Lebanese University
Faculty of Engineering- III
Department: Electricity and Electronics
Semester: VIII – option T

Lecture Notes
Telecom II

A. Samhat

Introduction to digital communications

1 TELECOMMUNICATIONS SYSTEMS

A communication system is, simply, any system in which information is transmitted from one physical location to a second physical location, e.g. one person talking to a second person.
Any communication system is made up of three parts, shown in Figure 1.

- The transmitter that includes two items: the source of the information, and the technology that sends the information out over the channel.
- The channel which is the medium that the information travels through in going from one physical location to a second physical location, e.g. copper wire, atmosphere etc.
- The receiver, the part of the communication system that gets all the information that the transmitter sends over the channel.

A telecommunication system is a communication system and a system which allows the information to be sent beyond the range of usual vocal or visual communications, e.g. one person talking to a second person over the telephone.

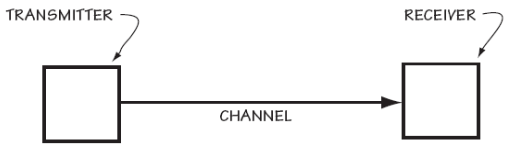


Figure 1: Parts of a communication system

An analog signal is a signal that can take on any amplitude and is well-defined at every time. See case (a) in Figure 2.
A discrete-time signal is a signal that can take on any amplitude but is defined only at a set of discrete times. See case (b).
A digital signal is a signal whose amplitude can take on only a finite set of values, normally two, and is defined only at a discrete set of times. See case (c).

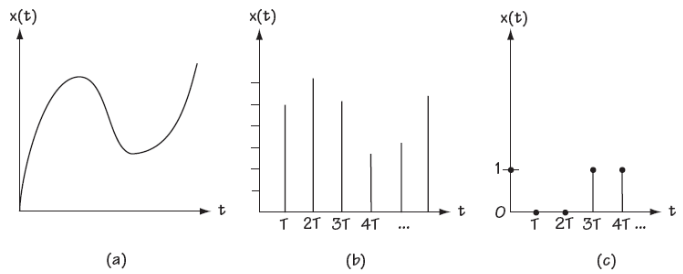


Figure 2: Different signals

An analog communication system is a communication system where the information signal sent from one point to another can only be described as an analog signal, e.g. speaking over the telephone.
A digital communication system is a communication system where the information signal sent from one point to another can be fully described as a digital signal. For example, data is sent from one computer to another over a wire.

1.1 DIGITAL COMMUNICATIONS

A typical digital communication system is shown in Figure 3 through the block diagram. A real system configuration could be more complicated where other features are added to the system. It could be simpler too where same blocks are not needed.

The message signal to be sent may be from an analog source (voice, etc.) or from digital source (computer data, etc.). The analog-to-digital (A/D) converter samples and quantizes the analog source and represents the samples in digital form (bit 1or 0).
The source encoder accepts the digital signal and encodes it into a shorter digital signal. This is called source encoding which reduces the redundancy and consequently reduces the bandwidth requirement of the system.
The channel encoder receives the output digital signal from the source encoder and encodes it into a longer digital signal. Redundancy is deliberately added into the coded digital signal so that some errors caused by the noise or interference during transmission through the channel can be corrected at the receiver most often the transmission is in high frequency.
The modulator then impresses the encoded digital symbols onto a carrier. (Sometimes the transmission is baseband see section 4).
Usually there is a power amplifier following the modulator and an antenna is the final stage of the transmission for wireless system. The transmission medium is usually called the channel which has a limited frequency bandwidth so that it can be seen as a filter. Noise adds to the signal and fading as well as attenuation affect the signal.
In the receiver, virtually the reverse signal processing happens. First the received weak signal is amplified and demodulated. Then the added redundancy is taken away by the channel decoder and the source decoder recovers the signal to its original form before being sent to the use. A Digital-to-Analog (D/A) converter is used for analog users.

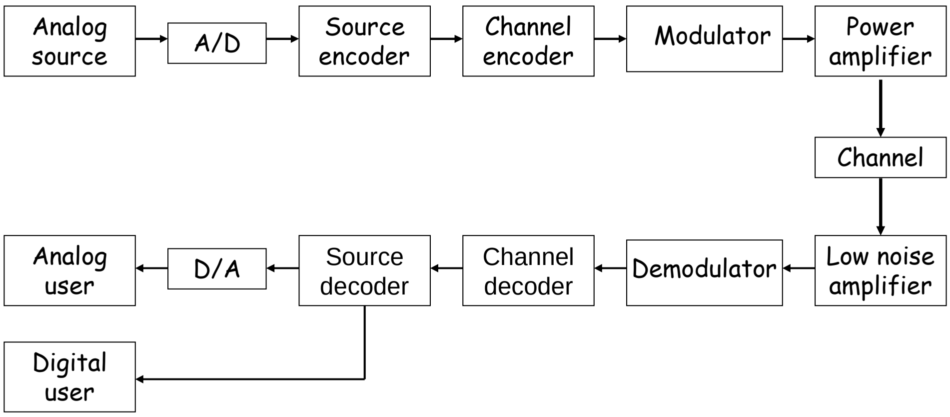


Figure 3: Block diagram of typical digital communication system

In this course, we will present the above blocks and we will mainly focus on the modulator and demodulator blocks to investigate the different modulations schemes used at this stage.

1.2 CRITERIA OF SELECTION

A wide variety of modulation techniques are used in wireless networking. The selection of the best digital modulation technique for a specific application is driven by a number of criteria, the most important being:

- Spectral efficiency: achieving the desired data rate within the available spectral bandwidth. The spectral efficiency of a digital signal is given by the number of bits per second of data that can be supported by each hertz of bandwidth $\eta=R/B$ (bits/s)/Hz where R is the data rate and B is the bandwidth.
- Bit error rate (*BER*) performance: achieving the required error rate given the specific factors causing performance degradation in the particular application (interference, fading, etc.)
- Power efficiency: particularly important in mobile applications where battery life is an important user acceptance factor.

- Modulation schemes with higher spectral efficiency (in terms of data bits per Hz of bandwidth) require higher signal strength for error-free detection.
- Implementation complexity: which translates directly into the cost of hardware to apply a particular technique. Some aspects of modulation complexity can be implemented in software, which has less impact on end-user costs.

2 A/D CONVERSION AND SOURCE CODING

To transmit an analog signal in digital form, an A/D converter is required at the sending side and D/A converter is required at the receiving side to reconstruct the analog signal. Note that we can find representations where the A/D conversion could be considered as a part of the source coding.

2.1 SAMPLING

Sampling is the changing of an analog signal to samples (or pieces) of itself. There are three methods of sampling: ideal sampling (or impulse train sampling), Zero-order Hold Sampling and natural sampling.

2.1.1 IDEAL SAMPLING

In ideal sampling, the sampler multiplies an analog input *x(t)* by the impulse train signal $p(t)=\sum_{k=-\infty}^{\infty}\delta(t-kT_s)$ (δ is the delta[1] function positioned at time $t=kT_s$, see Figure 4). The output of the sampler $x_s(t)$ is made up of impulses at times kT_s of height $x(kT_s)$; that is $x_s(t)=\sum_{k=-\infty}^{\infty}x(kT_s)\delta(t-kT_s)$

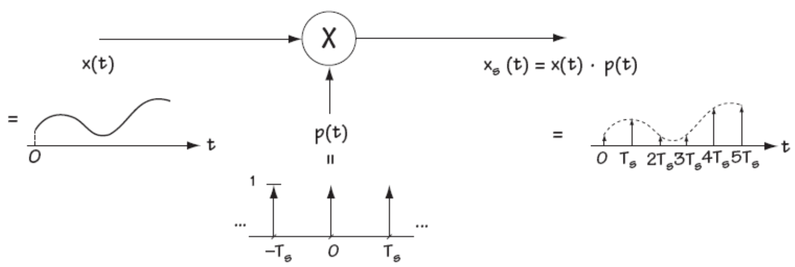


Figure 4: Ideal sampling

To figure out the information in the sampled signal, we evaluate $X_s(f)$ the Fourier transform of the output signal $x_s(t)$ is given by

$$X_s(f)=\frac{1}{T_s}\sum_{k=-\infty}^{\infty}X(f)*\delta(f-kf_s)=\frac{1}{T_s}\sum_{k=-\infty}^{\infty}X(f-kf_s)$$

Figure 5 shows the Fourier transform $X(f)$ of the input signal $x_s(t)$ and the Fourier transform *P(f)* of the impulse signal.

[1] A delta function is closely approximated by a rectangular pulse of duration Δt . The smaller Δt the better the approximation.

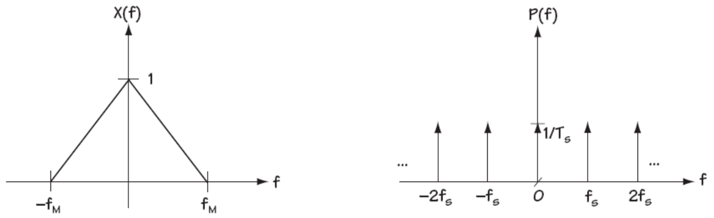


Figure 5: *X(f)* and *P(f)*

$X_s(f)$ is shown in Figure 6 and consists of a number of $X(f)$'s, shifted by multiples of f , and all added together.

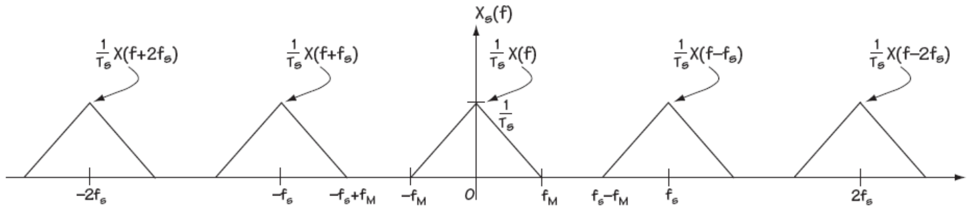


Figure 6: Fourier transform of the output signal $x_s(t)$

All of the information of $x(t)$, is in the sampled signal $x(t)$ *if* we just insure that $f_s > 2f_M$. To get back all the information from the samples (to recover $X(f)$ from $X_s(f)$), we use a low-pass filter (LPF) to cut off every thing outside of $-f_M$ and f_M , and we add a gain of T_s.

2.1.1.1 Sampling theorem

The sampling theorem states that a signal can be recovered from its samples as long as it is sampled at $f_s > 2f_M$.

2.1.1.2 Nyquist rate

The Nyquist rate f_N is the smallest sampling rate f_s that can be used if you want to recover the original signal from its samples (f_N=2 f_M). $\frac{1}{f_N}$ is called Nyquist interval.

2.1.1.3 Aliasing

Aliasing refers to the phenomenon when there is an overlapping of the $X(f)$ component with the $X(f - f_s)$ component. As a result, the original $X(f)$ is no longer preserved (see Figure 7).

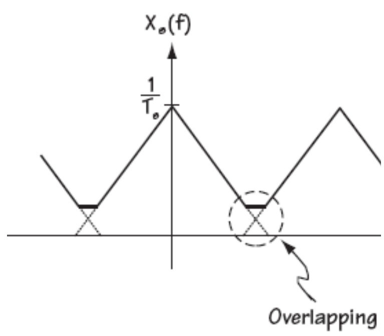


Figure 7: Aliasing

2.1.2 Zero-order Hold Sampling (Flat-top)

In this sampling also called falt-top sampling, the incoming signal $x(t)$ first goes into an ideal sample where it gets multiplied by $p(t)$. This leads to impulses of height $x_s(kT_s)$. Next, the output of the ideal sampler, $x_i(t)$, enters into a linear time invariant (LTI) system. The LTI system is described by the impulse response *h(t)* (a rectangular impulse response of amplitude 1 and duration T_s). This leads to the output: for each incoming sample of height $x(kT_s)$ (in $x_i(t)$), the output corresponds to "holding on" to the value $x(kT_s)$ for a duration of T_s. Briefly speaking, $x(t)$ is a signal with height $x_s(kT_s)$ in each time interval $[kT_s, (k+1)T_s)$ (see Figure 8).

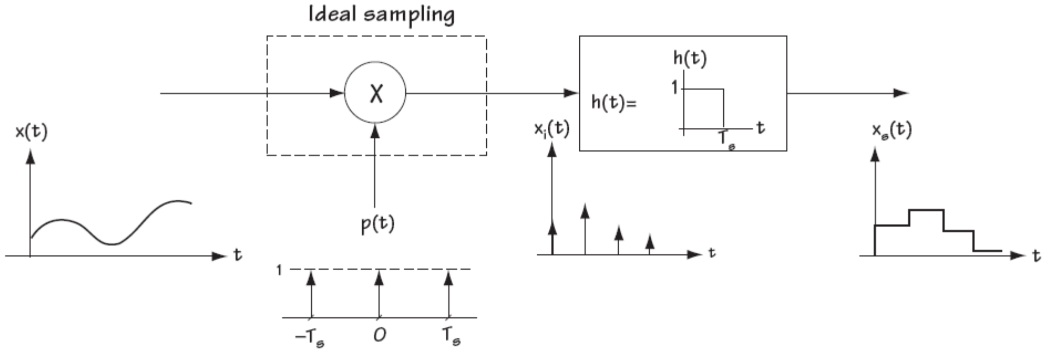


Figure 8: Zero–order hold sampling

2.1.3 NATURAL SAMPLING

Another practical method to sample the signal *x(t)* is natural sampling shown in . Here *p(t)* is made up of a bunch of tall, skinny, rectangular shapes of height $\frac{1}{T}$ and width *T*; these tall skinny rectangles are spaced T_s seconds apart. Figure 9 shows the natural sampling. The output of the sampler is simply $x_s(t) = x(t).p(t)$.

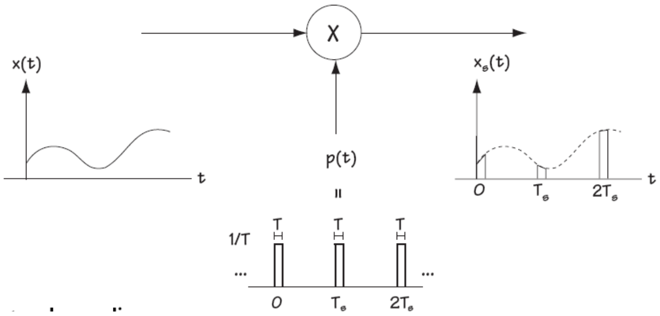


Figure 9: Natural sampling

If the sampling is done with $f_s > 2f_M$, we can get back the information from the sampled signal. Let *X(f)* the Fourier transform of *x(t)* and we will figure out $X_s(f)$ the Fourier transform of the output signal $x_s(t)$.

$$X_s(f) = F(x(t) * p(t))$$

Because *p(t)* is a periodic signal, it can be written using a Fourier series according to

$$p(t) = \sum_{k=-\infty}^{\infty} c_k e^{j2\pi k f_s t}$$

where c_k are the Fourier series coefficients. c_k can be computed by

$$c_k = \frac{1}{T_s} \frac{\sin(\frac{\pi nT}{T_s})}{\frac{\pi nT}{T_s}} = \frac{1}{T_s} \sin c(\frac{\pi nT}{T_s})$$

we get $X_s(f)$ by $X_s(f) = \sum_{k=-\infty}^{\infty} c_k.F(x(t).e^{j2\pi k f_s t}) = \sum_{k=-\infty}^{\infty} c_k.X(f - kf_s)$.

$X_s(f)$ consists of many copies of $X(f)$ added together, where the k^{th} copy is shifted by kf_s and multiplied by c_k . Figure 10 shows this in one simple picture. As in ideal sampling, if $f_s > 2f_M$ we get back all the information from the samples using a low-pass filter (LPF) to get $c_0.X(f)$ and we add a gain of $1/c_0$.

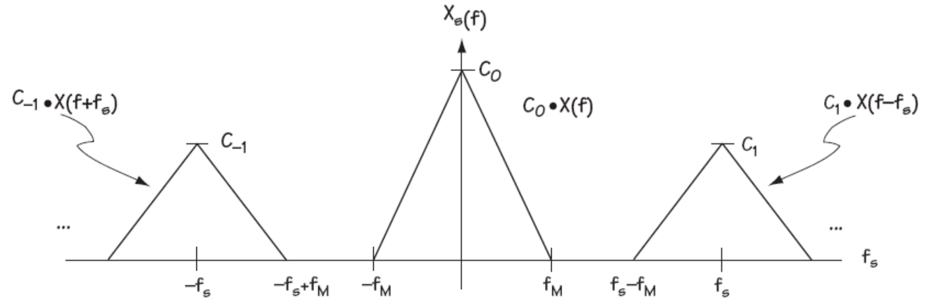


Figure 10: $X_s(f)$ the Fourier transform of the output signal

2.2 QUANTIZATION

After the sampling operation, the next operation carried out in source coding is called quantization, and the device which does it is called a quantizer. A quantizer is an amplitude changer that takes the incoming signal $x_s(t)$ and changes its amplitude (at every time) to the closest of one of L allowed values (see Figure 11). Each allowed output amplitude is called codeword and the set of L codewords is collectively called the codebook (L is always a power of 2.). Each quantized sample is digitally encoded to a sequence of l bits ($L=2^l$).

- A quantizer is called uniform if all its codewords are equally spaced. Otherwise it is called non-uniform.
- A quantizer is called a mid-tread if it has a 0 as one of its codewords. If it does not have 0 as one of its codewords, it is called mid-riser.

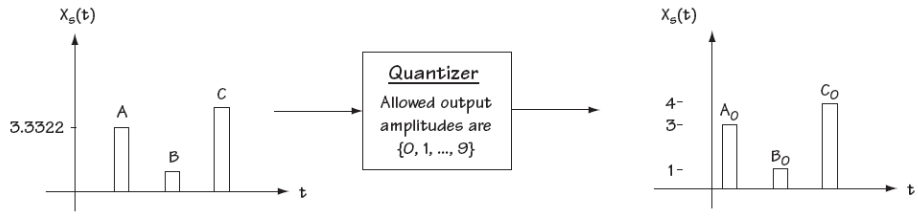


Figure 11: Quantizer with output amplitudes from the set {0, 1, 2, …, 9}

2.2.1 QUANTIZER PERFORMANCES

The quantization process introduces an error defined as the difference between the input and the output signals. This error is also referred to as quantization noise. A good quantizer has a small error term. If x is the input to the quantizer at time t and $\hat{x}$ is the output of the quantizer at that same time t the error signal is $e(x) = |x - \hat{x}|$.

- Mean Squared Error (MSE) : Instead of giving the error at each moment in time, we usually use the mean squared error to measure the performance of the quantizer. MSE is given by

$MSE = E[(x-\hat{x})^2] = \int_{-\infty}^{\infty} (x-\hat{x})^2 \, p_x(x)dx$, $p_x(x)$ is the probability density function of x which is the incoming amplitude into the quantizer.

- Signal to Quantization Noise Ratio (*SQNR):* refers to the ratio of the signal input power to the power of the error (or noise) introduced by the quantizer, $SQNR = \frac{\int_{-\infty}^{\infty} (x-x_m)^2 \, p_x(x)dx}{MSE}$

3 COMMUNICATIONS CHANNELS

On the channel, noise and fading affects the transmitted signal. Noise includes all kinds of random electrical disturbance from outside or within the system. Before investigation the modulations schemes in the next chapter, we briefly present several important channel models in communications to be familiar with their characteristics when studying the modulations schemes.

3.1 ADDITIVE WHITE GAUSSIAN NOISE CHANNEL

Additive white Gaussian noise (AWGN) channel is a simple channel model for analyzing modulations schemes. In this model, the channel does nothing but add a white Gaussian noise to the signal passing through it. The received signal is given by *r(t) = s(t) + n(t).* Gaussian means that the noise signal *n(t)* is a random process and a sample of is a random variable with a Gaussian distribution. The probability density function (PDF) of *n* can be written as

$$p(n=n_i) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{n_i^2}{2\sigma_n^2}}$$

Where σ_n^2 is the variance. White tells us that the noise *n(t)* has the autocorrelation function: $R_n(\tau) = \sigma_n^2 \delta(\tau)$ where $\delta(\tau)$ is the Dirac delta function. This shows that samples of *n(t)* are completely independent from one another.

Putting it all together, we know that *n(t)* is added to the sent signal, it's a random process, and we know $p(n_i)$ and $R_n(\tau)$.

Note that AWGN channel does not exist since no channel can have an infinite bandwidth. However, when the signal bandwidth is smaller than the channel bandwidth, many practical channels are approximately an AWGN channel. For example, the line-of-sight (LOS) radio channels are approximately AWGN channels when the weather is good. Wideband coaxial cables are also approximately AWGN channels because there is no other interference except the Gaussian noise.

3.2 MULTIPATH FADING CHANNEL

In a typical wireless environment, the signal from transmitter to receiver may undergo reflections from multiple scatterers, creating a multipath channel. The line of sight (LOS) path may or may not be available, depending on the propagation environment. Fading is a phenomenon that occurs when the amplitude and phase of the radio signal change rapidly over a short period of time or travel distance. Fading is caused by interference between two or more versions of the transmitted signal which arrive at the receiver at slightly different times. These waves called multipath waves are combined at the receiver antenna to give a resultant signal which can vary widely in amplitude and phase.

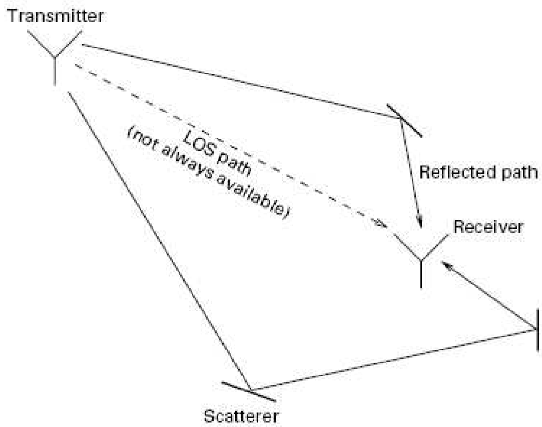


Figure 12: Multipath channel

4 BASEBAND AND BANDPASS MODULATIONS

The modulation is the process that turns the digital bit stream into a signal that is ready to be sent over the communication channel. There are two types of modulations: baseband modulations and bandpass modulations. The passband modulation and the baseband modulations exist for analog and digital signals. Baseband modulation is defined as a direct transmission transmission without frequency transform. Bandpass modulation takes incoming bits and outputs a waveform centered around new frequency.

Baseband modulations

1 BASEBAND PULSE MODULATIONS

A number of specific modulation techniques are used in pulse modulation. We first present pulse-amplitude and pulse-time modulations where outputs still analog, then we investigate the pulse-code modulation and its extensions including differential pulse-code modulation and Delta modulation.

