C3B Option Modern communications
Dominic O'Brien

Outline
Modern communications systems transmit digital signals using a combination of advanced coding, transmission and reception techniques. The performance of some wireless communications systems is close to theoretical limits, due to the availability of the sophisticated signal processing chips required to achieve this at very low cost, and the desire for bandwidth efficient communications. In this course the principles and techniques required are introduced, and examples of wireless LAN, ADSL and other systems in widespread use explained.

8 Lectures, 2 tutorial sheets

Suggested Texts
J. Dunlop and D. G. Smith, *Telecommunications engineering*, 3rd ed. Cheltenham: Nelson Thornes, 2001.

S. S. Haykin, *Communication systems*, 4th ed. New York ; Chichester: Wiley, 2001.

I. Glover and P. M. Grant, *Digital communications*, 2nd ed. Harlow: Prentice Hall, 2004.

Syllabus
Representation of information.
Information rate and system capacity. Shannon's law. Signal Space. Bandwidth and information rate.
Baseband signal transmission
NRZ,RZ, Manchester encoding and other examples. Generation and detection. Errors and performance.
Bandpass Signal transmission .
Information representation. ASK,FSK,PSK,DPSK. Generation and detection. QPSK, M-PSK,M-QAM, CPFSK (Continuous Phase Shift Keying), MSK, GMSK. Generation and detection. Errors and performance.
Coding.
Information rate and data rate. Bandwidth reduction techniques. Error correction techniques.
Orthogonal Frequency Division Multiplexing (OFDM).
Principles, transmission, detection.Errors and performance.
Applications and Examples.
ADSL, Wireless LAN, other applications in widespread use.

1. Modern Communications

1. Modern Communications 1
2. Introduction 4
3. Information content of signals 5
Hartleys law 7
Shannon limit 10
4. Digital Signal Transmission Through Linear Systems 13
Sampling of analogue signal 13
Quantisation and A/D conversion 14
Transmission 15
Reconstruction 15
Frequency considerations 15
Data bandwidth 15
Time considerations 16
Nyquist Channel Filtering 17
Raised Cosine Filtering 19
Eye diagrams 22
The Discrete Fourier Transform (DFT) 23
5. Baseband Data Transmission 25
Probability of Error for simple binary (two-level) system 25
An optimum detector-Matched Filtering 30
NRZ Signaling 32
Return to Zero Signaling 34
Manchester Coding 37
Alternative Mark Inversion (AMI) 39
Other Formats 40
Factors in choosing line codes: 40

6. Binary Bandpass Signalling 42
Amplitude Shift Keying (ASK) 42
Mixer 44
Switch 45
Detection 45
Phase Shift Keying (PSK) 46
Frequency shift keying 49
Differential Phase Shift Keying (DPSK) 53
Comparison of Binary Modulation Schemes 55

2. Introduction

Modern Communications

- Analogue information
- Digital transmission
- Enabled by ‘Moores Law’ scaling of processing power
- Performance approaching theoretical limits (so-called Shannon Limit)

Course will examine

- Information content of signals
- Bandwidth of signals
- Modulation schemes
- System performance
- Example systems: ADSL
- Example systems: Wireless LAN

3. Information content of signals

Consider sending a long stream of binary 'ones'- most of the time the receiver would be correct if it produced 1 at its output without actually looking at its input. The signal is not actually carrying much information. The more unpredictable a signal is the more information it carries.

The information content of a signal is measured in **binits**, shortened to **bits**. This is not the same quantity as the number of binary digits required to transmit the signal.

The information capacity of a channel is limited by bandwidth, which determines the maximum signalling speed, and by noise, which determines the number of distinguishable signal levels.

Consider a system which has a vocabulary of four signals ABCD. We have a possibility of transmitting four (4^1) different single symbol messages (A or B or C or D). If we send two symbols there are sixteen (4^2) (a choice of two from four)

In general if P symbols are sent, the number of possible messages is $4P$. Further, if there are n different symbols in the vocabulary and the message consists of P symbols, then we can have n^P different messages. So the greater the number of symbols, and the greater the length, the more **unpredictable** the message.

So we can write the information content H of a message is some function of n^P, or

$$H \propto f(n^P)$$

We can make this a function of time by saying that we transmit one symbol every t seconds for a message transmission time T. Then

$$P = \frac{T}{t}$$

And

$$H \propto f(n^{\frac{T}{t}})$$

If we assume linearity (ie a message of $2T$ duration contains twice as much information as one of T) then we can say

$$H = K \log_x (n^{\frac{T}{t}})$$

Or

$$H = K\left(\frac{T}{t}\right)\log_x n$$

Where K and x are constants to be determined. K can be set to one without loss of generality, as we can choose x to take this into account.

Consider only a two symbol vocabulary- A and B. Sending a single symbol is equivalent to sending one bit of information. Recall that $P = \frac{T}{t} = 1$ if one symbol is sent, and $n = 2$.

So

$H = 1 = \log_x 2$ and $x = 2$

Information transmitted by a source with n symbols in its vocabulary over a time T with a symbol time t, **assuming all symbols are equally probable** is

$$H = \left(\frac{T}{t}\right)\log_2 n \text{ bits}$$

The rate is therefore

$$H = \left(\frac{1}{t}\right)\log_2 n \text{ bits/sec}$$

For a single pulse $T = t$ and the information is therefore

$$H = \log_2 n = \frac{\log_{10} n}{\log_{10} 2} \text{ bits/pulse}$$

EXAMPLE

For a Pulse Amplitude Modulated (PAM) symbol set where 8 voltage levels can be transmitted $n = 8$ so a single pulse has $H = \log_2 8$ or 3 bits/pulse.

3.1. Hartleys law

So far have assumed all symbols carry equal information, but as far as the transmitter is concerned this may not be the case.

Assume

- A transmitter can transmit n equiprobable symbols. Assuming a message length n it transmits $n\log_2 n$ bits of information
- Receiver is only interested in the class to which the symbols belong. If class 1 contains n_1 symbols and class 2 contains n_2 symbols there are two messages which are of interest to the receiver.

- These have probability of $P_1 = \frac{n_1}{n}$ and $P_2 = \frac{n_2}{n}$
- Information in each message, **irrelevant to the receiver is concerned** is $n_1 \log_2 n_1$ and $n_2 \log_2 n_2$

Total information transmitted is therefore

$$H = n \log_2 n - n_1 \log_2 n_1 - n_2 \log_2 n_2$$

And H_{av} the average information per symbol sent is

$$\frac{H}{n} = \log_2 n - \frac{n_1}{n} \log_2 n_1 - \frac{n_2}{n} \log_2 n_2 \text{ bits}$$

or

$$H_{av} = \left(\frac{n_1 + n_2}{n}\right) \log_2 n - \left(\frac{n_1}{n}\right) \log_2 n_1 - \left(\frac{n_2}{n}\right) \log_2 n_2$$

or

$$H_{av} = P_1 \log_2 P_1 - P_2 \log_2 P_2$$

To find the maximum value of H_{av} with respect to message probability eliminate P_2 differentiate and set to zero.

$$\frac{dH_{av}}{dP_1} = -P_1 \frac{1}{P_1} - \log_2 P_1 + (P_1 - 1)(-\frac{1}{1 - P_1}) + \log_2 (1 - P_1)$$

Which is zero when $P_1 = P_2 = 0.5$

So the average information is a maximum when the symbols are equiprobable. Develop idea for n symbols in n classes to give

$$H_{av} = -\sum_{i=1}^{n} P_i \log_2 P_i$$

By analogy with statistical mechanics this is known as entropy. H_{av} is the entropy of the message. If all the symbols are equiprobable

$$H_{av} = -\sum_{i=1}^{n} P_i \log_2 n = \log_2 n$$

since

$$\sum_{i=1}^{n} P_i = 1$$

In any situation where the symbols are not equiprobable, the message contains redundancy. (useful in image compression and so forth). If there is knowledge of the probability distribution of the message this can be used to compress it, or assist with error correction.

Can define message redundancy R as

$$R = \frac{H_{av(\max)} - H_{av}}{H_{av(\max)}}$$

3.2. Shannon limit

The capacity of a channel is limited by noise. In a noiseless channel we could send an infinite number of voltage levels to represent symbols and a detector could always distinguish between them. Noise corrupts these levels, so that the capacity is finite. Claude Shannon (see the wikipedia entry) calculated the capacity of a channel in 1948. Only in 1995, with the invention of Turbo codes were capacities very close to this theoretical limit achieved.

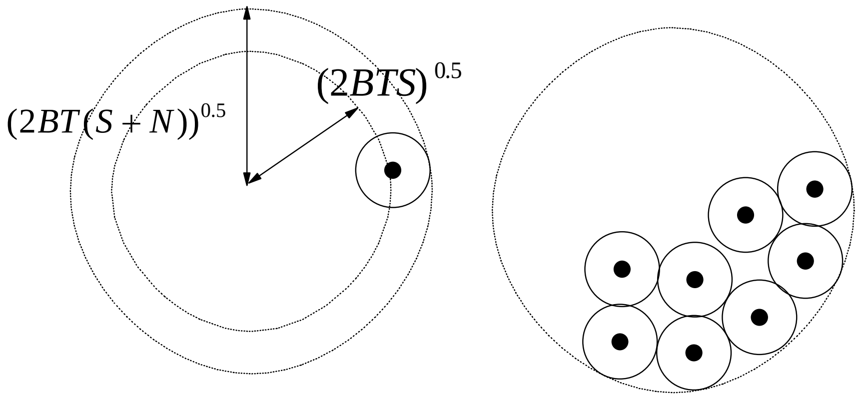


Figure 1. Signal representation

Figure 1 shows the representation he used. Consider a message consisting of n symbols, $x_1, x_2, ...x_n$. This can be represented as a point in $n-$dimensional space. Different potential messages are represented by different points in the space.

Assuming a message T seconds in length, with symbols transmitted at $2B$ symbols/sec. If the average energy per sample is S then the energy in the message is

$$E = 2BTS = k(x_1^2 + x_2^2 + x_3^2 + x_4^2 .... + x_n^2)$$

Where k is some constant of proportionality.

So the points representing the message are bounded by a hypersphere with a radius proportional to the message energy.

The noise adding to each of the symbols $x_1, x_2, ...x_n$ can be modelled as a small hypersphere around the noiseless symbol points. This represents the maximum the signal can be from its expected value in the symbol space.

Now have two hyperspheres-one representing the symbol space bounded by the signal energy, and one representing the effect of the noise.

Figure 1 shows the conceptual calculation of the maximum number of distinguishable levels. If the noise hyperspheres overlap then the signals cannot be distinguished, so the maximum number of levels is just the ratio of the volumes. The volume of a hypersphere of n dimensions is proportional to r^n but as $n \to \infty$ all the volume is concentrated on the surface, so all the message points will lie on the surface of the sphere.

Assuming there is N Joules of noise energy per symbol the noise can be represented by a hypersphere of radius

$$(2BTN)^{\frac{1}{2}}$$

And the signal + noise by a hypersphere of radius

$$(2BT(S+N))^{\frac{1}{2}}$$

The number of levels is therefore just the ratio of the volumes, or

$$= \left[\frac{S+N}{N}\right]^{\frac{n}{2}}$$

Where $n = 2BT$

Assuming the symbols are equiprobable then the number of bits that can be transmitted is given by

$$\log_2\left[\frac{S+N}{N}\right]^{BT} = BT\log_2(1+\frac{S}{N})bits$$

Leading to a channel capacity of

$$C = B\log_2(1+\frac{S}{N}) \text{ bits/sec}$$

THIS IS SHANNON'S LAW. For large signal to noise ratios this approximates to

$$C = B\log_2\left(\frac{S}{N}\right) \text{ bits/sec.}$$

This indicates that a linear change in bandwidth requires an exponential change in SNR to maintain a given information transmission rate.

- Need a large increase in signal power to compensate for relatively small reductions in bandwidth.
- In most systems noise power is a linear function of bandwidth i.e. $N = N_o B$ where N_o is the noise power spectral density (measured in W/Hz).
- Bandwidth is doubled the noise power doubles (or increases by 3dBs-recall what a dB is). So for a given required capacity competing effects-improved capacity for higher bandwidth, but usually leads to a decrease in SNR (which one wins?).
- So in reality if we increase the bandwidth to try and achieve a reduction in signal power we will induce an increase in noise power.

