3F4 Digital Modulation Course

Nick Kingsbury

February 10, 2016

Contents

Recommended Textbooks

Key Textbook:

- L. W. Couch, **Digital and Analog Communication Systems**, Prentice Hall, 6th edition, 2001.

Books covering digital communications only:

- R. E. Zeimer and R. L. Peterson, **Introduction to Digital Communication**, Prentice Hall, 2nd edition, 2001.
- J. G. Proakis, **Digital Communications**, McGraw Hill, 4th edition, 2001.
- I. A. Glover and P. M. Grant, **Digital Communications**, Prentice Hall, 1998.

Alternative textbooks on analogue and digital communications (do not cover all the course) and selected topics:

- K. S. Shanmugam, **Digital and Analog Communication Systems**, Wiley, 1979 and later editions.
- B. P. Lathi, **Modern Digital and Analog Communication Systems**, Oxford U P, 3rd edition, 1998.
- J. D. Gibson, **Principles of Digital and Analog Communications**, Prentice Hall, 2nd edition, 1993.

1 Modulation, Phasors and Bandlimited Noise

1.1 Types of Modulation - AM, FM, PM

We define a general sinusoidal wave as:

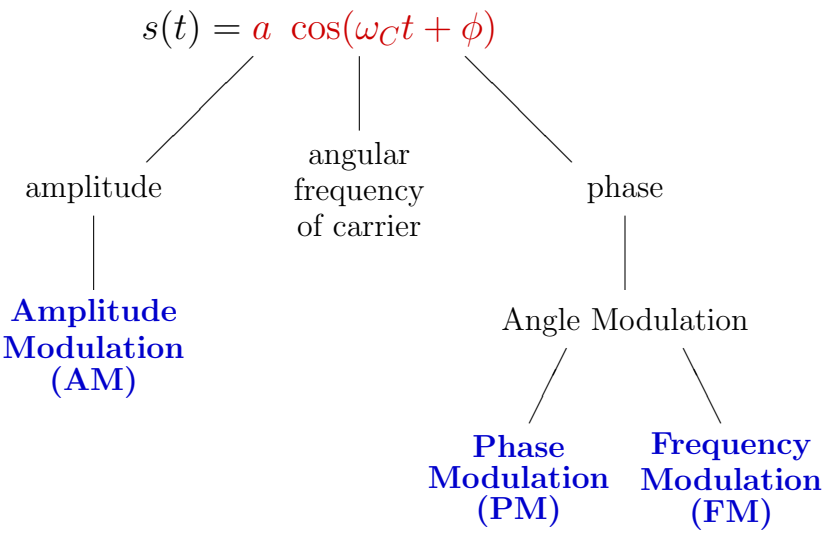

Fig 1.1 shows these three forms of modulation. The triangular wave at the top is the information signal, and below it is the carrier wave. Then the three modulated waves are shown.

If the information signal $= x(t)$, then:

For Ampl. Mod. (AM):

$$a = a_0 + K_A x(t)$$

For Phase Mod. (PM):

$$\phi = \phi_0 + K_P x(t)$$

For Frequency Mod. (FM):

$$\frac{d\phi}{dt} = K_F x(t)$$

Why use Modulation?

1. To transmit information via a **bandlimited channel**.
   e.g. Computer data via a 300 to 3000 Hz telephone circuit.
2. To pass **many information channels** via a common medium simultaneously.
   e.g. Radio and TV signals use different carrier frequencies to avoid interference.

For continuous analogue signals (e.g. TV, speech or music) AM or FM are most common.

For digital signals AM and PM are used and are often combined.

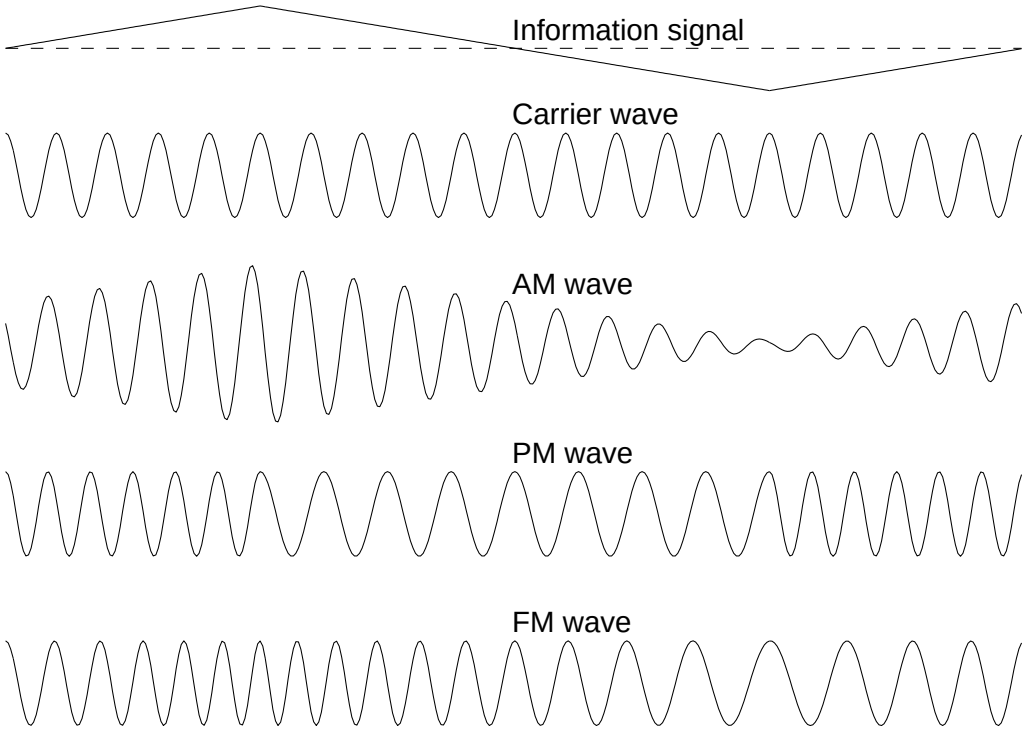

Fig 1.1: Main forms of analogue modulation

1.2 Important Features of Modulation Schemes:

There are 3 main criteria for assessing modulation schemes:

1. **Bandwidth required** – affects how close adjacent carriers can be without interference occuring.
2. **Demodulator / receiver complexity** – affects the cost of a system.
3. **Noise rejection properties** – affect the transmitter power needed and the maximum range.

