Information Theory - Entropy

Concept of Information

Figure 3

A typical binary coded digital communication system is shown in Figure 3. What is involved in the transmission of information?

- The system user wishes to transmit the full meaning or implication of the message, so that it may be received or retrieved without loss or distortion, and acted upon.

- The system designer wishes to extract that which is essential in a message, so as to transmit it efficiently.

- Neither is primarily concerned with the physical reality of the message generated by the source; e.g.,

  it is not necessary to transmit the piece of paper on which a telegram is written: it is only the letters or words need be sent.

Thus a concept of what constitutes meaningful information needs to be agreed so that the user and the designer can both be satisfied. When considering a definition of information, several points are relevant:-

1. Information source can be classified into two categories: analog and discrete.

   Analog information sources, such as a microphone output can be transformed into discrete information source through the process of sampling and quantising.

2. Information is only conveyed by the probability of change. Certainty does not convey information, unpredictability does.

   Eg. A black-board can only convey information when its clean surface is altered by writing symbols on it. To convey further information, more symbols should be used. Thus at least, two symbols are required to convey information. In our example, "no symbol" is only a particular kind of symbol. In other words, information is resolution of uncertainty, not knowledge, and is essentially discrete.

3. More information can be conveyed by using a larger set of symbols. This is because the uncertainty connected with the appearance of a symbol in the larger set

is greater.

In practice, large signal sets are difficult to implement, complicated statements must be broken down for transmission, implying either complexity or time delay. Another implication of the breaking down process is that the system may be unware of the relative importance of the symbols it is transmitting.

Eg. LOVE and VOLE will be treated alike if the letters are the symbols, in spite of the big difference in their meaning.

Also, the information in a book is proportional to its length, if the words are treated as symbols!

4. If the symbols occur in a statistically independent form, then the information represented by each symbol is related to the frequency of its occurrence from the source.

   The probability of occurrence of a symbol determines the amount of information it conveys.

5. If a message sequence has twice as many symbols as another message, then on average it can be expected to convey twice as much information. This is also another implication of point 3 above.

   Put another way, on average, the sum of the information represented by each symbol in a message should equal the information represented by the whole message.

Summary

Information source - analog or discrete

Larger uncertainty - more information

Sum of information represented by each symbol in a message - should equal the information represented by the whole message.

a message - a sequence of symbols.

Consider a set of M symbols (messages) m_1, m_2, ..., m_M with probabilities occurrence $p_1, p_2, \ldots, p_M$;

$$p_1 + p_2 + \ldots + p_M = 1 .$$

The information content or the amount of information in the k-th symbol (message) $I(m_k)$ is represented by

$$I(m_k) = -\log_b p_k. \qquad (29)$$

This is one of many functions which satisfy the concept of information and gives appropriate and convincing results in use.

The information associated with each symbol is

$$I(m_1), I(m_2), \ldots, I(m_M).$$

Then

$I(m_1) + I(m_2) + \ldots = I(m_1, m_2, \ldots)$ i.e. pt.5

$I(m_k)$ is a function of p_k i.e. pt.4

$M < \infty$ i.e. pt.3

$M \geq 2$ i.e. pt.2 .

From equation (29),

$I(m_k) \dashrightarrow 0$ if $p_k \dashrightarrow 1$ (likely event)

$I(m_k) \dashrightarrow \infty$ if $p_k \dashrightarrow 0$ (unlikely even) .

The information in M one-symbol messages is equal to the information in one M-symbol message, i.e.

$$I(m_1) + I(m_2) + \ldots = -\log_b(p_1) - \log_b(p_2) - \ldots \qquad (30)$$
$$I(m_1) + I(m_2) + \ldots = \log_b (p_1 p_2 \ldots)$$

$$I(m_1) + I(m_2) + \ldots = I(m_1, m_2, \ldots) \qquad (31)$$

The unit of information is in bit when the base b=2 or in nat when the base b is natural.

Eg.1 A binary system with two symbols 0 and 1 of equal probabilities.

M=2 , $p_1 = p_2 = 1/2$

Therefore, $I(m_1) = I(m_2) = -\log_2 (1/2)$

= 1 bit of information with base b=2 .

In general,

$I(m_k)$ is sometimes called the self-information associated with a symbol m_k.

Eg.2 A binary source, $p(m_1)$=1/4, $p(m_2)$ = 3/4.

$I(m_1) = -\log_2 (1/4) = 2$ bits

$I(m_2) = -\log_2 (3/4) = 0.415$ bits

From now on, we will assume b=2.

Summary

$I(m_k) > I(m_j)$ if $p_k < p_j$

$I(m_k)$ --> 0 as p_k --> 1

$I(m_k) > 0$ when $0 \leq p_k \leq 1$

$I(m_k) = -\log_{b=2} p_k$

$I(m_1) + I(m_2) + \ldots = I(m_1, m_2, \ldots)$

where m_1, m_2, ... are independent symbols (messages)

$I(m_k)$ - information content or self-information or amount of information associated with a symbol m_k.

Average Information Content (Entropy)

Case 1 : Entropy of symbols in long independent sequences

A message consists of a sequence of symbols. If we are sending messages via a communication channel, we often have to deal with individual symbols at the receiving end.

Thus, the communication system has to deal with individual symbols. Due to the randomness involved in the selection of symbols, we need to talk about average information content of symbols in a long message.

Assume we have a source that emits one of M possible symbols s_1, s_2, ... s_M in a statistically independent manner with probabilities of occurrence $p_1, p_2, \ldots p_M$.

Thus, the probability of occurrence of a particular symbol at a given interval does not influence the symbols emitted during the previous time intervals.

In a long message containing N symbols,

| message |

<------ N ------->

s_i will occur on the average $p_i N$ times. If we treat each symbol as a message of length one, the information content of the i-th symbol is

$$I(s_i) = -\log_2 p_i \quad \text{bits.} \tag{32}$$

With $p_i N$ times of occurrence, s_i contributes $p_i N \ I(s_i)$ bits of information content in the message of length N.

The total information content of the message of length N is equal to the sum of the contribution due to each of the M symbols, i.e.,

$$I_t = \sum_{i=1}^{M} N\, p_i I(s_i) \quad \text{bits.} \tag{33}$$

Eg. Source with symbols 0 and 1, with probabilities p_1=6/8, p_2=2/8, N=8

| <-- N = 8 --> |

$$I \quad = \sum_{i=1}^{M=2} N\, p_i I(s_i)$$
$$= 8(6/8) \log_2 (8/6) + 8(2/8) \log_2(8/2) \text{ bits}$$
$$= 2.49 \text{ bits} + 4 \text{ bits.}$$

The average information per symbol equals $I_t/N = H$.

$$H = \sum_{i=1}^{M} p_i \log_2 p_i \text{ bit/symbol.} \tag{34}$$

We can see that H is obtained on time-averaging concept because symbols were transmitted one after the other in real time-domain.

H is called the source entropy.

H is a measure of uncertainty of the source, since it is a maximum when $p_1 = p_2 = \ldots = p_M$; that is, when the uncertainty as to which symbol will occur next is at a maximum since no symbol is more likely than any other.

Eg. Consider a pack of cards, arranged in order of suit and magnitude. If used as a source the ordered pack has zero uncertainty; if shuffled, there is maximum uncertainty, i.e., maximum average information.

The ratio of H/H_{max} is called the source efficiency or relative entropy.

H/H_{max} -> source efficiency or relative entropy.

Eg. Source with symbols 0 and 1, with probabilities $p_1 = 0.25$ and $p_2 = 3/4$.

Then H $= -(1/4) \log_2(1/4) - (3/4)\log_2(1/4)$
$= 0.811$ bit/symbol.

If $p_1 = p_2 = 0.5$,

$H_{max} = (1/2) \log_2(1/2) - (1/2) \log_2(1/2)$
$H_{max} = 1$ bit/symbol.

