UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL ATOMIC ENERGY AGENCY
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

HANDOUT-3

ICTP - URSI - ITU/BDT WORKSHOP ON THE USE OF RADIO FOR DIGITAL COMMUNICATIONS IN DEVELOPING COUNTRIES

(17 - 28 February, 1997)

"Principles of Digital Modulation: Support Notes"

M.P. Fitton
University of Bristol
Bristol
UNITED KINGDOM

MAIN BUILDING STRADA COSTIERA,11 TEL. 2240111 TELEFAX 224163 TELEX 460392 ADRIATICO GUEST HOUSE VIA GRIGNANO,9 TEL. 224241 TELEFAX 224531 TELEX 460449
MICROPROCESSOR LAB. VIA BEIRUT,31 TEL. 2249911 TELEFAX 224463 TELEX 460392 GALILEO GUEST HOUSE VIA BEIRUT,7 TEL. 22401 TELEFAX 224559 TELEX 460392

Principles of Digital Modulation: Support notes

Mike Fitton
UNIVERSITY OF BRISTOL

These notes are intended as support material for the some elements of the "Principles of Digital Modulation" lecture course. They do not necessarily have to be read before the course.

Relevant Modulation Schemes

Quadrature Phase Shift Keying (QPSK)

Figure 1(i) shows the block diagram for a *QPSK* modulator. The input data is first divided into odd and even data channels with each channel operating at a rate corresponding to half the overall bit rate. These two NRZ data streams are then modulated onto the *in-phase* (I) and *quadrature* (Q) components of the carrier frequency using balanced modulators. The two transmissions can now be separated at the transmitter since, although they occupy the same spectral range, they are transmitted on uncorrelated quadrature components.

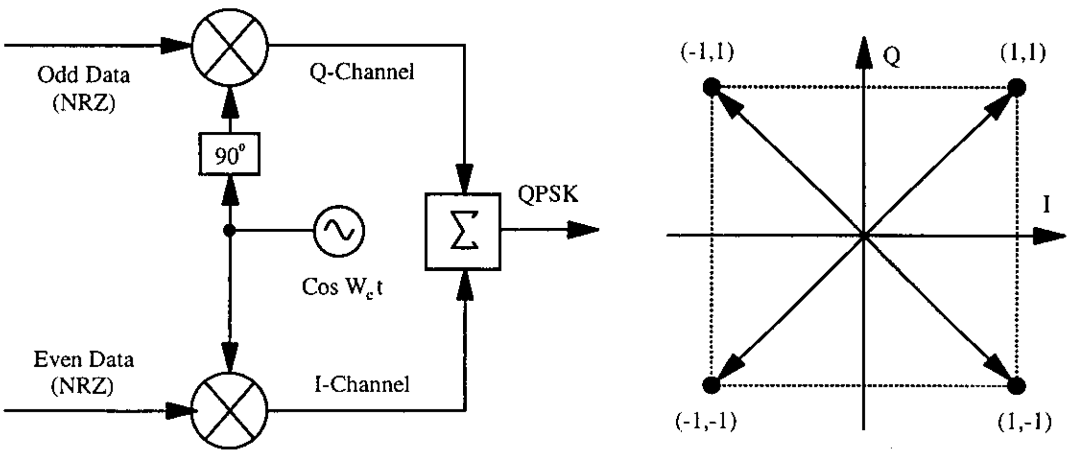
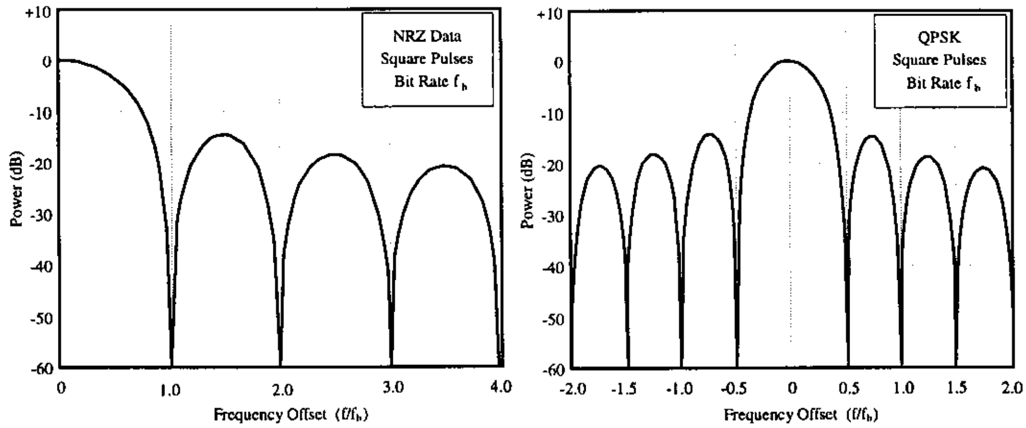


Figure 1(i) QPSK Modulator



(i) NRZ Data

(ii) QPSK

to use one of a family of *raised cosine filters* which closely approximate the performance of a Nyquist '*brick wall*' filter.