1.3 Phasor Representation of Modulated Signals

It is useful to handle AM, PM and FM in a unified way.

Let the modulated wave be:

$$s(t) = a(t)\cos(\omega_C t + \phi(t)) \tag{1.1}$$

Note $a(t)$ and $\phi(t)$ are **difficult to combine**.

So we consider the cosine term as the real part of a complex exponential:

$$\begin{aligned} s(t) &= Re[a(t)\ e^{j(\omega_C t+\phi(t))}] \\ &= Re[a(t)\ e^{j\phi(t)}\ e^{j\omega_C t}] \\ &= Re[\underbrace{p(t)}_{\substack{\text{modulation}\\ \text{phasor}}}\ \underbrace{e^{j\omega_C t}}_{\substack{\text{carrier}\\ \text{wave}}}\ ] \end{aligned} \tag{1.2}$$

$$\text{where}\quad p(t) = \underbrace{a(t)}_{\substack{\text{ampl.}\\ \text{of } p(t)}}\ \underbrace{e^{j\phi(t)}}_{\substack{\text{phase}\\ \text{of } p(t)}} \tag{1.3}$$

Now the modulation is completely separated from the carrier wave, and a and ϕ have been **combined into a single phasor waveform**, $p(t)$.

Phasors are very useful for defining any form of modulation that is to be applied to a carrier wave, without involving the carrier itself. Phasors vary much more slowly than the modulated wave, $s(t)$, and are easier to analyse.

See figs 1.2, 1.3 and 1.4 (3D plots of AM, PM and FM).

Significance of the Real and Imaginary Parts of Phasors:

If $p(t)$ has real and imaginary parts, $i(t)$ and $q(t)$, then:

$$\text{Let}\quad p(t) = i(t) + jq(t) \tag{1.4}$$

$$\therefore\ \text{from (1.3)}\quad i(t) = a(t)\cos\phi(t) \tag{1.5}$$

$$\text{and}\quad q(t) = a(t)\sin\phi(t) \tag{1.6}$$

We may obtain $s(t)$ in terms of i and q by substituting (1.4) into (1.2):

$$s(t) = Re[\{i(t) + jq(t)\}\ e^{j\omega_C t}] = i(t)\cos(\omega_C t) - q(t)\sin(\omega_C t) \tag{1.7}$$

Hence $i(t)$ is the **inphase** component of $s(t)$ and $-q(t)$ is the **quadrature** component.

