)-tuples where the \(l_y\)-th \(y\)-tuple in the \(m_y\)-th bin

\[
y(m_y, l_y), \quad m_y \in \{1, \ldots, 2^{nR_y}\}, \quad l_y \in \{1, \ldots, 2^{nR_y}\}
\]

is generated IID according to \(P_Y\) (the \(Y\)-marginal of (10)) independently of the other \(y\)-tuples. Additionally, generate
\[2^{nR_2}\] u-bins, each containing \(2^{nR_2}\) u-tuples, where the \(l_z\)-th u-tuple in the \(m_z\)-th u-bin
\[u(m_z, l_z), \ m_z \in \{1, \ldots, 2^{nR_2}\}, \ l_z \in \{1, \ldots, 2^{nR_2}\}\]
is drawn IID according to \(P_U\) (the \(U\)-marginal of (10)) independently of the other u-tuples and of the \(y\)-tuples.

**Encoder:** To send Message \(m_y \in \{1, \ldots, 2^{nR_2}\}\) to the deterministic receiver and Message \(m_z \in \{1, \ldots, 2^{nR_2}\}\) to the nondeterministic receiver, look for a u-tuple \(y(m_y, l_y)\) in \(y\)-bin \(m_y\) and a u-tuple \(u(m_z, l_z)\) in \(u\)-bin \(m_z\) such that \((y(m_y, l_y), u(m_z, l_z))\) is jointly typical with the state sequence:
\[(y(m_y, l_y), u(m_z, l_z), s) \in T_{T_e(n)}(P_{YUS}), \quad (11)\]
where \(T_{T_e(n)}(\cdot)\) denotes the \(\epsilon\)-strongly typical set with respect to a certain distribution. If such a pair can be found, send
\[x = g(y(m_y, l_y), u(m_z, l_z), s), \quad (12)\]
where in the above \(g(y, u, s)\) denotes the application of the function \(g(y, u, s)\) componentwise. (Note that in this case the sequence received by the deterministic receiver will be \((y(m_y, l_y))\).) Otherwise send an arbitrary codeword.

**Deterministic decoder:** Try to find the unique \(u\)-bin, say \(m'_y\) that contains the received sequence \(y\) and output its number \(m'_y\). If there is more than one such bin, declare an error.

**Nondeterministic decoder:** Try to find the unique \(u\)-bin \(m'_z\) which contains a \(u(m'_z, l'_z)\) that is jointly typical with the received sequence \(z\):
\[(u(m'_z, l'_z), z) \in T_{2\epsilon}(P_{UZ}), \quad (13)\]
and output \(m'_z\). If more than one or no such bin can be found, declare an error.

We next analyze the error probability of the above coding scheme. There are three types of errors:

**Encoder errors.** This happens only if there is no pair \((l_y, l_z) \in \{1, \ldots, 2^{nR_2}\} \times \{1, \ldots, 2^{nR_2}\}\) that satisfies (11).

To bound this probability, we use the Multivariate Covering Lemma [2, Lemma 8.2], which we restate as follows:

**Lemma 1:** Fix some joint distribution \(P_{A(0)\ldots A(k)}\) on \((A(0), \ldots, A(k))\), and fix positive \(\epsilon\) and \(\epsilon\) with \(\epsilon < \epsilon\). Let \(A^n(0)\) be a random sequence satisfying
\[\lim_{n \to \infty} \Pr [A^n(0) \in T_{T_e(n)}(P_{A(0)})] = 1. \quad (14)\]
For each \(j \in \{1, \ldots, k\}\), let \(A^n_{(j)}(m_j), m_j \in \{1, \ldots, 2^{nR_{(j)}}\}\), be pairwise independent conditional on \(A^n(0)\), each distributed according to \(\prod_{j=1}^k P_{A_{(j)}|A(0)=m_{(0)}, \ldots, a_{(j-1)}}\). Assume that
\[\left\{A^n_{(j)}(m_j): m_j \in \{1, \ldots, 2^{nR_{(j)}}\}\right\}, \ j \in \{1, \ldots, k\}\]
are mutually independent conditional on \(A^n(0)\). Then there exists \(\delta(\epsilon)\) which tends to zero as \(\epsilon\) tends to zero such that
\[\lim_{n \to \infty} \Pr [A^0(0), A^n_{(1)}(m_1), \ldots, A^n_{(k)}(m_k) \notin T_{T_e(n)}(\forall (m_1, \ldots, m_k))] = 0 \quad (15)\]
provided that, for all \(\mathcal{J} \subseteq \{1, \ldots, k\}\) with \(|\mathcal{J}| \geq 2\),
\[\sum_{j \in \mathcal{J}} r_j > \sum_{j \in \mathcal{J}} H(A_{(j)}|A(0)) - H(\{A_{(j)}: j \in \mathcal{J}\}|A(0)) + \delta(\epsilon), \quad (16)\]
where the conditional entropies are computed with respect to \(P_{A_{(0)}\ldots A_{(k)}}\).

