Coding Theory

Lector: Nikolai L. Manev

Institute of Mathematics and Informatics, Sofia, Bulgaria

A Bit of History

The twin disciplines of information theory and coding theory date back to the Claude Shannon's paper "A Mathematical Theory of Communication" written in 1948.
Indeed Shannon gave birth to the information theory since in his paper he considered the information transmission from a general aspect.
At the same time (1947-1948) Richard Hamming followed a more pragmatic approach, focusing on the construction of error detecting and error correcting codes. Because of patent considerations he was unable to announce his result until 1950 when his seminal paper "Error Detecting and Error Correcting Codes" was published.

Goals

The development of both theories have been motivated by the need of efficient and reliable communications in an uncooperative (and possibly hostile) environment, but they present two different directions of solving the problem.

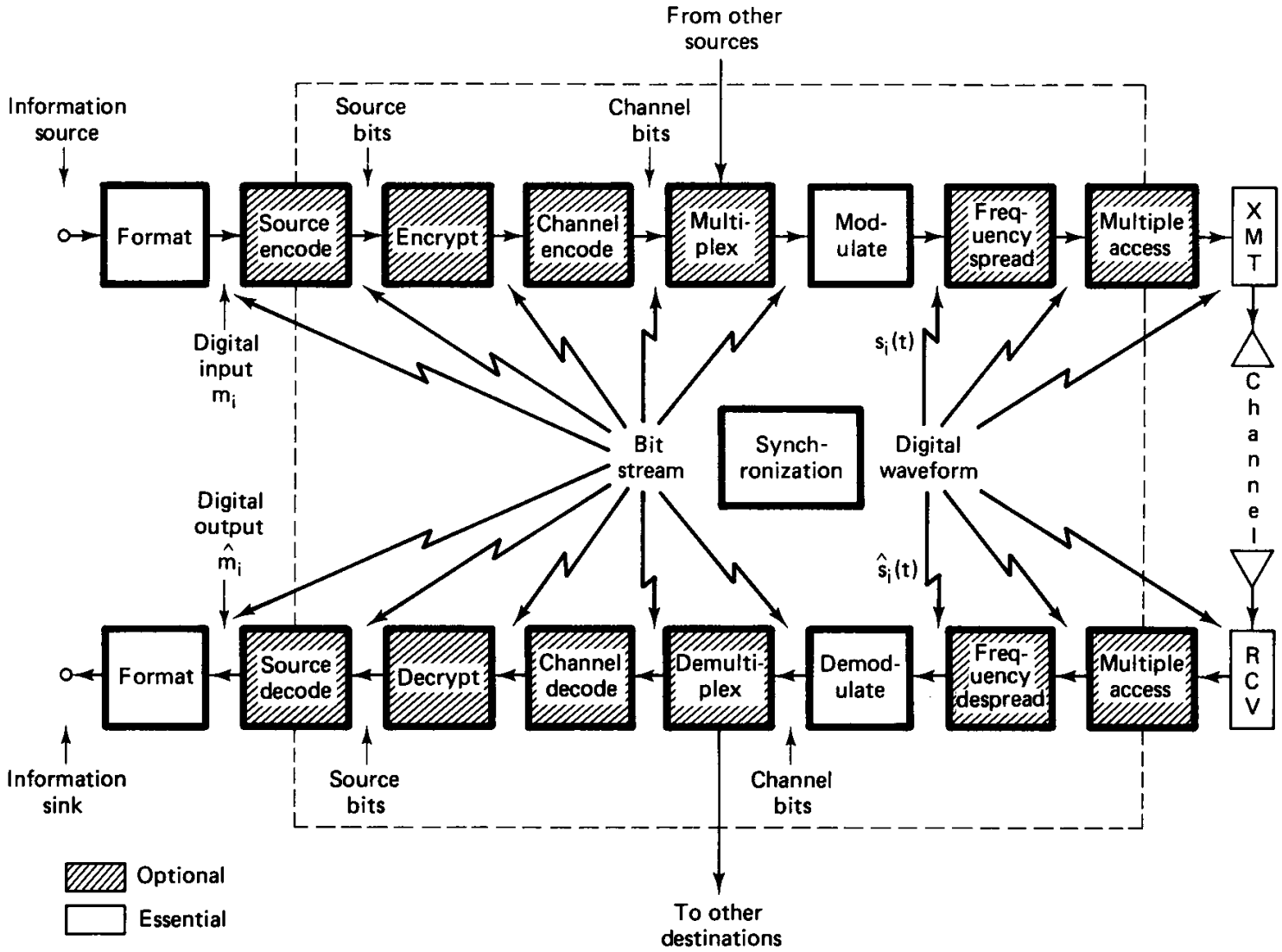
- *Information theory* is the study of achievable bounds for communication and is largely probabilistic and analytic in nature.
- In contrast, *coding theory* attempts to realize the promise of these bounds by models which are constructed through mainly algebraic means.

