Chapter 7
Channel Capacity and Coding
7.1 Channel models and channel capacity

7.1.1 Channel models
- Binary symmetric channel
- Discrete memoryless channels
- Discrete-input, continuous-output channel
- Waveform channels

7.1.2 Channel capacity
7.1.1 Channel Models

- Binary symmetric channel (BSC)
  - If the channel noise and other disturbances cause statistically independent errors in the transmitted binary sequence with average probability $p$, then

$$
P(Y = 0 \mid X = 1) = P(Y = 1 \mid X = 0) = p
$$

$$
P(Y = 1 \mid X = 1) = P(Y = 0 \mid X = 0) = 1 - p
$$
7.1.1 Channel Models

- Discrete memoryless channels (DMC)
  - BSC is a special case of a more general discrete-input, discrete-output channel.
  - Output symbols from the channel encoder are $q$-ary symbols, i.e., $X = \{x_0, x_1, \ldots, x_{q-1}\}$.
  - Output of the detector consists of $Q$-ary symbols, where $Q \geq M = 2^q$.
  - If the channel and modulation are memoryless, we have a set of $qQ$ conditional probabilities:
    \[
    P(Y = y_i \mid X = x_j) = P(y_i \mid x_j)
    \]
    where $i = 0, 1, \ldots, Q-1$ and $j = 0, 1, \ldots, q-1$.
  - Such a channel is called a discrete memoryless channel (DMC).
7.1.1 Channel Models

- **Discrete memoryless channels (DMC)**
  - Input: \( u_1, u_2, \ldots, u_n \)
  - Output: \( v_1, v_2, \ldots, v_n \)
  - The conditional probability is given by:

\[
P(Y_1 = v_1, \ Y_2 = v_2, \ldots, \ Y_n = v_n | X = u_1, \ldots, \ X = u_n)
= \prod_{k=1}^{n} P(Y = v_k | X = u_k)
\]

- In general, the conditional probabilities \( P(y_j | x_i) \) can be arranged in the matrix form \( P = [p_{ij}] \), called *probability transition matrix*. 

Discrete \( q \)-ary input, \( Q \)-ary output channel
7.1.1 Channel Models

- **Discrete-input, continuous-output channel**
  - Discrete input alphabet $X = \{x_0, x_1, \ldots, x_{q-1}\}$.
  - Output of the detector is unquantized ($Q = \infty$).
  - The most important channel of this type is the additive white Gaussian noise (AWGN) channel, for which
    \[
    Y = X + G
    \]
    where $G$ is a zero-mean Gaussian random variable with variance $\sigma^2$ and $X = x_k$, $k = 0, 1, \ldots, q-1$.

- \[
p(y \mid X = x_k) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-x_k)^2}{2\sigma^2}}
\]

- \[
p(y_1, y_2, \ldots, y_n \mid X_1 = u_1, X_2 = u_2, \ldots, X_n = u_n) = \prod_{i=1}^{n} p(y_i \mid X_i = u_i)
\]
7.1.1 Channel Models

Waveform channels

- Assume that a channel has a given bandwidth $W$, with ideal frequency response $C(f)=1$ within the bandwidth $W$, and the signal at its output is corrupted by AWGN: $y(t)=x(t)+n(t)$.
- Expand $y(t)$, $x(t)$, and $n(t)$ into a complete set of orthonormal functions:

$$
y(t) = \sum_i y_i f_i(t), \quad x(t) = \sum_i x_i f_i(t), \quad n(t) = \sum_i n_i f_i(t).
$$

$$
y_i = \int_0^T y(t) f_i^*(t) \, dt = \int_0^T \left[ x(t) + n(t) \right] f_i^*(t) \, dt = x_i + n_i
$$

$$
\int_0^T f_i(t) f_j^*(t) \, dt = \delta_{ij} = \begin{cases} 
1 & (i = j) \\
0 & (i \neq j)
\end{cases}
$$
7.1.1 Channel Models

Waveform channels

- Since \( y_i = x_i + n_i \), it follows that:

\[
p(y_i | x_i) = \frac{1}{\sqrt{2\pi \sigma_i}} e^{-\frac{(y_i - x_i)^2}{2\sigma_i^2}}, \quad i = 1, 2, \ldots
\]

- Since the functions \( \{f_i(t)\} \) are orthonormal, it follows that the \( \{n_i\} \) are uncorrelated.