We apply Lemma 1 by choosing \(k = 3, A(0) = 0\) (deterministic) so \(\epsilon = 0\), and
\[A(1) = Y, \ r_1 = \tilde{R}_y, \quad (17a)\]
\[A(2) = U, \ r_2 = \tilde{R}_z, \quad (17b)\]
\[A(3) = S, \ r_3 = 0. \quad (17c)\]
The joint distribution is chosen to be \(P_{YUS}\). We then obtain that the probability that the encoder errrs tends to zero as \(n\) tends to infinity provided that
\[\tilde{R}_y > I(Y; S) + \delta(\epsilon) \quad (18a)\]
\[\tilde{R}_z > I(U; S) + \delta(\epsilon) \quad (18b)\]
\[\tilde{R}_y + \tilde{R}_z > H(Y) + H(U) + H(S)
- H(Y, U, S) + \delta(\epsilon). \quad (18c)\]

**Deterministic decoder errors.** This happens only if there is more than one bin that contains the received \(y\). We may now assume that the encoding was successful so (11) is satisfied. Then \(y\) is in \(T_{T_e(n)}(P_Y)\), and
\[P_Y(y) \leq 2^{-nH(Y) - \delta(\epsilon)} \quad (19)\]
where \(\delta(\epsilon)\) tends to zero when \(\epsilon\) tends to zero. Hence the probability that a specific \(y\)-tuple in a bin that was not chosen by the encoder, which, by our code construction, was independently chosen from the received \(y\), happens to be the same as \(y\), is upper-bounded by the RHS of (19). Further note that the total number of \(y\)-tuples outside the bin chosen by the encoder is \(2^{nR_2} - (2^{nR_2} - 1)\). Using the union bound, we obtain that the probability that the deterministic decoder errrs is at most
\[2^{nR_2} (2^{nR_2} - 1) 2^{-n(H(Y) - \delta(\epsilon))}, \quad (20)\]
which tends to zero as \(n\) tends to infinity provided that
\[R_y + \tilde{R}_y < H(Y) - \delta(\epsilon). \quad (21)\]

