Communication Theory II

Lecture 8: Channel Coding

Channel Coding Theorem

Problem:

finding the maximum number of distinguishable signals for n uses of a communication channel.

The maximum number of distinguishable signals (M) for multi-level modulation (M-ary) is $M=2^K$.

This number grows exponentially with n.

Mathematical model

The mathematical analog of a physical signaling system is shown.
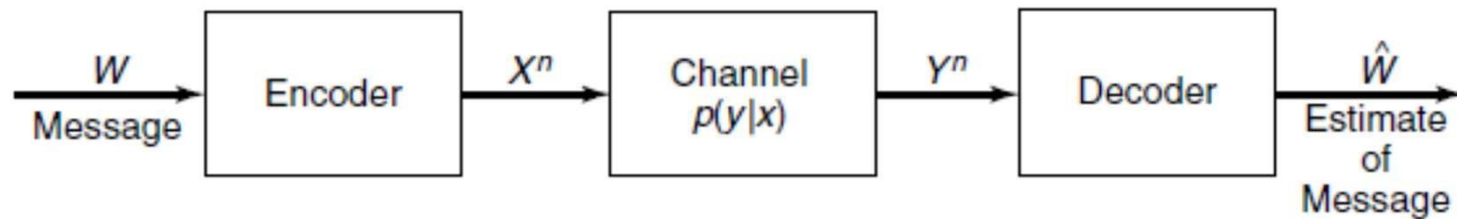

FIGURE 7.1. Communication system.

Problem: two different input sequences may give rise to the same output sequence; **the inputs are confusable**.

We show that we can choose a "nonconfusable" subset of input sequences so that with high probability there is only one highly likely input that could have caused the particular output.

Definitions

Definition *discrete channel :(X, p(y | x), Y)*

a system consisting of an input alphabet *X* and output alphabet *Y* (finite sets)

and a probability transition matrix *p(y | x)* that expresses the probability of

observing the output symbol y given that we send the symbol x.

The channel is said to be *memoryless* if the probability distribution of the output

depends only on the input at that time and is conditionally independent of

previous channel inputs or outputs.

Definition *"information" channel capacity* of a discrete memoryless channel:

$$C = \max_{p(x)} I(X; Y),$$

C is the channel capacity in bits per channel use.
I(X; Y) is the mutual information between the input random variable X and the output random variable Y.

Information capacity

The Information Capacity Theorem, also known as the Shannon's Channel Capacity Theorem or Shannon's Theorem, is one of the fundamental results in information theory. It was introduced by Claude Shannon in his landmark paper titled "A Mathematical Theory of Communication" published in 1948.

$$C = \max I(X; Y)$$

Mutual information

➢ Mutual information is a fundamental concept in information theory that measures the amount of information shared between two random variables. It quantifies how much knowing the value of one random variable reduces the uncertainty about the other random variable.

➢ The mutual information is always non-negative, and it is zero if and only if X and Y are independent random variables.

➢ When I(X; Y) is positive, it indicates that there is some dependency or relationship between X and Y. The larger the mutual information, the stronger the association or correlation between the two random variables.

➢ In the context of communication systems and channel capacity, mutual information plays a crucial role. It measures how much information is conveyed from the input to the output of the communication channel.

➢ To achieve high channel capacity, it is desirable to maximize the mutual information between the transmitted symbols and the received symbols, which indicates efficient use of the channel and better resistance to noise and distortion.

$$I(\mathcal{X}; \mathcal{Y}) = H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y})$$

conditional entropy

Uncertainty before observing Y   Uncertainty after observing Y

Mutual information

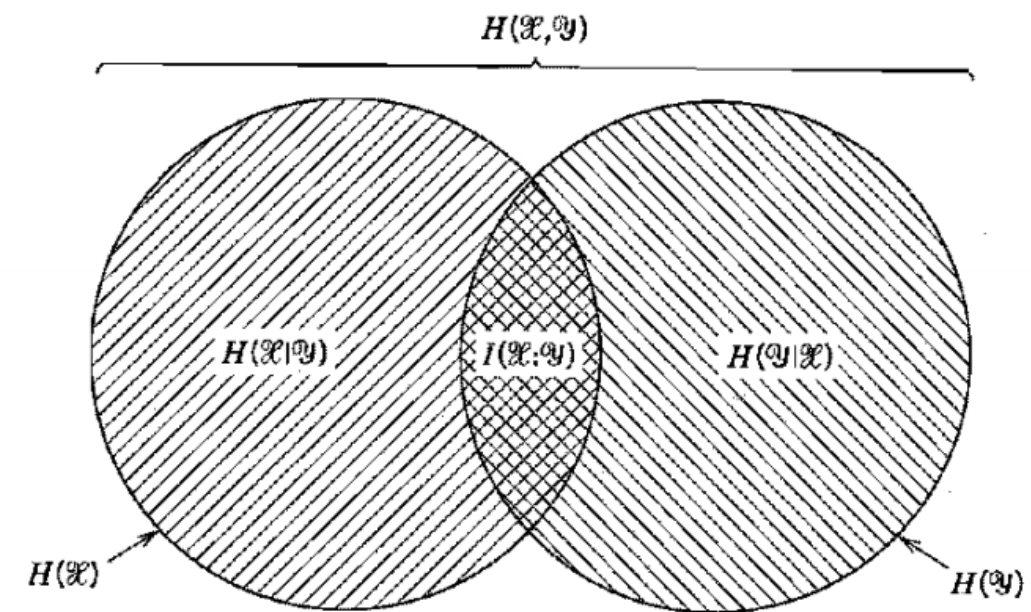➢ Non-Negativity $\quad I(\mathcal{X}; \mathcal{Y}) \geq 0$

➢ Symmetry $\quad I(\mathcal{X}; \mathcal{Y}) = I(\mathcal{Y}; \mathcal{X})$

➢ Connection to Entropy $\quad I(X; Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$

$$I(\mathcal{X}; \mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y}) - H(\mathcal{X}, \mathcal{Y})$$

$$H(\mathcal{X}, \mathcal{Y}) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2\left(\frac{1}{p(x_j, y_k)}\right)$$

Joint entropy is a measure of the total uncertainty or average information content in two or more random variables considered together.

$$\mathcal{X} = \{x_0, x_1, \ldots, x_{J-1}\}, \qquad \mathcal{Y} = \{y_0, y_1, \ldots, y_{K-1}\},$$

Conditional entropy

It measures the average amount of uncertainty or surprise in one random variable (X) given the value of another random variable (Y).

In other words, it quantifies how much uncertainty remains in X once we know the value of Y

$$H(\mathcal{X} \mid \mathcal{Y}) = \sum_{k=0}^{K-1} H(\mathcal{X} \mid Y = y_k) p(y_k)$$

$$= \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} p(x_j \mid y_k) p(y_k) \log_2 \left[\frac{1}{p(x_j \mid y_k)} \right]$$

$$= \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} p(x_j, y_k) \log_2 \left[\frac{1}{p(x_j \mid y_k)} \right]$$

Entropy - recap

➤ If the entropy of a random variable is low, it means that the variable's outcomes are more predictable, and there is less uncertainty associated with each outcome.

