# Information Theory for Wireless Communications:

## Lecture 11: Channel Capacity with Side Information

Instructor: Dr. Saif K. Mohammed
Scribe: Johannes Lindblom

In this lecture, we derive the channel capacities for the four scenarios of 1) no channels side information at the transmitter (CSIT) and no channels side information at the receiver (CSIR), 2) CSIR and no CSIT, 3) CSIT and no CSIR and 4) CSIT and CSIR. We consider the general system depicted in Fig. 1. Here, the channel output $y_k$ at time $k$ does not only depend on the input $x_k$, but also on the channel state $s_k$. We model $s_k \in \mathcal{S} \triangleq \{1, \ldots, t\}$ as a i.i.d. random variable with p.d.f. $p_S(\cdot)$. This lecture follows [1, Ch. 4.6]. A elaborated treatment of the topic can be found in [2].

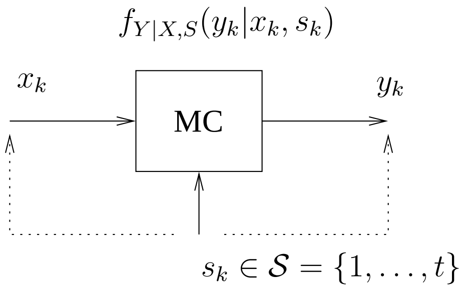


Fig. 1. General system model for the memoryless channel with and without channel side information.

### I. No CSIT, No CSIR

The case of no CSIT and no CSIR corresponds to the scenario where the channel's coherence time is to short to perform training. The fading state changes at each channel use. We assume that the p.m.f. $p_S(\cdot)$ is known at both the transmitter and the receiver. This situation is illustrated in Fig. 2. The dashed box constitutes the effective channel for which we have

$$p(j|i) = \Pr\{Y = b_j | X = a_i\} = \sum_{r=1}^{t} \Pr\{Y = b_j | X = a_i, S = r\} \, p_S(r)$$

and the capacity is

$$C = \max_{p_X(\cdot)} I(X;Y).$$

**Example 1.** *Consider the continuous fading AWGN channel illustrated in Fig. 3. We have $Y = SX + W$, where $S$ is the channel state and $W$ is the AWGN. For this setup, the c.d.f. of the effective channel is*

$$f_{Y|X}(y|x) = \int f_{Y|X,S}(y|x,s) f_S(s) ds$$

**This document is a property of Communication Systems Division, Department of Electrical Engineering, Linköping University, Sweden. Copyright mus be obtained by writing to** {saif,erik.larsson}@isy.liu.se **prior to usage.**

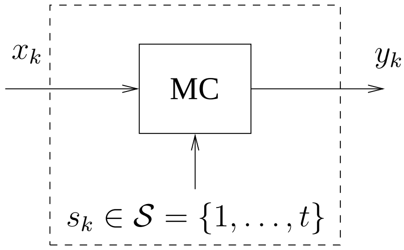


Fig. 2. System model for the memoryless channel without channel side information. The effective channel is inside the the dashed box.

*where $f_S(s)$ is the p.d.f. of the channel states. The capacity of this channel is*

$$C = \max_{f_x(\cdot),\ \mathrm{E}(X^2) \le P} I(X;Y)$$

*for some maximum power $P$.*

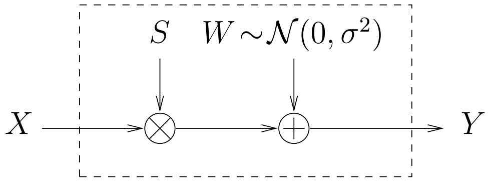


Fig. 3. The AWGN Channel for the case of no channel side information. The effective channel is the content inside the dashed rectangle.

## II. CSIR, No CSIT

This case corresponds to the scenario where we assume that the receiver can perfectly estimate the channel whereas the transmitter only knows the statistical distribution of the channel. We can model this scenario as depicted in Fig. 4, where the input to the channel is the transmitted symbol $X$ and the output consists of both the symbol $Y$ and the channel state $S$.