The Elements of a Digital Communication system

The modern communication systems are *digital*. It means that at a given stage of transmitting they use a sequence of symbols from a finite alphabet to represent the information. The transmission of data in digital form allows the use of a number of powerful information processing techniques that would otherwise be unavailable.

Detailed block diagram of a modern digital communication system is given in the book of Sklar.

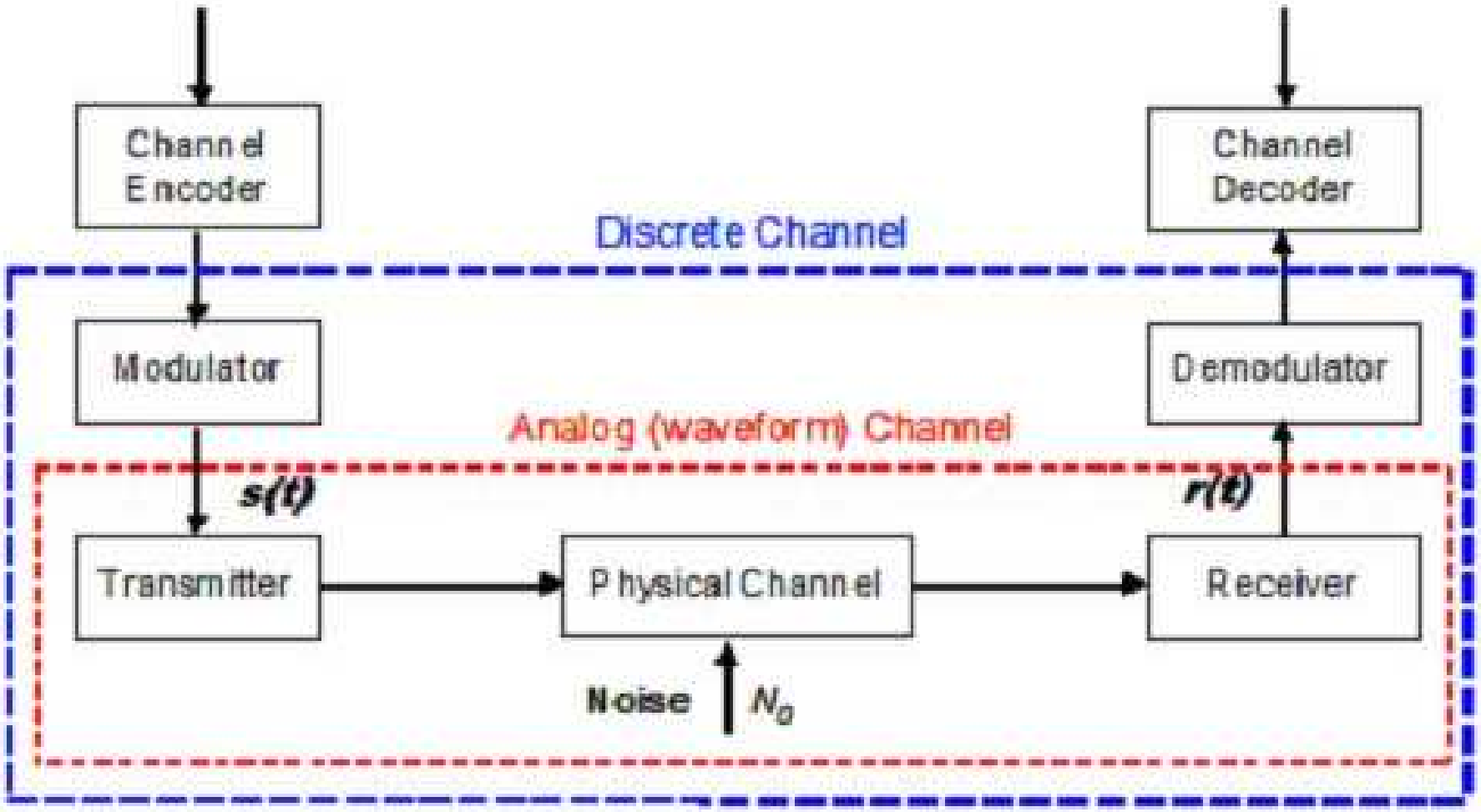
The Elements of a Digital Communication system



