Principles of Digital Modulation

Dr Mike Fitton,

mike.fitton@toshiba-trel.com

Telecommunications Research Lab

Toshiba Research Europe Limited

TOSHIBA-TREL

## Principles of Digital Modulation: Outline of Lectures

● Introduction to digital modulation

● Relevant Modulation Schemes (QPSK, GMSK, M-Ary Schemes)

● Coherent and Differential Reception

● The impact of the mobile channel on digital modulation

     – noise and interference

     – random FM (narrowband fading)

     – intersymbol interference (wideband fading)

Digital Modulation Basics

Digital Modulation Basics

- The *bit rate* defines the rate at which information is passed.
- The *baud* (or *signalling*) *rate* defines the number of symbols per second. Each symbol represents *n* bits, and has *M* signal states, where $M = 2^n$. This is called *M-ary signalling*.
- The maximum rate of information transfer through a baseband channel is given by:
- Capacity $f_b = 2\,W\,\log_2 M$ bits per second
- where W = bandwidth of modulating baseband signal

Amplitude Shift Keying (ASK)

Baseband Data: 1 0 0 1

ASK modulated signal

$A \cos \omega_c t$ 0   0 $A \cos \omega_c t$
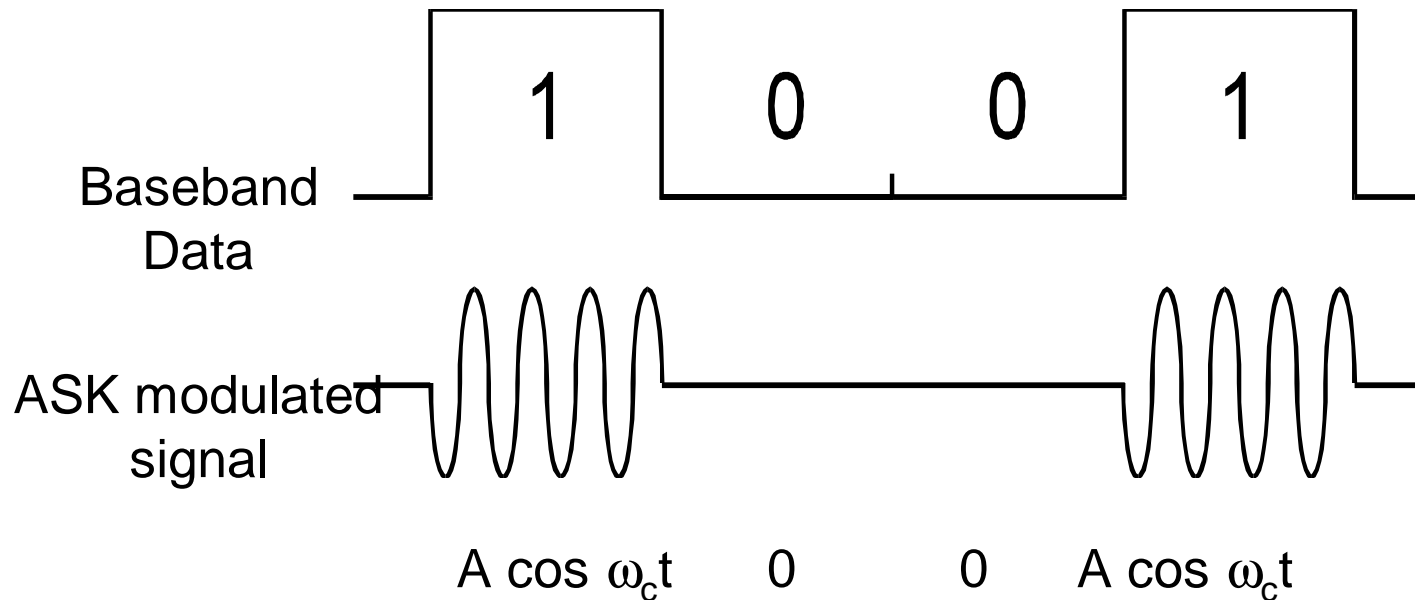
- Pulse shaping can be employed to remove spectral spreading.
- ASK demonstrates poor performance, as it is heavily affected by noise and interference.

Frequency Shift Keying (FSK)

Baseband Data

| 1 | 0 | 0 | 1 |

FSK modulated signal

f_1 f_0 f_0 f_1
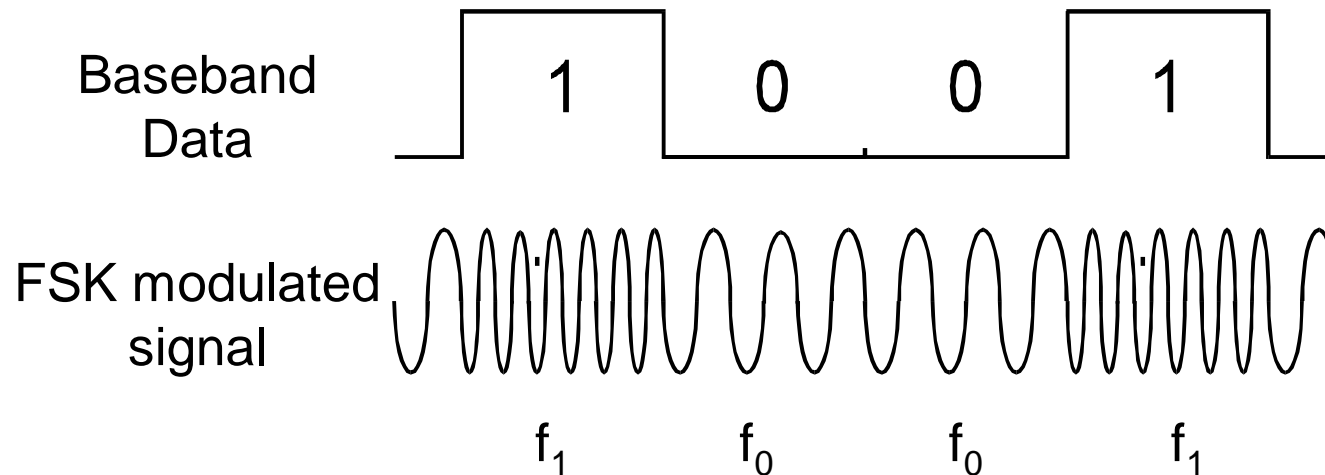
where $f_0 = A \cos(\omega_c - \Delta\omega)t$ and $f_1 = A \cos(\omega_c + \Delta\omega)t$

- Bandwidth occupancy of FSK is dependant on the spacing of the two symbols. A frequency spacing of 0.5 times the symbol period is typically used.
- FSK can be expanded to a M-ary scheme, employing multiple frequencies as different states.
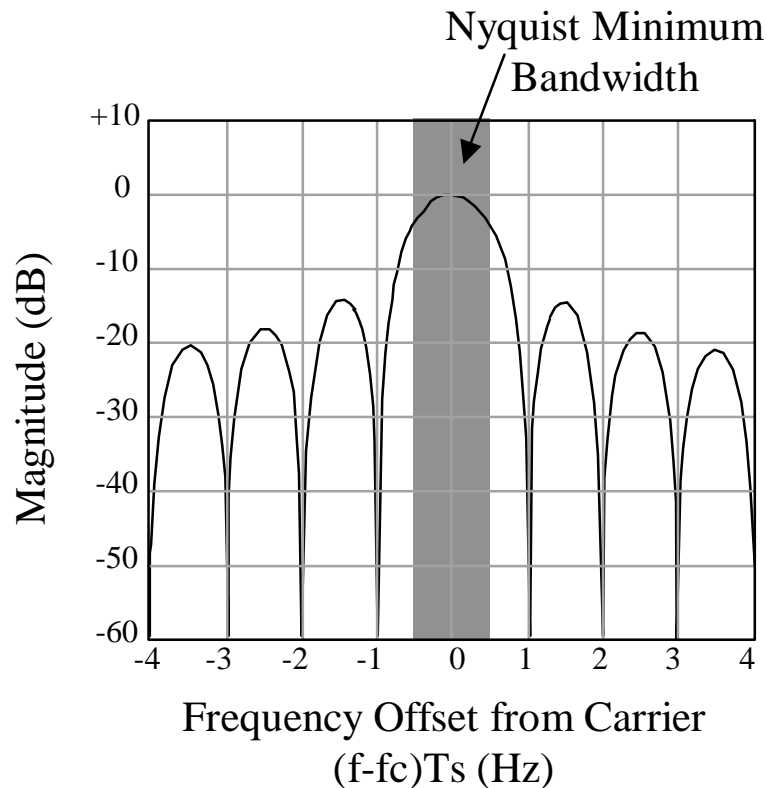
TOSHIBA-TREL

Phase Shift Keying (PSK)

