Appendix A: Error Probability in the Transmission of Digital Signals

The two main problems in the transmission of digital data signals are the effects of channel noise and inter-symbol interference (ISI) [2, 4]. In this appendix the effect of the channel noise, assumed to be additive white Gaussian noise (AWGN), is studied, in the absence of inter-symbol interference.

A.1 Digital Signalling

A.1.1 Pulse Amplitude Modulated Digital Signals

A digital signal can be described as a sequence of pulses that are amplitude modulated. The corresponding signal is of the form

$$x(t) = \sum_{k=-\infty}^{k=\infty} a_k p(t - kT) \tag{1}$$

where coefficient a_k is the kth symbol of the sequence, such that the coefficient a_k is one of the M possible values of the information to be transmitted, taken from a discrete alphabet of symbols. The pulse $p(t)$ is the basic signal to be transmitted, which is multiplied by a_k to identify the different signals that make up the transmission.

The signal $a_k p(t - kT)$ is the kth symbol that is transmitted at the kth time interval, where T is the duration of such a time interval. Thus, the transmission consists of a sequence of amplitude-modulated signals that are orthogonal in the time domain.

As seen in Figure A.1, the data sequence $a_k = A, 0, A, A, 0, A$, corresponding to digital information in binary format (101101), is a set of coefficients that multiply a normalized basic signal or pulse $p(t - kT)$. If these coefficients are selected from an alphabet $\{0, A\}$, the digital transmission is said to have a unipolar format. If coefficients are selected from an alphabet

Essentials of Error-Control Coding Jorge Castiñeira Moreira and Patrick Guy Farrell
 ISBN: 0-470-02920-X

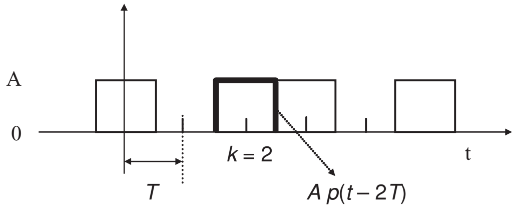

Figure A.1 A digital signal

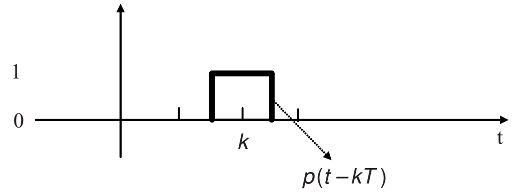

Figure A.2 A normalized pulse at time interval k multiplied by a given coefficient a_k

$\{-A/2, A/2\}$, the digital transmission is said to have a polar format. In this latter case, the sequence of this example would be given by $a_k = A/2, -A/2, A/2, A/2, -A/2, A/2$.

Index k adopts integer values from minus to plus infinity. As seen in Figure A.2, the basic signal $p(t)$ is normalized and of fixed shape, centred at the corresponding time interval k, and multiplied by a given coefficient that contains the information to be transmitted. This basic normalized pulse is such that

$$p(t) = \begin{cases} 1 & t = 0 \\ 0 & t = \pm T, \pm 2T, \ldots \end{cases} \tag{2}$$

The normalized pulse is centred at the corresponding time interval k and so its sample value at the centre of that time interval is equal to 1, whereas its samples obtained at time instants different from $t = kT$ are equal to 0. This condition does not necessarily imply that the pulse is time limited. Samples are taken synchronously at time instants $t = kT$, where $k = 0$, $\pm 1, \pm 2, \ldots$, such that for a particular time instant $t = k_1 T$,

$$x(k_1 T) = \sum_{\infty} a_{k_1} p(k_1 T - kT) = a_{k_1} \tag{3}$$

since $(k_1 T - kT) = 0$, for every k, except $k = k_1$.

Conditions (2) describe the transmission without ISI, and are satisfied by many signal pulse shapes. The classic rectangular pulse satisfies condition (2) if its duration τ is less than or equal to T. The pulse $\sin c(t)$ also satisfies the orthogonality condition described in the time domain by equation (2), but it is a pulse that is unlimited in time, however. Figure A.3 shows the transmission of the binary information sequence (11001), using $\sin c(t)$ pulses modulated in a polar format. At each sampling time instant $t = kT$, the pulse being sampled has amplitude different from 0, while the other pulses are all equal to 0.

