Capacity with Causal and Non-Causal Side Information
- A Unified View*

Syed Ali Jafar

Center for Pervasive Communications and Computing

Department of Electrical Engineering and Computer Science

University of California Irvine, CA 92697 USA

E-mail : syed@ece.uci.edu

Abstract

We identify the common underlying form of the capacity expression that is applicable to both cases where causal or non-causal side information is made available to the transmitter. Using this common form we find that for the single user channel, the multiple access channel, the degraded broadcast channel, and the degraded relay channel, the sum capacity with causal and non-causal side information are identical when all the transmitter side information is also made available to all the receivers. A genie-aided outerbound is developed that states that when a genie provides n bits of side information to a receiver the resulting capacity improvement can not be more than n bits. Combining these two results we are able to bound the relative capacity advantage of non-causal side information over causal side information for both single user as well as various multiple user communication scenarios. Applications of these capacity bounds are demonstrated through examples of random access channels. Interestingly, the capacity results indicate that the excessive MAC layer overheads common in present wireless systems may be avoided through coding across multiple access blocks. It is also shown that even one bit of side information at the transmitter can result in unbounded capacity improvement.

1 Introduction

Characterizing the capacity of reliable communication between a transmitter and a receiver in the presence of side information is a problem as old as information theory itself. Starting with Shannon's characterization of the capacity with causal side information [1] and Kusnetsov and Tsybakov's seminal work [2] on coding for computer memories with defective cells that lead to a general characterization of capacity with non-causal side information by Gel'fand and Pinsker [3], the theory of communication with side information has evolved as a dichotomy between the two cases of non-causal side information and causal side information. Capacity results with non-causal side information were generalized to the case of rate limited side information by Heegard and El Gamal [4] and more recently by Rosenzweig et. al. [5] for single user communication and by Cemal and Steinberg [6] for the multiple access channel. Study of non-causal side information has lead to interesting duality relationships such as that between source coding and channel coding with side information [7] as well as the duality between the Gaussian multiple input multiple output (MIMO) broadcast and multiple access channels [8–10]. In addition to data storage [2, 4] and data-hiding/watermarking/steganography [11] these results have found applications on Gaussian channels with additive Gaussian interference [3, 12–14] recently leading to the determination of first the sum capacity [9, 10, 15, 16] and then the entire capacity region for the MIMO broadcast channel [17–19].

The case of causal side information has been studied separately by researchers [1, 20–23]. The original capacity result due to Shannon [1] requires coding over an extended alphabet of mappings from the channel state to the input alphabet. Caire and Shamai [21] showed that when the side information at the transmitter is a deterministic function of the side information available to the receiver, capacity achieving codes can be constructed directly on the input alphabet. The capacity of the time varying multiple access channel with causal side information was explored by Das and Narayan [22] and more recently by Kim and Sigurjonsson [23].

With the availability of all the above mentioned capacity results on causal and non-causal side information, there is a need to develop a unified view of communication with side information. A unified view would allow us to relate, combine and extend the existing results to new applications. In this paper, we approach this objective through the following questions:

1. *Common Framework:* What is the fundamental connection between the capacity characterizations with causal and non-causal side information at the transmitter?

2. *Value of a bit of side information:* What is the maximum possible capacity improvement with n bits of any kind (causal, non-causal, memoryless or correlated) of side information provided by a genie to the transmitter/receiver?

3. *Non-causal versus Causal:* What is the relative capacity advantage of non-causal side information over causal side information?

4. *Impact of Correlation:* How does correlation, either temporal or between the transmitter and receiver side information, affect the capacity?

5. *Multiple users:* Finally, how do the causal and non-causal capacity characterizations extend to multiple user channels, e.g. the multiple access, broadcast and relay channels?

In this work we focus on both causal and non-causal side information, the relationship between them, and their extensions to multiuser communications. Our interest is in general capacity expressions with finite states. Specialized expressions for AWGN or fading channels may be obtained as special cases from these general expressions, subject to input distribution optimizations.

2 Background and Channel Model

The channel is a discrete memoryless channel (DMC) with message W, input X_i, output Y_i, state S_i, transmitter side information $S_{T,i}$ and receiver side information $S_{R,i}$ so that $P(S^n, S_T^n, S_R^n) = \Pi_{i=1}^n P(S_i, S_{T,i}, S_{R,i})$ and $P(Y^n|X^n, S^n) = \Pi_{i=1}^n P(Y_i|X_i, S_i)$. The only difference between the causal and non-causal side information cases is that with causal side information the input to the channel at time i can only depend on the present and past states but not the future states, $X_i(W, S_T^i)$, whereas with non-causal side information all inputs can depend on the entire state sequence $X_i(W, S_T^n)$. Notice that for the receiver it does not matter if the side information is made available causally or non-causally. This is because the receiver can wait till the end of transmission to decode the message. Figure 1 illustrates the two scenarios. Probability of error, achievable rates and the capacity of this channel are defined in the standard sense [21, 24].
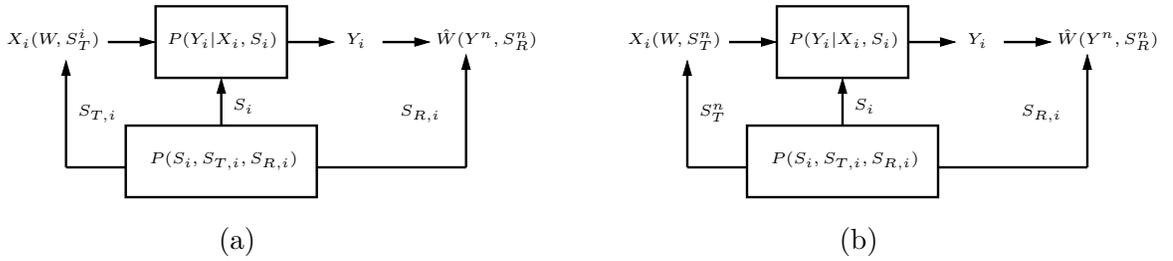

Figure 1: Channel Model with (a) Causal, and (b) Non-causal side information.

For non-causal side information at the transmitter, the single user capacity is known to be [2–4]:

$$C^{\text{non-causal}} = \max_{\mathcal{P}_{\text{non-causal}}} I(U;Y) - I(U;S_T) \tag{1}$$

3

where $\mathcal{P}_{\text{non-causal}} = \{P(U, X|S_T) = P(U|S_T)P(X|U, S_T)\}$. Comparing this to the case where no side information is available ($S_T = \phi$),

$$C = \max_{P(U,X)} I(U;Y) = \max_{P(X)} I(X;Y), \tag{2}$$

note that the availability of side information at the transmitter is helpful in that the transmitter can match its input to the channel information by picking the input alphabet U, X conditioned on S_T, as opposed to (2) where the input can not be matched to the channel state. However, the benefit of matching the input to the channel state comes with the cost of the subtractive term in (1), i.e., $I(U; S_T)$ which can be interpreted as the overhead required to communicate to the receiver, the adaptation to the channel state at the transmitter. For the case where the side information is available at the transmitter only causally, the capacity expression has been found by Shannon as [1]

$$C^{\text{causal}} = \max_{P(t)} I(T;Y) \tag{3}$$

where T is an extended alphabet of mappings from the channel state to the input alphabet.

The capacity expressions (1), (2), (3) explicitly account for side information at the transmitter. Side information at the receiver, S_R, is easily incorporated into the same expressions by replacing Y with (Y, S_R) in the corresponding expressions.

We start by finding a common form for the causal and non-causal cases.