**Nondeterministic decoder errors.** This happens if either the \(u\)-tuple \((m_z, l_z)\) is not jointly typical with the received \(z\)-tuple, or if a \(u\)-tuple in a different bin happens to be jointly typical with the received \(z\)-tuple. Assuming that the encoding was successful, the probability of the former case tends to zero as \(n\) tends to infinity by (11) and by the Markov Lemma [2, Lemma 12.1]. To upper-bound the probability of the latter case, note that any \(u(m'_z, l'_z)\), where \(m'_z \not= m_z\), is chosen independently of \((m_z, l_z)\) and \(y(m_y, l_y)\), and is hence also independent of the received \(z\). By the Joint Typicality Lemma [2, p.29] we have
\[\Pr [U(m'_z, l'_z), Z) \in T_{2\epsilon}(P_{UZ})] = 2^{-n(I(U; Z) - \delta(\epsilon))} \quad (22)\]
where the probability is computed with respect to the randomly chosen codebook. Next note that the total number of such
entropy.\footnote{Let \(\epsilon_n\) tend to zero as \(n\) tends to infinity. Here, (25) follows from Fano’s Inequality; (26) and (27) because \(M_y\) and \(S^n\) are independent; (28) from the chain rule; (29) by dropping negative terms; and (30) because conditioning cannot increase entropy. We next bound \(R_z\) as in [12]:
\[
R_z \leq H(M_z) = H(M_z; Z^n) + n\epsilon_n
\]
\[
= \sum_{i=1}^{n} I(M_z; Z_i | Z^{i-1}) - \sum_{i=1}^{n} I(S^n_{i+1}; Z_i | M_z, Z^{i-1}) + n\epsilon_n
\]
\[
= \sum_{i=1}^{n} I(M_z, S^n_{i+1}; Z_i | Z^{i-1}) - \sum_{i=1}^{n} I(S^n_{i+1}; Z_i | M_z, Z^{i-1}) + n\epsilon_n
\]
\[
\leq \sum_{i=1}^{n} I(M_z, Z^{i-1}; S^n_{i+1}, Y^n_{i+1}; S_i) + n\epsilon_n
\]
\[
\leq \sum_{i=1}^{n} I(M_z, Z^{i-1}; S^n_{i+1}, Z_i) + n\epsilon_n
\]
\[
\leq \sum_{i=1}^{n} I(V_i; Z_i) - I(V_i; S_i) + n\epsilon_n
\]
Here, (32) follows from Fano’s Inequality; (33) and (34) from the chain rule; (35) from Csiszár’s Identity [15]
\[
\sum_{i=1}^{n} I(C^n_{i+1}; D_i | D^{i-1}) = \sum_{i=1}^{n} I(D^{i-1}; C_i | C^n_{i+1});
\]
(36) because \(S_i\) and \((M_z, S^n_{i+1})\) are independent; (37) from the chain rule and by dropping negative terms; and (38) by defining the auxiliary random variables
\[
V_i \triangleq (M_z, Z^{i-1}, S^n_{i+1}), \quad i \in \{1, \ldots, n\}.
\]
We next bound the sum rate \(R_y + R_z\):
\[
R_y + R_z = H(M_y, M_z) = H(M_z) + H(M_y | M_z)
\]
\[
\leq I(M_z; Z^n) + I(M_y; Y^n | S^n) + H(M_y | M_z)
\]
\[
\leq I(M_z; Z^n) + I(M_y; Y^n | S^n) + n\epsilon_n,
\]
Here, (44), (45), and (46) follow from the chain rule; (47) by applying Csiszár’s Identity (39) between \((S^n, Y^n)\) and \(Z^n\); and (48) again from the chain rule.

We next study the sum of the last term on the RHS of (48) and the second mutual information on the RHS of (43):

\[
\sum_{i=1}^{n} I(M_z, S^n_{i+1}, Y^n_{i+1}; S_i, Y_i) + I(M_y; Y^n, S^n|M_z)
\]

\[
= \sum_{i=1}^{n} I(M_z, S^n_{i+1}, Y^n_{i+1}; S_i, Y_i)
\]

\[
+ \sum_{i=1}^{n} I(M_y, S^n_{i+1}, Y^n_{i+1})
\]

\[
= \sum_{i=1}^{n} I(M_y, M_z, S^n_{i+1}, Y^n_{i+1}; S_i, Y_i)
\]

\[
+ \sum_{i=1}^{n} I(S^n_{i+1}; S_i, Y_i|M_z, M_y, S^n_{i+1}, Y^n_{i+1})
\]

(49)

\[
= \sum_{i=1}^{n} I(M_y, M_z, S^n_{i+1}, Y^n_{i+1}; S_i, Y_i)
\]

(50)

(51)

Here, (49) and (50) follow from the chain rule; (51) by applying Csiszár’s Identity between \((S^n, Y^n)\) and \(S^n\); (52) from the chain rule; (53) because \(S_i\) and \((M_y, M_z, S^n_{i+1})\) are independent; (54) again from the chain rule; and (55) because, given \((M_y, M_z, S^n)\), the channel inputs \(X^n\) are determined by the encoder, and hence \(Y^n\) are also determined, so

\[
H(Y_i|S_i) = 0.
\]

Combining (43), (48), and (55), using the definitions (40), and further defining

\[
T_i \triangleq Y^n_{i+1}, \quad i \in \{1, \ldots, n\}
\]

(57)

we obtain

\[
n(R_y + R_z) \leq \sum_{i=1}^{n} I(V_i, T_i; Z_i) - \sum_{i=1}^{n} I(V_i, T_i; S_i, Y_i)
\]

\[
+ \sum_{i=1}^{n} H(Y_i|S_i) + n\epsilon_n.
\]

(58)

Summarizing (30), (38), and (58) and letting \(n\) tend to infinity, we obtain that any achievable rate-pair \((R_y, R_z)\) must be contained in the convex closure of the union of rate-pairs satisfying

\[
R_y < H(Y|S)
\]

(59a)

\[
R_z < I(V; Z) - I(V; S)
\]

(59b)

\[
R_y + R_z < H(Y|S) + I(V, T; Z) - I(V, T; S, Y)
\]

(59c)

where, given \((X, S)\), the outputs \((Y, Z)\) are drawn according to the channel law (1) independently of the auxiliary random variables \((V, T)\).