4. Digital Signal Transmission Through Linear Systems

Figure 2 shows a block diagram of a digital transmission system

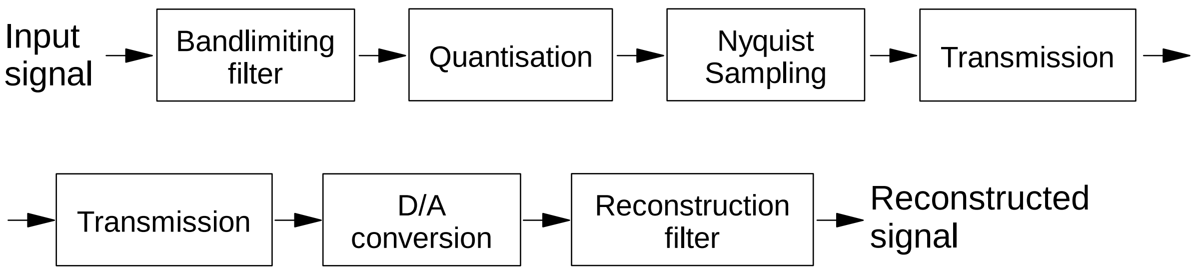


Figure 2. Digital transmission system

4.1. Sampling of analogue signal

Consider an analogue signal $h(t)$ with a frequency spectrum $H(f)$. First stage is to bandlimit so it lies in the range $-W \leq F \leq W$ Hz. Leads to a signal $h_b(t)$ with frequency domain representation $H_b(f)$.

If this is sampled with a train of dirac-δ functions with period T_s and frequency $f_s = 1/T_s$ Then Nyquist sampling requires that $f_s > 2W$.

Figure 3 shows a ‘picture proof’. See HLT for the individual transforms

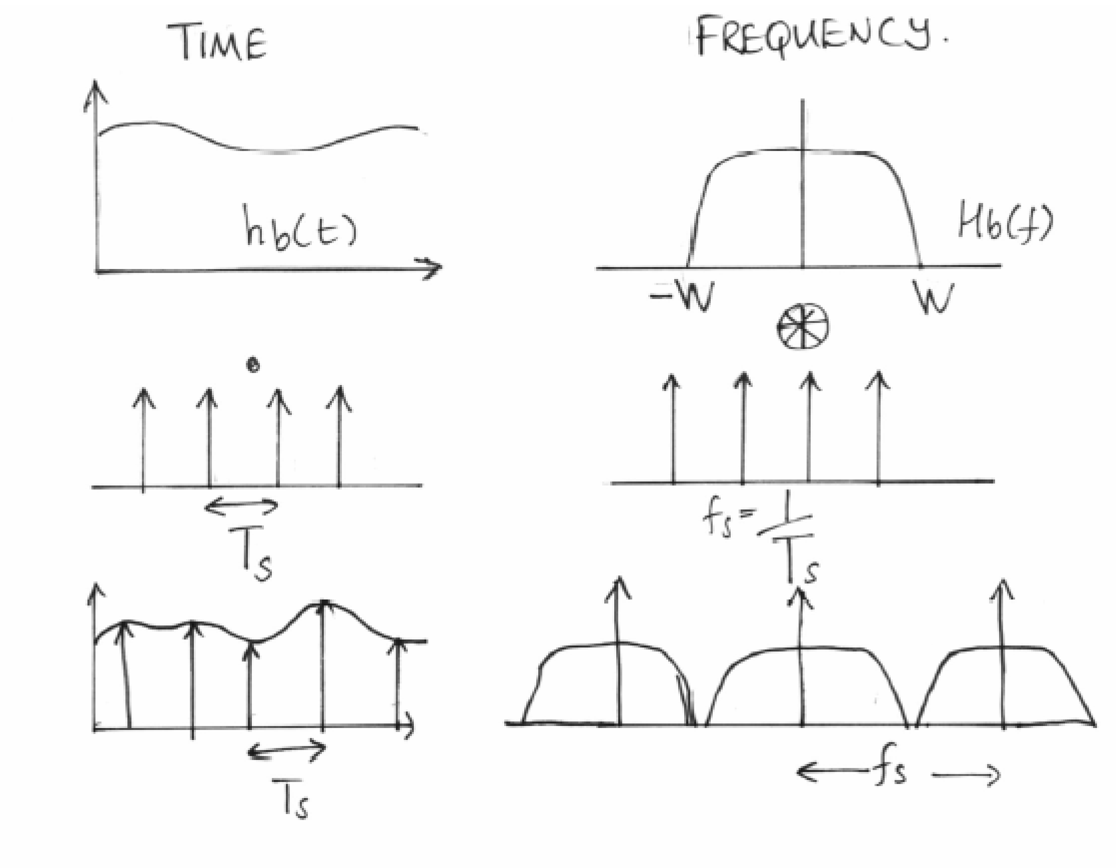


Figure 3. Nyquist sampling

Two considerations in choosing A/D converter and sampling rate.

4.2. Quantisation and A/D conversion

- Nyquist criteria
- ‘slew-rate’ of signal into the A/D converter and sampling window of converter. (Recall that is signal changes rapidly more than 0.5 bit of signal change could occur during the time the sampling window is open-thus causing error)-Recall the A3 instrumentation course.
- Will still be loss of information and error due to quantisation.
- For a quantisation resolution (LSB) ΔV the RMS error due to quantisation (assuming uniformly distributed samples) is $\Delta V^2/12$.

4.3. Transmission

Dealt with in later parts of the course.

4.4. Reconstruction

Assume D/A converter produces one analogue sample every T_s seconds.

- Know that the spectrum of this is multiple copies of the required bandlimited frequency spectrum $H_b(f)$.
- Required spectrum is just the copy centred about zero frequency, so need to filter this copy out with a **reconstruction** filter. Two ways to consider this.
- In frequency domain an ideal 'brickwall' filter cutting off at W Hz
- In time domain each sample from the A/D is shaped by the impulse response of the reconstruction filter. (An 'impulse' of a height (strictly weight) corresponding the the analogue conversion of the binary input to the A/D) enters the filter which responds with its impulse response. This impulse response correctly 'fills in the gaps' between the samples.
- Seems impossible that the filter would 'know' how to interpolate correctly, but the bandlimiting of the input spectrum might be thought to 'constrain' the waveform shape between sampling points so the interpolation is an exact reconstruction.

4.5. Frequency considerations

4.6. Data bandwidth

- Random information (one of m levels –usually binary and equiprobable).
- Fixed pulse shape.
- Need to calculate the spectrum.

Not possible to get amplitude spectrum (non-deterministic signal)

Find Power Spectral Density (PSD)

1. Find autocorrelation of signal (Often using ‘picture maths’)
2. Fourier transform to obtain PSD

More specific examples will follow, but generally the PSD is the square of the Fourier Transform of the pulse shape

Key decisions are therefore

- what pulse shape?
- what channel bandwidth do I need for the required bit rate and pulse shape? The ratio of bit-rate to bandwidth is a function of the pulse shape and this should be minimised.

Will always be the case that bandwidth is finite

4.7. Time considerations

A single pulse that has a fixed duration in time will have an unbounded frequency response.

When transmitted through a real channel the pulse becomes bounded in frequency, and thus unbounded in time.

Figure 4 shows a schematic of transmission of pulses. In the time domain the information can be represented by a train of δ functions convolved with the pulse shape,together with a weighting function representing the ‘ones’ and ‘zeros’ of the data stream.

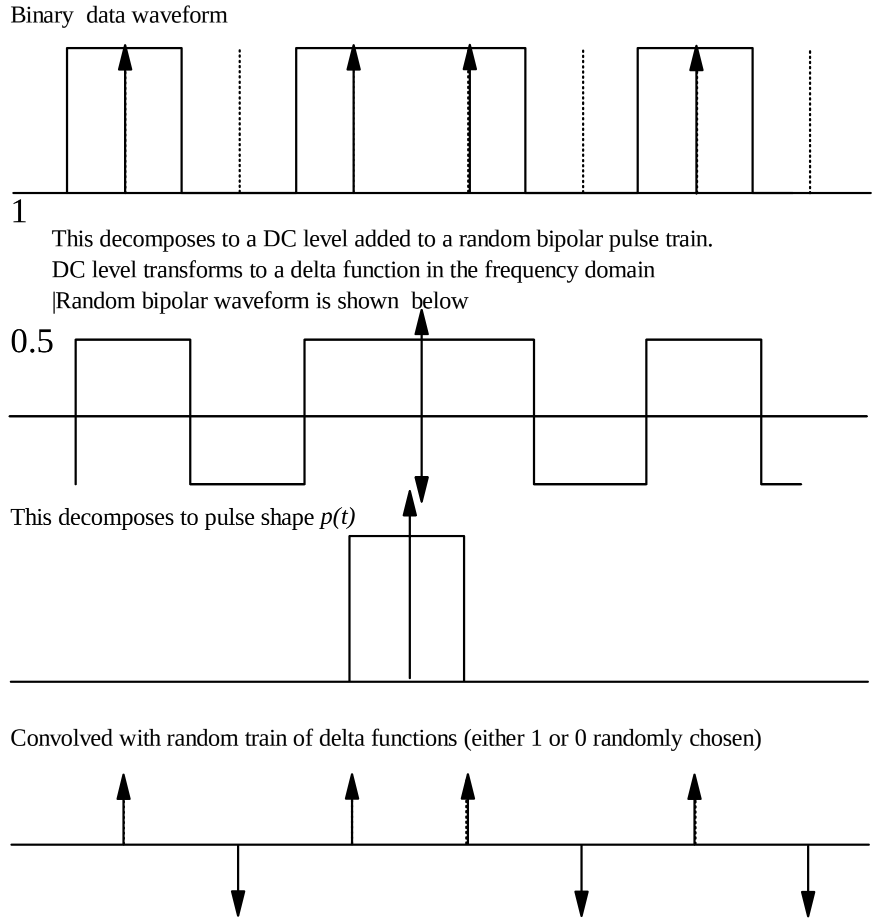


Figure 4. Decomposition of data transmission waveform

4.8. Nyquist Channel Filtering

When this waveform passes through the bandlimited channel (or any filtering process that limits bandwidth) each of the pulses is spread in time and overlaps with adjacent pulses, This is intersymbol interference. This can be caused by a number of factors. (which will be discussed later).

At the receiver,for each pulse, a decision is made about its level, usually at the middle of the pulse, so as long as the contribution from each of the adjacent pulses is zero at this time instant there is no intersymbol interference ISI. These are called Nyquist pulses and are achieved if the net channel response (i.e. the effect of transmission pulse shaping, channel response, and receiver filtering result in a Nyquist channel response when measured at the decision point in the receiver).

In the trade the characteristic transfer function is described as having a transition between pass band and stop band that is symmetrical about a frequency equal to $0.5/T_s$.

Figure 5 shows the required symmetry.

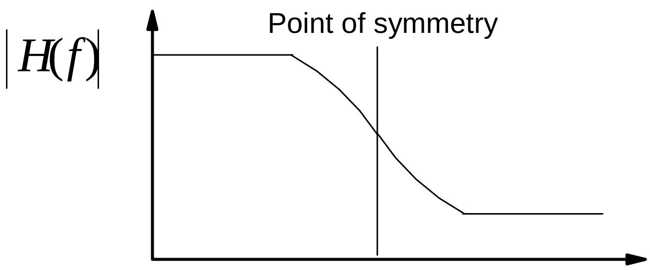


Figure 5. Symmetry required for Nyquist filtering

Nyquist responses include

- A brickwall response will generate this- creating a sinc envelope impulse response. Figure 6 shows this response.
- Raised cosine filter.

Often filtering done at transmitter and receiver (rather than one end- why?) so in this case a root-raised cosine response is used (so they add together to give the net raised cosine desired).