➤ Conversely, if the entropy is high, it means that the variable's outcomes are less predictable, and there is more uncertainty associated with each outcome.
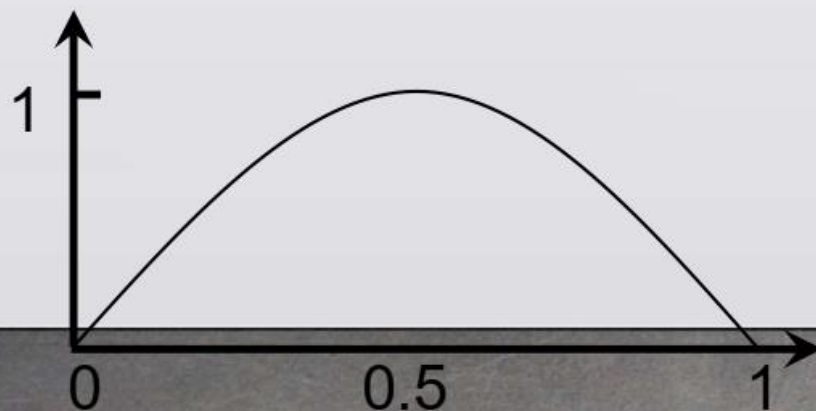
$$p = P(X_k = 1)$$

$$q = P(X_k = 0) = 1 - p$$

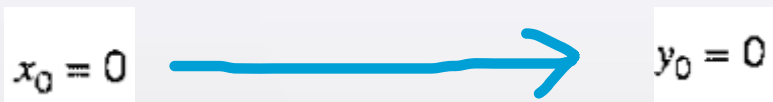$$H(\mathcal{S}) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right)$$

$$H = p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p}$$

$$I(X; Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$$

What is Entropy? and its relation to Compression - YouTube

Examples Of Channel Capacity

Any transmitted bit is received without error

Noiseless Binary Channel

$x_0 = 0$ $\longrightarrow$ $y_0 = 0$
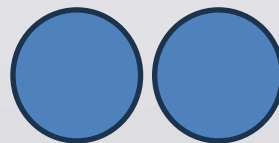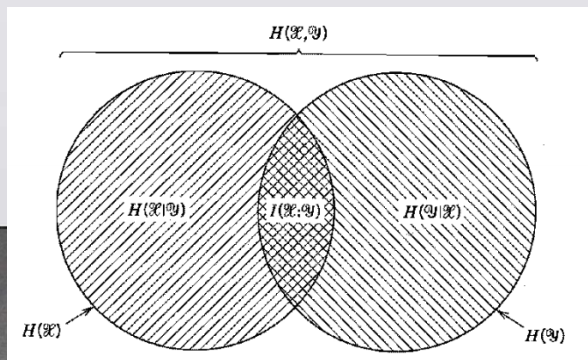
$x_1 = 1$ $\longrightarrow$ $y_1 = 1$

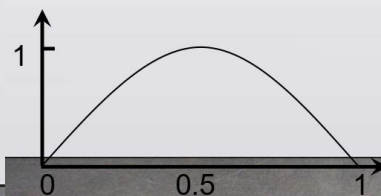$(x_0 = 0, x_1 = 1)$ $(y_0 = 0, y_1 = 1)$

Input symbols   output symbols

One error-free bit can be transmitted per use of the channel, so the capacity is 1 bit.

Or, by the definition of C :

C = max I (X; Y)

$$I(\mathcal{X}; \mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y}) - H(\mathcal{X}, \mathcal{Y})$$

$$H = p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p}$$

$H(\mathcal{X}, \mathcal{Y})$

$H(\mathcal{X}|\mathcal{Y})$ $I(\mathcal{X};\mathcal{Y})$ $H(\mathcal{Y}|\mathcal{X})$

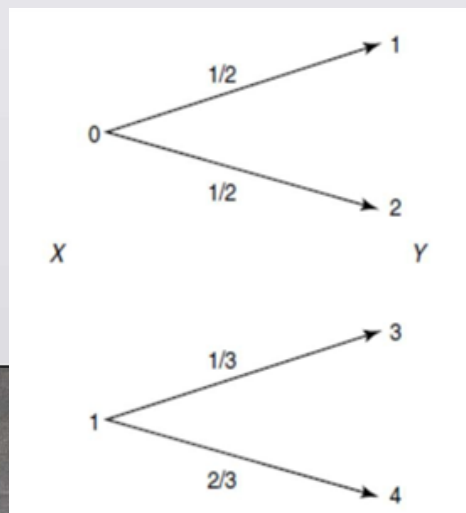$H(\mathcal{X})$ $H(\mathcal{Y})$

achieved by: *p(x) = (1/2, 1/2)*

C = max I (X; Y) =1 bit,
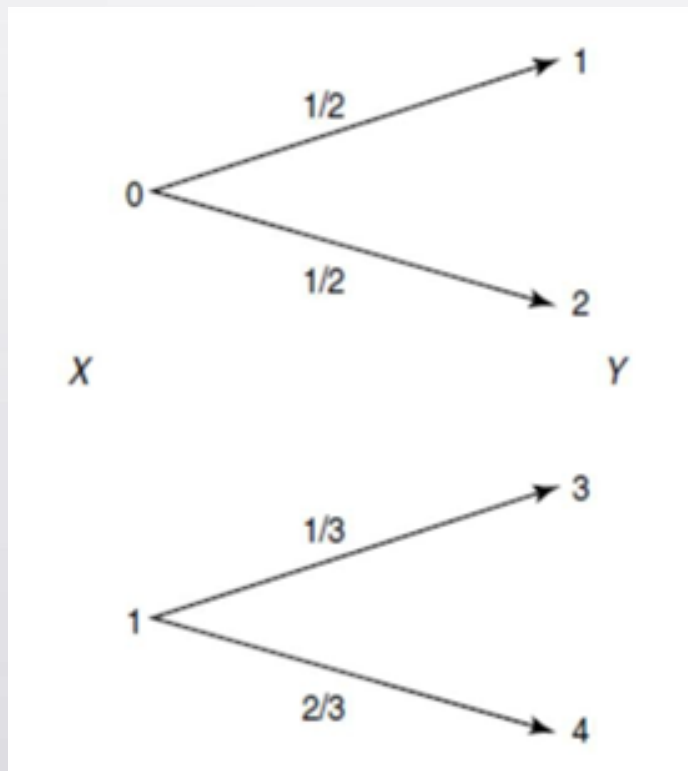
Examples Of Channel Capacity

Noisy channel

- In this context, the noisy channel represents an imperfect communication channel where errors or distortions can occur during transmission.

- Noisy channel with non-overlapping outputs

    - The non-overlapping outputs mean that each transmitted symbol or message corresponds to a unique received symbol, and there is no ambiguity or confusion between different transmitted symbols leading to the same received symbol

Examples Of Channel Capacity

Noisy channel with nonoverlapping outputs

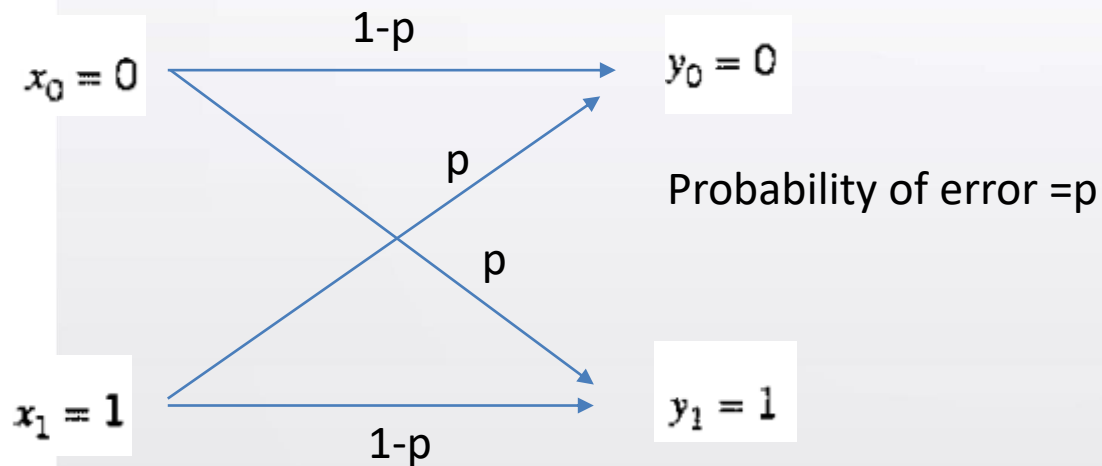**The input can be determined from the output**

$$I(\mathscr{X}; \mathscr{Y}) = H(\mathscr{X}) + H(\mathscr{Y}) - H(\mathscr{X}, \mathscr{Y})$$

C = max I (X; Y) = 1 bit,

achieved by :

p(x) = (1/2, 1/2).