3   Relating Capacity with Causal and Non-causal Side Information

Comparing the non-causal case (1) with the causal case (3) the two capacity expressions are in different forms so that their relationship is not obvious. We start by making the relationship clearly apparent by representing the causal case in a different form, comparable to (1). The following result has also been observed and derived independently in parallel work by [23] and [6].

$$
\begin{aligned}
C^{\text{non-causal}} &= \max_{\mathcal{P}_{\text{non-causal}}} I(U;Y,S_R) - I(U;S_T), \\
C^{\text{causal}} &= \max_{\mathcal{P}_{\text{causal}}} I(U;Y,S_R) - I(U;S_T),
\end{aligned}
$$

with

$$
\begin{aligned}
\mathcal{P}_{\text{non-causal}} &= \{P(U,X|S_T) = P(U|S_T)P(X|U,S_T)\} \\
\mathcal{P}_{\text{causal}} &= \{P(U,X|S_T) = P(U)P(X|U,S_T)\}
\end{aligned}
$$

In the non-causal case, the choice of U can be made conditional on the channel state S_T. In the causal case U is picked independent of S_T. This makes the subtractive term equal to zero for the causal case. In both cases, it suffices for the optimal input symbol X to be just a function of U, S_T, i.e. $P(X|U, S_T)$ is either 0 or 1.

Sketch of Proof: [Converse] Achievability of the capacity expression with causal side information is straightforward. Interestingly, the converse allows a very simple proof, as follows. Starting with Fano's inequality,

$$
\begin{align}
nR &\leq I(W; Y^n, S_R^n) + n\epsilon_n \tag{4}\\
&\leq \sum_{i=1}^{n} I(W, S_T^{i-1}; Y_i, S_{R,i}|Y^{i-1}, S_R^{i-1}) \tag{5}\\
&= \sum_{i=1}^{n} H(Y_i, S_{R,i}|Y^{i-1}, S_R^{i-1}) - H(Y_i, S_{R,i}|Y^{i-1}, S_R^{i-1}, W, S_T^{i-1}) \tag{6}\\
&\leq \sum_{i=1}^{n} H(Y_i, S_{R,i}) - H(Y_i, S_{R,i}|S_T^{i-1}, W) \tag{7}\\
&= \sum_{i=1}^{n} I(U_i; Y_i, S_{R,i}) \tag{8}
\end{align}
$$

where (7) follows because the current output is independent of the past outputs, conditioned on all the past inputs. $U_i = (W, S_T^{i-1})$ is the auxiliary random variable, independent of current channel state $S_{T,i}$. ■

The capacity expression (1) has been shown [7,25] to be the common form of single user capacity for all four cases of non-causal side information as well as the corresponding cases for rate-distortion. In other words, for the capacity problem, whether the non-causal side information is available at the transmitter, the receiver, both, or neither, the capacity expression has the common form $I(U; Y, S_R) - I(U; S_T)$. The only difference is in the constraints on the distribution of the auxiliary random variable U, the input alphabet X and the state variable S. Thus, combining the results of [7,25] with the common expression obtained above we find that the expression $I(U; Y, S_R) - I(U; S_T)$ is indeed the common expression for not only all cases of non-causal side information but also for causal side information as well.

$$
\begin{align}
C^{\text{non-causal}}(S_T, S_R) &= \max_{P(U,X|S_T)} & I(U; Y, S_R) - I(U; S_T) & \\
C^{\text{non-causal}}(\phi, S_R) &= \max_{U=\phi, P(X)} & I(U; Y, S_R) - I(U; S_T) &= \max_{P(X)} I(X; Y, S_R) \\
C^{\text{non-causal}}(S_T, \phi) &= \max_{P(U,X|S_T)} & I(U; Y, S_R) - I(U; S_T) &= \max_{P(U,X|S_T)} I(U; Y) - I(U; S_T) \\
C^{\text{non-causal}}(\phi, \phi) &= \max_{U=\phi, P(X)} & I(U; Y, S_R) - I(U; S_T) &= \max_{P(X)} I(X; Y)
\end{align}
$$

$$C^{\text{causal}}(S_T, S_R) = \max_{P(U)P(X|U,S_T)} \quad I(U;Y,S_R) - I(U;S_T) \quad = \max_{P(U)P(X|U,S_T)} I(U;Y,S_R)$$
$$C^{\text{causal}}(\phi, S_R) = \max_{U=\phi, P(X)} \quad I(U;Y,S_R) - I(U;S_T) \quad = \max_{P(X)} I(X;Y,S_R)$$
$$C^{\text{causal}}(S_T, \phi) = \max_{P(U)P(X|U,S_T)} \quad I(U;Y,S_R) - I(U;S_T) \quad = \max_{P(U)P(X|U,S_T)} I(U;Y)$$
$$C^{\text{causal}}(\phi, \phi) = \max_{U=\phi, P(X)} \quad I(U;Y,S_R) - I(U;S_T) \quad = \max_{P(X)} I(X;Y)$$

Investigating the relationship between causal and non-causal information further, we prove that the two capacities are the same if the transmitter side information is also available to the receiver.

Theorem 1 (Relationship between causal and non-causal side information capacity) *If the side-information at the transmitter is a deterministic function of the side-information at the receiver, i.e., if $S_T = f(S_R)$, then capacity with causal side information is equal to the capacity with non-causal side information.*

Capacity achieving codes, in both cases, can be constructed directly on the input alphabet and the auxiliary random variable U is not required.

Proof:

$$
\begin{aligned}
C^{\text{non-causal}} &= \max_{P(U,X|S_T)=P(U|S_T)P(X|U,S_T)} I(U;Y,S_R) - I(U;S_T) \\
&= \max_{P(U,X|S_T)=P(U|S_T)P(X|U,S_T)} I(U;S_R) + I(U;Y|S_R) - I(U;S_T) \\
&= \max_{P(U,X|S_T)=P(U|S_T)P(X|U,S_T)} I(U;S_R,S_T) + I(U;Y|S_R) - I(U;S_T) \\
&= \max_{P(U,X|S_T)=P(U|S_T)P(X|U,S_T)} I(U;S_R|S_T) + I(U;Y|S_R) \\
&= \max_{P(U,X|S_T)=P(U|S_T)P(X|U,S_T)} I(U;Y|S_R) \\
&= \max_{P(X|S_T)} I(X;Y|S_R) \\
&= C^{\text{causal}}
\end{aligned}
$$

where the last equality follows from the results of [21] for causal side information. It is shown in [21] that with causal side information at the transmitter, if $S_T = f(S_R)$, i.e., if the side-information at the transmitter is a deterministic function of the side-information at the receiver then the auxiliary random variable U is not needed and coding can be performed directly on the input alphabet X.

Thus, if the side information at the transmitter is a deterministic function of the side information at the receiver, then capacity with non-causal side information is the same as the capacity with causal side-information. ∎

Next we investigate the value of side information through capacity bounds.

4 Genie bits and the Value of Side Information

In this section we answer the question: what is the maximum capacity gain that can result from a fixed number of bits of side information? For example, suppose a genie provides a fixed number of bits of side information G to the transmitter or the receiver per channel use. There are no constraints on the genie provided side information, i.e., it could be causal, non-causal, temporally correlated, or i.i.d. A fundamental question is whether the capacity increase, $C_G - C$, due to a fixed number of genie bits is bounded and if so, then what is the maximum capacity benefit. The following results show that transmitter and receiver side information are fundamentally different in their potential capacity advantages.

4.1 Receiver Side Information

Theorem 2 *The maximum possible capacity improvement due to the availability of receiver side information is bounded by the amount of the side information itself.*

$$C_G - C \leq H(\mathcal{G}) = \lim_{N \to \infty} \frac{1}{N} H(G_1, G_2, \cdots, G_N).$$

Proof:

$$
\begin{aligned}
C_G &= \sup_{p()} \lim_{N \to \infty} \frac{1}{N} I(W; Y^N, G^N) \\
&= \sup_{p()} \lim_{N \to \infty} \frac{1}{N} \left(I(W; Y^N) + I(W; G^N | Y^N) \right) \\
&= C + \Delta C
\end{aligned}
$$

where the sup is over the multi-letter distributions of allowed input strategies. C_G is the capacity with the side information provided by the genie, C is the capacity without the side information and ΔC, the capacity improvement, is bounded by the entropy rate $H(\mathcal{G})$. In other words, if the genie provides one bit of side information per channel use, the capacity benefit $\Delta C = C_G - C$ can not be more than 1 bit, regardless of the kind of side information. ∎

4.2 Transmitter Side Information

Theorem 3 *The maximum possible capacity improvement from a fixed number of genie bits (per channel use) of side information is unbounded, even if the side information is causal.*

Proof: The proof is in the form of the following example. Consider the channel with input alphabet $X \in \mathcal{X} = \{0, 1, 2, \cdots, 2N - 1\}$, output alphabet $Y \in \mathcal{Y} = \mathcal{X} \cup \{\phi\}$ where ϕ is the erasure

7

symbol, and i.i.d. state sequence drawn from the alphabet $S \in \mathcal{S} = \{0,1\}$. The input output relationship is given by,

$$Y = \begin{cases} X & \text{when } X + S \text{ is even,} \\ \phi & \text{when } X + S \text{ is odd.} \end{cases}$$

In simple words, when the channel state is $S = 0$, the channel conveys even inputs noise-free and erases odd inputs. Similarly, when the channel state is $S = 1$, the channel conveys odd inputs noise-free and erases even inputs. The channel state S is unknown to the receiver.