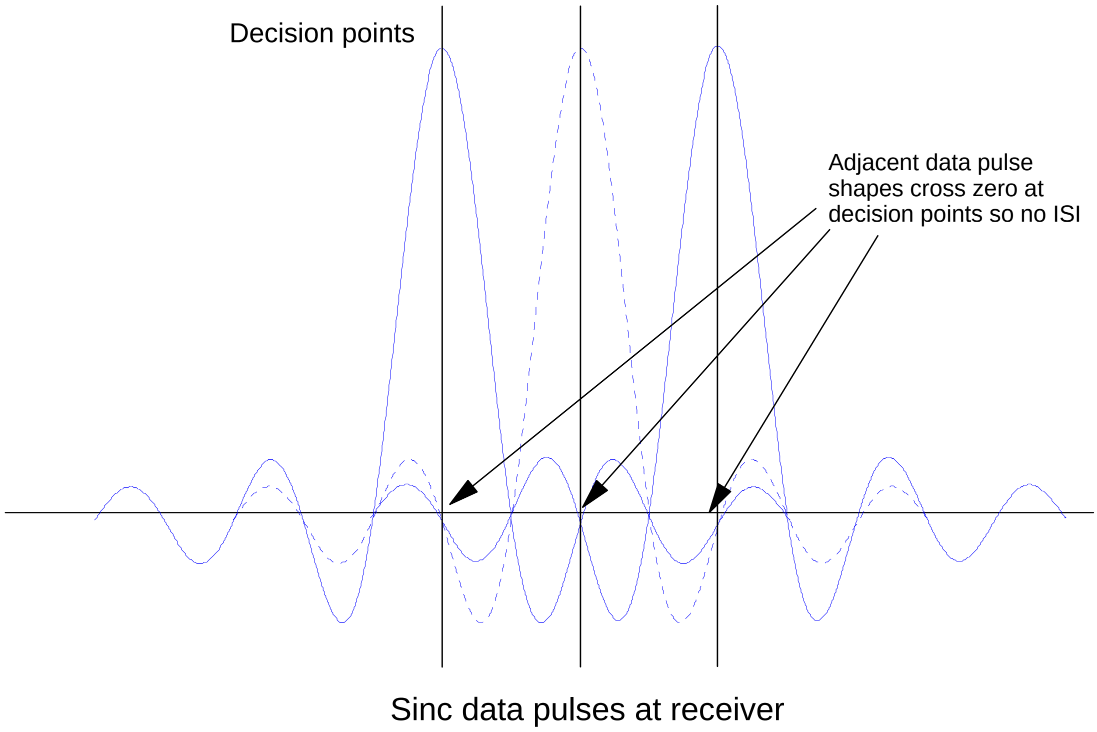


Figure 6. Sinc pulses showing Nyquist response

4.9. Raised Cosine Filtering

The task is therefore to design a pulse shaping scheme to provide for minimal ISI and minimum BER. Such schemes can be found in the raised cosine family of filters (so named because of their shape in *time*). The sharpness of the filter is controlled by the parameter α, called the roll off factor. This factor is used to describe several filter families, but here the bandwidth occupied by the data signal is increased from its minimum value of $B_{\min} = 0.5/T_s$ to $B_{\min}(1+\alpha)$.

The frequency response is given by

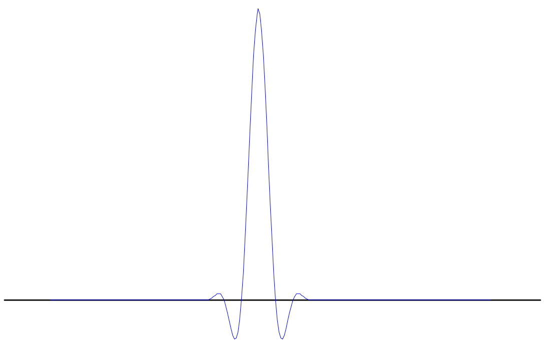
$$H(f) = \begin{cases} T & 0 \le |f| \le \dfrac{1-\alpha}{2T} \\ \dfrac{T}{2}\left\{1+\cos\left(\dfrac{\pi T}{\alpha}\left(|f| - \dfrac{1-\alpha}{2T}\right)\right)\right\} & \dfrac{1-\alpha}{2T} \le |f| \le \dfrac{1+\alpha}{2T} \\ 0 & |f| \ge \dfrac{1+\alpha}{2T} \end{cases}$$

This has a pulse shape

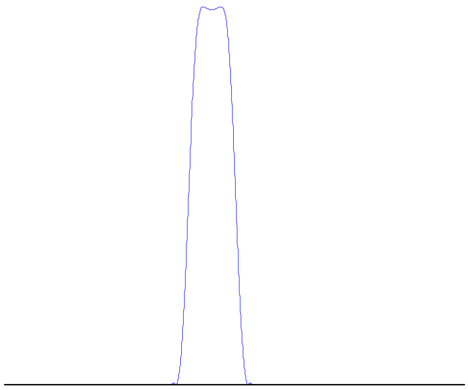
$$x(t) = \frac{\sin(\pi t / T)}{\pi t / T} \frac{\cos(\pi \alpha t / T)}{1 - 4\alpha^2 t^2 / T^2}.$$

Note that the zero crossings as were seen in the sinc pulses are maintained, but the envelope dies away quicker.

Figure 7 shows the response. n many channels a transmit and receive filter is used, so each uses a root-raised cosine pulse shape so that the net multiplicative effect of the filters has the required raised cosine response.



Impulse response of raised cosine filter



Frequency response of raised cosine filter

Figure 7. Response of raised cosine filter

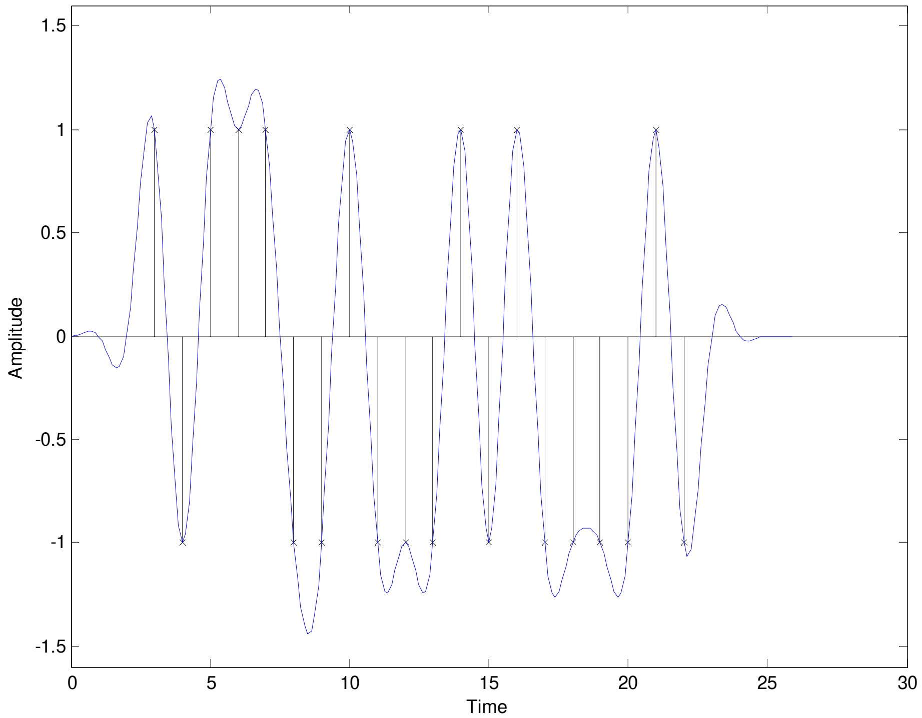
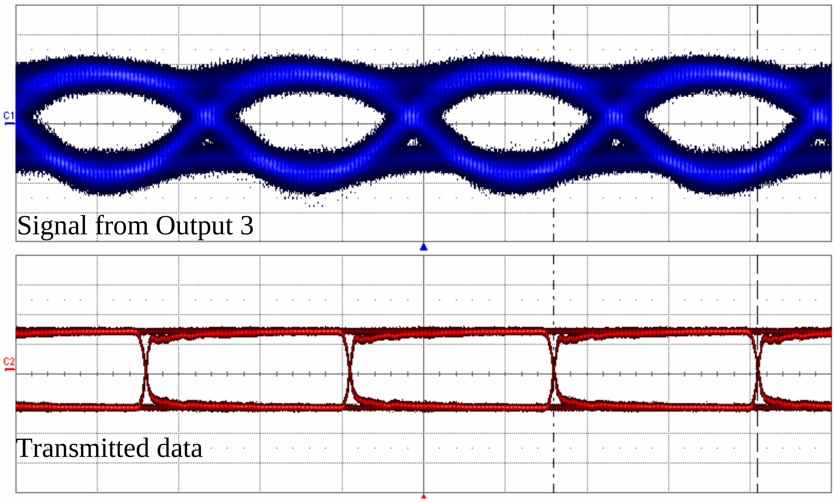


Figure 8. Filtered data stream showing no ISI at sample points

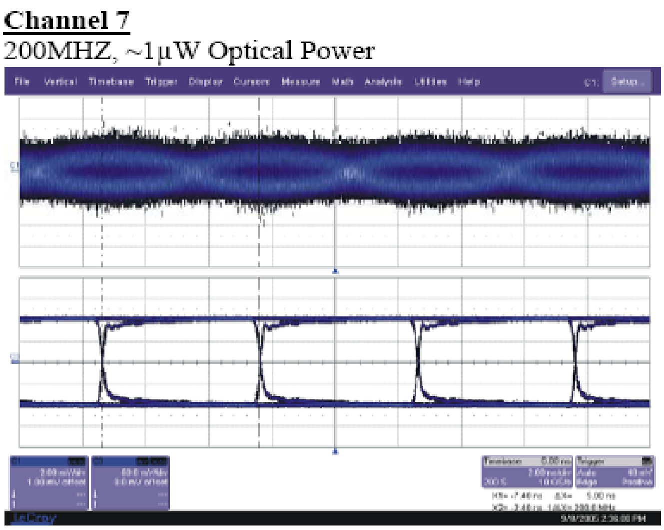
In Figure 8 the data transmitted is delta functions and these are convolved with the raised cosine filter, causing ISI. However, this is zero at sampling instants, as the received data matches the transmitted delta functions.

4.10. Eye diagrams

Examining received digital data can be achieved using an 'eye' diagram as shown in Figure 9.



Open eye (100Mb/s)



Closed eye (100Mb/s)

Figure 9. Eye diagram

The oscilloscope is triggered of the data clock, so if the received data is mis-timed relative to this the data transitions do not occur at the same time instant causing blurring on the vertical lines. If the system does not respond to levels properly the horizontal levels can blurr. An open 'eye' therefore shows good received timing and distinguishable one and zero levels. A closed ‘eye’ shows this is a problem. High-end oscilloscopes can obtain quantitative measurements from eyes, by comparing with a mask corresponding to the desired performance standard. More commonly the analysis is ‘it looks fairly open’.

4.11. The Discrete Fourier Transform (DFT)

The DFT is an extension of the continuous Fourier Transform designed for regularly sampled signals. Figure 10 shows representative spectra.

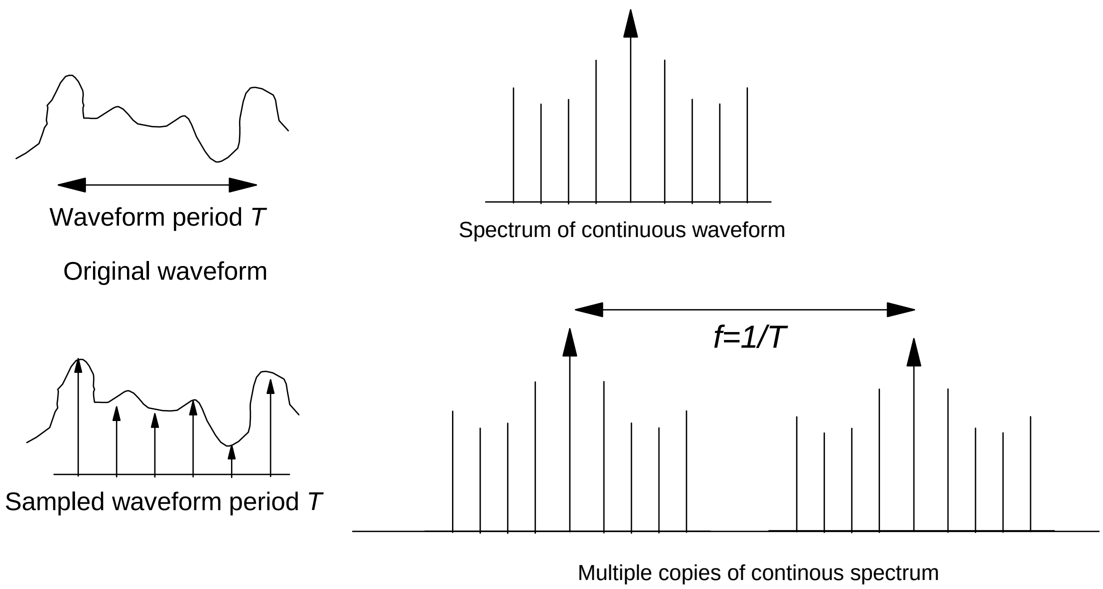


Figure 10. Spectrum of continous and sampled spectra

Now develop the DFT for a periodic signal $h(t)$ with no components at or above a frequency f_x=x/*T*, *x* being an integer and *T* being the period of the signal waveform.