Examples Of Channel Capacity

$x_0 = 0$

$1-p$ → $y_0 = 0$

p

Probability of error $=p$

p

$x_1 = 1$

$1-p$ → $y_1 = 1$

Conditional probabilities

$$p_{10} = P(y = 1 \mid x = 0)$$
$$p_{01} = P(y = 0 \mid x = 1)$$
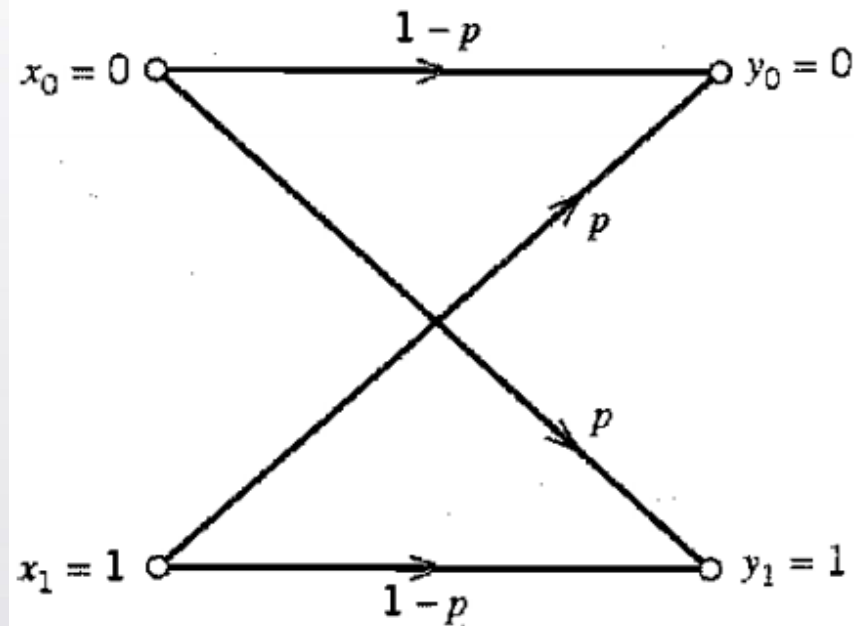$$p_{10} = p_{01} = p$$

$(x_0 = 0, x_1 = 1)$

$(y_0 = 0, y_1 = 1)$

Input symbols

output symbols

Transition probability diagram of binary symmetric channel.

Questions

Find the probability of symbol 0 appearing in the output.
Find the probability of symbol 1 appearing in the output.
Calculate the above values for p=0.3 and p=0.5.

$$p(y_0) = (1 - p)p(x_0) + p\, p(x_1)$$

$$p(y_1) = p\, p(x_0) + (1 - p)\, p(x_1)$$

$$P(x_0) = P(x_1) = \frac{1}{2}$$

$$p(y_0) = (1 - p)p(x_0) + p\, p(x_1) = (1 - p)\left(\frac{1}{2}\right) + p\left(\frac{1}{2}\right) = \frac{1}{2}$$
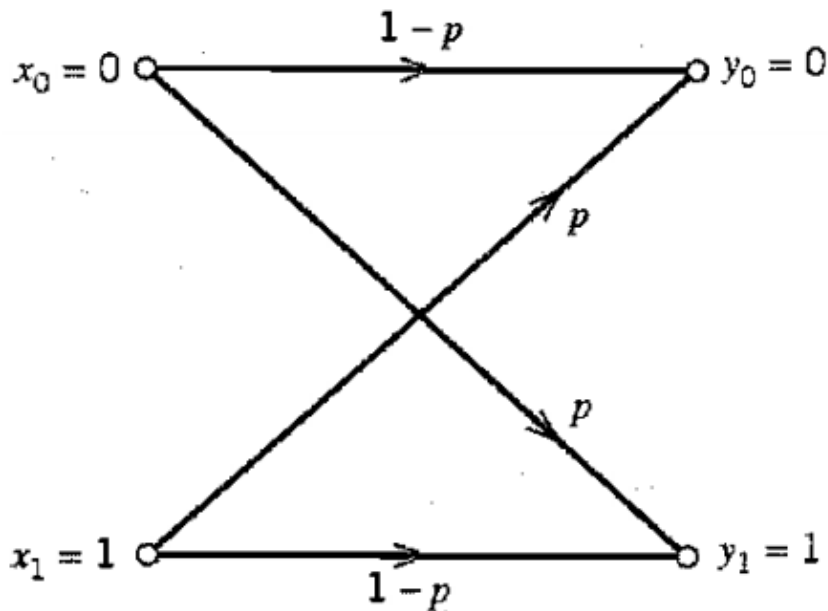
Examples Of Channel Capacity

This is a model of a channel with errors, all the bits received are unreliable.

Binary Symmetric Channel

$I(X; Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$

$I(X; Y) = H(X) - H(X \mid Y)$

$$H(\mathscr{X} \mid Y = y_k) = \sum_{j=0}^{J-1} p(x_j \mid y_k) \log_2 \left[\frac{1}{p(x_j \mid y_k)} \right]$$

$$H(\mathscr{X} \mid \mathscr{Y}) = \sum_{k=0}^{K-1} H(\mathscr{X} \mid Y = y_k) p(y_k)$$

$p(x_0) = p_0$

$p(x_1) = 1 - p_0$

$H(X) = \sum_{k=0}^{K-1} p_k \log_2 \left(\frac{1}{p_k} \right)$

$H(X) = -p_0 \log_2 p_0 - (1 - p_0) \log_2(1 - p_0)$ bits

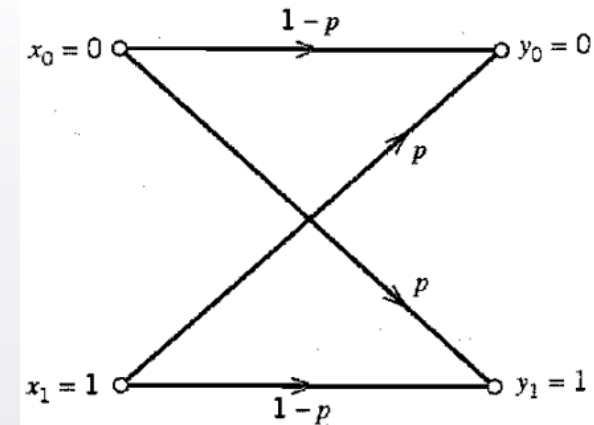Entropy function $\quad \mathcal{H}(p_0) = -p_0 \log_2 p_0 - (1 - p_0) \log_2(1 - p_0)$