Suppose the genie provides one bit of side information in the form of $G = S$ to the transmitter every channel use. With perfect knowledge of S at the transmitter, the capacity of this channel is clearly $C_G = \log(N + 1)$ where $N + 1$ is the number of distinct outputs that can be affected by the transmitter. On the other hand, with no state information at the transmitter the optimal input distribution is uniform on $\mathcal{X}$, the corresponding output distribution is

$$Y = \begin{cases} \phi & \text{with probability } \frac{1}{2} \\ i & \text{with probability } \frac{1}{4N}, \quad i \in \mathcal{X} \end{cases}$$

and the capacity is

$$C = I(X;Y) = H(Y) - H(Y|X) = \frac{1}{2}\log(N) + \frac{1}{2}.$$

The capacity benefit of one genie bit therefore is

$$C_G - C = \log(N + 1) - \frac{1}{2}\log(N) - \frac{1}{2} \geq \frac{1}{2}\log(N) - \frac{1}{2}$$

which is unbounded, i.e. goes to infinity as N goes to infinity. Notice that C_G is obtained with only causal side information at the transmitter. Thus, the example shows that the capacity benefit of one bit of side information at the transmitter can be unbounded even when the side information is causal. ∎

The contrasting potential capacity benefits of side information are summarized in Fig. 2.

While we have shown that, in theory, unbounded capacity improvement can result from a finite amount of causal side information at the transmitter, it is not clear how such side information and the associated capacity benefits can be obtained in practice. For practical systems, side information at the transmitter is often obtained through a feedback channel from the receiver. However, it is well known that for a DMC, even if all the past channel outputs obtained at the receiver are made available to the transmitter through a noise and delay free feedback channel, the capacity is unaffected. Thus, the unbounded potential capacity benefits of causal side information are in sharp contrast with the zero capacity benefit of causal feedback for a DMC. The key to reconciling these results lies in the definition of causal in both cases. While causal side information allows the

(a) Receiver Side Info.: One genie bit cannot improve capacity by more than one bit.

(b) Transmitter Side Info.: One genie bit can improve capacity by an unbounded amount.
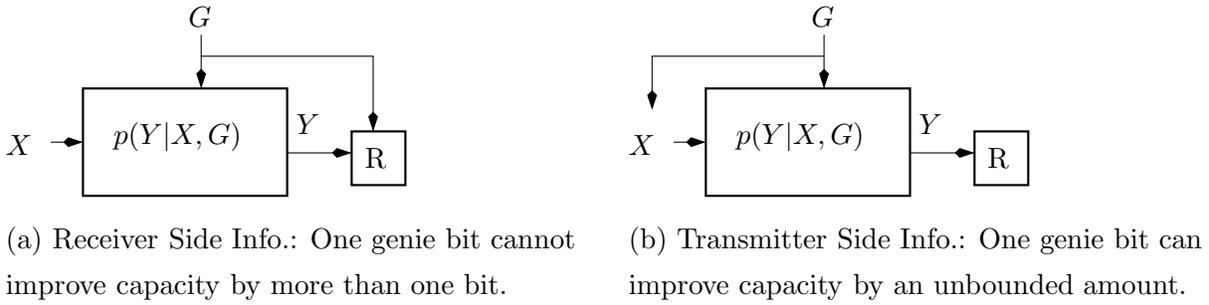
Figure 2: Value of Side Information at the Transmitter and Receiver

transmitter knowledge of the current channel state, causal feedback allows information only about the past channel outputs. For a memoryless channel past outputs do not provide any information about the current channel state. In the example considered above, causal feedback of the channel outputs will provide the transmitter precise knowledge of all the past channel states, but no information about the present channel state. Thus, the timing of the side information at the transmitter can make the difference between unbounded capacity improvements and no improvement at all.

5 Advantage of Non-Causal Side Information over Causal Side Information

The preceding section shows that the present state information at the transmitter can be invaluable compared to the information of all the past states. The natural question then is, what is the advantage of knowing the *future* channel states over knowing just the present and past states? In other words, what is the capacity advantage of non-causal side information over causal side information? It turns out, the answer to this question is already implicit in the results of Theorems 1 and 2.

Theorem 4 *The capacity benefit of non-causal side information over causal side information is bounded as follows:*

$$C^{non\text{-}causal}(S_T, S_R) - C^{causal}(S_T, S_R) \leq H(S_T|S_R)$$

Proof: From the preceding sections we have the following two results.

- If the transmitter side information is also available to the receiver, capacity with causal and non-causal side information is identical. (Theorem 1)

- If a genie provides a bit of side information to the receiver it can not improve capacity by more than a bit. (Theorem 2)

9

Combining these two results, suppose the transmitter side information is made available to the receiver by a genie. This requires $H(S_T|S_R)$ genie bits and therefore cannot improve capacity by more than $H(S_T|S_R)$ bits. Using the results stated above, we have:

$$
\begin{aligned}
C^{\text{non-causal}}(S_T, (S_T, S_R)) - H(S_T|S_R) &\leq C^{\text{non-causal}}(S_T, S_R) \leq C^{\text{non-causal}}(S_T, (S_T, S_R)) \\
C^{\text{causal}}(S_T, (S_T, S_R)) - H(S_T|S_R) &\leq C^{\text{causal}}(S_T, S_R) \leq C^{\text{causal}}(S_T, (S_T, S_R)) \\
C^{\text{non-causal}}(S_T, (S_T, S_R)) &= C^{\text{causal}}(S_T, (S_T, S_R)) \\
C^{\text{non-causal}}(S_T, S_R) - C^{\text{causal}}(S_T, S_R) &\leq H(S_T|S_R)
\end{aligned}
$$

In simple words, the advantage of non-causal side information over causal side information is bounded by the number of genie bits required to make the transmitter side information available to the receiver as well.

5.1   Example: Random Access

Consider a single user DMC characterized by $P(Y|X)$ and with a capacity $C_0 = \max_{P(X)} I(X; Y)$ when the input is directly controlled by a transmitter. Now, suppose instead that the transmitter is only able to access the channel (control the input) in a random manner as:

$$
X = X_T S + X_r(1 - S),
$$

where $S \in \{0, 1\}$ is a switch state that determines when the transmitter can access the channel with the symbol X_T, and X_r is a randomly generated input.
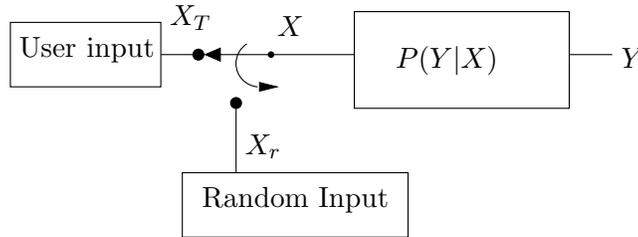

Figure 3: Random Access Channel

Suppose the state S is known to the transmitter and *not known* to the receiver. Such a channel is relevant to cognitive communication scenarios [26] and is also similar to the "memory with stuck-at defects" problem considered in [4]. Clearly, if the switch state is provided to the receiver by a genie the resulting capacity is $C(S, S) = \text{Prob}(S = 1)C_0 = \overline{S}C_0$. Since the extra information provided by the genie is only one bit we have

$$
\overline{S}C_0 \geq C^{\text{non-causal}}(S, \phi) \geq \overline{S}C_0 - 1
$$