There are three distinct types of QPSK, usually referred to *as non-offset QPSK, Offset (or staggered) QPSK and π/4-QPSK*. The difference in these systems lies in the transients that can occur at the end of each bit interval. Normally the odd data channel takes new values on the odd bit transients while the even channel changes on the even transients. As can be seen in the right-hand diagram of figure 2, the even channel takes new values at transients 0,2,4,6 etc. while the odd channel updates its value on the odd transients, i.e. 1,3,5,7 etc. This results in a situation where a transient on the odd channel never occurs together with a transient on the even channel. Due to the effective offset between these channels, this form of modulation is known as Offset QPSK (usually abbreviated to OQPSK). For this modulation scheme, after each bit interval, the transmitted signal phase will change by a multiple of 90 degrees rather than by 180 degrees as in BPSK (and non-offset QPSK). However, if either the I or the Q channel is delayed by one bit period, the transients will occur together and this results in non-offset QPSK, which is shown in the left-hand section of figure 2.

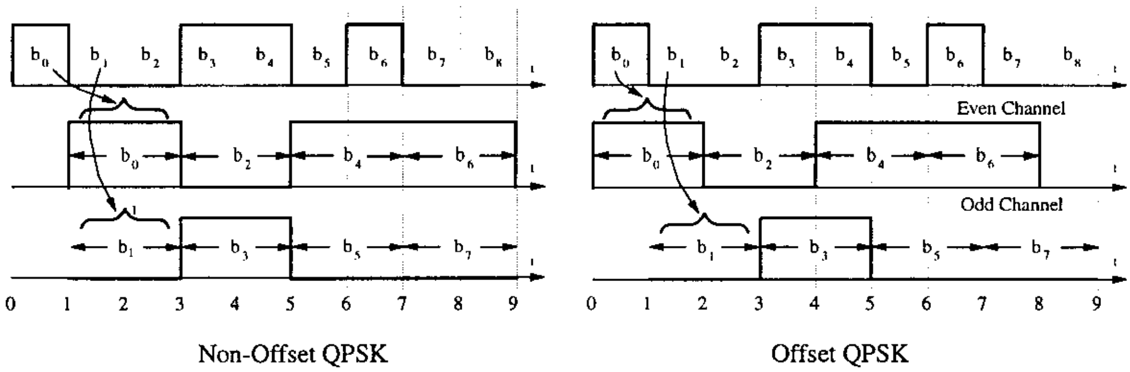


Figure 2: Offset and Non-Offset QPSK

In this configuration, phase changes of 180 degrees can occur for certain combinations of bit transient. The third variant, known as π/4-QPSK, makes use of two QPSK constellations, the second rotated by π/4 radians relative to the first. Even data bits are mapped into I and Q using the first constellation while odd data bits use the second rotated constellation diagram. This has the effect of generating eight possible phases at the receiver (rather than the more usual four), although these are grouped into two four-phase constellations, one for even symbols and one for odd symbols.

If coherent demodulaton is assumed, then the receiver needs to know whether the current symbol is odd or even, without this information it is not possible to hard limit the incoming symbols in the most efficient manner. This is one reason why $\pi/4$-QPSK is often received non-coherently using either differential or discriminator detection. For differential detection, four received phase differences are generated at the receiver while, with discriminator detection, four distinct amplitude levels are created. The various phase transients are summarised in figure 3.

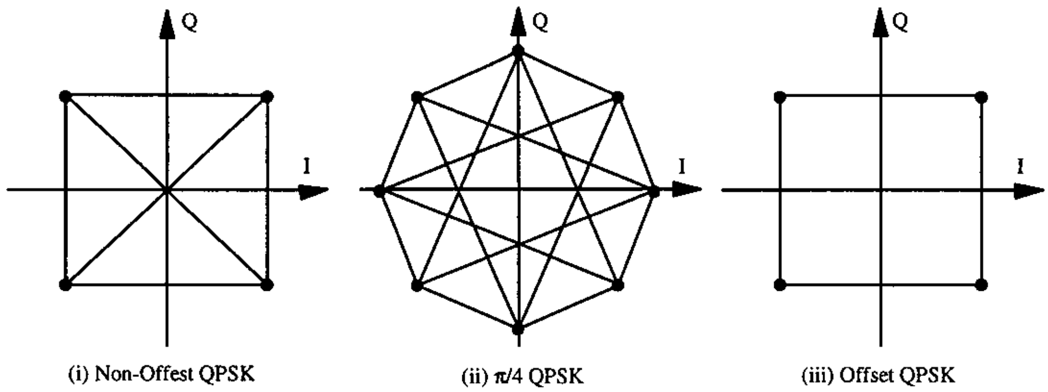


Figure 3: QPSK Constellations

MSK/GMSK (Gaussian Minimum Shift Keying)

Rather than jumping rapidly from one frequency to another, in MSK filtered transients are used to reduce the out-of-band spectral components. Although MSK is inherently an FSK modulation scheme, it is more easily analysed by observing its phase transients with time.

$$s(t) = A\sin(\omega_c t + \gamma(t))$$

Figure 4 shows all the possible paths that $\gamma(t)$ can take for the first eight bits of transmission, assuming $\gamma(0)$ to equal zero. These transients are very easy to analyse with the phase ramping *up* through 90 degrees for a binary one, and *down* 90 degrees for a binary zero.