The Elements of a Digital Communication system

- Formatting and source coding: transform source data into a suitable digital form and remove the uncontrolled redundancy that naturally occurs in most source information stream.
- Encryption: If it is necessary some measure to prevent unauthorized access to the transmitted data to be taken.
- Modulator/Demodulator: It maps digital symbols onto signals that can be efficiently transmitted over the communication channel.

The Elements of a Digital Communication system



The Elements of a Digital Communication system

- Analog (continuous) channel: the part of communication chain between the modulator's output and the input of the demodulator. The information stream flowing through this part is represented by waveform signals.
- Discrete channel: The enlargement of the analog channel obtained by including the modulator and demodulator. Both the input and output data streams of discrete channel have digital format.

In both cases the influence of the channel is described in the terms of the probability theory: the channel transforms one random variable into another

$$\text{input } X \overset{channel}{\longrightarrow} Y \text{ output}$$

The Elements of a Digital Communication system

Communication channel and their behavior are subject of the information theory but some basic knowledge is necessary in order to make sense of the constructions in coding theory. Although we shall mainly concentrate on mathematics of coding theory we have always to keep in mind the fundamental bounds of information theory and the practical desires of engineering.

Analog (continuous) channel 1

Signals transmitted through the analog channel are mathematically described by real or complex functions in the time. The Fourier transform of these functions into frequency domain expresses the fact that the signals are composed from periodic signals with different frequencies.

Analog (continuous) channel 1

Signals transmitted through the analog channel are mathematically described by real or complex functions in the time. The Fourier transform of these functions into frequency domain expresses the fact that the signals are composed from periodic signals with different frequencies.

The physical channel attenuates the transmitted signal and introduces noise. The attenuation is generally caused by energy absorbtion and scattering in the propagation medium. The most common noise source is the ambient heat in the transmiter/receiver hardware and the propagation medium.

The mathematical tools for modeling noisy characteristic of the analog channel are continuous probability distributions.

Continuous probability distribution 1

Let $X = X(s)$ be a random variable taking real values, when s runs through the space of events (e.g.,possible outputs of an experiment). The *cumulative distribution function* (**cdf**), also called *probability distribution function or just distribution function* of X, is defined by

$$F_X(x) \stackrel{\text{def}}{=} \Pr(X \leq x),\ x \in \mathbb{R}.$$

Continuous probability distribution 1

Let $X = X(s)$ be a random variable taking real values, when s runs through the space of events (e.g.,possible outputs of an experiment). The *cumulative distribution function* (**cdf**), also called *probability distribution function or just distribution function* of X, is defined by

$$F_X(x) \stackrel{\text{def}}{=} \Pr(X \leq x),\ x \in \mathbb{R}.$$

Properties:

1. $0 \leq F_X(x) \leq 1$
2. $F_X(x_1) \leq F_X(x_2)$, if $x_1 \leq x_2$
3. $F_X(-\infty) = 0$
4. $F_X(\infty) = 1$.

