An Introduction to Digital Communication

1. Basic Signal to Noise Calculation

In this introductory chapter, we are going to consider very simple communication systems. The transmitter consists on a plain power amplifier, the communication medium is modeled as a lossy system, the noise is considered as bandlimited white noise and the receiver is also modeled as a plain amplifier.

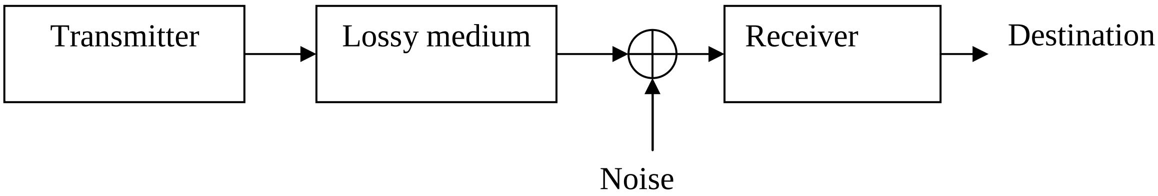


Figure 1 Simple communication system

Let us call S_T the transmitted power, S_R the signal power at the receiver input and S_D the power at the receiver output (at the destination). The lossy medium has a loss L. The bandlimited noise has a power density $\frac{N_0}{2}$ and a bandwidth B_N.

The transmission loss L is of course the inverse of a gain. This means that if the lossy medium has an input power P_{in} and an output power P_{out}, the loss is:

$$L = \frac{P_{in}}{P_{out}} \tag{1}$$

or in dB:

$$L_{dB} = 10\log_{10} L \tag{2}$$

For a transmission line, the loss depends exponentially on the length of the line. This means that the loss in dB is proportional to the length of the line. It is usually expressed in dB/km. A twisted pair, such as the ones used in telephony or in LAN wiring has an average loss per km of $3\,dB/km$ at a frequency of 100 kHz. Optical fiber is also a transmission line and as such has the same type of loss (but much smaller). Radiowave channels, on the other hand, attenuate the power in a $1/l^2$ law.

For example, 20 km of twisted pair produce a loss $L_{dB} = 3\times 20 = 60\,dB$ or $L = 10^6$. This means that the output power is one millionth times smaller than the input one. If we double the length of the cable, the loss increases to $L_{dB} = 120\ dB$ or $L = 10^{12}$. In contrast, if a radio channel has an attenuation of 10^6 for 20 km, at 40 km, the attenuation is only 4×10^6.

Signal to noise ratios

The quality of an analog communication system is often quantified by the signal to noise ratio (SNR). The signal to noise ratio is defined as the ratio of the useful signal power and the noise power at some point in a communication system.

In the system described by Figure 1, the signal to noise ratio of interest is the one at the destination. The receiver is modeled as a simple amplifier with a power gain G_R. So, if the transmitted signal power is S_T, the signal power at the input of the receiver is $S_R = S_T / L$. At the destination, the power is $S_D = G_R S_T / L$.

The noise power at the input of the receiver is $N_0 B_N$. The noise power at the destination is $N_D = G_R N_0 B_N$. The signal to noise power at the destination is:

$$SNR_D = \frac{S_T}{L N_0 B_N} \tag{3}$$

If we evaluate the different quantities in dB, the signal to noise ratio is given by:

$$\left(SNR_D\right)_{dB} = S_{TdBW} - L_{dB} - \left(N_0 B_N\right)_{dBW} \tag{4}$$

In equation(4), the powers are expressed in dBW, i.e. dB referred to 1 W. The same relation is valid if we express the powers in dBm.

Example:

A cable having a loss of 3 dB/km is used to transmit a signal for a distance of 40 km. The noise power at the input of the receiver is $\left(N_0 B_N\right)_{dBW} = -157\ dBW$. What is the minimum transmitter power required in order to have a signal to noise ratio larger than 50dB at the destination?

Using equation(4), the required signal power is:

$$S_{TdBW\min} = \left(SNR_D\right)_{dB} + L_{dB} + \left(N_0 B_N\right)_{dBW}$$

The loss introduced by the cable is $L_{dB} = 3 \times 40 = 120\,dB$. This gives a minimum transmitter power $S_{TdBW} = 13\ dBW$ or $S_T = 10^{1.3} = 20\ W$.

We see that the required power is quite high and the transmitter requires some costly electronic amplifier. A solution to this problem is the use of repeaters.

Analog repeaters

Instead of using one single length of cable, we can divide it into M sections. After each cable length, we insert an amplifier.

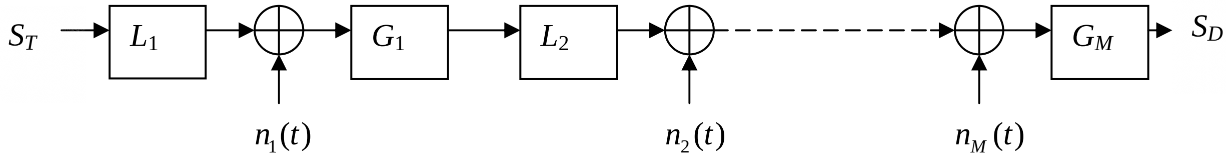


Figure 2 Analog Repeater System

In order to simplify the analysis of the above system, we assume that all sections of the cable have the same loss L_1 and every amplifier compensates exactly the loss introduced by the preceding section. So:

$$L_1 = L_2 = \cdots = L_M = G_1 = G_2 = \cdots = G_M$$

Because of the above assumption, we have $S_D = S_T$. We have also to assume that the different noises are independent and have the same power $N_0 B_N$. So, the total variance of the noise at the destination is the sum of the variances of all the different noises (multiplied by the gain chain from the summer point to the destination). The variance due to the first noise source is then $(N_0 B_M) \times G_1 \times {}^1\!/_{L_2} \times G_2 \times \cdots \times {}^1\!/_{L_M} \times G_M = G_1 N_0 B_N = L_1 N_0 B_N$. The other noise sources produce the same noise power at the output of the last amplifier. So, the noise power at the destination is $N_D = M L_1 N_0 B_N$. Finally, the signal to noise power is:

$$SNR_D = \frac{S_T}{M L_1 N_0 B_N} \tag{5}$$

or in dB:

$$(SNR_D)_{dB} = S_{TdBW} - 10\log_{10} M - L_{1dB} - (N_0 B_N)\, dBW \tag{6}$$

Example:

If we use the same data as in the previous example and we divide the 40 km cable into two sections with repeaters (M=2), we obtain $L_1 = 60\ dB$ and $S_{TdBW} = -44\ dBW$ or $S_T = 40\ \mu W$. This amount of power can be provided with a very cheap single transistor amplifier.

Digital repeaters

If we use digital communications, the quality of the transmission is measured using the probability of error. The repeaters in this case will consist on complete receivers that will demodulate the signal, recover the digital information and then retransmit it amplified. If we divide the path into M sections, the probability of error of the M sections in cascade is the probability to have an error in at least one section. So, the total probability of error is the sum of the probability that one section is in error or two, etc up to the probability of M sections being in error. If p is the probability that one section is in error, the total probability of error is given by the binomial distribution:

$$P[E] = \binom{M}{1} p(1-p)^{M-1} + \binom{M}{2} p^2 (1-p)^{M-2} + \cdots + \binom{M}{M} p^M \tag{7}$$

In general, the probability p is very small, the above expression is then reduced to the first term along with $1 - p \approx 1$. So,

$$P[E] \approx Mp \tag{8}$$

2. Pulse transmission

Many modern communication systems transmit digital information as a train of pulses. For example, the bipolar NRZ signal in binary communication corresponds to transmitting a positive pulse for one symbol ("1") and the pulse reversed (minus) for the other symbol ("0"). The pulse has a duration T. The problem that we have to answer is: What is the minimum bandwidth required to detect the presence of a positive or a negative pulse? Another problem is to determine the minimum bandwidth required to preserve the shape of the pulse.

The above signal has been studied previously and its power spectrum is a sinc square function. However, in order to answer the above questions, we have to look at one member of the process. Let us consider the signal produced by a succession of ones and zeroes. It is a square wave. Since the symbol duration is T, the fundamental period of the wave is $2T$.