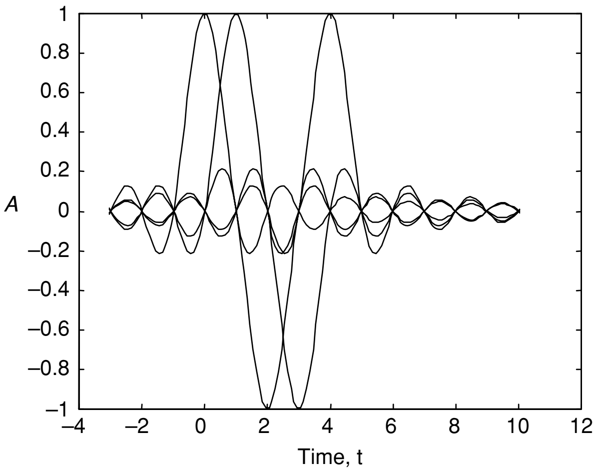

Figure A.3 A digital transmission using sinc (t) pulses modulated in polar format

Each pulse occurs in a time interval of duration T. The inverse of this duration is the symbol rate of the transmission, since it is the number of symbols that are transmitted in a unit of time (usually a second). The symbol rate r is then equal to

$$r = 1/T \text{ (symbols per second)} \tag{4}$$

which is measured in symbols per second. When the discrete alphabet used in the transmission contains only two symbols, $M = 2$, then it is binary transmission, and the corresponding symbol rate $r = r_{\mathrm{b}}$ is the binary signalling rate

$$r_{\mathrm{b}} = 1/T_{\mathrm{b}} \text{ (bit per second)} \tag{5}$$

where $T = T_{\mathrm{b}}$ is the in time duration of each bit. The binary signalling rate is measured in bits per second (bps).

A.2 Bit Error Rate

Figure A.4 shows the basic structure of a binary receiver.

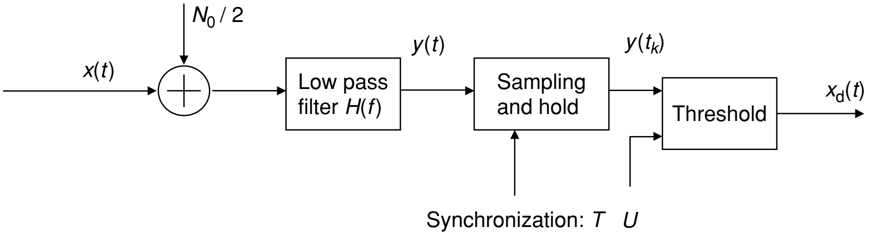

Figure A.4 A binary receiver

The signal $x(t)$ is a digital signal $\sum_k a_k p(t - kT)$, that is, an amplitude-modulated pulse signal. This signal is affected by AWGN noise in the channel and is then input to the receiver. The first block in this receiver is a low pass filter that eliminates part of the input noise without producing ISI, giving the signal $y(t)$. The receiver takes synchronized samples of this signal, and generates after the sample-and-hold operation a random variable of the form

$$y(t_k) = a_k + n(t_k) \tag{6}$$

Sampled values $y(t_k)$ constitute a continuous random variable Y, and noise samples $n(t_k)$ taken from a random signal $n(t)$ form a random variable n.

The lowest complexity decision rule for deciding the received binary value is the so-called hard decision, which consists only of comparing the sampled value $y(t_k)$ with a threshold U, such that if $y(t_k) > U$, then the receiver considers that the transmitted bit is a 1, and if $y(t_k) < U$ then the receiver considers that the transmitted bit is a 0. In this way the received sampled signal $y(t_k)$ is converted into a signal $x_{\mathrm{hd}}(t)$, basically of the same kind as that expressed in equation (1), an apparently noise-free signal but possibly containing some errors with respect to the original transmitted signal.

The probability density function of the random variable Y is related to the noise, and to conditional probability of the transmitted symbols. The following hypotheses are relevant:

H_0 is the hypothesis that a '0' was transmitted $\quad a_k = 0,\ Y = n$

H_1 is the hypothesis that a '1' was transmitted $\quad a_k = A,\ Y = A + n.$

The probability density function of the random variable Y conditioned on the event H_0 is given by

$$p_{\mathrm{Y}}(y/H_0) = p_{\mathrm{N}}(y) \tag{7}$$

where $p_{\mathrm{N}}(y)$ is the Gaussian probability density function.