To prove the converse part of Theorem 1, it remains to replace \(V\) and \(T\) with a single auxiliary random variable. I.e., it remains to find an auxiliary random variable \(U\) such that

\[
I(V; Z) - I(V; S) \leq I(U; Z) - I(U; S).
\]

(60a)

and

\[
H(Y|S) + I(V, T; Z) - I(V, T; S, Y) \leq H(Y|S) + I(U; Z) - I(U; S, Y).
\]

(60b)

In fact, as we shall see, either choosing \(U\) to be \(V\) will satisfy (60) or else choosing it to be \((V, T)\) will satisfy (60). If we choose \(U = V\), then (60a) is satisfied with equality, and the requirement (60b) becomes

\[
I(T; Z|V) - I(T; S, Y|V) \leq 0.
\]

(61)

On the other hand, if we choose \(U = (V, T)\), then (60b) is satisfied with equality, and the requirement (60a) becomes

\[
I(T; Z|V) - I(T; S|V) \geq 0.
\]

(62)

It remains to show that at least one of the two requirements (61) and (62) must be satisfied: if it is (61), then we shall choose \(U\) as \(V\), and if it is (62), then we shall choose \(U\) as \((V, T)\). To this end we note that for all random variables \(T, Z, V, S, Y\)

\[
I(T; Z|V) - I(T; S, Y|V) \leq I(T; Z|V) - I(T; S|V).
\]

(63)

because the RHS minus the left-hand side is \(I(T; Y|S, V)\), which is nonnegative. This implies that at least one of (61) and (62) must hold. We have thus shown that there must exist a \(U\) which satisfies both inequalities in (60), hence the bounds (59) can be relaxed to (2). This concludes the proof of the converse part of Theorem 1.

IV. AN EXAMPLE

Consider a broadcast channel whose input, output, and state alphabets are all binary and whose law is

\[
P_S(1) = 1 - P_S(0) = \sigma
\]

(64a)

\[
Y = x \oplus S
\]

(64b)

\[
W(Z = x|x, s) = 1 - W(Z = x \oplus 1|x, s) = 1 - p
\]

(64c)

for some constants \(0 \leq p, \sigma \leq 1\). The deterministic output \(Y\) of this channel is the modulo-two sum of the input \(x\) and the state \(S\), and the channel from \(x\) to the nondeterministic
output $Z$ is unaffected by the state and is a binary symmetric channel with crossover probability $p$.

To cancel the state’s effect, the encoder could flip the input $x$ whenever $S = 1$, but this would hurt the nondeterministic receiver. In fact, if the state is unbiased ($\sigma = 0.5$), and if only causal state-information is available at the encoder, then one cannot do better than time-sharing:

**Proposition 2:** The capacity region of the channel (64) with $\sigma = 0.5$ when the states are known causally to the transmitter but not to the receivers, is the union over $\lambda \in [0, 1]$ of rate-pairs $(R_y, R_z)$ satisfying

\[
R_y \leq \lambda \quad (65a) \\
R_z \leq (1 - \lambda)(1 - H_b(p)). \quad (65b)
\]

I.e., it is the collection of rate pairs satisfying

\[
R_y + \frac{R_z}{1 - H_b(p)} \leq 1. \quad (66)
\]

**Proof:** See Appendix B.

However, with noncausal state-information the transmitter can cancel the effect of the state without hurting the nondeterministic receiver:

**Proposition 3:** The capacity region of the channel (64) when the states are known noncausally to the transmitter but not to the receivers, is the union over $\alpha \in [0, 1]$ of rate-pairs $(R_y, R_z)$ satisfying

\[
R_y \leq H_b(\alpha) \quad (67a) \\
R_z \leq 1 - H_b(\beta) \quad (67b)
\]

where

\[
\beta \triangleq \alpha(1 - p) + (1 - \alpha)p. \quad (67c)
\]

The capacity regions of the channel (64) when $\sigma = 0.5$ and $p = 0.2$ with noncausal and with causal state-information are depicted in Figure 1.

We present two different proofs for Proposition 3: the first is based on the achievability part of Theorem 1; the second is based on the fact that revealing the states to the deterministic receiver does not increase the capacity region.