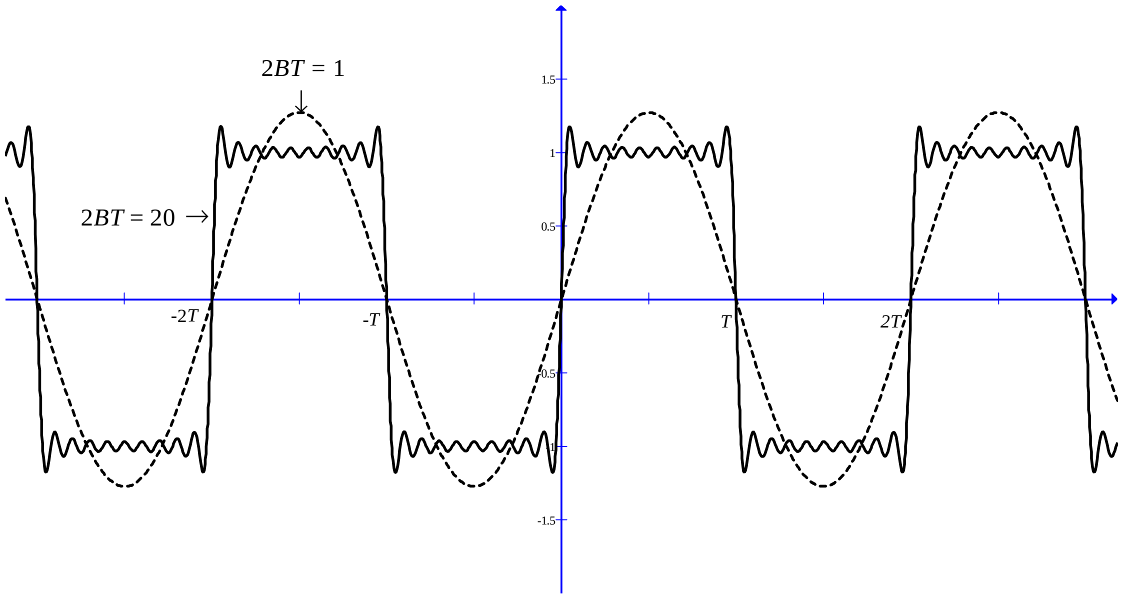


Figure 3 Filtered Square Wave

If we filter this wave and keep only the fundamental, we obtain a sinewave having a period $2T$. This signal allows us to distinguish easily the different symbols. So, the answer to the first question is:

The minimum bandwidth B required to detect the presence of a positive or a negative pulse must satisfy:

$$2BT \geq 1 \tag{9}$$

On the other hand, If we want to preserve the general shape of the pulse, we must keep much more harmonics. In Figure 3, the curve drawn with a plain line contains 19 harmonics. The general shape of the pulse is preserved. This curve is such that:

$$2BT \geq 20 \tag{10}$$

3 Binary Communication systems

In this course, we are going to study digital communication systems. We say that a communication system is digital if the source of information produces discrete "symbols" taken from some finite set called "alphabet". For example, an alphabet can be the set of different symbols that can be produced by a computer keyboard. A commonly used alphabet is one with two symbols (binary). In many cases, these symbols are called "0" and "1" or "mark" and "space" or "high" and "low".

In this section, we are going to consider binary communication systems. The transmitter is going to convert the symbol "1" to a signal and the symbol "0" to another signal. The signals are assumed to have a finite duration T. The symbol rate in our case is then

$R = 1/T$ symbols/s. Since the communication system is binary, a symbol is encoded with one bit, the symbol rate is equal to the bit rate (number of bits/s). The symbol rate is also called the Baud rate and is measured in bauds. If the source is not binary, we can encode every symbol into N binary digits. At that time, the two rates will be different.

Simple Sampling receiver

In this part, we consider a binary communication system using a bipolar NRZ signaling. The two symbols are assumed to be equiprobable. White Gaussian noise is added to the signal and the receiver filters this signal plus noise. The filtering is such that the signal is preserved and the noise power is reduced. The noise bandwidth of the receiver is B_N.

The symbol duration is T. If we want to preserve the shape of the wave, the receiver bandwidth should satisfy the inequation(10). We select

$$B_N = \frac{10}{T}$$

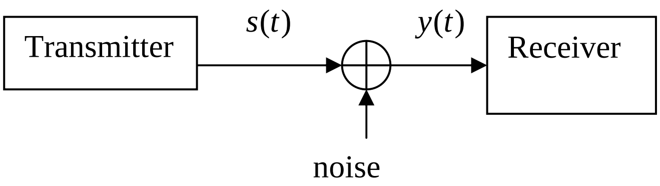


Figure 4 Communication System

The receiver is the simple following system:

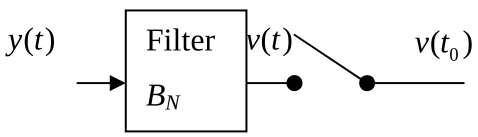


Figure 5 Sampling Receiver

We sample the signal at some instant t_0 situated in the middle of a symbol interval.

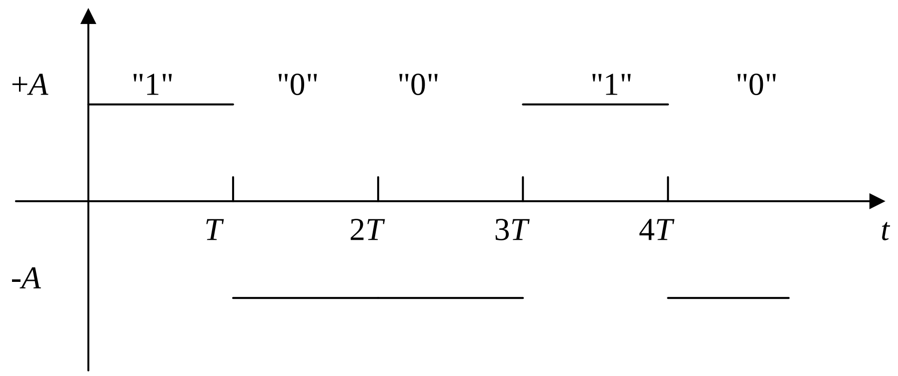


Figure 6 Waveform corresponding to the message 10010...

According to Figure 4 , Figure 5 and Figure 6, the signal $s(t)$ produced by the transmitter is constant during the whole symbol period and takes the values of $+A$ or $-A$, the noise $n(t)$ is white Gaussian with power spectrum $N_0/2$. The output of the filter $v(t)$ is then the sum of $s(t)$ and a filtered Gaussian noise $n_D(t)$ (the signal is not affected if the bandwidth of the filter is large enough). So, the sampled signal is:

$$V = v(t_0) = \pm A + n_D(t_0)$$

The variable V is then a Gaussian random variable with a mean equal to $\pm A$ (depending on which symbol is sent) and a variance $\sigma^2 = N_0 B_N$. We can write the two likelihood functions (see the communication example given in the previous set of notes):

$$f_{V|"1"}(v|"1") = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(v-A)^2}{2\sigma^2} \right] \tag{11}$$

and

$$f_{V|"0"}(v|"0") = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(v+A)^2}{2\sigma^2} \right] \tag{12}$$

Since the two symbols are assumed to be equiprobable, the MAP receiver is equivalent to a maximum likelihood one. The decision rule is then:

Decide "1" if the random variable V takes a value $v \geq 0$, decide "0" otherwise.

The probability of error is given by:

$$P[E] = P["1"]P[E|"1"] + P["0"]P[E|"0"]$$

The conditional probabilities of error are:

$$P\left[E\,|\,"1"\right]=\int_{-\infty}^{0} f_{V|"1"}\left(v\,|\,"1"\right)dv=\frac{1}{2}\operatorname{erfc}\left(\frac{A}{\sqrt{2}\sigma}\right) \tag{13}$$

and

$$P\left[E\,|\,"0"\right]=\int_{0}^{+\infty} f_{V|"0"}\left(v\,|\,"0"\right)dv=\frac{1}{2}\operatorname{erfc}\left(\frac{A}{\sqrt{2}\sigma}\right) \tag{14}$$

So, the probability of error is:

$$P\left[E\right]=\frac{1}{2}\operatorname{erfc}\left(\frac{A}{\sqrt{2}\sigma}\right)=\frac{1}{2}\operatorname{erfc}\left(\frac{A}{\sqrt{2N_0B_N}}\right) \tag{15}$$

In order to be able to compare signaling schemes using different type of waveforms, it is better to use the average energy of the pulse instead of the peak amplitude. For a simple rectangular pulse of duration T and amplitude $\pm A$, the average energy is: $E=A^2T$. So, the probability of error is:

$$P\left[E\right]=\frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E}{2N_0B_NT}}\right) \tag{16}$$

and for our choice of bandwidth:

$$P\left[E\right]=\frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E}{20N_0}}\right) \tag{17}$$

Integrate and dump receiver

The previous receiver is not very efficient. This is due to the fact that it does not take into account all the available information. During a symbol interval, the signal $s(t)$ is constant ($\pm A$) while the noise takes many different values that average to zero. If we take many samples of the signal during the time T and add them, the signal portion is going to add coherently while the noise samples will have a tendency to add to zero. In the next structure, we are going to add continuously all values of the received signal during T. We use an integrator.