1.1 PULSE-AMPLITUDE MODULATION (PAM)

Pulse amplitude modulation (see Figure 1) is a process that represents a continuous analog signal with a series of discrete analog pulses. The amplitude of the information signal at a given time could be coded as a binary number. PAM is the output of the sampling process. Two operations involved in the generation of the PAM signal are:

- Instantaneous sampling of the signal *s(t)* every T_s seconds, where $f_s = 1/T_s$ is selected according to the sampling theorem.
- Lengthening the duration of each sample obtained to some constant value T.

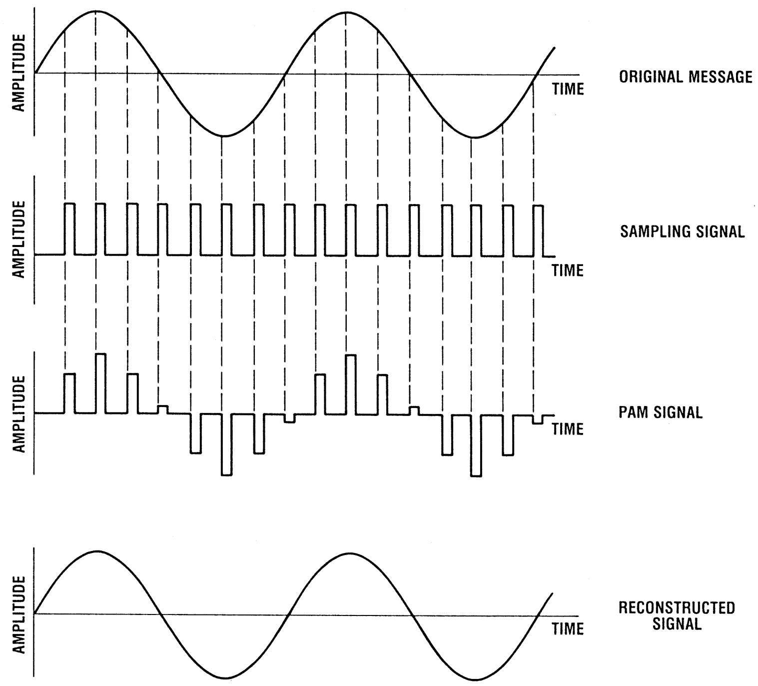


Figure 1: Pulse-Amplitude Modulation

These operations are jointly referred to as sample and hold (flat top sampling). One important reason for intentionally lengthening the duration of each sample is to avoid the use of an excessive channel bandwidth, since bandwidth is inversely proportional to pulse duration. The Fourier transform of the rectangular pulse $h(t)$ is shown in Figure 2.

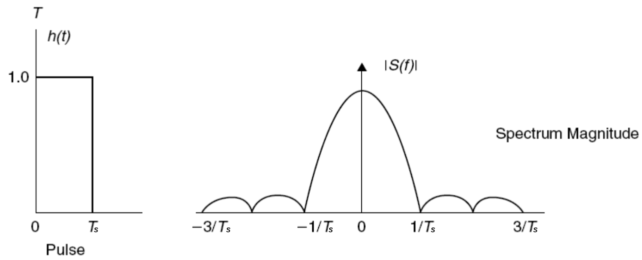


Figure 2: *h(t)* and its Fourier transform

Using flat-top sampling of an analog signal with a sample-and-hold circuit introduces amplitude distortion as well as a delay. The distortion caused by the use of PAM to transmit an analog signal is called the aperture distortion. The distortion may be corrected by use of an equalizer.

1.2 PULSE-TIME MODULATION (PTM)

The result with PAM modulation is a signal consisting of regularly-spaced constant width pulse whose amplitudes vary in proportion to the message signal. Another type of pulse modulation is pulse-time modulation (PTM). With PTM, samples of the signal are used to vary a parameter of pulse timing. If we use the pulse width as a timing parameter, the modulation is called Pulse-width modulation (PWM) or Pulse-Duration modulation (PDM). The result is signal consisting of constant amplitude pulses whose width is proportional to the message signal. Pulse-position is also another timing parameter that can be used in PTM. Pulse-position modulation (PPM) uses the samples of the message signal to modulate the positions of constant-width, constant-amplitude pulses. PWM and PPM are closely related each other. PPM signals are usually generated from PWM signals.

PTM, like PAM, requires the sampling rate to be greater than the Nyquist rate in order to respect the sampling theorem. PTM signals are much less sensitive to noise than PAM signals. In addition, PPM is very efficient because the pulses can be very narrow and therefore require very little power.

To generate PTM signals, the signal is sampled using a sawtooth signal. Then a fixed-amplitude pulse is generated for each sample. With PWM, the width of the generated pulse is proportional to the sampled value. It is determined by the amplitude of the message signal at the sampling instant (see Figure 3). To ensure proper pulse generation, a minimum $\tau_{\min}$ and a maximum $\tau_{\max}$ pulse width must be set. A guard time τ_g is required between pulses of maximum width to avoid one pulse to run into the next. To determine the modulation constant we calculate $\dfrac{\tau_{\max}-\tau_{\min}}{A_{\max}-A_{\min}}$ where $A_{\max}$ ($A_{\min}$) is the amplitude of the signal when the width of the pulse is $\tau_{\max}$ ($\tau_{\min}$).

In PPM, the position of the generated pulse is proportional to the sampled value and deduced from the PWM as shown in Figure 3.

The demodulation of both PWM and PPM signals is done using low-pass filtering because the spectra of these signals contain the original signal spectrum. Note that the pulses in a PPM signals are often made very narrow for efficient transmission and low-pass filtering of such narrow pulses would yield a reconstructed signal of very small amplitude. For this reason, the timing of the PPM signal is usually recovered before demodulation using a clock signal. It may be done using a flip-flop circuit as shown in Figure 4. To reconstruct the PWM signal, the trailing edge of a clock signal pulse sets the flip-flop and the next PPM pulse resets the flip-flop.

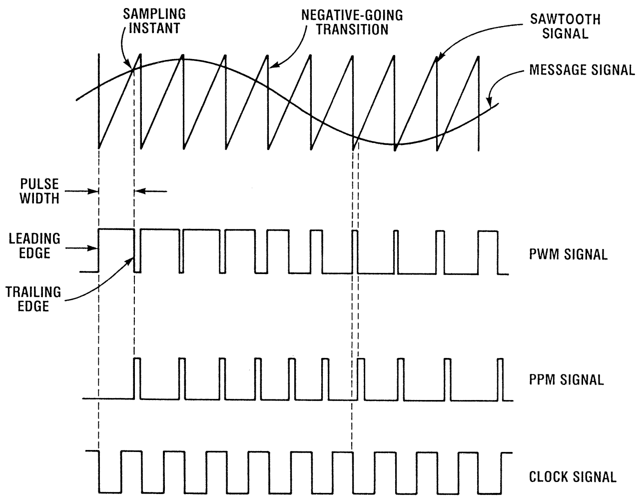


Figure 3: PWM and PPM signals

Note that the clock signal required for PPM timing recovery must either be transmitted along with the PPM signal or generated electronically. Any error in the timing of the regenerated clock signal, with respect to the PPM signal causes a phenomenon called offset error.

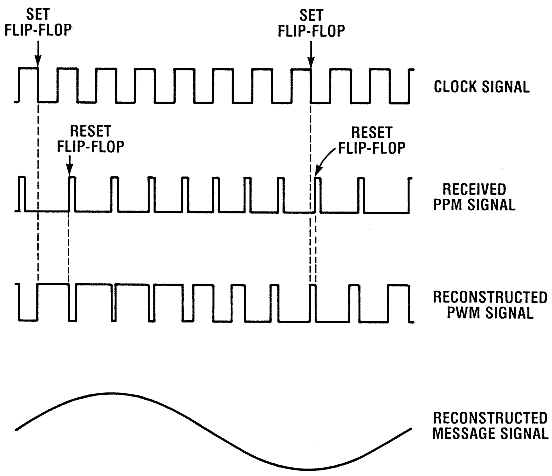


Figure 4: Demodulation of a PPM and PWM signals

1.3 PULSE-CODE MODULATION (PCM) AND EXTENTION FORMS

1.3.1 PCM

Pulse code modulation (PCM) is a digital scheme for transmitting analog data. In PCM the analog signal is sampled and each sample is quantized as detailed in chapter 1. The quantization process may be modeled mathematically as $\hat{x}_n = x_n + e_n$ where $\hat{x}_n$ represents the quantized value of x_n and e_n is the quantization (distortion) error.

1.3.1.1 CHOICE OF THE NUMBER OF QUANTIZATION LEVELS

Each analog sample is transmitted into a PCM word consisting of groups of bits. The PCM word size can be described by the number of quantization levels that are used for each sample. In digital telephone, PCM

samples the analog waveform 8000 times per second and converts each sample into an 8-bit number yielding 256 levels. The choice of the number of quantization levels *L*, or bits per sample, depends on the magnitude of quantization distortion that one is willing to tolerate with the PCM format.
It is useful to develop a general relationship between the PCM word (the required number of bits *l* per analog sample) and the allowable quantization distortion. For a uniform quantization with *L* levels, let $|e|$ the magnitude of quantization error specified as a fraction *p* of the peak-to-peak analog voltage V_{pp}, $|e| \le pV_{pp}$. Since the quantization error can not be larger than *q/2* (q volts is the step size between two levels, see Figure 5), we can write
$|e|_{\max} = \frac{q}{2} = \frac{V_{pp}}{2L}$. We have

$$\frac{V_{pp}}{2L} \le pV_{pp} \Leftrightarrow L \ge \frac{1}{2p} \Leftrightarrow 2^l \ge \frac{1}{2p}$$

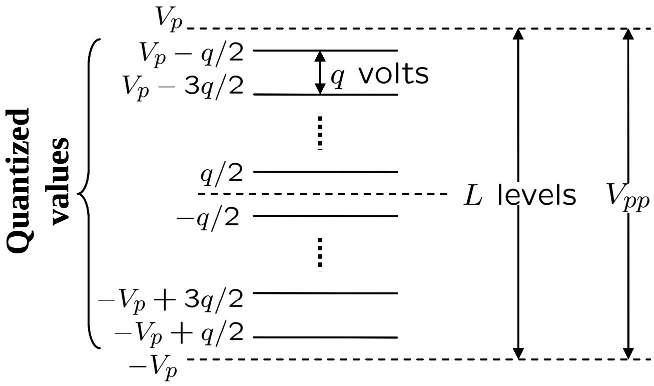


Figure 5: Peak-to-peak V_{pp} voltage for uniform quantization with step size q volts between two levels

1.3.1.2 QUANTIZATION ERROR

With L-level uniform quantizer for an analog signal with peak-to-peak voltage range $V_{pp}=2V_p$ volts. As in the previous subsection, the step size between quantization levels is denoted q volts and the quantization error e is uniformly distributed over the interval q. The probability of error *p(e)* is then equal to 1/q. Since the mean of quantization error *e* is zero, its variance will be

$$\sigma^2 = \int_{-q/2}^{q/2} e^2 p(e) de = \int_{-q/2}^{q/2} e^2 \left(\frac{1}{q}\right) de = \frac{q^2}{12}$$

and the power of the analog signal (normalized to 1Ω) is

$$(V_p)^2 = \left(\frac{V_{pp}}{2}\right)^2 = \left(\frac{L_q}{2}\right)^2 = \frac{L^2 q^2}{4}$$

The signal to quantization noise error (SNQR) = $3L^2$

1.3.1.3 SYMBOL

To transmit the sequence of bits (also called bit stream), groups of *b* bits form symbols from a finite symbol set $M = 2^b$. A system using a symbol set size of *M* is referred to as an M-ary system. When pulse modulation is applied to a binary symbol (*b=1*), the resulting binary wave form is called a pulse code modulation waveform. When pulse modulation is applied to a nonbinary symbol (*b>1*), the resulting waveform is called M-ary pulse modulation waveform.

1.3.1.4 THE TELEPHONE QUANTIZER

The standard quantizer used in the telephone system achieves a non uniform quantization (Figure 6 a). It is done by first distorting the original signal with a logarithmic compression characteristic (Figure 6b), then using a uniform quantizer (Figure 6 c). For small magnitude signals the compression characteristic has much steeper slope than for large magnitude signals. Thus, a given signal change at small magnitude will carry the uniform quantizer through more steps than the same change at large magnitudes. The compression characteristic effectively changes the distribution of the input signal magnitudes so that there is no a preponderance of low magnitude signals at the output of the compressor. After the compression, the distorted signal is used as an input to a uniform quantizer. At the receiver, an inverse phenomenon called expansion is applied so that the overall transmission is not distorted. This processing is usually referred to as companding with the ‘com’ coming from compression and the ‘panding’ coming from expanding.

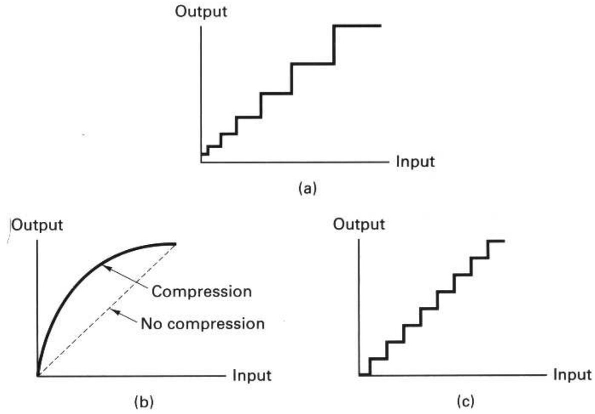


Figure 6: non uniform quantization for telephone systems

1.3.1.5 COMPANDING

The implemented PCM systems use a piecewise linear approximation to the logarithmic compression characteristic. In North America, a μ-law compression characteristic is mainly used. It is described by the expression:

$$|y| = y_{\max} \frac{\ln(1+\frac{\mu|x|}{x_{\max}})}{\ln(1+\mu)}, \ 0 \le |x| \le x_{\max}$$

where μ is a positive constant to define the amount of compression typically = 255, x and y represent input and output voltages, and x_{max} and y_{max} are the maximum positive values of the input and output voltages. Note that μ =0 corresponds to uniform quantization.

A-law is another compression characteristic used mainly in Europe. It is described by the expression:

$$|y| = \begin{cases} y_{\max} \dfrac{A(\frac{|x|}{x_{\max}})}{1+\ln(A)}, & 0 < \dfrac{|x|}{x_{\max}} \le \dfrac{1}{A} \\ y_{\max} \dfrac{1+\ln(\frac{A|x|}{x_{\max}})}{1+\ln(A)}, & \dfrac{1}{A} < \dfrac{|x|}{x_{\max}} < 1 \end{cases}$$

where A is a positive constant to define the amount of compression. In Europe the standard value for A is 87.6. Figure 7 shows μ-law and A-law for several values of μ and A respectively.

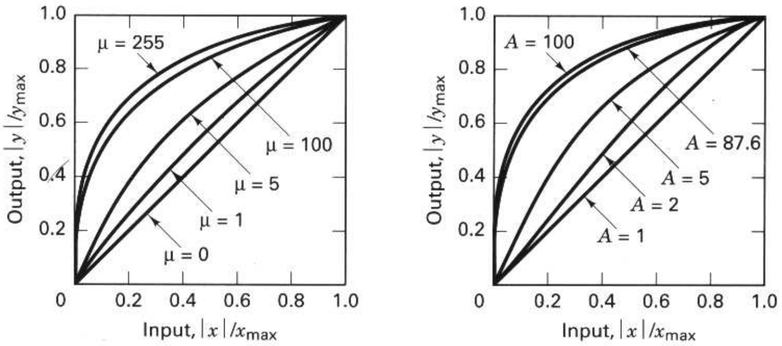


Figure 7: μ -law and A-law compression characteristics

1.3.2 DIFFERENTIAL PCM (DPCM)

In PCM, each sample of the signal is encoded independently to all the others. However, most source signals sampled at the Nyquist rate or faster exhibit significant correlation between successive samples and the change in amplitude between successive samples is relatively small. Consequently, an encoding scheme that exploits the redundancy in the samples will result in a lower bit rate for the source output because differences between samples are expected to be smaller than the actual sampled amplitudes; fewer bits are required to present the differences. Briefly speaking, DPCM encodes the PCM values as differences between the current and the previous value

This general approach is refined by predicting the current sample based on the p previous samples. Let $\hat{x}_n$ the predicted value of x_n defined as $\hat{x}_n = \sum_{i=1}^{p} a_i x_{n-i}$. a_i are the predictor coefficients and selected to minimize some functions of the error between $\hat{x}_n$ and x_n.

Consider the block diagram of a practical DPCM including both encoder and decoder (see Figure 8).

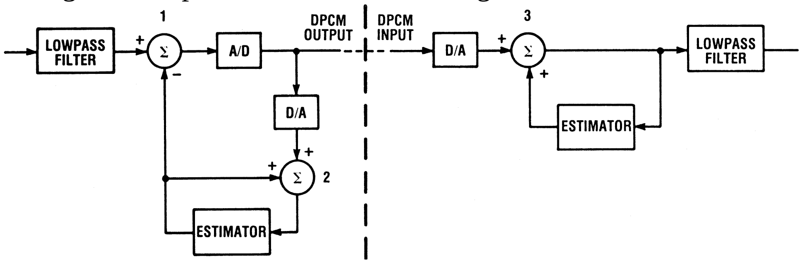


Figure 8: Encoder and decoder

In this configuration, the input of the estimator is $\tilde{x}_n$ and the output is denoted $\hat{\tilde{x}}_n$. The difference d_n between x_n and $\hat{\tilde{x}}_n$ is the input of the Quantizer (A/D block).

Note that for audio, this type of encoding reduces the number of bits required per sample by about 25% compared to PCM. Adaptive Differential PCM (ADPCM) is a variant of DPCM that varies the size of the quantization step to allow further reduction of the required bandwidth for a given signal-to-noise ratio.

1.3.3 DELTA MODULATION (DM)

DM may be viewed as a simplified form of DPCM in which a two-level (1-bit) quantizer is used in conjunction with a fixed predictor. The simplest form of DM will use just one bit to code each sample depending on whether the sample value is greater than or less than the previous sample. The encoder transmits 1 to indicate that the amplitude of the original message signal is greater than the amplitude of the estimated message signal or 0 to indicate that the amplitude of the original message signal is less than the amplitude of the estimated message signal. The decoder reconstructs the message signal using the serial output of the 1-bit words form the encoder. The decoder forms the constructed signal the same way as the encoder formed the estimated message: adding or subtracting a quantization interval from a running total of the previous quantization interval. The block diagram of a DM encoder called linear DM is shown in Figure 9.

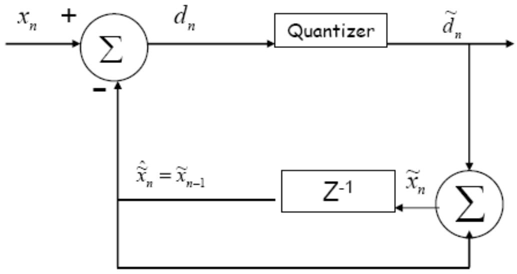


Figure 9: Block diagram of a delta modulation encoder

The estimated (predicted) value of x_n is the previous sample x_{n-1} modified by the quantization noise e_{n-1}.

$\hat{\hat{x}}_n = \tilde{x}_{n-1} = \hat{\hat{x}}_{n-1} + \tilde{d}_{n-1} = x_{n-1} + e_{n-1}$, because $e_n = \tilde{d}_n - d_n = \tilde{d}_n - (x_n - \hat{\hat{x}}_n)$

At any given sampling rate, the performance of the DM encoder is limited by two types of distortions (see Figure 10)

- Slope-overload distortion due to the use of step size that is too small to follows portions of the waveform that have a steep slope.
- Granular noise that results from using a step size that is too large in parts of the wave form having a small slope.

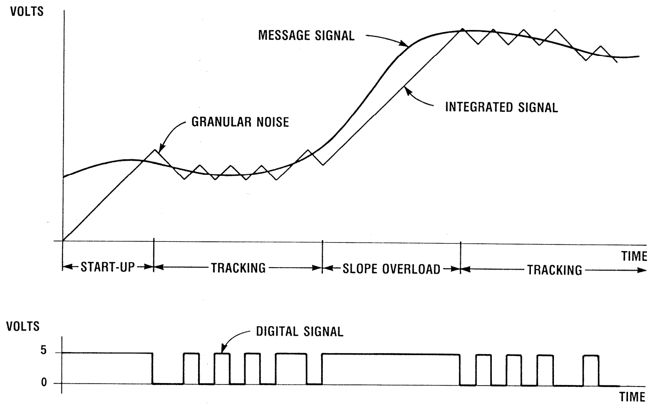


Figure 10: Examples of distortion in DM encoder

The need to minimize both of these two types of distortion results in conflicting requirements in the selection of the step size. One solution is to employ a variable step size that adapts itself to the short-term characteristics of the source signal.

2 BASEBAND MODULATIONS AND CODING

Baseband modulation is defined as a direct transmission without frequency transform. It is usually used for short distance transmission (local area networks, data storage devices). The baseband modulator turns the bit stream into a waveform centered around 0 Hz (see Figure 11). There are many baseband modulations, each one uses its own unique waveform around 0 Hz to send across the channel.

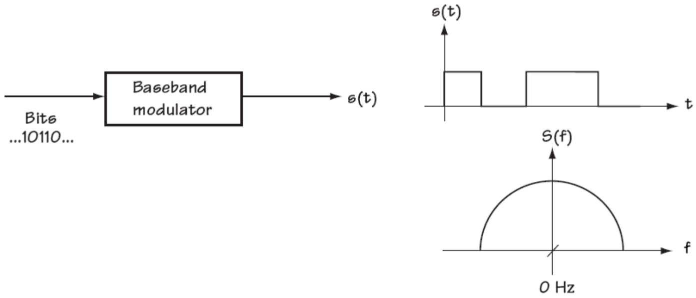


Figure 11: Baseband modulation

2.1 NON-RETURN TO ZERO

With NRZ modulation, the waveform created never returns to 0 (zero) volts. The first one is the non-return to zero-level (NRZ-L) modulation that represents a symbol 1 by a positive square pulse (+V) with length T_b and a symbol 0 by a negative square pulse (-V) with length T_b (see Figure 12).

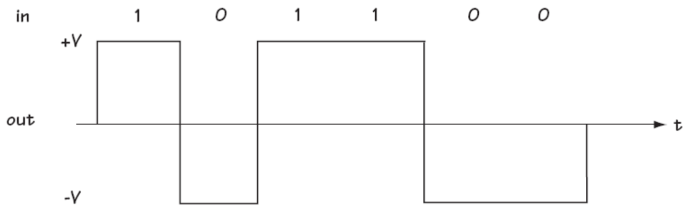


Figure 12: NRZ-L modulation

Another one is the NRZ-Mark (NZR-M) modulation where +V for a bit time T_b, or –V for bit time T_b are used as follows: if the bit is a 0, the signal level doesn't change (see Figure 13). If the bit is a 1, the signal level changes (e.g., from +V to –V).