Consider a source emitting two symbols with probabilities $p_1 = p$ and $p_2 = 1\text{-}p$ where $0 < p < 1$.

$$H = p_1 \log_2(1/2) + p_2 \log_2 [1/(1-p)] \quad (35)$$

H reaches its maximum value when $dH/dp = 0$ and $p = 1/2$, where $H_{max} = 1$ bit/symbol.

Figure 4

A plot of equation (35) with $p_1 = p$ is shown in Figure 4.

When the symbols are emitted at a rate of r_s symbols per second, the average source information rate - R is

$$R = r_s H \text{ bit/second}. \quad (36)$$

Case 2 : Entropy of symbols in long dependent sequences

We have seen the source entropy of symbols in long independent sequences. In practice, source does not emit symbols in a statistically independent way.

For example, the English symbol E occurs more often than the symbol Y in a letter or a word. It can be seen that the dependence reduces the amount of information generated by the source.

Markoff statistical model may be used to calculate such source information rate and its entropy.

Markoff Statistical Source Model

For the purpose of analysis, we will assume that a discrete information source emits symbols once every T seconds according to certain probabilities. These probabilities depend on previously emitted symbols as well as the current symbol.

Further, the source has assumed to be an ergodic source and therefore the source is also stationary.

Ergodic implies stationary (not true in reverse direction)

Eg. 4 waveforms $x_1(t), x_2(t), \ldots, x_4(t)$

Figure 5

Time average - we take a function from a collection of functions and average over a defined time interval.

Ensemble average - we take the value of each function at time $t= \tau$ and average the values.

A process is said to be ergodic if the time average equals to the ensemble average. Modelling of a source can be viewed as a class of random process. The process is known as Discrete Stationary Markoff process and has the following properties:

1. The source has n states, $S_1, S_2, \ldots, S_n$.

2. The source is in one of n states.

3. As the source changes its state from i to j, a symbol is emitted and the probability of this occurrence (transition) is p_{ij}, which only depends on the initial state S_i and the final states S_j. It does not depend on the preceding states. Of course p_{ij} is a fixed constant.

$$S_i \xrightarrow{p_{ij}} S_j.$$

4. $\sum_{j=1}^{n} p_{ij} = 1.$

Figure 6

5. For a set of M symbols $\{s_1, s_2, ..., s_M\}$, the probability of emitting the k-th symbol in a sequence X_k, which is s_q, is

$$P(X_k = s_q / S_k) = P(X_k = s_q / X_1 X_2 \ldots X_{k-1}). \tag{37}$$

That is, this probability depends only on the state before emitting s_q.

Example : M=3, k=5, q=1 and $\{s_1, s_2, s_3\}$ maps to {Letters A, B, C}, resp.

Figure 7

6. Before emitting the first symbol from the source, the system is in one of the n states with probabilities $P_1(1)$, $P_2(1)$, $P_3(1)$, ... $P_n(1)$ and

$$\sum_i P_i(1) = 1.$$

Example:

Figure 8

A Markoff source can be represented by a state transition or a tree diagram. The above example is called a tree diagram and can be represented by a state transition diagram of the following form.

Figure 9

In general, the probability of the source emitting a particular symbol sequence is the sum of the product of probabilities in the tree diagram along all paths that yield the sequence.

Figure 10

Entropy and Information Rate of Markoff Sources

We only consider the Ergodic Markoff random process, i.e., the probability of being in state i at the beginning of the k-th symbol interval (before emitting k-th symbol) is the same as the probability of being in state i before $(k+j)$-th symbol (Figure 9).

The entropy of the source, H, is defined as the weighted average entropy of the symbols emitted from each state. With n states

$$H = \sum_{i=1}^{n} P_i H_i \ \text{bit/symbol} \tag{38}$$

where P_i is the probability that the source is in state i. Therefore $P_i H_i$ gives the contribution.

H_i - The entropy of state i is defined as the average information content of the symbols emitted from state i, i.e.

$$H_i = - \sum_{j=1}^{n} p_{ij} \log_2 p_{ij} \quad \text{bit/symbol} \tag{39}$$

where p_{ij} is the transition probability from state i to state j.

R - The average information rate of the source is defined as

$$R = r_s H \quad \text{bit/second} \tag{40}$$

where r_s is the symbol rate or the number of state-transitions per second.

Eg.1 Single symbol message

Figure 11

From equation (39),

$$H_1 = -p_{11} \log_2 p_{11} - p_{12} \log_2 p_{12}$$

$$H_2 = -p_{21} \log_2 p_{21} - p_{22} \log_2 p_{22}.$$

From equation (38)

$$H = P_1(1)H_1 + P_2(1)H_2.$$

Note that $P_i(k) = P_i(k+j)$ as mentioned before and this value remains unchanged. For convenience, we let $P_i(k) = P_i$.

Theorem 1

Let $P(m_i)$ be the probability of a sequence m_i of N symbols from a source. Let

$$G_N = -(1/N) \sum_i P(m_i) \log_2 P(m_i) \tag{41}$$

i.e.,

Figure 12

Then G_N is a monotonic decreasing function of N and

$$\lim_{N \to \text{infinity}} G_N = H \quad \text{bit/symbol}. \tag{42}$$

We will illustrate this concept by the following example.

Eg.2 Two-symbol message with the following source state diagram.

Figure 13

The above source emits two symbols and the tree diagram for sequences with two symbols is shown below.

Figure 14

To find $G_{N=2}$, we take the following steps.

Step 1 : Calculate the probability of each sequence of two symbols.

Eg. The sequence CC, the probability of that sequence in the upper part of the tree is

$P_1(1)p_{12}p_{21}$ = (1/3)(1/4)(1/4) = 1/48

and the probability of the sequence in the lower part of the tree is

$P_2(1)p_{21}p_{12}$ = (2/3)(1/4)(1/4) = 1/24.

The total probability of the sequence CC is

$P(CC)$ = (1/48) + (1/24) = 3/48.

Step 2 : Calculate the information content of each sequence, I(sequence).

$I(AA)$ = $-\log_2 P(m_i)$
=$-\log_2$(prob. of that sequence)
= $-\log_2$ 3/16

Eg. $I(AA) = -\log_2 3/16 = 2.42$ bits

$I(AC) = -\log_2 1/16 = 4$ bits

$I(CC) = -\log_2 3/48 = 4$ bits

$I(CB) = -\log_2 1/16 = 4$ bits

$I(CA) = -\log_2 2/16 = 3$ bits

$I(BC) = -\log_2 2/16 = 3$ bits

$I(BB) = -\log_2 6/16 = 1.42$ bits .

Step 3 : Calculate G_N where

$$G_N = -(1/N) \sum_i P(m_i) \log_2 P(m_i)$$

$= -(1/N)(P(AA)\ -[I(AA)] + \ldots)$
$= -(1/N)(P(AA) \log_2 P(AA) + \ldots)$
$= -(1/N)[(3/16)(2.42) + \ldots]$

$= 1.243$ bit/symbols message with N=2 symbols.

Exercise : Find G_1.

Summary

For long independent sequences, M symbols

1. H --> average information per symbol, source entropy

$$H = \sum_{i=1}^{M} p_i \log_2 p_i \quad \text{bit/symbol .}$$

2. H/H_{max} --> source efficiency or relative entropy .

3. R --> average source information rate

$$R = r_s H \qquad \text{bit/second .}$$

For long dependent sequences, n states

1. $H = \sum_{i=1}^{n} P_i H_i$ bit/symbol

   P_i --> Probability of the source at state i.

2. $H_i = -\sum_{j=1}^{n} p_{ij} \log_2 p_{ij}$ bit/symbol

   H_i --> average information content (source entropy) of the symbols emitted from state i.

3. $G_N = -(1/N) \sum_i P(m_i) \log_2 P(m_i)$

   $\lim_{N \to \infty} G_N = H$ bit/symbol

   G_N --> average information per symbol in a message containing N symbols.

4. $\lim_{N \to \infty} G_N = H$ bit/symbol

5. H/H_{max} --> source efficiency or relative entropy .

6. R --> average source information rate

   r_s --> symbol rate

   $R = r_s H$ bit/second.