- Since they are Gaussian, they are also statistically independent:

\[
p(y_1, y_2, \ldots, y_N | x_1, x_2, \ldots, x_N) = \prod_{i=1}^{N} p(y_i | x_i)
\]

- Samples of \( x(t) \) and \( y(t) \) may be taken at the Nyquist rate of \( 2W \) samples per second. Thus, in a time interval of length \( T \), there are \( N = 2WT \) samples.
7.1.2 Channel Capacity

Consider a DMC having an input alphabet \( X = \{x_0, x_1, \ldots, x_{q-1}\} \), an output alphabet \( Y = \{y_0, y_1, \ldots, y_{Q-1}\} \), and the set of transition probabilities \( P(y_i|x_j) \).

The mutual information provided about the event \( X = x_j \) by the occurrence of the event \( Y = y_i \) is \( \log \frac{P(y_i|x_j)}{P(y_i)} \), where

\[
P(y_i) = P(Y = y_i) = \sum_{k=0}^{q-1} P(x_k)P(y_i|x_k)
\]

Hence, the average mutual information provided by the output \( Y \) about the input \( X \) is:

\[
I(X; Y) = \sum_{j=0}^{q-1} \sum_{i=0}^{Q-1} P(x_j)P(y_i|x_j) \log \frac{P(y_i|x_j)}{P(y_i)}
\]
The value of $I(X;Y)$ maximized over the set of input symbol probabilities $P(x_j)$ is a quantity that depends only on the characteristics of the DMC through the conditional probabilities $P(y_i|x_j)$. This quantity is called the *capacity* of the channel and is denoted by $C$:

$$C = \max_{P(x_j)} I(X; Y)$$

$$= \max_{P(x_j)} \sum_{j=0}^{q-1} \sum_{i=0}^{Q-1} P(x_j)P(y_i | x_j) \log \frac{P(y_i | x_j)}{P(y_i)}$$

The maximization of $I(X;Y)$ is performed under the constraints that

$$P(x_j) \geq 0 \quad \text{and} \quad \sum_{j=0}^{q-1} P(x_j) = 1.$$
7.1.2 Channel Capacity

- Example 7.1-1 BSC with transition probabilities $P(0|1)=P(1|0)=p$.
  - The average mutual information is maximized when the input probabilities $P(0)=P(1)=\frac{1}{2}$.
  - The capacity of the BSC is
    \[ C = p \log 2p + (1-p)\log 2(1-p) = 1 - H(p) \]
    where $H(p)$ is the binary entropy function.
7.1.2 Channel Capacity

Consider the discrete-time AWGN memoryless channel described by

\[ p(y \mid X = x_k) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(y-x_k)^2}{2\sigma^2}} \]

The capacity of this channel in bits per channel use is the maximum average mutual information between the discrete input \( X = \{x_0, x_1, \ldots, x_{q-1}\} \) and the output \( Y = \{\infty, -\infty\} \):

\[
C = \max_{P(x_i)} \sum_{i=0}^{q-1} \int_{-\infty}^{\infty} p(y \mid x_i) P(x_i) \log_2 \frac{P(y \mid x_i)}{P(y)} dy
\]

where

\[ p(y) = \sum_{k=0}^{q-1} p(y \mid x_k) P(x_k) \]
7.1.2 Channel Capacity

Example 7.1-2. Consider a binary-input AWGN memoryless channel with possible inputs $X=A$ and $X=-A$.

The average mutual information $I(X;Y)$ is maximized when the input probabilities are $P(X=A)=P(X=-A)=\frac{1}{2}$.

$$C = \frac{1}{2} \int_{-\infty}^{\infty} p(y | A) \log_2 \frac{p(y | A)}{p(y)} dy$$

$$+ \frac{1}{2} \int_{-\infty}^{\infty} p(y | -A) \log_2 \frac{p(y | -A)}{p(y)} dy$$
7.1.2 Channel Capacity

- It is not always the case to obtain the channel capacity by assuming that the input symbols are equally probable.
- Nothing can be said in general about the input probability assignment that maximizes the average mutual information.
- It can be shown that the necessary and sufficient conditions for the set of input probabilities \( \{P(x_j)\} \) to maximize \( I(X;Y) \) and to achieve capacity on a DMC are:

\[
I(x_j; Y) = C \quad \text{for all } j \text{ with } P(x_j) > 0
\]

\[
I(x_j; Y) \leq C \quad \text{for all } j \text{ with } P(x_j) = 0
\]

where \( C \) is the capacity of the channel and

\[
I(x_j; Y) = \sum_{i=0}^{Q-1} P(y_i \mid x_j) \log \frac{P(y_i \mid x_j)}{P(y_i)}
\]
7.1.2 Channel Capacity

Consider a band-limited waveform channel with AWGN.

