

Consider a discrete memoryless channel $\Pr(Y|X)$ shown in Figure 1.

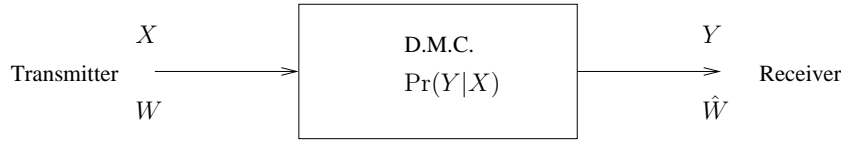


Figure 1: Discrete Memoryless Channel

The transmitter takes a message W from a message set and sends to the receiver at rate R , which is defined by:

$$R = \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 H(W) \quad (1)$$

Let \hat{W} be the decoding result computed by the receiver.

Theorem 1 (Shannon's Coding Theorem for Discrete Memoryless Channel) :

$$\text{If } R < I(X; Y), \text{ then } \lim_{n \rightarrow \infty} \Pr(W \neq \hat{W}) = 0.$$

Achievability Proof: Assume the communication spans over n channel uses. Without loss of generality, we assume the message set is the integers between 1 and 2^{nR} . $A_{\varepsilon, Pr_{X,Y}}^n$ denotes the set of joint typical sequences for a given joint distribution $\Pr(X, Y)$.

- (a). Codebook Generation: Sample 2^{nR} sequences, each of length n , in an i.i.d. fashion from $\Pr(X)$. We use these sequences to construct the codebook and denote it as \mathcal{C} . Both the sender and the receiver knows the codebook \mathcal{C} .

The i th codeword in the codebook \mathcal{C} is denoted by X_i^n . With this notation, the probability of using a certain codebook c , $\Pr(\mathcal{C} = c)$, can be computed as:

$$\Pr(\mathcal{C} = c) \quad (2)$$

$$= \Pr(\{X_j^n = x_j^n, j = 1 \dots 2^{nR}\}) \quad (3)$$

$$= \prod_{j=1}^{2^{nR}} \Pr(X_j^n = x_j^n) \quad (4)$$

- (b). Encoder: We denote the encoder for a codebook \mathcal{C} as f :

$$X^n = f_{\mathcal{C}}(W) \quad (5)$$

Given \mathcal{C} and W , we can find out X^n .

Given \mathcal{C} and X^n , we can find the corresponding W .

- (c). Decoder: The decoder is denoted by g :

$$\hat{W} = g_{\mathcal{C}}(Y^n) \quad (6)$$

where $g_{\mathcal{C}}$ is defined as:

- (a) If $\exists X^n \in \mathcal{C}$, such that $X^n, Y^n \in A_{\varepsilon, Pr_{X,Y}}^n$, then g outputs $f_{\mathcal{C}}^{-1}(X^n)$.
(b) Otherwise, g outputs the first message.

(d). Decoding error probability computation:

Define the binary random variable E such that

$$E = \begin{cases} 1 & \text{if } g_{\mathcal{C}}(Y^n) \neq W \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Then $\Pr(E = 1|\mathcal{C} = c)$ is the probability of decoding error, which is difficult to compute. Hence we compute $\Pr(E = 1)$ instead. From the law of total probability, $\Pr(E = 1)$ can be written as:

$$\begin{aligned} \Pr(E = 1) &= \sum_c \Pr(\mathcal{C} = c) \sum_w \Pr(W = w) \sum_{y^n, x^n} \Pr(Y^n = y^n, X^n = x^n | W = w, \mathcal{C} = c) \\ &\quad \Pr(E = 1 | W = w, \mathcal{C} = c, Y^n = y^n, X^n = x^n) \end{aligned} \quad (8)$$

Since \mathcal{C} and W determines the codebook X^n , (8) can be re-written as:

$$\begin{aligned} &\sum_c \Pr(\mathcal{C} = c) \sum_w \Pr(W = w) \sum_{y^n} \Pr(Y^n = y^n | W = w, \mathcal{C} = c) \\ &\quad \Pr(E = 1 | W = w, \mathcal{C} = c, Y^n = y^n) \end{aligned} \quad (9)$$

Define binary random variable $E'_j, j = 1, \dots, 2^{nR}$ such that

$$E'_j = \begin{cases} 1 & \text{if } \{f_{\mathcal{C}}(j), Y^n\} \in A_{\epsilon, P_{X,Y}}^n \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