Examples Of Channel Capacity

$$H(\mathcal{X}|\mathcal{Y}) = \sum_{k=0}^{K-1} H(\mathcal{X}|Y = y_k)p(y_k)$$

$$H(\mathcal{X}|Y = y_k) = \sum_{j=0}^{J-1} p(x_j|y_k) \log_2\left[\frac{1}{p(x_j|y_k)}\right]$$

$$J = K = 2$$

$$\begin{aligned}
H(\mathcal{X}|\mathcal{Y}) &= \sum_{k=0}^{K-1} H(\mathcal{X}|Y = y_k)p(y_k) \\
&= \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} \boxed{p(x_j|y_k)}p(y_k) \log_2\left[\frac{1}{\boxed{p(x_j|y_k)}}\right] \\
&= \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j, y_k) \log_2\left[\frac{1}{p(x_j|y_k)}\right]
\end{aligned}$$

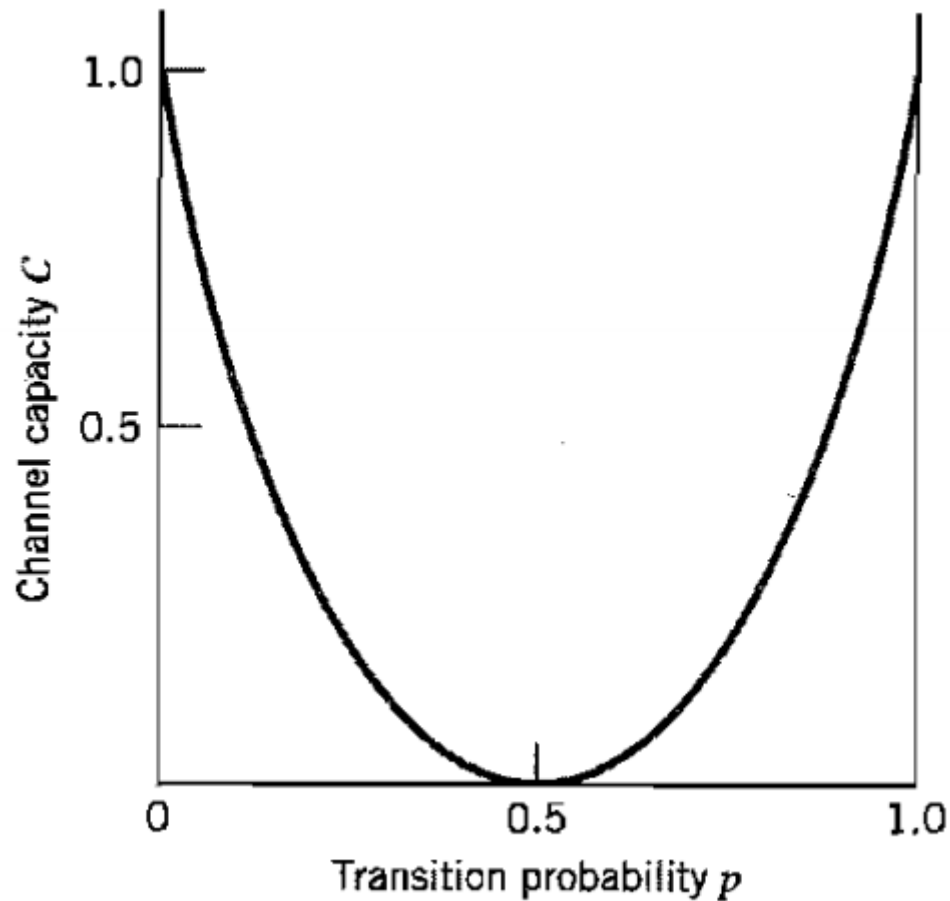$$I(X; Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$$

$$\begin{aligned}
I(X; Y) &= H(Y) - H(Y|X) \\
&= H(Y) - \sum p(x)H(Y|X = x) \\
&= H(Y) - \sum p(x)H(p) \\
&= H(Y) - H(p) \\
&\leq 1 - H(p),
\end{aligned}$$

Equality is achieved when the input distribution is uniform.

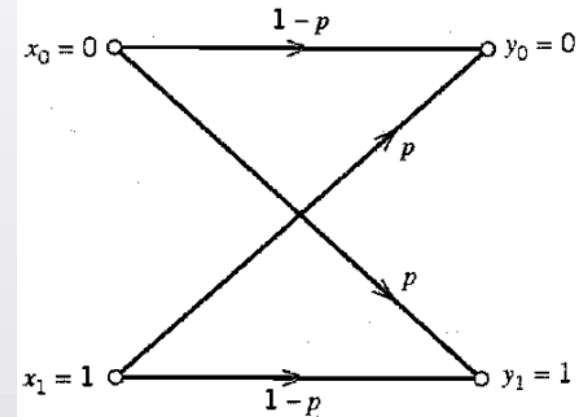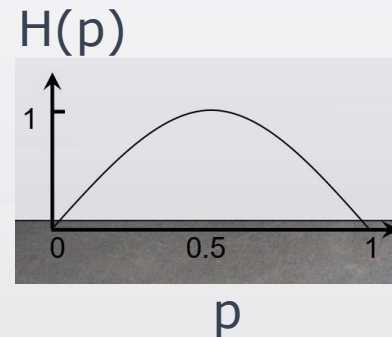Hence, the information capacity of a binary symmetric channel with parameter p is $C = 1 - H(p)$ bits.   We lose H(p) fraction of bits per transmission

Examples Of Channel Capacity

$C = 1 - H(p)$ bits
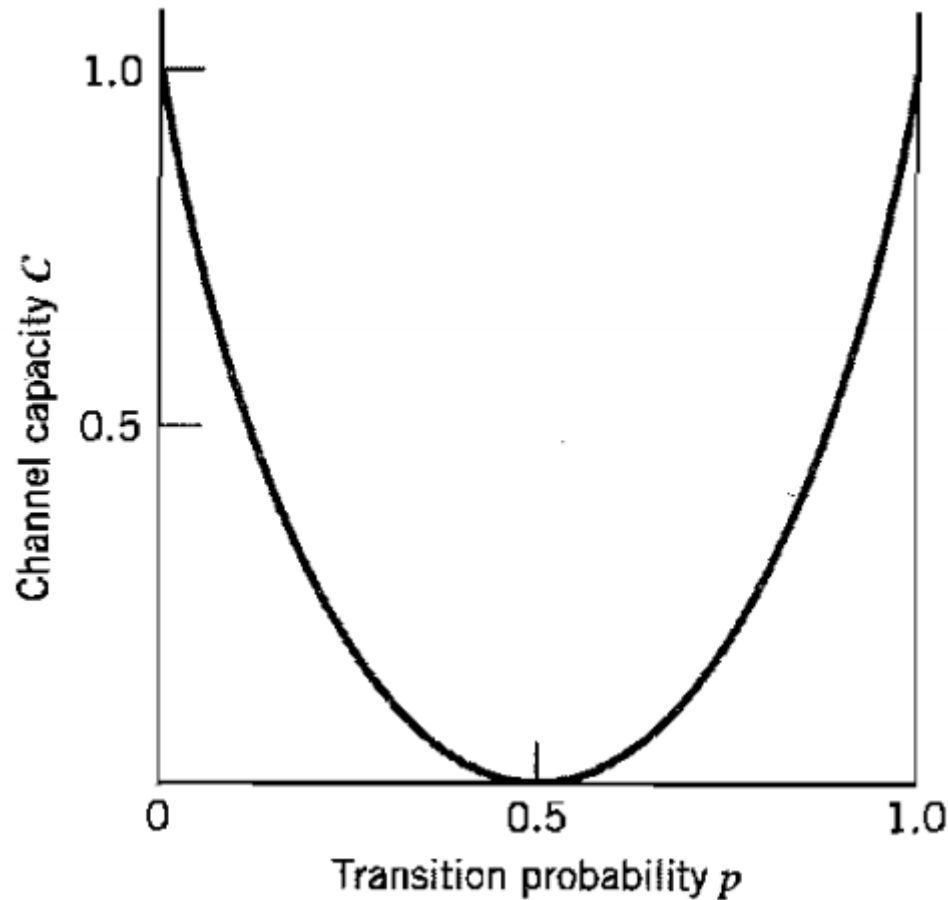
$H(p) = -p \log_2 (p) - (1-p) \log_2 (1-p)$
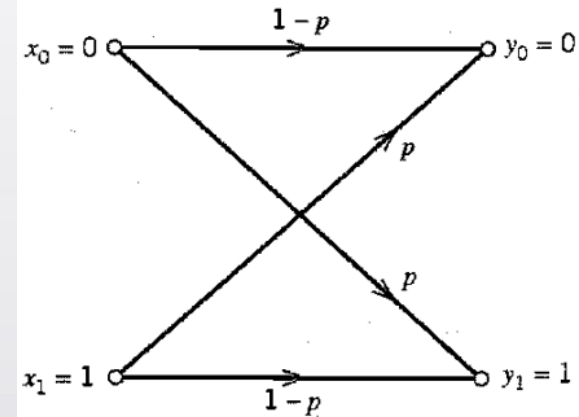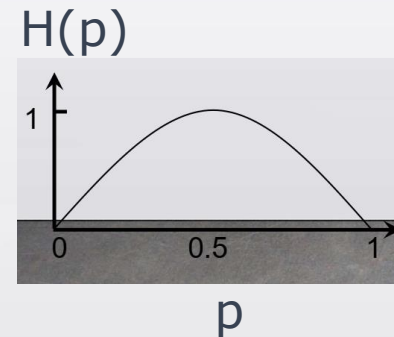
H(p)