Baseband Data

1   0   0   1

Binary PSK modulated signal

s_1 s_0 s_0 s_1
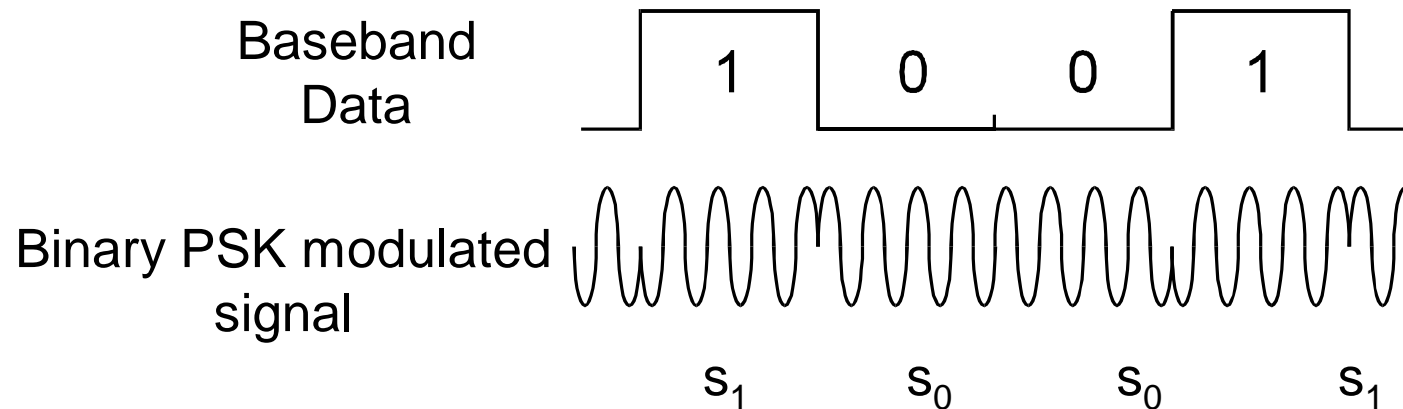
where $s_0 = -A \cos \omega_c t$ and $s_1 = A \cos \omega_c t$

- Binary Phase Shift Keying (BPSK) demonstrates better performance than ASK and FSK.

- PSK can be expanded to a M-ary scheme, employing multiple phases and amplitudes as different states.

- Filtering can be employed to avoid spectral spreading.

Nyquist & Root-Raised Cosine Filters

Nyquist Minimum
Bandwidth

Magnitude (dB)

+10
0
-10
-20
-30
-40
-50
-60

-4 -3 -2 -1 0 1 2 3 4

Frequency Offset from Carrier
(f-fc)Ts (Hz)

- The Nyquist bandwidth is the minimum bandwidth than can be used to represent a signal.

- It is important to limit the spectral occupancy of a signal, to improve bandwidth efficiency and remove adjacent channel interference.

- Root raised cosine filters allow an approximation to this minimum bandwidth.

Nyquist bandwidth on the QPSK spectrum

8

TOSHIBA-TREL

Modulation - QPSK
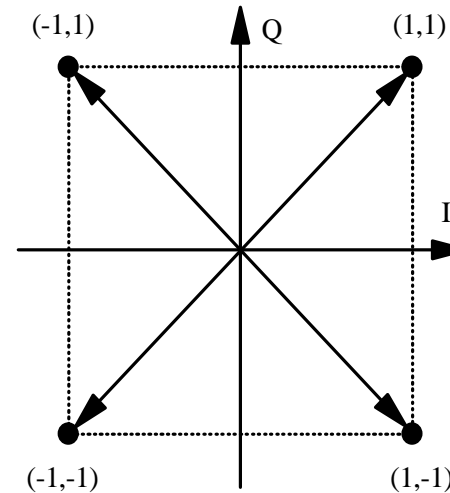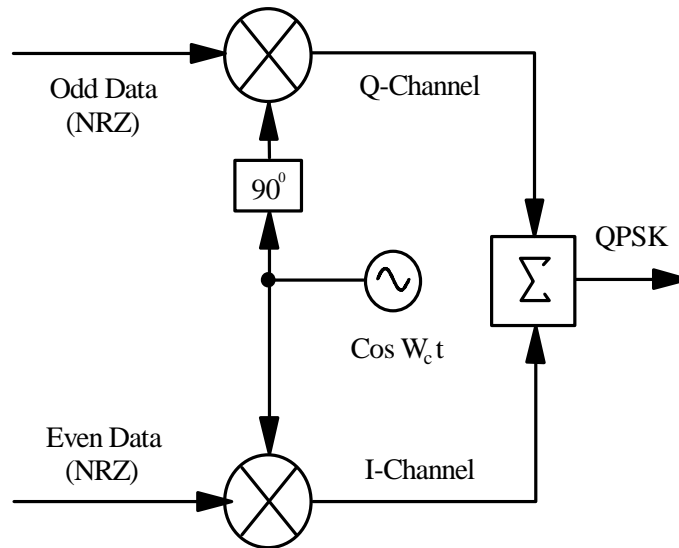
Wc = Carrier Frequency, I = In phase channel, Q = Quadrature channel

● Quadrature Phase Shift Keying is effectively two independent BPSK systems (I and Q), and therefore exhibits the same performance but *twice* the bandwidth efficiency.

● Quadrature Phase Shift Keying can be filtered using raised cosine filters to achieve excellent out of band suppression.

● Large envelope variations occur during phase transitions, thus requiring *linear* amplification.

# Types of QPSK

Conventional QPSK
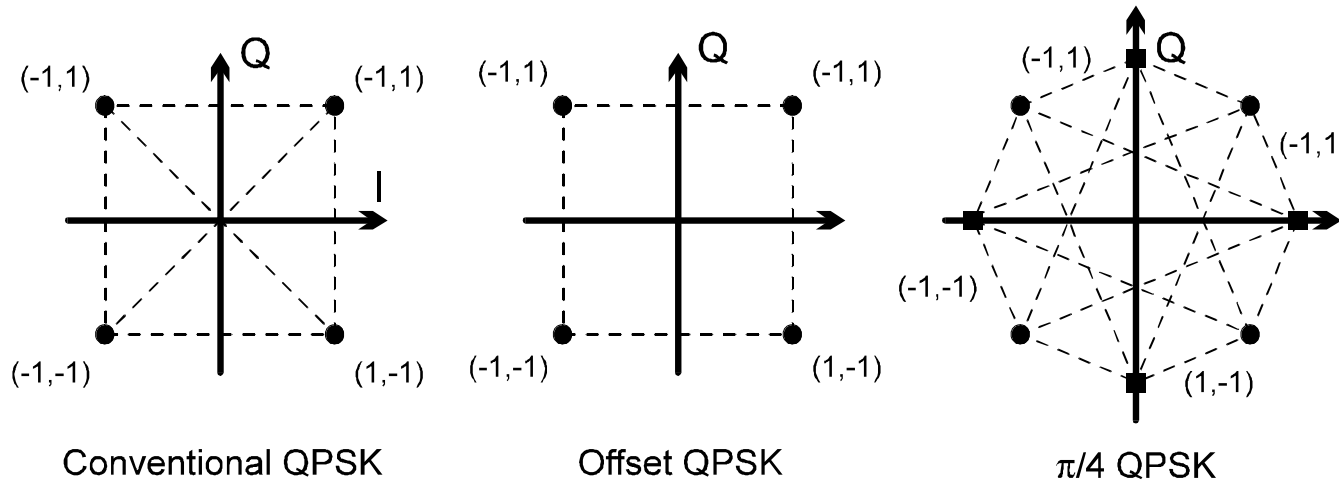
Offset QPSK

$\pi/4$ QPSK

● Conventional QPSK has transitions through zero (ie. 180° phase transition). Highly linear amplifier required.

● In Offset QPSK, the transitions on the I and Q channels are staggered. Phase transitions are therefore limited to 90°.

● In $\pi/4$-QPSK the set of constellation points are toggled each symbol, so transitions through zero cannot occur. This scheme produces the lowest envelope variations.

● All QPSK schemes require linear power amplifiers.

GMSK - Gaussian Minimum Shift Keying

● GMSK is a form of continuous-phase FSK, in which the phase is changed between symbols to provide a constant envelope. Consequently, it is a popular alternative to QPSK.

● The RF bandwidth is controlled by the Gaussian low-pass filter bandwidth.

● The degree of filtering is expressed by multiplying the filter 3dB bandwidth by the bit period of the transmission, ie. by BT.

● As BT is lowered the amount of *intersymbol-interference* introduced increases and this results in either a fixed power penalty or an irreducible error floor.