From the definition of the decoder, it is easy to see:

$$\{E = 1\} \subseteq \{E'_W = 0\} \cup \bigcup_{j=1 \dots 2^{nR}, j \neq W} \{E'_j = 1\} \quad (11)$$

Therefore (9) is upper bounded by

$$\begin{aligned} &\sum_c \Pr(\mathcal{C} = c) \sum_w \Pr(W = w) \sum_{y^n} \Pr(Y^n = y^n | W = w, \mathcal{C} = c) \\ &\quad (\Pr(E'_W = 0 | W = w, \mathcal{C} = c, Y^n = y^n) \\ &\quad + \sum_{j \neq w} \Pr(E'_j = 1 | W = w, \mathcal{C} = c, Y^n = y^n)) \end{aligned} \quad (12)$$

The first term in (12) can be re-written as

$$\begin{aligned} &\sum_w \Pr(W = w) \sum_c \Pr(\mathcal{C} = c) \sum_{y^n} \Pr(Y^n = y^n | W = w, \mathcal{C} = c) \\ &\quad \Pr(E'_W = 0 | W = w, \mathcal{C} = c, Y^n = y^n) \end{aligned} \quad (13)$$

Applying (2)-(4), we find (13) equals:

$$\begin{aligned} &\sum_w \Pr(W = w) \sum_{x_j^n, j=1 \dots 2^{nR}} \Pr(\{X_j^n = x_j^n, j = 1 \dots 2^{nR}\}) \\ &\quad \sum_{y^n} \Pr(Y^n = y^n | W = w, \{X_j^n = x_j^n, j = 1 \dots 2^{nR}\}) \\ &\quad \Pr(E'_W = 0 | W = w, Y^n = y^n, \{X_j^n = x_j^n, j = 1 \dots 2^{nR}\}) \end{aligned} \quad (14)$$

Since

$$\Pr(Y^n = y^n | W = w, \{X_j^n = x_j^n, j = 1 \dots 2^{nR}\}) = \Pr(Y^n = y^n | X_w^n = x_w^n) \quad (15)$$

and the value of E'_W only depends on X_W^n and Y^n , (14) can be re-written as:

$$\begin{aligned} & \sum_w \Pr(W = w) \sum_{x_j^n, j=1 \dots 2^{nR}} \Pr(\{X_j^n = x_j^n, j = 1 \dots 2^{nR}, j \neq w\}) \Pr(X_w^n = x_w^n) \\ & \sum_{y^n} \Pr(Y^n = y^n | X_w^n = x_w^n) \Pr(E'_w = 0 | Y^n = y^n, X_w^n = x_w^n) \end{aligned} \quad (16)$$

$$\begin{aligned} & = \sum_w \Pr(W = w) \sum_{x_j^n, j=1 \dots 2^{nR}, j \neq w} \Pr(\{X_j^n = x_j^n, j = 1 \dots 2^{nR}, j \neq w\}) \sum_{x^n} \Pr(X_w^n = x^n) \\ & \sum_{y^n} \Pr(Y^n = y^n | X_w^n = x^n) \Pr(E'_w = 0 | Y^n = y^n, X_w^n = x^n) \end{aligned} \quad (17)$$

$$\begin{aligned} & = \sum_w \Pr(W = w) \sum_{x^n} \Pr(X_w^n = x^n) \\ & \sum_{y^n} \Pr(Y^n = y^n | X_w^n = x^n) \Pr(E'_w = 0 | Y^n = y^n, X_w^n = x^n) \end{aligned} \quad (18)$$

$$= \sum_w \Pr(W = w) \sum_{x^n, y^n} \Pr(X_w^n = x^n, Y^n = y^n) \Pr(E'_w = 0 | Y^n = y^n, X_w^n = x^n) \quad (19)$$

$$= \sum_w \Pr(W = w) \sum_{x^n, y^n} \Pr(X^n = x^n, Y^n = y^n) \Pr(\{X^n, Y^n\} \notin A_{\varepsilon, P_{X,Y}}^n | Y^n = y^n, X^n = x^n) \quad (20)$$

In (17) we simply change the notation from x_w^n to x^n . In (20) we replace $\Pr(X_w^n = x^n, Y^n = y^n)$ with $\Pr(X^n = x^n, Y^n = y^n)$ since $\Pr(X_w^n = x^n)$ is the same for all w . $E'_w = 0$ is replaced accordingly with the definition of this event. Since the term inside the second sum in (20) no longer includes W , equation (20) can be written as:

$$\sum_{x^n, y^n} \Pr(X^n = x^n, Y^n = y^n) \Pr(\{X^n, Y^n\} \notin A_{\varepsilon, P_{X,Y}}^n | Y^n = y^n, X^n = x^n) \leq n\varepsilon_1 \quad (21)$$

Here $\varepsilon_1 \geq 0$ and $\lim_{n \rightarrow \infty} \varepsilon_1 = 0$. The inequality is because: Since X^n, Y^n are sampled in an i.i.d. fashion from a joint distribution $\Pr(X, Y)$, the probability that they are not jointly typical divided by n should vanish as n goes to ∞ .

