6.441 Transmission of Information	Mar 14, 2006
Lecture 10	
Lecturer: Madhu Sudan	*Scribe: Srujan Linga*

1 Last lecture

- Universal Coding
- Lempel-Ziv Algorithm

2 Today

- Channel capacity
- Sample channels and their capacities
- AEP for channels

3 Communication System Overview

The block diagram in Fig.1 shows an overview of the communication system.

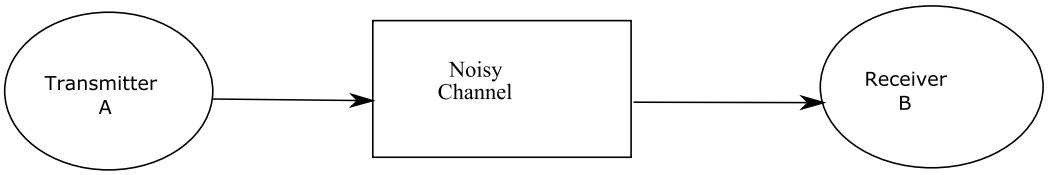

Figure 1: Block diagram of the communication system

The transfer of information from the transmitter to the receiver is a physical process and therefore is subject to noise and imperfections of the signalling process itself. Hence it is a property of physics that there are no perfect channels. A modified block diagram with an encoder and decoder introduced into the communication system is shown in Fig.2. Source symbols S from some finite alphabet Ω_S are mapped into some sequence of channel symbols X of alphabet Ω_X. The output sequence Y of the channel (input to the decoder) is random but has a distribution that depends on the input sequence X. From the output sequence, we attempt to recover the transmitted message.

3.1 Basic features of the channel

Considering a block of n channel symbols, let

1. $p_{\mathbf{X}^n}(\mathbf{x}^n)$ denote the probability distribution on an n-element sequence from Ω_X which is under the designer's control.
2. $p_{\mathbf{Y}^n|\mathbf{X}^n}(\mathbf{y}^n|\mathbf{x}^n)$ denote the probability that $\mathbf{y}^n$ is received given $\mathbf{x}^n$ was transmitted.
3. $\mathbf{P}_{\mathbf{Y}^n|\mathbf{X}^n}(\mathbf{y}^n|\mathbf{x}^n)$ denote an $|\Omega_X|^n \times |\Omega_Y|^n$ stochastic probability transmission matrix of the channel.

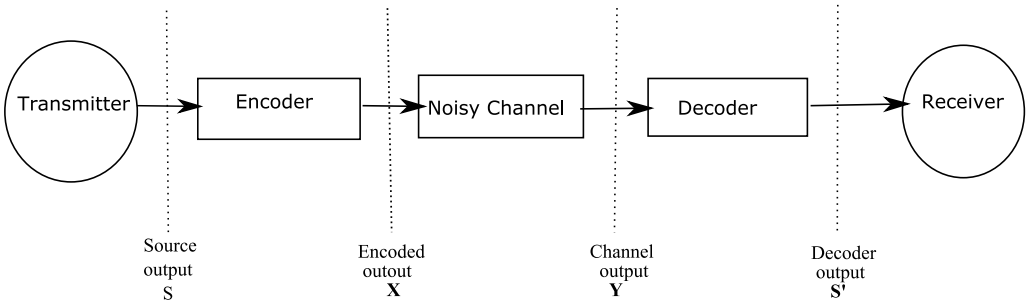

Figure 2: Block diagram with encoder and decoder

3.2 Channel Capacity

Let the capacity of the channel transmitting n-length sequences be given by $\mathbb{C}^{(n)}$, the n-fold capacity of the channel. Then,

$$\mathbb{C}^{(n)} = \max_{p_{\mathbf{X}^n}(\mathbf{x}^n)} (\frac{1}{n}\mathbf{I}(\mathbf{X}^n; \mathbf{Y}^n))$$

We would like to understand the behavior of $\mathbb{C}^{(n)}$ as $n \to \infty$.