p

Noise free channel

- p=?
- C=?

In a noise-free binary channel, the channel capacity c represents the maximum rate at which information can be reliably transmitted without any errors. P=0 and C= 1bit. Here, H(p)=0 (its minimum value)

Examples Of Channel Capacity

Channel capacity C vs Transition probability p

$C = 1 - H(p)$ bits

$$H(p) = -p \log_2(p) - (1-p) \log_2(1-p)$$

$H(p)$

p

P=1/2
- H(p)=?
- C=?

H(p) at its maximum. C at its minimum value zero
Here, channel said to be useless.

When a channel has zero capacity, it means that no matter how much information the sender tries to transmit through the channel, the receiver will not be able to correctly reconstruct the original message. In practical terms, this means that any communication over such a channel would be subject to errors, and the transmitted information would be completely unreliable.

Examples Of Channel Capacity

The Binary Erasure Channel (BEC) is a specific type of communication channel used in information theory. It is a simple and widely studied channel that models the transmission of binary data (0s and 1s) between a sender and a receiver, where some of the bits may get erased (lost) during transmission with a certain probability.
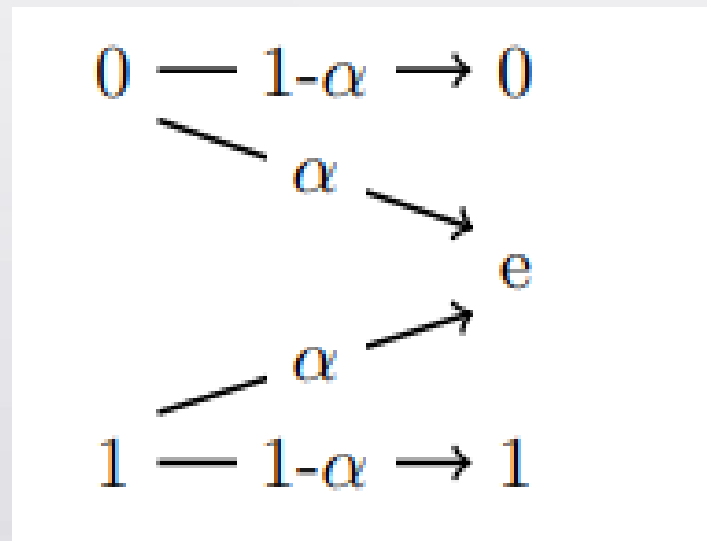
Erasures: The channel introduces erasures, where some transmitted bits may be lost during transmission. An erasure is represented by the symbol "e."

The probability of an erasure occurring (a bit being lost) is denoted by "∝."

Input {0,1}   Output {0,1,e}
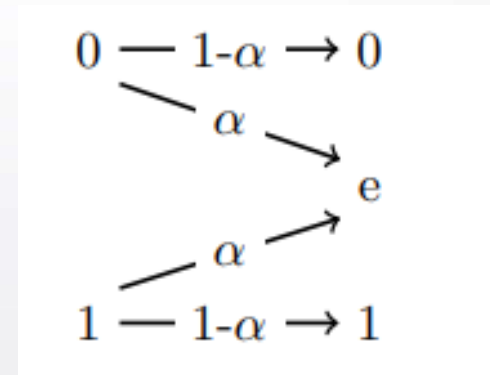
Examples Of Channel Capacity

$I(X; Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$

$X = \{0, 1\}$ and $Y = \{0, 1, e\}$

- If we know $Y = 0$ or 1, then $H(X \mid Y)$?

In the Binary Erasure Channel (BEC), when we observe the output Y to be either "0" or "1" (i.e., $Y \neq e$), the input X is known exactly, and there is no uncertainty about it. Therefore, the conditional entropy $H(X|Y \neq e)$ in such cases is indeed equal to 0.

Additionally, when we observe the output Y to be "e" (i.e., $Y = e$), it means that the transmitted bit got erased, and we have no information about the exact value of the input X. In this case, the conditional entropy $H(X|Y = e)$ is equal to the entropy of the input X, denoted as $H(X)$.

$H(X|Y) = 0 \cdot P(Y \neq e) + H(X) \cdot P(Y = e) = \propto H(X)$ ⟹ $I(X; Y) = H(X) - \propto H(X) = (1 - \propto)H(X)$
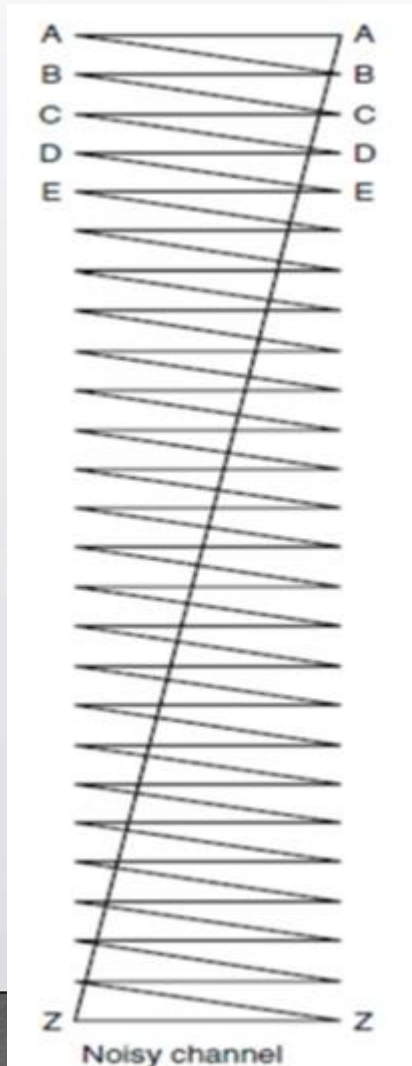
$I(X; Y) \leq 1 - \propto$ and when $\propto$ is uniform max $(I(X;Y)) = 1 - \propto$

We lose α fraction of bits per transmission

Examples Of Channel Capacity

Noisy Typewriter

Noisy channel

The channel input is either received unchang at the output with probability 1/2 or is transformed into the next letter with probability of 1/2

1. If you send the letter "A" through the channel, there is a 50% chance it will be received as "A" and a 50% chance it will be received as "B."