The Fourier series of $h(t)$ can be written, for values $t = kT$, where $k = 1,2,3,$

$$h(k\Delta t) = \frac{1}{T}\sum_{n=-(x-1)}^{x-1} C_n \exp(j2\pi nk\Delta t/T)$$

but over the range -(x-1)$\leq$n $\leq$ (x-1) the coefficients C_n are identical to $C(n\Delta f)$; hence

$$h(k\Delta t) = \frac{1}{T}\sum_{n=-(x-1)}^{x-1} C(n\Delta f)\exp(j2\pi nk\Delta t/T)\,.$$

If there is a total of N samples in the interval *T,* then *T* = *N*Δ*t* and the range of *k* is 0, $\pm$1, $\pm$2,.. $\pm$ [(N/2) –1]. Since we can write *t* = 1/2f_x = *T/2x* then *N* = *2x* and we have

$$h(k\Delta t) = \frac{1}{N\Delta t}\sum_{n=-(x-1)}^{x-1} C(n\Delta f)\exp(j2\pi nk/N)$$

and we see that *C(n*Δ*f)* is periodic and we can change the range of n to make *C(n*Δ*f)* symmetrical about the frequency f_x.

$$h(k) = \frac{1}{N}\sum_{n=0}^{N-1} C(n)\exp(j2\pi nk/N)$$

We can now work to the Discrete Fourier Transform representation by noting that the amplitude spectrum *C(n*Δ*f)* is obtained from the Fourier transform expression and remembering **that *h(t) only exists for discrete values of t,*** we can write the FT as a summation:-

$$C(n\Delta f) = \sum_{k=-N/2-1}^{N/2-1} h(k\Delta t)\exp(-j2\pi nk\Delta t/T)\Delta t$$

Note Δ*t* = *T/N* and *h(k*Δ*t)* is a periodic function so that the limits of the summation can be changed to give:-

$$C(n\Delta f) = \Delta t\sum_{k=0}^{N-1} h(k\Delta t)\exp(-j2\pi nk/N)$$

We usually write this as :-

$$C(n) = \sum_{k=0}^{N-1} h(k)\exp(-j2\pi nk/N)\,.$$

5. Baseband Data Transmission

Baseband transmission means sending the digital pulses ‘as they are’, using the bandwidth from DC to the frequency which is required for proper transmission.

Use in optical fibre systems (though a region very close to DC is cut-off) and telephone transmission (for speech anyway). Ethernet and many short-distance data transmission schemes are baseband

Most wireless transmission systems take the baseband data and then modulate it into some allowed band (away from the baseband) before transmission.

Recall that

Shannon gave capacity (error free) in terms of signal to noise ratio. In real channels noise will cause errors and the SNR required to achieve a certain error rate must be determined.

As much (most) transmission is of digital data the Bit Error Rate (BER) is used.

- For fibre systems 10^{-9} was typically used as a target. This is very hard to achieve at 40Gb/s and rates closer to 10^{-6} (a lot of errors per second) are used with error correcting codes
- For RF transmission 10^{-3} is more typical and this requires error correction.

In any case need to be able to estimate (and it is an estimate) the BER of a transmission system.

5.1. Probability of Error for simple binary (two-level) system

- Consider the transmission of a unipolar non-return to zero (NRZ) digital signal
- Binary 1 is a high voltage, binary 0 is a low voltage

- If noiseless probability distributions would be δ functions
- Random noise will add to the signal and disturb this. Noise is random and white in frequency, which creates a Gaussian distribution o f likely voltages with zero mean- known as Additive White Gaussian Noise (AWGN)
- Now possibility of error- that noise added to the zero level will climb above the mid-point and be detected as a one, or similar for the one level.

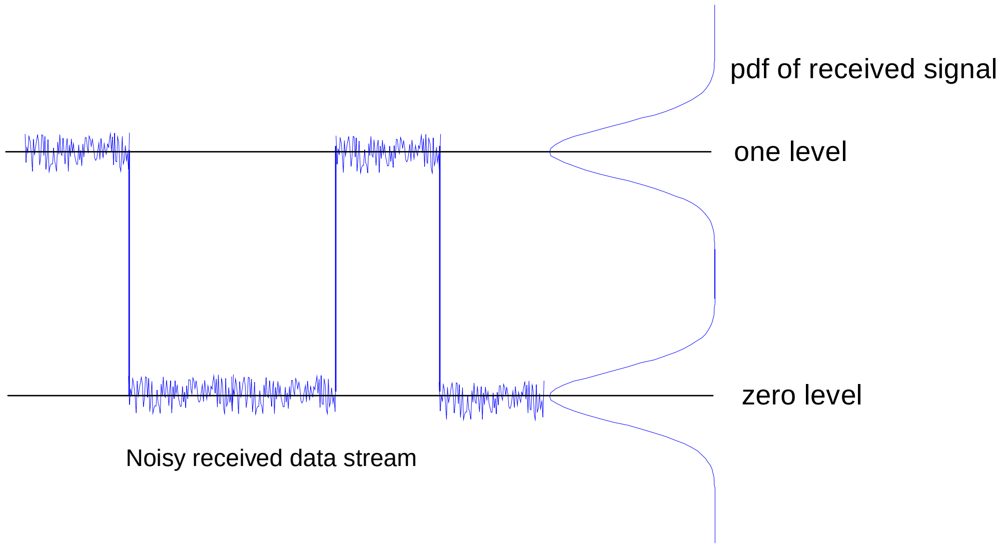


Figure 11. Received data corrupted by noise

Figure 11shows the received data waveform.

Calculating the probability of error.

Assume noiseless data waveform has levels $\{0, A\}$. The received waveform $y(t)$, will be of the form:-

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

where the $s(t)$ is the signal waveform and $n(t)$ is the noise component.

The two possible signals received are

$y(t_i) = A + n(t_i)$ for a "1" sent

and $y(t_i) = n(t_i)$ for a "0" sent

where t_i is the time when the decision is made. The decision detector operates by comparing *y(t$_i$)* against a threshold level μ such that : $y(t_i) > \mu$ "1" output

: $y(t_i) < \mu$ "0" output

Where would you normally place μ and when would you not do this?

Then the probability density function of the "0" waveform is :

$$p_o(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2 / 2\sigma^2}$$

Where the variance of the noise is σ^2. Similarly the PDF of the "1" is:

$$p_o(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-A)^2 / 2\sigma^2}$$

The probability of a 0 being mistaken for a 1 is the area under the 0 PDF between μ and infinity:

$$P_{e0} = \int_{\mu}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2 / 2\sigma^2} dy$$

Similarly the false alarm probability of a 1 being mistaken for a 0 is given by:

$$P_{e1} = \int_{-\infty}^{\mu} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-A)^2 / 2\sigma^2} dy$$

The total probability of error is the probability that a zero was transmitted multiplied by the probability that it will be misdetected (if you don't actually send a zero you can't misdetect it) plus the probability that a one was transmitted multiplied by the probability that a one will be misdetected.

Total probability of error is therefore:

$$P_e = P_0 P_{e0} + P_1 P_{e1}$$

Normally assume that the decision point μ is in the middle of the voltage levels and both distributions are identical, so $P_{e0} = P_{e1}$.

Then

$$P_e = P_{e1}(P_0 + P_1) = P_{e1}$$

Thus the probability of error is given by :

$$P_e = \int_{-\infty}^{-A/2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/2\sigma^2} dy$$

This can be written as

$$P_e = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/2\sigma^2} dy - \int_{-A/2}^{0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/2\sigma^2} dy$$

First term is just half the Gaussian pdf, so expression is

$$P_e = \frac{1}{2} + \int_{0}^{-A/2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/2\sigma^2} dy$$

This can then be written as

$$P_e = \frac{1}{2}\left\{1 - erf\left[A/(2\sqrt{2}\sigma)\right]\right\}$$

Where

$$erf(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy$$

Or

$$P_e = \frac{1}{2} erfc\left[A/(2\sqrt{2}\sigma)\right]$$

Where $erfc(x) \equiv 1 - erf(x)$

Need this in terms of signal and noise powers.

Signal power

On average half pulses are amplitude A so have power proportional to A^2 and other half have zero power so signal power

$$S = 0.5A^2$$

Noise power is

$$N = \sigma^2 .$$

So for our Unipolar pulse transmission with NRZ pulses

$$P_e = \frac{1}{2}\left\{1 - erf\left(\frac{1}{2}\sqrt{\frac{S}{N}}\right)\right\}$$

For bipolar NRZ coding with the same peak-to-peak amplitude $\{A/2, -A/2\}$ the average signal power $S = A^2/4$, giving

$$P_e = \frac{1}{2}\left\{1 - erf\left(\frac{1}{\sqrt{2}}\sqrt{\frac{S}{N}}\right)\right\}$$

So bipolar is better (has lower error) for the same average power or requires $\sqrt{2}$ as much (1.5dB) more average signal power to achieve the same BER. (Why?).

Other similar cases exist (see D&S p153 for the ASK/PSK case).

A further advantage of polar NRZ is that the decision threshold (assuming equiprobable 1s and 0s) is zero. This is helpful if the channel attenuates the signal in an unpredictable way (why?).

5.2. An optimum detector-Matched Filtering

Want to improve SNR by passing through a filter ‘matched’ to the characteristics of the signal.

Consider the input to a filter $x(t)+n(t)$ where *x(t)* is the signal and *n(t)* is the additive white noise. The impulse response of the filter is $h(t)$ such that the output $y(t)$ is:

Assuming the filter output at T is required to make the decision aim is to maximise

$$\frac{x_o(T)}{\overline{n_o^2(t)}}$$

This occurs if the impulse response is given by (see D&S p 156)

$h(t) = x(T-t)$.

Which is the time reversed version of the symbol we want to detect.

Several different interpretations

- The convolution of a signal and a time reversed version of itself is equivalent to correlation, and the shifting by T creates the maximum output at the end of the pulse.
- Expect the frequency response of a filter to be matched to the frequency spectrum of input signal, so the matched filter in this domain would be:

$$H(f) = F[h(t)] = X^*(f)e^{-j2\pi fT}$$

where $F[]$ denotes Fourier Transform, and * the complex conjugate.

Thus the magnitude response of *H(f)* is the same as the original signal (as we would expect) but on transforming to the time domain this gives us the time reversal element.

The spectral amplitude matching gives the peak value its optimum SNR, whilst the spectral phase matching produces the desired peak at time *T*. Signal output of the matched filter is the total energy contained in $x(t)$.