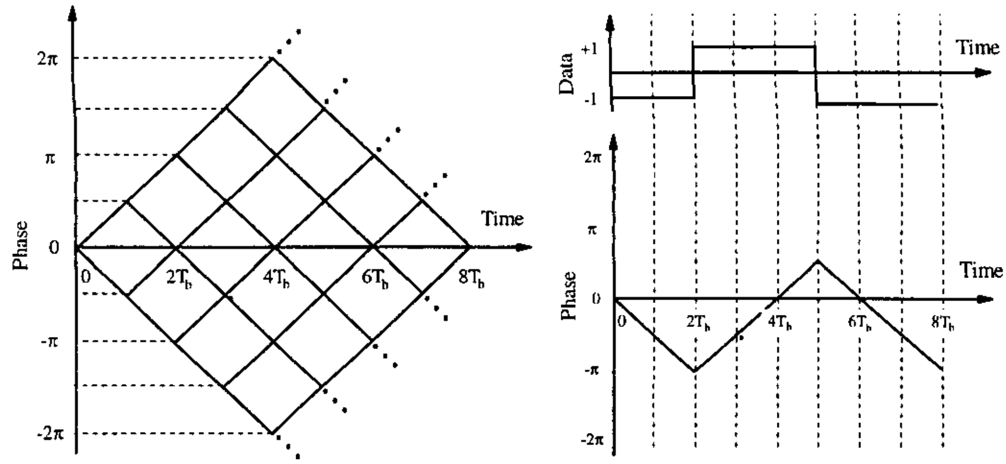
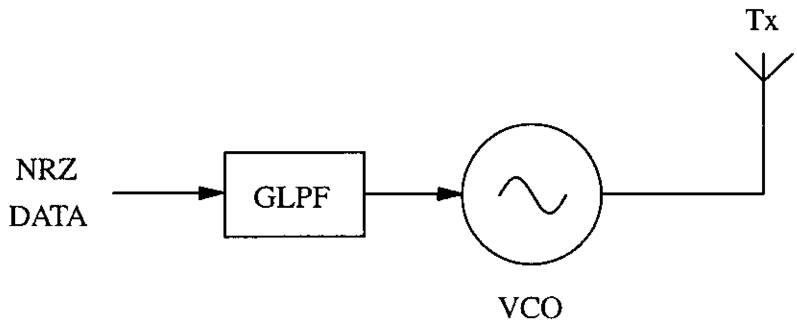


Figure 4: Phase Transients in MSK

Figure 5 shows the modulator required for the generation of either MSK or GMSK waveforms. For GMSK transmission, a *Gaussian pre-modulation baseband filter* is used to suppress the high frequency components in the data. The degree of out-of-band suppression is controlled by the tightness of this filter which, in turn, is governed by its *BT* product, where *B* represents the 3 dB filter bandwidth and *T* the system's bit duration.