● GMSK allows efficient class C non-linear amplifiers to be used, however even with a low BT value its bandwidth efficiency is less than filtered QPSK.
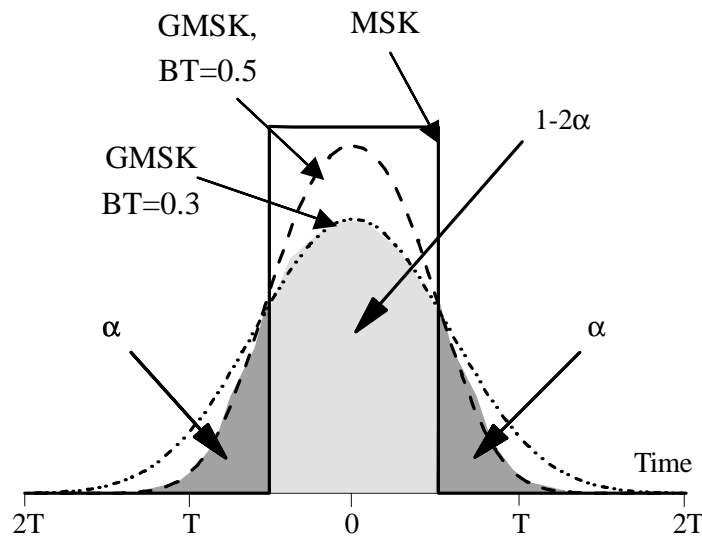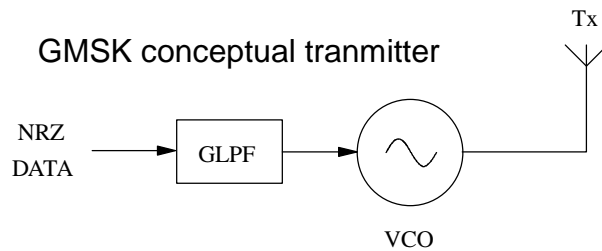
Minimum Shift Keying (MSK)
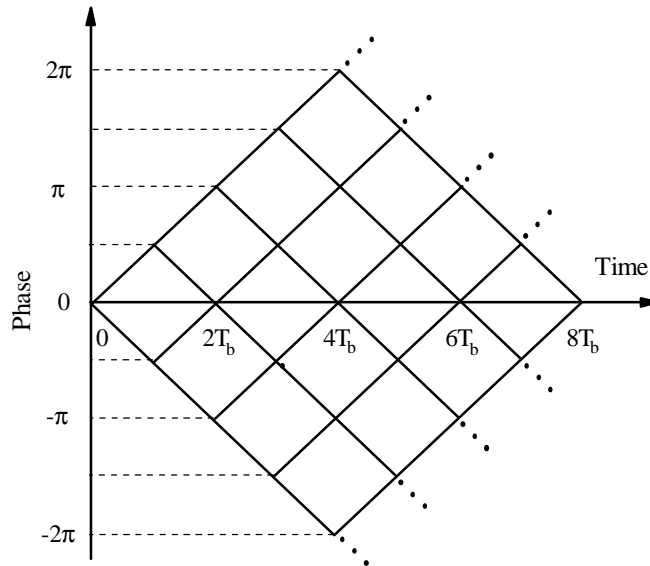
MSK possible phase transitions
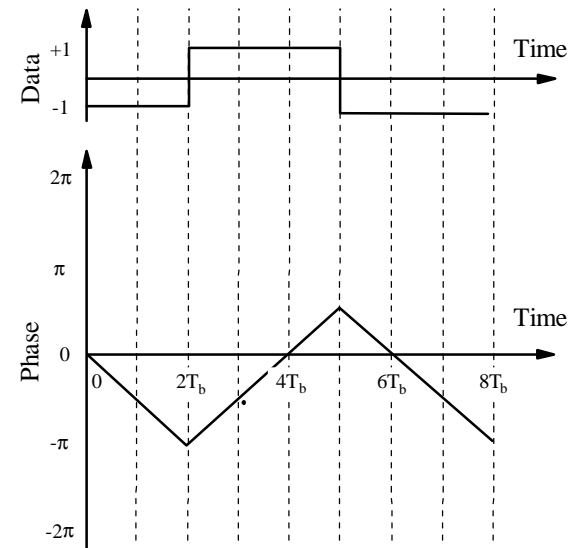
MSK phase transitions for data:
(00111000...)

- In MSK phase ramps up through 90 degrees for a binary one, and down 90 degrees for a binary zero.

- 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 BT product.
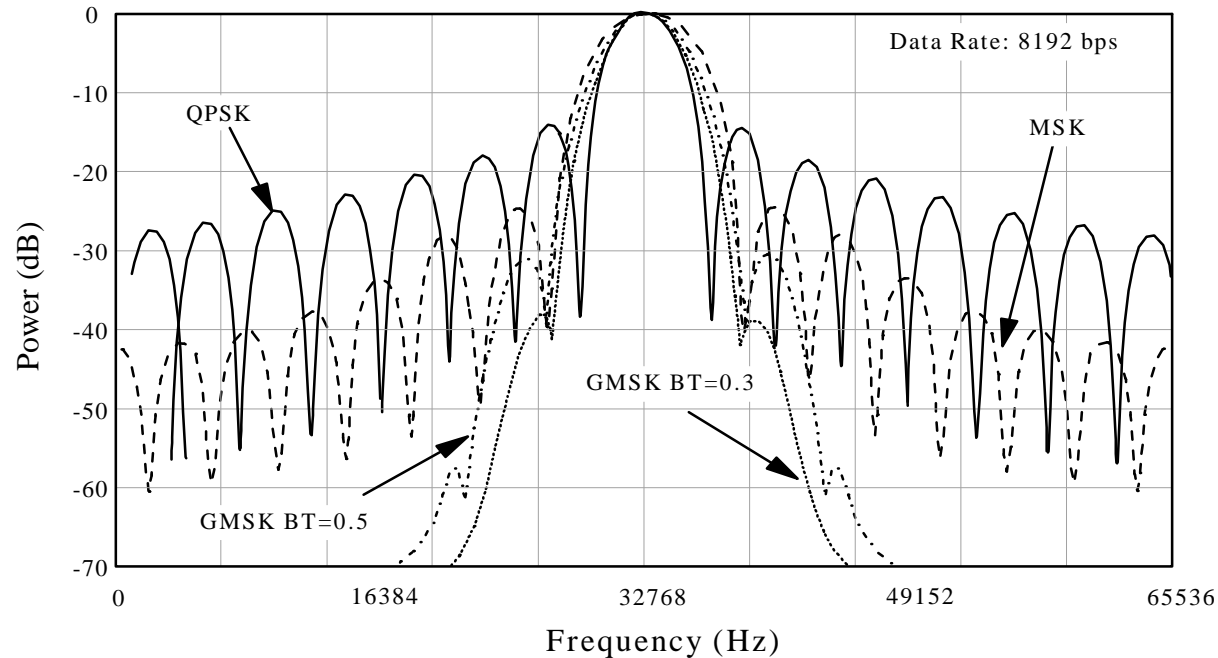
GMSK Signals

GMSK conceptual tranmitter

Tx

NRZ
DATA → GLPF → VCO

GMSK,
BT=0.5

MSK

GMSK
BT=0.3

$1-2\alpha$

α

α

Time

2T   T   0   T   2T

GMSK Pulse Shapes and ISI

- In MSK , the BT is infinity and this allows the square bit transients to directly modulate the VCO.

- In GMSK, low values of BT create significant intersymbol interference (ISI). In the diagram, the portion of the symbol energy α acts as ISI for adjacent symbols.

- If BT is less than 0.3, some form of combating the ISI is required.

GMSK Spectra

The chart shows Power (dB) versus Frequency (Hz). Data Rate: 8192 bps. Curves labeled: QPSK, MSK, GMSK BT=0.3, GMSK BT=0.5.

- GMSK has a main lobe 1.5 times that of QPSK.

- GMSK generally achieves a bandwidth efficiency less than 0.7 bits per second per Hz (QPSK can be as high as 1.6 bits per second per Hz).

Multi-level (M-ary) Phase and Amplitude Modulation

●Amplitude and phase shift keying can be combined to transmit several bits per symbol (in this case M=4). These modulation schemes are often refered to as *linear*, as they require linear amplification.

●16QAM has the largest distance between points, but requires very linear amplification. 16PSK has less stringent linearity requirements, but has less spacing between constellation points, and is therefore more affected by noise.