7. It can be shown that the average information per symbol in messages containing N symbols decreases as N increases, i.e.

   $G_1 \geq G_2 \geq G_3 \geq \ldots \geq H$ bit/symbol. (43)

Transfer of Information

Figure 15

So far we have only considered the input, or source, of the system. In fact, the digital communication system shown in Figure 3 is degenerated into a much more simpler model. This is shown in Figure 15.

The amount of information available from source X can be calculated: how much is lost during transmission, storage or processing, i.e., how much arrives or is recovered? The general case is presented and an example is given.

Consider transmitting any one of m source symbols, which may be transformed by noise on the channel into one of n sink symbols ($m \leq n$ in general).

Figure 16

The source symbol probabilities are given by:-

$$p(x_1),\ p(x_2),\ \ldots,\ p(x_m).$$

Hence, the source entropy $H(X)$ is

$$H = \sum_{i=1}^{m} p(x_i) \log_2 p(x_i)\ .$$

The effect of channel noise on the source symbols can be measured, to give a set of conditional probabilities:-

$$p(y_1/x_1),\ p(y_1/x_2),\ \ldots,\ p(y_1/x_m);$$

i.e., the probabilities of y_j being received given that x_i was transmitted where $1 \leq i \leq m$ and $1 \leq j \leq$ n.

Note that for a particular value of i

$$\sum_{j=1}^{n} p(y_j/x_i) = 1,$$

since a source symbol has to go somewhere!

The average effect of noise on the symbols may now be found; that is, the average amount of information associated with the conditional probabilities can be calculated, to give the error entropy.

Step 1

Averaging over j for a given x_i:

$$H(Y/x_i) = -\sum_{j=1}^{n} p(y_j/x_i)\log_2 p(y_j/x_i)\ . \tag{44}$$

Step 2

Averaging over i:

$$H(Y/X) = \sum_{i=1}^{m} p(x_i)\, H(Y/x_i) \tag{45}$$

$$= -\sum_{i=1}^{m}\sum_{j=1}^{n} p(x_i)p(y_j/x_i)\log_2 p(y_j/x_i).$$

$H(Y/X)$ --> Error Entropy or Conditional Entropy.

Equivocation - $H(X/Y)$

It is also possible to define another conditional entropy $H(X/Y)$ in terms of the conditional probabilities $p(x_i/y_j)$.

$$H(X/y_j) = -\sum_{i=1}^{m} p(x_i/y_j)\log_2 p(x_i/y_j) \tag{46}$$

and

$$H(Y/X) = \sum_{j=1}^{n} p(y_j)\log_2 p(X/y_j) \tag{47}$$

$$= -\sum_{i=1}^{m}\sum_{j=1}^{n} p(y_j)p(x_i/y_j)\log_2 p(x_i/y_j).$$

Sink Symbol Probabilities

The sink symbol probabilities, $p(y_j)$, are given by

$$p(y_j) = \sum_{i=1}^{m} p(y_j/x_i)p(x_i) \tag{48}$$

and obviously

$$\sum_{j=1}^{n} p(y_j) = 1 \;. \tag{49}$$

Sink Entropy - $H(Y)$

$$\underline{H(Y)} = - \sum_{j=1}^{n} p(y_j) \log_2 p(y_j) \;. \tag{50}$$

Similarly, one can find the source entropy $H(X)$

$$\underline{H(X)} = - \sum_{i=1}^{m} p(x_i) \log_2 p(x_i) \;. \tag{51}$$

For $m > 2$, the channel is called a discrete memoryless channel (DMC). The channel is memoryless because the occurrence of an error at a given interval does not affect the behaviour of the communication system at any other intervals.

Example $n = 3$, $m = 3$

Figure 17

$H(Y/X)$ $=-p(x_1)[p(1/1)\log p(1/1)+p(2/1)\log(2/1)+$
$p(3/1)\log p(3/1)]$
$-p(x_2)\ [p(1/2)\log p(1/2) + p(2/2)\log(2/2)$
$+ \mathrm{p}(3/2)\log p(3/2)]$
$-p(x_3)[p(1/3)\log p(1/3) + p(2/3)\log(2/3)$
$+ p(3/3) \log p(3/3)]$

$H(Y)$ $= -p(y_1)\log p(y_1) - p(y_2)\log p(y_2) -$
$p(y_3) \log p(y_3)$

$p(y_3)$ $= p(y_3/x_1)p(x_1) + p(y_3/x_2)p(x_2) +$
$p(y_3/x_3)p(x_3)$.

Further Interpretation of DMC

1. Assume $n = m$. When the channel shown in Figure 16 is error-free (noiseless), then the source symbols are transmitted to the sink without change.

$$p(y_j/x_i) = 1 \quad \text{for } i = j$$

and

$$p(y_j/x_i) = 0 \quad \text{for } i \neq j \ .$$

Therefore

$$H(Y/X) = 0$$

and

$$H(Y) = H(X)$$

There is no uncertainty about the output where the input is known.

Figure 16 becomes

$$\begin{array}{rcl} & 1 \;\_\_\_\_\_\_\_\_\_\_\_\_\; 1 & \\ H(X) & 2 \;\_\_\_\_\_\_\_\_\_\_\_\_\; 2 & H(Y) \\ & \cdot \qquad \cdot & \\ & \cdot \qquad \cdot & \\ & m \;\_\_\_\_\_\_\_\_\_\_\_\_\; n & \end{array}$$

2. At the other extreme, all the $p(y_j/x_i)$ are equal; that is, each source symbol has equal change of being converted by the noise. The source and the sink are statistically independent of each other. We can therefore view the sink as a random noise generator.

Thus,

$$p(y_j/x_i) = 1/n \qquad \text{for all } i \ \& \ j$$

$$p(y_j) = 1/n \ , \ p(x_i) = 1/m$$

and

$$H(Y/X) = H(Y) = \log_2 n.$$

This is the maximum value of $H(Y/X)$ can ever have.

$$H(Y) = -\sum_{j=1}^{n} p(y_j)\log_2 p(y_j)$$

$$= -\sum_{j=1}^{n} (1/n)\log_2(1/n)$$

$$= -\log_2(1/n)$$

$$= \log_2(n)$$

$$H(Y/X) = \sum_{i=1}^{m} p(x_i)H(Y/x_i)$$

$$= -\sum_{i=1}^{m=n} p(x_i)\sum_{j=1}^{n} p(y_j/x_i)\log_2 p(y_j/x_i)$$

$$= -\sum_{i=1}^{n} (1/n)\sum_{j=1}^{n} (1/n)\log_2(1/n)$$

$$= \log_2(n)\ .$$

Summary

$$H(X) = -\sum_{i=1}^{m} p(x_i)\log_2 p(x_i) \ \text{bit/symbol}$$

$$H(Y) = -\sum_{j=1}^{n} p(y_j)\log_2 p(y_j) \ \text{bit/symbol}$$

Source entropy - $H(X)$

Sink entropy - $H(Y)$

$$H(Y/x_i) = -\sum_{j=1}^{n} p(y_j/x_i)\log_2 p(y_j/x_i)$$

$$H(Y/X) = \sum_{i=1}^{m} p(x_i)H(Y/x_i)$$

Conditional entropy - $H(Y/X)$
Error entropy

$$H(X/y_j) = -\sum_{i=1}^{m} p(x_i/y_j)\log_2 p(x_i/y_j$$

$$H(X/Y) = \sum_{j=1}^{n} p(y_j)H(X/y_j)$$

Equivocation - $H(X/Y)$

Sink symbol probabilities - $p(y_j)$
Source symbol probabilities - $p(x_i)$

$$p(y_j) = \sum_{i=1}^{m} p(y_j/x_i)p(x_i)$$

$$p(x_i) = \sum_{j=1}^{n} p(x_i/y_j)p(y_j)$$

$$\sum_{j=1}^{n} p(y_j) = 1 \qquad \sum_{i=1}^{m} p(x_i) = 1$$

Mutual Information

The entropy $H(Y/X)$ represents the amount by which $H(X)$ has been distorted during transmission. In other word, it is the amount of information $H(Y)$ that one cannot be relied upon.