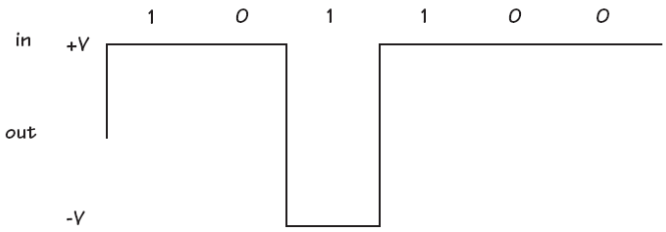
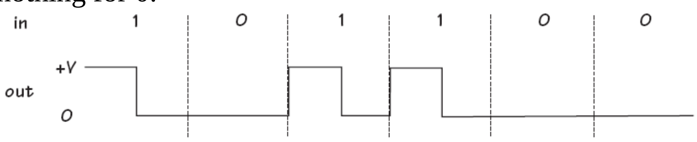


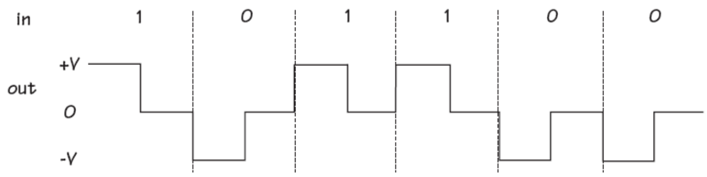
Figure 13: NRZ-M Modulation

2.2 RETURN TO ZERO

These modulators make sure that, for at least some of the time, the transmitted signal sits at 0. First is the unipolar return to zero modulation (unipolar RZ) that represents a symbol 1 by a positive pulse with length $T_b /2$ and nothing for 0.

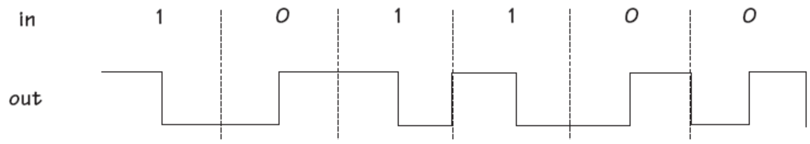


Another one is the bipolar return to zero modulation (bipolar RZ) that represents a 1 by a positive pulse with length $T_b /2$ and 0 by a negative pulse with length $T_b /2$.

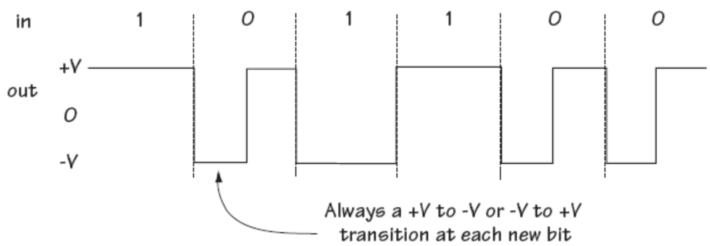


2.3 PHASE-ENCODED MODULATIONS

This group of modulation includes the biphase codes that use half period pulses with different phases according to certain encoding rules in the waveform. Among the most popular is the Manchester coding modulator, in which 1 is mapped to a waveform that starts at +V and ends up at 0, and where bit 0 becomes the waveform starting out at 0 and ending at +V.

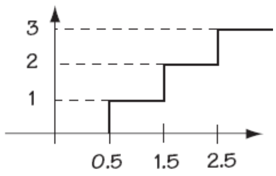


Another popular member is the Miller Coding modulator. The output signal is done so that there’s a transition (e.g., from +V to –V) between bits. A 0 is represented by a transition in the midpoint of the bit. A 1 is represented by no transition in the midpoint of the bit. A transition is placed at the end of each bit. (Another choice could be midpoint transition for 1 and no transition for 0).

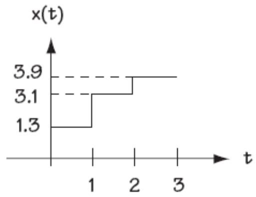


Chapter 2
Problems

I. Consider the quantizer shown in figure (a) below. Draw the output signal of the quantizer $\hat{x}$(t) as a function of *t* if the input signal is x(t) given in figure (b).



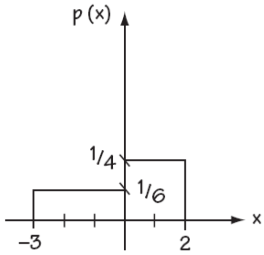
(a)



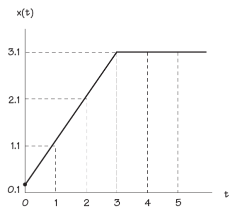
(b)

II. Consider the signal s(t) given by: $s(t) = s_1(t) * s_2(t)$, where $S_1(f) = 0, |f| > 2KHz$, $S_2(f) = 0, |f| > 3KHz$. What is the minimum sampling rate that ensures that *s(t)* is completely recovered from its samples. Justify your answer.

III. Consider a quantizer with an input signal x(t). The probability density function of x, p(x), is given in the following figure
   a) Draw a mid-tread uniform quantizer with 7 levels and step size equal to 1 and –3 is its smallest output value.
   b) Evaluate the MSE of this quantizer given the input x(t).
   c) Evaluate the SQNR.



IV. Consider a Delta Modulation (DM) with fixed step equal to 1. The figure shows the input x(t) (t in seconds). Determine the output of this modulator when the input x(t) is sampled at a rate of one sample per second.



V. Consider a Delta Modulation (DM) with fixed step equal to 0.2 . The input x(t)=t^2+0.01 (t in seconds). x(t) is sampled at a rate of 5 samples per second. Over the time 0 s to 2 s, determine the input to the DM modulator, the output of the DM modulator and the SQNR over this period.

VI. Given the sequence S= 11101010110000 generated by delta Δ-modulator. Binary 0 and 1 represent decrements and increments respectively. The input of the modulator is y(t) sampled at a rate of 1 sample per second.
   i. Assume zero initial conditions and Δ=1, find the values of the demodulator output and draw it. (6 pts)
   ii. If the sequence S given above is at the output of an adaptive Δ-Modulator with two step size Δ=1 and Δ_2=2 working as follows: Δ_2=2 is used if the generated bit is similar to the previous one, otherwise Δ=1 is used. Draw in this case the demodulator output assuming zero initial conditions.

VII. Consider a DPCM encoder. At its input, the samples x(n) are assumed to be stochastic variables: x(n)= u(n)+u(n -1) where u(n) is a zero mean white signal with variance =1. If each sample is predicted based on one previous sample, calculate the optimal predictor coefficient that minimizes the mean squared error. Then determine the mean squared error.

VIII. The information in an analog waveform, whose maximum frequency f_m = 4KHz, is to be transmitted using a 16-ary PAM system. The quantization distortion must not exceed ±1% of the peak-to-peak analog signal.

a) What is the minimum number of bits per sample or bits per PCM word that should be used in digitizing the analog signal?
b) What is the minimum required sampling rate, and what is the resulting bit transmission rate?
c) What is the 16-ary PAM symbol transmission rate?
d) What is the bandwidth efficiency for this system if the transmission bandwidth equals to 12 KHz.

IX. Consider an analog signal with maximum frequency f_m=3 KHz to be transmitted over M-ary system on a bandwidth equals to 10 KHz.

a) What is the used sampling rate if the bandwidth efficiency is equal to 4 bits/s/Hz and the quantization error should not exceed ± 2% of the peak-to peak analog signal?
b) What is the symbol rate when M=16.
c) If we use the same number of levels given in a) but the signal is sampled using Nyquist rate, determine the max quantization error as a % of the peak-to peak analog signal and the bandwidth efficiency.

X. Consider the sequence of bits 1110100101 at the input of a baseband modulator. Draw the waveform at the output of the modulator for the following cases:

a) NRZ-L modulator
b) NRZ-M modulator
c) Bipolar RZ
d) Manchester coding modulator

Demodulation and detection of signals in AGWN channel

1 INTRODUCTION

In case of baseband modulation, the received waveforms are in a pulse form and suffer from the intersymbol interference (ISI). The arriving pulses are not in the form of ideal pulses shapes; in addition to the noise, the channel adds timing offset, a phase offset, a frequency offset. The model of the received signal is: $r(t) = h(t) * s_i(t) + n(t)$ and the simple model in AWGN is $r(t) = s_i(t) + n(t)$ (Figure 1). The demodulator is needed to recover the pulse waveforms with the best possible signal-to noise ratio (SNR) free of any ISI. Equalization is a technique used to help accomplish this goal. The demodulation is then the recovery of a waveform to an undistorted baseband pulse. We define detection as the decision-making process of selecting the digital meaning of that waveform. (Note that some books use the terms demodulation and detection without difference).

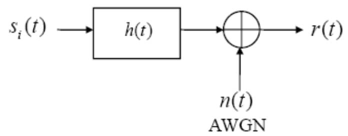


Figure 1: Received signal

As we will see in next chapters, for bandpass case the model of the detection process is virtually identical to the baseband model. This is because the received bandpass waveform is first transformed to a baseband waveform before the detection takes place.

1.1 SYNCHRONIZATION

The demodulation of a signal requires that the receiver be synchronized with the transmitter signal as it is received at the input of the receiver. The synchronization must be for:

- Carrier synchronization (for bandpass modulation): The receiver is on the same frequency as the transmitted signal, adjusted for the effects of Doppler shifts.
- Bit synchronization: The receiver is aligned with the beginning and end of each bit interval.
- Word synchronization (also known as frame synchronization): The receiver is aligned with the beginning and end of each word in the transmitted signal.

If the synchronization in the receiver is not precise for any of these operations, then the bit error probability of the receiver will not be the same. The design of a receiver is an area for which standards are traditionally not specified. It is an art that enables one manufacture to offer better performance in its equipment compared to a competitor.

1.2 SNR PARAMETER FOR DIGITAL COMMUNICATIONS SYSTEMS

The SNR is defined as the ratio of the signal power to the noise power $SNR = \frac{Signal\,power\,(P)}{Noise\,power\,(N)}$ (often expressed in dB). For simplicity we assume in this chapter that the noise is an AGWN that has a probability density function of amplitude that is Gaussian and that the noise spectral density is flat with frequency with the two sided Power Spectral Density of $N_0/2$, where N_0 is the thermal noise directly proportional to bandwidth (B) and temperature (T_a). The amount of thermal noise to be found in 1 Hz of bandwidth is given as: $N_0 = kT_a$ (W/Hz). Where k is Boltzmann's constant (1.3803 $.10^{-23}$ J/K), T_a is the absolute temperature (K).

In digital communication, we more often use E_b/N_0, which is a normalized version of SNR. E_b is bit energy and can be described as signal power P times the bit time T_b. N_0 is the noise power spectral density and can be described as noise power by bandwidth B. Let R_b=1/T_b be the bit rate, we can write: $\frac{E_b}{N_0} = \frac{P/R_b}{N/B}$.

1.3 CAPACITY OF AWGN CHANNEL

According to Shannon's information theorem, the capacity of an AWGN channel free from intersymbol interference is defined by: $C = B\log_2(1+SNR)\ bits/s$. Where B is the channel bandwidth and SNR is measured at the channel output.

1.4 NYQUIST BANDWIDTH CONSTRAINT AND RC FILTER

Nyquist investigated the problem of specifying a received pulse shape so that no ISI occurs at the detector. He showed that the theoretical minimum required system bandwidth to detect *Rs* [symbols/s] without ISI is *Rs/2* [Hz]. Equivalently, a system with bandwidth *W=1/2T=Rs/2* [Hz] can support a maximum transmission rate of *2W=1/T=Rs* [symbols/s] without ISI :

$$\frac{1}{2T} = \frac{R_s}{2} \le W \Rightarrow \frac{R_s}{W} \ge 2 \ \text{[symbol/s/Hz]}.$$

The names Nyquist filter and Nyquist pulse are often used to describe the general class of filtering and pulse-shaping that satisfy zero ISI at the sampling points. The transfer function of the Nyquist filter in frequency domain is obtained by convolving a rectangular function with any real even-symmetric frequency function. Its shape can be represented by a sinc(t/T) function multiply by another time function. Amongst the class of Nyquist filters, the most popular one is the raised cosine. The transfer function of the raised cosine (RC) filter can be expressed by

$$H_f(f) = \begin{cases} 1 & \text{for } |f| < 2W_0 - W \\ \cos^2\left[\frac{\pi}{4}\frac{|f|+W-2W_0}{W-W_0}\right] & \text{for } 2W_0 - W < |f| < W \\ 0 & \text{for } |f| > W \end{cases} \quad \text{and} \quad h_f(t) = 2W_0(\text{sinc}(2W_0 t))\frac{\cos[2\pi(W-W_0)t]}{1-[4(W-W_0)t]^2}$$

Where W is the absolute bandwidth and W_0 = 1/2T represents the minimum Nyquist bandwidth for the rectangular spectrum and the -6dB bandwidth (or half amplitude point) for the raised-cosine spectrum. The difference W-W_0 is termed Excess bandwidth which means additional bandwidth beyond the Nyquist minimum (for the rectangular spectrum, W=W_0). The Roll-off factor $r = \frac{W-W_0}{W_0}$ where $0 \le r \le 1$ specifies the required exess bandwidth as a fraction of W_0 and charaterizes the steepness of the filter roll off. Figure 2 shows the raised cosine characteristic for roll-off values of r=0, 0.5 and 1. r=0 roll-off is the Nyquist minimum bandwidth case. When r=1 the required excess bandwidth is 100% and tails are quite small.

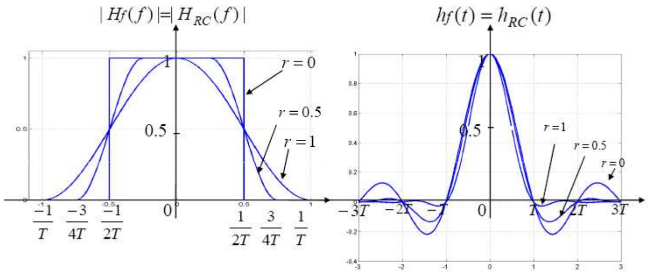


Figure 2: Raised-Cosine Filter: system transfer function and system impulse response

The general relationship between required bandwidth and symbol transmission rate involves the filter roll-off factor r and can be stated as $W = (1+r)\frac{R_s}{2}$. For double-sideband signals, the required DSB bandwidth and the symbol rate are related by $W_{DSB} = (1+r)R_s$.

2 DEMODULATION

We will consider two demodulators: the correlation demodulator and the matched filter demodulator.

2.1 CORRELATION DEMODULATOR

A correlation demodulator is a bank of correlators, each of which computes the projection of the received signal r(t) onto a particular orthogonal basis function {$\varphi_k(t)$} (Figure 3). For the k^{th} correlator, we can write in case of AWGN channel:

$$r_k = \int_0^T r(t)\varphi_k(t)dt = \int_0^T (s_i(t) + n(t))\varphi_k(t)dt = s_{i,k} + n_k$$

The noise samples n_k are uncorrelated, and since they are Gaussian random variables, this implies they are statistically independent. As a result, r_k conditioned on $s_{i,k}$ is a Gaussian random variable with mean $s_{i,k}$ and variance $N_0/2$, and all {r_k} conditioned on the transmitted signals $s_{i,k}$ are statistically independent. The samples {r_k} are sufficient statistics for detection of the transmitted signal. Thus, the correlation demodulator converts the continuous-time system into a discrete-time system without losing the information.

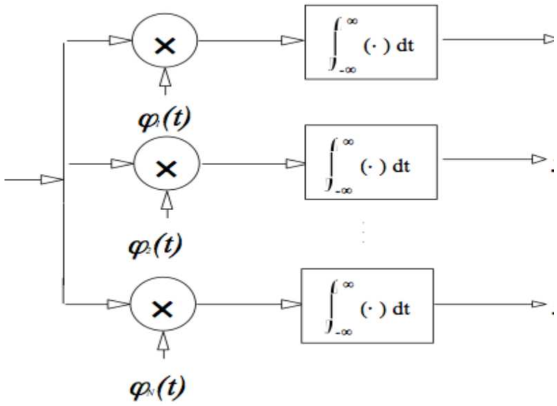


Figure 3: Correlation demodulator

2.2 MATCHED FILTER DEMODULATOR

In this case, instead of performing a set of correlations, we can pass the received signal through a set of linear filters and sample at t = T.

$$r_k = \int_0^T r(t)h_k(T-t)dt$$

Where $h_k(t)$ is the k^{th} filter impulse response ($h_k(t) = \varphi_k(T-t)$ yields the k^{th} output of the correlation demodulator.). This matched-filter demodulation is illustrated in Figure 4.

The matched filter output at the sampling time is realized as the correlator output:

$$r_k = r(t) * h(t) = \int_0^T r(t)h_k(T-t)dt\,.$$

The matched filter is designed to provide the maximum signal to noise power ratio at its output for a given transmitted symbol waveform. We need to calculate the SNR of the output signal in function of the input signal and the noise, then we determine the transfer function of the filter so that the SNR is maximal. In other words, the filter response is matched to the transmitted signal.

Let s(t) be the signal and E_s is its energy. The two sided power spectral density of the input noise is $N_0/2$. We can demonstrate that the max SNR is equal to $2E_s/N_0$ and we determine the optimum filter transfer function that produces this maximum SNR.

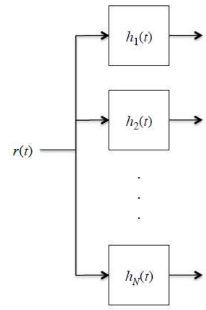


Figure 4: Matched filter demodulator

The problem is to design the receiver filter *h(t)* such that the SNR is maximized at the sampling time when $s_i(t), i = 1,...,M$ is transmitted. The optimum filter, the Matched filter, is the time-reversed and delayed version of the conjugate of the transmitted signal. It is given by[1]:

$h_{opt}(t) = s_i^*(T-t)$ and $H_{opt}(f) = S_i^*(f)\exp(-j2\pi fT)$.

The Fourier transform of a matched filter output with the matched signal as input is, except for a time delay factor, proportional to the ESD of the input signal: $Z(f) = |S(f)|^2 \exp(-j2\pi fT)$

The output signal of a matched filter is proportional to a shifted version of the autocorrelation function of the input signal to which the filter is matched: $z(t) = R_s(t-T) \Rightarrow z(T) = R_s(0) = E_s$

The output SNR of a matched filter depends only on the ratio of the signal energy to the PSD of the white noise at the filter input: $\max\left(\frac{S}{N}\right)_T = \frac{E_s}{N_0/2}$.

3 OPTIMUM DETECTION

Both the correlation demodulator and the matched-filter demodulator produce as outputs the sequence $r = (r_1,\dots, r_k, \dots)$ for each received signal. These are sufficient statistics that can be used to decide what signal was transmitted, i.e., they can be used to detect the transmitted signal. This is done optimally by determining the most probable transmitted signal given the received statistics. That is, we wish to find the transmitted signal s_i that maximizes the following probability:

$$P(s_i / r) = \frac{P(r/s_i)P(s_i)}{\sum_k P(r/s_k)P(s_k)}$$

The probability $P(s_i / r)$ is known as the a posteriori probability, and so this detector is called the maximum a posteriori (MAP) detector. When the a priori probabilities are all equal, the MAP detector reduces to the maximum likelihood (ML) detector.

3.1 DETECTION OF BINARY SIGNAL

During a given signaling interval T, a binary baseband/bandpass system will transmit one of two kinds of signals denoted $s_1(t)$ and $s_2(t)$.

The energy of the two signals are $E_1 = \int_0^T s_1(t)^2 dt$ and $E_2 = \int_0^T s_2(t)^2 dt$. In general these two signals may be correlated. We define $\rho_{12} = \frac{1}{\sqrt{E_1E_2}}\int_0^T s_1(t)s_2(t)dt$ as the correlation coefficient of $s_1(t)$ and $s_2(t)$. For any binary channel, the transmitted signal over a symbol duration T is represented by :

[1] The proof is an application of the Cauchy-Schwarz Inequality.

$$\begin{cases} s_1(t) & 0 \le t \le T \quad \textit{for bit } 1 \\ s_2(t) & 0 \le t \le T \quad \textit{for bit } 0 \end{cases}$$

The received signal degraded by noise *n(t)* assumed to be a zero mean AWGN, is given by

$r(t) = s_i(t) + n(t)$

Figure 5 shows a typical digital receiver where demodulation and detection functions are represented in two steps: demodulate and sample block and detection block. Within the demodulate and sample block, the receiving filter performs waveform recovery in preparation for the next step i.e. the detection. The goal of the receiving signal is to recover a baseband pulse with the best possible SNR free of any ISI. The optimum receiving filter could be a matched filter or a correlator filter. An equalizing filter may follow the receiving filter. It is used only for the systems where channel induced ISI can distort the signals. In most case, when equalizer is used, a single filter would be designed to incorporate both functions, best possible SNR and compensation of the distortion caused by the transmitter and the channel. Such composite filter is simply called the receiving filter.

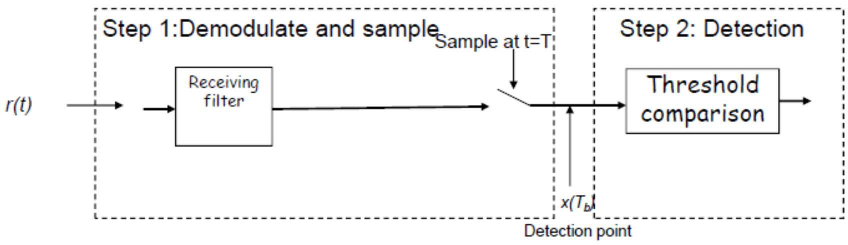


Figure 5: Demodulation and detection

At the end of each symbol duration (T = T_b because each symbol consists of one bit), the output of the sampler, the predetection point (end of step 1), yields a sample x which has a voltage directly proportional to the energy of the received symbol and inversely proportional to the noise. In step 2, a decision (detection) is made regarding the digital meaning of the sample. The output of step 1 yields $x=r(T) = u_i(T) + n_0(T)$ (in simple form $x=u_i+n_0$), where u_i is the desired signal component, and n_0 the noise component is a zero mean Gaussian random variable. x is then a Gaussian random variable with a mean of either u_1 or u_2 depending on whether a binary 1 or binary 0 was sent. As given in chapter 1, the probability density function (PDF) of the Gaussian random noise n_0 can be expressed as

$$p(n_0) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{n_0^2}{2\sigma_0^2}}$$

Where σ_0^2 is the noise variance. We can express the conditional PDFs $p(x/s_1)$ and $p(x/s_2)$ by

$$p(x/s_1) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(x-u_1)^2}{2\sigma_0^2}} \text{ and } p(x/s_2) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(x-u_2)^2}{2\sigma_0^2}}$$

These conditional PDF are illustrated in Figure 6. The rightmost conditional PDF $p(x/s_1)$ called likelihood of s_1 illustrate the PDF of the random variable x given that symbol s_1 was transmitted. Similarly, the leftmost conditional PDF $p(x/s_2)$ called likelihood of s_2 illustrate the PDF of the random variable x given that symbol s_2 was transmitted. The threshold used to make the decision is noted γ.