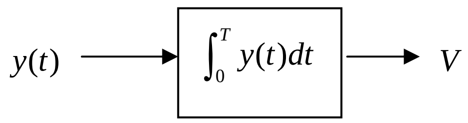


Figure 7 Integrate and Dump Receiver

With this structure, the receiver converts the stochastic process $y(t) = s(t) + n(t)$ present at its input to a random variable *V*. Since the noise is Gaussian and white, the random variable *V* is also Gaussian.

When a "1" is transmitted, the signal *s*(*t*) is constant and is equal to *A* during *T* seconds. So, the random variable *V* is:

$$V = \int_0^T \left(s(t) + n(t) \right) dt = AT + N$$

And when "0" is transmitted, we have

$$V = -AT + N$$

where *N* is

$$N = \int_0^T n(t) dt$$

The random variable *N* is Gaussian because it is formed by a linear combination of Gaussian random variables (integrated Gaussian process). Its mean is

$$E[N] = E\left[\int_0^T n(t) dt \right] = \int_0^T E[n(t)] dt = 0$$

The variance is

$$E\left[N^2 \right] = E\left[\int_0^T n(t) dt \int_0^T n(u) du \right] = \int_0^T \int_0^T E[n(t) n(u)] dt du = \int_0^T \int_0^T R_n(t-u) dt du$$

where $R_n(t-u)$ is the autocorrelation function of the noise. Being white, its autocorrelation is:

$$R_n(\tau) = \frac{N_0}{2} \delta(\tau)$$

So, the variance is:

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

The likelihood functions are:

$$f_{V|"1"}\left(v | "1" \right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\left(v - AT \right)^2}{2\sigma^2} \right] \quad (18)$$

and

$$f_{V|"0"}\left(v | "0" \right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\left(v + AT \right)^2}{2\sigma^2} \right] \quad (19)$$

The probability of error is:

$$P[E]=\frac{1}{2}\mathrm{erfc}\left(\frac{AT}{\sqrt{2}\sigma}\right)=\frac{1}{2}\mathrm{erfc}\left(\frac{AT}{\sqrt{N_0T}}\right) \quad (20)$$

Introducing the average energy $E = A^2T$, we obtain

$$P[E]=\frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{E}{N_0}}\right) \quad (21)$$

If we compare the expressions (21) and(17), we see that for the same amount of noise, the sampling receiver needs 20 times the energy of integrate and dump in order to achieve the same probability of error.

Integrate and dump receivers can be implemented using analog electronic or digital electronic circuits. However, the digital implementation requires bandlimited signals. We will study later bandlimited communication. So, we will show simply an analog implementation for this circuit. We need to build a circuit that is able to integrate a signal for a finite time. This integration can be achieved using an Op-Amp integrator.

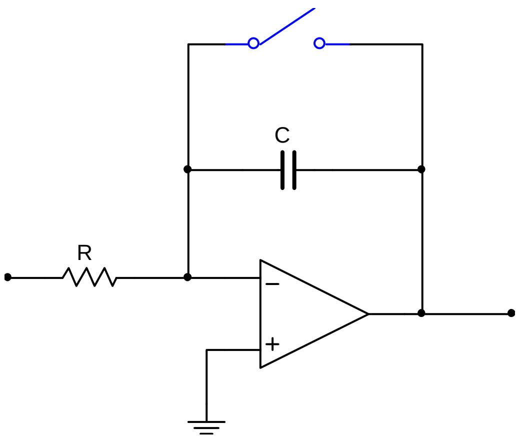


Figure 8 Resetable Integrator

The switch across the capacitor is used to discharge it before the start of the integration period. The output of the integrator is read after T seconds.

Matched Filter Receiver

Up to now, we have imposed the shape of the signal generated by the transmitter. In this part, we are going to use a transmitter that outputs a signal $s_1(t)$ when the information source produces a "1" and a signal $s_0(t)$ when the source produces a "0". The transmitted signal is then a sequence of these two signals every T seconds. This signal is added to a white Gaussian noise having a psd $S_n(f)=\frac{N_0}{2}$. The receiver consists on a filter with impulse response $h(t)$ and transfer function $H(f)$. The output of the filter is sampled at the end of the symbol duration (i.e. after T seconds).

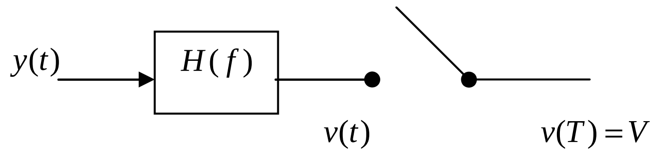


Figure 9 Matched Filter Receiver

The filter is chosen in order to minimize the probability of error of the receiver. In order to compute this probability, we have first to determine the likelihood functions. If a "1" is transmitted, the input of the receiver is:

$$y(t) = s_1(t) + n(t)$$

The output of the filter is:

$$v(t) = s_{o1}(t) + n_o(t)$$

And at the sampling instant:

$$V = v(T) = s_{o1}(T) + n_0(T) = s_{o1}(T) + N$$

The signal $s_{o1}(t)$ is the signal part of the output of the filter and N is a Gaussian random variable equal to a sampled value of the filtered noise. When "0" is transmitted, we have:

$$y(t) = s_0(t) + n(t)$$

The output of the filter is:

$$v(t) = s_{o0}(t) + n_o(t)$$

And at the sampling instant:

$$V = v(T) = s_{o0}(T) + n_0(T) = s_{o0}(T) + N$$

In this case, $s_{o0}(t)$ is the signal part of the output of the filter.

In order to determine the likelihood functions, we have to find the mean and the variance of the random variable N. Since $n(t)$ is a white noise, it is zero mean. So, the output of the filter is also zero mean. The variance of the noise at the output of the filter is given by the integral of output power spectrum. So, the variance of the random variable N is:

$$\sigma_o^2 = \int_{-\infty}^{+\infty} |H(f)|^2 S_n(f)df = \frac{N_0}{2}\int_{-\infty}^{+\infty} |H(f)|^2 df \qquad (22)$$

The likelihood functions are the conditional pdf's of the random variable V given "1" and of V given "0". They are:

$$f_{V|"1"}\left(v\,|\,"1"\right)=\frac{1}{\sqrt{2\pi}\sigma_o}\exp\left\{-\frac{\left(v-s_{o1}(T)\right)^2}{2\sigma_o^2}\right\} \tag{23}$$

and

$$f_{V|"0"}\left(v\,|\,"0"\right)=\frac{1}{\sqrt{2\pi}\sigma_o}\exp\left\{-\frac{\left(v-s_{o0}(T)\right)^2}{2\sigma_o^2}\right\} \tag{24}$$

The decision rule is:

Decide "1" if the random variable V takes a value $v \geq k$, decide "0" otherwise. The threshold k is given by:

$$k=\frac{s_{o1}(T)+s_{o0}(T)}{2} \tag{25}$$

We have assumed that $s_{o1}(T) > s_{o0}(T)$. The conditional probabilities of error are given by:

$$P\left[E\,|\,"1"\right]=\int_{-\infty}^{k}\frac{1}{\sqrt{2\pi}\sigma_o}\exp\left\{-\frac{\left(v-s_{o1}(T)\right)^2}{2\sigma_o^2}\right\}dv \tag{26}$$

and

$$P\left[E\,|\,"0"\right]=\int_{k}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma_o}\exp\left\{-\frac{\left(v-s_{o0}(T)\right)^2}{2\sigma_o^2}\right\}dv \tag{27}$$