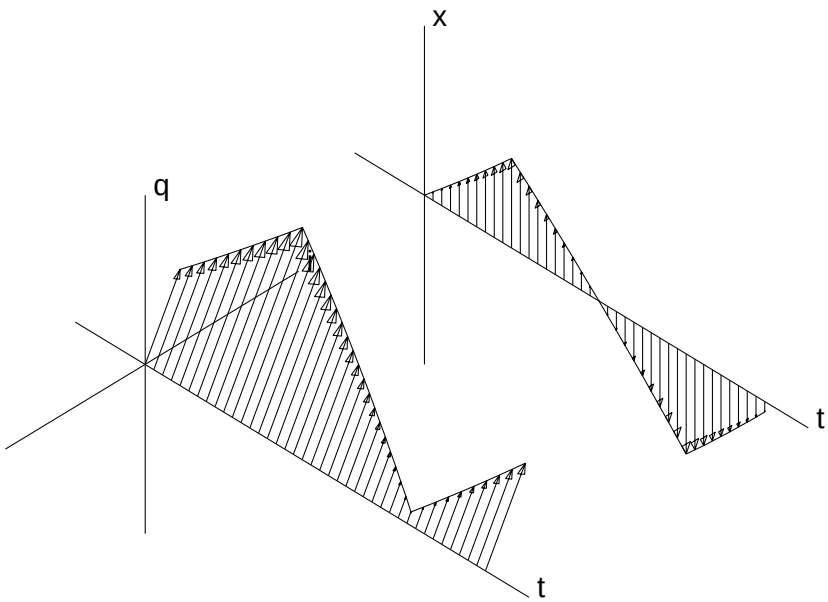

Fig 1.2: 3-D plot of AM phasor and input signal

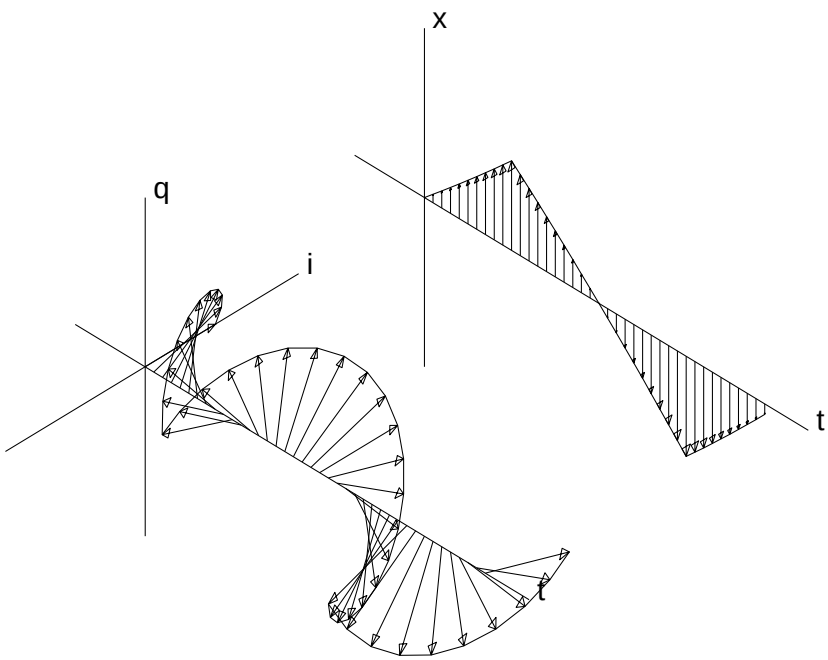

Fig 1.3: 3-D plot of PM phasor and input signal

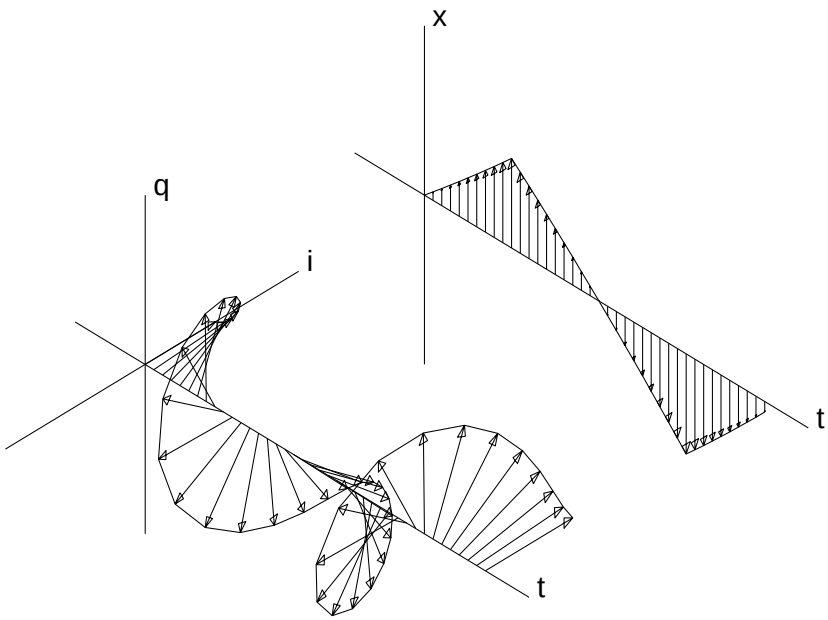

Fig 1.4: 3-D plot of FM phasor and input signal

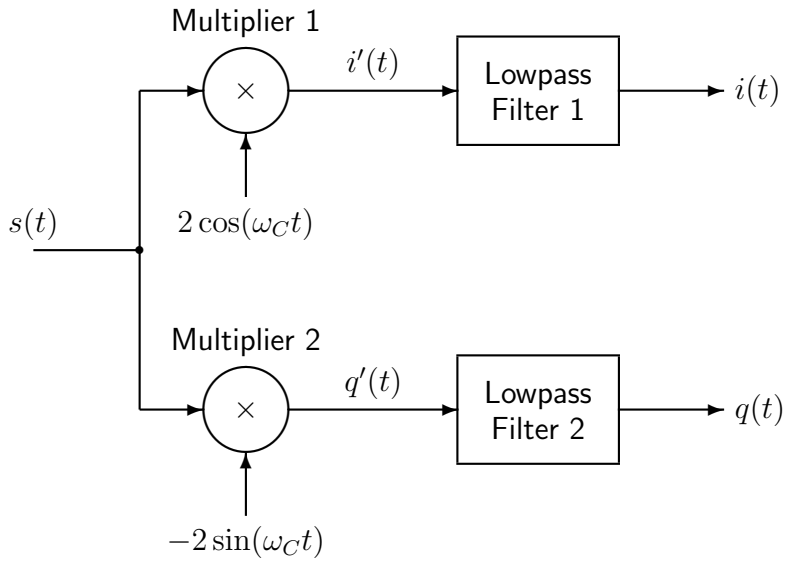

Fig 1.5: Quadrature Demodulator.

1.4 Quadrature Demodulation:

Given $s(t)$, how do we obtain $i(t)$ and $q(t)$ for the phasor waveform which generated s?

Fig 1.5 shows the **quadrature demodulator** which achieves this.

From multiplier 1:

$$\begin{aligned} i'(t) &= s(t) \times 2\cos(\omega_C t) \\ &= [i(t)\cos(\omega_C t) - q(t)\sin(\omega_C t)] \times 2\cos(\omega_C t) \\ &= 2i(t)\cos^2(\omega_C t) - 2q(t)\sin(\omega_C t)\cos(\omega_C t) \\ &= i(t) + i(t)\cos(2\omega_C t) - q(t)\sin(2\omega_C t) \end{aligned}$$

Hence the output of lowpass filter 1 is $i(t)$, since the two terms modulated onto carriers at $2\omega_C$ are rejected by the filter.

Similarly from multiplier 2:

$$\begin{aligned} q'(t) &= s(t) \times [-2\sin(\omega_C t)] \\ &= [i(t)\cos(\omega_C t) - q(t)\sin(\omega_C t)] \times [-2\sin(\omega_C t)] \\ &= -2i(t)\cos(\omega_C t)\sin(\omega_C t) + 2q(t)\sin^2(\omega_C t) \\ &= q(t) - q(t)\cos(2\omega_C t) - i(t)\sin(2\omega_C t) \end{aligned}$$

and the output of lowpass filter 2 is $q(t)$.

The quadrature demodulator is the basis for demodulating many types of modulation, although often only one output component is required, so only half of the demodulator is needed.

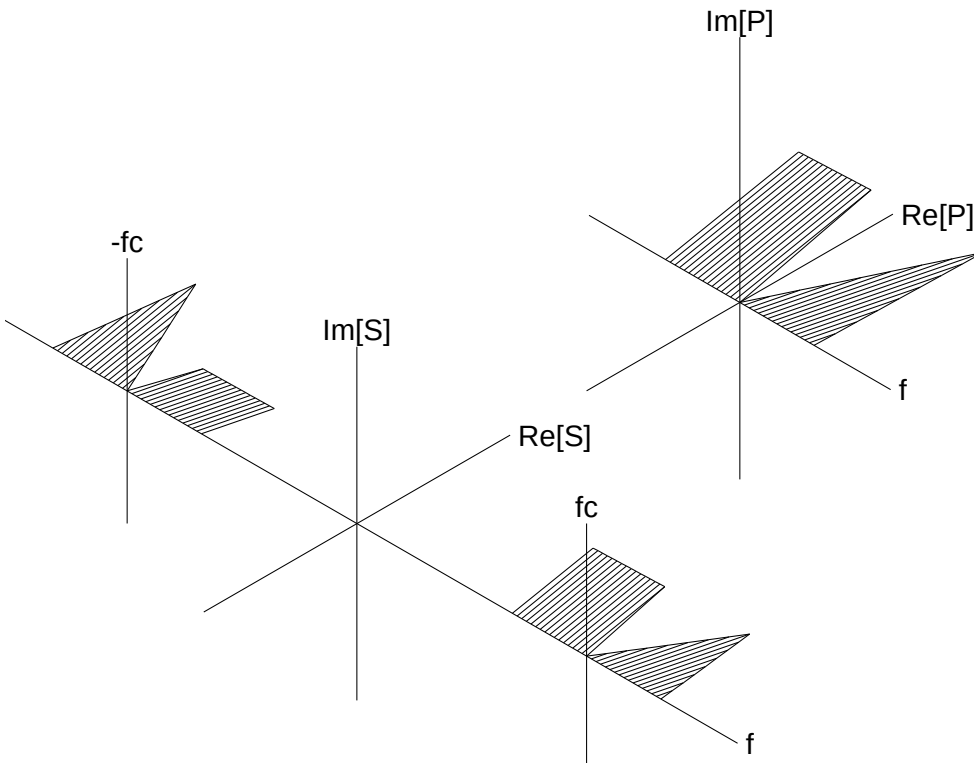

Fig 1.6: Phasor spectrum, P, and the corresponding signal spectrum, S.