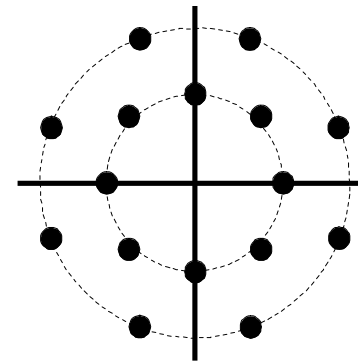●M-ary schemes are more bandwidth efficient, but more susceptible to noise.
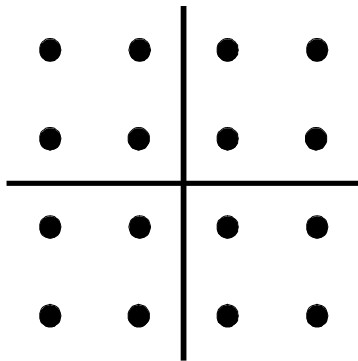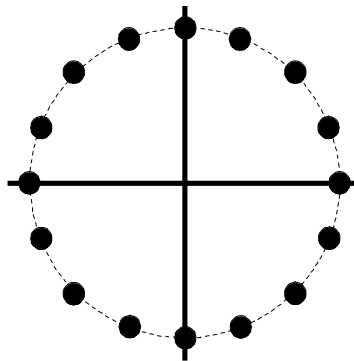
16 QAM   16 PSK   16 APSK

Shannon-Hartley Capacity Theorem

For error free communication, it is possible to define the capacity which can be supported in an additive white gaussian noise (AWGN) channel.

$$f_b/W = \log_2(1 + E_b f_b /\eta W)$$

where f_b = Capacity (bits per second)

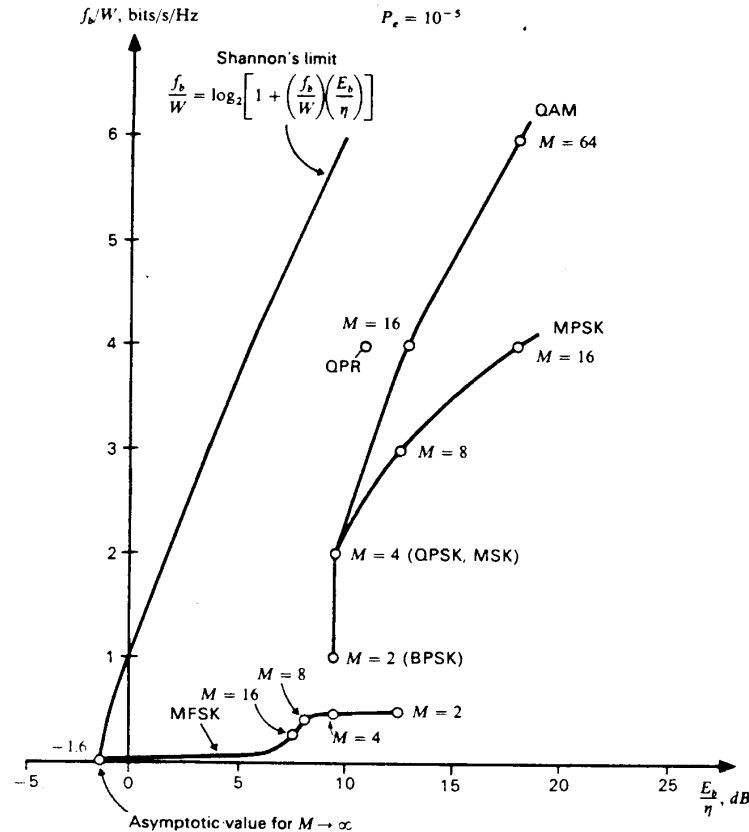W = bandwidth of the modulating baseband signal (Hz)

E_b = energy per bit

η = noise power density (watts/Hz)

thus $E_b f_b$ = total signal power

ï ηW = total noise power

f_b/W = bandwidth efficiency (bits per second per Hz)

Comparison of Modulation Schemes

f_b/W, bits/s/Hz $P_e = 10^{-5}$

Shannon's limit

$$\frac{f_b}{W} = \log_2\left[1 + \left(\frac{f_b}{W}\right)\left(\frac{E_b}{\eta}\right)\right]$$

QAM
M = 64

M = 16
QPR

MPSK
M = 16

M = 8

M = 4 (QPSK, MSK)

M = 2 (BPSK)

M = 8
M = 16
MFSK
M = 4
M = 2

Asymptotic value for M → ∞

$\frac{E_b}{\eta}$, dB

bits/s/Hz vs. E_b/η for Probability of Error = 10^{-5}
taken from "Principle of Communication Systems"
Taub & Schilling, page 482

- •This graph shows that *bandwidth efficiency* is traded off against *power efficiency*.
- ● MFSK is power efficient, but not bandwidth efficient.
- ● MPSK and QAM are bandwidth efficient but not power efficient.

- ● Mobile radio systems are *bandwidth limited*, therefore PSK is more suited.

Comparison of Modulation types

Modulation Format	Bandwidth efficiency (C/B)	log2 (C/B)	Error free Eb/No
16 PSK	4	2	18dB
16 QAM	4	2	15dB
8PSK	3	1.585	14.5dB
4PSK	2	1	10.1dB
4QAM	2	1	10.1dB
BFSK	1	0	13dB
BPSK	1	0	10.5dB

TOSHIBA-TREL

Spectral Efficiencies in practical radios

- ## GSM- Digital Cellular
  - Data Rate = 270kb/s, bandwidth = 200kHz
  - Bandwidth Efficiency = 270/200   =1.35bits/sec/Hz
  - Modulation: Gaussian Minimum Shift Keying (FSK with orthogonal frequencies).
  - "Gaussian" refers to filter response.

- ## IS-54 North American Digital Cellular
  - Data Rate = 48kb/s, bandwidth = 30kHz
  - Bandwidth Efficiency = 48/30   =1.6bits/sec/Hz
  - Modulation: $\pi/4$ DPSK

Coherent Reception

An estimate of the channel phase and attenuation is recovered. It is then possible to reproduce the transmitted signal, and demodulate. It is necessary to have an accurate version of the carrier, otherwise errors are introduced. Carrier recovery methods include:

● Pilot Tone (such as Transparent Tone in Band)

• Less power in information bearing signal

• High peak-to-mean power ratio

● Pilot Symbol Assisted Modulation

• Less power in information bearing signal

● Carrier Recovery (such as Costas loop)

• The carrier is recovered from the information signal

TOSHIBA-TREL

Differential Reception

- In the transmitter, each symbol is modulated relative to the previous symbol, for example in differential BPSK:

- 0 = no change   1 = +180°

- In the receiver, the current symbol is demodulated using the previous symbol as a reference. The previous symbol acts as an estimate of the channel.

- Differential reception is theoretical 3dB poorer than coherent. This is because the differential system has *two* sources of error: a corrupted symbol, and a corrupted reference (the previous symbol).

- Non-coherent reception is often easier to implement.

TOSHIBA-TREL

Modulation Summary

- Phase Shift Keying is often used, as it provides a highly bandwidth efficient modulation scheme.

- QPSK, modulation is very robust, but requires some form of linear amplification. OQPSK and π/4-QPSK can be implemented, and reduce the envelope variations of the signal.

- High level M-ary schemes (such as 64-QAM) are very bandwidth-efficient, but more susceptible to noise and require linear amplification.

- Constant envelope schemes (such as GMSK) can be employed since an efficient, non-linear amplifier can be used.

- Coherent reception provides better performance than differential, but requires a more complex receiver.

TOSHIBA-TREL

Problems in the wireless environment: noise, interference and the mobile channel

TOSHIBA-TREL

Noise in the mobile radio channel

- Noise arises from a variety of sources, including automobile ignitions and lightning, or thermal noise in the receiver itself. Thermal noise can be modelled as *Additive White Gaussian Noise* (AWGN).

- The ratio of the signal strength to the noise level is called the *signal-to-noise ratio* (SNR). If the SNR is high (ie. the signal power is much greater than the noise power) few errors will occur. However, as the SNR reduces, the noise may cause symbols to be demodulated incorrectly, and errors will occur.

- The *bit error rate* (BER) of a system indicates the quality of the link. Usually, a BER of 10^{-3} is considered acceptable for a voice link, and 10^{-9} for a data link. A coherent QPSK system requires a SNR of greater than approximately 12dB for a BER of better than 10^{-3}.

Interference in the mobile radio channel

- Interference is the result of other man-made radio transmissions.

    - for example in the ISM band at 2.4GHz a large number of systems co-exist, such as Wireless LAN, Bluetooth, Microwave ovens, etc

- *Adjacent channel interference* occurs when energy from a carrier spills over into adjacent channels. *Co-channel interference* occurs when another transmission on the same carrier frequency affects the receiver. This will often arise from transmissions in another cell in ther network.

- The ratio of the carrier to the interference (from both sources) is called the *carrier-to-interference ratio* (C/I). A certain C/I ratio is required to provide adequate quality transmission.

- Increasing the carrier power at the receiver will increase the interference for other mobiles in the network.