The probability of error is

$$P\left[E\right]=\frac{1}{2}\operatorname{erfc}\left(\frac{s_{o1}(T)-s_{o0}(T)}{2\sqrt{2}\sigma_o}\right) \tag{28}$$

Since the complementary error function is monotonic decreasing, the probability of error is minimized when the argument of the erfc function is maximized. So, the problem at hand is to find the optimum filter that maximizes the following positive quantity:

$$\xi=\frac{s_{o1}(T)-s_{o0}(T)}{\sigma_0} \tag{29}$$

Let $g(t)=s_1(t)-s_0(t)$, then $g_o(t)=s_{o1}(t)-s_{o0}(t)$. The quantity to be maximized becomes:

$$\xi^2=\frac{\left|g_o(T)\right|^2}{\sigma_o^2} \tag{30}$$

We have used the square because we have the expression of the variance and not the standard deviation. The denominator of the above expression is given by (22). The numerator is the output of the filter at the instant T when the input is $g(t)$. Using inverse Fourier transforms, we have

$$g_0(t) = \int_{-\infty}^{+\infty} H(f)G(f)e^{j2\pi ft} df$$

and

$$g_0(T) = \int_{-\infty}^{+\infty} H(f)G(f)e^{j2\pi fT} df \qquad (31)$$

So,

$$\xi^2 = \frac{\left|\int_{-\infty}^{+\infty} H(f)G(f)e^{j2\pi fT} df\right|^2}{\frac{N_0}{2}\int_{-\infty}^{+\infty} |H(f)|^2 df} \qquad (32)$$

We can find an upper bound for ξ^2 by using the Schwartz inequality.

Given two complex functions X and Y of the real variable f, we can write:

$$\left|\int_{-\infty}^{+\infty} X(f)Y^*(f)df\right|^2 \leq \int_{-\infty}^{+\infty} |X(f)|^2 df \int_{-\infty}^{+\infty} |Y(f)|^2 df \qquad (33)$$

Relation (33)becomes an equality if $X(f) = \alpha Y(f)$, i.e. if they are proportional. We recognize the left hand side of (33) as being the numerator of (32)if $X(f) = H(f)$ and $Y^*(f) = G(f)e^{j2\pi fT}$. So, we obtain:

$$\xi^2 \leq \frac{2}{N_0} \frac{\int_{-\infty}^{+\infty} |H(f)|^2 df \int_{-\infty}^{+\infty} |G(f)|^2 df}{\int_{-\infty}^{+\infty} |H(f)|^2 df} = \frac{2}{N_0}\int_{-\infty}^{+\infty} |G(f)|^2 df \qquad (34)$$

The upper bound does not depend on the filter. If (33)becomes an equality, the upper bound is reached and ξ^2 is maximized. To obtain equality, we must have:

$$H(f) = G^*(f)e^{-j2\pi fT} \qquad (35)$$

We use $\alpha = 1$ as a constant of proportionality.

To obtain the impulse response of the filter, we compute the inverse Fourier transform:

$$\begin{aligned} h(t) &= \int_{-\infty}^{+\infty} H(f)e^{j2\pi ft} \\ &= \int_{-\infty}^{+\infty} G^*(f)e^{-j2\pi fT}e^{j2\pi ft} df \\ &= \int_{-\infty}^{+\infty} G(-f)e^{-j2\pi f(T-t)} df \\ &= \int_{-\infty}^{+\infty} G(f)e^{j2\pi f(T-t)} df \end{aligned}$$

So, the impulse response is given by:

$$h(t) = g(T - t) = s_1(T - t) - s_0(T - t) = h_1(t) - h_0(t) \tag{36}$$

The impulse response of the optimum filter is the difference between impulse responses of two filters. Each filter is "matched" to a signal used by the transmitter.

$$h_k(t) = s_k(T - t) \qquad k = 0,1 \tag{37}$$

The impulse response of a filter matched to a signal $s(t)$ is obtained by mirroring the signal with respect to the ordinate axis and then translating it by T seconds.

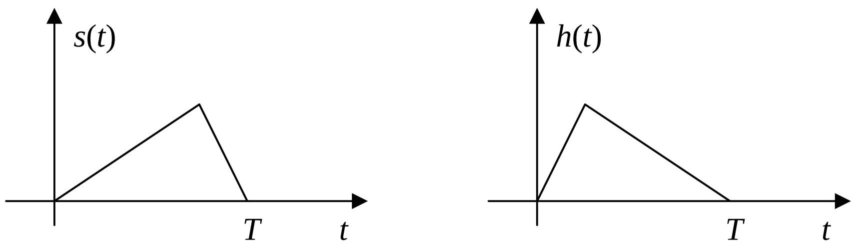


Figure 10 Matched Filter

When we use matched filters, the probability of error is given by:

$$P[E] = \frac{1}{2}\text{erfc}\left(\frac{\xi}{2\sqrt{2}}\right) \tag{38}$$

and

$$\xi = \sqrt{\frac{2}{N_0}\int_{-\infty}^{+\infty} |G(f)|^2 \, df}$$

$|G(f)|^2$ depends on the two signals $s_0(t)$ and $s_1(t)$.

Using Parseval's theorem, we can write:

$$\int_{-\infty}^{+\infty} |G(f)|^2 \, df = \int_{-\infty}^{+\infty} |g(t)|^2 \, dt = \int_{-\infty}^{+\infty} |s_1(t) - s_0(t)|^2 \, dt$$

$$\int_{-\infty}^{+\infty} |G(f)|^2 \, df = \int_{-\infty}^{+\infty} \left(s_1(t) - s_0(t)\right)^2 dt = \int_{-\infty}^{+\infty} s_1^2(t)dt + \int_{-\infty}^{+\infty} s_0^2(t)dt - 2\int_{-\infty}^{+\infty} s_0(t)s_1(t)dt$$

We recognize the energies of the signals $s_0(t)$ and $s_1(t)$. The integral of the product defines a correlation coefficient ρ_{12}:

$$\rho_{01} = \frac{1}{\sqrt{E_0 E_1}}\int_{-\infty}^{+\infty} s_0(t)s_1(t)dt \tag{39}$$

where

$$E_0 = \int_{-\infty}^{+\infty} s_0^2(t)dt \tag{40}$$

and

$$E_1 = \int_{-\infty}^{+\infty} s_1^2(t)dt \tag{41}$$

The correlation coefficient satisfies: $-1 \le \rho_{01} \le 1$ and it is equal to +1 when $s_1(t) = \alpha s_0(t)$ with α being a real positive number and it is equal to –1 when $s_1(t) = \alpha s_0(t)$ with α being a real negative number (you can use the Schwartz inequality to show the above result).

So,

$$\xi^2 = \frac{2}{N_0}\left[E_0 + E_1 - 2\sqrt{E_0 E_1}\rho_{01}\right]$$

If we introduce the average energy:

$$E = \frac{E_0 + E_1}{2} \tag{42}$$

and a generalized correlation coefficient:

$$R_{01} = \frac{2\sqrt{E_0 E_1}}{E_0 + E_1}\rho_{01} \tag{43}$$

We obtain:

$$\xi^2 = \frac{4E}{N_0}[1 - R_{12}] \tag{44}$$

Plugging (44)into(38), we can obtain the expression of the probability of error as a function of the average energy and the relationship that can exist between the signals.

$$P[E] = \frac{1}{2}\text{erfc}\sqrt{\frac{E}{2N_0}(1 - R_{01})} \tag{45}$$

A close look at the definition (43)of R_{01} shows that it is equal to ρ_{01} multiplied by the ratio of the geometric mean of E_0 and E_1 by the arithmetic mean of the two energies. The geometric mean of two positive numbers is always smaller than the arithmetic mean. The two means are equal if the two numbers are equal. So, we can conclude that $-1 \le R_{01} \le 1$. So, for a given average SNR, the probability of error is minimized if $R_{01} = -1$. This can occurs if $\rho_{01} = -1$ and $E_0 = E_1$.