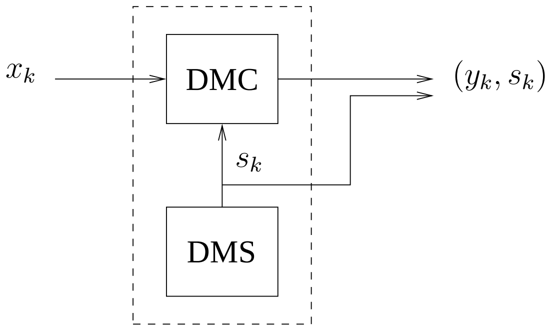


Fig. 4. System model for the case of CSIR and no CSIT.

We have the mutual information

$$I(X;(Y,S)) = H(Y,S) - H(Y,S|X) = H(S) + H(Y|S) - \underbrace{H(S|X)}_{=H(S)} - H(Y|S,X)$$

$$= H(Y|S) - H(Y|S,X) = \sum p_S(r) p(y,x|r) \log \frac{p(y|x,r)}{p(y|r)}.$$

For the case of CSIR, we have $I(X;(Y,S)) = H(X) - H(X|Y,S)$. For the no CSI case we have $I(X;Y) = H(X) - H(X|Y)$. Since $H(X|Y,S) \leq H(X|Y)$, we have $I(X;(Y,S)) \geq I(X;Y)$, i.e., the mutual information is higher for the case of CSIR than for the case of no CSI.

**Example 2.** *Consider the continuous fading AWGN depicted in Fig. 5. We have* $Y = SX + W$, *where* $S$ *is the channel state and* $W$ *is the AWGN. The capacity for this channel is*

$$C = \max_{f_X(\cdot),\ \mathrm{E}(X^2)\leq P} h(Y|S) - \underbrace{h(Y|S,X)}_{=h(W)}.$$

*We have*

$$h(Y|S) = -\int f_S(r) f_{Y|S}(y|r) \log f_{Y|S}(y|r) dy dr = \int f_S(r) \left( \int -f_{Y|S}(y|r) \log f_{Y|S}(y|r) dy \right) dr, \quad (1)$$

*where* $f_{Y|S}(y|r)$ *is the c.d.f. of the received symbol given the channel state. The inner integral of* (1) *is maximized if* $Y|S$ *is Gaussian. In order to achieve that,* $X$ *must be Gaussian. Moreover, in order to maximize* (1), *we must have* $\mathrm{Var}(X) = P$. *Now,* (1) *becomes*

$$h(Y|S) = \frac{1}{2} \int f_S(r) \log\left(2\pi e \left(r^2 P + \sigma^2\right)\right) dr.$$

*Hence, the capacity is*

$$C = \frac{1}{2} \int f_S(r) \log\left(1 + \frac{r^2 P}{\sigma^2}\right) dr \leq \frac{1}{2} \log\left(1 + \frac{\mathrm{E}(S^2) P}{\sigma^2}\right), \quad (2)$$

*where we used Jensen's inequality. The right-hand side of* (2) *is the AWGN capacity with the same average received power.*

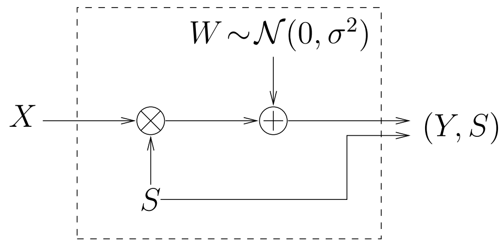


Fig. 5. The AWGN Channel for the case of only CSIR. The effective channel is the content inside the dashed rectangle.

## III. Causal CSIT, No CSIR

We consider the case of causal CSIT and no CSIR. For this scenario, the transmitted codeword will not only depend on the message that we want to convey but on the current state as well. The system model for this scenario is depicted in Fig. 6.