Thus, the amount of information at the sink, represented by $H(Y)$, must be reduced by the uncertainty caused by noise, $H(Y/X)$ - the error entropy, to give the true amount of information which reaches the sink. The mutual information or the trans-information is

$$I(X;Y) = H(Y) - H(Y/X)\ . \qquad (52)$$

$I(X;Y)$ can also be derived by noting that the information emitted by the source X, $H(X)$, is reduced by the loss in information $H(X/Y)$ suffered by the source symbols during transmission over the noisy channel.

$$I(X;Y) = H(X) - H(X/Y)\ . \qquad (53)$$

In our early examples:

1. Noiseless channel -- $I(\mathrm{X};Y) = H(Y) - H(Y/X)$

$= H(Y)$.

2. Very noise channel -- $p(y_j/x_i) = (1/n)$

$$I(\mathrm{X};Y) = H(Y) - H(Y/X) = 0.$$

Summary

$I(X;Y)$ - mutual/trans-information

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

$$= H(X) - H(X/Y)$$

Channel Capacity

1. Discrete Memoryless Channel

The concepts presented above can conveniently be summarised in the following diagram:-

Figure 18

In a practical communication system, $p(y_j/x_i)$ are fixed for a given discrete channel unless the noise characteristics are time-varying.

The channel capacity C is defined as

$$C = \max_{p(x_i)} \{I(X;Y)\} \tag{54}$$

$$C = \max_{p(x_i)} \{H(Y) - H(Y/X)\}$$

$$= \max_{p(x_i)} \{H(X) - H(X/Y)\}.$$

In general, C is a function of source and the noise statistics. The channel redundancy and the channel efficiency can be defined as

Channel Redundancy $= C\text{-}I(X;Y)$ bit/symbol (55)

$= [C- I(X;Y)] / C\,\%$

Channel Efficiency $= I(X;Y) / C\,\%$ (56)

Note that C is expressed in bit/symbol. If the symbol rate is r_s symbols/second, C can be expressed in bit/second by multiplying r_s to equation (54).

2. Binary Symmetric Channel

Binary symmetric channel is a sub-class of discrete memoryless channel. The ideas developed in the previous sections will now be applied to this particular transmission channel.

Figure 19

The BSC has two input (source) symbols, and two output (sink) symbols. The probability of a symbol being received correctly, q, is the same for each symbol, as is the probability that each symbol may be received incorrectly, p; hence the term symmetric.

p -- cross-over probability or error rate of the BSC.

Let the source probabilities be

$p_0 = \alpha$, $p_1 = 1- \alpha$.

From equation (45),

$$H(Y/X) = -p(x_0)p(y_0/x_0) \log_2 p(y_0/x_0)$$

$$-p(x_1)p(y_0/x_1) \log_2 p(y_0/x_1)$$

$$-p(x_0)p(y_1/x_0) \log_2 p(y_1/x_0)$$

$$-p(x_1)p(y_1/x_1) \log_2 p(y_1/x_1)$$

$$= -\alpha q \log_2 q - (1-\alpha)p \log_2 p - \alpha p \log_2 p -$$

$(1-\alpha)q \log_2 q$

$= -p\log_2 p - q\log_2 q$

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

From equation (50),

$H(Y) = -p(y_0) \log_2 p(y_0) - p(y_1) \log_2 p(y_1)$

Because $p(y_0)$ $= p(y_0/x_0)p(x_0) + p(y_0/x_1)p(x_1)$

$= (1-p)\alpha + p(1-\alpha)$

$= \alpha + p - 2\alpha p$

and

$p(y_1)$ $= p(y_1/x_0)p(x_0) + p(y_1/x_1)p(x_1)$

$= 1 - \alpha - p + 2\alpha p.$

Therefore, $H(\mathrm{Y})$ $= -(\alpha + p - 2\alpha p) \log_2(\alpha + p - 2\alpha p) -$

$(1-\alpha-p+2\alpha p)\log_2(1-\alpha-p+2\alpha p)$.

Thus,

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

$= H(Y) + p \log_2 p + (1-p) \log_2 (1-p)$.

max$\{H(Y)\}$ when $p(y_0) = p(y_1) = 1/2$ and $H(Y) = 1$.

$p(y_0) = p(y_1)$

$\Rightarrow \alpha + p - 2\alpha p = 1/2$

$1 - \alpha - p + 2\alpha p = 1/2$

Therefore, $\alpha = 1/2$.

The result is that capacity is at maximum when $p(x_0) = p(x_1) = \alpha = 1/2$. This is a characteristic of symmetric channel.

$$C = 1 + p \log_2 p + (1-p) \log_2 (1-p) .$$

Figure 20

BSC is a useful model which approximates the behaviour of many practical binary channels.

Other types of channel models are

1. Binary Erasure Channel (BEC)

Figure 21

2. Symmetric Erasure Channel (SEC), a combination of BSC and BEC.

Figure 22

3. Continuous Channel

Figure 23

Theorem 2 (Shannon's Theorem)

Given a source of M equally likely messages, with $M >> 1$, which is generating information at a rate of R bits per second. If $R \leq C$, the channel capacity, there exists a channel coding technique such that the communication system will transmit information with an arbitrarily small probability of error.

Theorem 3 (Shannon-Hartley Theorem)

For a white, bandlimited Gaussian channel, the channel capacity is

$$C = B \log_2 [1 + (S/N)] \text{ bit/second} \tag{57}$$

where

S - average signal power at the continuous channel output

N - average noise power at the continuous channel output.

$$N = (N_o/2)\,(2B) \quad \text{Watts} \tag{58}$$

$N_o/2$ - two sided power spectral density (PSD) of the noise in Watts/Hz

B - channel bandwidth.

Figure 24

Equation (57) can be found by using noiselike signals and maximised the information rate R with respect to the noiselike signals. Hence, we consider the communication system as a random process.

Equation (57) often provides a lower bound on the performance of a system operating over a non-Gaussian channel, i.e., $C_{\text{(datum line)}} \leq C_{\text{non-Gauss}}$.

Example

Given a sequence of messages which can be represented by a waveform $s(t)$ as shown in Figure 25, find C if there are M possible messages, the noise power $N = \sigma^2$ and $r_s = 1/T = 2B$ messages/second.

Figure 25

M possible messages => M possible levels.

If the messages are equally likely,

$$S = 2\{(1/M)[(\lambda\sigma^2/2)^2 + (3\lambda\sigma^2/2)^2 + \ldots + ((M-1)\lambda\sigma^2/2)^2]\}$$

$$= [(M^2-1)/12](\lambda\sigma)^2$$

$$M = (1 + (12/\lambda^2)(S/N))^{1/2}$$

$$H = \sum_{i=1}^{M} p_i \log_2 p_i$$

$$= \sum_{i=1}^{M} (1/M) \log_2 (1/M)$$

$$= \log_2 M \qquad \text{bit/message .}$$

Therefore, $R = r_s H$

$$= B \log_2 (1 + (12/\lambda^2)(S/N) \text{ bit/second} \qquad (59)$$

For acceptable probability of error with max. R, we simply take the value of R as the approximated value of C. With $\lambda = 3.5$, equation (59) becomes equation (57).

Equation (57) - With channel coding, transmission at C is possible with arbitrarily free of errors.

Equation (59) - Specifies the information transmission rate with SMALL error.

From the channel capacity expression, equation (57), a noiseless channel has an infinite capacity. On the other hand, while the channel capacity increases, it does not go to infinity as B tends to infinity. This is because the noise power also increases.

Thus for a fixed signal power in the presence of white Gaussian noise, C becomes

$$C_{\text{infinity}} = \lim_{B \to \text{infinity}} B \log_2 [1 + (S/N_o B)]$$

$$= \lim_{B \to \text{infinity}} (S/N_o) \log_2 [1 + (S/N_o B)]^{N_o B/S}$$

$$= (S/N_o) \log_2 e$$

$$C_\infty = 1.44 (S/N_o) \quad \text{bit/second.} \qquad (60)$$

This expression gives the best possible channel capacity and is taken as a reference datum line for the design of a practical communication system.