2. If you send the letter "B" through the channel, there is a 50% chance it will be received as "B" and a 50% chance it will be received as "C."

Task
1. How many bits need to transmit 26 letters?
2. Sending 26 letter is a good method? How to minimize uncertainty?

Examples Of Channel Capacity

Noisy Typewriter

Noisy channel   Noiseless subset of inputs
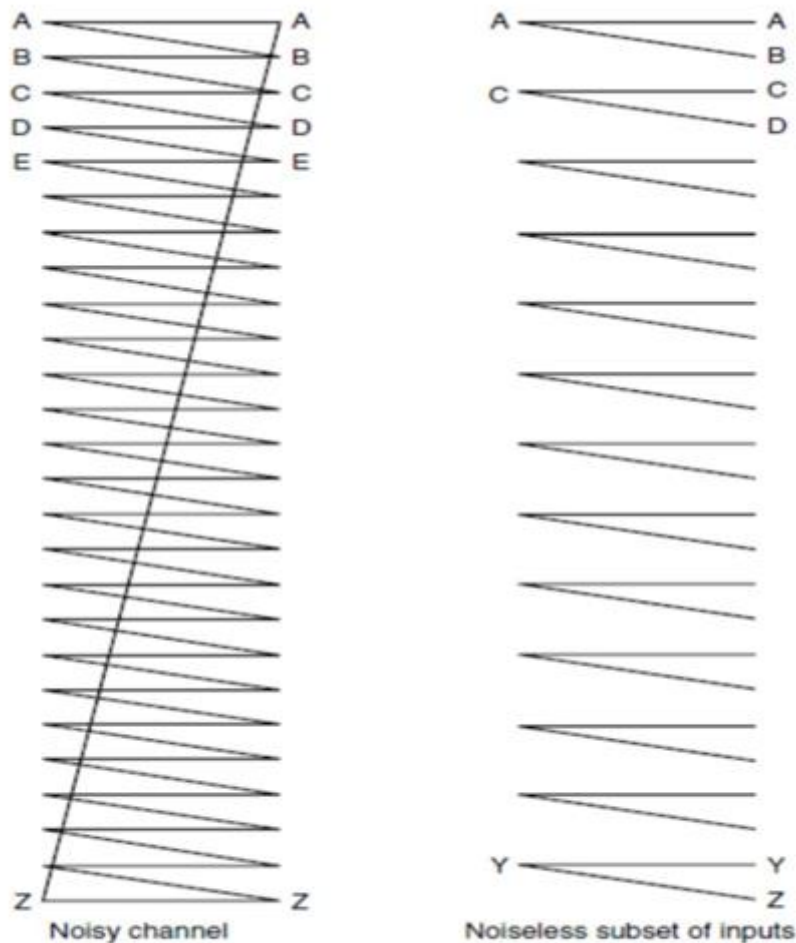
Noisy Typewriter. $C = \log 13$ bits.

$C = \max I\,(X; Y)$

$$I(X;Y) = H(Y) - H(Y|X)$$

$$= H(Y) - \sum p(x)H(Y|X=x)$$

$$H(\mathcal{Y}) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right)$$

Max H(y)=$\log_2(26)$ $H(Y) \le \log_2 26.$

achieved by using *p(x)* distributed uniformly over all the inputs.

$$H(Y|X) = \sum_x p(x)H(Y|X=x)$$

$$= \sum_x p(x) = 1.$$

$C = \max I\,(X; Y) = \log_2 26 - 1 = \log_2 26 - \log_2 2$

$= \log_2 13$

$\log(a) - \log(b) = \log(a/b)$

Definitions

A message W, is drawn from the index set {1, 2, . . . , M}.

Definition The *nth extension of the discrete memoryless channel (DMC)*

is the channel $(X^n, p(y^n \mid x^n), Y^n)$, where

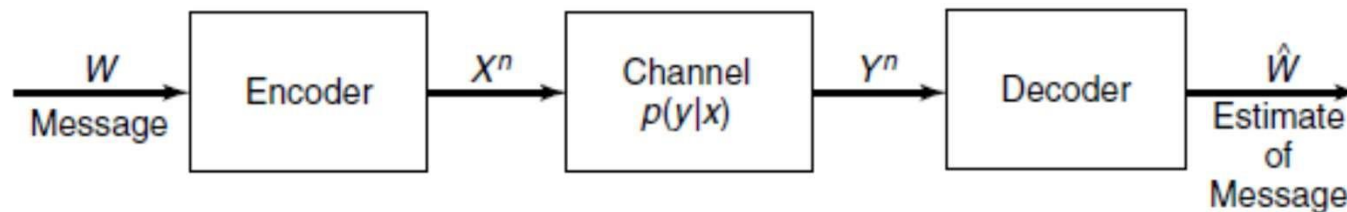$p(y_k \mid x_k, y_{k-1}) = p(y_k \mid x_k)$, k = 1, 2, . . . , n.

FIGURE 7.8. Communication channel.

without feedback:

$$p(y^n \mid x^n) = \prod_{i=1}^{n} p(y_i \mid x_i).$$

Definitions

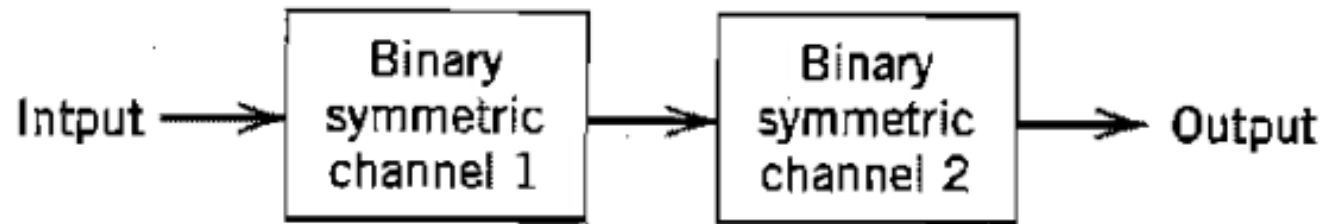**Definition** An *(M, n)* code for the channel *(x, p(y|x), y)* consists of the following:

1. An index set {1, 2, ... , M}.

2. An encoding function X^n : {1, 2, ...,M} → X^n, yielding codewords $x^n(1)$, $x^n(2)$, . . ., $x^n(M)$. The set of codewords is called the *codebook*.