**First proof of Proposition 3:** We let $U$ be a uniform binary random variable that is independent of $S$, and let $X$ be the outcome of feeding $U$ into a binary symmetric channel of crossover probability $\alpha$ (independently of $S$). Note that now the channel from $U$ to $Z$ is a binary symmetric channel with crossover probability $\beta$ as defined in (67c). Using Theorem 1 we obtain that the capacity region contains all rate-pairs $(R_y, R_z)$ satisfying

\[
R_y < H(Y|S) = 1 \\
R_z < I(U; Z) - I(U; S) = (1 - H_b(\beta)) - 0 \quad (70) \\
= 1 - H_b(\beta) \quad (71) \\
R_y + R_z < H(Y|S) + I(U; Z) - I(U; S, Y) = 1 + (1 - H_b(\beta)) - I(U; X) \quad (72)
\]

\[
W(Z = x|x) = 1 - W(Z = x \oplus 1|x) = 1 - p. \quad (76a)
\]

The capacity region of (76) can be found in [1, Example 15.6.5] and is the same as the region characterized by (67). □

**Second proof of Proposition 3:** By Theorem 1, the capacity region of the channel (64) with states known noncausally to the transmitter is unchanged if the states are also revealed to the deterministic receiver. When $S$ is revealed to the deterministic receiver, it can form $Y \oplus S$ and thus recover $x$. This reduces the channel to the one without states (76). Hence the capacity region of interest is the same as the capacity region of (76), which is given by the union over $\alpha \in [0, 1]$ of rate-pairs satisfying (67) [1, Example 15.6.5]. □

\[
Pr[Y = y, Z = z|X = x, S = s] = W(y, z|x, s) \quad (77a) \\
Pr[S = s] = P_S(s). \quad (77b)
\]
We let the state sequence $S$ be known noncausally to the transmitter and also known to the receiver which observes $Y$. The capacity region is defined in the same way as for the semideterministic broadcast channel. In particular, we consider only two private messages.

Applying the techniques of Section III, we obtain the following outer bound on the capacity region of the channel (77). (The bound is tight for semideterministic channels.)

**Proposition 4:** The capacity region of the channel (77), with the state sequence being revealed noncausally to the transmitter and also revealed to the receiver which observes $Y$, is contained in the convex closure of rate-pairs $(R_y, R_z)$ satisfying

$$R_y < I(X; Y|S)$$  
$$R_z < I(U; Z) - I(U; S)$$

$$R_y + R_z < I(X; Y|S) + I(U; Z) - I(U; S, Y)$$

for joint distributions of the form

$$P_{XY|ZSU}(x, y, z, s, u) = P_S(s) P_{X|S}(x, u|s) W(y, z|x, s).$$

(79)

**Proof:** To bound $R_y$, we note that (28) holds also for the general broadcast channel (77), and we continue (28) as follows:

$$nR_y \leq \sum_{i=1}^{n} I(M_y; Y_i|Y_i^{i-1}, S^n) + n\epsilon_n$$

$$\leq \sum_{i=1}^{n} I(M_y, X_i; Y_i|Y_i^{i-1}, S^n) + n\epsilon_n$$

$$= \sum_{i=1}^{n} H(Y_i|Y_i^{i-1}, S^n) - H(Y_i|X_i, S_i) + n\epsilon_n$$

$$\leq \sum_{i=1}^{n} I(X_i; Y_i|S_i) + n\epsilon_n.$$  

(80)

Here (82) follows because, given $(X_i, S_i)$, the channel output $Y_i$ is independent of $(M_y, Y_i^{i-1}, S^n_{i+1}, Y_{i+1}|S_i)$.

We bound $R_z$ exactly as (38) with $V_i, i \in \{1, \ldots, n\}$, defined as in (40).

To bound the sum-rate $R_y + R_z$, note that (43), (48), and (54) still hold, but (55) should be replaced by

$$\sum_{i=1}^{n} I(M_y, M_z, S_i^{i-1}, S_{i+1}^{n}, Y_i^{n+1}, Y_i|S_i) = \sum_{i=1}^{n} I(X_i; Y_i|S_i),$$

which is true because $(M_y, M_z, S^n)$ determines $X_i$, and because, without feedback, given $(X_i, S_i)$, the output $Y_i$ is independent of $(M_y, M_z, S_i^{i-1}, S_{i+1}^{n}, Y_{i+1}|S_i)$. These together yield

$$n(R_y + R_z) \leq \sum_{i=1}^{n} I(V_i, T_i; Z_i) - \sum_{i=1}^{n} I(V_i, T_i; S_i, Y_i)$$

$$+ \sum_{i=1}^{n} I(X_i; Y_i|S_i) + n\epsilon_n,$$  

(85)

where $T_i, i \in \{1, \ldots, n\}$, are defined in (57).