Interestingly, the capacity with causal side information in this case is the same as the capacity with no side information. This is easily seen by rewriting $X_T = f(U, S)$ as follows:

$$X_T = \begin{cases} f_0(U), S = 0, \\ f_1(U), S = 1 \end{cases} \equiv f_1(U), S = 0, 1.$$

In other words, the choice of input symbol does not matter when the switch is open $(S = 0)$. Thus, we have

$$C^{\text{causal}}(S, \phi) = C(\phi, \phi) \geq C(S, S) - 1.$$

The effect of memory in side information is also revealed by this example. Suppose the switch changes state in a block static model, i.e. it retains its state for N symbols and then changes to an i.i.d. realization. In this case, the genie only needs to provide one bit to the receiver every N channel uses and the bounds are tighter.

$$\overline{S}C_0 = C(S, S) \geq C^{\text{non-causal}}(S, \phi) \geq C^{\text{causal}}(S, \phi) = C(\phi, \phi) \geq \overline{S}C_0 - \frac{1}{N}$$

6 Multiple Access Channel with Independent Side Information

Achievable regions with causal side information are straightforward to obtain because the codewords can always be constructed on mappings from the side information to the channel input alphabet. In the multiple access channel, the two transmitters have (possibly correlated) side information S_{T1}, S_{T2} respectively, and the common receiver has side information S_R. The characterization of the achievable region with side information is the same as without side information, with the codes defined on auxiliary random variables that are independent of the side information, and the actual channel input symbols chosen as a function of the auxiliary random variable and the instantaneous side information. Thus the following achievable region is obtained.

$$\begin{aligned} R_1 &\leq I(U_1; Y, S_R | U_2) = I(U_1; Y | U_2, S_R) \\ R_2 &\leq I(U_2; Y, S_R | U_1) = I(U_2; Y | U_1, S_R) \\ R_1 + R_2 &\leq I(U_1, U_2; Y, S_R) = I(U_1, U_2; Y | S_R) \end{aligned}$$

where U_1, U_2 are mutually independent as well as independent of the side information and the channel inputs are given by $X_1 = f_1(U_1, S_{T1}), X_2 = f_2(U_2, S_{T2})$.

While a full converse is not known, upperbounds for the MAC with causal side information are obtained in [23] in terms of the capacity achieved with transmitter cooperation. In this work, we focus on the multiple access channel with independent side information at the two transmitters. We prove that the sum capacity of the multiple access channel with independent side information is given by the corresponding constraint in the achievable region provided above.

Theorem 5 *The sum capacity of the discrete memoryless multiple access channel with causal side information S_{T1}, S_{T2} (S_{T1}, S_{T2} independent) and S_R available to transmitter 1, transmitter 2 and the receiver, respectively, is given by*

$$R_1 + R_2 = \max I(U_1, U_2; Y, S_R) \tag{9}$$

with $P(U_1, U_2, X_1, X_2, S_{T1}, S_{T2}) = P(U_1)P(U_2)P(X_1|U_1, S_{T1})P(X_2|U_2, S_{T2})P(S_{T1}, S_{T2})$.

Proof: The converse is proved as follows:

$$
\begin{aligned}
n(R_1 + R_2) &\leq I(W_1, W_2; Y^n, S_R^n) + n\epsilon \\
&\leq \sum_{i=1}^{n} I(W_1, W_2, S_{T1}^{i-1}, S_{T2}^{i-1}; Y_i, S_{R,i} | Y^{i-1}, S_R^{i-1}) + n\epsilon \\
&\leq \sum_{i=1}^{n} H(Y_i, S_{R,i}) - H(Y_i, S_{R,i} | W_1, S_{T1}^{i-1}, W_2, S_{T2}^{i-1}, Y^{i-1}, S_R^{i-1}) + n\epsilon \\
&= \sum_{i=1}^{n} I(U_{1,i}, U_{2,i}; Y_i, S_{R,i}) + n\epsilon
\end{aligned}
$$

where $U_{1,i} = W_1, S_{T1}^{i-1}$ and $U_{2,i} = W_2, S_{T2}^{i-1}$ are independent of $S_{T1,i}, S_{T2,i}$. Independence of messages and transmitter side information implies independence of the auxiliary random variables U_1, U_2 as well. ∎

For non-causal side information, an achievable region is readily obtained when the side information at the two transmitters is independent.

$$
\begin{aligned}
C \supseteq \{(R_1, R_2) \quad : \quad & R_1 \leq I(U_1; Y, S_R | U_2) - I(U_1; S_{T1}) \\
& R_2 \leq I(U_2; Y, S_R | U_1) - I(U_2; S_{T2}) \\
& R_1 + R_2 \leq I(U_1, U_2; Y, S_R) - I(U_1; S_{T1}) - I(U_2; S_{T2})\}
\end{aligned}
$$

for all $P(U_1, X_1, U_2, X_2 | S_{T1}, S_{T2}) = P(U_1, X_1 | S_{T1})P(U_2, X_2 | S_{T2})$. However, to the best of the author's knowledge a converse has not been shown for independent side information[1]. Correlation of the side information makes even the achievable region non-trivial, as the possibility of Slepian Wolf coding of correlated side information at the two transmitters can be exploited.

For our purpose however, we show that with independent side information at the transmitters, if all the transmitter side information is also made available to the receiver, i.e. $(S_{T1}, S_{T2}) = f(S_R)$, then the MAC capacity region with causal side information is identical to the capacity region with non-causal side information.

---

[1]The problem with extending the single user approach appears to be that including Y^{i-1} into the auxiliary random variables makes them correlated.

Theorem 6 *For the discrete memoryless multiple access channel with side information S_{T1}, S_{T2} (independent) and S_R available to transmitter 1, transmitter 2 and the receiver, respectively, if $(S_{T1}, S_{T2}) = f(S_R)$ then the capacity region for both causal or non-causal side information is given by the convex hull of all rates pairs (R_1, R_2) satisfying the following inequalities:*

$$
\begin{aligned}
R_1 &\leq I(X_1; Y | S_R, X_2) \\
R_2 &\leq I(X_2; Y | S_R, X_1) \\
R_1 + R_2 &\leq I(X_1, X_2; Y | S_R) \}
\end{aligned}
$$

for all $P(X_1 | S_{T1}), P(X_2 | S_{T2})$.

Proof: [Achievability] Achievability is easily established as follows. Starting with the achievable region for the causal side information case, we have:

$$
\begin{aligned}
R_1 &\leq I(U_1; Y | U_2, S_R) & (10) \\
&= H(Y | U_2, S_R) - H(Y | U_2, U_1, S_R) & (11) \\
&= H(Y | U_2, S_R, X_2) - H(Y | U_2, U_1, S_R, X_2, X_1) & (12) \\
&= H(Y | X_2, S_R) - H(Y | X_1, X_2, S_R) & (13) \\
&= I(X_1; Y | S_R, X_2). & (14)
\end{aligned}
$$

Equation (12) follows from the fact that X_1 (resp. X_2) is a function of U_1, S_{T1} (resp. U_2, S_{T2}) and S_{T1} (resp. S_{T2}) is a function of S_R. Thus, X_1 (resp. X_2) is a function of U_1, S_R (resp. U_2, S_R). The corresponding inequalities for R_2 and the sum rate $R_1 + R_2$ are similarly obtained. Clearly, what is achievable with causal side information is also achievable with non-causal side information. ∎

Proof: [Converse] For the converse, we start with the individual rate constraints. For both causal and non-causal side information we have:

$$
\begin{aligned}
nR_1 &\leq I(W_1; Y^n, S_R^n | W_2) + n\epsilon & (15) \\
&= \sum_{i=1}^{n} I(W_1; Y_i | W_2, Y^{i-1}, S_R^n) + n\epsilon & (16) \\
&= \sum_{i=1}^{n} H(Y_i | S_R^n, W_2, Y^{i-1}, X_{2,i}) - H(Y_i | W_2, W_1, S_R^n, Y^{i-1}, X_{1,i}, X_{2,i}) & (17) \\
&\leq \sum_{i=1}^{n} H(Y_i | S_{R,i}, X_{2,i}) - H(Y_i | W_2, W_1, S_R^n, Y^{i-1}, X_{1,i}, X_{2,i}) & (18) \\
&= \sum_{i=1}^{n} H(Y_i | S_{R,i}, X_{2,i}) - H(Y_i | X_{1,i}, X_{2,i}, S_{R,i}) + n\epsilon & (19) \\
&= \sum_{i=1}^{n} I(X_{1,i}; Y_i | X_{2,i}, S_{R,i}) + n\epsilon & (20)
\end{aligned}
$$

13

Independence of side information ensures that $X_{1,i} \to S_{1,i} \to S_{2,i} \to X_{2,i}$ form a Markov chain as required.

