Advanced Communication Systems

ELEN90051 (Lecturer Margreta Kuijper)

Channel capacity of discrete channels

1st Semester 2018

Written by Margreta Kuijper; see Chapters 6 of "Digital Communications" by Proakis & Salehi, 2008

All scanned tables and text are from the textbook "Digital Communications" by Proakis and Salehi, 2008

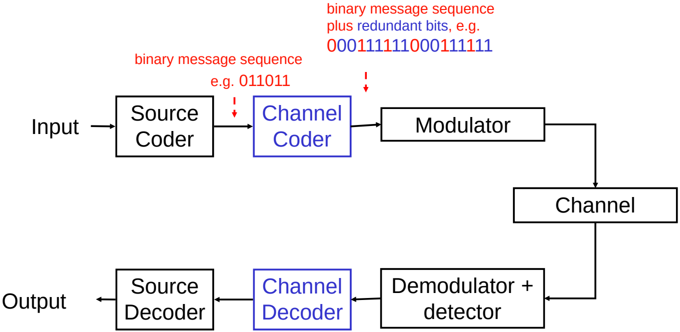
IN LAST CHAPTER:

- The entropy $H(X)$ as a notion that measures the uncertainty in a discrete source X.
- Important in Shannon's source coding theorem

IN THIS CHAPTER:

- The received signal y is different from the transmitted signal x because of channel noise.
- We need a notion that measures the channel quality
- This will be channel capacity
- Important in forthcoming Shannon's channel coding theorem

Channel coding put into context:



In this "repetition channel code" example the **rate** of the code equals 1/3 because for every single message bit there are 3 coded bits to be modulated and sent through the channel

Let's look at the channel as a **discrete channel** by including "modulation" as well as "demodulation and detection", see next figure:

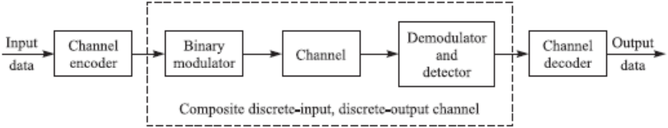


FIGURE 6.5–1
A composite discrete input, discrete output channel formed by including the modulator and the demodulator as part of the channel.

- Thus we are only interested in input X and output Y being discrete sources with alphabets $\mathcal{X} = \{x_0, \ldots, x_{q-1}\}$ and $\mathcal{Y} = \{y_0, \ldots, y_{Q-1}\}$, respectively.

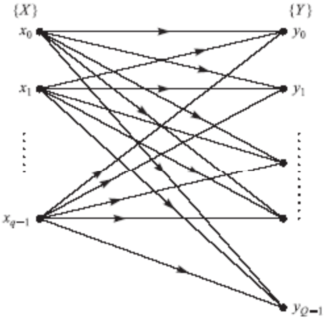


FIGURE 6.5–3
Discrete memoryless channel.

Example

Consider binary antipodal modulation for an AWGN continuous channel with corresponding demodulator plus detector. Then the bit error probability equals $p = Q(\sqrt{\frac{2E_b}{N_0}})$. This can be modeled as a discrete channel (called a binary symmetric channel BSC):

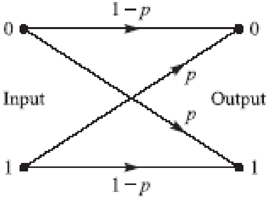


FIGURE 6.5–2
Binary symmetric channel.

- Each channel use is assumed to be independent from the previous, that is, the channel output at time t depends only on the channel input at time t. We call such a channel a **discrete memoryless channel (DMC)**.
- A DMC is completely characterized by its conditional probability matrix $P = (p_{ij})$ where $p_{ij} = P(Y = y_i | X = x_j)$.