In practice, we maximise the transmission rate over the continuous portion of the channel.

For the discrete portion of the channel shown in Figure 23, we try to maximise the data rate such that it is approaching the discrete channel capacity with less error probability, using error-control coding (channel coding).

Implication of Shannon-Hartley Theorem

1. Equation (57) represents the upper limit that one can reliably transmit signals over Gaussian channels.

2. It indicates that we may trade-off bandwidth for signal-to-noise ratio.

Example 1

If $S/N = 7$ $B = 4$ kHz $C = 12$ k bit/second .

If $S/N = 15$ $B = 3$ kHz $C = 12$ k bit/second .

With a 3 kHz bandwidth, the noise power will be 3/4 as large as with 4 kHz. Since, by equation (58)

$$N_3 \text{ kHz} / N_4 \text{ kHz} = (N_o . 3) / (N_o . 4)$$

and

$$(S/N)_{3 \text{ kHz}} / (S/N)_{4 \text{ kHz}} = 15/7.$$

Therefore,

$$S_{3 \text{ kHz}} / S_{4 \text{ kHz}} = (3/4)\ (15/7).$$

The signal power will have to increase by (3/4) . (15/7) = 1.6. A bandwidth reduction results in an increase in signal power.

Example 2

Can we transmit an analog signal of bandwidth $B_S = f_S$ Hz over a channel having a bandwidth less than f_S Hz?

Suppose we sample the analog signal at 3 times the Nyquist sampling rate, i.e. 3 f_S samples/second and quantised the signal value into one of M levels. Then the data rate is $R = 3 f_S \log_2 M$ bit/second.

If the channel bandwidth is B bits per second, we can achieve a channel capacity $C \geq R$ by an appropriate choice of signal power S, using equation (57).

Say $M = 64$,

$R = 18\ f_s$ bit/second .

If $B = B_s/2 = f_s/2$ Hz for channel bandwidth ;

i.e., half the signal bandwidth.

From equation (57) and let $R = C$, then

$$18\, f_s = (f_s/2) \log_2 [1 + (S/N)]$$

$$S/N \simeq 109 \text{ dB.}$$

Summary

$$C = B \log_2 [1 + (S/N)] \quad \text{bit/second}$$

$$N = (N_o/2)(2B) \quad \text{Watts .}$$

$$C_\infty = 1.44\, (S/N_o) \quad \text{bit/second .}$$

References

[1] Shanmugam, K. S., ‘Digital and Analog Communication Systems’, John Wiley & Sons, 1979.

[2] Taub, H. and Schilling, D. L., ‘Principles of Communication Systems’, 2/e, McGraw-Hill, 1987.