Let $\{1, \ldots, w, \ldots, N\}$ be the set of messages, let $\mathcal{X} \triangleq \{1, 2, \ldots, M\}$ denote the alphabet of the channel inputs $x_k$, and let $\mathcal{S} \triangleq 1, 2, \ldots, t$ be the set of channel states. Let us assume that we at the $k$th channel use want to transmit message $w$. At the $k$th channel use we send the symbol

$$x_k = g_k(w, s_1, \ldots, s_{k-1}, s_k), \quad (3)$$

where $g_k(\cdot)$ is the precoding function. In (3) we assume that we have causal CSIT, i.e., only the history and the current channel state are known. Since the channel is memoryless, it suffices that the precoding function only depends on the current state, i.e.,

$$x_k = g_k(w, s_k)$$

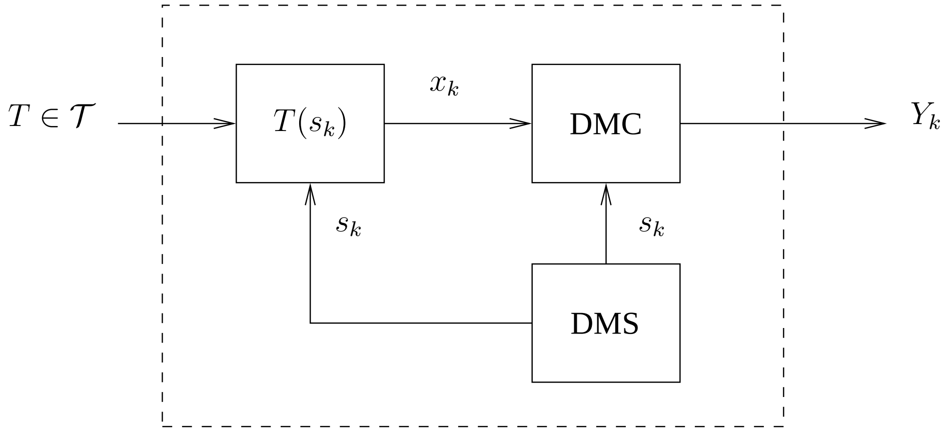


Fig. 6. System model for the case of causal CSIT and no CSIR. The dashed box mark the effective DMC

For a given message $w$, $g_k(w, \cdot)$ is a mapping $\mathcal{S} \to \mathcal{X}$. Let $T$ be the set of mappings. Since $|\mathcal{S}| = t$ and $|\mathcal{X}| = M$, there are $|\mathcal{T}| = M^t$ possible mappings. This mapping is illustrated in Tab. I. Each codeword is generated randomly and is nothing but a sequence of mappings. The decoder finds the "codeword" which is the unique codeword typical to $y^n$.

| | Time instance | | | |
|---|---|---|---|---|
| Message | 1 | 2 | $\cdots$ | $n$ |
| 1 | $g_1(1, s_1)$ | $g_2(1, s_2)$ | $\cdots$ | $g_n(1, s_n)$ |
| $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ |
| $w$ | $g_1(w, s_1)$ | $g_2(w, s_2)$ | $\cdots$ | $g_n(w, s_n)$ |
| $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ |
| $N$ | $g_1(N, s_1)$ | $g_2(N, s_2)$ | $\cdots$ | $g_n(N, s_n)$ |

TABLE I
MAPPING FROM CHANNEL STATE TO CHANNEL INPUT FOR A GIVEN MESSAGE.

**Example 3.** *We have $N = 4$ messages that we want to encode with a $n = 3$ code. The states belong to the set $\mathcal{S} = \{0, 1\}$ and the channel input alphabet is $\mathcal{X} = \{0, 1\}$. Hence, we have $2^2 = 4$ mappings $\mathcal{S} \to \mathcal{X}$:*

$$\xi_1 : \begin{array}{ccc} 0 & \to & 0 \\ 1 & \to & 1 \end{array} \quad \xi_2 : \begin{array}{ccc} 0 & \to & 0 \\ 1 & \to & 0 \end{array} \quad \xi_3 : \begin{array}{ccc} 0 & \to & 1 \\ 1 & \to & 1 \end{array} \quad \xi_4 : \begin{array}{ccc} 0 & \to & 1 \\ 1 & \to & 0 \end{array} \tag{4}$$