Example 1: BSC=Binary Symmetric Channel

- $\mathcal{X} = \mathcal{Y} = \{0, 1\}$
- $p_{10} = p_{01}$

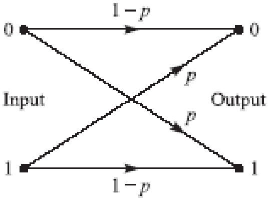
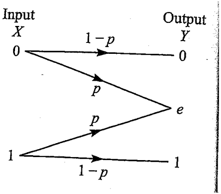


FIGURE 6.5–2
Binary symmetric channel.

Example 2: BEC=Binary Erasure Channel

$\mathcal{X} = \{0, 1\}$; $\mathcal{Y} = \{0, 1, e\}$



How to achieve perfectly reliable communication over a DMC?

A NAIVE APPROACH: THE $(n,1)$ REPETITION CODE

Suppose $\mathcal{X} = \{A, B\}$; $\mathcal{Y} = \{0, 1\}$ with cross-over probability $p < 0.5$.

- channel encoder maps into codewords of length n as:

$$\begin{aligned} A &\mapsto 00\cdots 0 \\ B &\mapsto 11\cdots 1 \end{aligned}$$

- channel decoder uses majority vote: consider $N_0 := \#$ 0's in received word and decide as follows:

$$\begin{aligned} N_0 > \frac{n}{2} &\Rightarrow A \\ N_0 \leq \frac{n}{2} &\Rightarrow B \end{aligned}$$

- Then $P_e = P(error|A)P(A) + P(error|B)P(B) = P(N_0 > \frac{n}{2}|B)$ approaches 0 as $n \to \infty$ (**Quiz**: Derive an approximate expression for P_e in terms of n, p and the Q-function)
- Thus reliable communication is achieved asymptotically

- But it comes at a cost since the code rate $1/n$ (in message bits per channel use) also approaches 0 asymptotically

Question:

Can we do better?

Shannon observed that there exist sequences of codes (with increasing blocklength n) that achieve reliable communication asymptotically as $n \to \infty$ but also have an asymptotic rate **strictly** larger than zero.

What is the maximum value of this rate?

Let X be a random variable with values in $\{x_1, \ldots, x_q\}$ with corresponding probabilities $p_1, p_2, \ldots, p_q$

RECALL:

The **entropy** of X is defined as

$$H(X) := -\sum_{i=1}^{q} p_i \log_2 p_i$$

The entropy $H(X)$ expresses "the amount of uncertainty" in X...

Now let Y be a random variable with values in $\{y_1, \ldots, y_Q\}$; let XY denote the vector-valued random variable $\begin{bmatrix} X \\ Y \end{bmatrix}$. Then

$$H(XY) := -\sum_{x,y} P(x,y) \log_2 P(x,y).$$

We call $H(XY)$ the **joint entropy** of X and Y.

Tute question 5.1

Let X and Y be random variables. Show that the joint entropy $H(XY)$ satisfies

$$H(XY) \leq H(X) + H(Y)$$

Hint: first show that

$$H(X) = -\sum_{x,y} P(x,y) \log_2 P(x),$$

and then use the inequality $\ln w \leq w - 1$.

Tute question 5.2

When does equality hold in the previous tute question?

The remaining uncertainty in X after observing $Y = y_i$ is

$$H(X|Y = y_i) = -\sum_{j=1}^{q} P(x_j|y_i)\log_2 P(x_j|y_i)$$

Note that $H(X|Y = y_i)$ is a function of y_i.

DEFINITION

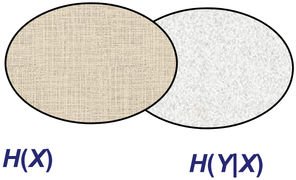
The **conditional entropy** of X given Y is defined as the expected value of the above expression:

$$\begin{aligned} H(X|Y) &:= \sum_{i=1}^{Q} H(X|Y = y_i)P(y_i) \\ &:= -\sum_{i=1}^{Q}\left(\sum_{j=1}^{q} P(x_j|y_i)\log_2 P(x_j|y_i)\right)P(y_i) \\ &= -\sum_{i=1}^{Q}\sum_{j=1}^{q} P(x_j, y_i)\log_2 P(x_j|y_i) \end{aligned}$$

Tute question 5.3

Let X and Y be random variables. Show that the joint entropy $H(XY)$ equals

$$H(XY) = H(X) + H(Y|X).$$



The above equality is called the **chain rule for entropies**; note that it is symmetric: $H(XY) = H(Y) + H(X|Y)$ also holds.