3. A decoding function g : Y^n → {1, 2, . . . , M}.

**Definition** (*Conditional probability of error*) Let
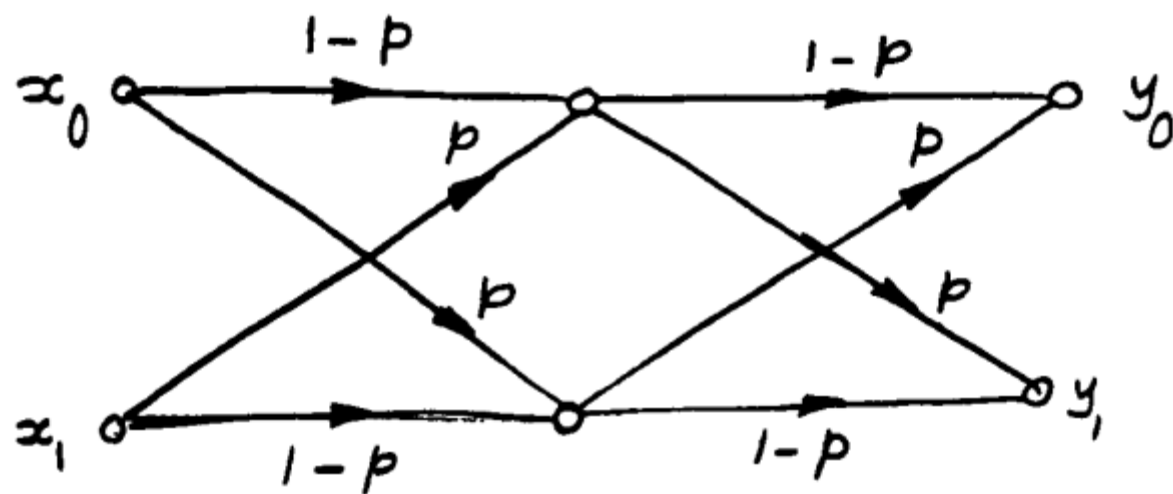
$$\lambda_i = \Pr(g(Y^n) \neq i | X^n = x^n(i))$$

be the *conditional probability of error* given that index *i* was sent.

Examples

In the given communication setup, two binary symmetric channels are connected in cascade. Each channel has the same transition probability diagram as shown in below figure. The task is to find the overall channel capacity of the cascaded connection.
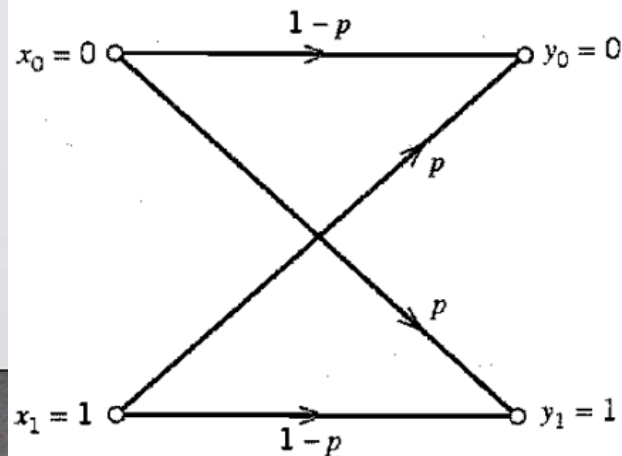
$P(y_0 \mid x_0) = (1 - p)^2 + p^2 = 1 - 2p(1 - p)$

$P(y_0 \mid x_1) = p(1 - p) + (1 - p)p = 2p(1 - p)$
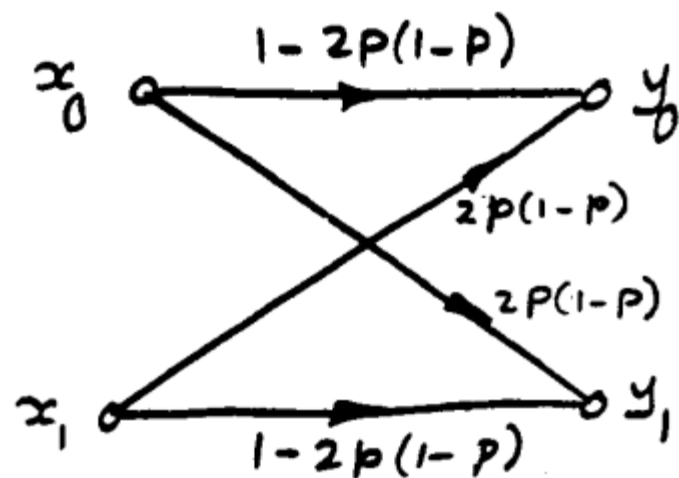
$C = 1 - H(2p(1 - p))$

$= 1 - 2p(1 - p) \log_2 [2p(1 - p)] - (1 - 2p + 2p^2) \log_2 (1 - 2p + 2p^2)$
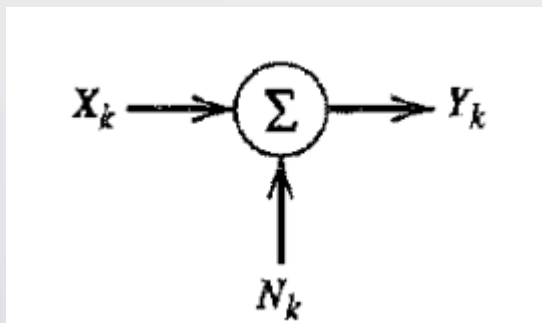
Information capacity

The Information Capacity Theorem, also known as the Shannon's Channel Capacity Theorem or Shannon's Theorem, is one of the fundamental results in information theory. It was introduced by Claude Shannon in his landmark paper titled "A Mathematical Theory of Communication" published in 1948.

$$C = \max I(X; Y)$$

capacity C for a Gaussian channel
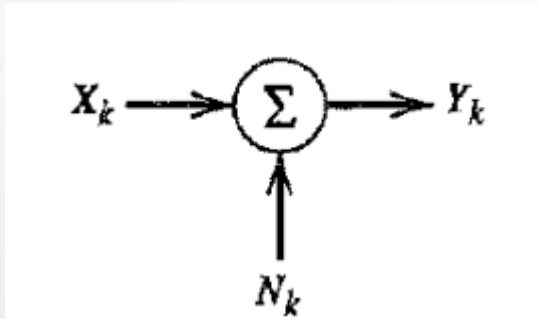

$$C = \frac{1}{2} \log_2\left(1 + \frac{P}{\sigma^2}\right) \text{ bits per transmission}$$

σ^2 is the noise power (variance of the Gaussian noise)

Noise: The Gaussian noise N has a probability distribution that follows a Gaussian (normal) distribution with mean zero and a certain variance, denoted as $N(0, \sigma^2)$, where σ^2 represents the noise power.

Information capacity

capacity C for a Gaussian channel

$$C = \frac{1}{2} \log_2\left(1 + \frac{P}{\sigma^2}\right) \text{ bits per transmission}$$

$$C = B \log_2\left(1 + \frac{P}{N_0 B}\right) \text{ bits per second}$$

P is the average transmitted power

The capacity of a Gaussian channel represents the maximum data rate achievable for reliable communication in the presence of Gaussian noise. The capacity is determined by Shannon's Capacity Formula, which considers the signal power and noise power in the channel

Information capacity
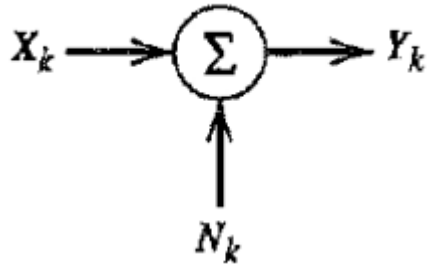
capacity C for a Gaussian channel

$$C = B \log_2\left(1 + \frac{P}{N_0 B}\right) \text{ bits per second}$$

P is the average transmitted power

$$C = B \ \log_2(1 + SNR)$$

Bandwidth (B) Influence: The bandwidth (B) of the channel directly scales the channel capacity. A wider bandwidth allows for a higher capacity, enabling the transmission of more data in a given time. The capacity increases linearly with bandwidth, as evident from the formula.

Logarithmic Relationship: As seen in the formula, the channel capacity (C) is proportional to the logarithm of (1 + SNR). This means that the capacity increases as the logarithm of the SNR increases. As SNR increases, the capacity increases but at a diminishing rate. The capacity does not increase linearly with SNR; instead, it approaches a maximum value as SNR becomes very large.