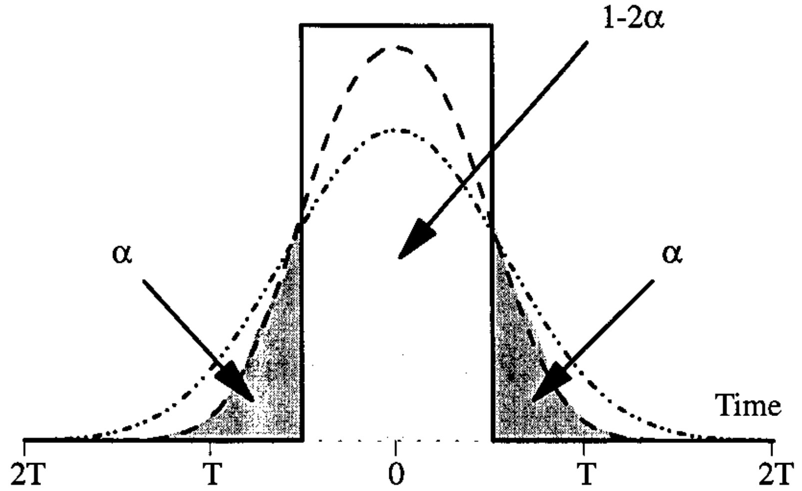


Figure 6: Gaussian Pulse Shapes & ISI

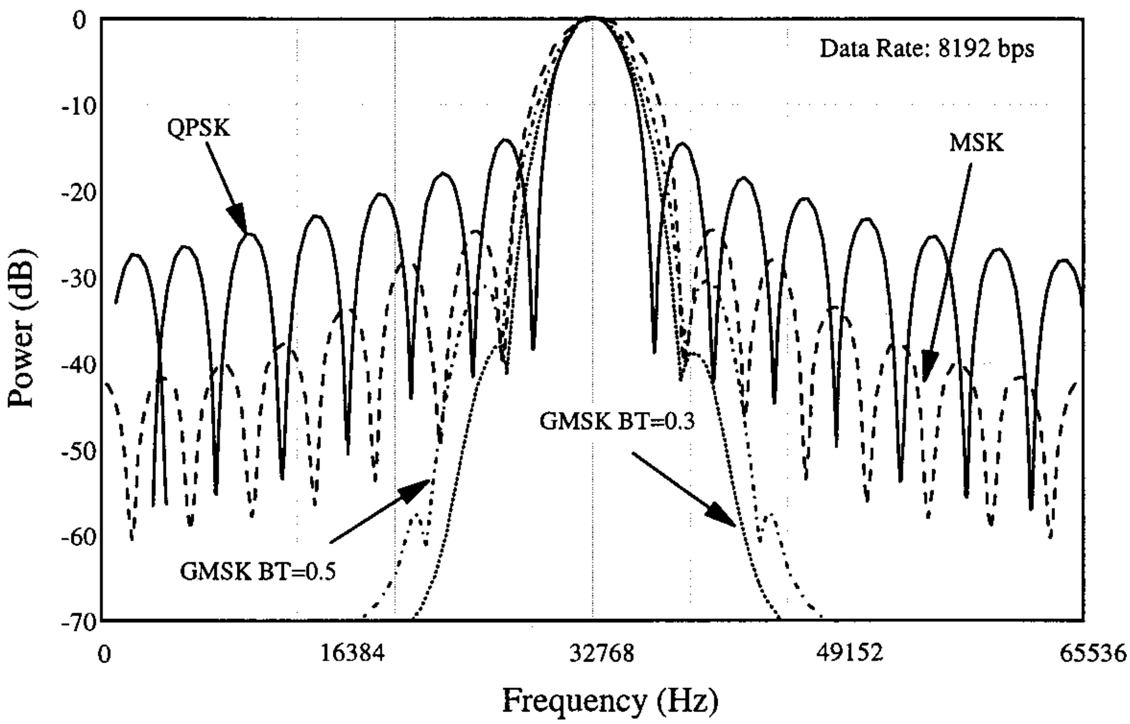


Figure 7: GMSK/MSK Spectrum

Benefits of GMSK

The main advantage of GMSK lies in the fact that a constant power envelope is maintained at all times. This, coupled with continuous phase, produces very good out-of-band frequency suppression which can then be maintained through non-linear amplifiers. Figure 7 shows the output power spectrum for QPSK and also GMSK for various values of *BT*. For unfiltered QPSK the sidebands are as high as -14 dB relative to the main lobe and this is obviously unacceptable due to *adjacent channel* interference. The spectrum for MSK shows a sizeable improvement relative to QPSK (due to the use of sinusoidal pulse weightings in I and Q) with sidebands attenuated to approximately -24 dB. However, using a Gaussian pre-modulation filter, the first sideband can be further reduced and all higher frequencies attenuated by more than 70 dB relative to the main lobe. The figure shows how the value of *BT* may be used to directly control the overall bandwidth efficiency of the system. The spectrum is more compact with a *BT* of 0.3 (relative to 0.5) and this also results in almost total suppression of the first set of sidebands. The drawback with such tight filtering is the ISI introduced into the system, which can seriously degrade the BER performance of the modem in a noisy channel.

Modulation Summary

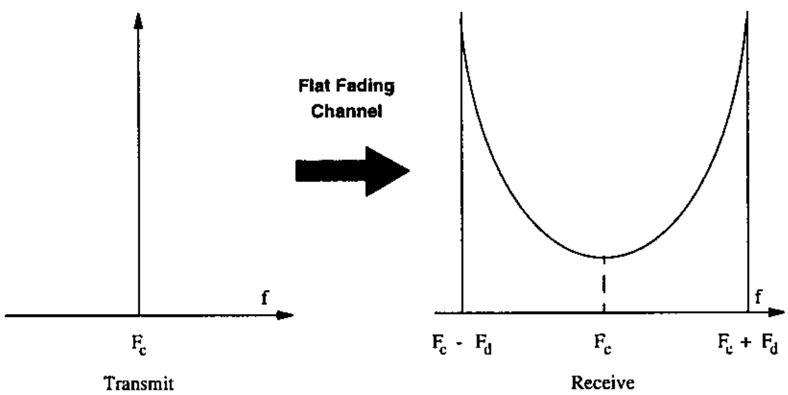
Modulation schemes fall into two distinct categories, those which exhibit a *constant power envelope* and those that require some degree of *linear amplification*. If constant envelope amplification is required, then MSK or GMSK represents an excellent compromise between power and spectral efficiency. However, it is possible to receive MSK and GMSK as if they were offset QPSK transmissions. This tends to blur the boundary between these modulation schemes and demonstrates that MSK is just a special case of QPSK exhibiting constant envelope.

If some degree of linearity is permitted in the transmitter then the choice of modulation increases. At present the trend seems to be towards differentially or discriminator detected root-raised cosine $\pi/4$-QPSK. This is currently being used in systems such as the European PMR *TETRA* standard, the interim American cellular standard (*D-AMPS or IS-54*) and the Japanese *Personal Handy Phone* (PHP). For satellite and microwave point to point communications, 16-QAM is already very popular and there is much interest, particularly in Japan, in applying these ideas to the mobile radio market.

The Impact of the Mobile Channel on Digital Communications

The Effect of Multipath Fading

The mobile receiver suffers from periodic phase shifts which change with time as the mobile moves, arising from the multipath fading of the channel. The rate of change of phase gives rise to a Doppler frequency which varies as a result of the changes in the mobile's speed and also the arrival angle of the rays. The resulting power spectral density for a single transmitted tone is shown below.