3.3 Channel Classes

The classes of channels we want to consider today are Discrete Memoryless Channels (DMC's). These channels have the following properties:

1. Discreteness: Both Ω_X and Ω_Y are finite sets.
2. Memoryless: The behavior of the channel at time t is independent of the time and past inputs/outputs of the channel. More precisely,

$$p_{\mathbf{Y}^n|\mathbf{X}^n}(\mathbf{y}^n|\mathbf{x}^n) = \prod_{i=1}^{n} p_{\mathbf{Y}|\mathbf{X}}(y_i|x_i)$$

Roughly, outputs of memoryless channels capture the same idea as *i.i.d* outputs of the source.

3.4 Capacity of a Discrete Memoryless Channel

$$\mathbb{C}^{(n)} = \max_{p_{\mathbf{X}^n}(\mathbf{x}^n)} (\frac{1}{n}\mathbf{I}(\mathbf{X}^n;\mathbf{Y}^n)) \tag{1}$$

$$= \max_{p_{\mathbf{X}^n}(\mathbf{x}^n)} (\frac{1}{n}\mathbf{I}(\mathbf{Y}^n;\mathbf{X}^n)) \tag{2}$$

$$= \frac{1}{n} \max_{p_{\mathbf{X}^n}(\mathbf{x}^n)} (\mathbf{H}(\mathbf{Y}^n) - \mathbf{H}(\mathbf{Y}^n|\mathbf{X}^n)) \tag{3}$$

$$= \frac{1}{n} \max_{p_{\mathbf{X}^n}(\mathbf{x}^n)} (\sum_{i=1}^{n} \mathbf{H}(Y_i|Y_{i-1}, Y_{i-2}...Y_1) - \mathbf{H}(\mathbf{Y}^n|\mathbf{X}^n)) \tag{4}$$

$$\leq \frac{1}{n} \max_{p_{\mathbf{X}^n}(\mathbf{x}^n)} (\sum_{i=1}^{n} \mathbf{H}(Y_i) - \mathbf{H}(\mathbf{Y}^n|\mathbf{X}^n)) \tag{5}$$

$$= \frac{1}{n} \max_{p_{\mathbf{X}^n}(\mathbf{x}^n)} (\sum_{i=1}^{n} \mathbf{H}(Y_i) - \sum_{i=1}^{n} \mathbf{H}(Y_i|X_i)) \tag{6}$$

$$= \frac{1}{n} \max_{p_{\mathbf{X}^n}(\mathbf{x}^n)} (\sum_{i=1}^{n} (\mathbf{H}(Y_i) - \mathbf{H}(Y_i|X_i))) \tag{7}$$

$$= \max_{p_{\mathbf{X}^n}(\mathbf{x}^n)} (\sum_{i=1}^{n} \frac{\mathbf{I}(Y_i;X_i)}{n}) \tag{8}$$

where equation (2) arises due to symmetry of mutual information, equation (4) arises because of the chain rule of entropy, inequality (5) arises because conditioning only reduces the entropy and the property of discrete memoryless channel was used in (6) to expand $\mathbf{H}(\mathbf{Y}^n|\mathbf{X}^n)$. The result in (8) tells us that maximizing $\mathbf{I}(\mathbf{Y}^n;\mathbf{X}^n)$ over $p_{\mathbf{X}^n}(\mathbf{x}^n)$ is equivalent to maximizing $\mathbf{I}(Y_i;X_i)$ for each i from 1 to n. Therefore $p_{\mathbf{X}^n}(\mathbf{x}^n)$ may well be a product distribution, i.e. we may choose X_i to be $i.i.d$ random variables so that $p_{\mathbf{X}^n}(\mathbf{x}^n) = \prod_{i=1}^{n} p_{X_i}(x_i)$ and $p_{X_i}(x_i) = p_X(x)$ for each i. Now, since

$$\max_{p_X(x)} (\mathbf{I}(Y;X)) = \mathbb{C}^{(1)}$$

we have,

$$\mathbb{C}^{(n)} = \frac{\sum_{i=1}^{n} \mathbb{C}^{(i)}}{n} = \mathbb{C}^{(1)}$$

Therefore, n-fold usage of the channel is no greater than 1-fold usage in terms of the channel capacity. Note that if we choose $X_i's$ to be independent, $Y_i's$ are also independent due to the property of memoryless channel and hence the channel capacity can be achieved with equality. So, for independent X_i, $C^{(n)} = C^{(1)} = \max_{p_X(x)}(\mathbf{I}(Y;X))$.