5.3. NRZ Signalling

Figure 12 shows a bipolar NRZ data stream

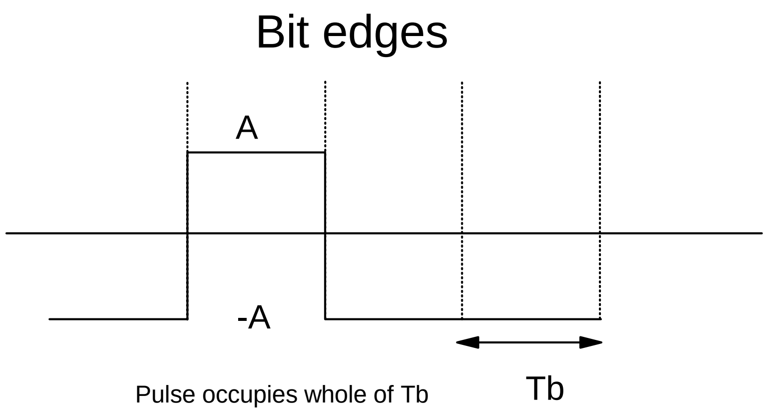


Figure 12. Bipolar NRZ transmission

Bandwidth

PSD Spectral Density $S(f)$ for bipolar. With the output level at a specified constant value $\{A,-A\}$ over the bit interval T_b (corresponding bit-rate R_b),

$$S(f) = A^2 T_b \operatorname{sinc}^2\left(\frac{2\pi f T_b}{2}\right)$$

Where $\operatorname{sinc}(x) = \sin(x)/x$ -NB this is not the same definition as HLT but is the one used in Proakis.

For a bit-rate R_b the bandwidth required depends on energy.

$-R_b/4 > f > R_b/4$ captures 70%.

$-R_b/2 > f > R_b/2$ captures 90%.

$-R_b > f > R_b$ captures 98%.

The 3dB bandwidth is when the bandwidth is 44% of the bit rate. Figure 13 shows the PSD.

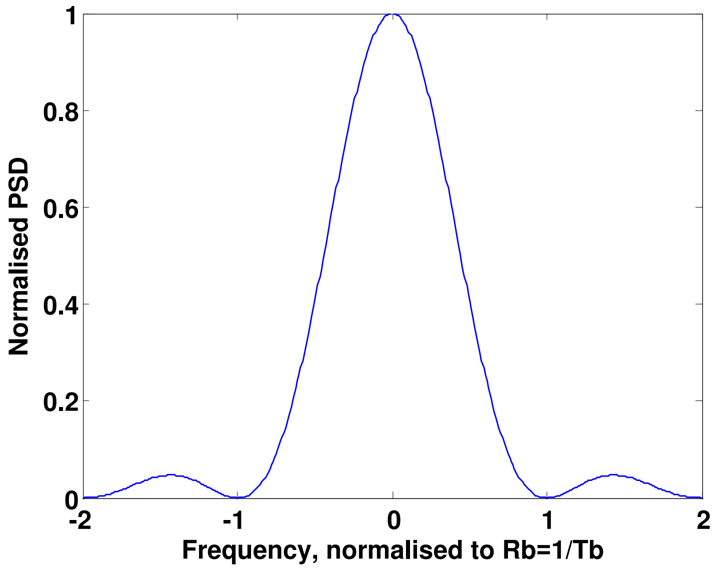


Figure 13. PSD of bipolar NRZ

Most communications systems choose a bandwidth of approximately 70% of the bit-rate. This factor is called the Personick factor in the field of Optical Communications.

In optical systems unipolar NRZ is used on the fibre (why?) although the electrical receivers use AC coupling to remove the average value.

Advantages

- Simple to implement .
- Bandwidth efficient.
- Highly suited to optical communications.

Disadvantages

- No timing information in the signal, so finding where to make the decisions can be complicated.
- Code is not intrinsically 'balanced'. A balanced code has no net average value-Most communications systems have no DC path-for

reasons of electrical design and performance, so a long stream of one symbol will begin to look like a DC level, and the output of the receiver will start to drop when this waveform is received. In practice protocols (line coding) restrict this to ensure the system works.

- In addition a lot of signal energy at low frequency, so if too much of this is cut-off then low frequencies not properly represented and the baseline of the signal changes with time (Baseline-wander). This is a function of the bit-rate relative to the AC coupling high-pass filter cut-off frequency

Used in

Fibre systems

5.4. Return to Zero Signaling

Bandwidth

Figure 14 shows the scheme. Levels are the same as for NRZ, but a one bit returns to zero within the bit period T_b -normally at half this time.

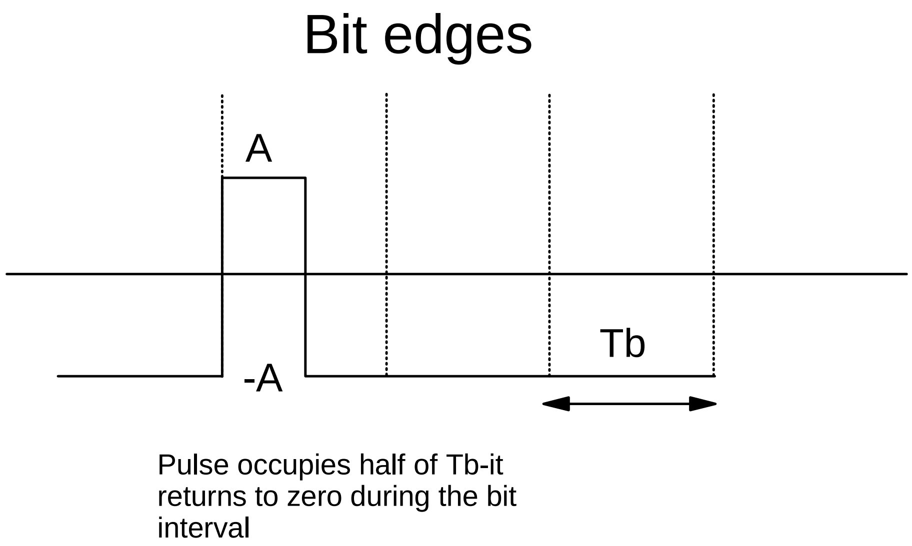


Figure 14. RZ data communication waveform

PSD is given by

$$S(f) = \frac{A^2 T_b}{16} \operatorname{sinc}^2\left(\frac{2\pi f T_b}{4}\right) + \frac{\pi A^2}{8} \sum_{n=-\infty}^{n=\infty} \operatorname{sinc}^2(n\pi/2)\delta[f - n/T_b]$$

Note that the sinc function is twice as wide as in the case of RZ and there are also delta functions representing specific frequencies where there is always energy present.

Figure 15 shows the PSD.

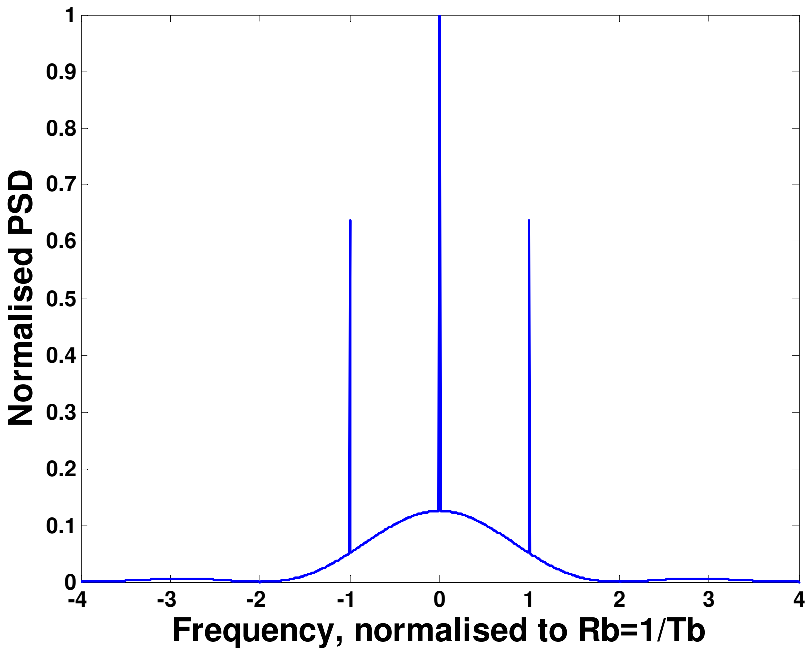


Figure 15. PSD of RZ waveform

The Probability of Error of unipolar RZ is given by:

$$P_e = \frac{1}{2} erfc\left(\sqrt{\frac{E}{2N_o}}\right).$$

Advantages

- RZ has discrete harmonics in the signal which can make clock recovery .
- Has found favour in very high speed optical communications (40Gb/s) due to the properties of the optical fibre channel.
- Has a high peak signal to average power ratio (as there is only energy for 1/3 time on average).

Disadvantages

- Requires twice as much bandwidth as NRZ.
- Again an unbalanced code.
- As with NRZ we have a significant amount of power at low frequencies causing baseline wander in ac-coupled systems. Also there is no error detection capability and the bandwidth required is double that for the equivalent NRZ scheme.

Used in

High bit-rate fibre systems.

5.5. Manchester Coding

Figure 16 shows the waveform.

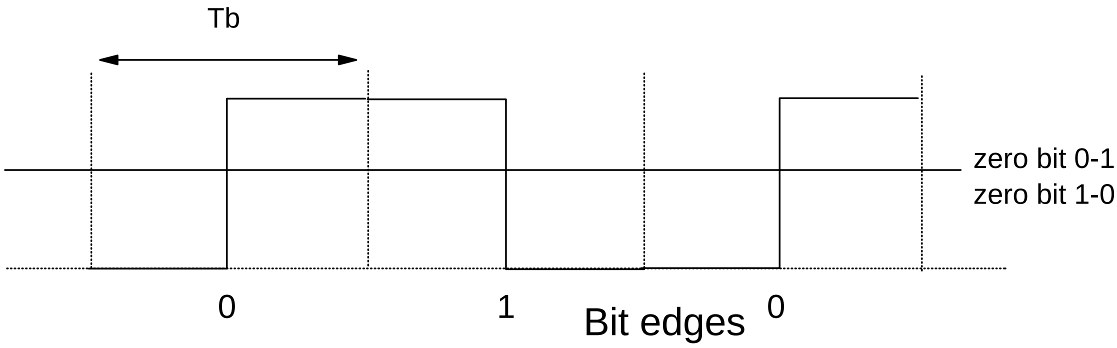


Figure 16. Manchester coded data stream

Bandwidth

The PSD is given by

$$S(f) = A^2 T_b \operatorname{sinc}^2\left(\frac{2\pi f T_b}{4}\right)\sin^2\left(\frac{2\pi T_b}{4}\right)$$

Figure 17shows the PSD.

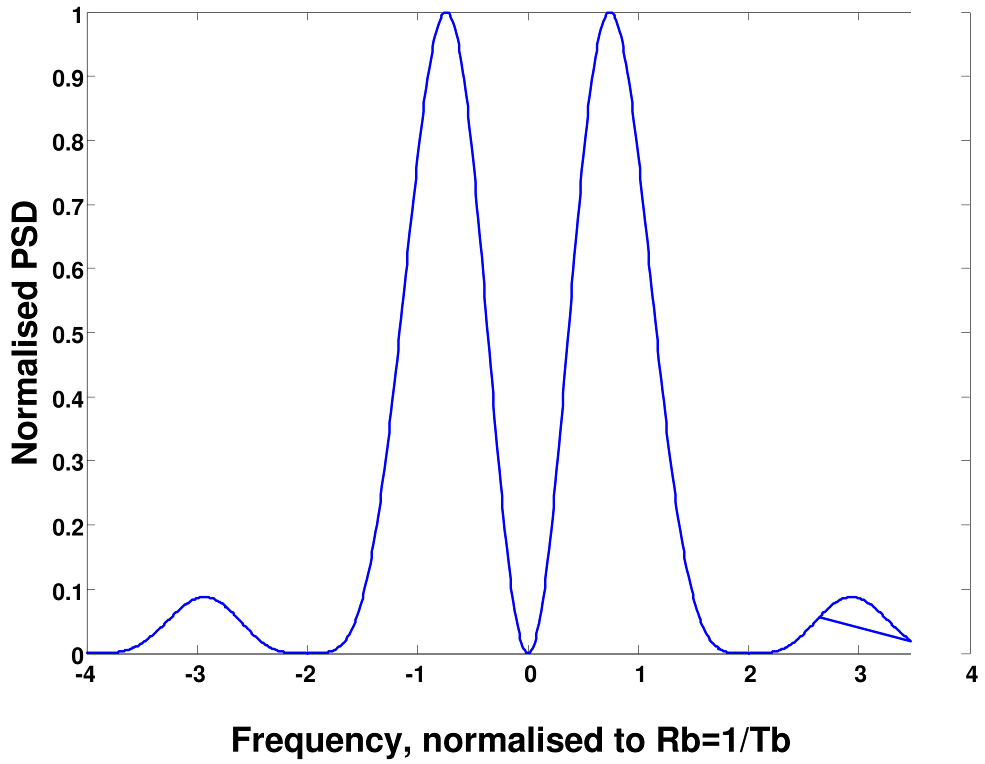


Figure 17. PSD of Manchester data

It can be seen Manchester has no DC content, and has discrete spectral components. The PSD has a maxima at $0.743/T_b$. The bandwidth occupancy is similar to RZ at the higher frequencies.