1.5 Spectra of Phasors

To assess bandwidth, we relate the spectrum of any modulated signal $s(t)$ to the spectrum of its phasor waveform $p(t)$.

Expanding (1.2):

$$s(t) = Re[p(t)\ e^{j\omega_C t}] = \tfrac{1}{2}[p(t)\ e^{j\omega_C t} + p^*(t)\ e^{-j\omega_C t}] \tag{1.8}$$

$$\begin{aligned} \text{If}\ \ p(t) &\rightleftharpoons P(\omega) \qquad \text{then}\ \ p^*(t) \rightleftharpoons P^*(-\omega) \\ \text{By frequency shift:}\ \ p(t)\ e^{j\omega_C t} &\rightleftharpoons \int_{-\infty}^{\infty} p(t)\ e^{-j(\omega-\omega_C)t} dt = P(\omega-\omega_C) \\ \text{and}\ \ p^*(t)\ e^{-j\omega_C t} &\rightleftharpoons \int_{-\infty}^{\infty} p^*(t)\ e^{-j(\omega+\omega_C)t} dt = P^*(-(\omega+\omega_C)) \\ \therefore\ \text{taking transforms:}\ \ S(\omega) &= \tfrac{1}{2}[P(\omega-\omega_C) + P^*(-(\omega+\omega_C))] \end{aligned} \tag{1.9}$$

Fig 1.6 shows how an arbitrarily shaped phasor spectrum, $P(\omega)$, is transformed into the spectrum, $S(\omega)$, of the real modulated signal, $s(t)$. Note that:

1. Spectra are normally plotted in terms of frequency $f = \omega/(2\pi)$ rather than angular frequency ω, since it is usually more convenient to think of **bandwidth in Hz** rather than radians per second.

2. If $P(\omega)$ is narrowband compared with ω_C, then $S(\omega)$ comprises **two components** centred on ω_C and $-\omega_C$, such that $P(\omega)$ is shifted to ω_C and $P^*(-\omega)$ is shifted to $-\omega_C$.

3. $S(\omega)$ is a **linear function** of $P(\omega \pm \omega_C)$; i.e. if $P(\omega)$ comprises a linear sum of certain components, then $S(\omega)$ will comprise the same linear combination of components, shifted up and down by ω_C. Hence we can easily determine the spectrum of any modulated signal once we have calculated $P(\omega)$.

4. Since $p(t)$ is a complex waveform, $P(\omega)$ **need not possess any symmetries** between positive and negative frequencies. Hence the upper sidebands of $S(\omega)$ need not be related to its lower sidebands.

5. $P(\omega)$ depends only on the modulation method and the input signal – it is **independent of the carrier frequency** ω_C.

1.6 One-sided and Two-sided Spectra:

- Two-sided spectra include **negative as well as positive** frequencies and are used for complex waveforms (and optionally for real waveforms).
- One-sided spectra consider frequencies as being **purely positive** quantities and may only be used for real waveforms whose 2-sided spectra are symmetrical about zero.

Note on Negative Frequencies

A real cosine wave can be written: $\cos(\omega t) = \frac{1}{2}(e^{j\omega t} + e^{-j\omega t})$

If $\omega > 0$, then $e^{j\omega t}$ is the positive frequency component of the cosine wave, and $e^{-j\omega t}$ is the negative frequency component. These two components form two contra-rotating unit phasors which, when summed together, produce the purely real cosine wave.

A purely imaginary sine wave is produced if the phasors are subtracted, instead of added.

Lowpass Noise

A **real lowpass noise signal** containing frequencies from 0 to B Hz has a 1-sided spectrum from 0 to B Hz (fig 1.7a) and a 2-sided spectrum from $-B$ to $+B$ Hz (fig 1.7b).

Hence its 1-sided bandwidth = B Hz, and its 2-sided bandwidth = $2B$ Hz.

If it is a noise signal with a flat power spectrum and its power is P watt, then its 1-sided power spectral density (PSD) = P/B watt/Hz, and its 2-sided PSD = $P/2B$ watt/Hz. (Tuned voltmeters or spectrum analysers measure 1-sided PSD, since they respond to positive and negative frequency components, summed together.)

Bandpass Noise

A **real bandpass noise signal** is shown in figs 1.7c and 1.7d. Its spectrum extends from $f_C - B$ to $f_C + B$ Hz (and equivalent negative frequencies) and has zero energy elsewhere, so it has a 1-sided bandwidth of $2B$ Hz.

If the total noise power is P watt, the 1-sided PSD of the noise near f_C is $P/2B$ watt/Hz and the 2-sided PSD near $\pm f_C$ is $P/4B$ watt/Hz.

This bandlimited noise $s_N(t)$ may be represented by a carrier of frequency f_C, modulated by a complex noise phasor $p_N(t)$:

$$s_N(t) = Re[p_N(t)\ e^{j\omega_C t}] \tag{1.10}$$

Fig 1.7e shows the 2-sided spectrum $E\{|P_N(\omega)|^2\}$ of the phasor $p_N(t)$ which represents the bandlimited noise $s_N(t)$.

$p_N(t)$ has a flat spectrum from $-B$ to $+B$ Hz, and therefore a 2-sided bandwidth of $2B$ Hz. Since it is a complex phasor waveform, we only use 2-sided descriptions for its spectrum and bandwidth.