The Multipath Environment
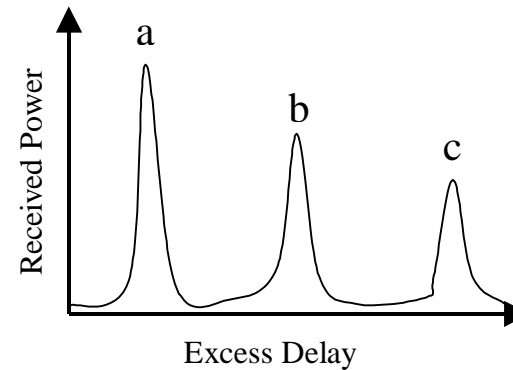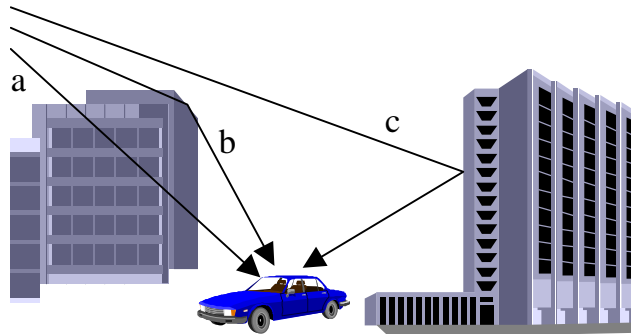
- The received signal is made up of a sum of attenuated, phase-shifted and time delayed versions of the transmitted signal.

- Propagation modes include diffraction, transmission and reflection.

Narrowband fast fading
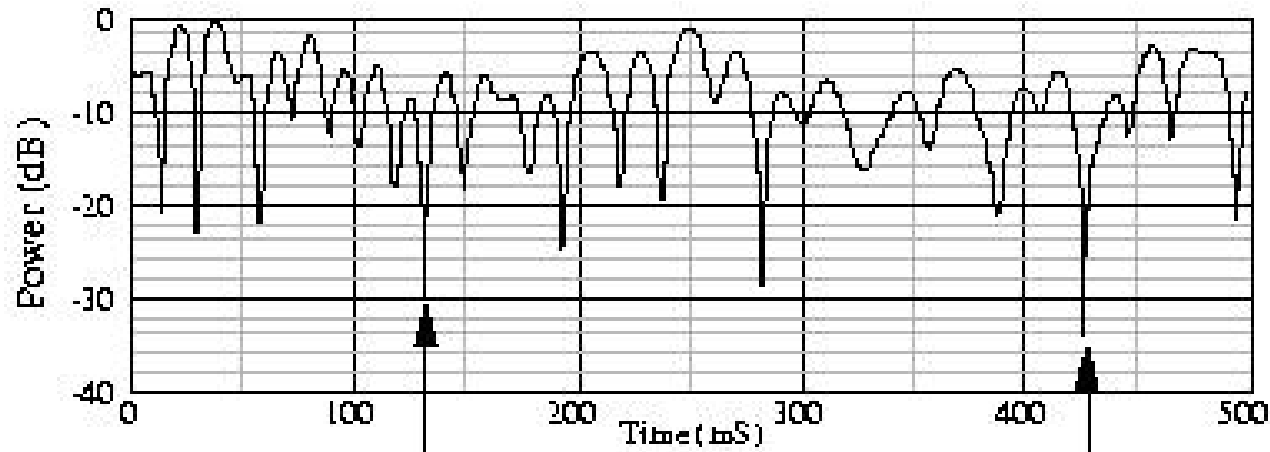
- If time dispersion is small, vector sum of rays occurs
- The arrival phase of each path alters as the receiver moves, resulting in a different vector sum.
- Rapid phase and amplitude shifts are observed (up to 40dB). Magnitude is modelled as Rayleigh (no line-of-sight) or Rician (more deterministic).

Shadowing (Slow Fading)
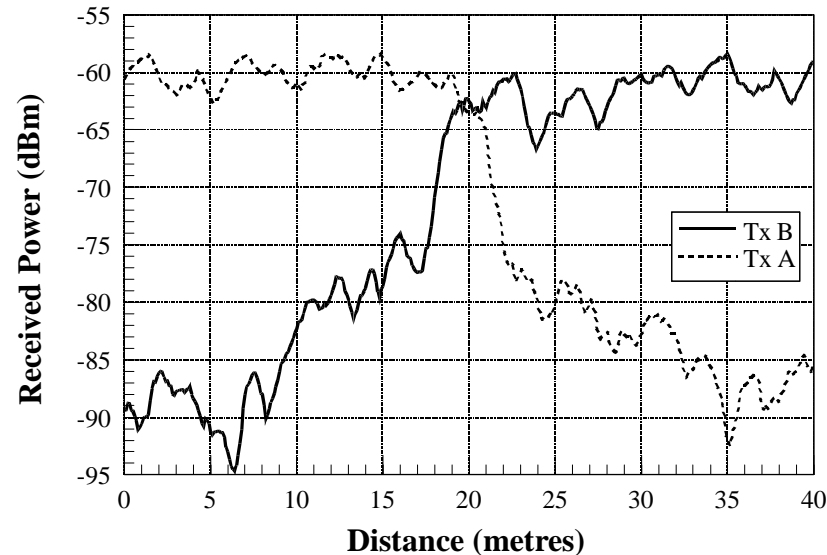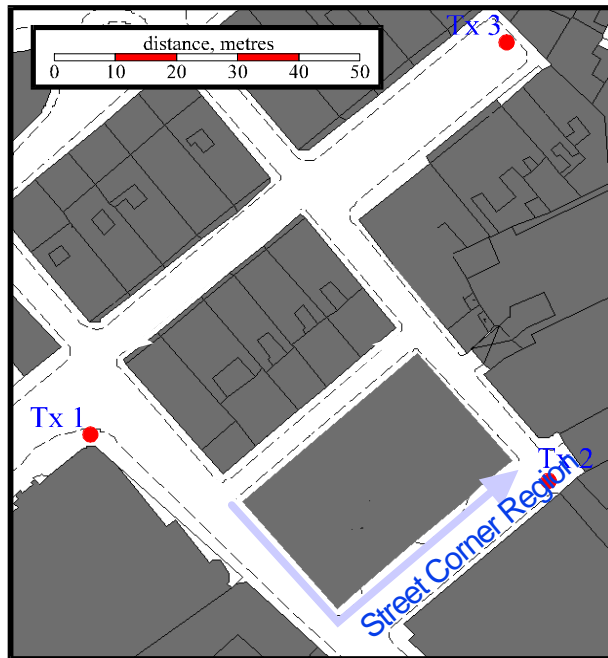
Reciprocal uplink/downlink

- Amplitude variation occurs as the receiver moves behind buildings and the propagation paths are obscured
- Variations of up to 20dB will cause handovers and change quality-of-service

Noise and interference in the multipath channel

• The received signal in a multipath channel exhibits large variations in magnitude.

• Although the *mean* SNR (or C/I) might be acceptable, the variations experienced in the multipath channel mean that occasionally the noise will be far more significant. At these times the system will experience a large number of errors.


mean SNR = 10dB

Velocity=1m/ s
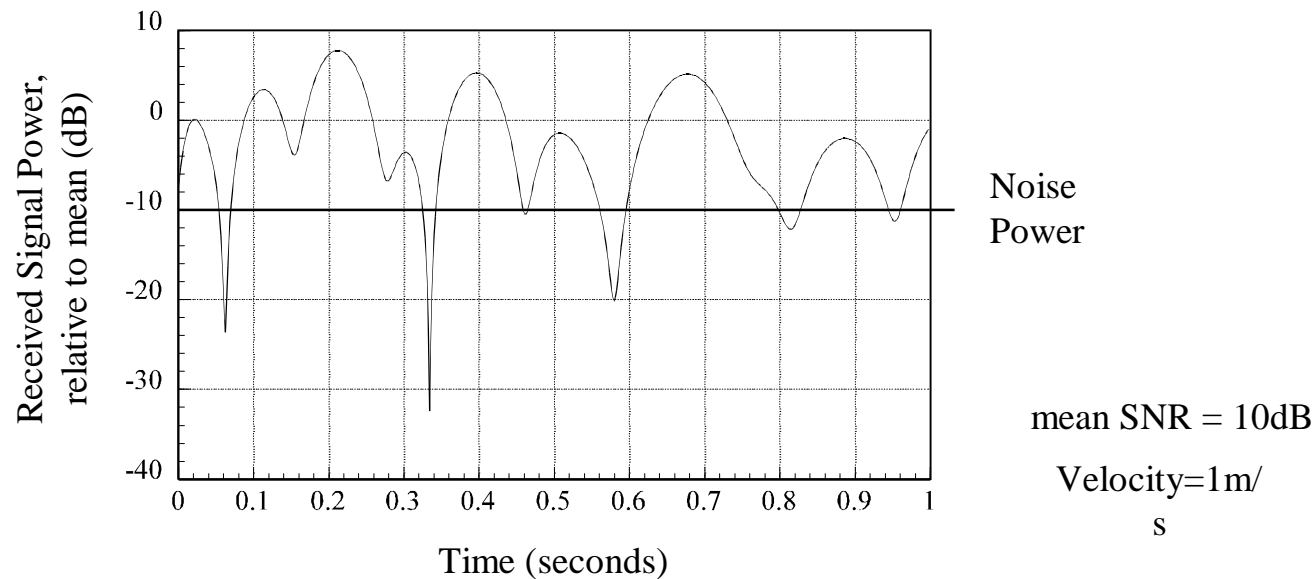
Rayleigh Distribution of the Multipath channel