For hypothesis H_1,

$$p_{\mathrm{Y}}(y/H_1) = p_{\mathrm{N}}(y - A) \tag{8}$$

The probability density function in this case is shifted to the value $n = y - A$. Thus, the probability density function for the noisy signal is the probability density function for the noise-free discrete signal 0 or A (unipolar format) added to the probability density function of the noise $p_{\mathrm{N}}(n)$. Figure A.5 shows the reception of a given digital signal performed using hard decision.

The probability density function for each signal is the Gaussian probability density function centred at the value of the amplitude that is transmitted.

Figure A.6 shows the shadowed areas under each probability density function that correspond to the probability of error associated with each hypothesis. Thus, the receiver assumes that if $Y < U$, hypothesis H_0 has occurred, and if $Y > U$, hypothesis H_1 has occurred. Error

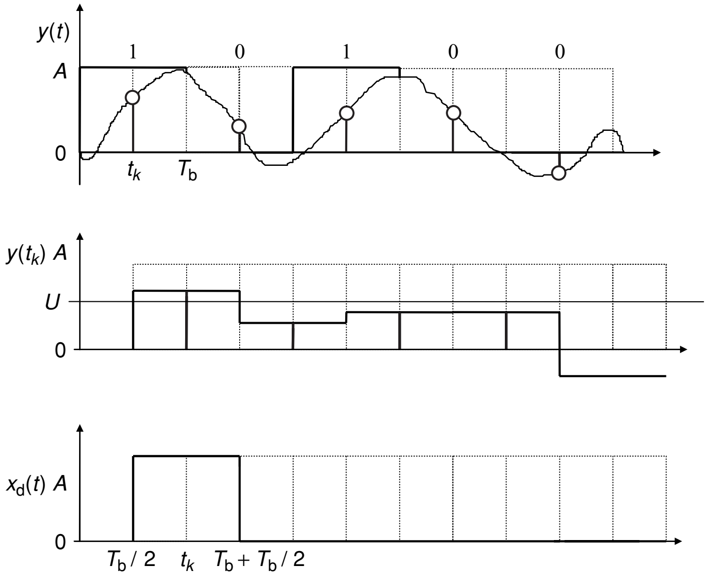

Figure A.5 Reception of a digital signal

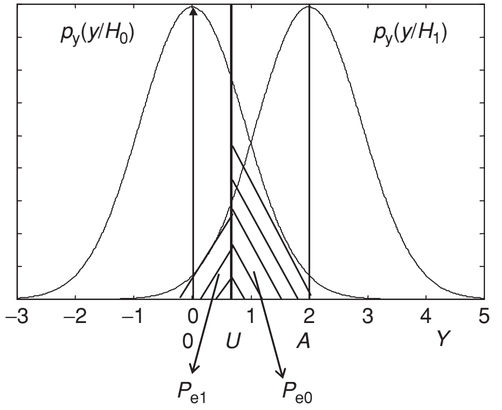

Figure A.6 Bit error rate calculation

probabilities associated with each hypothesis are described in Figure A.6, and are equal to

$$P_{e0} = P(Y > U/H_0) = \int_U^{\infty} p_Y(y/H_0)\,\mathrm{d}y \tag{9}$$

$$P_{e1} = P(Y < U/H_1) = \int_{-\infty}^{U} p_Y(y/H_1)\,\mathrm{d}y \tag{10}$$

The threshold value U should be conveniently determined. A threshold value U close to the amplitude 0 reduces the error probability associated with the symbol '1', but strongly increases the error probability associate with the symbol '0', and vice versa. The error probability of the whole transmission is an average over these two error probabilities, and its calculation can lead to a proper determination of the value of the threshold U:

$$P_e = P_0 P_{e0} + P_1 P_{e1} \tag{11}$$

where $P_0 = P(H_0)$, $P_1 = P(H_1)$.

P_0 and P_1 are the source symbol probabilities; that is, the probabilities of the transmission of a symbol '0' and '1'. The average error probability is precisely the mean value of the errors in the transmission that takes into account the probability of occurrence of each symbol.