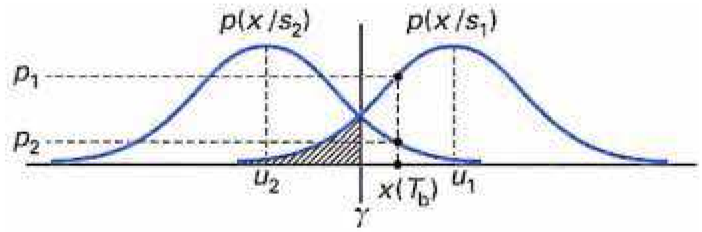


Figure 6: conditional PDFs *p(x/s₁)* and *p(x/s₂)*

In step 2 detection is performed by comparing x and γ. It is done by a function that takes input *x* and outputs a guess on the transmitted symbol. The decision-making criterion is as follows:

- Hypothesis H1: If $x > \gamma$, this means that the signal $s_1(t)$ was sent and hence a binary 1 is detected.
- Hypothesis H2: If $x < \gamma$, this means that the signal $s_2(t)$ was sent and hence a binary 0 is detected.
- If $x = \gamma$ the decision can an arbitrary one.

The value of threshold level γ is chosen to minimize the probability of error. For equally likely signals, the optimum threshold passes through the intersection of the likelihood functions. It is the mid-point of u_1 and u_2.

For the case where the signal classes are equally likely, the detector that minimizes the error probability is the maximum likelihood detector (ML detector). The decision stage effectively selects the hypothesis that corresponds to the signal with the maximum likelihood (ML). The detector chooses $s_1(t)$ if $p(x/s_1) > p(x/s_2)$.Otherwise the detector chooses $s_1(t)$.

3.2 BIT ERROR PROBABILITY

In binary decision-making , two errors can occur. An error *e* will occur

- when $s_1(t)$ is sent and the output signal $x(T)$ is in the decision zone of $s_2(t)$.
- when $s_2(t)$ is sent and the output signal $x(T)$ is in the decision zone of $s_1(t)$.

The probability of such errors are respectively

$$p(e/s_1) = \int_{-\infty}^{\gamma} p(x/s_1)dx \quad \text{and} \quad p(e/s_2) = \int_{\gamma}^{\infty} p(x/s_2)dx$$

The error probability is then $P(e) = P(s_1).p(e/s_1) + P(s_2).p(e/s_2)$. We can simplify this equation by making a simple realization that the signals are equally likely, which means that the probability to sent s_1 $P(s_1) =$ the probability to sent s_2 $P(s_2) = ½$. And because of the symmetry of the probability density functions $p(e/s_1) = p(e/s_2)$, consequently $P(e) = p(e/s_1) = p(e/s_2)$.

$P(e)$ is numerically equal to the area under the tail of either likelihood function $p(x/s_1)$ or $p(x/s_2)$ falling on the incorrect side of the threshold. $P(e)$ is computed by $\int_{-\infty}^{\gamma} p(x/s_1)dx$ or by $\int_{\gamma}^{\infty} p(x/s_2)dx$ where γ is chosen according to the strategy of minimum error criterion that gives the optimum threshold for minimizing the probability of making an incorrect decision ($\gamma = \frac{u_1 + u_2}{2}$).

$P(e) = \int_{\gamma}^{\infty} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(x-u_2)^2}{2\sigma_0^2}} dx$. Let $v = \frac{x - u_2}{\sigma_0}$ then $\sigma_0 dv = dx$ we can write $P(e)$ as

$$P(e) = \int_{(u_1-u_2)/2\sigma_0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} dv = Q(\frac{(u_1 - u_2)}{2\sigma_0})$$

Where Q(x), called co-error function is commonly used to represent the probability under the tail of the Gaussian PDF. It is defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{v^2}{2}} dv$$

To minimize *P(e)* in the context of an AWGN channel , we need to select the optimum receiving filter in step 1 and the optimum decision threshold in step 2. For the binary case, the optimum decision threshold is already given above according to the minimum error criterion. And we find $P(e) = Q(\frac{(u_1 - u_2)}{2\sigma_0})$. To minimize P(e), it is necessary to choose the filter that maximizes the argument of Q(x), i.e. $\frac{(u_1 - u_2)}{2\sigma_0}$. $(u_1 - u_2)$ is the difference of the desired signal components at the output of the filter at time t=T. The square of this difference signal is the instantaneous power of the difference signal. As shown in section 2, a matched filter achieve the maximum possible outpout SNR for a given known signal equal to $2E/N_0$.

Consider the case for binary signaling and the filter is matched to the input difference signal $[s_1(t)-s_2(t)]$, we can write the output SNR at time t=T as $(SNR)_T = \frac{(u_1 - u_2)^2}{\sigma_0^2} = \frac{2E_d}{N_0}$

$E_d = \int_0^T [s_1(t) - s_2(t)]^2 dt$ is the energy of the difference signal at the filter input.

We can rewrite for binary case the bit error probability (also known by Bit Error Rate BER):

$$P_b = Q(\frac{(u_1 - u_2)}{2\sigma_0}) = Q(\sqrt{\frac{E_d}{2N_0}})$$

By maximizing the output SNR, the matched filter provides the maximum distance (normalized by noise) between the two signals u_1 and u_2.

We have $E_d = \int_0^T [s_1(t) - s_2(t)]^2 dt = \int_0^T s_1(t)^2 dt + \int_0^T s_2(t)^2 dt - 2\int_0^T s_1(t)s_2(t)dt = E_1 + E_2 - 2\rho_{12}\sqrt{E_1E_2}$.

$$P_b = Q(\sqrt{\frac{E_1 + E_2 - 2\rho_{12}\sqrt{E_1E_2}}{2N_0}})$$

This expression indicates that P_b depends not only on the individual signal energies, but also on the correlation between them -1< ρ_{12} <1. It is interesting to discover that when $\rho_{12} = -1$, P_b is the minimum. Binary signals with $\rho_{12} = -1$ are called antipodal. When $\rho_{12} = 0$, the signals are orthogonal. If ρ_{12} = 1 the signals are correlated.

To develop a more general relationship in terms of received bit energy (E_b), if the two signals $s_1(t)$ and $s_2(t)$ have equal energies $E_b = E_1 = E_2$, $E_d = 2E_b(1-\rho)$, $\rho_{12} = \rho$

$$P_b = Q(\sqrt{\frac{E_b(1-\rho)}{N_0}})$$

3.3 BIT ERROR PROBABILITY OF NON RETURN-TO-ZERO

The NRZ-L is antipodal (bipolar) with

$$\begin{cases} s_1(t) = A & 0 \le t \le T \quad \textit{for bit } 1 \\ s_2(t) = -A & 0 \le t \le T \quad \textit{for bit } 0 \end{cases}$$

ρ is then equal to -1, $E_d = \int_0^T [s_1(t) - s_2(t)]^2 dt = 4A^2T$ and $E_b = A^2T$,

$$P_b = Q(\sqrt{\frac{2A^2T}{N_0}}) = Q(\sqrt{\frac{2E_b}{N_0}})$$

For NRZ-L, the optimum threshold is chosen as $\gamma = 0$.

If we consider a unipolar case where

$$\begin{cases} s_1(t) = A & 0 \le t \le T \quad \textit{for bit } 1 \\ s_2(t) = 0 & 0 \le t \le T \quad \textit{for bit } 0 \end{cases}$$

$$E_d = \int_0^T [s_1(t) - s_2(t)]^2 dt = A^2T \text{ , and } E_b = A^2T/2 \text{ , } P_b = Q(\sqrt{\frac{A^2T}{2N_0}}) = Q(\sqrt{\frac{E_b}{N_0}})$$

Because $s_2(t)$ is equal to 0 during the time T, this unipolar pulse fulfills the condition that $s_1(t)$ and $s_2(t)$ have zero correlation i.e. they form a set of orthogonal signal (ρ = 0).

We can see in Figure 7 a 3 dB error performance improvement for bipolar compared with unipolar signaling.

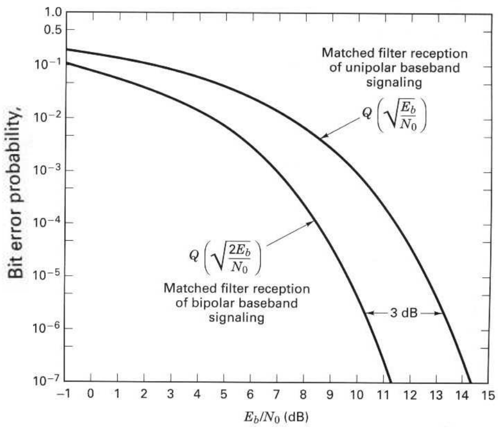


Figure 7: bit error probability as function of E_b/N_0 for unipolar and bipolar case.

3.4 BIT ERROR PROBABILITY OF RETURN-TO-ZERO

The bipolar RZ is

$$s_1(t) = -s_2(t) == \begin{cases} A & 0 \le t \le T/2 \\ 0 & \textit{elsewere} \end{cases}$$

ρ is then equal to -1, $E_d = \int_0^T [s_1(t) - s_2(t)]^2 dt = 2A^2T$ and $E_b = A^2T/2$,

$P_b = Q(\sqrt{\frac{A^2T}{N_0}}) = Q(\sqrt{\frac{2E_b}{N_0}})$ is the same as that of NRZ-L in terms of E_b/N_0 and the optimum threshold is also 0.

4 POWER SPECTRAL DENSITY OF BASEBAND MODULATIONS

In this part, we consider a general formula for power spectral density (PSD) calculation for digitally modulated baseband waveforms (also referred by binary line codes) used for most of the line codes. We know that most signals like voice signal and image signal are essentially random. Therefore digital signals derived from these signals are also random. Data signals are also essentially random. Assume the digital signal can be represented by

$$s(t) = \sum_{k=-\infty}^{\infty} a_k g(t-kT)$$

where a_k are discrete random data bits, g(t) is a signal of duration T (i.e., nonzero only in [0, T]. g(t) could be named as symbol function. It could be any signal with a Fourier transform. For example it could be a baseband symbol shaping pulse (or a burst of modulated carrier at passband). The random sequence $\{a_k\}$ could be binary or nonbinary. The power spectral density of s(t) is

$$\Psi_s(f) = \frac{|G(f)|^2}{T} \sum_{n=-\infty}^{\infty} R(n) e^{-jn\omega T}$$

where $w=2\pi f$. G(f) is the Fourier transform of g(t) and R(n) is the autocorrelation function of random sequence $\{a_k\}$ defined as R(n) = E$\{a_k a_{k+n}\}$ where E{x} is the probabilistic average of x. The equation of PSD shows that the PSD of a digitally modulated signal is not only determined by its symbol function but also is affected by the autocorrelation function of the data sequence. This equation is more convenient to be used for modulated baseband waveforms with R(n)=0 for n different de 0.

If we assume that the original binary data sequence has 1s and 0s equally likely. That is $P(0) = P(1) = 0.5$. For uncorrelated sequence $\{a_k\}$,

$$R(n) = \begin{cases} \sigma_a^2 + m_a^2, & n = 0 \\ m_a^2, & n \neq 0 \end{cases}$$

where σ_a^2 is the variance and m_a is the mean of the sequence. R_b =1/T is the data bit rate, the PSD expression (more convenient to be used for modulated signals with R(n) different de 0 for n different de 0) can be written as

$$\Psi_s(f) = \frac{|G(f)|^2}{T} \left[\sigma_a^2 + m_a^2 R_b \sum_{n=-\infty}^{\infty} \delta(f - nR_b) \right]$$

4.1 PSD OF NON RETURN-TO-ZERO

NRZ formats' symbol function is a square pulse with amplitude A in the interval [0,T] which can be expressed as *g(t)* and its Fourier transform can be easily found as G(f)

$$g(t) = \begin{cases} A, & 0 \le t \le T \\ 0, & \text{elsewhere} \end{cases} \qquad G(f) = AT \left(\frac{\sin \pi f T}{\pi f T} \right) e^{-j\pi f T}$$

For this waveform, the binary data sequence $\{a_k\}$: $a_k = \begin{cases} 1, & \text{for binary 1}, \quad p_1 = 1/2 \\ -1, & \text{for binary 0}, \quad p_0 = 1/2 \end{cases}$

Thus the autocorrelation function R(n):

$$R(n) = \begin{cases} p_1(1)^2 + p_0(-1)^2 = 1, & n = 0 \\ (p_0)^2(1)^2 + (p_1)^2(1)^2 + 2p_0p_1(1)(-1) = 0, & n \neq 0 \end{cases}$$

The PSD is then

$$\Psi_s(f) = A^2 T \left(\frac{\sin \pi f T}{\pi f T} \right)^2$$

In Figure 8 we set A = 1 and T = 1 for unity symbol pulse energy. This PSD is a squared sinc function with its first null at fT = 1.

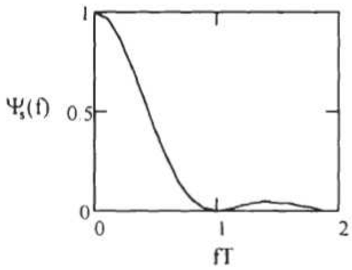


Figure 8: PSD of bipolar NRZ

4.2 PSD OF RETURN-TO-ZERO

For RZ formats, the pulse function is a square pulse with half-bit duration g(t) and its Fourier transform is G(f)

$$g(t) = \begin{cases} A, & 0 \le t \le \frac{T}{2} \\ 0, & \text{elsewhere} \end{cases} \qquad G(f) = \frac{AT}{2}\left(\frac{\sin \pi f T/2}{\pi f T/2}\right) e^{-j\pi f T/2}$$

The binary data sequence $\{a_k\}$ is: $$a_k = \begin{cases} 1, & \text{for binary 1}, \quad p_1 = 1/2 \\ -1, & \text{for binary 0}, \quad p_0 = 1/2 \end{cases}$$

Thus the autocorrelation function R(n): $$R(n) = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$$

The PSD is then

$$\Psi_s(f) = \frac{A^2 T}{4}\left(\frac{\sin \pi f T/2}{\pi f T/2}\right)^2$$

And Figure 9 shows the PSD. For unity average symbol energy $A^2T/2=1$ => $A=\sqrt{2}$. Compared with the PSD of NRZ format, this PSD is a stretched version with frequency axis scaled up twice and its bandwidths are double that of NRZ.

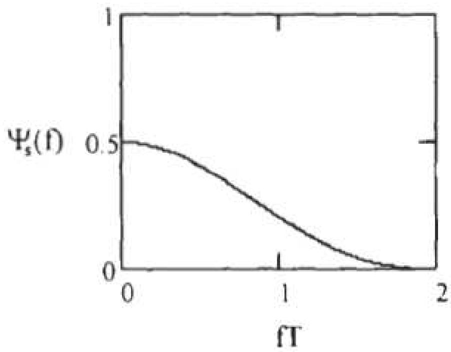


Figure 9: PSD of bipolar RZ

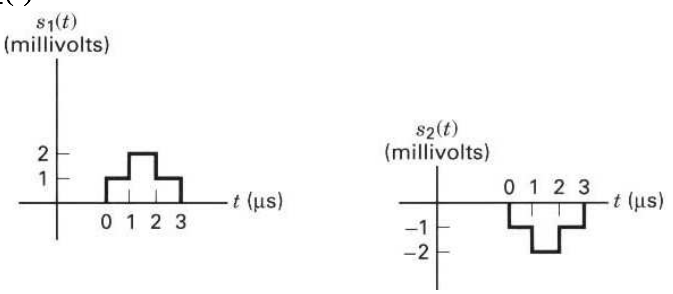
Chapter 3
Exercises

I. An audio signal in the frequency range 300 to 3300 Hz is sampled at 8000 samples/second. We may transmit these samples directly as PAM pulses or we may first convert each simple to a PCM format and use binary PCM format for transmission.

a) What is the minimum system bandwidth required for the detection of PAM pulses with no ISI and with a filter roll off characteristic of r=1?
b) Using the same filter roll off characteristic, what is the minimum system bandwidth required for the detection of binary PCM waveforms if the samples are quantized using 8 levels? 256 levels?

II. Consider a binary communication system where the signal component at the correlator receiver is u_1=1or u_2=-1 with equal probability. If the Gaussian noise at the correlator output has unit variance, find the probability of a bit error.

III. Consider a binary communication system that receives equally likely signals s_1(t) and s_2(t) plus AWGN. The receiving filter is a matched filter and the noise power spectral density N_0 is equal to 10^{-12} Watt/Hz. Calculate the bit error probability if s_1(t) and s_2(t) are as follows:



IV. Consider a binary communication system transmitting pulses $s_1(t)$ and $s_2(t)$ using NRZ modulation. The pulses are transmitted along a cable that attenuates the signal power from transmitter to receiver by 3 dB. Assuming a normalization relative to $1\,\Omega$ resistive load and the signals are equally likely and are coherently detected over a Gaussian channel with N_0 =10^{-6} Watt/Hz. The data rate is 56 kbits/s. What is the minimum amount of power needed at the transmitter in order to maintain a bit-error probability of 10^{-3}.

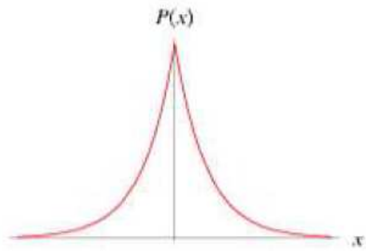
V. Consider a binary communication system transmitting signals $s_1(t)$ and $s_2(t)$ that are equally likely. At the receiver, the output of the demodulate and sample step yields $x(T) = u_i + n_0$, where the signal component u_i is either u_1 =+1 or u_2 =-1. The noise component n_0 is uniformly distributed yielding the conditional probability density functions $p(x/s_1)$ and $p(x/s_2)$ given by

$$p(x/s_1)=\begin{cases}0.5 & for \quad -0.3\le x\le 1.7\\ 0 & otherwise\end{cases} \quad \text{and} \quad p(x/s_2)=\begin{cases}0.5 & for \quad -1.7\le x\le 0.3\\ 0 & otherwise\end{cases}$$

a) Calculate the bit error probability when using an optimum decision threshold γ=0.

b) How the bit error probability changes when γ varies between 0 and 0.4

VI. Consider a binary communication system transmitting signals $s_1(t)$ and $s_2(t)$. At the receiver, the output of the demodulate and sample step yields $x(T) = u_i + n_0$, where the signal component u_i is either u_1 =+1 or u_2 =-1. Assume that the noise component n_0 has a Laplacian distribution where the PDF is $p(n_0)=\frac{1}{\sqrt{2}\sigma}e^{-\frac{\sqrt{2}|n_0|}{\sigma}}$, σ is the root mean square value of the noise. Calculate the bit error probability as a function of σ for the case of equally likely signaling $s_1(t)$ and $s_2(t)$ and using an optimum decision threshold γ=0.



VII. Consider a binary communication system transmitting pulses $s_1(t)$ and $s_2(t)$ using a baseband modulation with a bit rate of 10 Kbits/s. The pulses have amplitude of 200 mVolts and are transmitted along a cable that attenuates the signal power from transmitter to receiver by 6 dB. We assume a normalization relative to 1Ω resistive load. The signals are equally likely and are coherently detected over a Gaussian channel with N_0 =2.10^{-7} Watt/Hz. Which modulation. NRZ or bipolar RZ, should be used to maintain a bit error probability no worse that 10^{-3}? Justify your answer.

VIII. Consider a binary communication system transmitting through an additive white Gaussian noise channel the signals $s_1(t)$ and $s_2(t)$ that have prior probability $p(s_1) = p(s_2)= 0.5$. At the receiver, the output of the demodulate and sample step yields $x(T) = u_i + n_0$, where the signal component u_i is either u_1 =+1 V or u_2 =-1 V. The noise component n_0 is zero-mean with unit variance.

a) Calculate the bit error probability when using a decision threshold equal to 0.1 .

b) If $p(s_2)=1-p(s_1)=0.7$, what is the optimum decision threshold and the bit error probability in this case. Use Leibniz's rule:

$$\frac{d}{dy}\int_{g_1(y)}^{g_2(y)} f(x,y)dx = \int_{g_1(y)}^{g_2(y)} \frac{\partial f}{\partial y}(x,y)dx + g_2'(y)f(g_2(y),y) - g_1'(y)f(g_1(y),y)$$

$Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-\frac{v^2}{2}}dv$, the PDF of the Gaussian noise n is $p(n)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{n^2}{2\sigma^2}}$

Bandpass modulations

A stated in the chapter 2, the baseband signals (NRZ, Manchester, etc.) have applications in local area networks, data storage devices etc. The signal spectra of these schemes contain low frequency components that are attenuated quickly with distance in most transmission media. For long transmissions over wire and wireless media, we usually need to shift the waveform spectrum to fit within the channel bandwidth. In this context, a bandpass modulation takes incoming bits and outputs a waveform centered around frequency f_c. It is usually used for long distance and wireless transmissions and also called carrier modulation i.e. modulation of the waveform onto an RF carrier (recall that modulation is the process of mapping bits to analog waveforms). To transmit the digital bit stream, the bit stream is converted to the sinusoidal signal; for digital transmissions such a sinusoid of duration T is referred to as a digital symbol. It has tree features that can be used to distinguish it from other sinusoids: amplitude, frequency and phase. Thus bandpass modulation can be defined as the process whereby the amplitude, frequency, or phase of a carrier, or a combination of them, is varied in accordance with the information to be transmitted.

The general form of the wave carrier $s(t) = A\cos(\omega t + \phi)$ where ω [1] is a frequency at ω_c (or near). The modulator stores the information bits in either the amplitude *A*, the frequency (ω), or the phase (ϕ).We can change any of these three characteristics to formulate the modulation scheme. The basic forms of the three modulation methods used for transmitting digital signals are:

Amplitude shift keying (ASK), Frequency shift keying (FSK), Phase shift keying (PSK). Based on these basic schemes, a variety of modulation techniques can be derived from their combination including QPSK, QAM, etc.

Note that in the concepts of orthogonality can be used to advantage in the design and selection of single and multiple carriers for modulation, transmission and reception.

1 AMPLITUDE SHIFT KEYING (ASK)

The first bandpass modulation we'll look at is called amplitude shift-keying ASK modulation (also called OOK (On-Off keying)). This refers to the modulators that, given the input bits, create the waveform $s(t) = A\cos(\omega t + \phi)$, where the input bits are stored in the amplitude (*A*).