*An example codebook is given in Tab. II. Assume that the state sequence is $s_1 = 0$, $s_2 = 1$, and $s_3 = 0$, and that we want to transmit message $3$. Therefore, using Tab. II, the transmitted sequence is $\xi_4(0) = 1$, $\xi_2(1) = 0$, and $\xi_2(0) = 0$.*

The equivalent channel, i.e., what is inside the dashed rectangle in Fig. 6, is characterized by the conditional probability functions $p_{Y|T}(y|t)$ for all $y \in \mathcal{Y}$ and $t \in \mathcal{T}$. We would like to compute the mutual

| | Time instance | | |
|---|---|---|---|
| Message | 1 | 2 | 3 |
| 1 | $\xi_1$ | $\xi_4$ | $\xi_3$ |
| 2 | $\xi_2$ | $\xi_3$ | $\xi_1$ |
| 3 | $\xi_4$ | $\xi_2$ | $\xi_2$ |
| 4 | $\xi_1$ | $\xi_3$ | $\xi_4$ |

TABLE II
CODEBOOK FOR EXAMPLE 3.

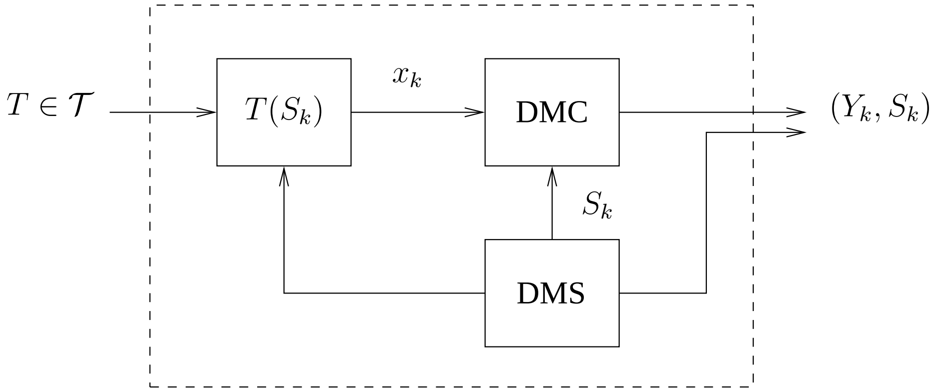


Fig. 7. System model for the case of both CSIT and CSIR. The dashed box mark the effective DMC. Note that the output contains both the received symbol $Y_k$ and the state $S_k$.

information $I(T;Y)$. First, note that $\mathcal{T} = \{\xi_1, \dots, \xi_{|\mathcal{T}|}\}$. We have

$$\begin{aligned} p_{Y|T}(Y = b_j | T = \xi_l) &= \sum_{r=1}^{t} p_{Y,S|T}(Y = b_j, S = r | T = \xi_l) \\ &= \sum_{r=1}^{t} p_{S|T}(S = r | T = \xi_l) p_{Y,|S,T}(Y = b_j | S = r, T = \xi_l) \\ &= \sum_{r=1}^{t} p_S(S = r) p_{Y,|S,T}(Y = b_j | S = r, T = \xi_l). \end{aligned}$$

The capacity for channel depicted in Fig. 6 is then

$$C = \max_{p_T(\cdot)} I(T;Y).$$

## IV. CAUSAL CSIT AND CSIR

Here, we consider the case with causal CSIT and (possibly) non-causal CSIR. This situation is depicted in Fig. 7. We extend the model given in Sec. III by letting the output also contain the channel state.