Cumulative Density Function
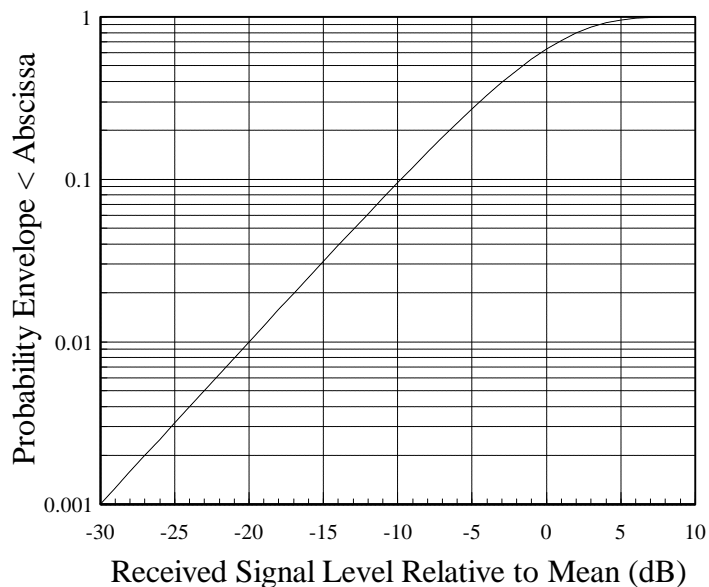
- A multipath channel without a significant deterministic component can be approximated to a *Rayleigh* distribution.

- The received signal experiences large variations in magnitude. For example, there is a 0.1% chance of the signal being 30dB below the mean level

- Consequently, even a system with a high SNR can experience errors as the signal fades.

System Performance in AWGN
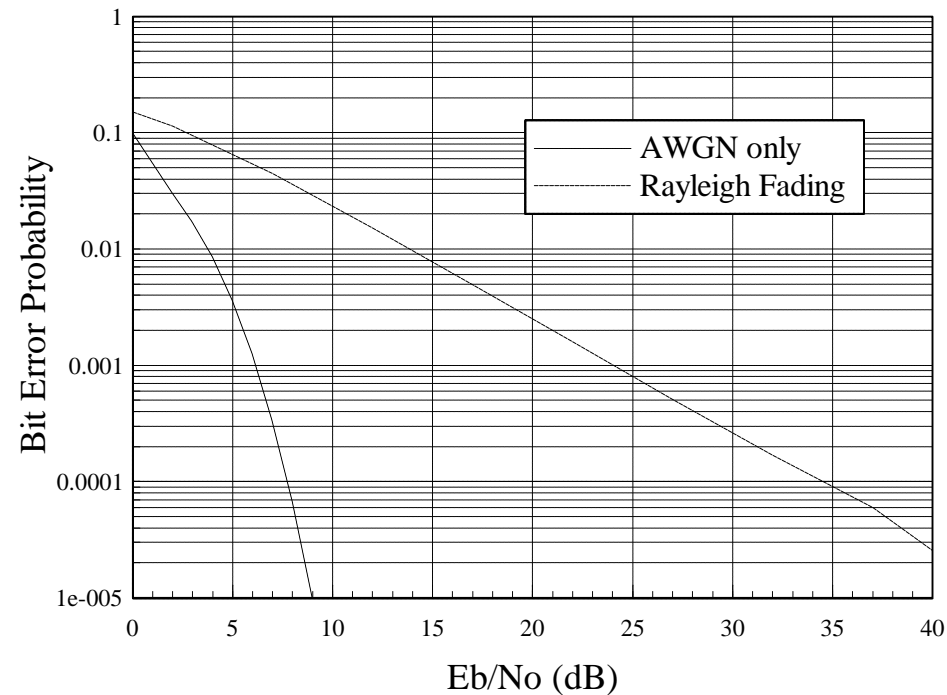
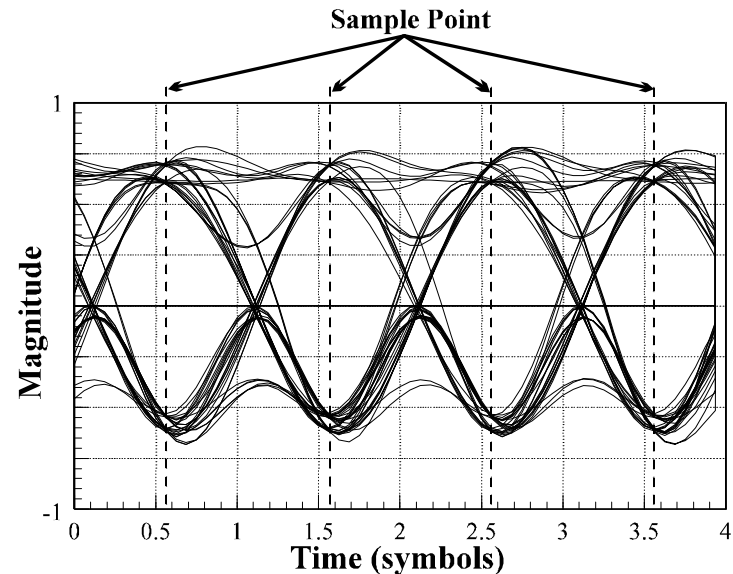

The effects of the multipath channel (*Rayleigh fading*) severely degrade the system performance in the presence of Additive White Gaussian Noise.
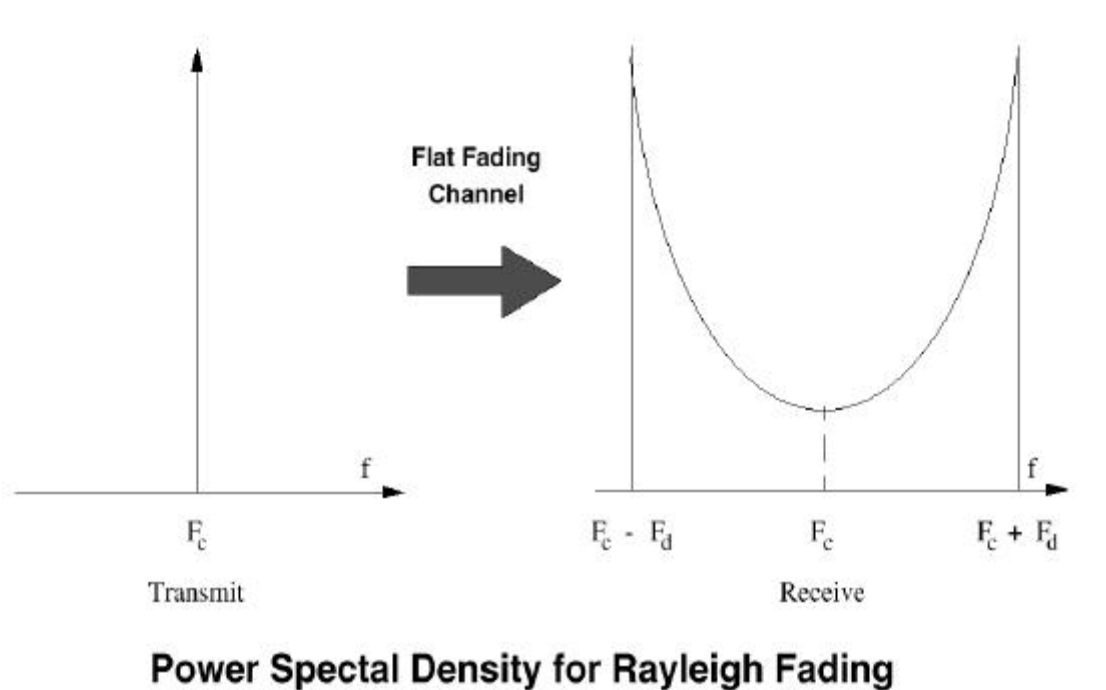
Eye Diagrams
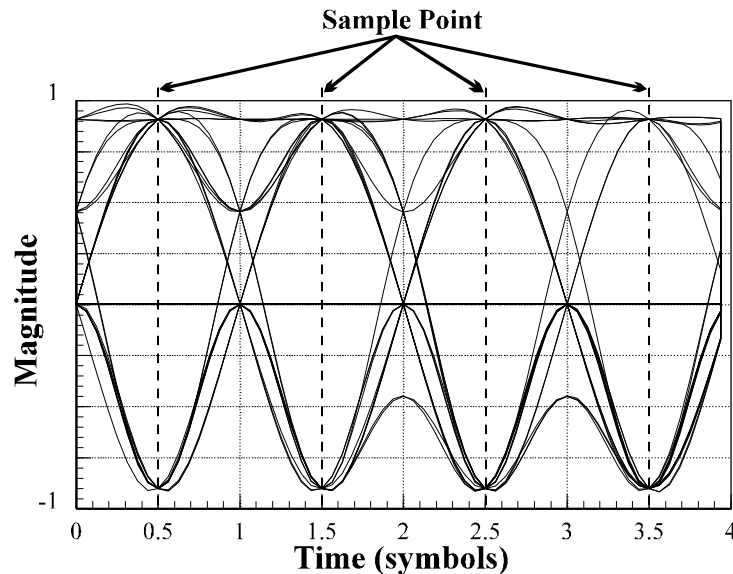
(i) No noise or interference

(ii) System corrupted by interference

- Eye diagrams show the signal superimposed on itself many times.