Advantages

- Manchester is 'transition' rich and there are straightforward clock recovery schemes.
- Scheme is balanced at a bit level-individual bits have no DC components.
- Code can do some error correction as it has memory.

Disadvantages

- Twice as much bandwidth required as NRZ.

- Logic required to code and decode (though relatively straightforward).

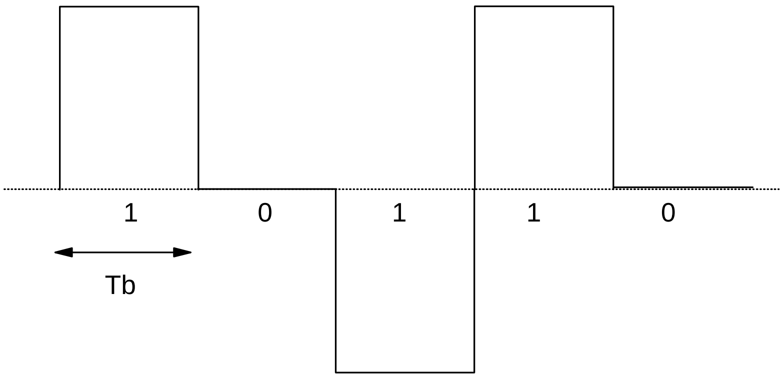
Used in

Infra-red optical links.

5.6. Alternative Mark Inversion (AMI)

Bandwidth

Figure 18 shows the scheme. A zero is coded as a zero and ones are coded alternately as +1 and -1.



Alternate occurences of one are inverted-zero is a 0 level

Figure 18. AMI waveform

The PSD is given by

$$S(f) = \frac{A^2 T_b}{4}\,\text{sinc}^2\left(\frac{2\pi T_b}{4}\right)\sin^2\left(\frac{2\pi T_b}{2}\right).$$

Figure 19 shows the PSD.

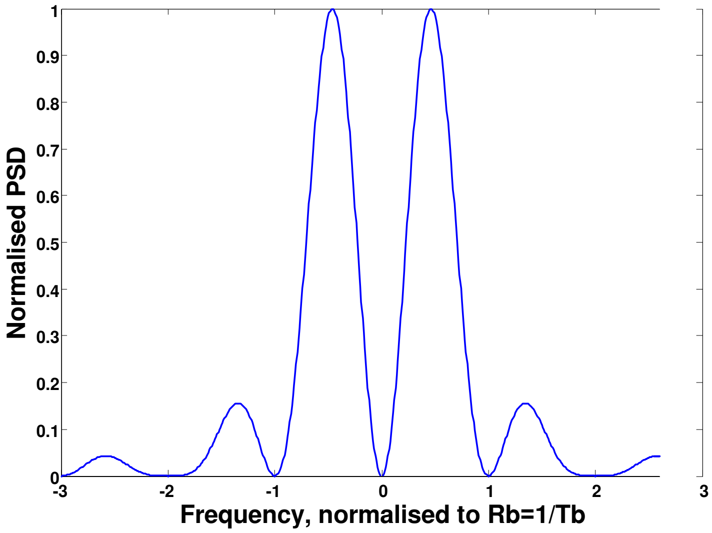


Figure 19. PSD of AMI waveform

Advantages

- Scheme is nearly DC balanced, usually over a small number of bits due to the inversion.
- Timing is recovered by squaring to obtain unipolar RZ and using the discrete harmonics.
- Code can do some error correction as it has memory.

Used in

Copper cable based digital transmission schemes

5.7. Other Formats

Code Mark Inversion (CMI).

Binary N Zero Substitution (BNZS).

High Density Bipolar N (HDBN).

.

5.8. Factors in choosing line codes:

- Bandwidth requirements
- DC content and the need for a DC path in the system
- Timing and receiver synchronization. Receiver must be able to make decisions about ones and zeros.

- Error detection capability. Useful, but more often provided by error correcting codes than using line codes.
- Transparency. Ability to transmit any data pattern. Many modern schemes have restrictions that are taken care of by extra coding before they reach the line coding stage.

6. Binary Bandpass Signalling

A **bandpass** channel is one where a specific range of frequencies (starting at a frequency away from 0Hz) is available for data transmission. Technique is to **modulate** the information onto a carrier frequency, or series of carrier frequencies in order to best use the available bandwidth.
Large number of different **modulation formats** of varying complexity that allow efficient use of the often scarce and expensive bandwidth available.
All RF (wireless) communications systems use some form of bandpass signalling, as the RF channel is away from 0Hz.

6.1. Amplitude Shift Keying (ASK)

Waveform

In ASK the amplitude of the carrier is modulated by two or more values associated with the original PCM code representation of the data. For the simple binary case, the carrier is usually modulated as an on-off keyed process (OOK) which essentially means a 1 is represented by the presence of the carrier and a 0 by the absence.

Bandwidth

ASK can be represented as a baseband unipolar data stream $s_b(t)$ (consisting of two levels $\{A,0\}$multiplied with a cosine function:

$$s_m(t) = s_b(t)\cos\omega_c t$$

Where ω_c is the carrier frequency. Figure 20 shows the relationship with the binary data.

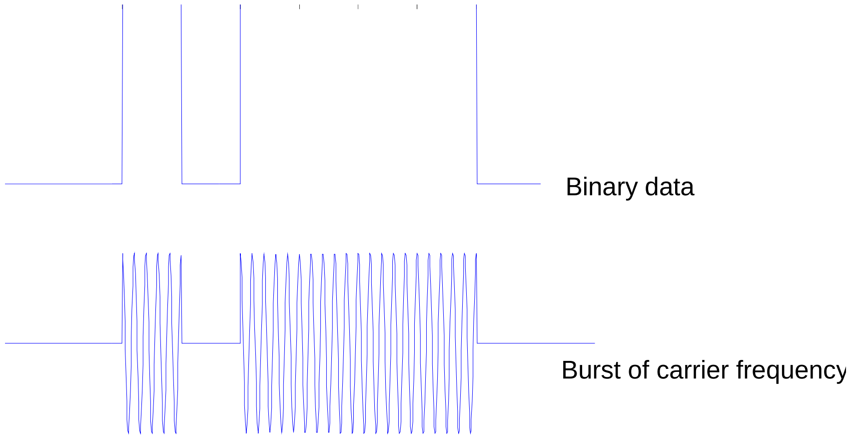


Figure 20. ASK waveform (fully modulated)

The power spectrum of this is the convolution of the Fourier transform of the cosine function and the PSD of the unipolar data stream.

Just a shifted version of the PSD of $s_b(t)$. Bandwidth required is approximately the distance between the two first minima of the sinc squared PSD. This is approximately twice that required for the baseband unipolar scheme (why?). **Figure 21** shows the PSD.

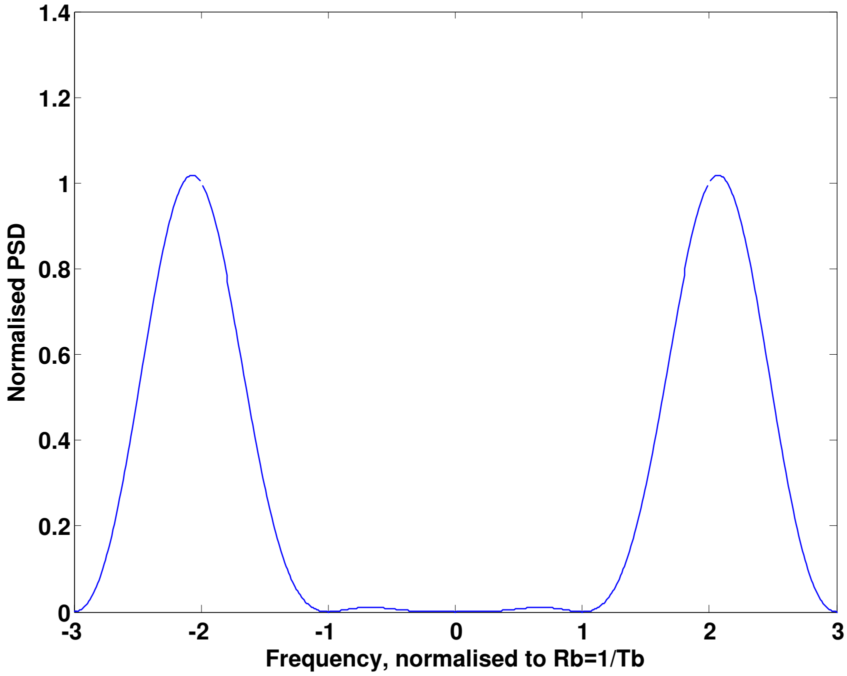
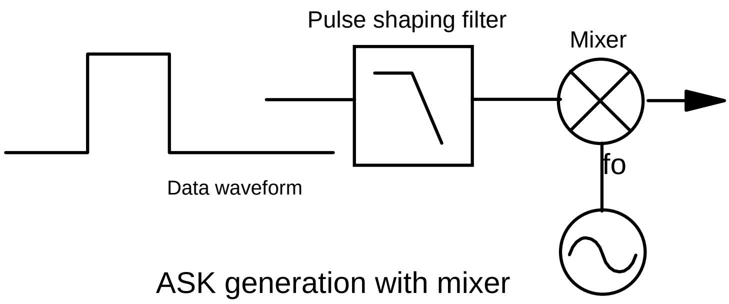


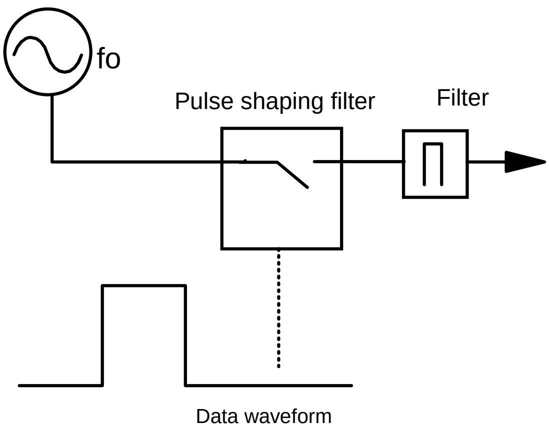
Figure 21. PSD of ASK where carrier frequency= 2*Rb

Generation

Generate ASK by two main processes, that is mixing and switching. Figure 22 shows schematics of the process.



ASK generation with mixer



ASK generation with switch

Figure 22. ASK generation

6.2. Mixer

- Multiplication of the baseband waveform takes place in an analogue mixer.
- Achieved by using a double balanced mixer (DBN) or a Gilbert Cell mixer.
- Baseband data can be pulse shaped and the filtered waveform is used to directly drive the IF port of the double balanced mixer.
- Baseband data preprocessing achieved using a DSP- possible up to 100MHz or so.

6.3. Switch

- A single pole double throw (SPDT) such as an analogue multiplexer, RF switch or PIN diode can be used to switch between the carrier oscillator and ground (no carrier).
- Note this is a digital process, so pulse shaping hard.

Mathematics for both processes is equivalent.

6.4. Detection

Incoherent detection:

- Bandpass filter and power detector will detect presence of signal power from data.
- Poor noise rejection

Coherent detection.

For coherent detection in the presence of AWGN, the input to the detector will be :

$$V_{in}(t) = s_b(t)\cos(2\pi f_c t) + x(t)\cos(2\pi f_c t) + y(t)\sin(2\pi f_c t)$$

This uses a narrowband representation of the AWGN. $x(t)$ and $y(t)$ are both Gaussian zero mean random variables, so that the $x(t)\cos()$ and $y(t)\sin()$ together create a noise voltage with random phase and amplitude. This is a quadrature representation.