Consider the Binary ASK (BASK), which is the simplest of the ASK modulators (*M*=2). Figure 1 shows how the bits 010 are stored in the amplitude. When a bit 0 is input, the modulator gets out $s_0(t) = -A\cos(\omega_c t)$ for the symbol duration. Whenever a bit 1 arrives, the modulator throws out $s_1(t) = +A\cos(\omega_c t)$ for the symbol duration.

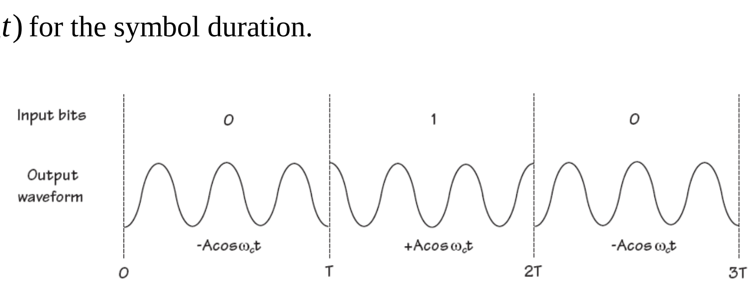


Figure 1: BASK modulation

The table summarizes the work of a BASK modulator. The times *iT* to *(i+1)T* refer to the time duration of the incoming symbol that is equal to the duration of the bit (T_b).

Input bits	Output waveform
0	$s_0(t) = -A\cos\omega_c t,\ iT \le t < (i+1)T$
1	$s_1(t) = +A\cos\omega_c t,\ iT \le t < (i+1)T$

[1] When frequency *f* is used, Hertz is intended. When ω is used, radians/s is intended. Recall that $\omega = 2\pi f$

When M=4, 4-ASK modulator maps each set of two bits to a waveform with a different amplitude. Figure 2 shows the input bits are 1011. First bit pair 10 comes into the modulator, which gets out the output waveform with amplitude A. Then bit pair 11 jumps into the modulator, and the modulator responds by throwing out the output waveform with amplitude $3A$.

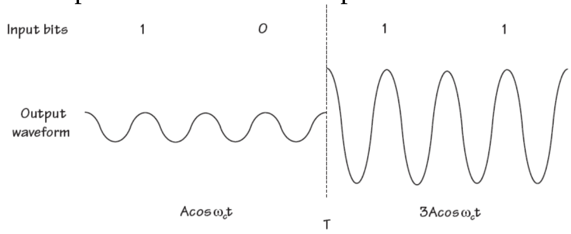


Figure 2: 4-ASK

The table provides a summary of 4-ASK work. Note that in this case each symbol is two bits, $T=2T_b$.

We can extend the ideas of B-ASK and 4-ASK to *M-ary ASK* where $b=\log_2 M$ bits are input to the modulator at the same time, and the modulator outputs one of M possible waveforms.

Input bits	Output waveform
00	$s_0(t) = -3A\cos\omega_c t,\ iT \le t < (i+1)T$
01	$s_1(t) = -A\cos\omega_c t,\ iT \le t < (i+1)T$
10	$s_2(t) = A\cos\omega_c t,\ iT \le t < (i+1)T$
11	$s_3(t) = 3A\cos\omega_c t,\ iT \le t < (i+1)T$

The value of A, the peak value of a sinusoidal waveform, equals $\sqrt{2}$ times the root-mean-square (rms) value, we can write:

$$s(t) = A\cos(\omega t) = \sqrt{2A_{rms}^2}\cos(\omega t)$$

If we assume that the signal is a voltage or current waveform, A_{rms}^2 represents average power P normalized to 1 Ω and replacing P [watts] by $\frac{E}{T}$ [joules/seconds] we get: $s(t) = \sqrt{2P}\cos(\omega t) = \sqrt{\frac{2E}{T}}\cos(\omega t)$, where E is the energy of the signal over the symbol duration T.

Therefore, $A=\sqrt{\frac{2E}{T}}$.

2 FREQUENCY SHIFT KEYING (FSK)

In FSK modulation, the information bits are stored in the frequency. The simplest form is the BFSK (binary FSK) where two signals with different frequencies are used to represent binary 1 and 0. Figure 3 shows how the bits 010 are stored in the frequency.

When a bit 1 is input, the modulator gets out $s_1(t) = A\cos(\omega_1 t + \phi_1)$ for the symbol duration. Whenever a bit 0 arrives, the output of the modulator is $s_2(t) = A\cos(\omega_2 t + \phi_2)$ for the symbol duration. $\omega_i = \omega_c + \Delta\omega_i$, i=1,2.

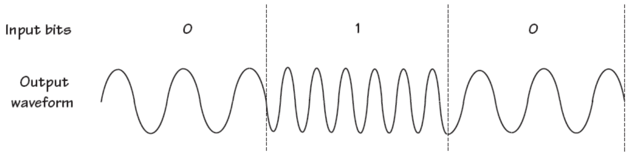


Figure 3: BFSK

Note that ϕ_1 and ϕ_2 are not the same in general, the two signal are not coherent. The waveform is not continuous at bit transitions. This form of FSK is therefore called noncoherent or discontinuous FSK.

It can be generated by switching the modulator output line between two different oscillators as shown in Figure 4 . It can be noncoherently demodulated.

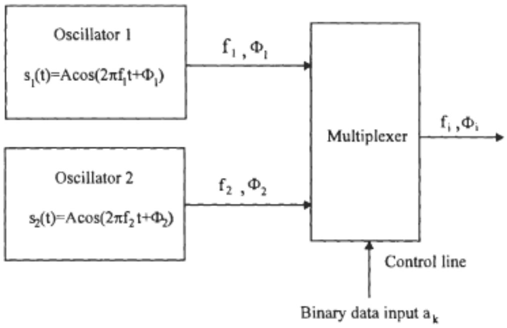


Figure 4: noncoherent FSK modulator

The two signal are coherent if they have the same initial phase ($\phi_1 = \phi_2$, assuming =0) at t = 0. It can be coherently or noncoherently demodulated. This type of FSK can be generated by the modulator as shown in Figure 5. The frequency synthesizer generates two frequencies, f_1 and f_2, which are synchronized. The binary input data controls the multiplexer. The bit timing must be synchronized with the carrier frequencies. Briefly speaking, if a 1 is present, $s_l(t)$ will pass and if a 0 is present, $s2(t)$ will pass. Note that $s_l(t)$ and $s2(t)$ are always there regardless of the input data. So when considering their phase in any bit interval $kT \leq t \leq (k+1)T$, the starting point of time is 0 not kT.

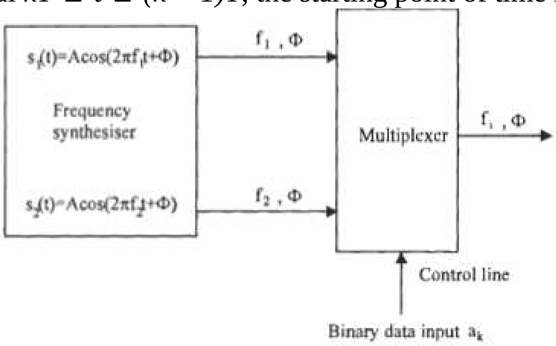


Figure 5: coherent FSK modulator

For coherent demodulation of the coherent FSK signal, the two frequencies are so chosen that the two signals are orthogonal:

$$\int_{kT}^{(k+1)T} s_1(t)s_2(t)dt = 0$$

This requires that $2\pi(f_1 + f_2)T = 2n\pi$ and $2\pi(f_1 - f_2)T = m\pi$ where n and m are integers. This leads to $f_1 = \dfrac{2n+m}{4T}$ and $f_2 = \dfrac{2n-m}{4T}$.

Let $2\Delta f = f_1 - f_2 = \dfrac{m}{2T}$, we may conclude that for orthogonality f_1 and f_2 must be integer multiple of 1/4T and their difference must be integer multiple of 1/2T. Using Δf we can rewrite the two frequencies as

$$f_1 = f_c + \Delta f \ ; \ f_1 = f_c - \Delta f \ ; \ f_c = \frac{f_1 + f_2}{2} = \frac{n}{2T}$$

where f_c the carrier frequency must be integer multiple of 1/2T for orthogonality. When the separation between f_l and f_2 is chosen as 1/T, then the phase continuity will be maintained at bit transitions, the FSK is called Sunde's FSK (example in Figure 3). As a matter of fact, if the separation is k/T, where k is an integer, the phase of the coherent FSK signal is always continuous. The minimum separation for orthogonality between f_l and f_2 is 1/2T. But this separation cannot guarantee continuous phase (because 1/2T < 1/T). (see section 5 for continuous phase)

In an M-ary FSK modulation, there are M signals with different frequencies to represent M symbols of b= $\log_2 M$ bits. The expression of the i^{th} signal is $s_i(t) = A\cos(\omega_i t + \phi_i)$ where T is the symbol period equal to bT_b. If the initial phases are the same for all i, the signal set is coherent. As in the binary case we can always assume $\phi_i = 0$ for coherent MFSK. The demodulation could be coherent or noncoherent. Otherwise the signal set is noncoherent and the demodulation must be noncoherent. The minimum separation between two adjacent frequencies is the same as that of the binary case. The noncoherent modulator for binary FSK in Figure 4 can be easily extended to noncoherent MFSK by simply increasing the number of independent oscillators. The coherent modulator for binary FSK in Figure 5 can also be extended to MFSK. Then the frequency synthesizer generates M signals with the designed frequencies and coherent phase, and the multiplexer chooses one of the frequencies, according to the *b* data bits.

3 PHASE SHIFT KEYING (PSK)

Phase shift keying (PSK) is a large class of digital modulation schemes. PSK is widely used in the communication industry. With these modulations, input bits are mapped into output waveforms of the form $s(t) = A\cos(\omega t + \phi)$, and the information bits are stuffed in the phase ϕ.

3.1 BINARY PSK (BPSK)

In BPSK, binary data are represented by two signals with different phases. Typically these two phases are 0 and π. When bit 0 goes into the modulator, the modulator gets out the waveform $s_1(t) = A\cos(\omega_c t) = \sqrt{\frac{2E}{T}}\cos(\omega_c t)$. If bit 1 struts into the modulator, it gets out $s_2(t) = A\cos(\omega_c t + 180) = \sqrt{\frac{2E}{T}}\cos(\omega_c t + 180)$. Figure 6 shows you what happens when bits 010 enter a BPSK modulator.

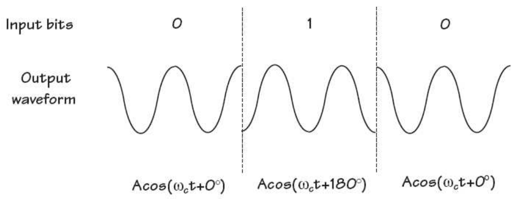


Figure 6: BPSK modulation

PSK signals can be graphically represented by a signal constellation in a two-dimensional coordinate system with $\phi_1(t) = \sqrt{\frac{2}{T}}\cos(\omega_c t)$ and $\phi_2(t) = -\sqrt{\frac{2}{T}}\sin(\omega_c t)$ as its horizontal and vertical axis, respectively (The two axes sometimes are simply labeled as I-axis (inphase) and Q-axis (Quadrature)). The signal constellation of BPSK is shown in Figure 7 where $s_1(t)$ and $s_2(t)$ are represented by two points on the horizontal axis, respectively, where

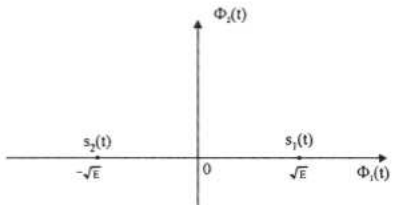


Figure 7: BPSK signal constellation

In general the phase is not continuous at bit boundaries. The modulator which generates the BPSK signal is quite simple (see Figure 8). First a bipolar data stream *a(t)* is formed from the binary data stream $a(t) = \sum_{k=-\infty}^{\infty} a_k p(t-kT)$, where $a_k \in \{1,-1\}$, p(t) is the rectangular pulse with unit amplitude defined on [0, T]. Then *a(t)* is multiplied with a sinusoidal carrier $A\cos(\omega_c t)$. The result is the BPSK signal $s(t) = Aa(t)\cos(\omega_c t)$.

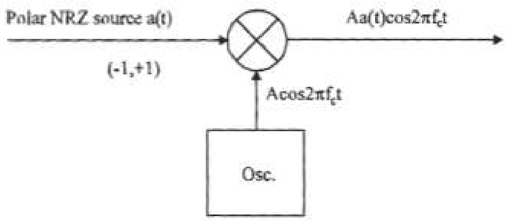


Figure 8: BPSK modulator

The demodulator of BPSK is a coherent detector because BPSK requires that the reference signal at the receiver to be synchronized in phase and frequency with the received signal. Figure 9 shows the coherent receiver. The reference signal must be synchronous to the received signal in frequency and phase. It is generated by the carrier recovery (CR) circuit. Noncoherent detection of BPSK is realized in the form of differential BPSK which will be studied in the next section.

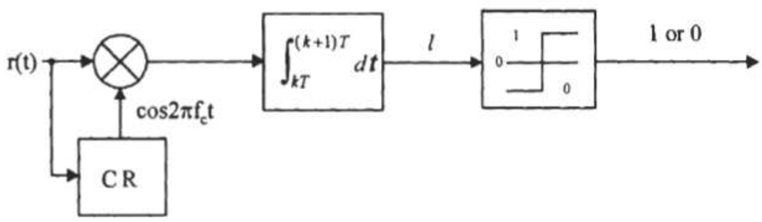


Figure 9: Coherent BPSK demodulator

3.2 DIFFERENTIAL BPSK

DBPSK uses the technique of differential encoding and decoding of binary data in PSK modulation. In this technique the phase reference for the present signaling interval is provided by a delayed version of the signal that occurred during the previous signaling interval. The DPSK receiver does not require a carrier synchronizer circuit. In practice, DBPSK is often used instead of BPSK. Figure 10 shows a DBPSK modulator and an example when an arbitrary reference bit 1 is selected. The differential encoding rule is $d_k = \overline{a_k \oplus d_{k-1}}$.

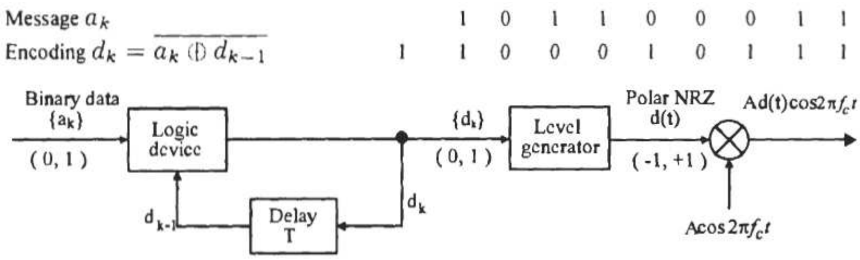


Figure 10: DBPSK modulator

Inversely we can recover a_k from d_k using $a_k = \overline{d_k \oplus d_{k-1}}$. If d_k and d_{k-1} are the same, then they represent a 1 of a_k. if d_k and d_{k-1} are different they represent a 0 of a_k. The differential demodulator (see Figure 11) uses the previous symbol as the reference for demodulating the current symbol. The bandpass filter reduces noise power but preserves the phase of the signal. The integrator output is positive if the current signal is the same as the previous one, otherwise the output is negative. This is to say that the demodulator makes decisions based on the difference between the two signals. Thus information data must be encoded as the difference between adjacent signals, which is exactly what the differential encoding accomplished.

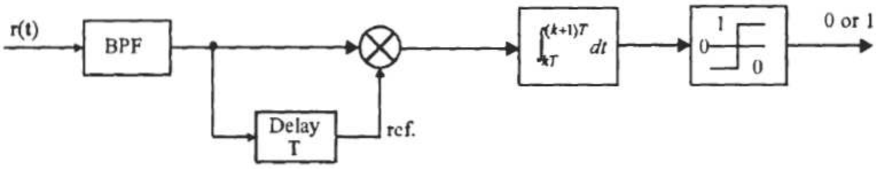


Figure 11: DBPSK demodulator

3.3 M-ARY PSK

The motivation behind M-ary PSK or MPSK is to increase the bandwidth efficiency of the PSK modulation schemes. In MPSK, $b = \log_2 M$ bits are represented by a symbol. M-ary PSK signal set is defined as: $s_i(t) = A\cos(\omega_c t + \varphi_i)$, $0 \le t \le T$, $i=1,\ldots,M$. where $\varphi_i = \frac{2(i-1)\pi}{M} + \theta$, θ is an arbitrary phase.

The signal expression can be written as:

$s_i(t) = A\cos\varphi_i \cos(\omega_c t) - A\sin\varphi_i \sin(\omega_c t) = s_{i1}\phi_1(t) + s_{i2}\phi_2(t)$, where $\phi_1(t)$ and $\phi_2(t)$ given above are orthonormal basis functions. $s_{i1} = \sqrt{E}\cos\varphi_i$ and $s_{i2} = \sqrt{E}\sin\varphi_i$. We can calculate $\varphi_i = Arctg\frac{s_{i2}}{s_{i1}}$.

The MPSK signal constellation is therefore two-dimensional. Each signal $s_i(t)$ is represented by a point S_i (s_{i1},s_{i2}) in the basis $\phi_1(t)$ and $\phi_2(t)$. The polar coordinates of the signal are $(\sqrt{E},\varphi_i)$.The signal points are equally spaced on a circle of radius $\sqrt{E}$ and centered at the origin. For M=4, the modulation is called quadrature phase-shift keying (QPSK) as shown in Figure 12.

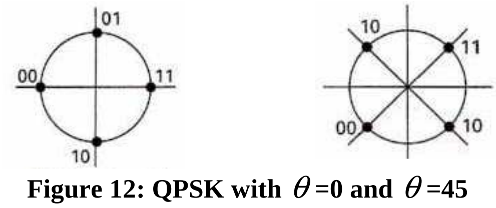


Figure 12: QPSK with θ=0 and θ=45

The bits-signal mapping could be arbitrary provided that the mapping is one-to-one. However, a method called Gray coding is usually used in signal assignment in MPSK. Gray coding assigns *b* bits with only one-bit difference to two adjacent signals in the constellation. When an M-ary symbol error occurs, it is more likely that the signal is detected as the adjacent signal on the constellation, thus only one of the b input bits is in error. Figure 13 shows the constellation of 8-PSK, where Gray coding is used for bit assignment.

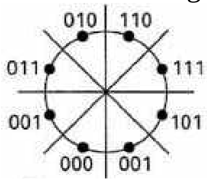


Figure 13: Constellation of 8-PSK with Gray coding

Other modulations schemes derived from QPSK are Offset QPSK and $\pi/4$ QPSK. Offset QPSK is essentially the same as QPSK except that the I- and Q-channel pulse trains are staggered. Only an extra delay of half symbol period (T/2 seconds) is added in the Q-channel (see the OQPSK in Figure 14). In comparison to QPSK, OQPSK signals are less susceptible to spectral side lobe restoration in satellite transmitters.

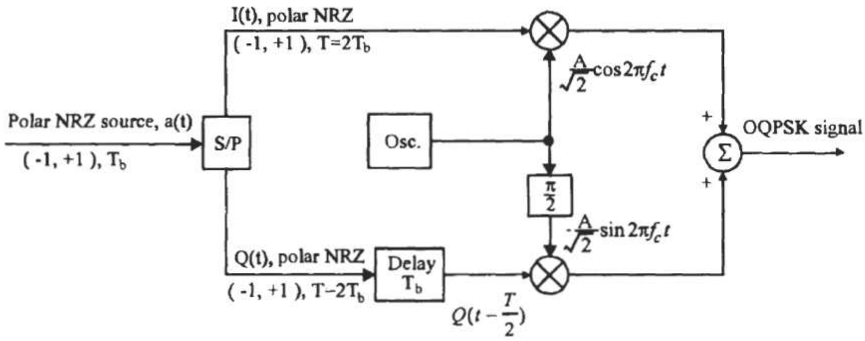


Figure 14: OQPSK modulator

Although OQPSK can reduce spectral restoration caused by nonlinearity in the power amplifier, it cannot be differentially encoded and decoded. $\pi/4$-QPSK is a scheme which not only has no 180 phase shifts like OQPSK, but also can be differentially demodulated. These properties make it particularly suitable to mobile communications where differential demodulation can reduce the adversary effect of the fading channel. $\pi/4$- QPSK has been adopted as the standard for the digital cellular telephone system in the United States and Japan.

Since MPSK signals are two-dimensional, for $M \geq 4$, the modulator can be implemented by a quadrature modulator. The MPSK modulator is shown in Figure 15. The only difference for different values of M is the level generator.

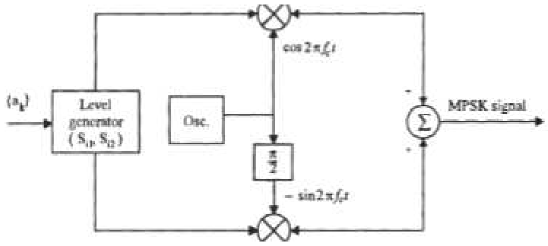


Figure 15: MPSK modulator

4 QAM

Quadrature amplitude modulation (QAM) is a composite modulation technique that combines both phase modulation and amplitude modulation. QAM is a scheme that can achieve higher bandwidth efficiency than MPSK with the same average signal power. QAM is widely used in modems designed for telephone channels in in satellite, wireless and mobile systems. In QAM, the information bits are stored into both the phase φ and the amplitude A of the waveform. M-QAM signal is defined as: $s_i(t) = A_i \cos(\omega_c t + \varphi_i)$, $0 \le t \le T$, $i=1,\ldots,M$. where A_i is the amplitude and φ_i is the phase of the ith signal in the M-ary signal set.

Pulse shaping is usually used to improve the spectrum and for ISI control purpose in QAM. With *p(t)* pulse shaping, the QAM signal is $s_i(t) = A_i p(t)\cos(\omega_c t + \varphi_i)$.