Summarizing (83), (38), and (85) we conclude that the desired capacity region is contained in the convex closure of rate-pairs $(R_y, R_z)$ satisfying

$$R_y < I(X; Y|S)$$

$$R_z < I(V; Z) - I(U; S)$$

$$R_y + R_z < I(X; Y|S) + I(V; Z) - I(U; S, Y)$$

where, given $(X, S)$, the outputs $(Y, Z)$ are drawn according to the channel law (77) independently of the auxiliary random variables $(V, T)$. Now to prove Proposition 4 it remains to find a single auxiliary random variable $U$ satisfying

$$I(V; Z) - I(V; S) \leq I(U; Z) - I(U; S)$$

(87a)

and

$$I(X; Y|S) + I(V, T; Z) - I(V, T; S, Y)$$

$$\leq I(X; Y|S) + I(U; Z) - I(U; S, Y)$$

(87b)

to replace both $V$ and $T$. Now note that (87) is equivalent to (60). Hence, according to our arguments in Section III, such a $U$ can always be found.

**APPENDIX A**

**PROOF OF PROPOSITION 1**

It suffices to show that, given any joint distribution $P_{XY|ZSU}$ of the form (3), there exists another distribution $\tilde{P}_{XY|ZSU}$ of the same form

$$\tilde{P}_{XY|ZSU}(x, y, z, s, u) = P_S(s) \tilde{P}_{X|S}(x, u|s) W(y, z|x, s)$$

(88)

satisfying

$$\left| \left\{ u : \tilde{P}_U(u) > 0 \right\} \right| \leq |\mathcal{X}| \cdot |\mathcal{S}| + 1,$$

(89)

where $\tilde{P}_U$ denotes the marginal of $\tilde{P}_{XY|ZSU}$ on $U$, and

$$H(Y|S) |_{\tilde{P}} = H(Y|S) |_{P}$$

$$I(U; Z) - I(U; S) |_{\tilde{P}} = I(U; Z) - I(U; S) |_{P}$$

$$H(Y|S) + I(U; Z) - I(U; S, Y) |_{\tilde{P}}$$

$$= H(Y|S) + I(U; Z) - I(U; S, Y) |_{P}.$$  

(90)

To this end, consider the following $|\mathcal{X}| \cdot |\mathcal{S}| + 1$ functions of $u$, all of which are determined by the conditional distribution $P_{XY|ZSU}$ and are independent of the marginal $P_U$:

$$h_{0}(u) \triangleq H(S|U = u) - H(Z|U = u)$$

$$h_{1}(u) \triangleq H(Y, S|U = u) - H(Z|U = u)$$

$$h_{x,s}(u) \triangleq P_{X|S}(x|s, u),$$

$$x \in \mathcal{X}, s \in \mathcal{S}, (x, s) \neq (1, 1).$$  

(91)

We now look for a $\tilde{P}_U$ (which will replace $P_U$) such that

$$\sum_{u \in U} \tilde{P}_U(u) h_{0}(u) = H(S|U) - H(Z|U) |_{\tilde{P}}$$

$$\sum_{u \in U} \tilde{P}_U(u) h_{1}(u) = H(Y, S|U) - H(Z|U) |_{\tilde{P}}$$

$$\sum_{u \in U} \tilde{P}_U(u) h_{x,s}(u) = P_{XS}(x, s),$$

$$x \in \mathcal{X}, s \in \mathcal{S}, (x, s) \neq (1, 1).$$  

(92)
By the Support Lemma [2, p.631], such a \( \tilde{P}_U \) can be found whose support-size is at most the total number of constraints, which equals \(|\mathcal{X}| \cdot |\mathcal{S}| + 1\). Choosing

\[
\tilde{P}_{XYZSU}(x, y, z, s, u) \triangleq \tilde{P}_U(u) P_{XYZS|U}(x, y, z, s|u) \tag{93}
\]

for all \((x, y, z, s, u)\) yields a joint distribution that satisfies (89). We next show that this choice also satisfies (88) and (90). First note that (92c) implies that \(\tilde{P}_{XYZSU}\) has the same marginal on \((X, S)\) as \(P_{XYZS}\). In particular,

\[
P_{S}(s) = P_{S}(s), \quad s \in \mathcal{S}. \tag{94}
\]