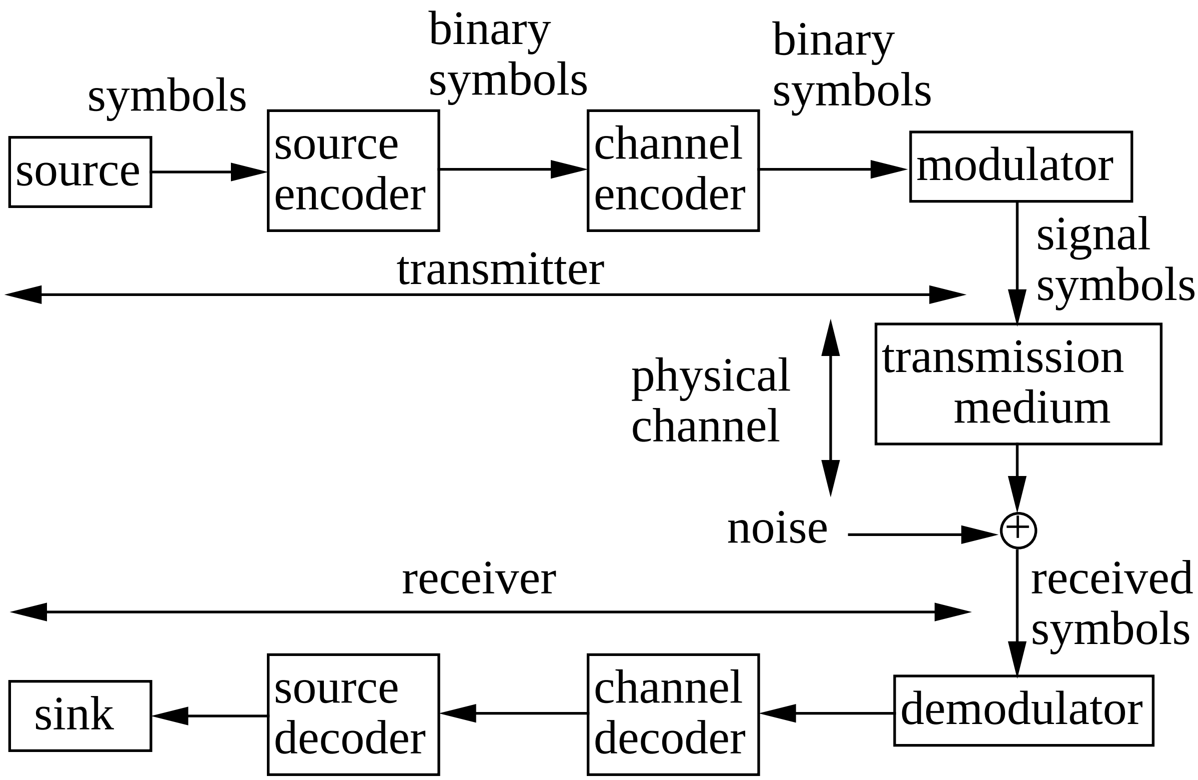


Figure 3

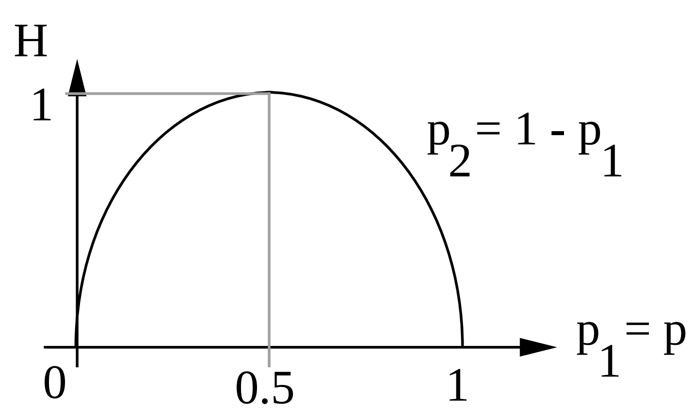


Figure 4

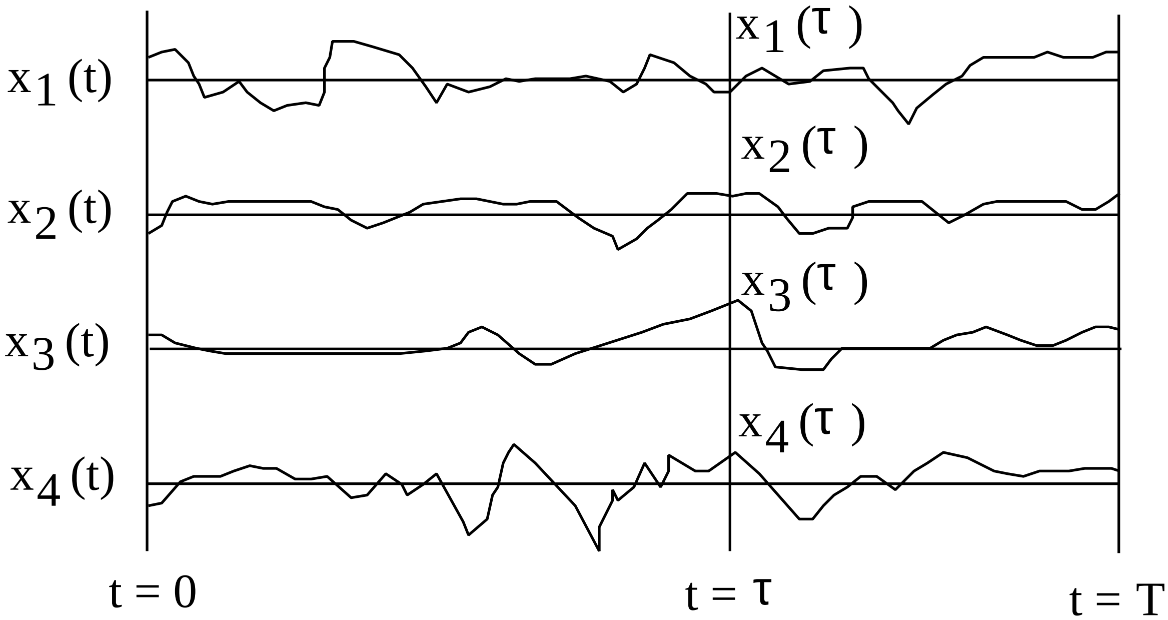


Figure 5

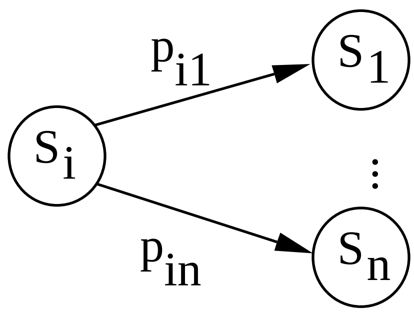


Figure 6

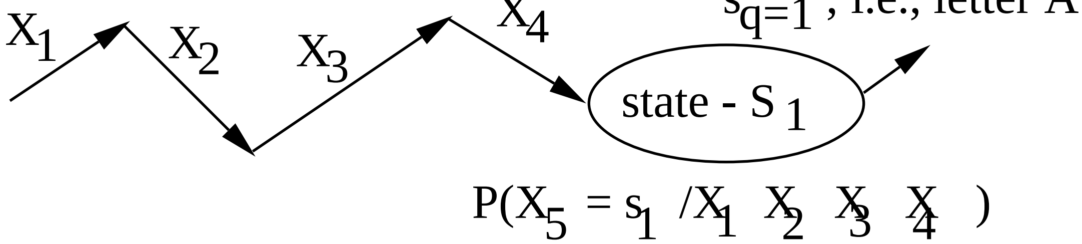
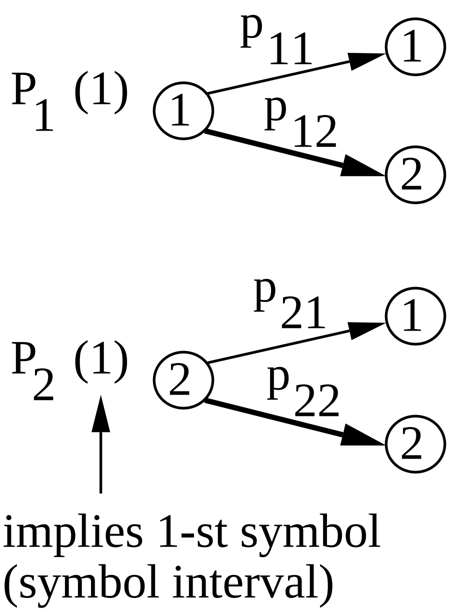


Figure 7



$$P_2(2) = P_1(1)p_{12} + P_2(1)p_{22}$$

$$P_1(2) = P_1(1)p_{11} + P_2(1)p_{21}$$

$$p_{11} + p_{12} = 1$$

$$p_{21} + p_{22} = 1$$

Figure 8

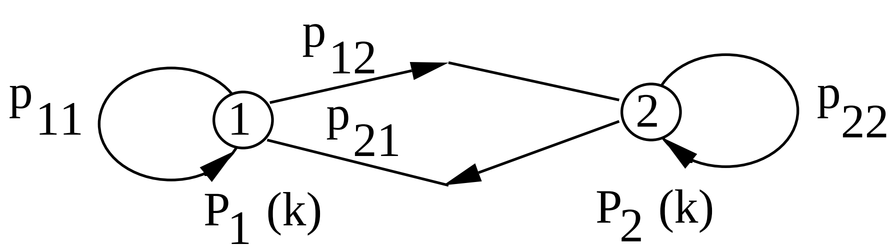
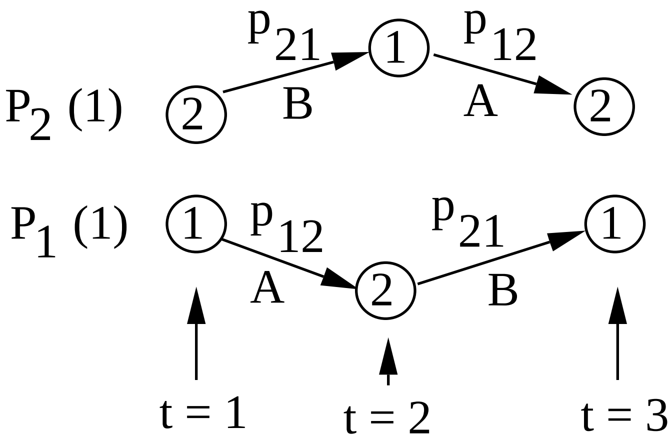


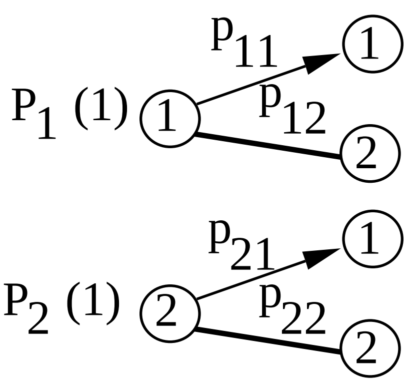
Figure 9



P(symbols AB)