The derivative with respect to the threshold U of the average error probability is set to be equal to zero, to determine the optimal value of the threshold:

$$\mathrm{d}P_e/\mathrm{d}U = 0 \tag{12}$$

This operation leads to the following expression:

$$P_0 p_Y(U_{\mathrm{opt}}/H_0) = P_1 p_Y(U_{\mathrm{opt}}/H_1) \tag{13}$$

If the symbols '0' and '1' of the transmission are equally likely

$$P_0 = P_1 = \frac{1}{2} \tag{14}$$

then

$$P_e = \frac{1}{2}(P_{e0} + P_{e1}) \tag{15}$$

and the optimal value of the threshold is then

$$p_{\mathrm{Y}}(U_{\mathrm{opt}}/H_0) = p_{\mathrm{Y}}(U_{\mathrm{opt}}/H_1) \tag{16}$$

As seems reasonable, the optimal value of the threshold U is set to be in the middle of the two amplitudes, $U_{\mathrm{opt}} = A/2$, if the symbol source probabilities are equal; that is, if symbols are equally likely (see Figure A.6).

The Gaussian probability density function with zero mean value and variance σ^2 characterizes the error probability of the involved symbols if they are transmitted over the AWGN channel. This function is of the form

$$p_{\mathrm{N}}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} \tag{17}$$

In general, this probability density function is shifted to a mean value m and has a variance σ^2, such that

$$p_{\mathrm{N}}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-m)^2}{2\sigma^2}} \tag{18}$$

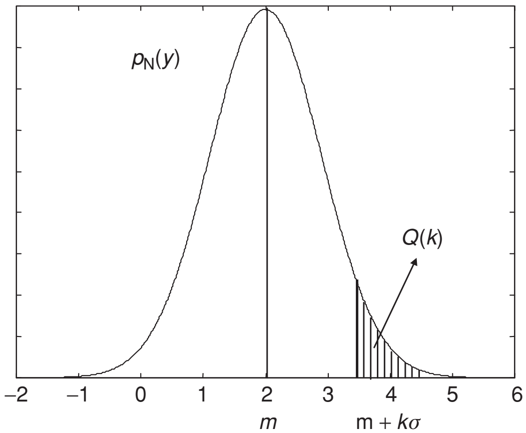

Figure A.7 Normalized Gaussian probability density function $Q(k)$

The probability that a given value of the random variable Y is larger than a value $m + k\sigma$ is a function of the number k, and it is given by

$$P(Y > m + k\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{m+k\sigma}^{\infty} e^{-\frac{(y-m)^2}{2\sigma^2}}\, \mathrm{d}y \tag{19}$$

These calculations are simplified by using the normalized Gaussian probability density function, also known as the function $Q(k)$ (Figure A.7):

$$Q(k) = \frac{1}{\sqrt{2\pi}} \int_{k}^{\infty} e^{-\frac{(\lambda)^2}{2}}\, \mathrm{d}\lambda \tag{20}$$

obtained by putting

$$\lambda = \frac{y - m}{\sigma} \tag{21}$$

This normalized function can be used to calculate the error probabilities of the digital transmission described in equations (9) and (10).

$$P_{e0} = \int_{U}^{\infty} p_{\mathrm{N}}(y)\, \mathrm{d}y = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{U}^{\infty} e^{-\frac{y^2}{2\sigma^2}}\, \mathrm{d}y = Q(U/\sigma) \tag{22}$$

and

$$P_{e1} = \int_{-\infty}^{U} p_{\mathrm{N}}(y - A)\, \mathrm{d}y = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{U} e^{-\frac{(y-A)^2}{2\sigma^2}}\, \mathrm{d}y = Q((A - U)\,/\,\sigma) \tag{23}$$

If $U = U_{\mathrm{opt}}$, the curves intersect in the middle point $U_{\mathrm{opt}} = A/2$.

In this case these error probabilities are equal to

$$P_{e0} = P_{e1} = Q\left(\frac{A}{2\sigma}\right) \tag{24}$$

$$P_e = \frac{1}{2}(P_{e0} + P_{e1}) = Q\left(\frac{A}{2\sigma}\right)$$

This is the minimum value of the average error probability for the transmission of two equally likely symbols over the AWGN channel. As seen in the above expressions, the term $A/2\sigma$ (or equivalently its squared value) defines the magnitude of the number of errors in the transmission, that is, the error probability or bit error rate of the transmission.