Information capacity

How to energy per bit related to capacity?

$$P = E_b C$$

$$C = B \log_2 \left(1 + \frac{P}{N_0 B} \right) \text{ bits per second}$$

$$\frac{C}{B} = \log_2 \left(1 + \frac{E_b}{N_0} \frac{C}{B} \right) \quad \Longrightarrow \quad \frac{E_b}{N_0} = \frac{2^{C/B} - 1}{C/B}$$

signal energy-per-bit to noise power spectral density ratio

Information capacity

$$\frac{E_b}{N_0} = \frac{2^{C/B} - 1}{C/B}$$

$$C = B \log_2 \left(1 + \frac{P}{N_0 B} \right) \text{ bits per second}$$

$$C_\infty = \lim_{B \to \infty} C$$

$$= \frac{P}{N_0} \log_2 e$$

$$\left(\frac{E_b}{N_0} \right)_\infty = \lim_{B \to \infty} \left(\frac{E_b}{N_0} \right)$$

$$= \log 2 = 0.693$$

Shannon limit for an AWGN channel

e is the base of the natural logarithm.

Bandwidth efficiency diagram

Bandwidth Efficiency: Bandwidth efficiency refers to how efficiently a communication system uses the available frequency spectrum to transmit data. It is typically measured in bits per second per hertz (bps/Hz). Higher bandwidth efficiency indicates that the system can transmit more data per unit of bandwidth.

Data Rate (Capacity): The data rate represents the amount of data transmitted per unit of time, typically measured in bits per second (bps)

Trade-off: The bandwidth efficiency diagram shows that there is a trade-off between data rate and bandwidth. In general, as the data rate increases, the occupied bandwidth also increases. Different communication schemes achieve different trade-offs on this graph, with some being more bandwidth-efficient at the expense of lower data rates, and others achieving higher data rates with wider bandwidth usage

Bandwidth efficiency diagram
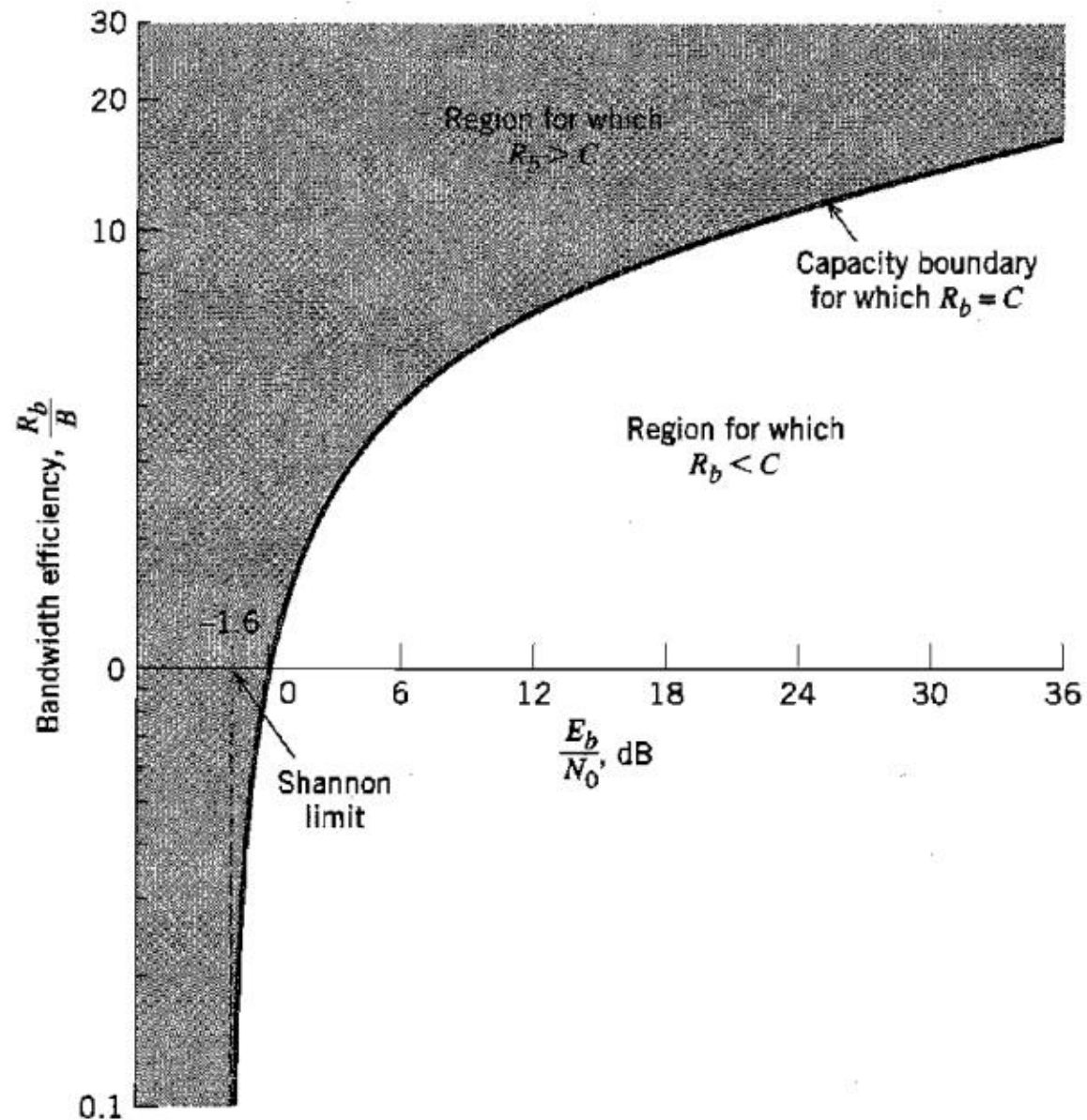
Example

We have a communication system operating in a frequency band with a bandwidth of 20 kHz (B = 20,000 Hz). The system is affected by additive white Gaussian noise. We want to calculate the channel capacity for three different scenarios with varying SNR values.
Scenario 1: SNR (dB) = 20 dB
Scenario 2: SNR (dB) = 30 dB
Scenario 3: SNR (dB) = 40 dB

Convert SNR (dB) to linear scale:

$$\text{SNR} = 10^{(\text{SNR (dB)}/10)}$$

$$\text{SNR} = 10^{(20/10)} = 100$$

Calculate the channel capacity:

$$C = B \cdot \log_2(1 + \text{SNR})$$

$$C = 20000 \cdot \log_2(1 + 100) \approx 20000 \cdot \log_2(101) \approx 20000 \cdot 6.6612 \approx$$

$$133,224 \text{ bps}$$

Example

Example: LTE (Long-Term Evolution) Mobile Communication

LTE is a widely used wireless communication technology for 4G mobile networks. In LTE, the available frequency band is divided into multiple subcarriers to enable simultaneous data transmission for multiple users. Each subcarrier has a specific bandwidth, and the goal is to maximize the data rate (channel capacity) for each user.

In LTE, the available frequency band is divided into multiple subcarriers, and each subcarrier typically has a bandwidth of 15 kHz. LTE uses a range of bandwidths depending on the specific deployment scenario and network configuration, such as 1.4 MHz, 3 MHz, 5 MHz, 10 MHz, 15 MHz, or 20 MHz.

For example, let's consider an LTE network that uses a bandwidth of 10 MHz

Assuming a specific LTE cell scenario, let's consider three different SNR values for three users:
•User 1: SNR = 10 dB
•User 2: SNR = 20 dB
•User 3: SNR = 30 dB

•Calculate the channel capacity for three different scenarios