$$= P_2(1)p_{21}\,p_{12} + P_1(1)p_{12}p_{21}$$

Figure 10



$P_1(1) = 1/3$

$P_2(1) = 2/3$

$p_{11} = 3/4$ $p_{21} = 1/4$

$p_{12} = 1/4$ $p_{22} = 3/4$

Figure 11

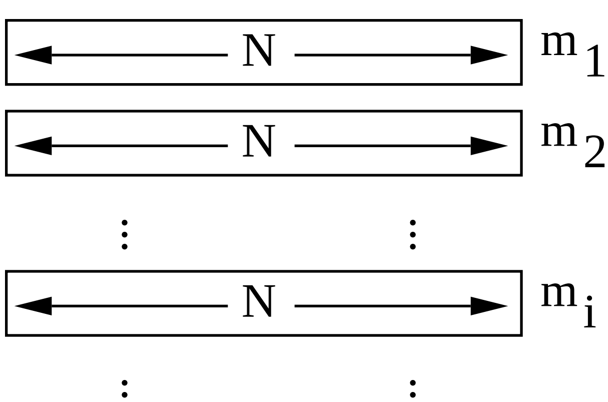


Figure 12

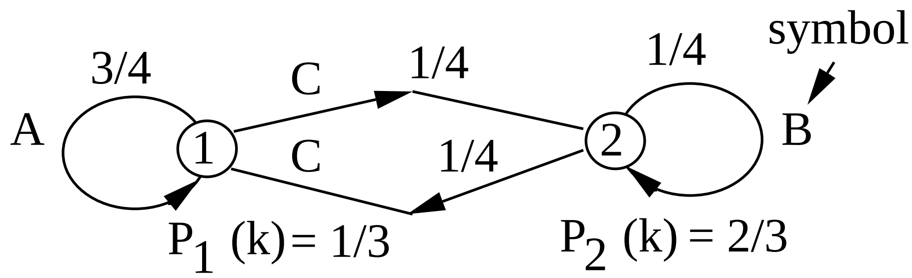


Figure 13

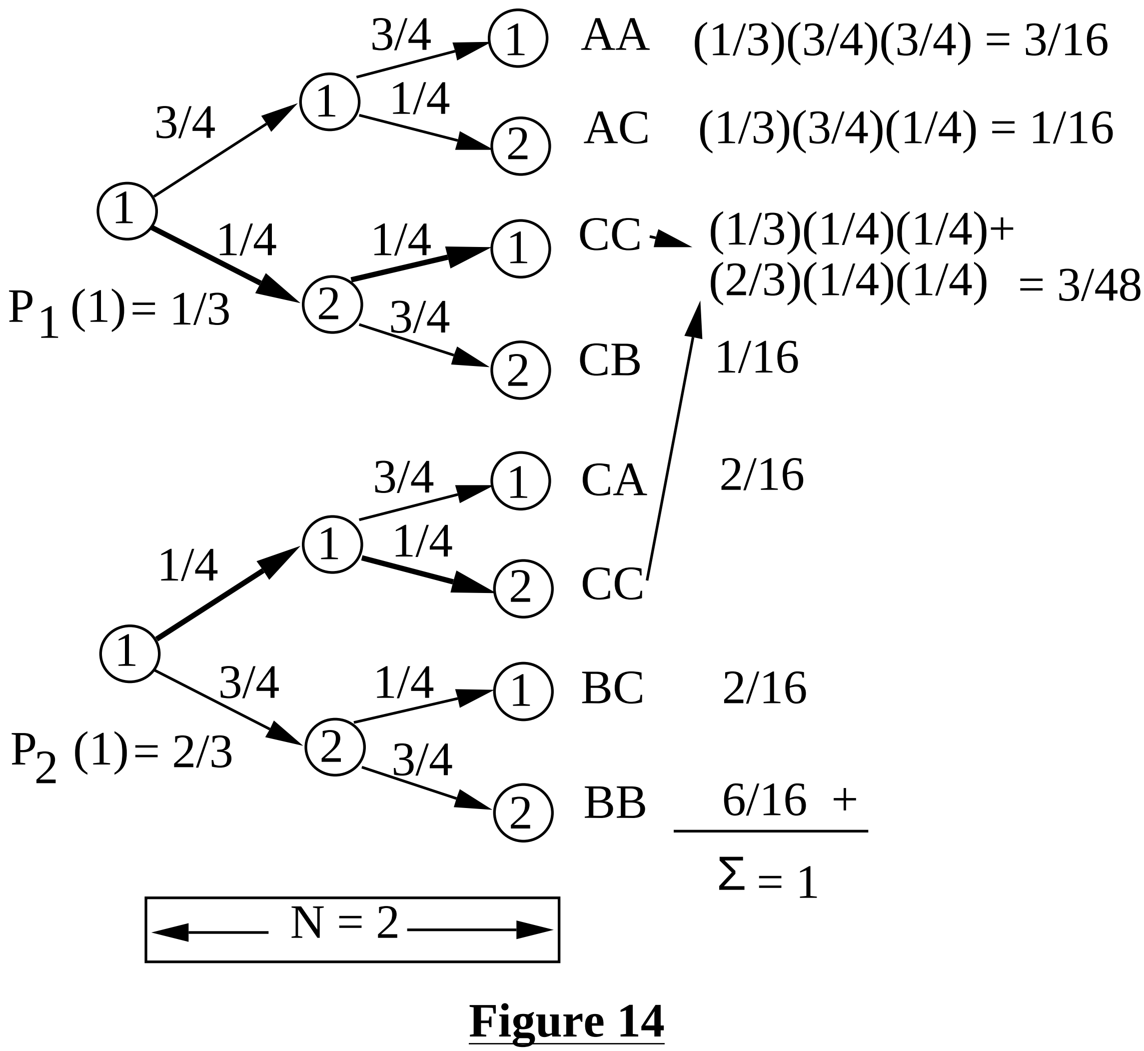


Figure 14

Figure 15

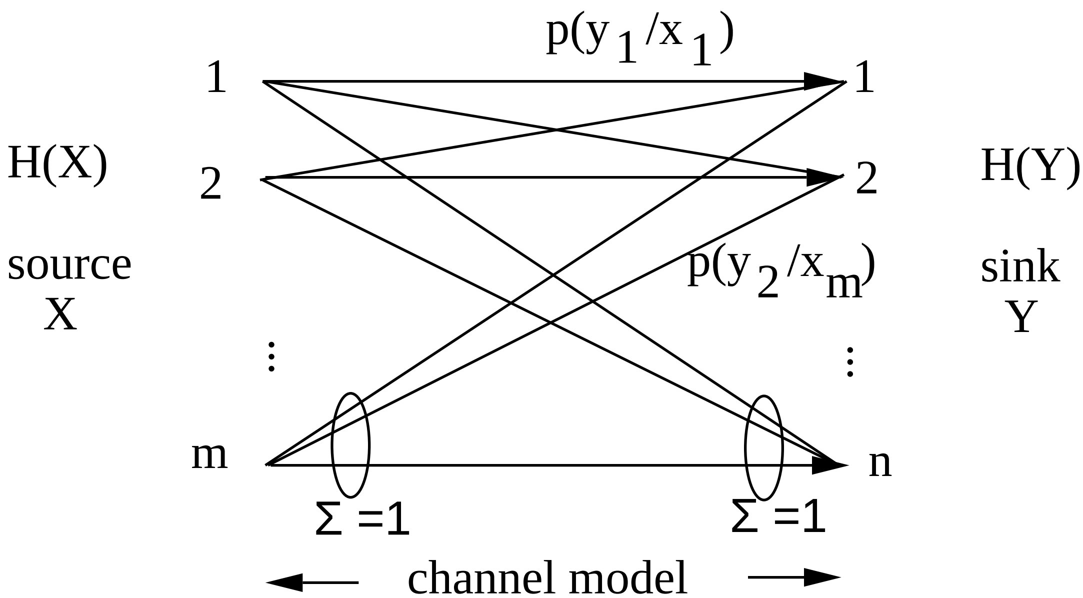


Figure 16

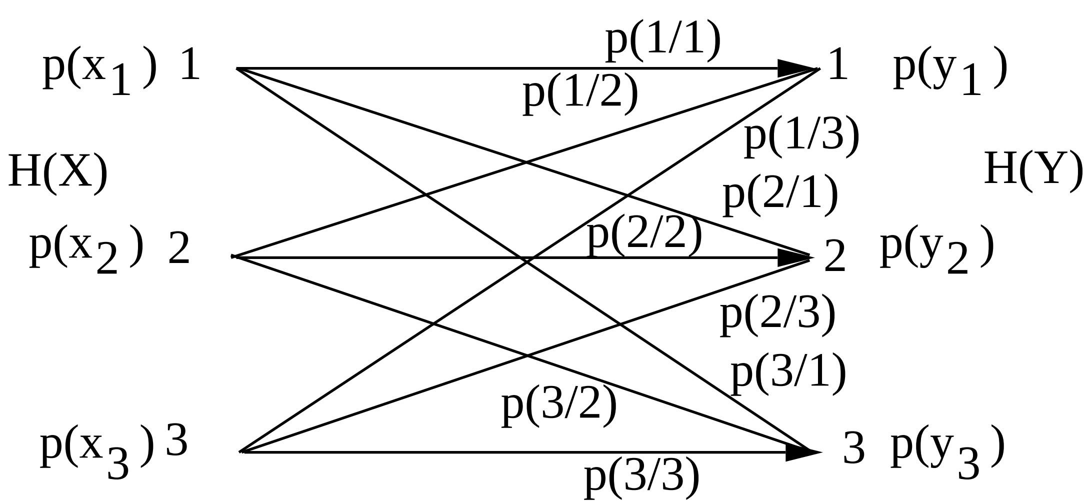