The result is the same for transmission using the polar format ($a_k = \pm A/2$), if the symbol amplitudes remain the same distance A apart.

The above expressions for the error probability can be generalized for the transmission of M symbols taken from a discrete source, and they can also be described in terms of the signal-to-noise ratio. The power associated with the transmission of the signal described in equation (1) is useful for this latter purpose. Let us take a sufficiently long time interval T_0, such that $T_0 = NT$, and $N \gg 1$. The amplitude-modulated pulse signal uses the normalized pulse

$$p(t) = \begin{cases} 1 & |t| < \tau/2 \\ 0 & ||t| > \tau/2 \end{cases} \tag{25}$$

where $\tau \leq T$. Then the power associated with this signal is equal to

$$\begin{aligned}
S_\mathrm{R} &= \frac{1}{T_0}\int_{-T_0/2}^{T_0/2}\left(\sum_k a_k p(t-kT)^2\right)\mathrm{d}t = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2}\sum_k a_k^2 p^2(t-kT)\,\mathrm{d}t \\
&= \frac{1}{T_0}\int_{-T_0/2}^{T_0/2}\sum_{k=-N/2}^{k=N/2} a_k^2 p^2(t-kT)\,\mathrm{d}t \\
S_\mathrm{R} &= \sum_k \frac{1}{NT}\int_{-T/2}^{T/2} a_k^2 p^2(t)\,\mathrm{d}t = \frac{N_0}{NT}\int_{-T/2}^{T/2} a_0^2 p^2(t)\,\mathrm{d}t + \frac{N_1}{NT}\int_{-T/2}^{T/2} a_1^2 p^2(t)\,\mathrm{d}t \\
S_\mathrm{R} &= P_0\frac{1}{T}\int_{-\tau/2}^{\tau/2} a_0^2 p^2(t)\,\mathrm{d}t + P_1\frac{1}{T}\int_{-\tau/2}^{\tau/2} a_1^2 p^2(t)\,\mathrm{d}t
\end{aligned} \tag{26}$$

The duration of the pulse can be equal to the whole time interval $T = \tau = T_\mathrm{b}$, in this case it is said that the format is non-return-to-zero (NRZ), or it can be shorter than the whole time interval $\tau < T_\mathrm{b}$, then the format is said to be return-to-zero (RZ). For the NRZ format,

$$S_\mathrm{R} = A^2/2 \quad \text{unipolar NRZ}$$

$$S_\mathrm{R} = A^2/4 \quad \text{polar NRZ}$$

$$A = \begin{cases} \sqrt{2S_\mathrm{R}} & \text{unipolar} \\ \sqrt{4S_\mathrm{R}} & \text{polar} \end{cases} \tag{27}$$

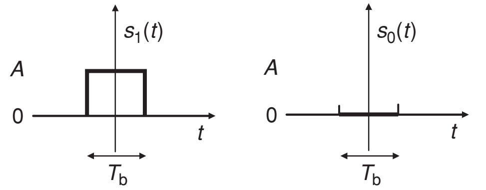

Figure A.8 Signalling in the NRZ unipolar format

If σ^2 is the noise power N_R at the output of the receiver filter, then

$$\left(\frac{A}{2\sigma}\right)^2 = \frac{A^2}{4N_R} = \begin{cases} (1/2)(S/N)_R & \text{unipolar} \\ (S/N)_R & \text{polar} \end{cases} \tag{28}$$

Thus, unipolar format needs twice the signal-to-noise ratio to have the same BER performance as that of the polar format.

The error probability was determined as a function of the parameter $A/2\sigma$. However, a more convenient way of describing this performance is by means of the so-called average bit energy-to-noise power spectral density ratio E_b/N_0. This new parameter requires the following definitions:

$$E_b = \frac{S_R}{r_b} \quad \text{average bit energy} \tag{29}$$

$$\frac{E_b}{N_0} = \frac{S_R}{N_0 r_b} \quad \text{average bit energy-to-noise power spectral density ratio} \tag{30}$$

The average bit energy of a sequence of symbols such as those described by the digital signal (1) is calculated as

$$E_b = E\left[a_k^2 \int_{-\infty}^{\infty} p^2(t-kD)\,\mathrm{d}t\right] = E\left[a_k^2 \int_{-\infty}^{\infty} p^2(t)\,\mathrm{d}t\right] = \overline{a_k^2} \int_{-\infty}^{\infty} p^2(t)\,\mathrm{d}t \tag{31}$$

The above parameters are calculated for the unipolar NRZ format. In this format a '1' is usually transmitted as a rectangular pulse of amplitude A, and a '0' is transmitted with zero amplitude as in Figure A.8.