This combined with the fact that we used the conditional distribution \(P_{XYZS|U}\) to generate \(\tilde{P}_{XYZSU}\) shows that \(\tilde{P}_{XYZSU}\) is indeed of the form (88). Furthermore, these imply that

\[
\tilde{P}_{XYZS}(x, y, z, s) = P_{XYZS}(x, y, z, s) \tag{95}
\]

for all \((x, y, z, s)\). Hence we have

\[
H(Y|S)|_P = H(Y|S)|_p \tag{96a}
\]

\[
H(Z) - H(S)|_P = H(Z) - H(S)|_p \tag{96b}
\]

\[
H(Z) - H(Y, S)|_P = H(Z) - H(Y, S)|_p. \tag{96c}
\]

On the other hand, (92a) and (92b) imply

\[
H(S|U) - H(Z|U)|_P = H(S|U) - H(Z|U)|_p \tag{97a}
\]

\[
H(Y, S|U) - H(Z|U)|_P = H(Y, S|U) - H(Z|U)|_p. \tag{97b}
\]

Combining (96) and (97) yields (90) and concludes the proof.

**APPENDIX B**

**PROOF OF PROPOSITION 2**

To prove Proposition 2, we need the following simple outer bound on the capacity region of any broadcast channel with causal state-information:

**Lemma 2**: The capacity region of any state-dependent two-receiver broadcast channel as in (77) with causal state-information at the transmitter is contained in the convex closure of the union of the rate pairs satisfying

\[
R_y < I(T; Y) \tag{98a}
\]

\[
R_z < I(T; Z) \tag{98b}
\]

where the union is over all joint distributions of the form

\[
P_{XYZST}(x, y, z, s, t) = P_{S}(s) P_{T}(t) 1\{x = g(t, s)\} W(y, z|x, s). \tag{99}
\]

**Proof**: We bound \(R_y\) as for single-user channels with causal state-information [2], [16] as follows:

\[
nR_y \leq I(M_i; Y^n) + n\epsilon_n \tag{100}
\]

\[
\leq I(M_i, M_i^n; Y^n) + n\epsilon_n \tag{101}
\]

\[
= \sum_{i=1}^{n} I(M_i, M_i^n; Y_i^{i-1}) + n\epsilon_n \tag{102}
\]

\[
\leq \sum_{i=1}^{n} I(M_i, M_i^n, S_i^{i-1}; Y_i^{i-1}) + n\epsilon_n \tag{103}
\]

\[
= \sum_{i=1}^{n} I(M_i, M_i^n, S_i^{i-1}; X_i^{i-1}, Y_i^{i-1}) + n\epsilon_n \tag{104}
\]

\[
= \sum_{i=1}^{n} I(M_i, M_i^n, S_i^{i-1}; Y_i) + n\epsilon_n \tag{105}
\]

\[
= \sum_{i=1}^{n} I(M_i, M_i^n, S_i^{i-1}; Y_i) + n\epsilon_n. \tag{106}
\]

Here, (105) and (107) follow because \(X_i^{i-1}\) is a function of \((M_i, M_i^n, S_i^{i-1})\); and (106) because, given \((M_i, M_i^n, S_i^{i-1}, X_i^{i-1})\), the output \(Y_i\) is independent of \(Y_i^{i-1}\).

In the same way we can obtain

\[
nR_z \leq \sum_{i=1}^{n} I(M_i, M_i^n, S_i^{i-1}; Z_i). \tag{107}
\]

We define

\[
T_i \triangleq (M_i, M_i^n, S_i^{i-1}), \quad i \in \{1, \ldots, n\} \tag{109}
\]

which clearly satisfy the conditions

\[
T_i \perp S_i, \quad T_i \rightarrow (X_i, S_i) \rightarrow (Y_i, Z_i), \quad i \in \{1, \ldots, n\}. \tag{110}
\]

We now have

\[
nR_y \leq \sum_{i=1}^{n} I(T_i; Y_i) + n\epsilon_n \tag{111a}
\]

\[
nR_z \leq \sum_{i=1}^{n} I(T_i; Z_i) + n\epsilon_n, \tag{111b}
\]

which imply that the capacity region of interest is contained in the convex closure of (98) for distributions on \((X, Y, Z, S, T)\) satisfying

\[
T \perp S, \quad T \rightarrow (X, S) \rightarrow (Y, Z). \tag{112}
\]

It now only remains to show that, to exhaust this region, it suffices to consider joint distributions in which \(X\) is a function of \((T, S)\). This is indeed the case because, given \(P_{TS}(t, s)\) and the channel law, both terms on the RHS of (98) are convex in \(P_X|TS\).