The second term in (12) can be re-written as

$$\begin{aligned} & \sum_w \Pr(W = w) \sum_{\substack{j \neq w \\ j=1 \dots 2^{nR}}} \sum_c \Pr(\mathcal{C} = c) \sum_{y^n} \Pr(Y^n = y^n | W = w, \mathcal{C} = c) \\ & \Pr(E'_j = 1 | W = w, \mathcal{C} = c, Y^n = y^n) \end{aligned} \quad (22)$$

$$\begin{aligned} & = \sum_w \Pr(W = w) \sum_{\substack{j \neq w \\ j=1 \dots 2^{nR}}} \sum_{x_k^n, k=1 \dots 2^{nR}} \Pr(\{X_k^n = x_k^n, k = 1 \dots 2^{nR}\}) \\ & \sum_{y^n} \Pr(Y^n = y^n | X_w^n = x_w^n) \Pr(E'_j = 1 | X_j^n = x_j^n, Y^n = y^n) \end{aligned} \quad (23)$$

Since each codeword is independently generated, (23) can be written as:

$$\sum_w \Pr(W = w) \sum_{\substack{j \neq w \\ j=1 \dots 2^{nR}}} \sum_{x_j^n} \Pr(X_j^n = x_j^n) \sum_{y^n} \Pr(Y^n = y^n) \Pr(E'_j = 1 | X_j^n = x_j^n, Y^n = y^n) \quad (24)$$

$$\leq \sum_w \Pr(W = w) \sum_{\substack{j \neq w \\ j=1 \dots 2^{nR}}} \sum_{x_j^n} \sum_{y^n} \Pr(X_j^n = x_j^n) \Pr(Y^n = y^n) \Pr(E'_j = 1 | X_j^n = x_j^n, Y^n = y^n) \quad (25)$$

For the term inside the sum, we have:

$$\sum_{x_j^n} \sum_{y^n} \Pr(X_j^n = x_j^n) \Pr(Y^n = y^n) \Pr(E'_j = 1 | X_j^n = x_j^n, Y^n = y^n) \leq 2^{-n(I(X;Y) - \varepsilon_2)} \quad (26)$$

where $\varepsilon_2 \geq 0$ and $\lim_{n \rightarrow \infty} \varepsilon_2 = 0$. Recall that this is because there are at most $2^{n(H(X,Y) + \varepsilon_2/3)}$ pairs of $\{x_j^n, y^n\}$ for which $E'_j = 1$ and for each sequence $\Pr(X_j^n = x_j^n)$ is bounded by $2^{-n(H(X) - \varepsilon_2/3)}$, $\Pr(Y^n = y^n)$ is bounded by $2^{-n(H(Y) - \varepsilon_2/3)}$.

Applying this result, we find (25) is upper bounded by:

$$\sum_w \Pr(W = w) \sum_{\substack{j \neq w \\ j=1 \dots 2^{nR}}} 2^{-n(I(X;Y) - \varepsilon_2)} \quad (27)$$

$$\leq \sum_w \Pr(W = w) 2^{nR} 2^{-n(I(X;Y) - \varepsilon_2)} \quad (28)$$

$$= 2^{nR} 2^{-n(I(X;Y) - \varepsilon_2)} \quad (29)$$

$$= 2^{-n(I(X;Y) - R - \varepsilon_2)} \quad (30)$$

Therefore we conclude

$$0 \leq \Pr(E = 1) \leq n\varepsilon_1 + 2^{-n(I(X;Y) - R - \varepsilon_2)} \quad (31)$$

This means

$$\lim_{n \rightarrow \infty} \Pr(E = 1) = 0 \quad (32)$$

Since

$$\Pr(E = 1) = \sum_c \Pr(E = 1 | \mathcal{C} = c) \quad (33)$$

this implies there must exist at least one codebook c^* , such that

$$\lim_{n \rightarrow \infty} \Pr(E = 1 | \mathcal{C} = c^*) = 0 \quad (34)$$

Hence we have proved the achievability part of the theorem.