The average bit energy E_b is equal to

$$E_1 = \int_0^{T_b} s_1^2(t)\,\mathrm{d}t = A^2 T_b$$

$$E_0 = \int_0^{T_b} s_0^2(t)\,\mathrm{d}t = 0$$

$$E_b = P_0 E_0 + P_1 E_1 = \frac{1}{2}(E_0 + E_1) = \frac{A^2 T_b}{2} \tag{32}$$

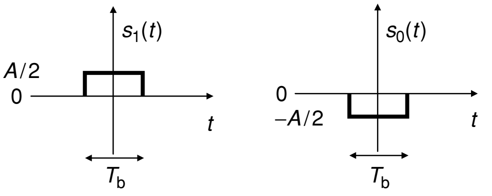

Figure A.9 Signalling in the NRZ polar format

Since the transmission is over the AWGN channel of bandwidth B and for the maximum possible value of the symbol or bit rate $r_b = 2B$, [1–4], the input noise is equal to

$$N_R = \sigma^2 = N_0 B = \frac{N_0 r_b}{2} \tag{33}$$

This is the minimum amount of noise that inputs the receiver if a matched filter is used [1–4]. The quotient $(A/2\sigma)^2$ can be now expressed as

$$\left(\frac{A}{2\sigma}\right)^2 = \frac{A^2}{4\sigma^2} = \frac{2E_b r_b}{4N_0 r_b/2} = \frac{E_b}{N_0} \tag{34}$$

In the case of the NRZ polar format, where a '1' is usually transmitted as a rectangular pulse of amplitude $A/2$ and a '0' is transmitted as a rectangular pulse of amplitude $-A/2$ (Figure A.9).

Then the average bit energy E_b is

$$E_1 = \int_0^{T_b} s_1^2(t)\,dt = \frac{A^2 T_b}{4}$$

$$E_0 = \int_0^{T_b} s_0^2(t)\,dt = \frac{A^2 T_b}{4}$$

$$E_b = P_0 E_0 + P_1 E_1 = \frac{1}{2}(E_0 + E_1) = \frac{A^2 T_b}{4} \tag{35}$$

and so

$$\left(\frac{A}{2\sigma}\right)^2 = \frac{A^2}{4\sigma^2} = \frac{4E_b r_b}{4N_0 r_b/2} = \frac{2E_b}{N_0} \tag{36}$$

It is again seen that the polar format has twice the value of $(A/2\sigma)^2$ for a given value of E_b/N_0 with respect to the unipolar format:

$$\left(\frac{A}{2\sigma}\right)^2 = \begin{cases} \frac{E_b}{N_0} & \text{unipolar} \\ \frac{2E_b}{N_0} & \text{polar} \end{cases} \tag{37}$$

Now expressing the error probabilities of the two formats in terms of the parameter E_b/N_0, we obtain

$$P_e = \begin{cases} Q\left(\sqrt{\frac{E_b}{N_0}}\right) & \text{unipolar} \\ Q\left(\sqrt{\frac{2E_b}{N_0}}\right) & \text{polar} \end{cases} \tag{38}$$

This is the minimum value of the error probability and is given when the receiver uses the matched filter. Any other filter will result in a higher bit error rate than that expressed in (38). The matched filter is optimum in terms of maximizing the signal-to-noise ratio for the reception of a given pulse shape, over a given channel transfer function, and affected by a given noise probability density function.

Bibliography

[1] Carlson, A. B., *Communication Systems: An Introduction to Signals and Noise in Electrical Communication*, 3rd Edition, McGraw-Hill, New York, 1986.
[2] Sklar, B., *Digital Communications: Fundamentals and Applications*, Prentice Hall, Englewood Cliffs, New Jersey, 1988.
[3] Couch, L. W., *Digital and Analog Communications Systems*, MacMillan, New York, 1996.
[4] Proakis, J. G. and Salehi, M., *Communication Systems Engineering*, Prentice Hall, Englewood Cliffs, New Jersey, 1994.