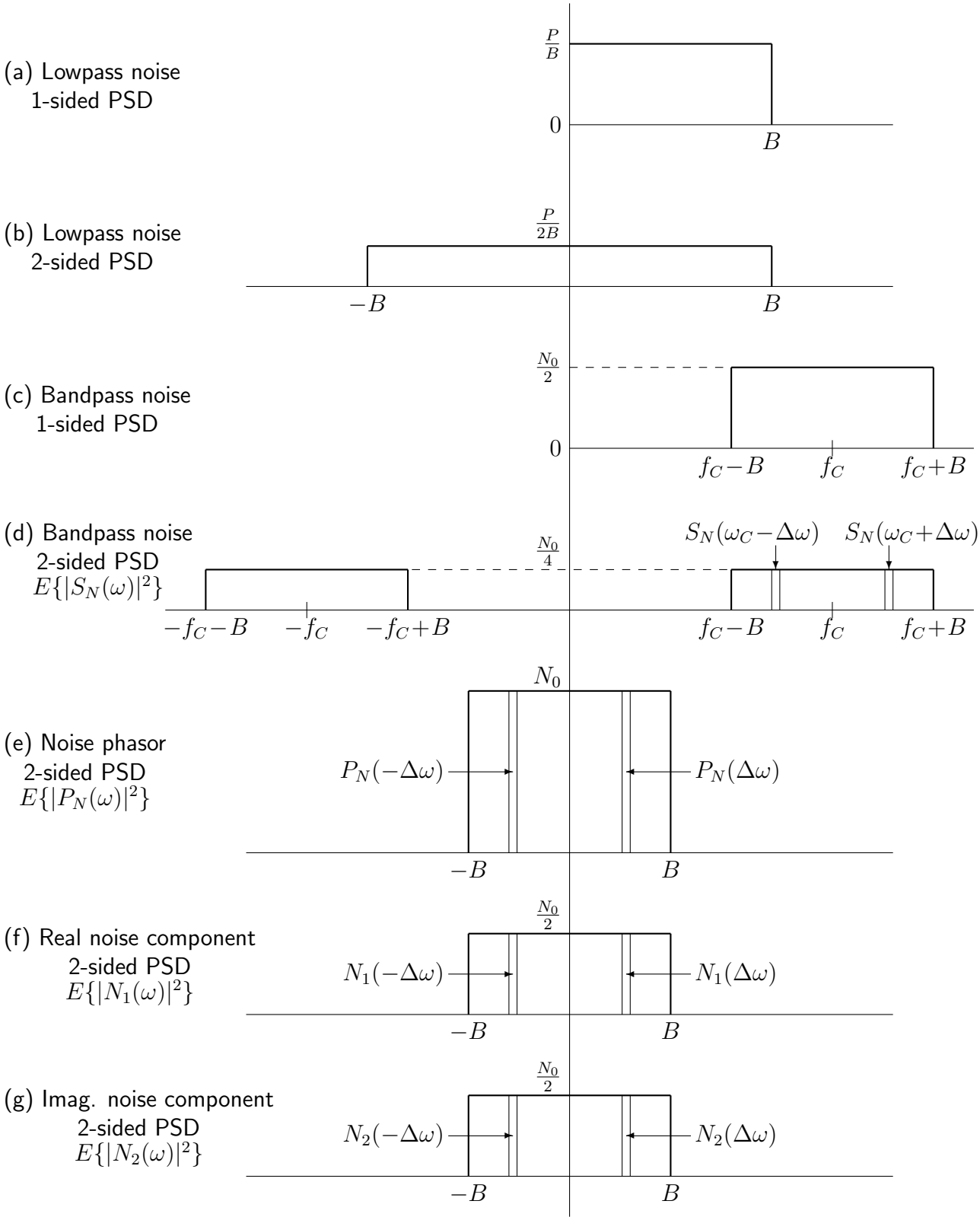

Fig 1.7: Power spectra of bandlimited noise and noise phasors.

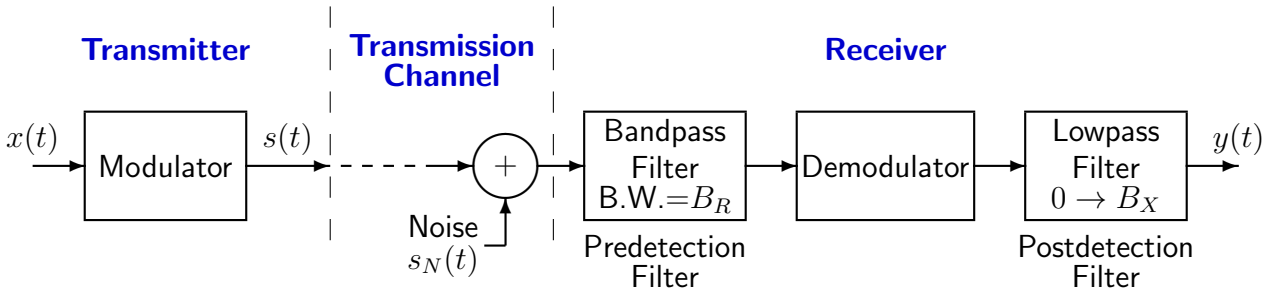

Fig 1.8: Simplified communication system.

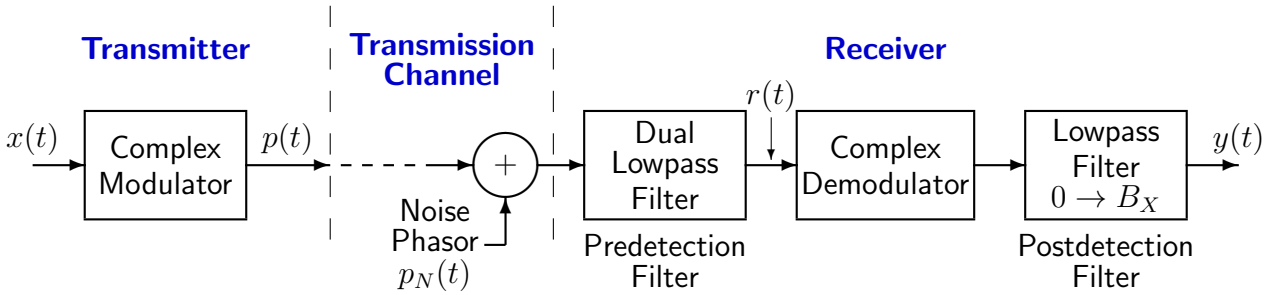

Fig 1.9: Complex baseband equivalent system.