The rate of this system is

$$\begin{aligned} R &= I(T;(Y,S)) = H(Y,S) - H(Y,S|T) \\ &= H(Y|S) + H(S) - H(S|T) - H(Y|S;T) = H(Y|S) - H(Y|S;T), \end{aligned} \tag{5}$$

where the last equality holds since $S$ and $T$ are independent. We have

$$H(Y|S) = -\sum_{r=1}^{t}\sum_{y=1}^{M'} p_S(r)p_{Y|S}(y|r)\log_2 p_{Y|S}(y|r) \tag{6}$$

and

$$H(Y|S,T) = -\sum_{r=1}^{t}\sum_{l=1}^{|\mathcal{T}|}\sum_{y=1}^{M'} p_S(r)p_T(\xi_l)p_{Y|S,T}(y|r,\xi_l)\log_2 p_{Y|S,T}(y|r,\xi_l). \tag{7}$$

By using the fact that $p_{Y|S}(y|r) = \sum_l p_{Y,T|S}(y,\xi_l|r) = \sum_l p_T(\xi_l)p_{Y|S,T}(y|r,\xi_l)$ and inserting (6)–(7) into (5), we get

$$I(T;(Y,S)) = \sum_{r=1}^{t} p_S(r)\sum_{l=1}^{|\mathcal{T}|}\sum_{y=1}^{M'} p_T(\xi_l)p_{Y|S,X}(y|r,x=\xi_l(r))\log_2\frac{p_{Y|S,X}(y|r,x=\xi_l(r))}{p_{Y|S}(y|r)}. \tag{8}$$

For fixed $S=r$, we have $X=T(r)$ with p.m.f. $p_T(\cdot)$ we have

$$p_X^r(x=a_i) = \sum_{\{\xi_l|\xi_l(r)=a_i\}} p_T(\xi_l). \tag{9}$$

Then, we can write the two inner sums of (8) as

$$\begin{aligned}
&\sum_{i=1}^{M}\left(\sum_{\{\xi_l|\xi_l(r)=a_i\}} p_T(\xi_l)\right)\sum_{y=1}^{M'} p_{Y|S,X}(y|r,x=a_i)\log_2\frac{p_{Y|S,X}(y|r,x=a_i)}{p_{Y|S}(y|r)}\\
&=\sum_{i=1}^{M} p_X^r(x=a_i)\sum_{y=1}^{M'} p_{Y|S,X}(y|r,x=a_i)\log_2\frac{p_{Y|S,X}(y|r,x=a_i)}{p_{Y|S}(y|r)}\\
&=I_{p_{X|S}(x|r)}(X;Y|r).
\end{aligned} \tag{10}$$

For this fixed state $r$ we have a DMC with input $X$, output $Y$, and c.d.f. $p_{Y|X,S}(u|x,r)$ with capacity

$$C(r) = \max_{p_{X(r)}(\cdot)} I(X;Y) \tag{11}$$

for the capacity achieving distribution $p^\star_{X(r)}$. By taking the expectation of (11), we get the achievable rate

$$R = \sum_{r=1}^{t} p_S(r)C(r)$$

which becomes the capacity of the channel. This follows from the functional representation lemma [3, Appendix B], which implies that we can find a distribution on mappings that induce a p.d.f. that achieves capacity.

The interpretation of the result above is that we have $t$ codebooks $\mathcal{C}_1,\ldots,\mathcal{C}_t$, i.e., one for each mapping, over the input alphabet $\mathcal{X}$. The message $w$ consists of $t$ sub-messages - one for each code book, i.e., $w=(w_1,\ldots,w_t)$. When we want to send $w$, we send $w_1$ $np_S(r=1)$ times and $w_t$ $np_S(r=t)$ times. Each code book $\mathcal{C}_r$ consists of $2^{np_S(r)C(r)}$ codewords. In total, we have

$$\prod_{r=1}^{t} 2^{np_S(r)C(r)}$$

possible messages. Note that

$$\frac{1}{n}\log_2\left(\prod_{r=1}^{t} 2^{np_S(r)C(r)}\right) = \sum_{r=1}^{t} p_S(r)C(r).$$

## References


[1] J. Wolfowitz, “Coding Theorems of Information Theory,” *Ergebnisse Der Mathematik Und Ihrer Grenzgebiete,* 2nd Ed. 1964.

[2] G. Keshet, Y. Steinberg and N. Merhav, “Channel Coding in the presence of side information: Subject Review,” Vol. 4, No. 6, Now Publishers, 2008.

[3] A. El Gamal and Y.-H. Kim, “Network Information Theory,” Cambridge University Press, 2011.