Continuous probability distribution 2

The *probability density function* (**pdf**) of the random variable X is the function

$$p_X(x) \stackrel{\text{def}}{=} \frac{dF_X(x)}{dx}.$$

Continuous probability distribution 2

The *probability density function* (**pdf**) of the random variable X is the function

$$p_X(x) \stackrel{\text{def}}{=} \frac{dF_X(x)}{dx}.$$

Properties $p_X(x)$:

1. $p_X(x) \geq 0$
2. $\Pr(x_1 < X \leq x_2) = \int_{x_1}^{x_2} p_X(x)\,dx$
3. $\int_{-\infty}^{+\infty} p_X(x)\,dx = 1.$

Continuous probability distribution 3

The *mathematical expectation (or expected value, or mean)* of X is defined by

$$\mathsf{E}(X) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} x p_X(x)\, dx,$$

The *mathematical expectation (or expected value, or mean)* of X is defined by

$$\mathsf{E}(X) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} x p_X(x)\, dx,$$

Let m_X be the expected value of a random variable X. The *variance* of X, denoted by $\mathrm{Var}(X)$, or σ_X^2 is

$$\sigma_X^2 \stackrel{\text{def}}{=} \mathsf{E}\{(X - m_X)^2\} = \int_{-\infty}^{\infty} (x - m_X)^2 p_X(x)\, dx.$$

σ_X *standard deviation*.

Continuous probability distribution 4

A random variable X is said to have *Gaussian distribution*, or to be *Gaussian with mean* m *and variance* σ^2, if its **pdf** is defined by

$$p_X(x) \stackrel{\text{def}}{=} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Continuous probability distribution 4

A random variable X is said to have *Gaussian distribution*, or to be *Gaussian with mean* m *and variance* σ^2, if its **pdf** is defined by

$$p_X(x) \stackrel{\text{def}}{=} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Rayleigh distribution:

$$p_X(x) \stackrel{\text{def}}{=} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, \ x \geq 0.$$

Rice (Rician) distribution; Nakagami distribution.

Analog (continuous) channel 2

A channel model is referred to be an *additive white Gaussian noise* (AWGN) channel if its impact on the transmitted signal consists in adding a noise that is Gaussian random variable with zero mean and variance σ^2. Hence, it can be represented by

$$r(t) = s(t) + n(t), \text{ where } p(n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{n^2}{2\sigma^2}}.$$

AWGN is "white" in the sense that its spectral density has the constant (two-side) value, $N_o/2$, for all frequencies of interest in communication systems - from dc to about 10^{12} Hz. It can be calculated that the variance is

$$\sigma^2 = \frac{N_0}{2}.$$

Analog (continuous) channel 3

AWGN channel model describes accurately satellite and line-of-sight channels. For wireless mobile communications, where signal propagation takes place near the ground, as well as for indoor communications this model is inadequate. In these cases large fixed or moving objects (buildings, hills, cars, etc.) make the signals to travel over multiple reflected “paths". This effect, referred to as *(multipath) fading*, causes fluctuation in received signal’s amplitude, phase, and angle of arrival. A channel, subject to fading, is called *fading-channel*. A more adequate model for such a channel is given by

$$r(t) = \alpha(t)e^{i\theta}s(t) + n(t),$$

where $n(t)$ is a complex Gaussian noise, $\alpha(t)$ has Rayleigh distribution, and $\theta(t)$ is uniformly distributed in $[0, 2\pi]$.

Analog (continuous) channel 4

An important characteristic of the analog channel is *signal-to-noise ratio* (SNR). It is defined as the ratio of the energy of signal to the spectral density of the noise:

$$\frac{E_s}{N_0}.$$

It is measured usually in dB. ($E_{dB} = 10 \log_{10} E$)

Analog (continuous) channel 5

Analog channel separates time intervals of length T seconds such that the signal during each time interval presents one transmitted symbol. The number of transmitted symbols

$$R_s = \frac{1}{T}$$

per second is called *symbol rate*. The transmission of one symbol in the analog channel may corresponds to more then one symbols (e.g. several bits) across the discrete channel.