1.7 Noise Phasors:

Figs 1.8 and 1.9 show a simplified comms. system and its baseband equivalent.

Just as the signal s can be represented by the phasor p, a bandlimited noise waveform s_N in the region of f_C can be represented by the noise phasor p_N.

Let:

$$p_N(t) = [n_1(t) + jn_2(t)]\ e^{j\phi_N} \tag{1.11}$$

where $n_1(t)$ and $n_2(t)$ are two real noise waveforms, and ϕ_N is an arbitrary constant phase offset. This process is illustrated in figures 1.10 to 1.15 . Figures 1.10 and 1.11 show two uncorrelated noise waveforms in real and imaginary directions, and figures 1.12 and 1.13 show the sum of these two waveforms in 2-D and 3-D respectively. Note the absence of any dominant direction to the sum, providing visual justification of the use of an arbitrary phase offset ϕ_N. Figures 1.14 and 1.15 show the effects of adding noise phasors to a constant phasor and the AM phasor of fig 1.2 .

We shall now show that if $n_1(t)$ and $n_2(t)$ are uncorrelated Gaussian processes with identical 1-sided power spectra, then $p_N(t)$ **accurately represents bandlimited gaussian noise** $s_N(t)$.

Let $n_1(t)$ and $n_2(t)$ each have a 2-sided noise PSD of $N_0/2$ watt/Hz from $-B$ to $+B$ Hz and zero elsewhere (see figs 1.7f and 1.7g).

This is expressed using expectations $E\{\}$, since noise is random, as:

$$E\{|N_1(\omega)|^2\} = E\{|N_2(\omega)|^2\} = \begin{cases} \dfrac{N_0}{2} & \text{for } |\omega| \le 2\pi B \\ 0 & \text{for } |\omega| > 2\pi B \end{cases}$$

Taking Fourier transforms of (1.11):

$$P_N(\omega) = [N_1(\omega) + jN_2(\omega)]\ e^{j\phi_N}$$

Since $n_1(t)$ and $n_2(t)$ are uncorrelated and Gaussian, the phases of $N_1(\omega)$ and $N_2(\omega)$ are uncorrelated and their powers add, so the PSD of p_N (fig 1.7e) is given by:

$$\begin{aligned} E\{|P_N(\omega)|^2\} &= E\{|N_1(\omega)|^2\} + E\{|N_2(\omega)|^2\} \\ &= \begin{cases} \frac{N_0}{2} + \frac{N_0}{2} = N_0 & \text{for } |\omega| \le 2\pi B \\ 0 & \text{for } |\omega| > 2\pi B \end{cases} \end{aligned}$$

Using (1.9) to get the spectrum of s_N:

$$S_N(\omega) = \tfrac{1}{2}[P_N(\omega - \omega_C) + P_N^*(-(\omega + \omega_C))]$$

If $B < f_C$, so that $P_N(\omega - \omega_C)$ does not overlap with $P_N^*(-(\omega + \omega_C))$:

$$\begin{aligned} E\{|S_N(\omega)|^2\} &= \tfrac{1}{4}E\{|P_N(\omega - \omega_C)|^2\} + \tfrac{1}{4}E\{|P_N(-(\omega + \omega_C))|^2\} \\ &= \begin{cases} \frac{N_0}{4} & \text{for } \omega_C - 2\pi B \le \omega \le \omega_C + 2\pi B \\ & \text{or } -\omega_C - 2\pi B \le \omega \le -\omega_C + 2\pi B \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

This gives the 2-sided PSD of s_N (fig 1.7d).

Thus its 1-sided PSD is $N_0/2$ watt/Hz from $f_C - B$ to $f_C + B$ Hz (fig 1.7c).

The need for $p_N(t)$ to be complex:

For $S_N(\omega)$ to be the spectrum of true bandlimited noise, $S_N(\omega_C - \Delta\omega)$ must be uncorrelated with $S_N(\omega_C + \Delta\omega)$ at any $\Delta\omega$.

Note the standard Fourier Transform result that, if $x(t)$ is purely real, its negative frequency components are the complex conjugates of its positive frequency components:

$$X(-\omega) = \int_{-\infty}^{\infty} x(t)\ e^{j\omega t}\ dt = \left(\int_{-\infty}^{\infty} x(t)\ e^{-j\omega t}\ dt\right)^* = X^*(\omega)$$

Hence if $p_N(t)$ were purely real:

$$P_N(-\Delta\omega) = P_N^*(\Delta\omega) \quad \text{and so} \quad S_N(\omega_C - \Delta\omega) = S_N^*(\omega_C + \Delta\omega)$$

showing undesirable strong correlation between $S_N(\omega_C - \Delta\omega)$ and $S_N(\omega_C + \Delta\omega)$.

Similarly there would be strong correlation if $p_N(t)$ were purely imaginary:

$$P_N(-\Delta\omega) = -P_N^*(\Delta\omega) \quad \text{and so} \quad S_N(\omega_C - \Delta\omega) = -S_N^*(\omega_C + \Delta\omega).$$

However if $p_N(t) = n_1(t) + jn_2(t)$ as proposed,

$$\begin{aligned} P_N(\Delta\omega) &= N_1(\Delta\omega) + jN_2(\Delta\omega) \\ \text{and } P_N(-\Delta\omega) &= N_1(-\Delta\omega) + jN_2(-\Delta\omega) \\ &= N_1^*(\Delta\omega) + jN_2^*(\Delta\omega) \end{aligned}$$

The cross-correlation between $P_N(\Delta\omega)$ and $P_N^*(-\Delta\omega)$ is proportional to

$$E\{P_N(\Delta\omega)P_N(-\Delta\omega)\}$$

Now:

$$\begin{aligned} P_N(\Delta\omega)P_N(-\Delta\omega) &= N_1(\Delta\omega)N_1^*(\Delta\omega) - N_2(\Delta\omega)N_2^*(\Delta\omega) \\ &\quad +j[N_1(\Delta\omega)N_2^*(\Delta\omega) + N_1^*(\Delta\omega)N_2(\Delta\omega)] \\ &= |N_1(\Delta\omega)|^2 - |N_2(\Delta\omega)|^2 + 2j\ Re[N_1(\Delta\omega)N_2^*(\Delta\omega)] \end{aligned}$$

The expected (or mean) value of each of the first 2 terms is $N_0/2$ and of the 3rd term is 0 since n_1 and n_2 are uncorrelated.

$$\therefore\ E\{P_N(\Delta\omega)P_N(-\Delta\omega)\} = \frac{N_0}{2} - \frac{N_0}{2} + 0 = 0$$

Hence there is no correlation between $P_N(\Delta\omega)$ and $P_N^*(-\Delta\omega)$, as desired.