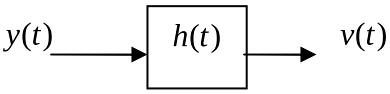
$\rho_{01} = -1$ when $s_1(t) = \alpha s_0(t)$ along with α negative. So, if we select antipodal signals ($s_1(t) = -s_0(t)$), the probability of error is minimum with

$$P[E] = \frac{1}{2}\operatorname{erfc}\sqrt{\frac{E}{N_0}} \qquad (46)$$

The probability given by (46)is the smallest achievable probability of error. It is attained when we use antipodal signals and a receiver matched to the signal. In the case of antipodal signaling, we don't need to implement two filters since $h(t) = g(T-t) = s_1(T-t) - s_0(T-t) = 2s(T-t)$ and the threshold is zero.

Correlation Receiver

The matched filter result can be obtained with a different structure: The correlation receiver. This is due to the fact that we don't need the output of the filter at times t before T. Consider the following system:



The filter is matched to $s(t)$. The signal $s(t)$ is a time limited signal. It is zero outside of the interval $[0, T]$. This means that the impulse response $h(t)$ is also time limited to T. The output $v(t)$ is given by:

$$v(t) = h(t) * y(t) = \int_0^T h(\tau)y(t-\tau)d\tau = \int_0^T s(T-\tau)y(t-\tau)d\tau$$

The output at time T is

$$v(T) = \int_0^T s(T-\tau)y(T-\tau)d\tau = \int_0^T s(\lambda)y(\lambda)d\lambda$$

The above relation can be obtained with the following system:

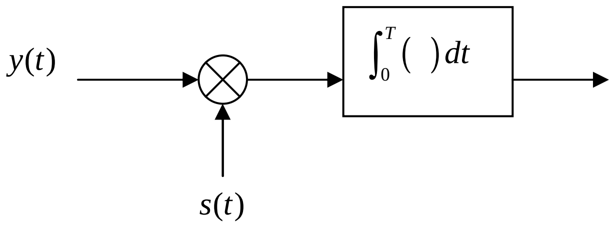


Figure 11 correlator

If the signal $s(t)$ is constant during the time T, the multiplication before integration is not needed. We see that integrate and dump receiver is in fact a matched filter receiver for antipodal pulses of constant amplitude A.

4. Simple Binary Keying Systems

In this section, we are going to analyze the performance of binary keying systems. A keying system consists on a constant amplitude sinewave transmitted for T seconds during a symbol interval. We distinguish three type of keying systems: ASK, PSK and FSK. The different systems are:

$$\text{ASK: } \begin{cases} s_0(t) = 0 & \\ s_1(t) = A\cos\omega_0 t & 0 \le t \le T \\ s_0(t) = s_1(t) = 0 & \text{elsewhere} \end{cases}$$

$$\text{PSK: } \begin{cases} s_0(t) = A\cos\omega_0 t & 0 \le t \le T \\ s_1(t) = A\cos\left(\omega_0 t + \pi\right) = -A\cos\omega_0 t & 0 \le t \le T \\ s_0(t) = s_1(t) = 0 & \text{elsewhere} \end{cases}$$

$$\text{FSK: } \begin{cases} s_0(t) = A\cos\omega_0 t & 0 \le t \le T \\ s_1(t) = A\cos\omega_1 t & 0 \le t \le T \\ s_0(t) = s_1(t) = 0 & \text{elsewhere} \end{cases}$$

It is clear that the matched filter strategy is the optimum way to go for the three keying systems.

For the ASK and the PSK modulations, we assume that the carrier has an integral number of periods during a symbol duration.

$$\omega_0 = \frac{2k\pi}{T} \qquad k \in \mathbb{N}$$

This condition implies that the energy of a burst of duration T and of amplitude A has an energy equal to $A^2T/2$. This also implies that the matched filter impulse response is:

$$h(t) = A\cos\omega_0\left(T - t\right) = A\cos\omega_0 t$$

For the FSK, we require that the same condition holds for both carriers.

For the structure of the receivers, we can use either matched filters or correlators. The ASK and PSK will have the same front end. They differ by the value of the threshold used to make a decision. The optimum threshold for PSK is zero will the optimum one for ASK is half of the maximum output of the matched filter (its value depends on the amplitude of the received signal and on the gain of the different stages before the matched filter. If the amplitude of the signal at the input of the matched filter is A and the impulse response of the filter is $A\cos\omega_0 t$, then the threshold value will be $k=\frac{A^2T}{4}$. In the three cases, the probability of error is given by equation(45). We also have:

$$E_1=\frac{A^2T}{2}$$

For PSK and FSK, we also have: $E_0 = E_1$, while $E_0 = 0$ for ASK. So, the average energy is:

$$E=\frac{E_1}{2}=\frac{A^2T}{4}\text{ for ASK}$$

$$E=E_1=E_0=\frac{A^2T}{2}\text{ for PSK and FSK}$$

Probability of error for the ASK system:

$$P\left[E\right]_{ASK}=\frac{1}{2}\operatorname{erfc}\sqrt{\frac{E}{2N_0}}=\frac{1}{2}\operatorname{erfc}\sqrt{\frac{A^2T}{8N_0}} \tag{47}$$

For the PSK, since $s_0(t) = -s_1(t)$, it is an antipodal signaling scheme. The probability of error is given by:

$$P\left[E\right]_{PSK}=\frac{1}{2}\operatorname{erfc}\sqrt{\frac{E}{N_0}}=\frac{1}{2}\operatorname{erfc}\sqrt{\frac{A^2T}{2N_0}} \tag{48}$$

In order to compute the probability of error for the FSK signaling, we need the value of R_{01}. In this case, $E_0 = E_1$, so $R_{01} = \rho_{01}$. Using equation(39), the correlation coefficient is

$$\rho_{01}=\frac{\sin\left(\omega_1-\omega_0\right)T}{\left(\omega_1-\omega_0\right)T}+\frac{\sin\left(\omega_1+\omega_0\right)T}{\left(\omega_1+\omega_0\right)T} \tag{49}$$

We define an average carrier

$$\omega_c=\frac{\omega_0+\omega_1}{2}$$

The correlation coefficient becomes:

$$\rho_{01}=\frac{\sin\left(\omega_1-\omega_0\right)T}{\left(\omega_1-\omega_0\right)T}+\frac{\sin 2\omega_c T}{2\omega_c T} \tag{50}$$

If the average carrier is very high or if $2\omega_c T=k\pi$, k being an integer, the second term of the above relation is zero. In order to have the smallest possible probability of error, the value of R_{01} should be negative and its absolute value must be as large as possible.

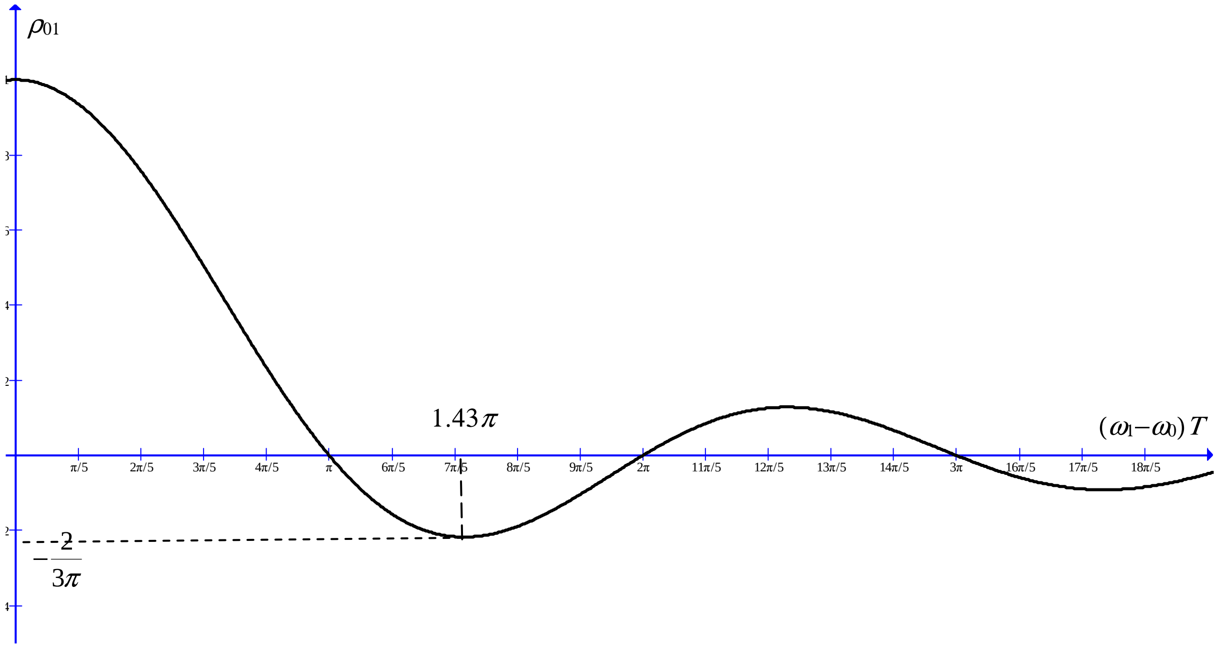


Figure 12 Correlation Coefficient

The correlation coefficient is plotted above in Figure 12. The minimum value of the correlation coefficient is