It can be written as: $s_i(t) = A_i p(t)\cos\varphi_i \cos(\omega_c t) - A_i p(t)\sin\varphi_i \sin(\omega_c t) = s_{i1}\phi_1(t) + s_{i2}\phi_2(t)$, where $\phi_1(t) = \sqrt{\frac{2}{E_p}}\, p(t)\cos(\omega_c t)$ and $\phi_2(t) = -\sqrt{\frac{2}{E_p}}\, p(t)\sin(\omega_c t)$ are the orthonormal basis functions. E_p is the energy of p(t) in [0,T]. $E_p = \int_0^T p^2(t)dt$. If $p(t)=1$, $E_p = T$ and we find the same formulas of $\phi_1(t)$ and $\phi_2(t)$ as given in section 3.1. $s_{i1} = \sqrt{\frac{E_p}{2}} A_i \cos\varphi_i$ and $s_{i2} = \sqrt{\frac{E_p}{2}} A_i \sin\varphi_i$. Similar to MPSK, a geometric representation called constellation is a very clear way of describing a QAM signal set. The horizontal axis of the constellation plane is $\phi_1(t)$ and the vertical axis is $\phi_2(t)$. A QAM signal is represented by a point (vector, or phasor) with coordinates S_i (s_{i1},s_{i2}). Alternatively, the two axes can be simply chosen as $p(t)\cos(\omega_c t)$ and $-p(t)\sin(\omega_c t)$. Then the signal coordinates are ($A_i \cos\varphi_i$, $A_i \sin\varphi_i$).

The distance d_{ij} between constellation points can be expressed as $d_{ij} = \sqrt{(s_{i1} - s_{j1})^2 + (s_{i2} - s_{j2})^2}$.

Depending what values ((s_{i1},s_{i2}) and A_i and φ_i are assigned with, a variety of QAM constellations can be realized.

In the constellation, a fixed number of signal points (or phasors) are equally spaced on each of the N circles, where N is the number of amplitude levels. This is called type I constellation in the literature Figure 16. In a type I constellation, the points on the inner ring are closest together in distance and are most vulnerable to errors . To overcome this problem, type II constellation was proposed. In a type II constellation, signal points are still on circles, but the number of points on the inner circle is less than the number of points on the outer circle, making the distance between two adjacent points on the inner circle approximately equal to that on the outer circle. Type III of QAM constellation is shown in Figure 16 , it has been the most widely used system because its implementation is considerably simpler than that of type I and II.

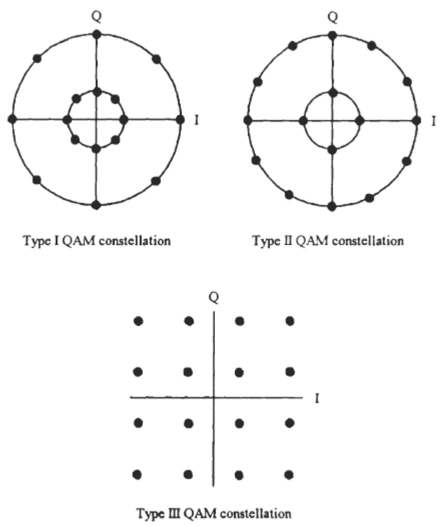


Figure 16: QAM constellations

When designing a constellation, some points must be considered:

- The minimum Euclidean distance among the signal points should be as large as possible under other constraints, since it determines the symbol error probability of the modulation scheme.
- The phase differences among the points should be as large as possible under other constraints.
- The average power of the points should be as small as possible under other constraints.
- The implementation complexity, etc.

M- QAM defines a constellation of M points, each with a particular phase and amplitude, and each representing a data symbol. Figure 17 shows several cases. Example 4-QAM: {1 + j,−1 + j, 1 − j,−1 − j}, etc.

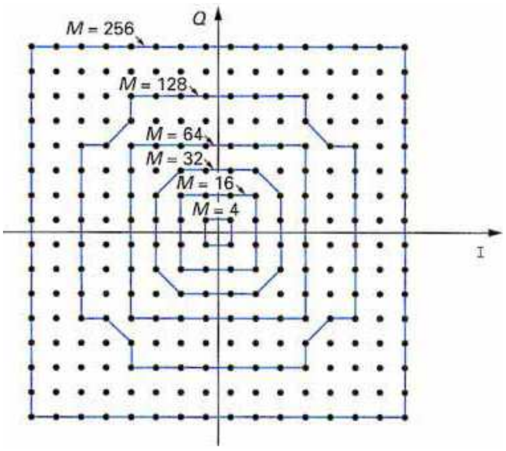
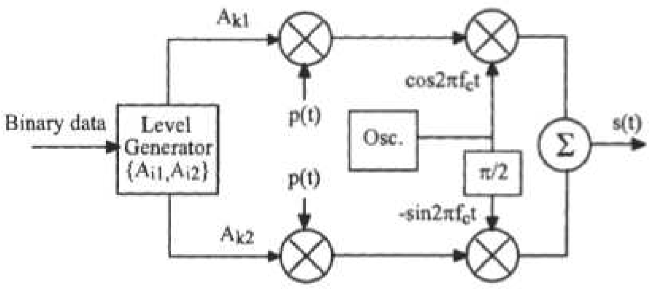


Figure 17: different QAM constellations

The QAM modulator is almost identical to that of MPSK since both of them are quadrature schemes (see Figure 18 (a)). If pulse shaping is not desired, the *p(t)* block will be absent. The mapping from bits to QAM points are usually Gray coded for minimizing bit errors. For square QAM, perfect Gray coding is possible. Figure 18 (b) is a Gray coded square 16-QAM constellation.



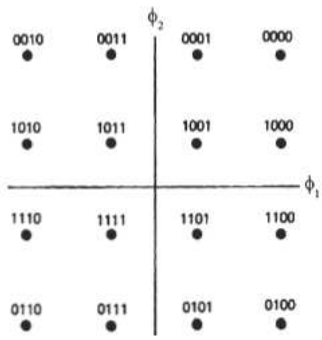


Figure 18: (a) QAM modulator **(b) Gray coded 16-QAM constellation**

5 CONTINUOUS PHASE MODULATION (CPM)

The motivation in designing continuous phase modulation (CPM) schemes is to avoid discontinuous phase transitions as this may happen in BPSK or QPSK etc. With proper choice of pulse shapes and other parameters, CPM schemes may achieve higher bandwidth efficiency than QPSK and higher order MPSK schemes. Minimum shift keying (MSK) is one such scheme that can be derived from OQPSK. MSK can also be viewed as a special case of FSK modulation that has continuous phase. GMSK (Gaussian MSK) is also an important modulation scheme and has been used in the European global system for mobile (GSM) system.

5.1 MSK

To create an MSK signal, we use the following signal: $s(t) = A\cos(2\pi f_c t + d_k \frac{\pi t}{2T} + \varphi_k)$, $kT \le t \le (k+1)T$. Where $d_k = \overline{a_k \oplus d_{k-1}}$, a_k is the original data stream. We can see that MSK signal is a special FSK signal with two frequencies $f_c + \frac{1}{4T}$ and $f_c - \frac{1}{4T}$. The frequency separation is equal to 1/2T. This is the minimum separation for two FSK signals to be orthogonal, hence the name minimum shift keying. MSK carrier phase is always continuous at bit transitions. To see this, we check the excess phase of the MSK signal, referenced to the carrier phase, which is given by

$\theta(t) = d_k \frac{\pi t}{2T} + \varphi_k = \pm \frac{\pi t}{2T} + \varphi_k$, $kT \le t \le (k+1)T$.

We can see that the excess phase $\theta(kT)$ is a multiple of $\pi/2$. It increases or decreases linearly with time during each period of T seconds. If $d_k = 1$ in the period, the carrier phase is increased by $\pi/2$ by the end of the period. This corresponds to an FSK signal at the higher frequency $f_c + \frac{1}{4T}$. If $d_k = -1$ in the period, the carrier phase is decreased by $\pi/2$ by the end of the period. This corresponds to an FSK signal at the lower frequency $f_c - \frac{1}{4T}$.

Because d_k is constant in the interval [(k-1)T, kT], $\theta(t)$ is linear and continuous in the interval [(k-1)T, kT], i.e. to maintain continuous phase at bit transition t = kT, we must require the following conditions

$d_{k-1}\frac{\pi t}{2T} + \varphi_{k-1} = d_k \frac{\pi t}{2T} + \varphi_k \mod(2\pi) \Rightarrow \varphi_k = \varphi_{k-1} + (d_{k-1} - d_k)\frac{k\pi}{2} \mod(2\pi)$. This means that $\varphi_k = \varphi_{k-1} \mod(2\pi)$ if $d_{k-1} = d_k$, and $\varphi_k = \varphi_{k-1} + k\pi \mod(2\pi)$ if $d_{k-1} \ne d_k$.

Figure 19 shows the MSK modulator implemented as a differential encoded FSK and the two frequencies are $f_c + \frac{1}{4T}$ and $f_c - \frac{1}{4T}$.

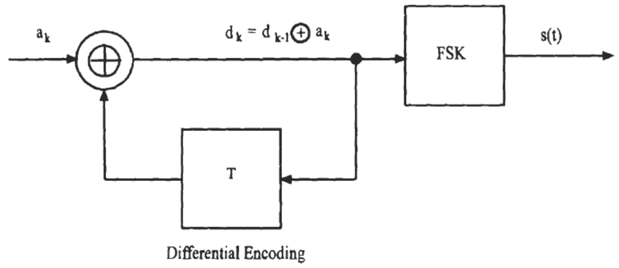


Figure 19: MSK modulator based on FSK

Another implementation of MSK modulation could be done using a sinusoidal weighted OQPSK modulator. This is shown in Figure 20. The data stream signal *a(t)* is demultiplexed into I and Q by the serial-to-parallel converter. The in-phase channel signal I(t) consists of even-numbered bits, and the quadrature channel signal Q(t) consists of odd numbered bits. Each bit in I(t) and Q(t) has a duration of 2T. Q(t) is delayed by T with respect to I(t). I(t) is multiplied by $A\cos(\frac{\pi t}{2T})$ and $A\cos(2\pi f_c t)$ in the two subsequent multipliers in the I-channel. Q(t) is multiplied by $A\sin(\frac{\pi t}{2T})$ and $A\sin(2\pi f_c t)$ in the two subsequent multipliers in the Q-channel. $A\sin(\frac{\pi t}{2T})$ and $A\sin(2\pi f_c t)$ are obtained through $\pi/2$ phase shifters from $A\cos(\frac{\pi t}{2T})$ and $A\cos(2\pi f_c t)$ respectively.

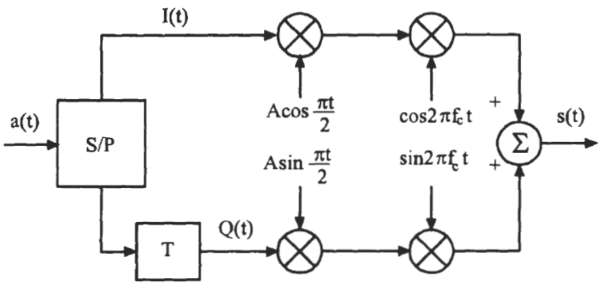


Figure 20: MSK modulator

5.2 GMSK

MSK is a spectral compact scheme where the phase changes linearly. This helps to control adjacent-channel interference and to provide narrow bandwidth. To further improve the spectrum of MSK, the data signal could be filtered before modulation. A wide range of filter responses is possible. The Gaussian filter, which leads to Gaussian MSK (GMSK), is the most popular. Briefly speaking, GMSK is another spectral compact MSK-type scheme which passes the data waveform through a Gaussian filter before sending it to a modulator. It is used in GSM systems. The values of the time-bandwidth product for the Gaussian filter (BT_b) vary between 0.3 and 0.5. B is the bandwidth of the Gaussian filter and T_b is the bit period time.

6 DEMODULATION AND DETECTION OF SIGNALS IN AGWN CHANNEL

Figure 21 shows a typical digital receiver for bandpass signal where demodulation and detection functions are represented in two steps: demodulate and sample block and detection block. The frequency down-conversion block performs frequency translation for bandpass signal operating at some radio frequency. Within the demodulate and sample block, the receiving filter performs waveform recovery in preparation for the next step i.e. the detection. The goal of the receiving signal is to recover a baseband pulse with the best possible SNR free of any ISI. The optimum receiving filter to accomplish this is called matched filter or correlator filter. An equalizing filter follows the receiving filter. It is used only for the systems where channel induced ISI can distort the signals. In most case, when equalizer is used, a single filter would be designed to incorporate both functions, best possible SNR and compensation of the distortion caused by the transmitter and the channel. Such composite filter is simply called the receiving and equalizing filter or simply equalizing filter.

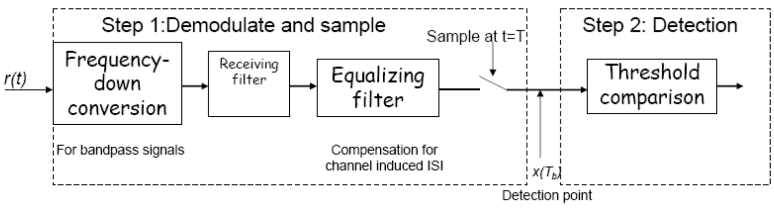


Figure 21: Demodulation and detection

6.1 BIT ERROR PROBABILITY OF BPSK

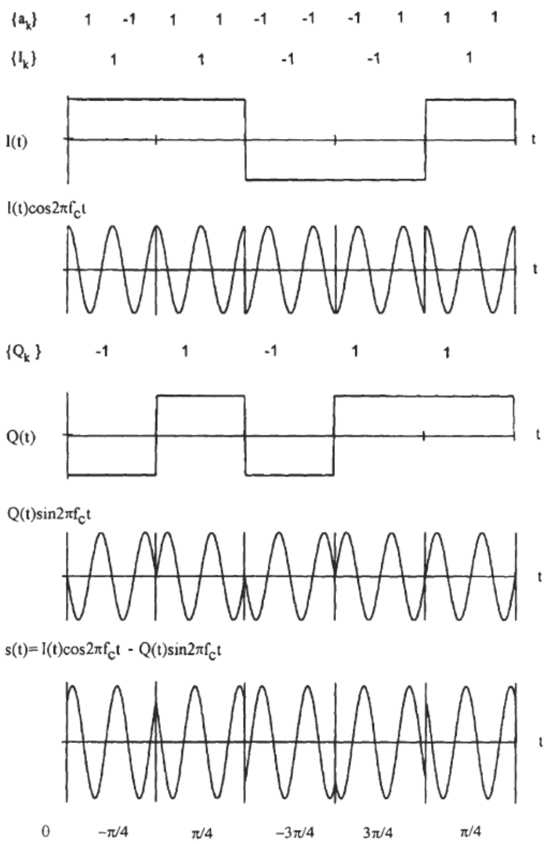
The bit error probability of the bandpass antipodal BPSK signaling can be given similar to the binary case of antipodal signaling (given in chapter 3) in general:

$$P_b = Q(\sqrt{\frac{2E_b}{N_0}})$$

6.2 BIT ERROR PROBABILITY OF QPSK

In this case we have two binary phase-modulation signals in phase quadrature. Since there is no interference between the signals on the two quadrature carries, the bit error probability is identical to $P_b = Q(\sqrt{\frac{2E_b}{N_0}})$. The symbol error probability is determined by $P_s = 1-(1-Q(\sqrt{\frac{2E_b}{N_0}}))^2$.

The channel with cosine reference is called in phase (I) channel and the channel with sine reference is called quadrature (Q) channel. The data sequence is separated by the serial-to-parallel converter (S/P) to form the odd-numbered-bit sequence for I-channel and the even-numbered-bit sequence for Q-channel. Then logic 1 is converted to a positive pulse and logic 0 is converted to a negative pulse, both have the same amplitude and a duration of T. Next the odd-numbered-bit pulse train is multiplied to cos() and the even-numbered-bit pulse train is multiplied to sin () . It is clear that the I-channel and Q-channel signals are BPSK signals with a symbol duration of 2Tb. Finally a summer adds these two waveforms together to produce the final QPSK signal. I- and Q-channel signals are demodulated separately as two individual BPSK signals



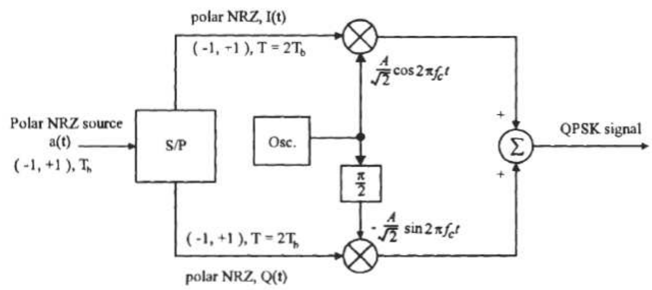


Figure 22:QPSK modulator

7 PROBABILITY OF ERROR FOR MPSK

For large energy-to-noise ratios, the symbol error performance $P_s(M)$ M>4, for equally coherently detected M-ary PSK signaling can be approximated by

$$P_s(M) \approx 2Q(\sqrt{\frac{2E_s}{N_0}}\sin\frac{\pi}{M})$$

The energy per symbol E_s is related to the energy per bit E_b as follows: $E_s = E_b(\log_2 M)$.

For Gray coded MPSK signals, since the erroneous symbols are the adjacent signals which only differ by one bit, the bit error rate can be related to the symbol error rate by $P_b(M) \approx \frac{P_s(M)}{\log_2 M}$.

8 PROBABILITY OF ERROR FOR M-ARY PAM

We consider the case of M-ary PAM signal set (M levels). M-ary PAM signals are represented geometrically as M one-dimensional signal points. For equal amplitude spacing, the amplitudes may be expressed as $s_i = (2i - 1 - M)V_0 \quad , i = 1,2,...,M$. Where V_0 is the smallest amplitude. Assuming s_i is transmitted, a symbol error occurs when the noise exceeds in magnitude one-half of the distance between two adjacent levels. This probability is the same for each s_i except for the two outside levels, where an error can occur in one direction only. Assuming all amplitude levels are equally likely, the average symbol error probability is simply the probability that the noise variable exceeds in magnitude one-half of the distance between levels. When either one of the two outside levels is transmitted, a error can occur in one direction only. Thus we have

$$P_s(M) = \frac{M-1}{M}\Pr(|r - s_i| > \frac{d}{2})$$

where d is the distance between adjacent signal levels, and also is the distance between adjacent thresholds. Then $d = |s_i - s_{i-1}| = 2V_0$. Thus

$$P_s(M) = \frac{M-1}{M}\Pr(|r - s_i| > V_0) = \frac{M-1}{M}\frac{2}{\sqrt{\pi N_0}}\int_{V_0}^{\infty} e^{-\frac{v^2}{N_0}}dv = \frac{M-1}{M}\frac{2}{\sqrt{2\pi}}\int_{V_0\sqrt{2/N_0}}^{\infty} e^{-\frac{v^2}{2}}dv = \frac{2(M-1)}{M}Q(\sqrt{\frac{2V_0^2}{N_0}})$$

The symbol error probability can be expressed in terms of the average energy or power of the signals. The average energy of the signals is

$$E_{avg} = \frac{1}{M}\sum_{i=1}^{M} E_i = \frac{1}{M}\sum_{i=1}^{M} s_i^2 = \frac{1}{3}(M^2 - 1)V_0^2$$

As a result

$P_s(M) = \frac{2(M-1)}{M}Q(\sqrt{\frac{6E_{avg}}{(M^2-1)N_0}}) = \frac{2(M-1)}{M}Q(\sqrt{\frac{(6\log_2 M)E_{bavg}}{(M^2-1)N_0}})$ where E_{bavg} is the average energy per bit $E_{bavg} = \frac{E_{avg}}{\log_2 M}$.

Figure 23 shows the curves of the symbol error probability of the M-ary PAM versus E_{bavg} / N_0.

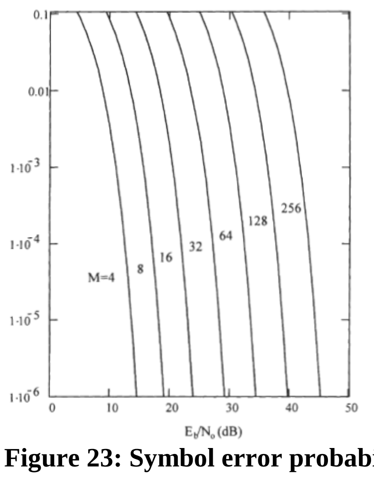


Figure 23: Symbol error probability of M-ary PAM

9 PROBABILITY OF ERROR FOR QAM

The general demodulator of the QAM signal the received signal is: $r(t) = s_i(t) + n(t)$. For QAM signal detection the sufficient statistic is the squared distance

$$l_i = (r_1 - s_{i1})^2 + (r_2 - s_{i2})^2$$

Where $r_1 = s_{i1} + n_1$ and $r_2 = s_{i2} + n_2$ are independent Gaussian random variables with mean values s_{i1} and s_{i2}, respectively. Their variance is $N_0/2$. The pair (r_1, r_2) determines a point in the QAM constellation plane, representing the received noisy signal. The detector compares the distances from (r_1, r_2) to all pairs of (s_{i1}, s_{i2}) and chooses the closest one.

For square QAM constellations with $M = 2^b$ where b is even, the QAM constellation is equivalent to two M-ary PAM signals on quadrature carriers, each having $\sqrt{M}$ signal points. A QAM symbol is detected correctly only when two M-ary PAM symbols are detected correctly. Thus the probability of correct detection of a QAM symbol is

$$P_c = (1 - P_s(\sqrt{M}))^2$$

where $P_s(\sqrt{M})$ is the symbol error probability of a $\sqrt{M}$-ary PAM with one-half the average power of the QAM signal. The symbol error probability of the square QAM is

$$P_s = 1 - (1 - P_s(\sqrt{M}))^2$$

At high SNR $P_s \approx 2P_s(\sqrt{M}) = \frac{4(\sqrt{M} - 1)}{\sqrt{M}} Q(\sqrt{\frac{3bE_{bavg}}{(M-1)N_0}})$

Note that this result is exact for square QAM with $M = 2^b$ where b is even. When b is odd there is no equivalent $\sqrt{M}$-ary PAM system. However, we can find a tight upper bound for any b,

$P_s \le Q(\sqrt{\frac{3bE_{bavg}}{(M-1)N_0}})$ where E_{bavg}/N_0, is the average SNR per bit.

To obtain bit error probability from the symbol error probability, we observe that square QAM can be perfectly Gray coded. That is, there is only one bit difference between adjacent symbols. Each symbol error most likely causes one bit error at large SNR. Thus $P_b \approx P_s / b$.

Figure 24 shows the bit error probability versus E_{bavg}/N_0 for M = 4, 8, 16, 32, 64, 128 and 256 where the curves for M = 8, 32 and 128 are tight upper bound.

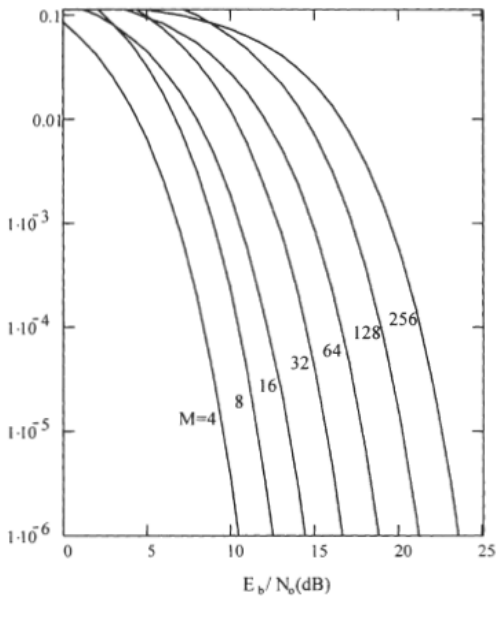


Figure 24: Bit error probability for square QAM

Chapter 4
Problems

I. Consider the following message (a_k) 100101101 to be transmitted using DBPSK. What is the output (d_k) of the differential encoding block when a reference bit 0 is used?