Discrete channel model

Now the input random variable X takes values from a finite alphabet $A = \{a_1, a_2, \ldots, a_m\}$ and the output random variable Y takes values from a finite alphabet $B = \{b_1, \ldots, b_n\}$. The channel impact is described by a stochastic matrix $\mathbf{P} = (p_{ij})$, where for $1 \leq i \leq m$, $1 \leq j \leq n$

$$p_{ij} \;=\; \Pr(Y = b_j \,|\, X = a_i), \qquad \text{and} \qquad \sum_{j=1}^{N} p_{ij} = 1.$$

Usually $A = B = \mathbb{F}$, where $\mathbb{F} = GF(q)$ is a finite field with q elements, e.g., $\mathbb{F} = \{0, 1\}$.

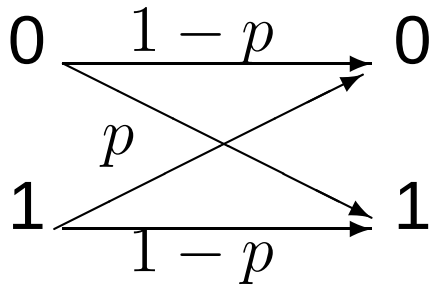
Discrete memoryless channel

Let $u_1, u_2, \ldots$ and $v_1, v_2, \ldots$ be the input, and output digital streams, respectively. That is, if X consecutively takes values $u_1, u_2, \ldots$ then Y takes $v_1, v_2, \ldots$ If the probabilities $\Pr(v_k = b_j \,|\, u_k = a_i)$ do not depend on the values $u_i,\ i < k$, then the channel is called *discrete memoryless channel (DMLC)*.

Discrete memoryless channel

Let $u_1, u_2, \ldots$ and $v_1, v_2, \ldots$ be the input, and output digital streams, respectively. That is, if X consecutively takes values $u_1, u_2, \ldots$ then Y takes $v_1, v_2, \ldots$ If the probabilities $\Pr(v_k = b_j \,|\, u_k = a_i)$ do not depend on the values $u_i,\ i < k$, then the channel is called *discrete memoryless channel (DMLC)*.

Binary Symmetric Channel (BSC):

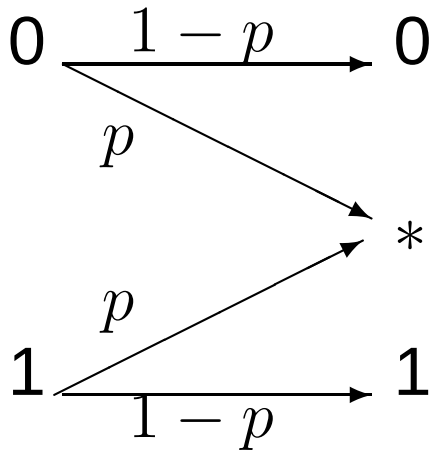


$$P = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

p is called *channel error* or *transmission probability*

Discrete memoryless channel

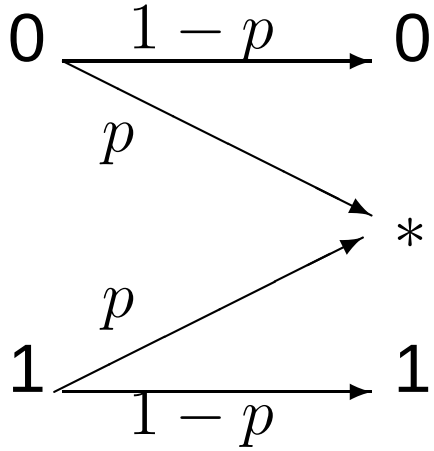
Binary Symmetric Erasure Channel (BSEC):



$$P = \begin{pmatrix} 1-p & 0 & p \\ 0 & 1-p & p \end{pmatrix}$$

Discrete memoryless channel

Binary Symmetric Erasure Channel (BSEC):