The capacity of the channel per unit time has been defined by Shannon (1948) as

\[
C = \lim_{T \to \infty} \max_{p(x)} \frac{1}{T} I(X; Y)
\]

Alternatively, we may use the samples or the coefficients \(\{y_i\}, \{x_i\}, \text{and} \{n_i\}\) in the series expansions of \(y(t), x(t), \text{and} n(t)\) to determine the average mutual informaiton between \(x_N = [x_1 \ x_2 \ldots x_N]\) and \(y_N = [y_1 \ y_2 \ldots y_N]\), where \(N=2WT, y_i = x_i + n_i\).

\[
I(X_N; Y_N) = \int_{x_N} \ldots \int_{y_N} \ldots p(y_N | x_N)p(x_N)\log \frac{p(y_N | x_N)}{p(y_N)} \, dx_N \, dy_N
\]

\[
= \sum_{i=1}^{N} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(y_i | x_i)p(x_i)\log \frac{p(y_i | x_i)}{p(y_i)} \, dy_i \, dx_i
\]
7.1.2 Channel Capacity

where

\[ p(y_i | x_i) = \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(y_i-x_i)^2}{N_0}} \]

- The maximum of \( I(X;Y) \) over the input PDFs \( p(x_i) \) is obtained when the \( \{x_i\} \) are statistically independent zero-mean Gaussian random variables, i.e.,

\[ p(x_i) = \frac{1}{\sqrt{2\pi \sigma_x}} e^{-\frac{x_i^2}{2\sigma_x^2}} \]

- \( \max_{p(x)} I(X_N; Y_N) = \sum_{i=1}^{N} \frac{1}{2} \log \left( 1 + \frac{2\sigma_x^2}{N_0} \right) = \frac{1}{2} N \log \left( 1 + \frac{2\sigma_x^2}{N_0} \right) \)

\[ = W T \log \left( 1 + \frac{2\sigma_x^2}{N_0} \right) \]
7.1.2 Channel Capacity

If we put a constraint on the average power in $x(t)$, i.e.,

$$P_{av} = \frac{1}{T} \int_0^T E\left[ x^2(t) \right] dt = \frac{1}{T} \sum_{i=1}^{N} E\left( x_i^2 \right) = \frac{N \sigma_x^2}{T}$$

$$\sigma_x^2 = \frac{TP_{av}}{N} = \frac{P_{av}}{2W}$$

$$\max_{p(x)} I\left( X_N ; Y_N \right) = WT \log \left( 1 + \frac{P_{av}}{WN_0} \right)$$

Dividing both sides by $T$ and we can obtain the capacity of the band-limited AWGN waveform channel with a band-limited and average power-limited input:

$$C = W \log \left( 1 + \frac{P_{av}}{WN_0} \right)$$
7.1.2 Channel Capacity

- Normalized channel capacity as a function of SNR for band-limited AWGN channel
- Channel capacity as a function of bandwidth with a fixed transmitted average power
7.1.2 Channel Capacity

- Note that as $W$ approaches infinity, the capacity of the channel approaches the asymptotic value
  
  \[ C_\infty = \frac{P_{av}}{N_0} \log_2 e = \frac{P_{av}}{N_0 \ln 2} \text{ bits/s} \]

- Since $P_{av}$ represents the average transmitted power and $C$ is the rate in bits/s, it follows that
  
  \[ P_{av} = C \varepsilon_b \]

- Hence, we have
  
  \[ \frac{C}{W} = \log_2 \left( 1 + \frac{C \varepsilon_b}{W N_0} \right) \]

- Consequently
  
  \[ \frac{\varepsilon_b}{N_0} = \frac{2^{C/W} - 1}{C/W} \]
7.1.2 Channel Capacity

- When $C/W = 1$, $\varepsilon_b/N_0 = 1$ (0 dB).

- When $C/W \to \infty$,
  
  $$\frac{\varepsilon_b}{N_0} \approx \frac{2^{C/W}}{C/W} \approx \exp \left( \frac{C}{W} \ln 2 - \ln \frac{C}{W} \right)$$

- When $C/W \to 0$
  
  $$\frac{\varepsilon_b}{N_0} = \lim_{C/W \to 0} \frac{2^{C/W} - 1}{C/W} = \ln 2$$
7.1.2 Channel Capacity

- The channel capacity formulas serve as upper limits on the transmission rate for reliable communication over a noisy channel.

- Noisy channel coding theorem by Shannon (1948)
  - There exist channel codes (and decoders) that make it possible to achieve reliable communication, with as small an error probability as desired, if the transmission rate $R < C$, where $C$ is the channel capacity. If $R > C$, it is not possible to make the probability of error tend toward zero with any code.