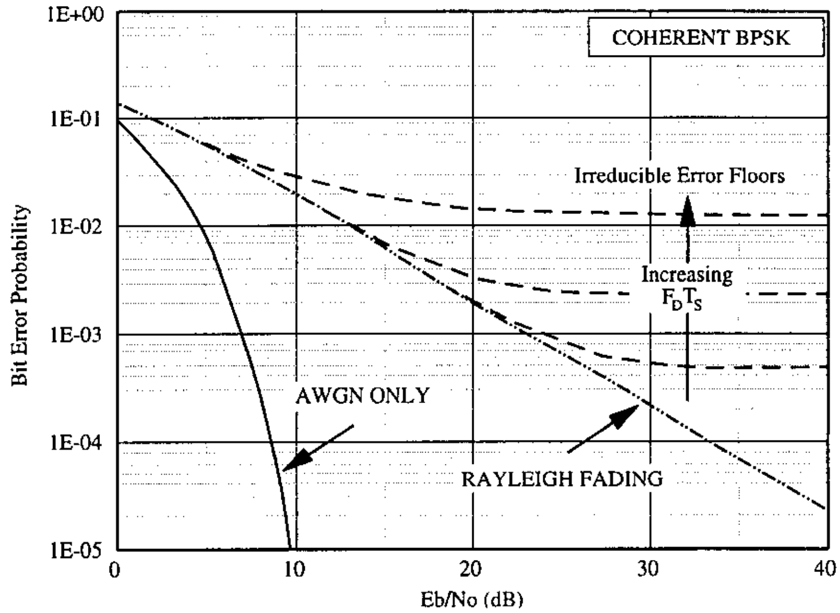
Power Spectal Density for Rayleigh Fading

This figure assumes the receiver to have an omni-directional gain pattern in the horizontal plane. The diagram shows that the received frequency is far more likely to take values in the region of $\pm f_D$ than values close to zero. This faded spectrum can be represented mathematically by the following equation:

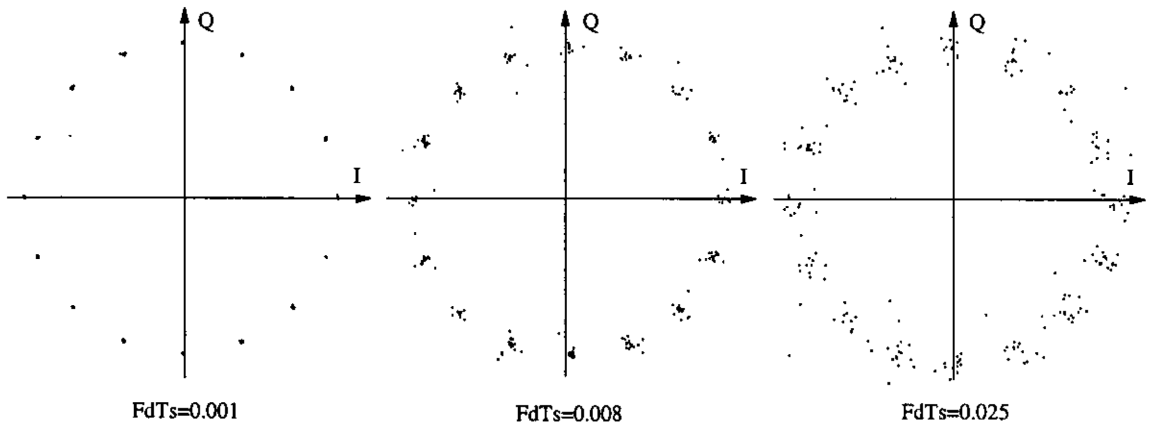
$$W(f) = \frac{\sigma_f^2}{\pi\sqrt{\left(f_D^2 + f^2\right)}} \quad \text{where} \quad |f| \le f_D$$

where σ_f^2 is the average power of the faded carrier and f_D the maximum Doppler frequency.

Assuming slow fading and perfect Automatic Gain Control (AGC) in the receiver, the observed bit error rate will depend only on the value of the received instantaneous signal strength. The diagram below shows the probability of error for coherent BPSK plotted against the energy per bit to noise power density (Eb/No) for an Additive White Gaussian Noise (AWGN) environment. For the fading environment the probability of error will be generated by averaging the error probabilities for the various values of Eb/No present at the receiver. Using this diagram it can be seen that to maintain an error probability of 1 in 10000 an Eb/No of greater than 8.5dB is required. For the Rayleigh fading case this value must be increased to approximately 34dB, which represents a considerable increase in the value required for the static case. For comparison the error probability for Rayleigh fading is also shown on the diagram together with the error floors produced by random FM.



The previous graph assumed that coherent detection had been used and that the signal was transmitted within the coherence bandwidth of the channel. Imperfect tracking of the random FM component (particularly during signal fades) prevents the required carrier phase from being generated and this can introduce symbol errors. As the fades get deeper the phase tends to flip 180 degrees thus producing errors. To demonstrate this problem the received 16PSK constellation diagram is shown below for various vehicle speeds.



16PSK Constellations for Various FdTs

The carrier recovery needs to track to an accuracy of 11.25 degrees (symbols are separated by 22.5 degrees). Alternatively, if differential detection is used, the phase difference in the channel between successive symbols must also remain below this threshold. At high speeds these conditions can no longer be maintained and this will result in the introduction of a short burst or errors. The overall bit error rate will depend on the modulation scheme, the data throughput and the number of fades encountered per second. The fading rate will also depend on the carrier frequency and the users speed. It is possible to combine this factors to produce a figure whose value is directly related to the error rate experienced in the channel. This figure, often referred to as the normalised bandwidth, is simply obtained by multiplying the Doppler frequency by the symbol period of the system.

In the above diagram the first value of FdTs represents a slow moving mobile and resulted in a symbol error rate below 1 error in 3000 (no noise was present in the receiver). The second diagram shows the received constellation for a larger value of FdTs, in this situation an irreducible error rate of 1 to 100 symbols was seen to arise. The constellation diagram reflects the increased variations in the channel with phase and amplitude jitter clearly seen to distort the received signal points. The third diagram shows the situation for an even higher value of normalised Doppler. The resulting error rate was now very high (greater than 1 in 10) with distortion making it difficult to distinguish individual symbol points.