TUTE QUESTION 5.4

Let X and Y be random variables. Show that

$$H(X|Y) \leq H(X).$$

When does equality hold?

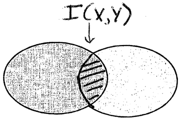
We saw in the previous tute question that

$$H(X) - H(X|Y) \geq 0.$$

This quantity has a name, namely the **mutual information** between X and Y:

$$I(X,Y) := H(X) - H(X|Y)$$

It can be interpeted as the reduction in uncertainty about X provided by observing Y.



TUTE QUESTION 5.5

Let X and Y be two binary random variables, distributed according to the joint distributions

$$P(X = Y = 0) = P(X = 0, Y = 1) = P(X = Y = 1) = \frac{1}{3}.$$

Compute $H(X), H(Y), H(X|Y), H(Y|X), H(XY)$ and $I(X, Y)$.

Let's now again consider a discrete memoryless channel

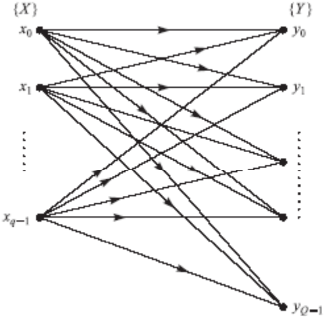


FIGURE 6.5–3
Discrete memoryless channel.

Definition

The **capacity** of the channel is given by

$$C = \max I(X, Y) \text{ bits/channel use},$$

where the maximum is taken over all possible distributions on the channel input X.

Example 1: BSC=Binary Symmetric Channel

$$
\begin{aligned}
C &= 1 + p\log_2 p + (1-p)\log_2(1-p) \\
&= 1 - H_b(p) \text{ bits/channel use,}
\end{aligned}
$$

where H_b is the binary entropy function. (**Quiz**: Derive this formula)

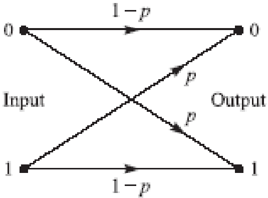


FIGURE 6.5–2
Binary symmetric channel.

Note that

- $C = 1$ for $p = 0$ (no uncertainty in transmission)
- $C = 0$ for $p = 0.5$ (maximum uncertainty in transmission)

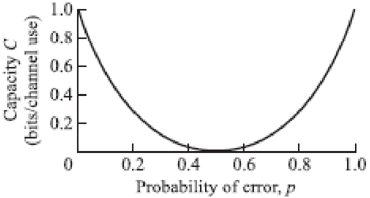
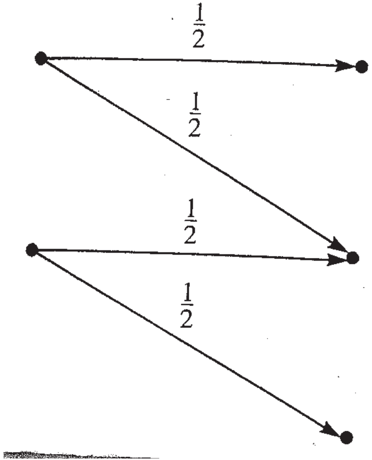


FIGURE 6.5–4
The capacity of a BSC.

Tute question 5.6

Consider a discrete memoryless channel given by the figure below. Determine the channel's capacity.