Equation (17) follows from the fact that $X_{2,i}$ is a function of W_2 and S_R^n (resp. S_R^i) for non-causal (resp. causal) side information. (18) follows because conditioning reduces entropy and because Y_i is independent of $Y^{i-1}, W_1, W_2, S_R^{i-1}, S_{R,i+1}^n$ given $S_{R,i}, X_{1,i}, X_{2,i}$. Note that the inequality in (18) is necessary because Y_i is not independent of $Y^{i-1}, W_2, S_R^{i-1}, S_{R,i+1}^n$ given $S_{R,i}, X_{2,i}$. Intuitively, Y^{i-1} contains some information about W_1 and thus $X_{1,i}$ which affects Y_i. The converse for R_2 follows similarly. Finally, for the sum rate we have:

$$
\begin{aligned}
n(R_1 + R_2) &\leq I(W_1, W_2; Y^n, S_R^n) + n\epsilon \\
&= I(W_1, W_2; Y^n | S_R^n) + n\epsilon \\
&= \sum_{i=1}^{n} H(Y_i | S_R^n) - H(Y_i | W_1, W_2, Y^{i-1}, S_R^n) + n\epsilon \\
&\leq \sum_{i=1}^{n} H(Y_i | S_{R,i}) - H(Y_i | X_{1,i}, X_{2,i}, S_{R,i}) + n\epsilon \\
&\leq \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_i | S_{R,i}) + n\epsilon
\end{aligned}
$$

$\blacksquare$

6.1 Genie Bits and Value of Side Information for a Multiple Access Channel

Theorem 6 extends the single user result of Theorem 1 to the multiple access channel. By analogy to the single user case, a relationship between causal and non-causal side information capacity can be obtained by extending the results of Theorem 2 to the multiple access channel. Recall that in Theorems 2 and 3 we characterized the potential capacity benefits of side information at the transmitter and receiver for single user communications. Clearly, Theorem 3 extends directly to the multiple access channel, because if unbounded capacity gains are possible with transmitter side information in a single user channel, then the same must be true of the multiple access channel which includes the single user capacity as a special point in its capacity region. The extension of Theorem 2 to the multiple access channel is only slightly less straightforward as the derivation for the single user channel can be easily modified to the multiple access case as follows.

Theorem 7 *For the multiple access channel, the maximum possible sum capacity improvement $C_G^\Sigma - C^\Sigma$ due to the availability of receiver side information is bounded by the amount of the side information itself.*

$$
C_G^\Sigma - C^\Sigma \leq H(\mathcal{G}) = \lim_{N \to \infty} \frac{1}{N} H(G_1, G_2, \cdots, G_N).
$$

Proof:

$$
\begin{aligned}
C_G^{\Sigma} &= \sup_{p(X_{1,1^N}(W_1),X_{2,1}^N(W_2))} \lim_{N\to\infty} \frac{1}{N} I(W,W_2;Y^N,G^N) \\
&= \sup_{p(X_{1,1^N}(W_1),X_{2,1}^N(W_2))} \lim_{N\to\infty} \frac{1}{N} \left(I(W_1,W_2;Y^N) + I(W_1,W_2;G^N|Y^N) \right) \\
&= C^{\Sigma} + \Delta C^{\Sigma}
\end{aligned}
$$

where C_G^{Σ} is the sum capacity with the side information provided by the genie, C^{Σ} is the sum capacity without the side information and ΔC, the capacity improvement, is bounded by the entropy rate $H(\mathcal{G})$. Thus, the single user capacity result extends directly to the multiple access sum capacity. If the genie provides one bit of side information to the common receiver per channel use, the *sum* capacity benefit $\Delta C = C_G - C$ can not be more than 1 bit, regardless of the kind of side information. ■

6.2 Advantage of Non-causal Side Information over Causal Side Information

We compare the causal and non-causal capacity regions in terms of the sum rate point C_{Σ}. Similar to the single user case, we have shown that the sum capacities (and the entire capacity regions) are identical when S_{T1}, S_{T2} (independent) are also available to the receiver. To make this information available to the receiver requires $H(S_{T1},S_{T2}|S_R)$ genie bits per symbol. And because we have shown for the multiple access channel that genie bits can not improve capacity by more than their own entropy, we have the following result:

Theorem 8 *For the multiple access channel with independent side information at the transmitters, the sum capacity benefit of non-causal side information over causal side information is bounded as follows:*

$$
C_{\Sigma}^{non\text{-}causal}(S_T,S_R) - C_{\Sigma}^{causal}(S_T,S_R) \leq H(S_{T1},S_{T2}|S_R)
$$

Proof:

$$
\begin{aligned}
C_{\Sigma}(S_{T1},S_{T2},(S_R,S_{T1},S_{T2})) &\geq C_{\Sigma}^{non\text{-}causal}(S_{T1},S_{T2},S_R) \\
&\geq C_{\Sigma}^{causal}(S_{T1},S_{T2},S_R) \\
&\geq C_{\Sigma}(S_{T1},S_{T2},(S_R,S_{T1},S_{T2})) - H(S_{T1},S_{T2}|S_R).
\end{aligned}
$$

■

6.3 Example: Random Access Channel with Multiple Users

The following example presents a scenario with correlated side information and shows how the results may be applicable in several such cases as well. Consider the random access channel described before, except now the channel input is controlled by two users as:

$$X = X_1 S + X_2(1 - S).$$

Thus, a scheduler randomly allows user 1 or user 2 to access the channel in bursts of N symbols. As
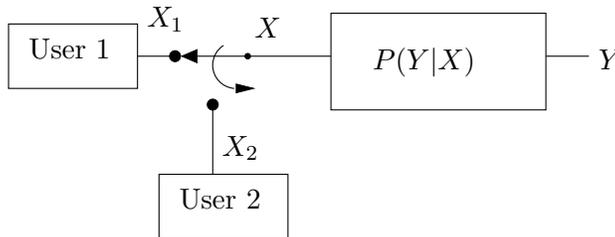

Figure 4: Random Access Channel with Two Users

before, the switch state S is known to the transmitters but not to the receiver. If a genie provides S to the receiver, the sum capacity is C_0. Using the same arguments as in the single user example, we have:

$$
\begin{aligned}
C_0 &= C_\Sigma(S, S, S) \\
&\geq C_\Sigma^{\text{non-causal}}(S, S, \phi) \\
&\geq C_\Sigma^{\text{causal}}(S, S, \phi) \\
&\geq C_\Sigma(S, S, S) - \frac{1}{N} H(S) = C_0 - 1/N.
\end{aligned}
$$

The practical implications of this example are quite interesting. In practice, random access is handled by the medium access layer (MAC layer) through explicit handshakes between the transmitter and receiver in the form of RTS-CTS (request to send, clear to send) messages. For a rapidly varying random access channel such exchanges can constitute a major overhead. Practical wireless systems such as the 802.11 Wireless LANs use nearly half the resources just for MAC layer overheads. However, the capacity results presented above show that even in the extremely rapidly varying random access channels where a transmission may or may not be made each symbol period without the knowledge of the receiver, the capacity loss is limited to less than a bit per channel use. For random access that fluctuates at a less rapid scale, the loss is negligible. This has interesting implications in how the RTS-CTS overhead can be minimized in practical systems through coding across access attempts.