For coherent detection multiply with a locally generated copy of the carrier

Assuming perfect synchronisation this gives

$$V_{out}(t) = \frac{1}{2}h(t) + \frac{1}{2}x(t).$$

Can now use baseband analysis. If data bits are equiprobable and the detection threshold is $A/2$, and the variance of $x(t)$ is σ^2 then the probability of error is

$$P_e = \frac{1}{2}\left\{1 - erf\left[A/(2\sqrt{2}\sigma\right]\right\}.$$

Converting this to the average energy per bit

$$E_b = \frac{A^2}{2}.$$

And the single sided noise power density $N_o = \sigma^2$

$$P_e = \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{2N_o}}\right).$$

6.5. *Phase Shift Keying (PSK)*

Waveform

In Phase shift keying schemes the phase of the carrier is switched between two levels corresponding to a unipolar baseband data input. A common scheme is to switch the carrier between 0 and π radians – known as phase reversal keying (PRK)

The signal waveform can be expressed as:

$$s_1(t) = A\sin(2\pi f_o t)$$

$$s_o(t) = A\sin(2\pi f_o t + \Delta\theta) = -A\sin(2\pi f_o t)$$

For the case where $\Delta\theta = \pi$.

If $\Delta\theta \neq \pi$ then can decompose to

$$s_o(t) = A\sin(2\pi f_o t)\cos(\Delta\theta) - A\cos(2\pi f_o t)\sin(\Delta\theta)$$

Figure 23 shows the phasor diagram.

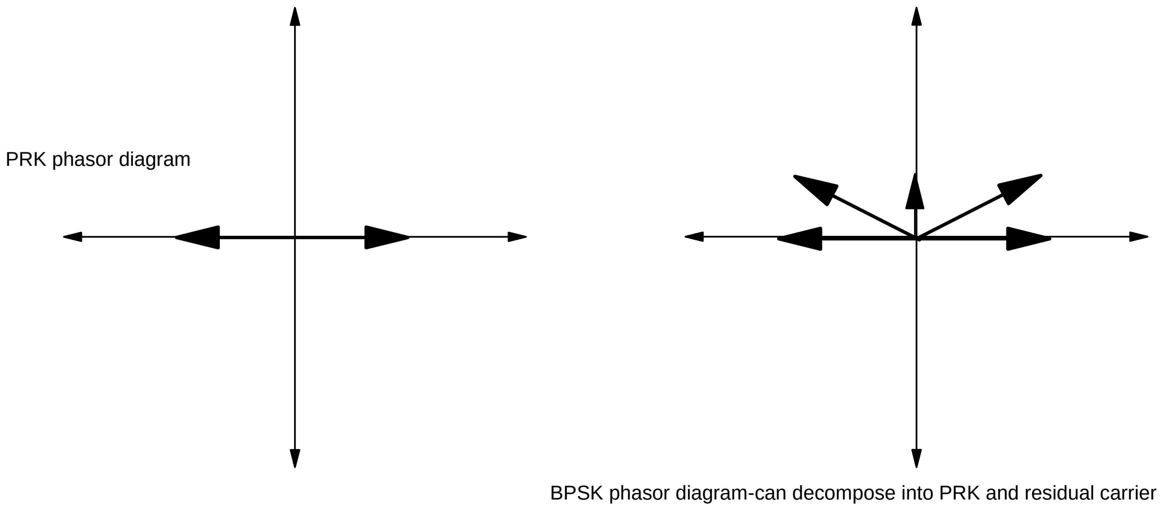


Figure 23. Phasor diagrams for BPSK and PRK

It can be seen this is equivalent to PRK and additional frequency component at the carrier frequency. This more general scheme is Binary Phase Shift Keying (BPSK)

Advantage that carrier can be used to lock coherent local oscillator at the receiver.

Bandwidth

Figure 24 shows the waveform for PRK.

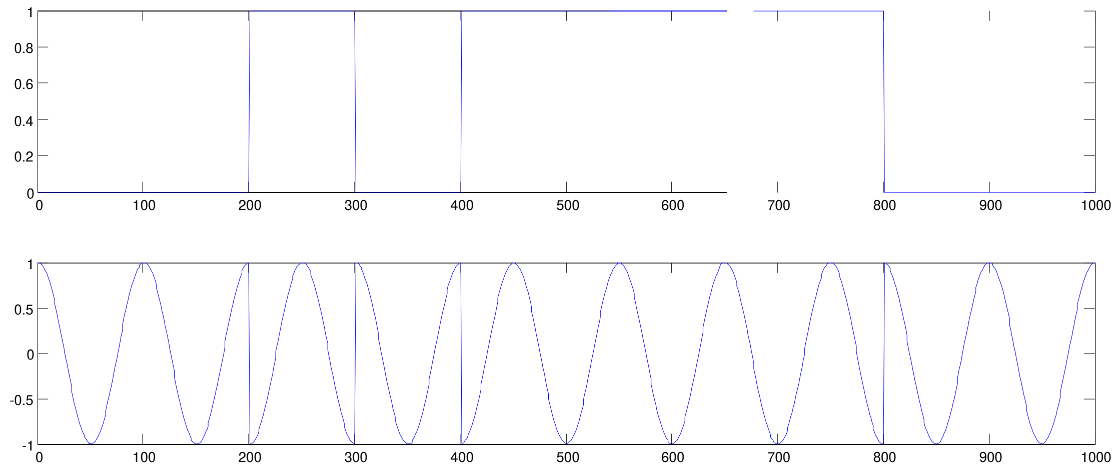


Figure 24. Binary PSK waveform

Power spectrum for PRK is the same as ASK. (why?). Bandwidth occupancy is therefore similar. Figure 25 shows the spectrum.

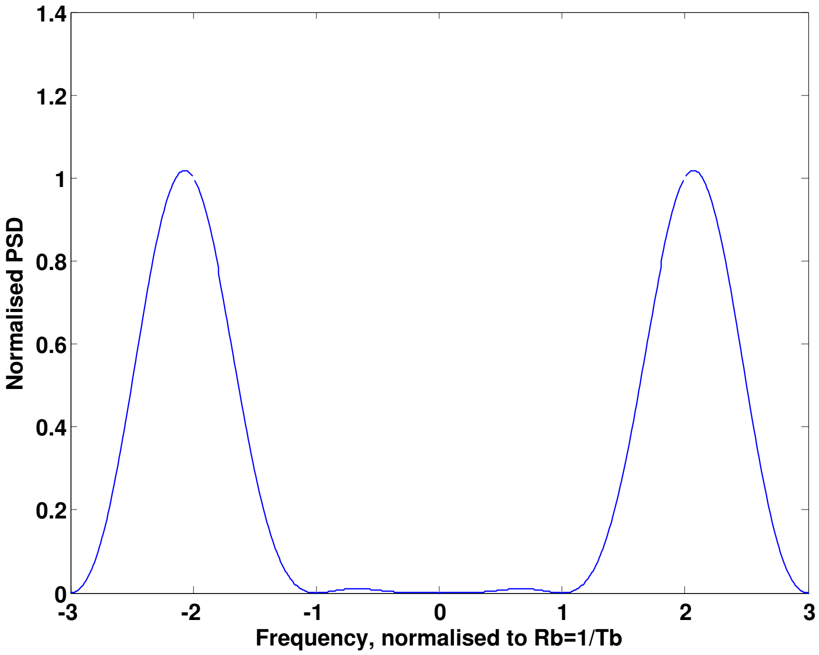


Figure 25. PSD for PRK

Generation

The simplest means of generating a binary Phase Shift Keyed (PSK) signal is to switch between the carrier oscillator signal and a 180 degree phase shifted version of itself.

However, accurate pulse shaping is required, then linear multiplication must be resorted to after pulse shaping at base band.

Just as for the bipolar vs unipolar baseband waveforms this scheme requires lower average power to send the information with the same BER as ASK.

Using similar arguments as for ASK BER is

$$P_e = \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{N_o}}\right).$$

Detection

Carrier recovery for BPSK can be accomplished using a squaring loop. The incoming PSK modulated waveform is of the form:

$$V_{in}(t) = \cos\left(\omega t + \{0,180^o\}\right)$$

Squaring the signal yields a continuous component at $2f_c$ (as the phase doubling makes both phase states identical), and DC components that can be filtered.

$$V_{in}^2(t) = \left[\cos\left(\omega t + \{0,180^o\}\right)\right]^2 = \cos\left(2\omega t + \{0,360^o\}\right)$$

A phase locked loop can be used to lock to this signal (using a 'divide by two' design (Glover and Grant p384))

Resulting wave could either be in phase with what was sent as a 'one' or what was sent as a 'zero'. Need a training sequence or preamble where the output is known to resolve this phase ambiguity.

Alternatively a Costas loop (see Glover and Grant) circuit can be used to achieve carrier recovery.

6.6. *Frequency shift keying*

Figure 26 shows the FSK waveform

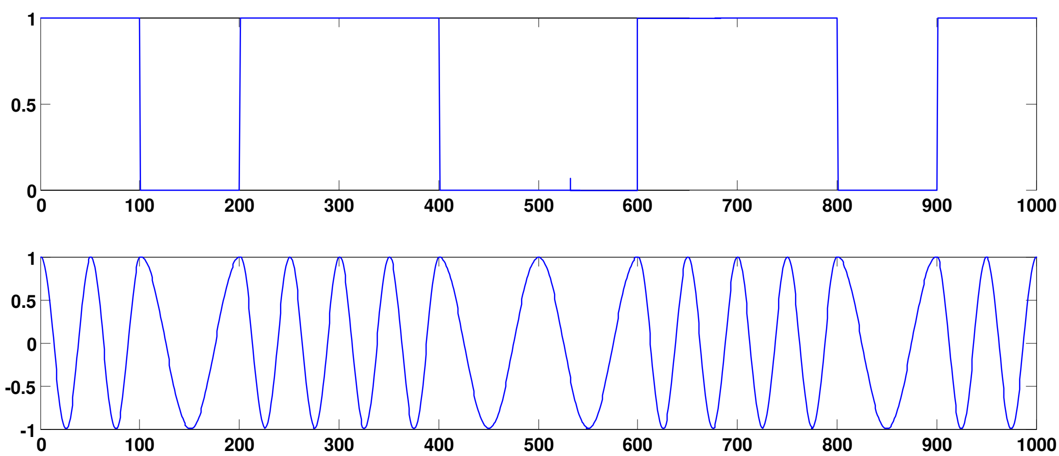


Figure 26. FSK waveform

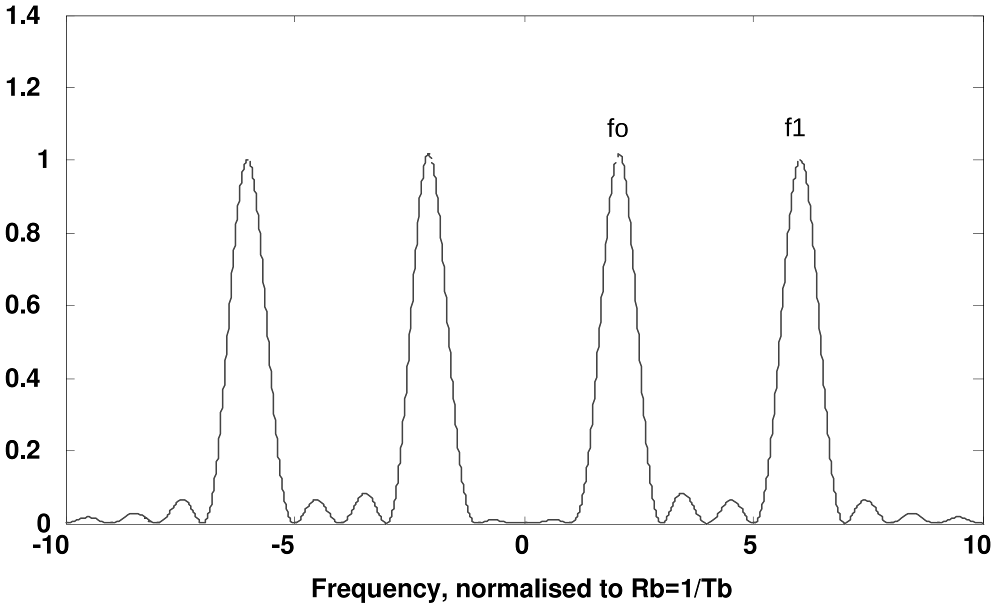


Figure 27 shows the PSD of FSK.