II. Consider a communication system using M-FSK modulation with a carrier frequency of 2446 MHz and a data rate of 4 Mbps.
   a) Calculate the transmitted frequencies when M=4 for coherent modulation and demodulation.
   b) What is the minimum required bandwidth and the bandwidth efficiency?
   c) Determine the bandwidth efficiency of the M-FSK signals in function of M.

III. Give the Gray coding for 64-QAM and plot the constellation diagram Type III.

IV. Consider a binary communication system using Sunde's FSK modulation with a data rate of 4 Mbps and a channel bandwidth of 5 MHz.
   a) If the carrier frequency is 2.4 GHz, calculate the transmitted frequencies and the bandwidth efficiency.
   b) Calculate the transmitted frequencies and the required bandwidth if we change the Sunde's FSK modulator and we use a MSK modulator with 0.3 Gaussian filter at its input.

V. An analog signal with maximum frequency f_m=3000 Hz will be transmitted using a modem that transmits the symbols at 4000 symbols/s.
   a) Using M-QAM system, what is the value of M if the signal is sampled at Nyquist rate and the quantization error should not exceed 4% of the peak-to-peak analog signal?
   b) Calculate the bandwidth efficiency if the channel bandwidth is about 4000Hz.
   c) If instead of using QAM system, we have to use a 16-PSK system and the signal is sampled at a sampling rate of 8000 samples/s. Calculate the max quantization error and the bandwidth efficiency in this case.

VI. Consider the MSK modulation for using in the GSM system with a channel bandwidth of 200 KHz and a data rate of 270 kbps.
   a) Calculate the frequency separation and the transmitted frequencies if the carrier frequency is 900 MHz.
   b) Calculate the bandwidth efficiency.
   c) What is the time-bandwidth product of the used filter?
   d) What is the required bandwidth channel if a Gaussian filter is used to generate 0.3 GMSK with the same data rate of 270 kpbs.

VII. What is the bit error probability at the reception in a communication system with bandwidth of 4 KHz and data rate of 10 Kbits/s if the received power to noise power is about 10 dB and the modulation is BPSK?

VIII. Consider a binary communication system transmitting equally likely signals $s_1(t)$ and $s_2(t)$ using BPSK modulation and a matched filter at the receiver . the matched filetr operates in AWGN with received $\frac{E_b}{N_0}$ = 6.8 dB. $s_1(t) = \sqrt{\frac{2E}{T}}\cos(\omega t)$ and $s_2(t) = \sqrt{\frac{2E}{T}}\cos(\omega t + \pi)$. We assume that E[(x(T)]=$\pm\sqrt{E}$.

a) Calculate the bit error probability when $\gamma = 0$.
b) If we change γ to $0.1\sqrt{E}$, what is the value of the bit error probability.

IX. Consider a digital system communication that transmits information through an additive white Gaussian noise channel via QAM Modulation at a rate of 2400 symbols/s. Determine the $\frac{E_b}{N_0}$ required to achieve a bit error probability of 10^{-5} at 9600 bits/s.

X. Consider a digital system communication that transmits information through an additive white Gaussian noise channel with noise-power spectral density $N_0 = 10^{-12}$ Watt/Hz. The information are sent at a rate of 2.4 Ksymbols/s using MPSK modulation.

a) Calculate the required energy per bit (E_b) to achieve a bit error probability of 1.35×10^{-3} at a bit rate of 4.8 Kbits/s.
b) What is the symbol error rate in this case?

XI. Consider a communication system supporting M-ary PSK modulation at 120 Ksymbols/s. The minimum required bit rate is 960 kbits/s.

a) What is the minimum $\frac{E_b}{N_0}$ required to maintain a reception with a bit error probability no worse that 10^{-3}?
b) Recalculate the new value for the minimum $\frac{E_b}{N_0}$ to maintain a reception with a bit error probability no worse that 10^{-3} when using M-ary QAM and comment this new result?

Introduction to Advanced Digital Communications

1 INTRODUCTION

Channel characteristic plays an important role in studying, choosing, and designing modulation schemes. Modulation schemes are studied for different channels in order to know their performance in these channels. Modulation schemes are chosen or designed according to channel characteristic in order to optimize their performance. A simple channel model for analyzing modulations schemes is the AWGN channel as seen in the previous chapters. In this model, the channel adds a white Gaussian noise to the signal passing through it and the received signal is r(t) = s(t) + n(t). The AWGN channel is the usual starting point for developing basic performances results.
When modeling practical systems, an accurate channel model describing the behavior of radio wave propagation is needed. Especially mobile radio channels are considered to be the most difficult channels, since they suffer from many imperfections like multi-path fading, Doppler shift, and interference. The multi-path channel model could be used in this context (see Figure 1).
In mobile radio channels, a signal from the transmitter may arrive at the receiver's antenna through several different paths. The transmitted electromagnetic wave may be reflected, diffracted, and scattered by surrounding buildings and by the terrain. As a result, the signal picked up by the receiver's antenna is a composite signal consisting of these multipath signals. Sometimes a line-of-sight (LOS) signal may exist. The multipath signals arrive at the receiver at slightly different delays and have different amplitudes. The different delays translate to different phases. This result is a composite signal which can vary widely and rapidly in amplitude and phase. This phenomena is called fading. Multipath also causes intersymbol interference for digital signals. For mobile radio channels, Doppler shift causes carrier frequency drift and signal bandwidth spread. All these adversaries cause degradation in performance of modulation schemes in comparison with that in AWGN channels.

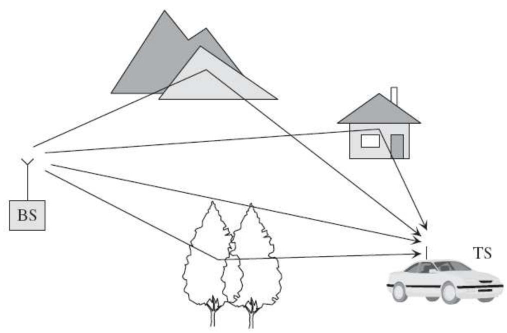


Figure 1: Multipath channel

2 FADING CHANEL CHRACTERITICS

Fading is the variation of the signal amplitude over time and frequency. The fading channels manifestations appear in 2 types of effects: Large-scale and small-scale fading. Large-scale fading occurs as the mobile moves through a large distance,for example, a distance of the order of cell size. It is caused by path loss of signal as a function of distance and shadowing by large objects such as buildings, intervening terrains, and vegetation. Shadowing is a slow fading process characterized by variation of median path loss between the transmitter and receiver. In other words, large-scale fading is characterized by average path loss and shadowing. On the other hand, small-scale fading refers to rapid variation of signal levels due to the constructive and destructive interference of multiple signal paths (multi-paths) when the mobile station moves short distances.

2.1 LARGE-SCALE FADING

Large-scale fading represents the path loss due to the propagation. It is proportional to $1/d^n$ where d is the distance between the transmitter and the receiver (n is the path loss exponent, varying from 2 (free-space) to 6 (urban environment) (see table below for details). If P_t is the transmitted power, The received power is given by

$$P_r(d) = P_t G_t G_r (\frac{\lambda}{4\pi d})^2$$

λ is the wavelength of the propagating signal and antennas are used with a transmit gain of G_t and a receive gain of G_r. Figure 2 shows the free-space path loss at the carrier frequency of f_c = 1.5 GHz for different antenna gains as the distance varies.

Environment	Path loss exponent (n)
Free space	2
Urban area cellular radio	2.7–3.5
Shadowed urban cellular radio	3–5
In building line-of-sight	1.6–1.8
Obstructed in building	4–6
Obstructed in factories	2–3

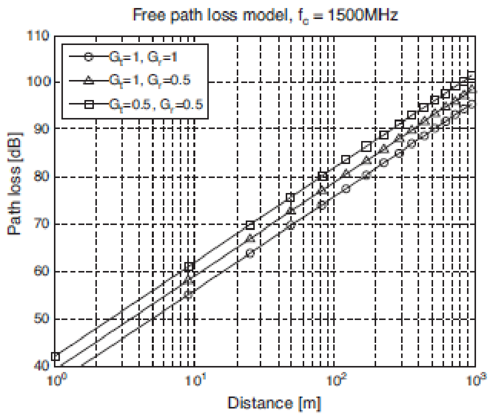


Figure 2: Free-space path loss

Using a reference distance, d_0, which is the signal attenuation at a standard distance from the antenna, the received power, P_r, at distance d is given as:

$$P_r(d) = P_0 (\frac{d_0}{d})^n$$

where P_0 =power at reference distance d_0, P_r = received power that is proportional to $1/d^n$
The received power can be written as: $P_r(d) = 10 \log [P_0(d_0)] + 10n\log (d_0/d)$ (dBm)
The accuracy of Pr can be improved by accounting for a random shadow effect caused by obstructions such as buildings or mountains. Shadow effect is described by a zero-mean Gaussian random variable, X_σ, with standard deviation σ (dB): path loss $L_p(d)$ = Path loss + Shadow effect X_σ.

$$L_p(d) = L_p(d0) + 10n\log (d/d_0) + X_\sigma \text{ (dB)}$$

2.2 SMALL-SCALE FADING

Small-scale fading is due to the small changes in position. It can be seen through time-spreading of the underlying digital pulses within the signal and the time-variant behavior of the channel due to the motion. In mobile communications, multipath interference occurs when a primary signal combines with delayed signals, typically caused by reflection or refraction from objects on or near the line-of-sight, resulting in constructive or destructive (fading) interference or phase shifts. If each path has a time variant propagation delay τ_i and a time variant multiplicative factor h_i (channel impulse response), the received signal can be written as

$$r(t) = \sum_i h_i(t)s(t-\tau_i(t)) + n(t)$$

A fading-multipath channel is characterized by several parameters:

2.2.1 DELAY SPREAD

In a multipath channel, the signal power at the receiver spreads over a certain range of time. The delay of the i^{th} signal component in excess of the delay of the first arriving component is called excess delay τ_i. Since τ is a random variable, the average $\overline{\tau}$ is the mean excess delay and the square root of the variance σ_τ is called root-mean-squared (rms) excess delay. The maximum excess delay (T_m) is defined as $\tau_X - \tau_0$, where τ_0 is the delay of the first arriving signal and τ_X is the maximum delay at which a multipath component is within X dB of the strongest signal. The delay spread is often characterized in terms of its rms value given by $\sigma_\tau = \sqrt{\overline{\tau^2} - (\overline{\tau})^2}$. $\overline{\tau^2}$ is the second moment.

(Figure 3 shows the TU 50 GSM model).

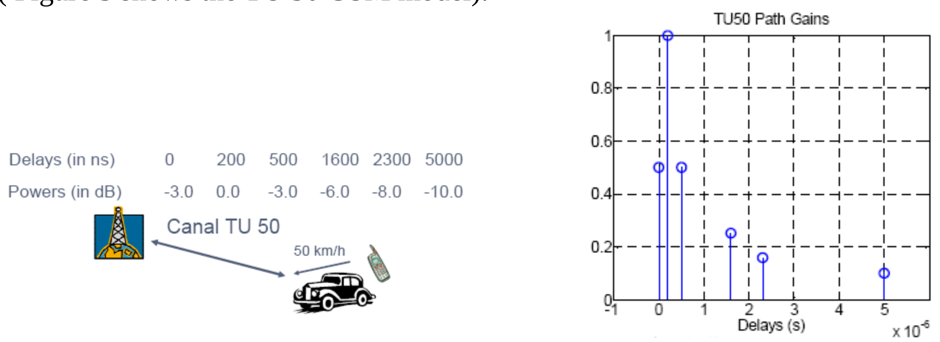


Figure 3: GSM Model

2.2.2 COHERENCE BANDWITH

The coherence bandwidth (B_c) is defined as the range of frequencies over which the channel can be considered flat, meaning the channel passes all spectral components with approximately equal gain. The coherence bandwidth of the channel is the inverse of the delay spread $B_c \approx 1/T_m$. We can relate coherence bandwidth to the rms delay spread using the following approximation: $B_c \approx \dfrac{1}{50\sigma_\delta}$.

2.2.3 DOPPLER SPREAD

Wireless mobile channel, can induce a Doppler shift in the signal of the order of vf_c/c , where v is the relative velocity between the transmitter and receiver, and c is the speed of light. Doppler spread is a measure of the spectral broadening caused by the relative movement of the mobile and base station, it is equal to the maximum Doppler frequency (f_d).

2.2.4 COHERENCE TIME

The coherence time (T_c) is defined as the time over which the channel can be assumed to be constant. The coherence time of the channel is the inverse of the Doppler spread $T_c \approx 1/f_d$.

2.2.5 CHANNEL CLASSIFICATION

We use the above parameters to classify fading channels:

- Flat fading: a fading channel is flat or frequency nonselective if the channel coherence bandwidth is much greater than the signal bandwidth, or equivalent the rms delay spread is much smaller than the signal symbol period (T) (Figure 4). Since the frequency response is flat, the received signal is as follows *r(t) = h(t)s(t)+n(t).*It is simply the transmitted signal multiplied by *h(t)* witch represents the time variant characteristics of the channel.

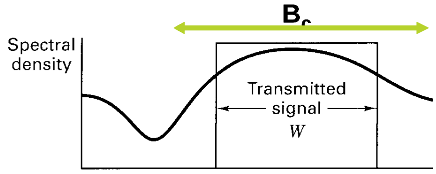


Figure 4: Flat fading

- Frequency selective: a fading channel is frequency selective if the channel coherence bandwidth is smaller than the signal bandwidth, or equivalently the rms delay spread is greater than the signal symbol period (T) (Figure 5). In this case the received signal contains multiple delayed versions of the transmitted signal. The multipath signals cause intersymbol interference (ISI). The received signal is .

$$r_k = \sum_i h_i s_{k-i} + n_k$$

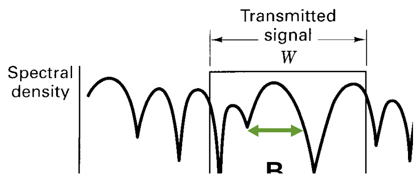


Figure 5: Frequency selective fading

- Fast fading when the channel coherence time is smaller than the symbol duration, or equivalently the Doppler spreading is greater than the signal bandwidth, a signal undergoes fast fading. The channel impulse response changes rapidly within a signal symbol duration. The received signal is

$$r(t) = \sum_i h_i(t)s(t-\tau_i(t)) + n(t)$$

- Slow fading: In a slow fading channel, the channel impulse response changes at a much slower rate than the symbol rate. The channel coherence time is much greater than the symbol duration, or equivalently. The Doppler spreading is much smaller than the signal bandwidth. The received signal is .

$$r(t) = \sum_i h_i s(t-\tau_i) + n(t)$$

- Figure 6 summarizes the different fading channel manifestations.

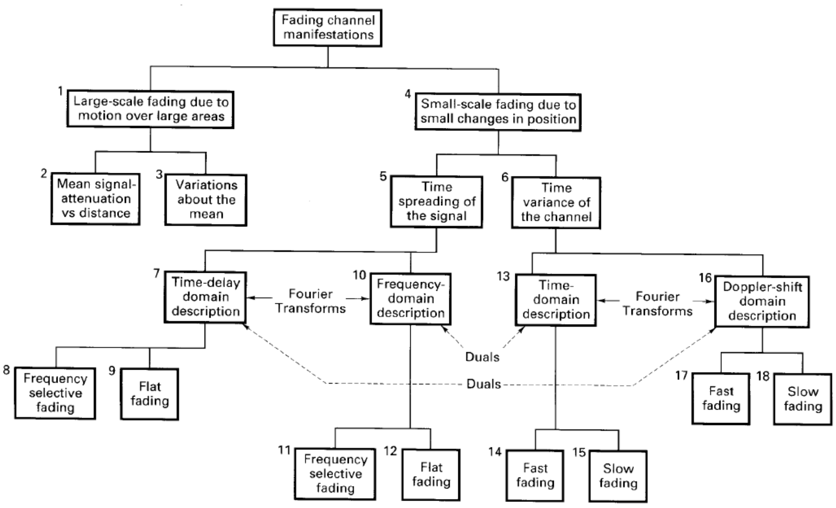


Figure 6: Different fading manifestations

2.3 STATISTICAL CHARACTERIZATION OF FADING CHANNEL

The probability of various fade depths depend upon the characteristics of the multipath environment. When there are a sufficient number of reflections present and they are adequately randomized (in both phase and amplitude), it is possible to use probability theory to analyze the underlying probability density functions (pdf's) and thereby determine the probability that a given fade depth will be exceeded. For the non-line-of-sight (LOS) case, where all elements of the received signal are reflections or diffraction components and no single component is dominant, the analysis draws on the central limit theorem since both the in phase and quadrature components of the received signal are the sum of many independent reflections. So the in-phase and quadrature components of the received signal will be independent, zero-mean Gaussian random variables with equal variance. The Rayleigh pdf is frequently used to describe the signal variations that are observed when there are many reflective paths to the receiver and no direct path and we can then determine the probability of any given fade depth being exceeded. The Rayleigh pdf is given by

$$p(r) = \begin{cases} \frac{r}{\sigma^2} e^{\frac{-r^2}{2\sigma^2}}, & r \geq 0 \\ 0, & r < 0 \end{cases}$$

The parameters of the Rayleigh density are expressed based on the variance of the underlying independent Gaussian random variables. The mean and variance of a Rayleigh random variable are given by

$$\mu_r = \sigma\sqrt{\frac{\pi}{2}}$$

$$\sigma_r^2 = \sigma^2\left(2 - \frac{\pi}{2}\right)$$

And the probabilities of any given fade depth being exceeded can be determined by evaluating the integral of the pdf directly. The expression for the probability of a fade of x dB is obtained by integrating the Rayleigh pdf from zero amplitude (the minimum value of the Rayleigh random variable) to the fade point where the fade point is referenced to the average signal power, σ^2.

$$P_x = \int_0^{\sigma 10^{-x/20}} \frac{r}{\sigma^2} e^{\frac{-r^2}{2\sigma^2}} dr$$

Using this formula, the probability of a 12-dB fade is found to be 0.031, which means that the signal will be more than 12 dB below the mean signal level about 3.1% of the time.

When there is a single dominant component, such as a line-of-sight path or a large specular reflection in the presence of multiple smaller-strength reflections, a Ricean probability density function is applicable.

The Ricean pdf is given by

$$p(r) = \begin{cases} \frac{r}{\sigma^2} e^{\frac{-(r^2+A^2)}{2\sigma^2}} I_0\left(\frac{Ar}{\sigma^2}\right), & A \geq 0, \quad r \geq 0 \\ 0, & r < 0 \end{cases}$$

Where σ^2 is the total reflection power that is received (i.e., the variance of the multipath signal), A is the amplitude of the direct or specular component and I_0 is the modified Bessel function of the first kind, order zero. The Ricean factor, K, is the ratio of the dominant component power ($A^2/2$) to the multipath power, σ^2 . It is expressed in dB by

$$K = 10\log\left(\frac{A^2}{2\sigma^2}\right) \quad \text{dB}$$

The determination of the probability of a fade is a little more complicated when the Ricean pdf is used, but it is still tractable. Figure 7 shows plots of the Ricean probability density function for several values of K. Note also that as K gets large, the Ricean pdf begins to look like a Gaussian pdf with a large mean. Of course, theoretically it can never become Gaussian because the Gaussian pdf has infinite tales and the Ricean pdf is zero for r less than zero. Nonetheless, for practical applications, once K exceeds about a factor of 10, the Gaussian pdf is a good approximation.

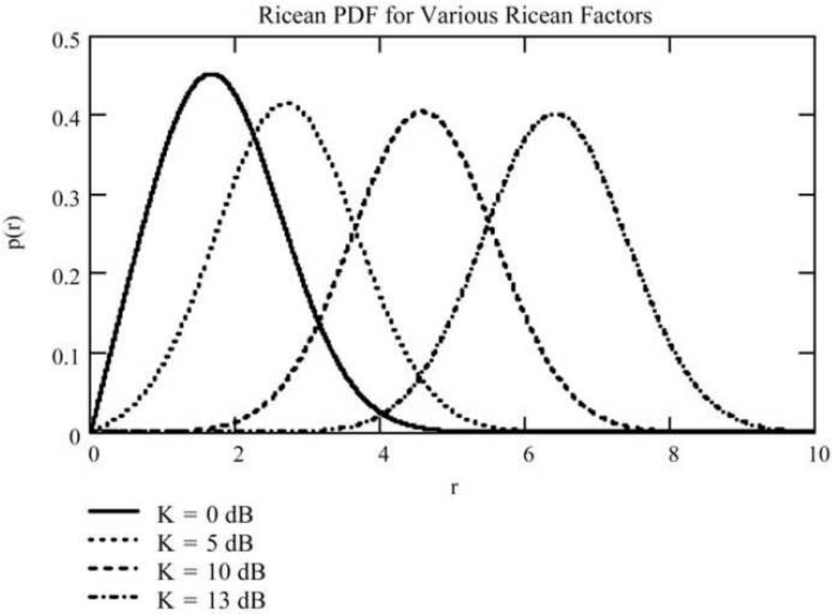


Figure 7: Ricean probability density function for several values of K

2.4 WHICH SOLUTIONS

As a result of fading, the channel introduces signal distortion and the system performance can exhibit an irreducible error rate higher than the acceptable error rate. In such cases, no amount of signal to noise ratio will help achieve the desired level of performance. Only approach available for improving performance is to use some other form of mitigation to remove or reduce the signal distortion. The mitigation method depends on whether the distortion is caused by frequency-selective fading or fast fading. Once the signal distortion has been mitigated, it is possible to further ameliorate the system to approach AWGN system performance by using some form of diversity to provide the receiver with a collection of uncorrelated replicas of the signal. In Figure 8 several mitigation techniques for combating the effects of both signal distortion and loss in SNR are listed. In the rest of this course we focus on the main techniques to combat frequency selective distortion.