$$P = \begin{pmatrix} 1-p & 0 & p \\ 0 & 1-p & p \end{pmatrix}$$

If the alphabets A and B are supplied with algebraic structure, which is the usual case, the distortions introduced by the channel can be expressed by the *error sequence* $e_1, e_2, e_3, \ldots$, where

$$e_j = v_j - u_j, \qquad \text{i.e.} \qquad v_j = u_j + e_j.$$

Example

Given a BSC with an error probability p find the average probability for correct transmission of a block of n bits. Assume that the input 0 and 1 are equally probable.

Solution: BSC is a memoryless channel, thus, the probability of a correct transmission of a block of n bits is the product of probabilities one bit to be correctly transmitted (transmissions of bits are independent events), that is

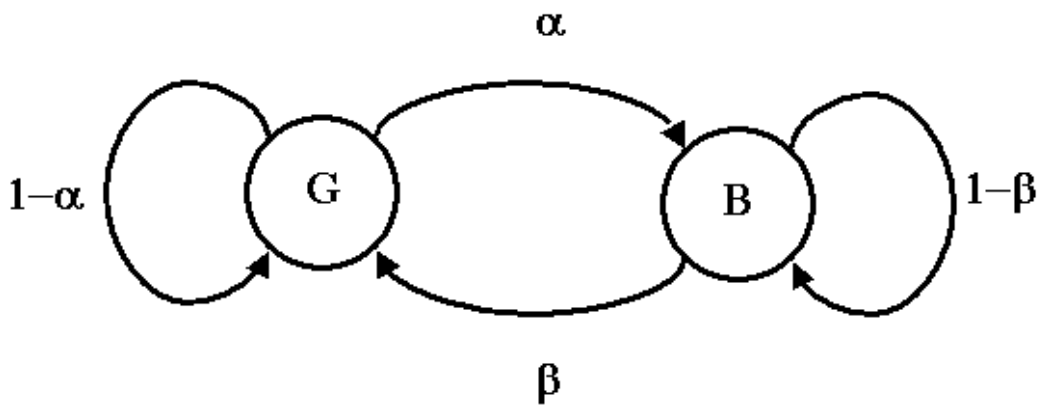
$$P_{correct} = (1-p)^n.$$

The probability only one bit to be incorrectly transmitted is $np(1-p)^{n-1}$, and so on, the probability of k incorrectly transmitted bits is

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

Discrete channel with memory

In many communication channels the error are not independent, and occur in bursts. A simple model of such channels is *Gilbert 2-state model*:



In each of states G (gap state) and B (burst state) the channel is BSC with transmission probability $p_G \approx 0$ and $p_B \approx 1/2$, respectively.

For example if $\alpha = 10^{-6}$ and $\beta = 0.056$, then the average length of a burst is $1/\beta \approx 18$ bits, while the channel is in the state G in average $1/\alpha = 10^6$ bits.

Channel capacity 1

The *channel capacity* C is defined to be the maximum of information about the input U of the channel given by the output V, that is

$$C = \max I(U \mid V).$$

Channel capacity 1

The *channel capacity* C is defined to be the maximum of information about the input U of the channel given by the output V, that is

$$C = \max I(U \mid V).$$

Shannon's Noisy Channel Coding Theorem:

There exist error control codes such that information can be transmitted across the channel at rate less than C with arbitrarily low bit error rate.