- If the "eye" is not "open" at the sample point, errors will occur.

- Eye diagram will be corrupted by noise and interference.

The Effect of Fading - Doppler

Flat Fading
Channel

Transmit — F_c

Receive — $F_c - F_d$, F_c , $F_c + F_d$

f

Power Spectal Density for Rayleigh Fading

Motion of the mobile causes periodic phase shifts which change with time. A typical spectrum for a Rayleigh channel is shown above. The rate of change of phase gives rise to a Doppler frequency (F_d), which varies with mobile speed (v) and the arrival angle of the rays (a_n).

- $F_d = v/l \cos a_n$ (l = wavelength)

33

The Effect of Fading - Random FM

FdTs=0.001   FdTs=0.008   FdTs=0.025
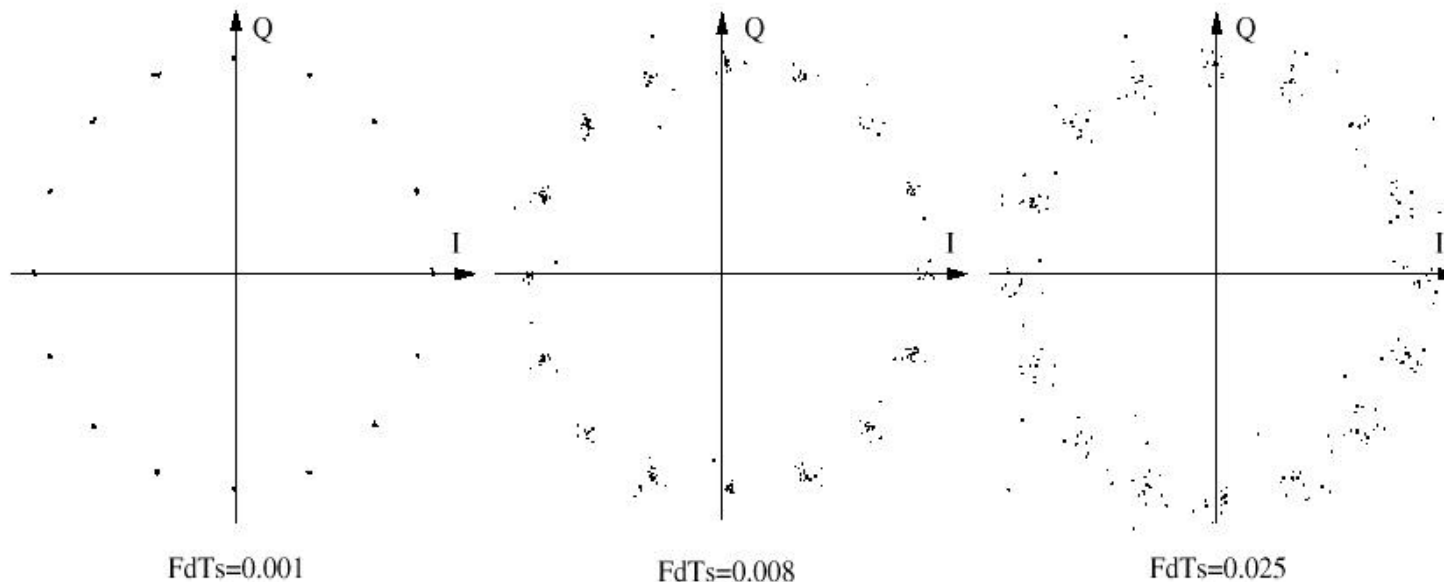
16PSK Constellations for Various FdTs

● The phase changes due to Doppler are superimposed on the received signal, and can cause errors if large. This phase change is often called *Random FM*.

● The phase error per symbol depends on both mobile Doppler frequency and symbol period. Consequently, Doppler frequency is often normalised to symbol period ($F_d\ T_s$).
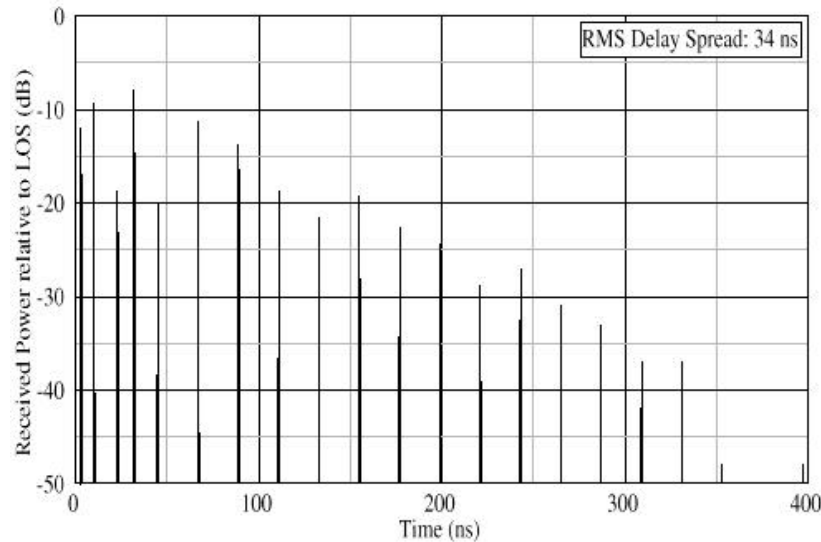
Irreducible Errors due to Random FM
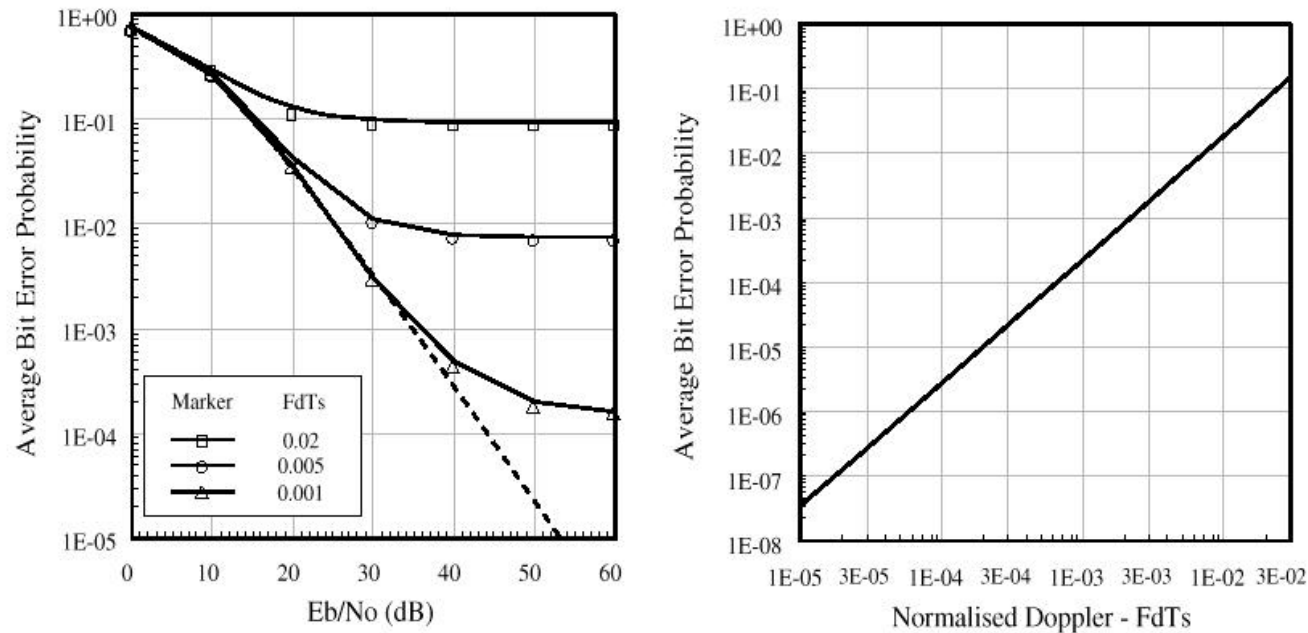
Irreducible Errors caused by Random FM

Random FM introduces an *irreducible* error floor which cannot be removed by increasing transmit power. In a differential system, this error floor depends on the phase change over a symbol period ($F_d T_s$).