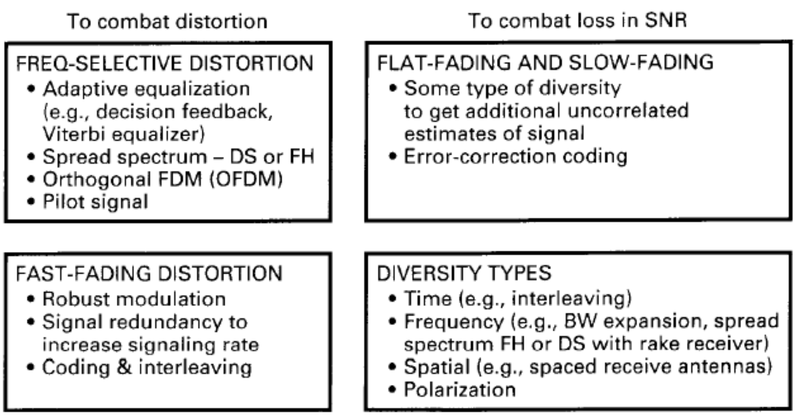


Figure 8: Mitigation techniques

2.5 EYE PATTERN

The amount of ISI, and noise in a digital communications system can be viewed on an oscilloscope. For PAM signals, we can display the received signal on the vertical input with the horizontal sweep rate set at 1/T . The resulting oscilloscope display is called an eye pattern because of its resemblance to the human eye. For example, Figure 9 illustrates the eye patterns for binary and four-level PAM modulation. The effect of ISI is to cause the eye to close, thereby reducing the margin for additive noise to cause errors.

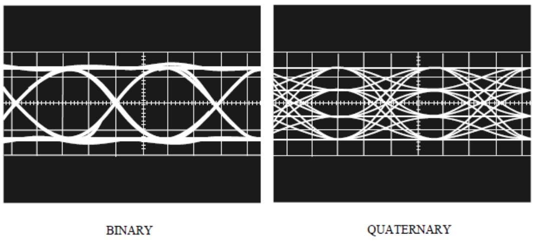


Figure 9: Examples of eye patterns for binary and quaternary amplitude shift keying (or PAM).

Figure 10 graphically illustrates the effect of ISI in reducing the opening of a binary eye. Note that intersymbol interference distorts the position of the zero crossings and causes a reduction in the eye opening.

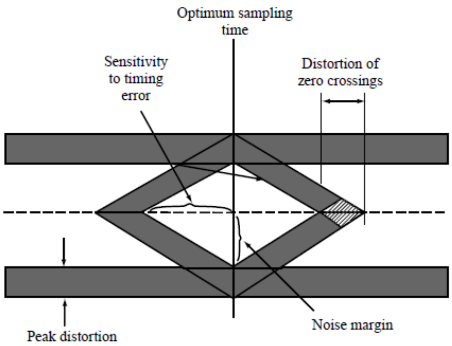


Figure 10: Effect of intersymbol interference on eye opening

3 EQUALIZATION

Channel equalizers are widely used in digital communication systems to mitigate the effects of ISI caused by channel distortion. Equalization can mitigate the effect of the ISI caused by the frequency selective fading using methods to gather the dispersed symbol energy back into its original time interval. An equalizer can be seen as an inverse filter of the channel. When the channel is frequency selective, the equalizer enhances the frequency components with small amplitudes and attenuates those with larges amplitudes. The goal is for the combination of channel and equalizer filter to provide a flat composite received frequency response. Because in a mobile system, the channel response varies with time, the equalizer filter must also change or adapt to the time-varying channel characteristics. The equalization uses different methods including:

3.1 MLSE (MAXIMUM LIKELIHOOD SEQUENCE ESTIMATION)

The MLSE equalizer tests all possible data sequences (rather than detecting each received symbol by itself) and chooses the data sequences that is most probable of all candidates (Maximum likelihood). For a sequence with length N and M states, the possible symbol's sequences are M^N. The advantage of this equalizer is the optimal error rate and the achievement of the equalization and the detection at the same time. But the disadvantage is the huge number of calculation when N increases. To reduce the number of calculation, the implementation of the MLSE equalizer is done using the Viterbi decoding algorithm which searches a state sequence through the trellis that minimizes the distance to the observation sequence.

3.2 FILTERING

Different filtering methods are used in the equalizer block shown in Figure 11 :

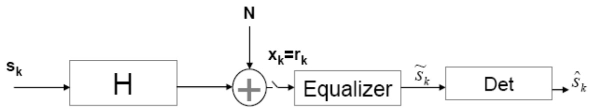


Figure 11: system model

3.2.1 TRANSVERSAL FILTERING (LINEAR):

The transversal filter shown is Figure 12 is the most popular form consisting of a delay line with T-second taps (T is the symbol duration). The current and past values of the received signal are linearly weighted with the equalizer coefficients or taps weights c_n and are summed to produce the output. Consider that there are $2N+1$ taps with weigth c_{-N}, c_{-N+1}, …,c_N. Output samples of the equalizer z_k are found by convolving the input samples x_k and the tap weights c_n as follows:

$$z_k = \sum_{n=-N}^{N} c_n x_{k-n} \qquad n = -N,...,N \quad k = -2N,...,2N$$

$$x_k = r_k = \sum_i h_i s_{k-i} + n_k$$

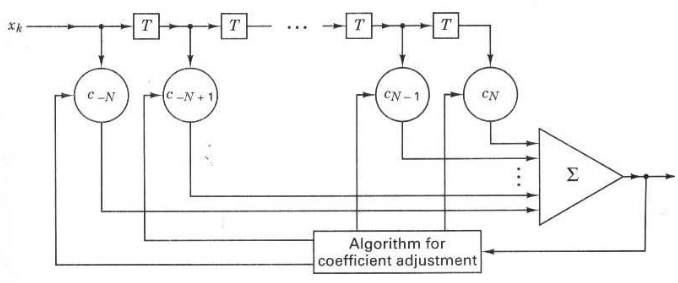


Figure 12: Transversal Filter

Two algorithms can be used to determine the (coefficients) tap weights c_n : zero-forcing and Minimum mean square error (MMSE).

<u>Zero-forcing equalizer</u>:

By defining the vector z and c and the matrix x, we can describe the relationship among z_k, x_k and c_n more compactly as

$$z = xc$$

where x is non-square matrix of dimension 4N+1 by 2N+1, z is a vector of dimension 4N+1 and c is a vector of dimension 2N+1.

One might be interested in the ISI at sample points around the symbol in question. The filter taps $\{c_n\}$ are then chosen such that the ZFE equalizer output is forced to be zero at N sample points on each side. Zero forcing solution starts by transforming x into a square matrix of dimension 2N+1 by 2N+1 and transforming z into a vector of dimension 2N+1 yielding a deterministic set of 2N+1 equations to determine the coefficients $\{c_n\}$ where

$$z_k = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{for } k = \pm 1, \pm 2, \ldots, \pm N \end{cases}$$

The solution is then given by $c = x^{-1} z$

The disadvantage of ZFE is that it neglects the effect of noise, thus not always the best solution.

<u>MMSE equalizer:</u> the filter taps are adjusted such that the MSE of ISI and noise power at the equalizer output is minimized.

$$\boxed{\text{Adjust } \{c_n\}_{n=-N}^{N}} \Rightarrow \boxed{\min E\left[(\tilde{s}_k - s_k)^2\right]}$$

3.2.2 DECISION FEEDBACK EQUALIZER:

The linear filter equalizers described in the preceding section are very effective on channels where the ISI is not severe. The basic limitation of a linear equalizer is that it performs poorly on channels having spectral nulls (severe ISI). A decision-feedback equalizer (DFE) is a nonlinear equalizer that employs previous decisions to eliminate the ISI caused by previously detected symbols on the current symbol to be detected. The DFE consists of two filters (Figure 13). The first filter is called a direct (feedforward) filter and it is generally a FIR filter with adjustable tap coefficients. This filter is identical in form to the linear equalizer already described. The second filter is a recursive (feedback) filter. It is implemented as a FIR filter with symbol-spaced taps having adjustable coefficients. Its input is the set of previously detected symbols. The output of the feedback filter is subtracted from the output of the direct filter to form the input to the detector. The tap coefficients of the direct and feedback filters are selected to optimize some desired performance measure. For mathematical simplicity, the MMSE criterion is usually applied.

DFEs offer ISI cancellation with reduces noise enhancement and may thus provide a significantly lower BER compared to linear equalizers.

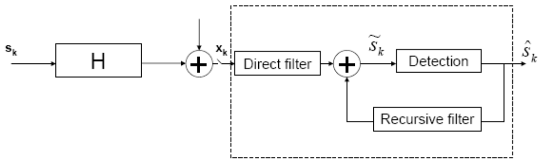


Figure 13: DFE

4 SPREAD SPECTRUM

The key to spread spectrum techniques is some function, independent of the data being transmitted, that is used to spread the information signal over a wide transmitted bandwidth. This process results in a transmitted signal bandwidth which is typically 20 to several 100 times the information bandwidth in commercial applications, or 1000 to 1 million times in military systems.

Several different methods of spread spectrum transmission have been developed, which differ in the way the spreading function is applied to the information signal: Direct Sequence Spread Spectrum, Frequency Hopping Spread Spectrum, and Time Hopping Spread Spectrum. DSSS and FHSS are most widely applied in wireless networking.

Hybrid systems also use combinations of spread spectrum techniques and are designed to take advantage of specific characteristics of the individual systems. For example, FHSS and THSS methods are combined to give the hybrid frequency division – time division multiple access (FDMA/TDMA) technique (see Section 4.2).

Of these alternative spread spectrum techniques, DSSS and FHSS are specified in the IEEE 802.11 wireless LAN standards, although DSSS is most commonly used in commercial 802.11 equipment.

FHSS is used by Bluetooth (IEEE 802.15.1), and FHSS is an optional technique for the IEEE 802.15.4a (ZigBee) specification.

4.1.1 DIRECT SEQUENCE SPREAD SPECTRUM

In Direct Sequence Spread Spectrum (DSSS), the spreading function is a code word, called a chipping code, that is XORed with the input bit stream to generate a higher rate "chip stream" that is then used to modulate the carrier.

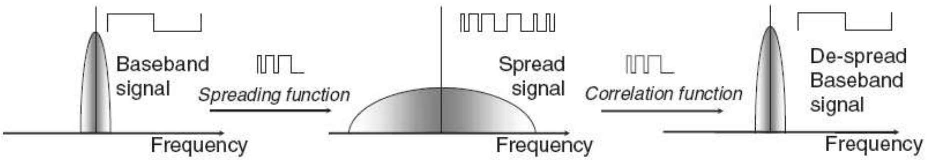


Figure 14: DSSS

4.1.2 FREQUENCY HOPPING SPREAD SPECTRUM

In Frequency Hopping Spread Spectrum (FHSS), the input data stream is used directly to modulate the carrier while the spreading function controls the specific frequency slot of the carrier within a range of available slots spread across the width of the transmission band.

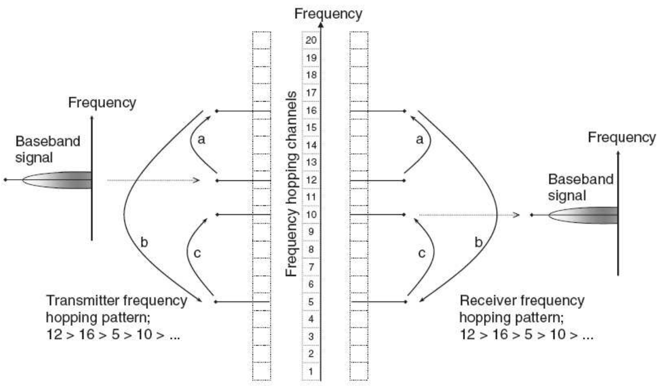


Figure 15: FHSS

4.1.3 TIME HOPPING SPREAD SPECTRUM

Time Hopping Spread Spectrum (THSS) is a third technique in which the input data stream is used directly to modulate the RF carrier which is transmitted in pulses with the spreading function controlling the timing of each data pulse.

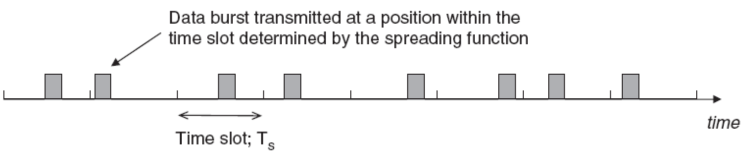


Figure 16: THSS

4.2 CDMA

Multiplexing techniques aim to increase transmission efficiency by transmitting multiple signals or data streams on a single medium. The resulting increased capacity can be used either to deliver a higher data rate to a single user, or to allow multiple users to access the medium simultaneously without interference. User access to the bandwidth can be separated by a numbers of means: in time (TDMA), in frequency (FDMA or OFDMA), or by assigning users unique codes (CDMA). These methods will be described in turn in the following sections.

In cellular networks, the combination of FDMA and TDMA is used in GSM (with eight time slots available in each 200 kHz radio channel). The combination of FDMA and CDMA is used in UMTS (R99) (Combination of FDMA, CDMA, and TDMA: HSDPA, HSUPA).

4.2.1 TIME DIVISION MULTIPLE ACCESS

Time division multiple access (TDMA) allows multiple users to access a single channel without interference by allocating specific time slots to each user. As shown in Figure 17, the time axis is divided into time slots that are assigned to users according to a slot allocation algorithm. A simple form of TDMA is time division duplex (TDD), where alternate transmit periods are used for uplink and downlink in a duplex communication system.

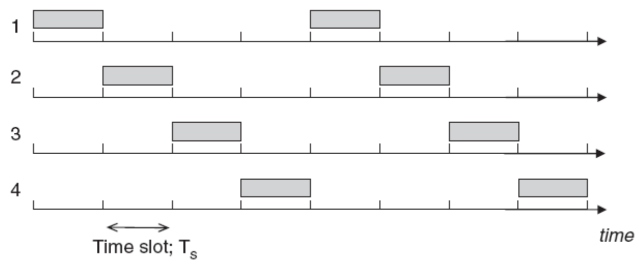


Figure 17: TDMA

4.2.2 FREQUENCY DIVISION MULTIPLE ACCESS

In contrast to TDMA, frequency division multiple access (FDMA) provides each user with a continuous channel that is restricted to a fraction of the total available bandwidth. This is done by dividing the available bandwidth into a number of channels that are then allocated to individual users as shown in Figure 18. Frequency division duplex (FDD) is simple form of FDMA in which the available bandwidth is divided into two channels to provide continuous duplex communication. Cellular phone systems such as GSM (2G) and UMTS (3G) use FDD to provide separate uplink and downlink channels.

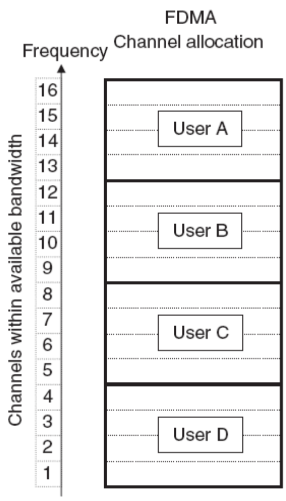


Figure 18: FDMA

4.2.3 CODE DIVISION MULTIPLE ACCESS

CDMA is closely related to DSSS, where a pseudo-noise code is used to spread a data signal over a wide bandwidth in order to increase its immunity to interference (see Figure 19). As noted above, if two or more transmitters use different, orthogonal pseudo-noise (PN) codes g_i in DS spread spectrum, they can operate on the same frequency band and in the same physical area without interfering. This is because a correlator using one PN code will not detect a signal encoded using another orthogonal code, since orthogonal codes by definition do not correlate with each other. Examples of an orthogonal code set are the Walsh codes. The property of orthogonality is the basis of CDMA and is used in 3G mobile telephony to ensure that many users, each assigned a unique orthogonal access code, can transmit and receive without interference within a single network cell.

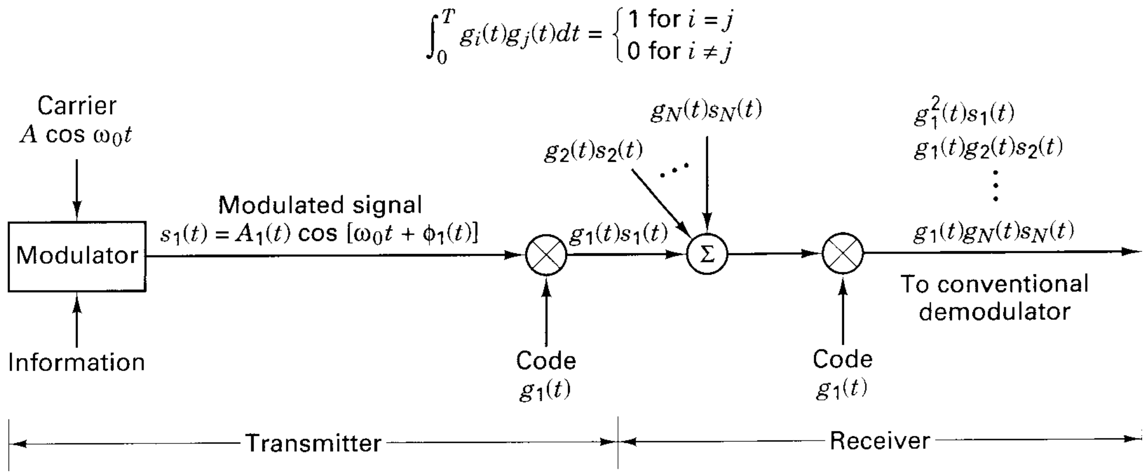


Figure 19: CDMA

Multiplication with the code sequence results in a chip stream with much wider spectrum. The ratio of the chipping rate to the information bit rate is called both the spreading factor and the processing gain of the CDMA system.

4.2.4 Capacity of a CDMA System

Consider a single cell CDMA system where a number of mobiles are simultaneously transmitting at the same frequency. Here, each mobile is assigned a unique PN code sequence. Let P be the carrier power, E_b the Energy per bit, B_c is the spread spectrum signal bandwidth, R_b be information bit rate, I is the power due to interference, N_o is noise power per bit. We know that $\frac{E_b}{N_0} = \frac{P/R_b}{I/B}$.

As stated above, we define $G_p = R_{chip}/R_b$ the *process gain,* where R_{chip} is the chip rate. The RF bandwidth $B_c = R_{chip}$. Thus $\frac{E_b}{N_0} = \frac{P}{I}G_p$

If there are N users, all transmitting at the same power and using the same chip rate, then *I=(N-1)P,*

$$\frac{E_b}{N_0} = \frac{P}{I}G_p = \frac{1}{N-1}G_p \Leftrightarrow N = 1 + \frac{G_p}{E_b/N_0}$$

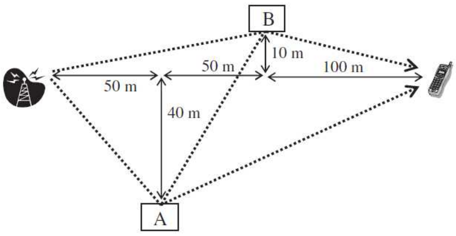
Notice that for a fixed bit error rate (that is, a fixed value of E_b/N_0), the greater the process gain, the larger the capacity N of the system. Similarly, with a fixed process gain, the capacity increases if the value of E_b/N_0 required to provide a satisfactory operation decreases. The capacity given by the previous equation is achieved only under ideal conditions. In actual practice, it may be significantly less for a number of reasons. Because the system is interference limited, the capacity of the system can be increased by reducing the interference.

Chapter 5
Problems

I. Consider a base station transmitting at 1800 MHz. Calculate the received power at a distance of 3 km from the transmitter if the path-loss exponent is 4. Assume a shadow effect of 10.5 dB and the power at reference distance (d_0=100m) of -32 dBm.

II. Consider a mobile communication system that is characterized by a power density profile made up of four impulse functions with relative power and time-delay locations 0 dB at 0 *ns*, -3 dB at 100 *ns*, -3 dB at 200 *ns* and -6 dB at 300 *ns*.
a) Calculate the mean excess delay
b) Calculate the second moment of the excess delay
c) Calculate the rms delay spread
d) Estimate the coherence bandwidth (corresponding to a correlation ≥ 0.9).

III. Given the communication system and the environment shown below, all of the buildings are good reflectors and both the transmitter and receiver are using wide-angle antennas.
a) What is the maximum expected delay spread value?
b) Based on that value, what is the highest symbol rate that you would recommend if the symbol duration greater than 2 times the delay spread?



IV. Consider a binary data sequence to be transmitted over a channel with ISI. The data signals are statistically independent and uniformly distributed and $\in \{-1,+1\}$. The transfer function of the channel is the following $H(z)=az+1+az^{-1}$.
a) What is the channel impulse response *h(t)* at symbol rate.
b) Determine the distribution density function of the received signal *r(t).*
c) When a= 0.7, determine the bit error probability if the threshold at the detector is 0.

V. Consider a BPSK data sequence to be transmitted over a multipath channel. The data signals are statistically independent and uniformly distributed and $\in \{-1,+1\}$. The discrete time model of the channel is given by $r_n=s_n+as_{n-1}$, where a is a fixed real coefficient, $0<a<1$.
a) Determine the distribution density function of the received signal *r(t).*
b) For $a = 0.5$, determine the bit error probability if the threshold at the detector is 0.
c) If the channel is with noise, the received signal is $r_n=s_n+as_{n-1}+w_n$ and w_n is i.i.d. zero-mean Gaussian variable with variance 1. Calculate the bit error probability (BER) at the receiver. when $a = 0.5$ and the threshold at the detector is 0. .

VI. Consider a discrete time model for a channel characterized by its impulse response: $h(i)= \delta(i)+0.5\delta(i-1)+0.5\delta(i-2)$
d) What is the signal to interference ratio at the receiver in the noise-less channel?
e) If we apply a linear equalizer with the impulse response $c(i)=0.6\,\delta(i)-0.2\delta(i-1)-0.1\delta(i-2)$, what is the signal to interference ratio in this case (also noise less case).

VII. Consider the output x_k of a channel sampled at kT where T is the symbol duration. $x_k=h_0a_k+h_1a_{k-1}+n_k$. a_k are statistically independent and uniformly distributed and$\in\{-1,+1\}$. n_k is zero mean noise with variance $\sigma^2=1$. A linear equalizer with 3-tap coefficients is used to minimize the ISI effects. Determine the coefficients if the equalizer is MMSE ($h_0=1$, $h_1=0.3$).

VIII. Consider a direct sequence CDMA system where 30 equal-power users share the data bandwidth. Each user transmits data at 10 kbits/s using BPSK modulation. Determine the minimum chip rate to obtain a signal to interference (E_b/N_0) of 7 dB. Assume that the receiver noise is negligible with respect to the interference from the other users.

IX. Consider a direct sequence CDMA system with a data bandwidth of 20 KHz. With only one user being transmitted, the received E_b/N_0 is about 16 dB.

a) What is the spreading factor if the system uses a spread bandwidth of 20MHz?
b) Using this spread bandwidth, how many equal-power users can share the band if the required ratio of the signal to interference and noise (SINR) is 10dB.
c) If each user's transmitted power is reduced by 3dB, how many equal-power users can share the band?