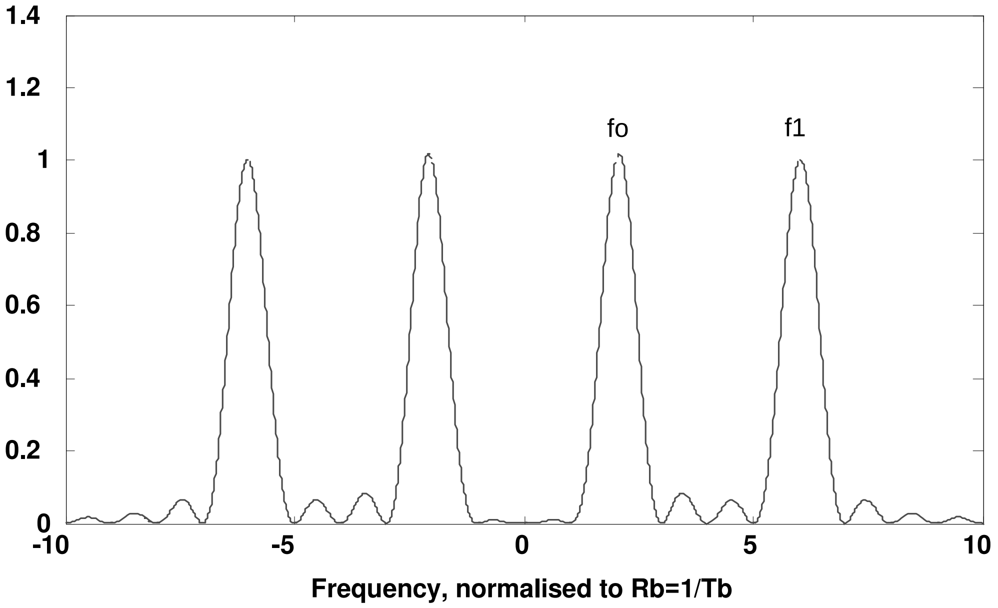


Figure 27. FSK PSD

Waveform

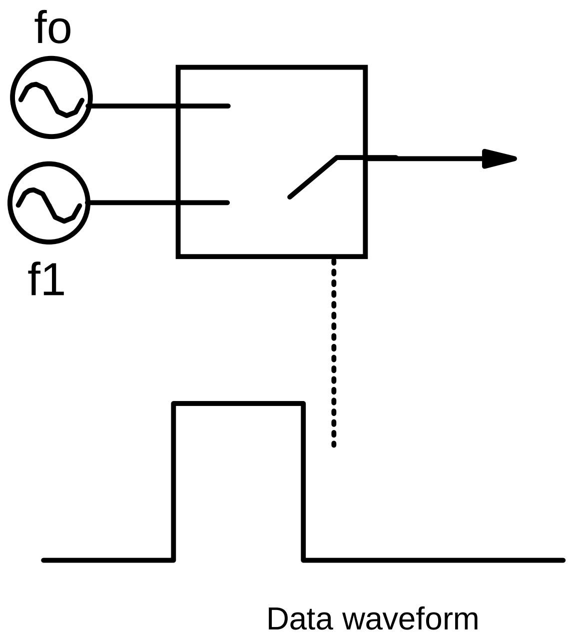


Figure 28. Generation of FSK

Figure 28 shows the generation of FSK. Instantaneous carrier frequency is switched between two or more levels corresponding to the baseband code.

Binary FSK (BFSK) can be considered to be the sum of two ASK signals with different carrier frequencies (see diagram). For the binary case , the signal waveform for a "1", $s_1(t)$ and that for a "0" , $s_0(t)$ can be expressed as:

$$s(t) = A\sin[2\pi f_o t + p(t)2\pi\Delta ft]$$

where $p(t)$ is the binary switching function $\{-1,1\}$ and $2\Delta f$ is the frequency separation between the two symbol states.

Bandwidth

Data bandwidth is the width of the sinc-squared pulse added to the carrier spacing. In orthogonal FSK the carrier spacing Δf_c is set to be equal to the symbol rate R_s so the first minima of the PSD overlaps aligns with the maximum of the next one. This means there is no overlap between the two spectra. More than two frequencies can be used in FSK, and there is an optimum bandwidth efficiency for this scheme.

Generation

FSK can be generated by a voltage controlled oscillator (VCO) or a switch between two separate oscillators. The problem with switching is that there is a likelihood of indeterminate phase discontinuities which give rise to out of band signals. In the other approach we can apply the baseband waveform directly to a VCO (the output frequency is than proportional the applied input voltage) and the transitions between symbols is continuous. By extension this approach can be used in multiple level FSK in which we have a range of baseband signal voltages representing our symbol set, and this is used to drive the VCO.

Detection

For the matched filter detection scheme (synchronous, coherent) the probability of error is given by (Glover and Grant P376):

$$P_e = \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{2N_o}}\right)$$

On an average energy per bit basis, the error probabilities for coherent ASK and FSK are the same.

However, there is a difference:

Assume that the FSK detector consists of two coherent detector chains with ideal bandpass characteristics, and use the quadrature representation of noise. The output voltage at each detector will be:

$$V_1(t) = h_1(t)\cos(2\pi f_1 t) + x_1(t)\cos(2\pi f_1 t) + y_1(t)\sin(2\pi f_1 t)$$

and

$$V_0(t) = h_0(t)\cos(2\pi f_0 t) + x_0(t)\cos(2\pi f_0 t) + y_0(t)\sin(2\pi f_0 t)$$

where $h_1(t)$ is the logical compliment of $h_o(t)$ and $x_1(t)$ and $x_o(t)$ are independent random variables, assuming that the pass band of the predetection filters do not overlap.

The output of the composite detector will then be :

$$V_{out}(t) = (\pm A) + [x_1(t) - x_0(t)]$$

Since $x_1(t)$ and $x_o(t)$ are independent Gaussian random variables, the term$\left[x_1(t) - x_0(t)\right]$ [$x_1(t)$ - $x_0(t)$] will also be Gaussian with a variance equal to the sum of the original variances. The noise variance at the output of the detector will thus be $2\sigma^2$ which gives an effective peak signal to rms noise ratio of $\frac{2A}{\sqrt{2}\sigma}$. This means that if FSK and PSK transmissions are to have the same error probability, then FSK requires a SNR which is 3dB above that required for PSK.

6.7. Differential Phase Shift Keying (DPSK)

Waveform

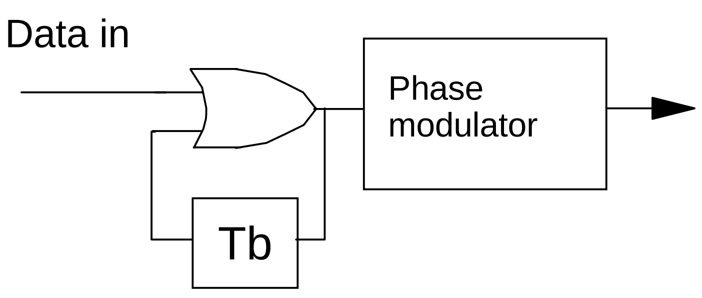


Figure 29. DPSK generation

Figure 29 shows how DPSK is generated. Data is differentially encoded into the phase of the carrier, using the difference between successive bits. This allows detection without a phase reference, as the reference is generated at the receiver using a delayed version of the received signal. The transmitted waveform is of the form

$$s(t) = A\cos\left(2\pi f_o t + \phi(t)\right)$$

The *k*^th^ bit of the differential code d_k can be generated from:

$$d_k = b_k \oplus d_{k-1}$$

where b_k is the *k*^th^ bit of the data, and $\oplus$ is exclusive OR. d_k is then transmitted.

b_k		1	1	0	1	0	1	1	0	0
d_k	0	1	0	0	1	1	0	1	1	1
$\phi(t)$	π	0	π	π	0	0	π	0	0	0

Detection

Figure 30 shows the receiver.

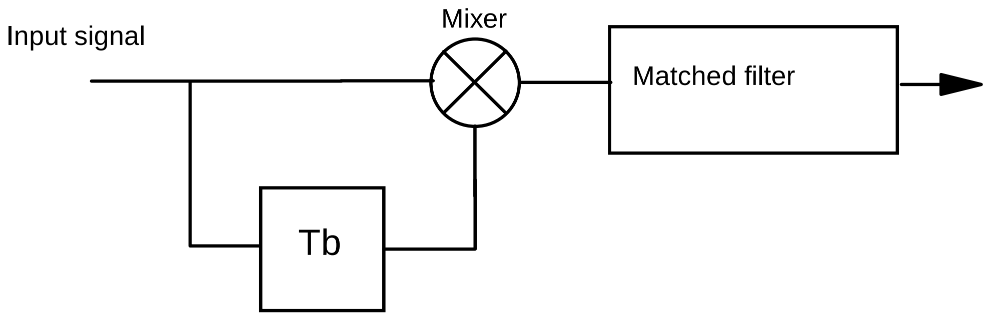


Figure 30. DPSK receiver

Achieved by splitting the received (bandpass) signal and using a copy that is delayed by one bit as a reference. The incoming signal is mixed with a version of itself that has been delayed by one bit period. The product of two *cos* terms yields the sum and difference, thus the output of the mixer is of the form $\cos\phi(t)$. The demodulation process is illustrated in the table below.

$\phi(t)$	π	0	π	π	0	0	π	0	0	0
$\cos(\phi_k - \phi_{k-1})$		-1	-1	+1	-1	+1	-1	-1	+1	+1
Decoded signal		1	1	0	1	0	1	1	0	0

Performance

- Bit determination relies on the current and previous bits
- Pairs of errors propagate, and reference is noisy, as no locally generated carrier

The probability of a bit error P_e for DPSK is (see Proakis P 274) CHECK 2

$$P_e = \frac{1}{2}\exp\left(\frac{-E_b}{N_o}\right)$$

Advantages and disadvantages

- Simpler than PSK. Compared to PSK systems, for P_e<10-4, the DPSK system suffers a power penalty of $\leq$ 1dB, making it a popular choice in many applications.
- Need a preamble as first few bits required to get system 'going'.

6.8. Comparison of Binary Modulation Schemes

PSK requires the least power and incoherent ASK the most. Note that although the difference between some characteristics is only on the order of 1 to 2 dB, remember that a change in signal power of only ~1dB results in a change in the order of magnitude of P_e (over ranges of practical interest).

	ASK	FSK	PSK
Ease of implementation	Noncoherent ASK receivers are relatively simple to construct (Coherent ASK unpopular as only slightly better than incoherent)	Non-coherent FSK receivers are simple to construct and popular for low speed data transmission	Complex. DPSK simplifies receiver design due to lack of local oscillator
	No power is transmitted when no data is sent (for the OOK case)	FSK system decision threshold operates symmetrically about a zero level regardless of received signal strength.	Least amount of power required for a given P_e *(for coherent PSK)*
Comments	Often found in applications such as short range telemetry systems.		A good compromise is DPSK since it retains fairly good power requirements whilst receiver complexity is much lower
	Disadvantage of ASK is that the threshold level must be adjusted with changes in received signal level- thus automatic gain control (AGC) is required.	Bandwidth requirements for coherent FSK can be made as small as desired by controlling Δf but for $2\Delta f < 0.5$ a signal to noise penalty is incurred. Coherent FSK schemes typically have a bandwidth occupancy which equal or are slightly greater than those required for ASK/PSK	Synchronous detection is required – making the receiver complex and hence more expensive.
Prob of error	Coherent ASK $P_e = \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{2N_o}}\right)$	$P_e = \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{2N_o}}\right)$	DPSK $P_e = \frac{1}{2}\exp\left(\frac{-E_b}{N_o}\right)$ PRK $P_e = \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{N_o}}\right)$
Bandwidth (between first zeros of PSD)	$2R_b$	$2R_b + \Delta f$	$2R_b$