Figure 17

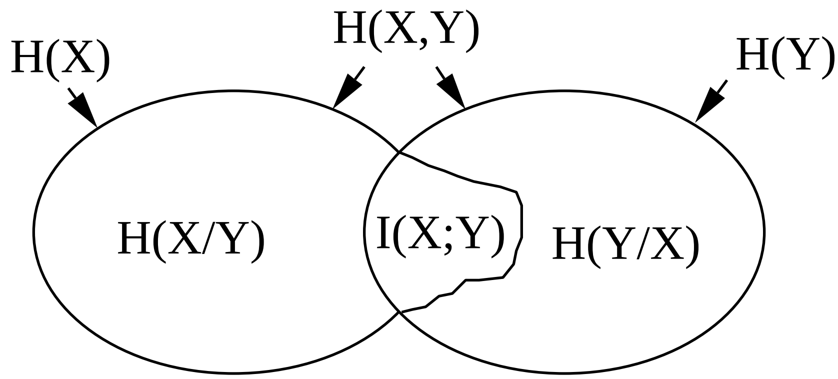


Figure 18

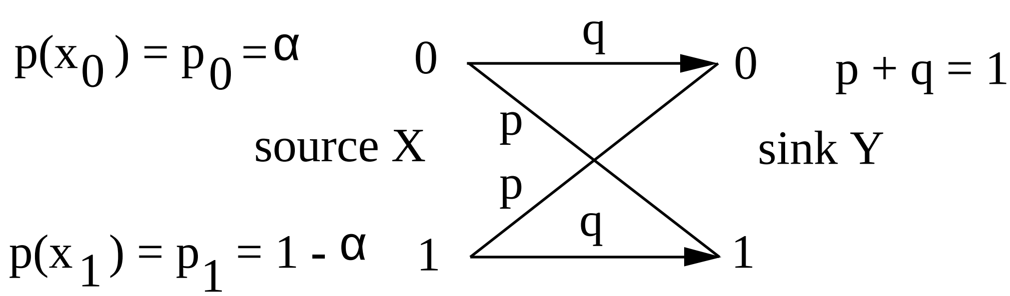


Figure 19

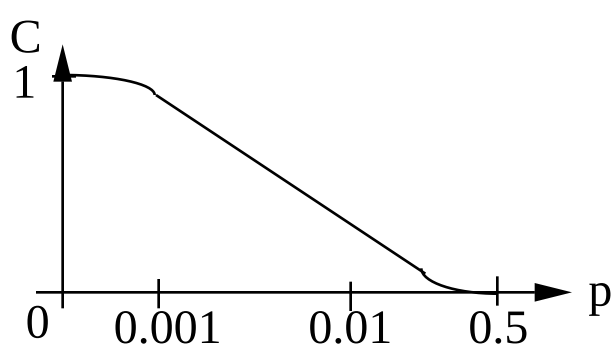


Figure 20

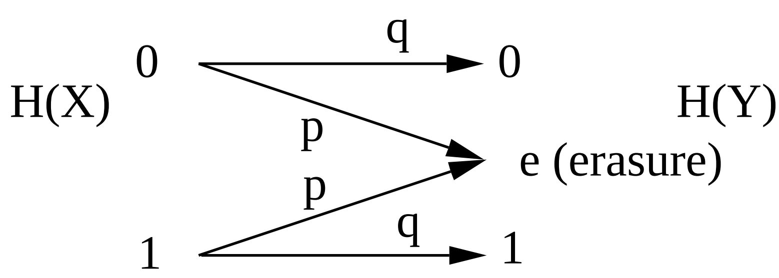


Figure 21

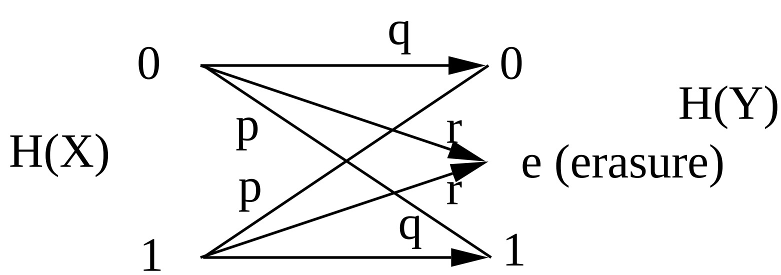


Figure 22

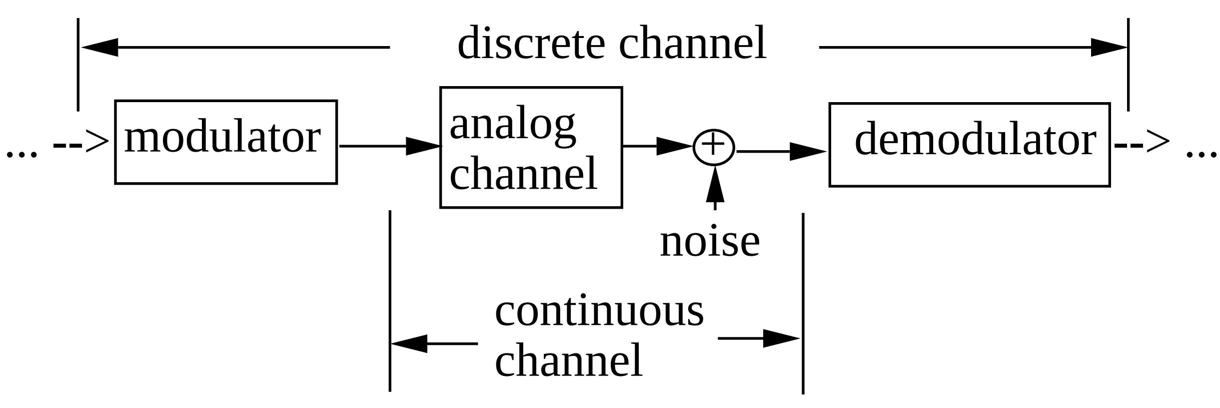


Figure 23

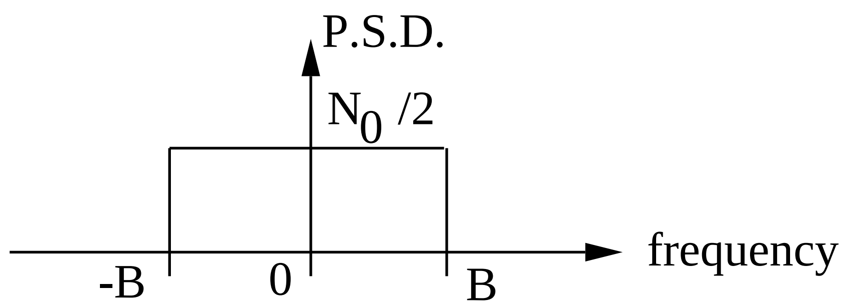


Figure 24

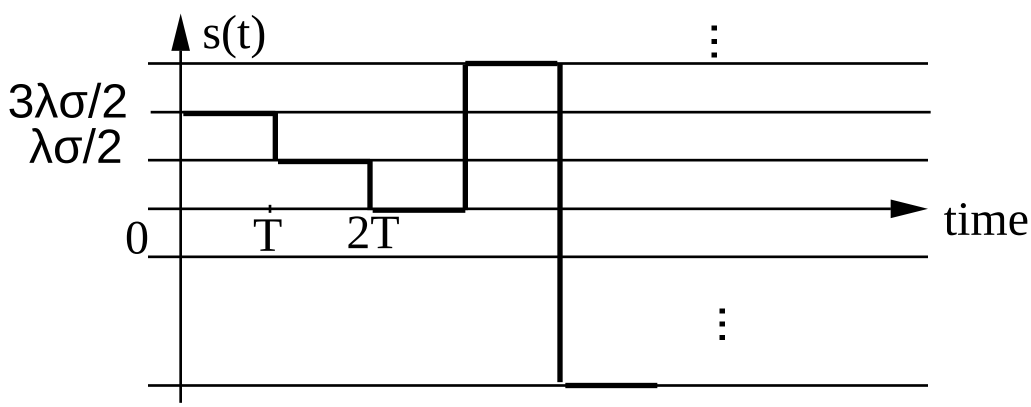


Figure 25