$$\rho_{01} = R_{01} = -\frac{2}{3\pi}$$

It corresponds to a frequency spacing of

$$(\omega_1 - \omega_0)T = 1.43\pi \text{ or } (f_1 - f_0)T = 0.715$$

The minimum probability of error for FSK is given by

$$P[E]_{FSK} = \frac{1}{2}\text{erfc}\sqrt{\frac{E}{2N_0}\left(1+\frac{2}{3\pi}\right)} = \frac{1}{2}\text{erfc}\sqrt{\frac{A^2T}{4N_0}\left(1+\frac{2}{3\pi}\right)} \qquad (51)$$

In many cases, the frequencies are selected so that $s_0(t)$ and $s_1(t)$ are orthogonal (correlation coefficient equal to zero). This corresponds to frequency spacing $(\omega_1 - \omega_0)T = k\pi$, k being a nonzero integer. The first zero of the correlation coefficient corresponds to the minimum frequency spacing that ensures orthogonality. In this case, the modulation is called Minimum Shift Keying (MSK).

If we compare the three different signaling schemes, we see that for the same amplitude of the carrier, ASK produces the worst result. However, if the waveforms are compared with respect to the average transmitted energy, ASK and orthogonal FSK produce the same probability of error. PSK provides the smallest probability of error with both criteria.

Comparison between digital and analog repeaters

Let us consider a cable system used in an antipodal communication system. It is divided into four (4) sections and each section has a signal to noise ratio $\frac{E}{N_0} = 9$. This means that the probability of error for a single section is:

$$p = \frac{1}{2}\text{erfc}\sqrt{9} = 1.1045\times10^{-5}$$

If we use analog repeater, the total signal to noise ratio for the whole system is divided by four. So, the probability of error of the communication system using analog repeaters is:

$$P[E]_{analog} = \frac{1}{2}\text{erfc}\sqrt{\frac{E}{4N_0}} = \frac{1}{2}\text{erfc}\sqrt{\frac{9}{4}} = \frac{1}{2}\text{erfc}\sqrt{2.25} = 1.69\times10^{-2}$$

On the other hand, if we use digital repeaters, the probability of error becomes:

$$P[E]_{digital} \approx 4p = 4.418\times10^{-5}$$

It is evident that we should not use analog repeaters when we want to transmit digital information.

5. Bandlimited Communications

When the communication system has a limited bandwidth, we cannot use time limited signals. We have to use signals that are themselves bandlimited. The main problem that one has to solve in bandlimited communication is the problem of intersymbol interference (ISI). An example is the filtering of a rectangular pulse by a first order lowpass filter:

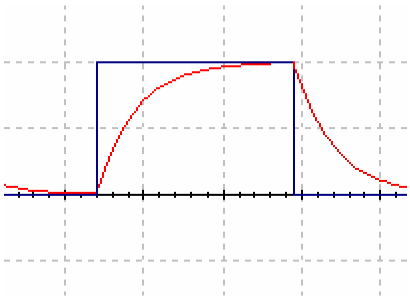


Figure 13 Filtered rectangular Pulse

In Figure 13, we clearly see that the filtered pulse leaks over the next interval. This ISI is a deterministic effect that modifies the decision threshold. To remain general, we consider that the receiver samples a train of pulses at multiples of the symbol period. Since the bandwidth is limited, we do not consider that the pulse is time limited. So, after signal processing by filters, the receiver takes values at times $t_k = kT$ and that the signal at the input of the sampler is

$$y(t) = \sum_{i=-\infty}^{+\infty} a_i p(t - iT) + n_0(t) \tag{52}$$

The signal $n_0(t)$ is the noise filtered by the receiver filter. So

$$y(t_k) = a_k p(0) + \sum_{\substack{i=-\infty \\ i \neq k}}^{+\infty} a_i p\left[\left(k - i\right)T\right] + n_0(t_k) \tag{53}$$

The first term in the above summation represents the data we want to recover, the second one represents the ISI and the third one is the added noise. In bandlimited communication, we must avoid ISI and at the same time, we must optimize the communication system from a probability of error point of view. In order to eliminate ISI, the choice of the pulse $p(t)$ is primordial. The following theorem indicates conditions for the possibility to transmit independent symbols at a given rate.

Nyquist theorem

Given an ideal lowpass channel of bandwidth B, it is possible to send independent symbol at a rate $r = {}^{1}\!/_{T}$ symbols per second if $r \leq 2B$. It is not possible to send independent symbols if $r > 2B$.

Proof:

If we try to send the signal corresponding to the sequence of alternating ones and zeroes …010101…, we are generating a periodic signal having as fundamental period $2T$. If the data rate is $r = 2\left(B + \varepsilon\right)$, $\varepsilon > 0$, the fundamental frequency of the generated wave is $B + \varepsilon$. Since the communication system is bandlimited to B, this waveform will be eliminated.

To show the possibility of such communication at a rate $r \leq 2B$, we just have to find a signal that is not affected by the filter and that does not produce ISI.

Let $p(t) = \mathrm{sinc}(rt)$, with $r = 2B$, then the signal:

$$x(t) = \sum_{k=-\infty}^{+\infty} a_k p(t - kT)$$

is such signal. If we sample it at times $t_k = kT$, we obtain:

$$x(kT) = \sum_{i=-\infty}^{+\infty} a_i \,\mathrm{sinc}\left(k - i\right) = a_k$$

by virtue of the fact that:

$$\operatorname{sinc} n = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$$

Furthermore, the Fourier transform of $p(t)$ is:

$$P(f) = F\left[\operatorname{sinc} rt\right] = \frac{1}{r}\Pi\left(\frac{f}{r}\right) = \frac{1}{2B}\Pi\left(\frac{f}{2B}\right)$$

It is not affected by the ideal lowpass filter of bandwidth B.

(Q.E.D)

The signal defined above does not produce ISI, however, it is an ideal case. It can only be approximated (for example using DSP FIR filters with a quite long delay). The sinc pulse also suffers from the fact that it has quite high side lobes. If the receiver clock is perfectly synchronized with the rate of the signal, there will not be any ISI. However, we can show that a very small amount of mistuning will produce a catastrophic ISI. So, we should look for bandlimited signals that have the same zero crossing as the sinc pulse but with much smaller side lobes.

So, the requirements for producing no ISI are to use bandlimited pulses such that:

$$p(kT) = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases} \tag{54}$$

These pulses can be produced using the following rule (VSB rule).

Let $P(f) = F\left[p(t)\right]$ be such that

$$P(f) = P_\beta(f) * \left\{\frac{1}{r}\Pi\left(\frac{f}{r}\right)\right\} \tag{55}$$

where $p_\beta(t) = F^{-1}\left[P_\beta(f)\right]$ is a pulse bandlimited to $\beta \leq r/2$ and normalized as

$$p_\beta(0) = \int_{-\infty}^{+\infty} P_\beta(f)\,df = 1 \tag{56}$$

In the time domain, the pulse is:

$$p(t) = p_\beta(t)\operatorname{sinc} rt \tag{57}$$

It satisfies the zero ISI criterion (54) and it is bandlimited to

$$B = \frac{r}{2} + \beta \leq r \tag{58}$$

Signals that satisfy the above requirement are called Nyquist shaped pulses. Among the different Nyquist shaped pulses, there is a class of easily realizable ones: "the raised cosine shaped pulses". They are characterized by:

$$P_\beta(f) = \frac{\pi}{4\beta}\cos\frac{\pi f}{2\beta}\Pi\left(\frac{f}{2\beta}\right) \qquad 0 \le \beta \le \frac{r}{2} \tag{59}$$