Discrete probability

A random variable X is discrete, if it can take only finite number of values $\{u_1, u_2, \ldots, u_n\}$ and

$$\sum_{j=1}^{n} \Pr(X = u_j) = 1.$$

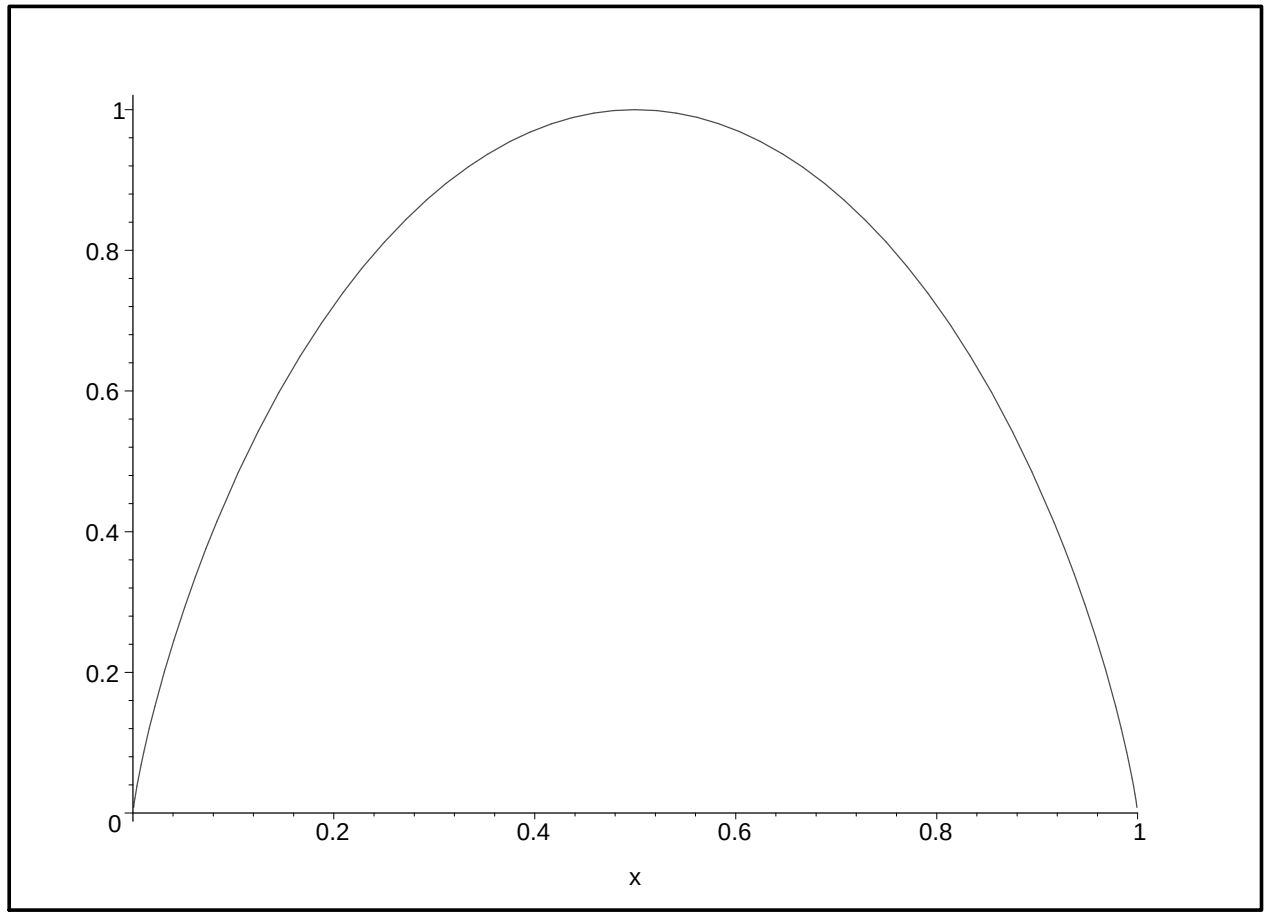
The set of probabilities $\{p_1, p_2, \ldots, p_n\}$, where $p_j = \Pr(X = u_j)$ is called the *probability distribution* of X.

Entropy of X is called the function

$$H(X) = -\sum_{k} p_k \log_2 p_k,$$

where the sum is taken over all k with $p_k > 0$.