Summary:

A real noise waveform $s_N(t)$, bandlimited from $(f_C - B)$ to $(f_C + B)$ Hz and with a 1-sided PSD of $N_0/2$ watt/Hz, may be represented by the noise phasor:

$$p_N(t) = [n_1(t) + jn_2(t)]\ e^{j\phi_N}$$

where $n_1(t)$ and $n_2(t)$ are real uncorrelated Gaussian noise waveforms of equal 2-sided PSD of $N_0/2$ watt/Hz, bandlimited from $-B$ to $+B$ Hz, and $p_N(t)$ has a 2-sided PSD of N_0 watt/Hz over the same bandwidth. The phase offset ϕ_N may be an arbitrary constant.

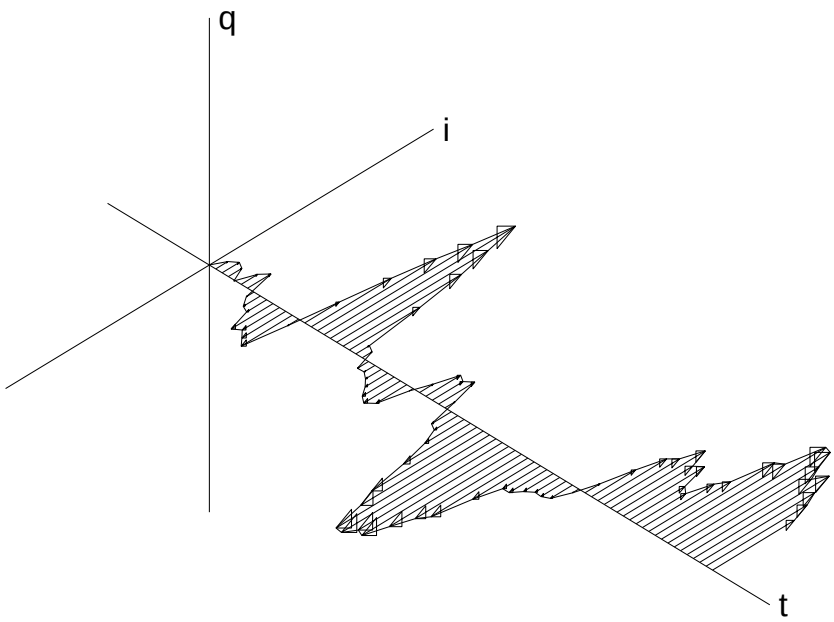

Fig 1.10: Real component of noise phasor.

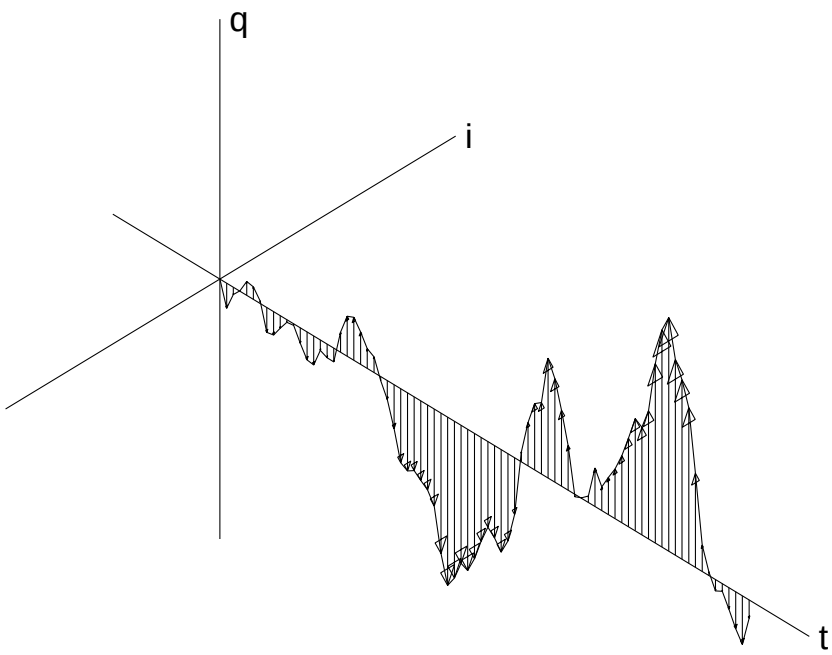

Fig 1.11: Imaginary component of noise phasor.

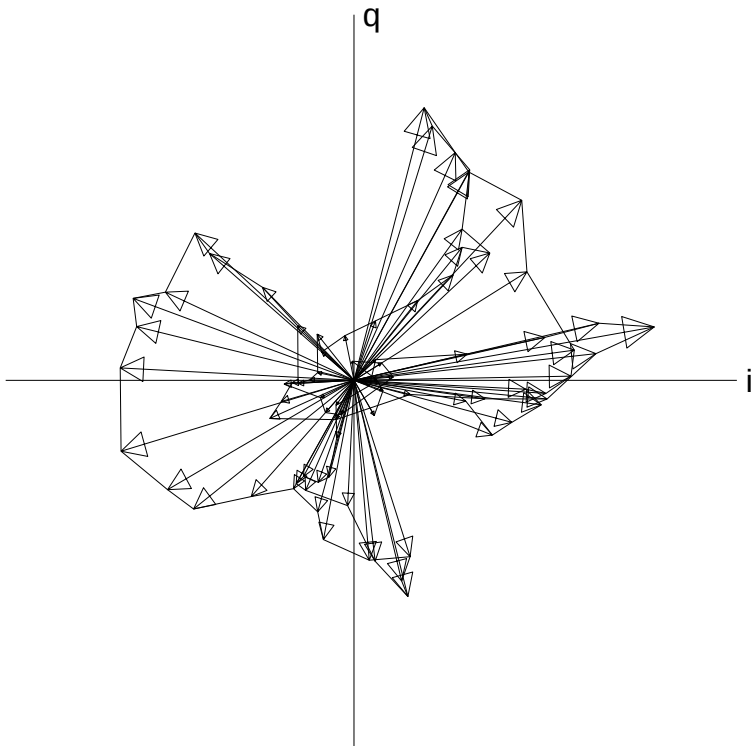

Fig 1.12: Complex noise phasor in 2−D.