Giving

$$P(f) = \begin{cases} \frac{1}{r} & |f| \le \frac{r}{2} - \beta \\ \frac{1}{r}\cos^2\left(|f| - \frac{r}{2} + \beta\right) & \frac{r}{2} - \beta < |f| \le \frac{r}{2} + \beta \\ 0 & |f| > \frac{r}{2} + \beta \end{cases} \tag{60}$$

In the time domain, the pulses are expressed as:

$$p(t) = \frac{\cos 2\pi\beta t}{1 - (4\beta t)^2}\operatorname{sinc} rt \tag{61}$$

The most commonly used pulse is the one that corresponds to $\beta = r/2$. It is:

$$p(t) = \frac{\operatorname{sinc} 2rt}{1 - (2rt)^2} \tag{62}$$

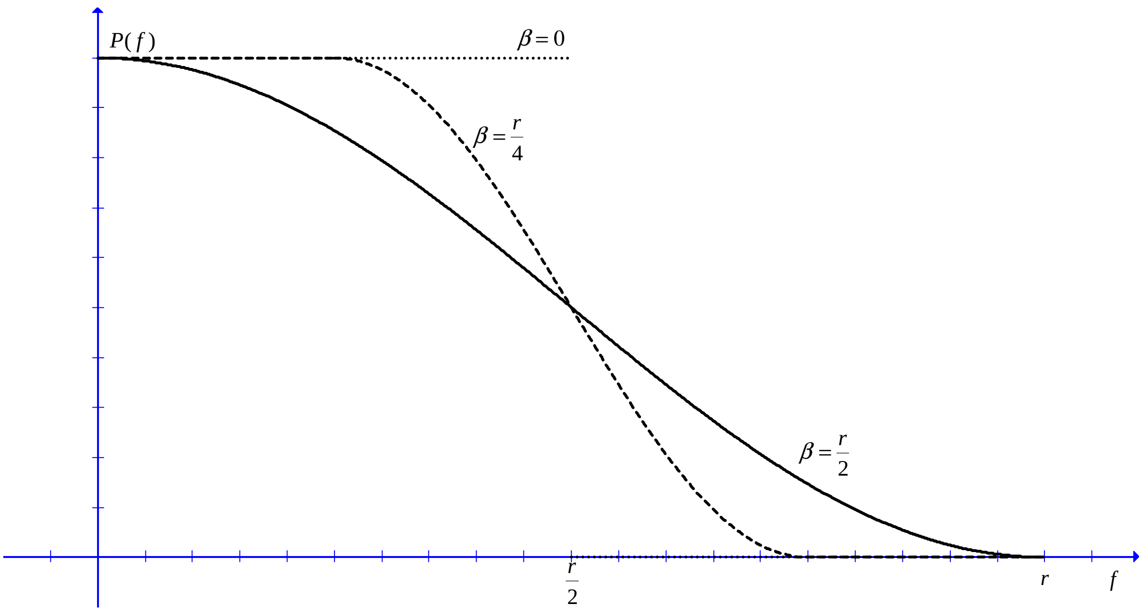


Figure 14 Spectrum of three raised cosine pulses

The next figure (Figure 15) shows both the sinc function ($\beta = 0$) and the raised cosine corresponding to $\beta = r/2$. We remark that the side lobes are very small in the raised cosine pulse. So, a small error in timing will not result in catastrophic ISI as in the sinc pulses.

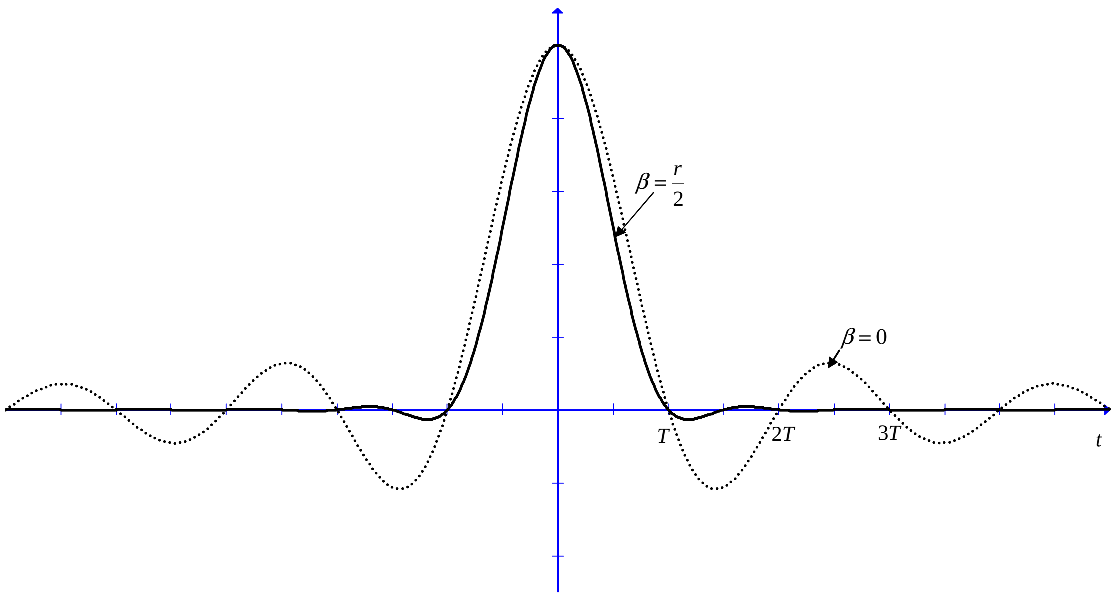


Figure 15 Pulses in the time domain

Design of a Basic bandlimited System

In this part, we are going to design a communication system with a strategy that will be different from the one of the matched filter. We have to design a system that minimizes the probability of error and at the same time avoids ISI at the sampling instant. This implies that the pulse shape at the sampler input has to be defined a priori.

The transmitter is also selected to shape the signal and we are going to assume that the transmission medium can be modelled by linear time invariant system. The noise is modelled as a zero mean Gaussian process. However, the assumption of whiteness is dropped. So, the block diagram of the communication system is:

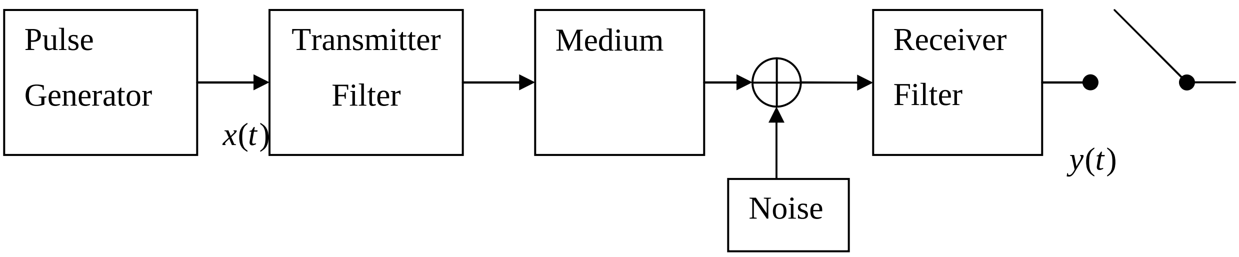


Figure 16 Bandlimited System

At the output of the receiver filter, the signal $y(t)$ is a train of Nyquist shaped pulses.

$$y(t) = \sum_{k=-\infty}^{+\infty} a_k\, p(t - kT) + n_0(t) \tag{63}$$

The signal $n_o(t)$ is the noise at the output of the receiving filter.

The pulse can be a raised cosine pulse with a given value of β compatible with the data rate *r* and the available bandwidth *B*. The pulse generator at the input of the system generates a pulse train:

$$x(t) = \sum_{k=-\infty}^{+\infty} a_k p_x(t - kT) \tag{64}$$

The sequence of numbers a_k is a sequence of independent numbers taking the values $+A$ and $-A$ with a probability of ½. The power spectrum of the above sequence is:

$$S_x(f) = A^2 r \left|P_x(f)\right|^2 \tag{65}$$

The transmitting filter has a transfer function $H_T(f)$. The medium is modeled as a filter with transfer function $H_C(f)$ and the receiving filter has a transfer function $H_R(f)$. The noise is a zero mean Gaussian process with a power spectrum $S_n(f)$. We assume that the total voltage gain of the transmission system is unity. This is why we have the same data at the input of the transmitter and at the output of the receiving filter.