Entropy function



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

Channel capacity 2

Consider a DMLC with an input alphabet A, an output alphabet B, and a stochastic matrix $\mathbf{P} = (p_{ij})$. Let the input U have probability distribution $\{p_1, p_2, \ldots, p_m\}$, that is

$$p_k = \Pr(U = a_k),\ k = 1, 2, \ldots, m.$$

Then the output V of the channel is determined by the probability distribution $(q_1, q_2, \ldots, q_n)$, where

$$q_j = \sum_{i=1}^{m} p_i \Pr(V = b_j \,|\, U = a_i) = \sum_{i=1}^{m} p_i p_{ij}.$$

Channel capacity 3

Theorem. The capacity of BSC with error probability p is given by

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

Channel capacity is calculated for each of proposed channel models.

Discrete vs. analog channel

Some correspondence between analog channel and discrete channel models exists. For example, such correspondence between AWGN channel and BSC can be observed if large number of bits are transmitted.
When we send 10 000, respect. 1 000, random bits across a channel with 4-PSK modulation and 3dB SNR per symbol (i.e. 2 bits) the transmission errors are as follows

p_0	p_1	p
0.1177	0.1151	0.1164
0.1190	0.1192	0.1161
0.1148	0.1187	0.1168
0.1190	0.1192	0.1161

p_0	p_1	p
0.1184	0.1333	0.1260
0.1000	0.0780	0.0890
0.1412	0.1306	0.1360
0.0806	0.1017	0.0920

Homework 1

Problem 1.1. Given a BSC with an error probability $p = 0.01$ find the average probability a block of 7 bits to be transmitted: a) correctly; b) with at most one error. (Assume that the input bits are equally probable.)

Problem 1.2. Calculate the capacity of BSC with channel error probability $p = 0.01$.

Problem 1.3. Demonstrate the relation between analog and discrete channel by determining the bit channel error of communication based on 4-PSK across AWGN channel with $E_s/N_0 = 3$ dB.

Problem 1.4. List codewords of binary $[6, 3]$ code: each block of 3 information bits $(a, b, c) \to (a, b, c, x, y, z)$, where

$$x = a + b,\ y = a + c,\ z = b + c.$$

Error control codes 1

Error control coding is accomplished by inserting control redundancy into the transmitted information stream. The added symbols (from the input alphabet A) allow the receiver to detect and possibly correct errors caused by the channel.
There are two main types of error control codes:

- Block codes: The information sequence is divided into blocks of k symbols and each block is enlarged to block of n symbols independently of the others.
- Convolutional codes: The information symbols are also manipulated in blocks of length k, but the resulted block depend on the previously transmitted ones.

Error control codes

Let A be the input alphabet. Recall that $A^n = \{(x_1, x_2, \ldots, x_n) \mid x_j \in A\}$. Any subset C of A^n is called *block code of length n over A*. If $M = \mid C \mid$ we say that C is (n, M) code.

Usually the alphabet $A = GF(q)$ is a finite field of q elements. In this case A^n is a linear space over A.

C is called *linear code of length n and dimension k, or q-ary $[n, k]_q$ code*, if it is a k-dimensional subspace of A^n. Then $\mid C \mid = q^k$. The *rate* of (n, M) code C is the ratio

$$R = \frac{\log_q M}{n}.$$

It is simplifies to $R = k/n$, when $M = q^k$, in partial, when C is linear code.

$r = n - \log_q M$ is called *redundancy*

Examples of binary codes

Example 1. *Simple Parity Check Codes* (high-rate codes with poor error performance): Given k information symbols $(a_1, \ldots, a_k)$, $a_i \in \mathbb{F}$, add a $(k+1)$-th symbol a_{k+1} so that $a_1 + a_2 + \cdots + a_{k+1} = 0$ (sum in $\mathbb{F}$). Therefore, the code is a linear $[k+1, k]$ code.

Example 2. *Simple Repetition Codes* (low-rate codes with good error performance): Every symbol is repeated n times. Usually n is odd. For instance, if $n = 5$ then $0 \to 00000$ and $1 \to 11111$. Obviously, a repetition code of length n is a linear $[n, 1]$ code with rate $R = 1/n$.

The end

Thank You for Attention!