The above errors are described as irreducible since, unlike those generated by noise, these cannot be lowered by simply increasing the output power of the transmitter. This type of error is obviously very restrictive since, for a particular application, only so much error can be tolerated. This then results in a maximum velocity being imposed on the mobile users.

QUESTION

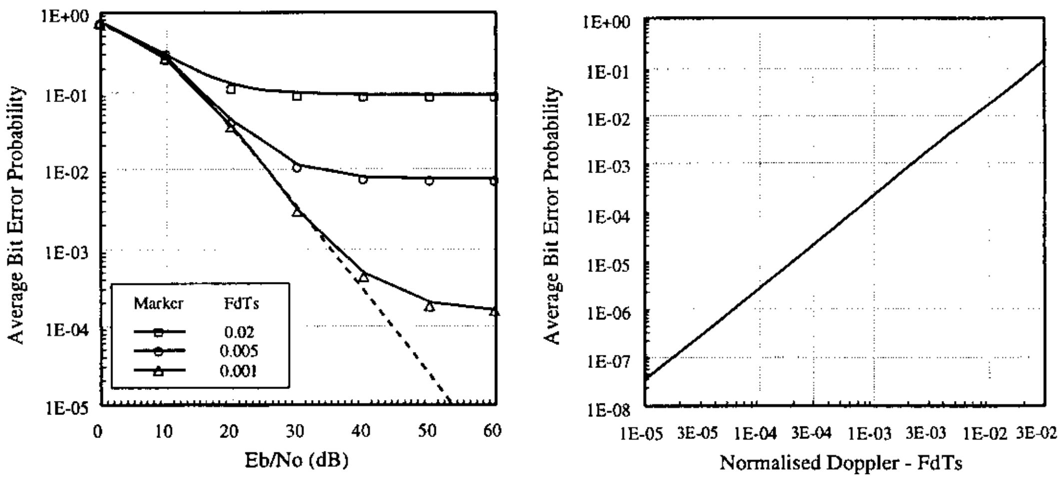
Using the following graph, calculate the maximum vehicle speed assuming a 1 in a 1000 error threshold and the following parameters:

(i) 150 MHz and 9.6Kb/s

(ii) 900 MHz and 72 kb/s

(iii) 5.1 GHz and 72 kb/s

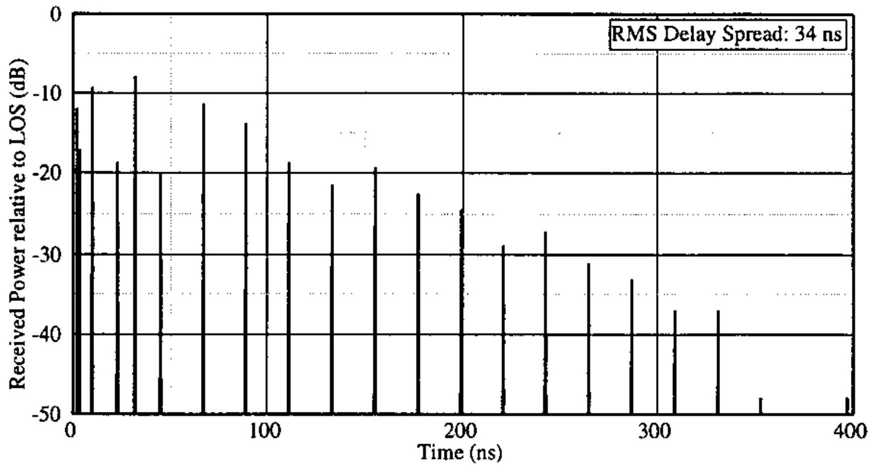
(iv) 5.1 GHz and 10 Mb/s



Irreducible Errors caused by Random FM

Delay Dispersion

The multipath signals arriving at the mobile receiver each take their own unique route and as such suffer from a particular time delay, attenuation anbd phase shift. This time delay may be different for each of the paths and the dispersion can result in a problem if high data rate digital systems are employed. The Power Delay Profile (PDP) is obtained by squaring the ampltide term for each path (to obtain power) and plotting its value against the time delay (which is usually normalised relative to the time of the first received ray). A fairly typical indoor PDP is shown below.



Typical Indoor Power Delay Profile

Usually the first rays to arrive have the largest amplitude and their magnitudes tend to fall away as their time delay increases (due to extra path loss and further reflections and diffractions). It is possible to quantify the degree of dispersion in a channel by evaluating *its Root Mean Square (RMS) delay spread.* This value can be calculated using the equation below, where α_k represents the received amplitude of the *k*-th ray after a time delay of τ_k seconds and τ_a represent the time for half the power to arrive.

$$\tau_a = \frac{\sum_{k=1}^{N} \tau_k \alpha_k^2}{\sum_{k=1}^{N} \alpha_k^2}, \qquad \tau_{rms} = \left[\frac{\sum_{k=1}^{N} [\tau_k - \tau_a]^2 \alpha_k^2}{\sum_{k=1}^{N} \alpha_k^2} \right]^{1/2}$$