7 The Broadcast Channel with Side Information

In the general broadcast channel, the transmitter has side information S_T and the two receivers have side information S_{R1} and S_{R2}. The capacity region of the general broadcast channel is not known even without side information. The best known achievable region is due to Marton [27] with auxiliary variables U, V and W. When the transmitter has causal side information, Marton's achievable region is directly extended to the convex hull of rate pairs (R_1, R_2) satisfying:

$$
\begin{aligned}
R_1 &\leq I(W,U;Y_1,S_{R1}) = I(W,U;Y_1|S_{R1}) \\
R_2 &\leq I(W,V;Y_2,S_{R2}) = I(W,V;Y_2|S_{R2}) \\
R_1 + R_2 &\leq \min\{I(W;Y_1,S_{R1}) + I(W;Y_2,S_{R2})\} + I(U;Y_1,S_{R1}|W) + I(V;Y_2,S_{R2}|W) - I(U;V|W) \\
&= \min\{I(W;Y_1|S_{R1}) + I(W;Y_2|S_{R2})\} + I(U;Y_1|S_{R1},W) + I(V;Y_2|S_{R2},W) - I(U;V|W)
\end{aligned}
$$

where $P(U,V,W,X|S_T) = P(U,V,W)P(X|S_T)$. Note that the only difference in the achievable region with causal side information versus Marton's innerbound without causal side information is that with causal side information the mapping from the auxiliary random variables U, V, W to the channel input alphabet X depends on the current value of the side information S_T, i.e. $X = f(U, V, W, S_T)$. In both cases U, V, W are independent of the side information. As usual the achievability proof for causal side information is straightforward as the mapping $f(.)$ can be incorporated into the channel to obtain a case with no side information where Marton's innerbound applies.

Extension of Marton's innerbound to case where non-causal side information is available at the transmitter is less straightforward. A simple and elegant proof of achievability of a subset of Marton's innerbound with only the auxiliary random variables U, V was provided by El Gamal and Van Der Meulen in [28] [2]. The achievable region with only U, V is the set of rates (R_1, R_2) satisfying

$$
\begin{aligned}
R_1 &\leq I(U;Y_1) \\
R_2 &\leq I(V;Y_2) \\
R_1 + R_2 &\leq I(U;Y_1) + I(V;Y_2) - I(U;V)
\end{aligned}
$$

for all $P(U, V, X)$.

El Gamal and van der Meulen's simple achievability proof was extended by the author in [29]

---

[2]While their proof directly addresses only the random variables U, V it can be extended to include W and hence the entire Marton's innerbound region

to incorporate non-causal side information as follows:

$$
\begin{aligned}
R_1 &\leq I(U; Y_1, S_{R1}) - I(U; S_T) \\
R_2 &\leq I(V; Y_2, S_{R2}) - I(V; S_T) \\
R_1 + R_2 &\leq I(U; Y_1, S_{R1}) + I(V; Y_2, S_{R2}) - I(U; V) - I(U, V; S_T)
\end{aligned}
$$

for all $P(U, V, X | S_T) = P(U, V) P(X | U, V, S_T)$. While a full extension of Marton's innerbound with side information has been found recently, the proof of the abovementioned limited extension is interesting for its simplicity and also illustrates the binning concept central to the full extension. Since, the same concept can be applied to obtain achievable regions for many multiuser communication scenarios with non-causal side information [29] we include a sketch of this proof here.

[Sketch of Proof:] Proceeding as in [28], $2^{n(I(U; Y_1, S_{R1}) - \epsilon)}$ (resp. $2^{n(I(V; Y_2, S_{R2}) - \epsilon)}$) i.i.d. U sequences (resp. V sequences) are generated independently according to $P(U)$ (resp. $P(V)$) and uniformly distributed over 2^{nR_1} (resp. 2^{nR_2}) bins. This forms the codebook that is shared by all parties prior to beginning of communication. During the communication phase, the message index $W_1 \in [1, 2^{nR_1}]$ (resp. $W_2 \in [1, 2^{nR_2}]$) marks the appropriate U (resp. V) sequence bin selected. Given the state sequence S_T^n, the transmitter finds a U sequence and a V sequence in the chosen bins so that U^n, V^n, S^n are jointly typical. The probability that independently generated U^n (resp. V^n) and S^n are jointly typical is bounded by $2^{-n(I(U; S_T) - \delta(\epsilon))}$ (resp. $2^{-n(I(V; S_T) - \delta(\epsilon))}$). This is reflected in the individual rate constraints. Finally, the probability that independently generated U^n, V^n and S^n are jointly typical is bounded by $2^{-n(H(U) + H(V) + H(S) - H(U, V, S)) - \delta(\epsilon)} = 2^{-n(I(U; V) + I(U, V; S)) - \delta(\epsilon)}$ which results in the sum rate constraint. ∎

The complete extension of Marton's innerbound extension to non-causal side information was obtained in an independent and parallel work by Steinberg and Shamai [30]. With non-causal side information at the transmitter, Marton's innerbound becomes the convex hull of all rate pairs (R_1, R_2) satisfying [30]:

$$
\begin{aligned}
R_1 &\leq I(W, U; Y_1, S_{R1}) - I(W, U; S_T) \\
R_2 &\leq I(W, V; Y_2, S_{R2}) - I(W, V; S_T) \\
R_1 + R_2 &\leq -[\max\{I(W; Y_1, S_{R1}), I(W; Y_2, S_{R2})\} - I(W; S_T)]_+ \\
&\quad + I(W, U; Y_1, S_{R1}) - I(W, U; S_T) + I(W, V; Y_2, S_{R2}) - I(W, V; S_T) - I(U; V | W, S_T)
\end{aligned}
$$

for all $P(U, V, W, X | S_T)$.

7.1 Genie Bounds and the Relative Advantage of Non-causal Side Information over Causal Side Information

Since the capacity region is not known for the general broadcast channel, there are two interesting alternatives to consider. As a first alternative, we could limit our attention to the special cases where the capacity region is known, such as the degraded broadcast channel. However, since our interest is to bound the *sum* capacity improvement with causal and non-causal side information, the degraded case becomes trivial. For the degraded broadcast channel the sum capacity is the single user capacity of the stronger user, and therefore all the single user results obtained in previous sections apply.

Another possibility is to limit our attention to Marton's innerbound region instead of the actual capacity region. In that case, the natural question is whether the extensions of Marton's innerbound for causal and non-causal side information become identical when the side information at the transmitter is also available at both the receivers. While this result was shown to be true for the general multiple access channel, it is interesting that the corresponding extension for the broadcast channel turns out to be less general.

Observation 1 *The extensions of Marton's innerbound with causal and non-causal side information (as described above) are identical if the following condition is satisfied:*

$$I(U; S_T | W) + I(V; S_T | W) = I(U, V; S_T | W)$$

Intuitively, one can interpret the condition as follows. The common state information is provided by W and therefore conditioned on W, the auxiliary random variables U and V should provide independent information about the state S_T. It is not known if this condition actually makes the achievable region smaller. A sketch of the proof of this observation is provided in the Appendix.

Following our previous results, the next question is to bound the maximum capacity benefit of genie bits. A straightforward extension of the single user result leads to the following bound: If a genie provides G_1 bits of side information to receiver 1 and G_2 bits of side information to receiver 2, the sum capacity can not improve by more than $G_1 + G_2$ bits. The proof is straightforward since the improvement in the individual rates R_1, R_2 is bounded by G_1, G_2 respectively.

7.2 Example: Broadcast Channel

Consider a two user fading broadcast channel where the users' channels experience i.i.d. block fading (block size N) given by h_1 and h_2 respectively.

$$Y_1 = h_1 X + n_1$$
$$Y_2 = h_2 X + n_2$$

where $n_1, n_2 \sim \mathcal{N}(0,1)$ are additive white Gaussian noise terms and a transmit power constraint of P is in place. It is assumed that each user knows only his own channel, while the transmitter knows only the side information variable S_T ($S_T = 1$ if $h_1 > h_2$ and 0 otherwise). While in a degraded broadcast channel it is well known that the sum rate maximizing policy is to transmit only to the strongest user, the problem faced here is that a receiver does not know when it has the stronger channel compared to the other receiver's channel. Clearly, if a genie provides S_T to both receivers then the sum capacity of this channel is $C_0 = C_\Sigma(S_T, (h_1, S_T), (h_2, S_T)) = \mathrm{E} \log(1 + P \max(|h_1|^2, |h_2|^2))$. Since the genie's bits can not increase capacity by more than 2 bits (one bit to each receiver), the sum capacity (with causal or non-causal side information) can be bounded as:

$$C_0 \geq C_\Sigma^{\text{non-causal}}(S_T, h_1, h_2) \geq C_\Sigma^{\text{causal}}(S_T, h_1, h_2) \geq C_0 - 2/N.$$

8   The Relay Channel

The discrete memoryless relay channel with side information consists of source input alphabet $\mathcal{X}_S$, relay input alphabet $\mathcal{X}_R$, state alphabet $\mathcal{S}_S, \mathcal{S}_R, \mathcal{S}_D$ at the source (and relay and destination (respectively), relay output alphabet $\mathcal{Y}_R$, the destination output alphabet $\mathcal{Y}_D$, and a probability transition function $P(Y_R, Y_D | X_S, X_R, S_S, S_R, S_D)$.