Time Dispersion in the Multipath Channel


Typical Indoor Power Delay Profile

- The time dispersion associated with the multipath channel can cause problems if high data rate digital modulation is employed.
- The Power Delay Profile shown here shows the power and delay of each ray arriving at the receiver.

- The dispersion of the channel is normally characterised using the *RMS delay spread*, which is defined as the standard deviation of the power delay profile, as shown here:

$$t_a = \frac{\sum\limits_{k=1}^{N} t_k a_k^2}{\sum\limits_{k=1}^{N} a_k^2} \qquad t_{rms} = \left[\frac{\sum\limits_{k=1}^{N} [t_k - t_a]^2 a_k^2}{\sum\limits_{k=1}^{N} a_k^2} \right]^{1/2}$$
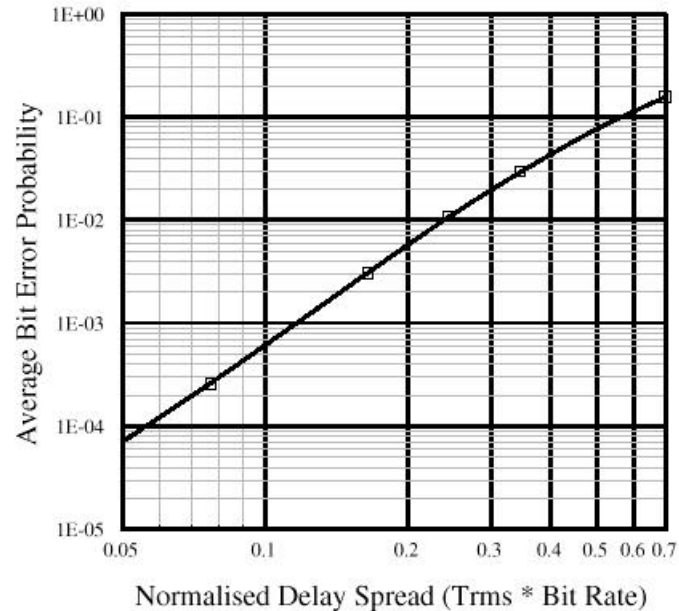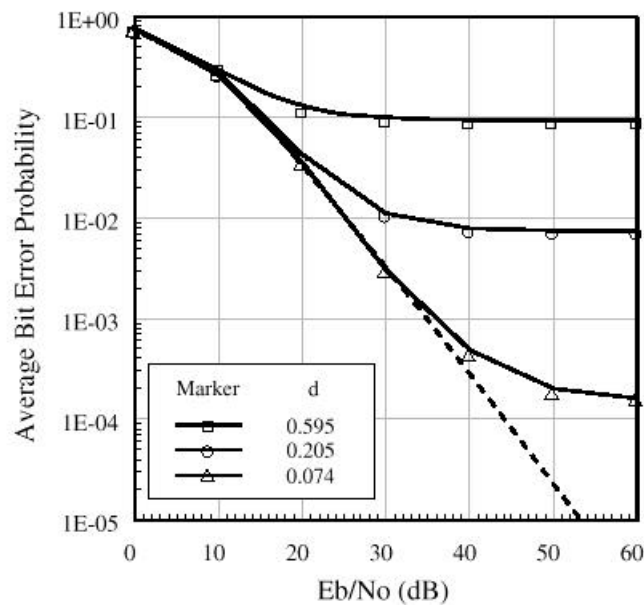
36

Intersymbol Interference (ISI)
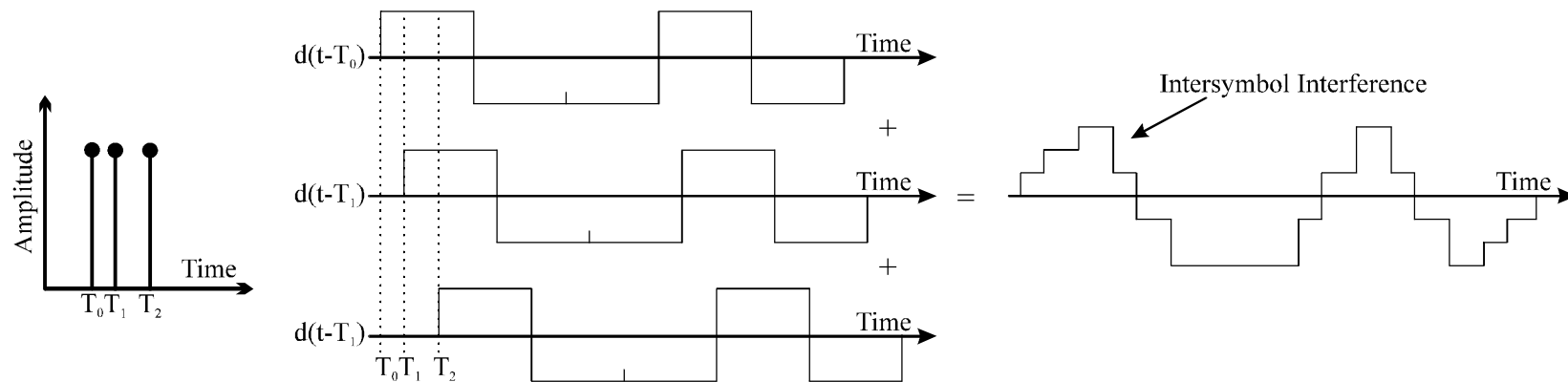
- ISI arises when energy from one symbol slot is spread out over neighbouring symbol slots.

- ISI is introduced by the channel when the RMS delay spread becomes an appreciable fraction of the bit period (say greater than 10%).

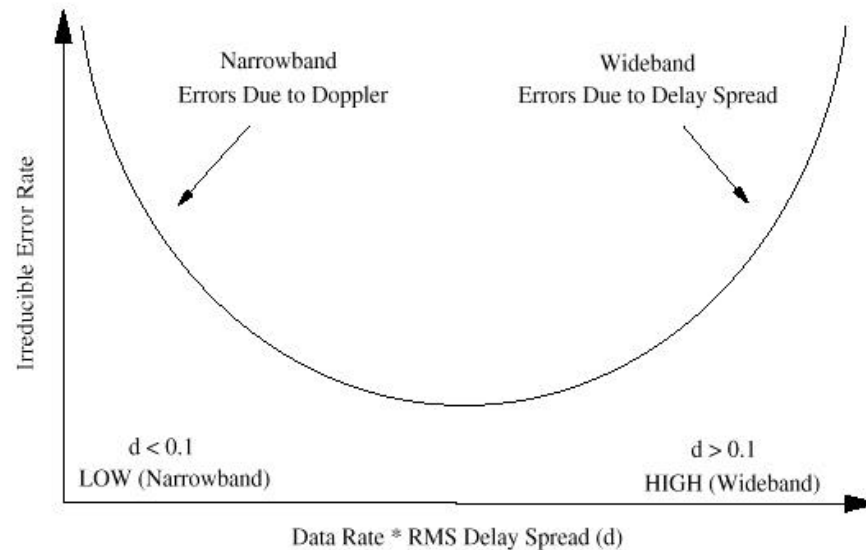# Irreducible Errors due to time dispersion in the multipath channel

Irreducible Errors caused by Delay Spread

- Intersymbol interference introduces an *irreducible* error floor which cannot be removed by increasing transmit power. This error floor degrades as symbol rate (or delay spread) increases.
- Delay spread is often *normalised* to symbol rate or bit rate.

Summary of Error Mechanisms

● Noise arises from a variety of sources, including ignition noise and thermal noise in the receiver. Man-made radio transmissions cause *adjacent channel interference* and *co-channel interference.*

● In the presence of noise and interference, it is necessary to increase signal power to reduce the possibility of errors.

● The multipath channel gives rise to irreducible errors form *random FM* and *intersymbol interference.* These errors are *irreducible* as they cannot be removed by increasing signal power.

Summary of Error Mechanisms: Irreducible Errors

Summary of Irreducible Error Rates

● Low symbol rate (*narrowband*): large phase change over long symbol period, therefore errors arise due to Doppler.

● High symbol rate (*wideband*), small phase change over short symbol period. Dispersion is large compared to the symbol period, therefore errors due to intersymbol interference.

● Whether narrowband or wideband depends on the symbol rate and the enironment.