4 Examples of Channel Capacity

4.1 Binary Erasure Channel (BEC)

Consider the Binary Erasure Channel shown in Fig.3. A BEC has two inputs 0 and 1 and a fraction p of the bits are erased. The receiver knows which of the bits have been erased. We calculate the capacity of the channel as follows,

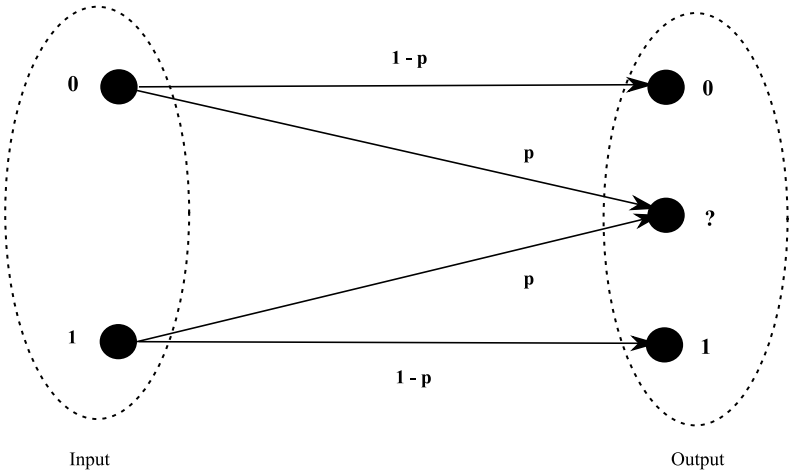

Figure 3: Binary Erasure Channel

$$
\begin{aligned}
\mathbb{C}^{(1)} &= \max_{p_X(x)} (\mathbf{I}(X;Y)) \\
&= \max_{p_X(x)} (\mathbf{H}(X) - \mathbf{H}(X|Y)) \\
&= \max_{p_X(x)} (\mathbf{H}(X) - \underbrace{\mathbf{H}(X|Y=?)}_{=\mathbf{H}(X)} \underbrace{Pr(Y=?)}_{=p}) \\
&= \max_{p_X(x)} (\mathbf{H}(X) - \mathbf{H}(X)p) \\
&= \max_{p_X(x)} (\mathbf{H}(X)(1-p)) \\
&= 1-p
\end{aligned}
$$

where the last equality was derived because maximum value of $H(X)$ is 1 bit. This result gives us the following intuition: *a)* One will be able to "push" through roughly $(1-p)$ bits every time unit; *b)* For every n bits one transmits, one would expect to see roughly $n(1-p)$ bits without erasures. This insight leads us to the following encoding scheme which asymptotically achieves zero error probability: If the encoder picks up source strings ($\mathbf{S}$) of length $n(1-p)$ and encodes them into channel symbols ($\mathbf{X}$) of length n, i.e. we add some form of redundancy to the source sequences, then by the above observations, the output sequence ($\mathbf{Y}$) has approximately $n(1-p)$ correct symbols which can then be decoded to $\mathbf{S}^{'}$ at the receiver. For this scheme it can be shown that, asymptotically, $Pr(\mathbf{S} = \mathbf{S}^{'}) = 1$ as $n \to \infty$. This scheme is depicted in Fig.4.

4.2 Binary Symmetric Channel (BSC)

Consider the Binary Symmetric Channel shown in Fig. 5. The capacity of a BSC can be calculated as follows:

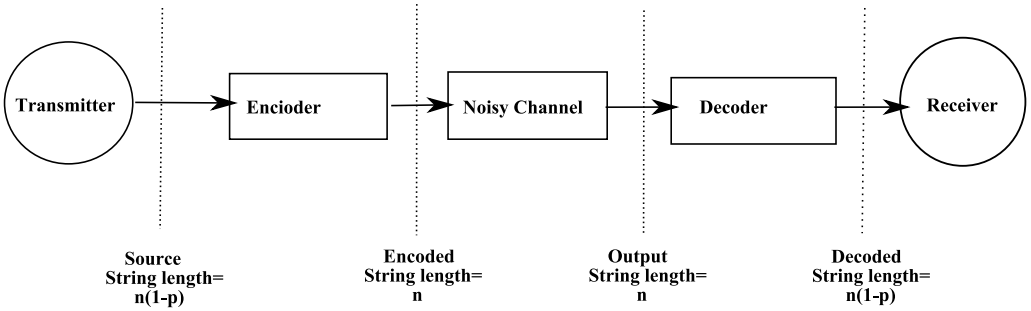

Figure 4: Proposed encoding scheme

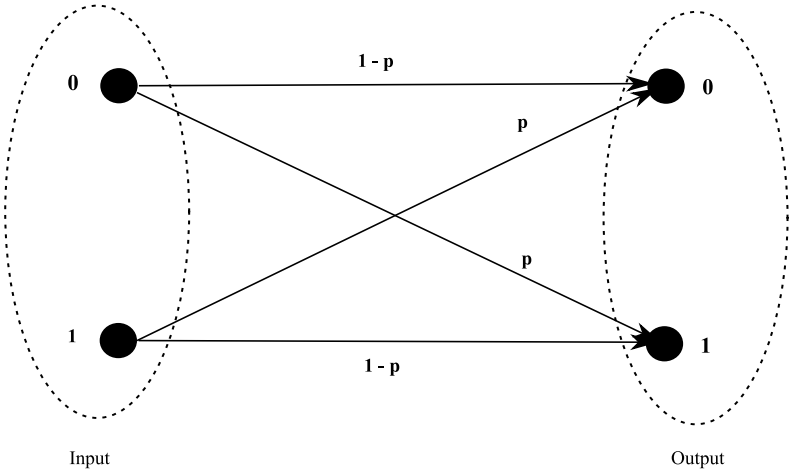

Figure 5: Binary Symmetric Channel

$$
\begin{aligned}
\mathbb{C}^{(1)} &= \max_{p_X(x)} (\mathbf{I}(X;Y)) \\
&= \max_{p_X(x)} (\mathbf{H}(Y) - \underbrace{\mathbf{H}(Y|X)}_{\mathbf{H}(p)}) \\
&= \max_{p_X(x)} (\mathbf{H}(Y) - \mathbf{H}(p)) \\
&= \max_{p_X(x)} (\mathbf{H}(Y)) - \mathbf{H}(p) \\
&\leq 1 - \mathbf{H}(p)
\end{aligned}
$$

where the final inequality is achieved if $p_X(x)$ is a uniform distribution.

4.3 Noisy Typewriter

In this case, the channel input is either received unchanged at the output with probability $\frac{1}{2}$ or transformed into the next letter with probability $\frac{1}{2}$. The channel transition probability matrix for such a channel is shown in Fig.6.

INPUT

OUTPUT

	A	B	C	D		Y	Z
A	½	½	0	0	.....	0	0
B	0	½	½	0	.....	0	0
C	0	0	½	½	.....	0	0
D	0	0	0	½	.....	0	0
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Y	0	0	0	0	.....	½	½
Z	1/2	0	0	0	.....	0	½

Figure 6: Channel matrix for the noisy typewriter

The channel transition matrix, $\mathbf{P}_{\mathbf{Y}^n|\mathbf{X}^n}(\mathbf{y}^n|\mathbf{x}^n)$ is symmetric and has the following properties:

1. Every row of $\mathbf{P}_{\mathbf{Y}^n|\mathbf{X}^n}$ is a permutation of the first row.
2. Every column of $\mathbf{P}_{\mathbf{Y}^n|\mathbf{X}^n}$ is a permutation of the first column.

For such a channel,

$$
\begin{aligned}
\mathbb{C}^{(1)} &= \max_{p_X(x)} (\mathbf{I}(X;Y)) \\
&= \max_{p_X(x)} (\mathbf{H}(Y) - \mathbf{H}(Y|X)) \\
&\leq \max_{p_X(x)} (\mathbf{H}(Y)) - \min_{p_X(x)} (\mathbf{H}(Y|X)) \\
&= \max_{p_X(x)} (\mathbf{H}(Y)) - (\text{Entropy of the first row}) \\
&\leq \log(|\Omega_Y|) - (\text{Entropy of the first row})
\end{aligned}
$$

where the final inequality is satisfied if $p_X(x)$ is a uniform distribution. Therefore for the noisy typewriter, $\mathbb{C}^{(1)} \leq \log(26) - 1 = \log(13)$ bits.