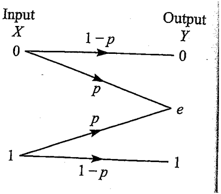


Example 2: BEC=Binary Erasure Channel

$$C = 1 - p \text{ bits/channel use}$$

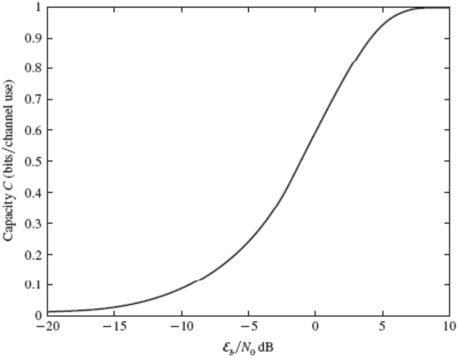
Quiz

Derive the above formula.



Example

Consider binary antipodal modulation for an AWGN continuous channel with corresponding demodulator plus detector. Then the bit error probability equals $p = Q(\sqrt{\frac{2E_b}{N_0}})$. As we saw before, this can be modeled as a discrete channel; its capacity is given by the following figure:



Consider a discrete memoryless channel with input X and output Y and capacity C. The following result is the second main result of Shannon's 1948 paper:

CHANNEL CODING THEOREM: Communication with arbitrarily small error probability is possible if the transmission rate R satisfies

$$R < C.$$

Furthermore, if $R > C$ then the error probability is bounded away from zero.

Note the analogy with the capacity of a water pipe: if we pump water at a rate larger than the pipe's capacity then water will be lost. Similarly, if we try to communicate at a rate $> C$ then information will be lost

So roughly speaking we conclude from the above theorem the following more practical statement:

> *When employing n channel uses, it is possible to reliably send nC data bits. And forget about reliable transmission of more than nC data bits.*

The theorem sets a fundamental limit, but doesn't tell us how to find the best channel code. The fundamental limit serves as a yardstick to measure performance of channel codes to come....

AWGN channel, BPSK coherent demodulator, ML detector. Recall the plot of the bit SNR versus the Binary Symmetric Channel capacity, using the channel capacity formula for the BSC (so here we have discrete-time discrete symbol values coming out of the detector):

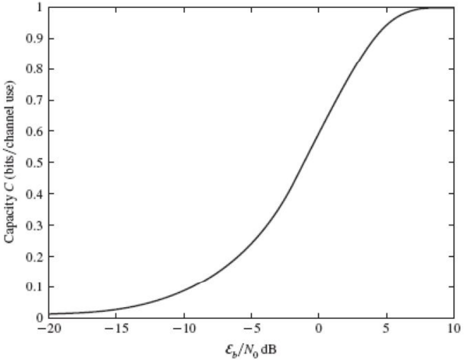


FIGURE 6.5–5
The capacity plot versus SNR per bit.

AWGN channel, BPSK coherent demodulator. See textbook, with discrete-time complex values coming out of the demodulator, the bit SNR versus capacity plot is:

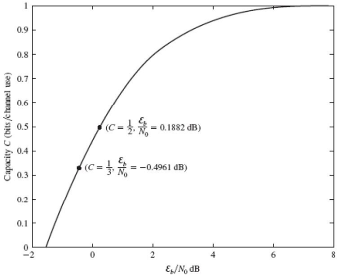


FIGURE 6.5–6
The capacity of binary input AWGN channel.

According to Shannon's channel coding theorem, to achieve error-free communication with (for example) a rate 1/2 code, the minimum required SNR equals 0.188 dB. This minimum value is referred to as the Shannon limit at rate 1/2. Some recently developed channel codes come close to this limit. (for more details, see "Apparent contradiction in the Shannon Limit", p. 533 Sklar book)

AWGN channel, no fixed modulation scheme. Then, with channel bandwidth W and average power P at the receiver:

SHANNON-HARTLEY THEOREM:

$$C = W \log_2(1 + \frac{E_b}{N_0}) = W \log_2(1 + \frac{P}{N_0 W}) \text{ bits/sec}$$

No error-free communication is possible for SNR values < -1.6 dB. This is the famous Shannon limit.