We next proceed to prove Proposition 2. We begin with the achievability part, which is straightforward. If the transmitter only communicates to the receiver which observes \(Y\), then it can cancel the interference of \(S\) by flipping the input symbol whenever \(S = 1\). In this way the rate-pair

\[
(R_y, R_z) = (1, 0) \tag{113}
\]
can be achieved. On the other hand, if the transmitter only communicates to the receiver which observes \( Z \), then it can ignore \( S \) and achieve the rate-pair
\[
(R_y, R_z) = (0, 1 - H_b(p)).
\] (114)

Time-sharing between (113) and (114) achieves the claimed capacity region.

To prove the converse part, we use Lemma 2. Note that the auxiliary random variable \( T \) in Lemma 2 can be restricted to take value in all “input strategies” [16]. Namely, its alphabet is the set of all mappings from \( S \) to \( X \). There are four such mappings:

\[
\begin{align*}
T = 0: & \text{ maps 0 to 0 and 1 to 0 } \quad (115a) \\
T = 1: & \text{ maps 0 to 1 and 1 to 1 } \quad (115b) \\
T = 2: & \text{ maps 0 to 0 and 1 to 1 } \quad (115c) \\
T = 3: & \text{ maps 0 to 1 and 1 to 0. } \quad (115d)
\end{align*}
\]

Here, \( T = 0 \) or \( T = 1 \) means sending a fixed \( x \) independently of \( S \), and \( T = 2 \) or \( T = 3 \) means flipping \( x \) whenever \( S = 1 \). Using the “fixed” strategies \( T = 0 \) or \( T = 1 \) one can transmit information to the receiver which observes \( Z \) but not to the receiver which observes \( Y \):

\[
\begin{align*}
H(Y|T = 0) &= H(Y|T = 1) = 1 \quad (116) \\
H(Z|T = 0) &= H(Z|T = 1) = 1 - H_b(p); \quad (117)
\end{align*}
\]

while using the “flipped” strategies \( T = 2 \) or \( T = 3 \) one can transmit information to the receiver which observes \( Y \) but not to the receiver which observes \( Z \):

\[
\begin{align*}
H(Y|T = 2) &= H(Y|T = 3) = 0 \quad (118) \\
H(Z|T = 2) &= H(Z|T = 3) = 1. \quad (119)
\end{align*}
\]

We now have

\[
R_y \leq I(T; Y)
\]
\[
= H(Y) - H(Y|T)
\]
\[
= H(Y) - P_T(0)H(Y|T = 0) \\
- P_T(1)H(Y|T = 1) \\
\leq 1 - \Pr[T \in \{0, 1\}] \cdot 1 \quad (122)
\]
\[
= \Pr[T \in \{2, 3\}] \quad (123)
\]

\[
R_z \leq I(T; Z)
\]
\[
= H(Z) - H(Z|T)
\]
\[
= H(Z) - P_T(0)H(Z|T = 0) - P_T(1)H(Z|T = 1) \\
- P_T(2)H(Z|T = 2) - P_T(3)H(Z|T = 3) \\
\leq 1 - \Pr[T \in \{0, 1\}] \cdot (1 - H_b(p)) \\
- \Pr[T \in \{2, 3\}] \cdot 1 \\
= (1 - \Pr[T \in \{2, 3\}]) \cdot (1 - H_b(p)). \quad (129)
\]

Denoting
\[
\lambda \triangleq \Pr[T \in \{2, 3\}] \quad (130)
\]
we see that \((R_y, R_z)\) indeed must satisfy (98). This ends our proof of Proposition 2.

ACKNOWLEDGMENTS

The authors thank the anonymous reviewers of both the conference and the journal versions of this paper, as well as the Associate Editor Yossef Steinberg for their helpful comments.

REFERENCES