Our first goal is to extend the result of Theorem 1 to the relay channel. For the general relay channel, the capacity is unknown. For the special case of the degraded relay channel with no side information the capacity was found by [31]. In recent work, Sigurjonsson and Kim [23] combine Shannon's proof technique [1] with Cover and El Gamal's [31] relay coding theorems to determine the capacity of the degraded relay channel with causal side information for the special case $S_R = S_S = S$, i.e. the same state information is available to both the source and the relay. From the perspective of Theorem 1 we are interested in the scenario where the side information available to the source and relay is also available to the destination. Mathematically, we define the physically degraded channel with (causal or non-causal) side information where $S_R = S_S = S = f(S_D)$ as follows:

$$P(Y_D, Y_R | X_S, X_R, S, S_D) = P(Y_R | X_S, X_R, S) P(Y_D | Y_R, X_R, S_D). \tag{21}$$

Figure 5 illustrates the relay channel with side information. The following theorem extends the result of Theorem 1 to the relay channel.

Theorem 9 *The capacity of the physically degraded relay channel described above with causal or non-causal side information is given by:*

$$C^{causal} = C^{non\text{-}causal} = \max_{P(X_S, X_R | S)} \min \left[I(X_S, X_R; Y_D | S_D), I(X_S; Y_R | X_R, S) \right]$$
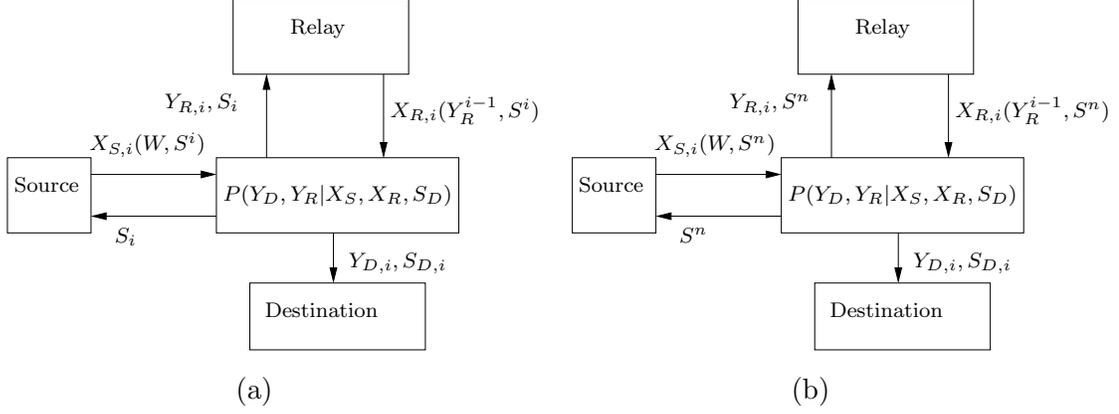
Figure 5: Relay channel with (a) Causal and (b) Non-causal Side Information

Proof: : [Achievability] Achievability follows directly from the idea of multiplexed codebooks. Separate codebooks are designed for each possible state S. As the states are revealed in a causal fashion, the transmitter and the relay switch between the corresponding codebooks. Since the receiver also knows S (a deterministic function of S_D), the receiver makes the corresponding switch as well. In this manner, for each state $S = s$, the capacity

$$C_s = \max_{P(X_S, X_R | S = s)} \min\left[I(X_S, X_R; Y_D | S_D, S = s), I(X_S; Y_R | X_R, S = s)\right]$$

is achieved. Averaging over S we obtain the capacity expression of Theorem 9. Note that codebook multiplexing requires only causal side information. What is achievable with causal side information is also achievable with non-causal side information. ∎

Proof: : [Converse]

$$
\begin{aligned}
I(W; Y_D^n, S_D^n) &\leq I(W; Y_D^n, Y_R^n | S_D^n) \\
&= \sum_{i=1}^{n} H(Y_{D,i}, Y_{R,i} | Y_D^{i-1}, Y_R^{i-1}, S_D^n) - H(Y_{D,i}, Y_{R,i} | Y_D^{i-1}, Y_R^{i-1}, S_D^n, W) \\
&= \sum_{i=1}^{n} H(Y_{D,i}, Y_{R,i} | Y_D^{i-1}, Y_R^{i-1}, S_D^n, X_{R,i}) - H(Y_{D,i}, Y_{R,i} | Y_D^{i-1}, Y_R^{i-1}, S_D^n, W, X_{R,i}, X_{S,i}) \\
&\leq \sum_{i=1}^{n} H(Y_{D,i}, Y_{R,i} | X_{R,i}, S_{D,i}) - H(Y_{D,i}, Y_{R,i} | S_{D,i} X_{R,i}, X_{S,i}) \\
&= \sum_{i=1}^{n} I(X_{S,i}; Y_{D,i}, Y_{R,i} | X_{R,i}, S_{D,i})
\end{aligned}
$$

Using Fano's inequality and the time sharing variable to get a single letter characterization we get

$$
\begin{aligned}
R &\leq I(X_S; Y_D, Y_R | X_R, S_D) + \epsilon_n \\
&= I(X_S; Y_R | X_R, S_D) + I(X_S; Y_D | X_R, S_D, Y_R) + \epsilon_n
\end{aligned}
$$

21

However, conditioned on S_D, the degraded channel condition implies that $X_S, (Y_R, X_R), Y_D$ is a Markov chain, i.e. $I(X_S; Y_D | X_R, S_D, Y_R) = 0$ and we have,

$$
\begin{aligned}
R &\leq I(X_S; Y_R | X_R, S_D) + \epsilon_n \\
&= H(Y_R | X_R, S_D, S) - H(Y_R | X_R, S_D, X_S, S) + \epsilon_n \\
&\leq H(Y_R | X_R, S) - H(Y_R | X_R, X_S, S) + \epsilon_n \\
&= I(X_S; Y_R | X_R, S) + \epsilon_n
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
I(W; Y_D^n, S_D^n) &= \sum_{i=1}^{n} H(Y_{D,i} | Y_D^{i-1}, S_D^n) - H(Y_{D,i} | Y_D^{i-1}, S_D^n, W) \tag{22} \\
&\leq \sum_{i=1}^{n} H(Y_{D,i} | S_{D,i}) - H(Y_{D,i} | Y_D^{i-1}, S_D^n, W, Y_R^{i-1}) \tag{23} \\
&= \sum_{i=1}^{n} H(Y_{D,i} | S_{D,i}) - H(Y_{D,i} | Y_D^{i-1}, S_D^n, W, Y_R^{i-1}, X_{S,i}, X_{R,i}) \tag{24} \\
&= \sum_{i=1}^{n} H(Y_{D,i} | S_{D,i}) - H(Y_{D,i} | S_{D,i}, X_{S,i}, X_{R,i}) \tag{25} \\
&= \sum_{i=1}^{n} I(X_{S,i}, X_{R,i}; Y_{D,i} | S_{D,i}) \tag{26}
\end{aligned}
$$

Combining (26) and (22) with Fano's inequality we have,

$$
R \leq \min \left[I(X_S, X_R; Y_D | S_D), I(X_S; Y_R | X_R, S) \right] + \tilde{\epsilon}_n
$$

and the converse proof is complete. ∎

8.1 Genie Bound and the Relative Advantage of Non-causal Side Information over Causal Side Information

For the relay channel with only one destination, the result of Theorem 2 applies with the identical proof. In other words, one bit of genie information at the receiver does not improve capacity by more than one bit. Combining with the result of Theorem 9 we have the following result: For the physically degraded relay channel with side information defined by (21) (in general without the constraint $S = f(S_D)$) the difference between capacity with non-causal and causal side information at the transmitter is bounded as: $C^{\text{non-causal}} - C^{\text{causal}} \leq H(S | S_D)$.