The problem we have to solve is to find the optimum transmitting and receiving filters given that the pulse p(*t*), the transfer function $H_C(f)$, the power spectrum $S_n(f)$ are known. The optimization is performed from a probability of error point of view. We also fix a constraint on the transmitted power S_T at the output of the transmitting filter. If the amount of transmitted power is not constrained, then a very large transmitted power will lead to a large signal to noise ratio at the receiver. The transmitted power is given by:

$$\begin{aligned} S_T &= \int_{-\infty}^{+\infty} \left|H_T(f)\right|^2 S_x(f)\,df \\ &= A^2 r \int_{-\infty}^{+\infty} \left|H_T(f)\right|^2 \left|P_x(f)\right|^2 df \end{aligned} \tag{66}$$

At the output of the receiving filter, we require that the generated pulse $p_x(t)$ produces a Nyquist shaped pulse $p(t)$. So:

$$P_x(f) H_T(f) H_C(f) H_R(f) = P(f) e^{-j2\pi f \tau} \tag{67}$$

The power of the noise at the output of the receiving filter is:

$$\sigma^2 = \int_{-\infty}^{+\infty} \left|H_R(f)\right|^2 S_n(f)\,df \tag{68}$$

At the sampling instant, the signal $y(kT)$ is equal to a Gaussian random variable with a mean equal to $+A$ or $-A$ (depending on the transmitted symbol) and a variance σ^2. So, the probability of error is

$$P[E] = \frac{1}{2} \operatorname{erfc}\left(\frac{A}{\sqrt{2}\sigma}\right) \tag{69}$$

Using equation (66) and(68), we obtain:

$$\left(\frac{A}{\sigma}\right)^2 = \frac{S_T}{r\int_{-\infty}^{+\infty}|H_T(f)|^2|P_x(f)|^2 df\int_{-\infty}^{+\infty}|H_R(f)|^2 S_n(f)df} \tag{70}$$

Let

$$I_{TR} = \int_{-\infty}^{+\infty}|H_T(f)|^2|P_x(f)|^2 df\int_{-\infty}^{+\infty}|H_R(f)|^2 S_n(f)df \tag{71}$$

Then

$$\left(\frac{A}{\sigma}\right)^2 = \frac{S_T}{rI_{TR}} \tag{72}$$

In order to minimize the probability of error, we have to maximize the above ratio. This means that we have to minimize the denominator of the above expression. In other words, we have to find the pair of filters $H_T(f)$ and $H_R(f)$ that minimize I_{TR}.

If we use the relation(67), we can write:

$$I_{TR} = \int_{-\infty}^{+\infty}|H_R(f)|^2 S_n(f)df\int_{-\infty}^{+\infty}\frac{|P(f)|^2}{|H_C(f)|^2|H_R(f)|^2}df \tag{73}$$

We can use the Schwartz inequality (33) to solve the above problem.

$$\left|\int_{-\infty}^{+\infty}X(f)Y^*(f)df\right|^2 \le \int_{-\infty}^{+\infty}|X(f)|^2 df\int_{-\infty}^{+\infty}|Y(f)|^2 df$$

Let $X(f) = |H_R(f)|\sqrt{S_n(f)}$ and $Y(f) = \frac{|P(f)|}{|H_C(f)||H_R(f)|}$, then

$$I_{TR} \ge \int_{-\infty}^{+\infty}\frac{|P(f)|\sqrt{S_n(f)}}{|H_C(f)|}df \tag{74}$$

So ITR will be minimized when equality is reached in(74). This can occur when $X(f) = \alpha Y(f)$, giving:

$$|H_R(f)|\sqrt{S_n(f} = \alpha\frac{|P(f)|}{|H_C(f)||H_R(f)|}$$

Finally, the optimum receiving filter is:

$$|H_R(f)|^2 = \alpha\frac{|P(f)|}{|H_C(f)|\sqrt{S_n(f)}} \tag{75}$$

and the optimum transmitting filter is obtained using (67):

$$|H_T(f)|^2 = \frac{|P(f)|\sqrt{S_n(f)}}{\alpha|P_x(f)|^2|H_C(f)|} \tag{76}$$

If we use the optimum filters, the probability of error is given by (69)

$$P[E]=\frac{1}{2}\operatorname{erfc}\left(\frac{A}{\sqrt{2}\sigma}\right)=\frac{1}{2}\operatorname{erfc}\sqrt{\frac{1}{2}\left(\frac{A}{\sigma}\right)^2}$$

Using (72) and(74), the probability of error is given by:

$$P[E]=\frac{1}{2}\operatorname{erfc}\sqrt{\frac{S_T}{2r\left|\int_{-\infty}^{+\infty}\frac{|P(f)|\sqrt{S_n(f)}}{|H_C(f)|}df\right|^2}} \qquad (77)$$

If we have a flat lossy channel (loss *L*), its transfer function is:

$$|H_C(f)|^2=\frac{1}{L}$$

And if the noise is white (psd = $\frac{N_0}{2}$), the probability of error becomes:

$$P[E]=\frac{1}{2}\operatorname{erfc}\sqrt{\frac{S_T}{rN_0L}} \qquad (78)$$

The signal power at the receiver is:

$$S_R=\frac{S_T}{L}$$

We can also define the average energy of a symbol at the input of the receiving filter by

$$E=S_RT=\frac{S_R}{r}$$

So, finally, the probability of error can be expressed as:

$$P[E]=\frac{1}{2}\operatorname{erfc}\sqrt{\frac{E}{N_0}} \qquad (79)$$

The expression of the probability of error is the same as the one obtained for antipodal signaling using matched filters. However, the two systems are completely different. In the matched filter, the pulse generated by the transmitter is assumed to have a finite duration *T*. The is no filtering during the transmission that will produce a leakage of the pulse outside of the symbol interval (the matched filter is reset at the beginning of each symbol interval). In bandlimited communication, the pulses are not time limited. The pulse generated by the pulse generator is shaped by both the transmitter and the receiver filters.

When the noise is not white, the transmitter filter preemphasizes the signal where the noise is strong ($\sqrt{S_n(f)}$ at the numerator) and the receiving filter deemphasizes the signal and eliminates the effect of the preemphasis ($\sqrt{S_n(f)}$ at the denominator).

Example:

Consider a bandlimited system with a bandwidth *B*. If we want to communicate at a rate *r* = *B*, we select a pulse

$$p(t) = \frac{\operatorname{sinc} 2rt}{1-(2rt)^2}$$

corresponding to $\beta = \frac{r}{2}$. The pulse produced by the pulse generator is a simple rectangular pulse of duration *T*: $p_x(t) = \Pi(rt)$. So:

$$P_x(f) = \frac{1}{r}\operatorname{sinc}\left(\frac{f}{r}\right)$$

and

$$P(f) = \frac{1}{r^2}\cos^2\left(\frac{\pi f}{2r}\right)\Pi\left(\frac{f}{2r}\right)$$

For a flat channel and white noise, the optimum transmitting and receiving filters (within a multiplicative constant) are

$$|H_T(f)| = \frac{\cos\left(\frac{\pi f}{2r}\right)}{\sqrt{\operatorname{sinc}\left(\frac{f}{r}\right)}}\Pi\left(\frac{f}{2r}\right)$$

and

$$|H_R(f)| = \cos\left(\frac{\pi f}{2r}\right)\Pi\left(\frac{f}{2r}\right)$$

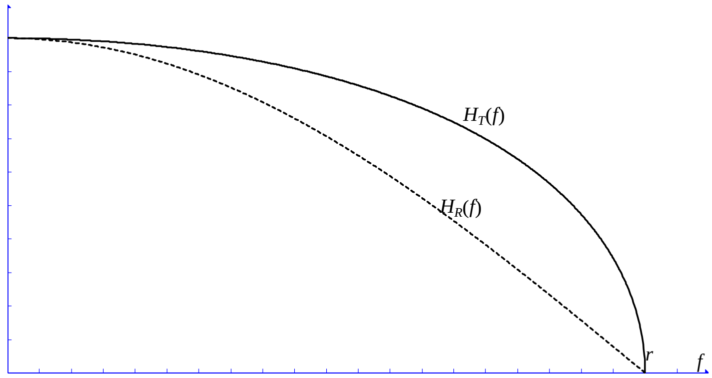


Figure 17 Optimum transmitting and receiving filters