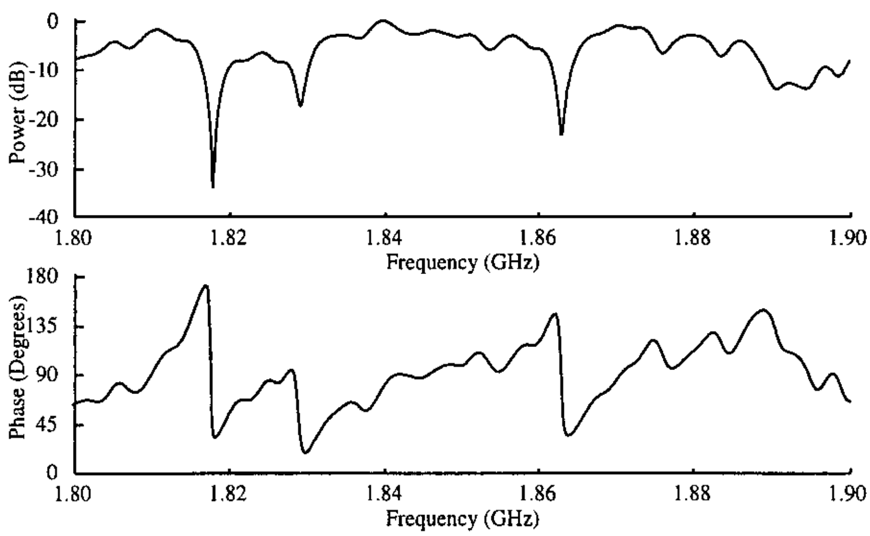
This value incorporates all the information from a power delay profile and expresses it as a single value whose magnitude allows direct comparisons between differing areas and environments. For digital transmission it is more common to quote this rms delay spread *normalised* to the transmission bit rate (R) or the bit period (T_b) as shown below.

As a rule of thumb, most systems generate a Bit Error Rate of around 1 in 1000 for a normalised delay spread in the region of 0.1.

$$d = \frac{\tau_{rms}}{\tau_n} = \tau_{rms} R$$

The maximum bandwidth that can be supported in a time dispersive or frequency selective channel is often referred to as the *coherence bandwidth.* This figure is loosely defined as the frequency separation required for the correlation of a single fading tone to fall below 0.5.

The diagram below shows the frequency domain for the power delay profile described earlier.

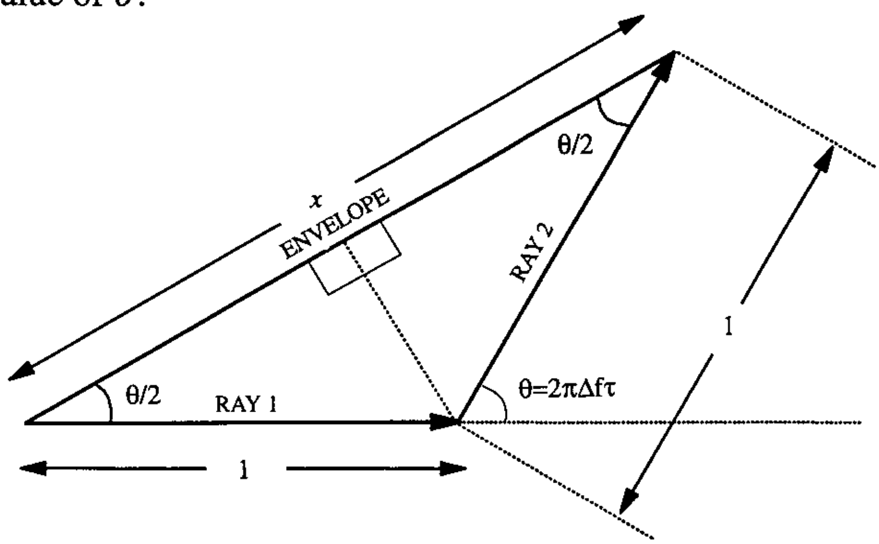


Frequency Selective Amplitude and Phase

From this plot the frequency selective nature of the channel can clearly be observed with deep notches being seen with widths of approximately 5 to 7MHz. If we assume a simple two ray model it becomes possible to estimate the impact of rms delay spread on the digital transmission of data. In the frequency domain, time delay spread creates deep frequency notches whereas in the time domain it produces Inter Symbol Interference (ISI).

If we assume a simple two ray model, it is then possible to estimate the impact of rms delay spread. Firstly, it is assumed that the coherence bandwidth can be estimated by evaluating the frequency shift at which the correlation in the fading envelope falls to 0.5.

The following figure shows the vector addition of these rays and the resulting vector envelope, x. The equation below can be used to calculate x for any value of θ.



where τ represents the time delay between the two rays. As we have already seen, it is common for a wideband channel to be described by its root mean square (rms) delay spread. For the simple two ray model this value can be shown to equal half the assumed time delay between the rays (i.e. $\tau_{rms} = \tau/2$). The coherence bandwidth can then be written as:

$$B_c = \frac{1}{6\tau}$$

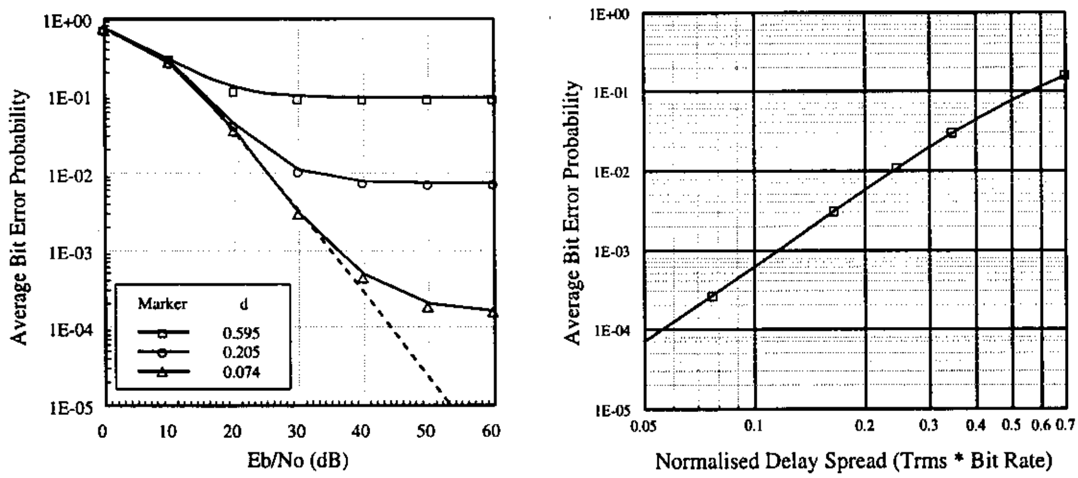
This equation can now be used in conjunction with typical rms delay spreads to estimate approximate values for the coherence bandwidth. For short range picocellular (30m radius) and micro-cellular (200m radius) environments, typical rms delay spreads would be in the order of 30 ns and 200 ns respectively. For larger cells, such as those used in GSM, rms delay spreads between 1 and 2 us might be expected (higher in hostile environments).

Inserting these values into the previous equation reveals that for 2-3 km transmissions, the usable bandwidth lies between 83-166 kHz. However, for transmissions over 200m and 30m, the coherence bandwidth increases to approximately 833 kHz and 5.5 MHz respectively.

For the large cell scenarios, data rates of up to 100 kb/s could probably be supported before the channel would result in an unacceptably high Bit Error Rate (BER) (greater than 1 error in 100). For the indoor environment, far larger bandwidths can be supported with data rates of around 833 kb/s and 5.5 Mb/s becoming possible.

What is meant by a Wideband System?

Wideband transmission is the term used to describe systems that operate far in excess of the channel's coherence bandwidth, B_c. All other systems are known as narrowband and use a frequency range that is less than or equal to this bandwidth.



Irreducible Errors caused by Delay Spread

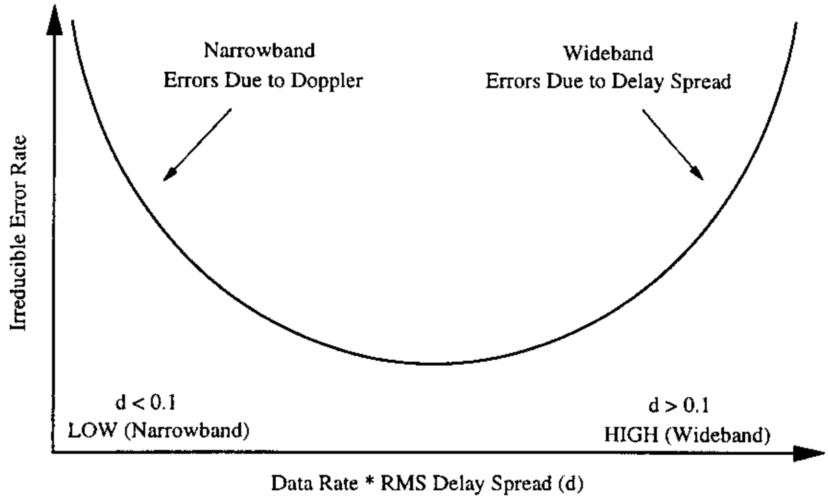
QUESTION

1. Assuming a maximum tolerable bit error rate of 1 in 1000, using the graph above calculate the maximum data rate that can be achieved for an rms delay spread of (i) 50ns and (ii) 2us.

2. If a data rate of (i) 100kb/s, (ii) 1Mb/s and (iii) 10Mb/s is required for the previous system, what is the maximum tolerable rms delay spread?

Propagation Summary

There are two main mechanisms that lead to the introduction of an irreducible error. Firstly, for the narrowband scenario, the amplitude and phase variations associated with the fading channel produce an error rate that is directly proportional the users speed and the frequency of transmission. As the data rate is increased the problems due to Doppler reduce although these are then replaced by the difficulties associated with delay spread. The data rate at which the channel can be considered wideband varies depending on the particular environment being considered. For example, GSM is regarded as a wideband system and operates at a data rate of 270 kb/s. However, DECT is considered narrowband and yet transmits with a data rate of 1.1 Mb/s. The major difference between these systems is the typical rms delay spread in which they operate (DECT and GSM operate with delay spreads as high as 100ns and 4us respectively). If the rms delay spread is normalised to the data rate then we obtain values of 0.11 (100ns*1.1Mb/s) and 1.08 (4us*270kb/s). Generally, for a normalised delay spread less than 0.1 a system can be considered narrowband and the majority of the irreducible errors will arise as a result of the time variations in the channel. Once this normalised value exceeds 0.1 then the channel will start to exhibit wideband characteristics and some form of equalisation may be required to maintain a low overall error rate. This is summarised graphically in the following diagram.



Summary of Irreducible Error Rates