9 Conclusion

Side information is a crucial factor in determining the capacity of a channel. Previous research has shown that for a single user memoryless channel even perfect feedback does not result in a capacity advantage. However, we find that even one bit of causal side information at the transmitter can increase capacity by an unbounded amount. The contrasting results are due to the fact that perfect feedback can only provide information about past channel states, whereas causal side information can provide information about the current channel state. Thus, for the transmitter the difference between the past state information and present state information is very significant. We further explore the relative advantage of knowing the future channel states (non-causal side information) relative to the knowledge of only the present channel state (causal side information). For a single user, we find that the knowledge of future channel states can increase capacity only if the state information is not available to the receiver. In other words, non-causal side information has no capacity benefit over causal side information when the side information is available to the receiver as well. Note that for the receiver it does not matter if the side information is made available causally or non-causally, because no delay constraint is assumed in the decoding operation for capacity results. We evaluate the benefits of receiver side information in the form of a genie bound, which states quite simply that the capacity advantage from one genie bit of side information at the receiver can not exceed one bit. This gives us a bound on the capacity benefit of non-causal side information at the transmitter over causal side information in the form of the number of bits required to inform the receiver of the transmitter side information that is not already available to the receiver.

We follow single user results with multiuser scenarios. All the results are found to extend to the general multiple access channel as well. For the broadcast channel, the extension is less general. In particular, for the broadcast channel it is not clear (even in terms of Marton's innerbound) if there is any benefit of non-causal side information over causal side information when both receivers know all the transmitter side information. This equivalence is established only subject to an interesting constraint. Finally, we show that the single user results extend to the degraded relay channel. Examples of random access channels are provided throughout as illustrations of the capacity bounds. While the MAC layer overheads in practical systems can be excessive the capacity results show that such overheads can be nearly eliminated through clever coding across multiple access blocks.

Appendix

A Sketch of Proof of Observation 1

When $S_T = f_1(S_{R1}) = f_2(S_{R2})$, the following simplifications can be made, as in the proof of Theorem 1:

$$I(W,U;Y_1,S_{R1}) - I(W,U;S_T) = I(W,U;Y_1|S_{R1}).$$
$$I(W,V;Y_2,S_{R2}) - I(W,V;S_T) = I(W,V;Y_2|S_{R2}).$$
$$\max\{I(W;Y_1,S_{R1}), I(W;Y_2,S_{R2})\} - I(W,U;S_T) = \max\{I(W;Y_1|S_{R1}), I(W;Y_2|S_{R2})]$$

and finally

$$-\max\{I(W;Y_1|S_{R1}), I(W;Y_2|S_{R2})] + I(W,U;Y_1|S_{R1}) + I(W,V;Y_2|S_{R2})$$
$$= \min\{I(W;Y_1|S_{R1}) + I(W;Y_2|S_{R2})\} + I(U;Y_1|S_{R1},W) + I(V;Y_2|S_{R2},W)$$

Using the simplifications it is easy to see that the causal case and the non-causal case are identical if $I(U;V|W,S) = I(U;V|W)$. But,

$$I(U;V|W,S) = I(U;V|W) + I(U,V;S|W) - I(U;S|W) - I(V;S|W)$$

and thus we have the condition of observation 1. ∎

References

[1] C. E. Shannon, "Channels with side information at the transmitter," in *IBMJ. Res. Develop*, pp. 289–293, 1958.

[2] A. Kusnetsov and B. Tsybakov, "Coding in a memory with defective cells," *Prob. Pered. Inform.*, vol. 10, pp. 52–60, April-June 1974.

[3] S. I. Gelfand and M. Pinsker, "Coding for channel with random parameters," in *Problems of Control and Information Theory*, pp. 19–31, 1980.

[4] C. Heegard and A. El Gamal, "On the capacity of computer memories with defects," *IEEE Trans. Inform. Theory*, vol. 29, pp. 731–739, Sep. 1983.

[5] A. Rosenzweig, Y. Steinberg, and S. Shamai, "On channels with partial channel state information at the transmitter," *IEEE Transactions on Information Theory*, vol. 51, pp. 1817–1830, May 2005.

[6] Y. Cemal and Y. Steinberg, "The multiple-access channel with partial state information at the encoders," *IEEE Trans. Inform. Theory*, vol. 51, pp. 3992–4003, Nov. 2005.

[7] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1629–1638, June 2002.

[8] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," in *Proceedings of Int. Symp. Inform. Theory*, p. 500, June 2002.

[9] N. Jindal, S. Vishwanath, S. Jafar, and A. Goldsmith, "Duality, dirty paper coding and capacity for multiuser wireless channels," in *Proceedings of DIMACS workshop on Signal Processing for Wireless Transmissions*, October 2002.

[10] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of MIMO broadcast channels," *IEEE Trans. Inform. Theory*, pp. 2895–2909, Oct. 2003.

[11] Y. Steinberg and N. Merhav, "Identification in the presence of side information with application to watermarking," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1410–1422, May 2001.

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

[13] Y. H. Kim, A. Sutivong, and S. Sigurjonsson, "Multiple user writing on dirty paper," *Proc. of IEEE Int. Symp. Information Theory*, p. 534, June/July 2004.

[14] Y. Steinberg and S. Shamai, "Achievable rates for the broadcast channel with states known at the transmitter," in *Proceedings of IEEE Int. Symp. on Info. Theory*, Sep. 2005.

[15] W. Yu and J. M. Cioffi, "Sum capacity of a Gaussian vector broadcast channel," in *Proceedings of Int. Symp. Inform. Theory*, p. 498, June 2002.

[16] P. Viswanath and D. N. Tse, "Sum capacity of the multiple antenna Gaussian broadcast channel," in *Proceedings of Int. Symp. Inform. Theory*, p. 497, June 2002.

[17] S. Vishwanath, G. Kramer, S. Shamai, S. Jafar, and A. Goldsmith, "Capacity bounds for Gaussian vector broadcast channels," in *Proceedings of DIMACS workshop on Signal Processing for Wireless Transmissions*, October 2002.

[18] P. Viswanath and D. Tse, "On the capacity of the multiple antenna broadcast channel," in *Proceedings of DIMACS workshop on Signal Processing for Wireless Transmissions*, October 2002.

[19] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," in *Proceedings of Conference on Information Sciences and System s*, March 2004.

[20] M. Salehi, "Capacity and coding for memories with real-time noisy defect information at encoder and decoder," *Proc. Inst. Elec. Eng. Pt. I*, vol. 139, pp. 113–117, April 1992.

[21] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2007–2019, Sept 1999.

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

[23] S. Sigurjonsson and Y. Kim, "On multiple user channels with causal state information at the transmitters," *cs.IT/0508096*, 2005. Download available at http://arxiv.org.

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[25] M. Chiang and T. Cover, "Unified duality between channel capacity and rate distortion with state information," in *Proceedings of IEEE Int. Symp. Inform. Theory*, p. 301, June 2001.

[26] S. Jafar and S. Srinivasa, "Capacity limits of cognitive radio with distributed and dynamic spectral activity," *cs.IT/0509077*, 2005. Download available at http://arxiv.org.

[27] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. 25, pp. 306–311, May 1979.

[28] A. El Gamal and E. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. 27, pp. 120–122, Jan. 1981.

[29] S. Jafar, "Degrees of freedom in distributed MIMO communication," *IEEE Communication Theory Workshop*, June 2005.

[30] Y. Steinberg and S. Shamai, "Achievable rates for the broadcast channel with state known at the transmitter," in *Proceedings of IEEE Int. Symp. Inform. Theory*, 2005.

[31] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, pp. 572–584, Sept. 1979.