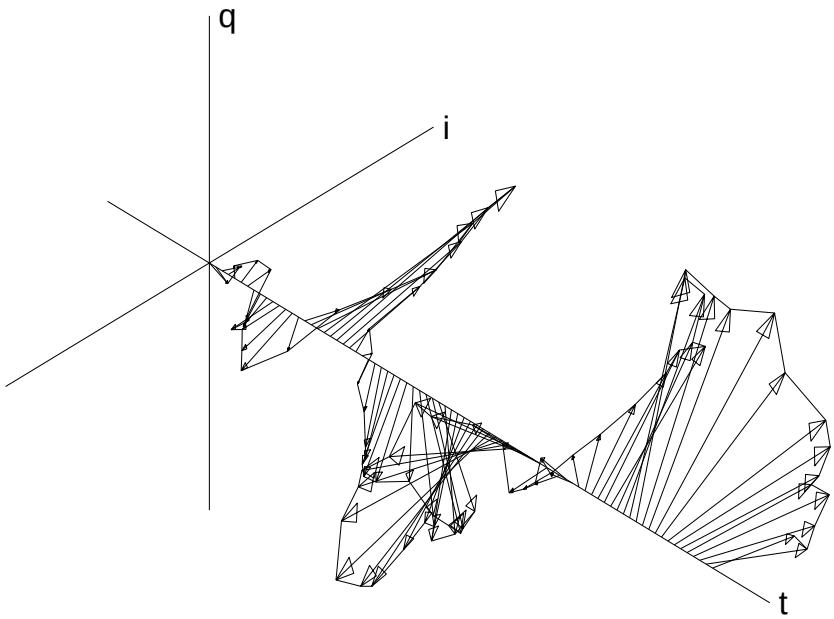

Fig 1.13: Complex noise phasor in 3−D.

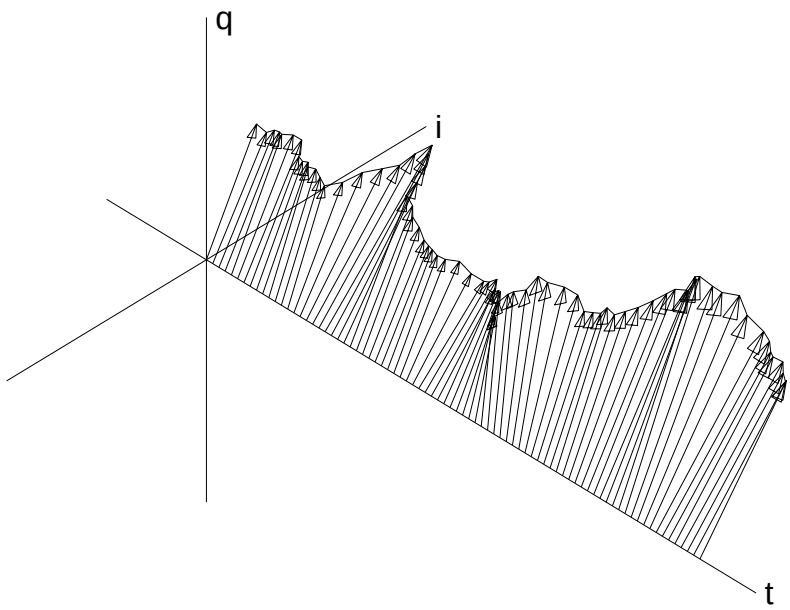

Fig 1.14: Constant signal phasor + (0.3 * Noise phasor).

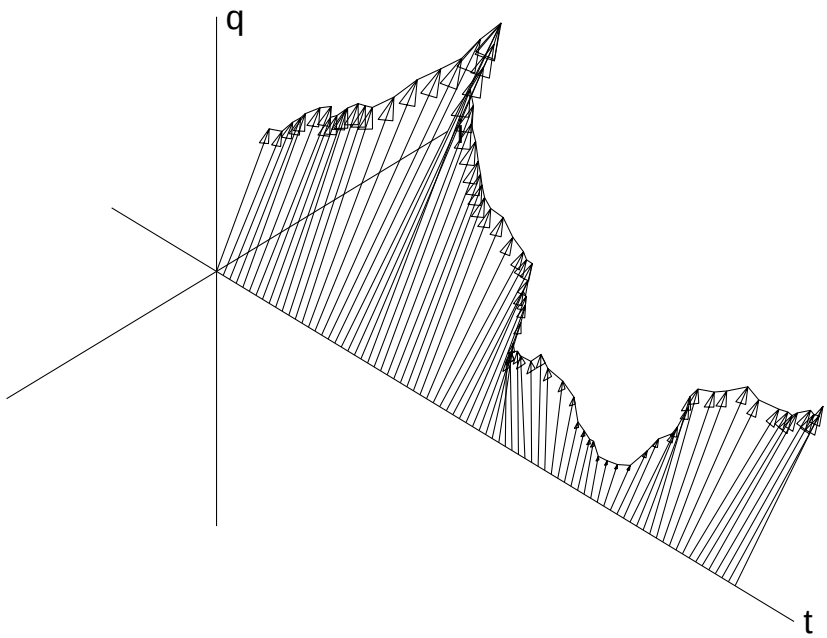

Fig 1.15: AM signal phasor + (0.3 * Noise phasor).

1.8 Signal/Noise Ratios for Signals and Phasors:

- How do the signal power P_S and 1-sided noise PSD N_S of the real signals $s(t)$ and $s_N(t)$ in fig 1.8 relate to the phasor power P_0 and 2-sided noise PSD N_0 for the phasors $p(t)$ and $p_N(t)$ of fig 1.9 ?

Let P_S be the power of $s(t)$.

$$\therefore \; P_S = \overline{(Re[p(t)\; e^{j\omega_C t}])^2} = \overline{[i\cos(\omega_C t) - q\sin(\omega_C t)]^2} \quad \text{if p = i + jq}$$

$$= \overline{[i\cos(\omega_C t)]^2} + \overline{[q\sin(\omega_C t)]^2} = \tfrac{1}{2}\overline{i^2} + \tfrac{1}{2}\overline{q^2} = \tfrac{1}{2}\overline{|p|^2} = \tfrac{1}{2}P_0$$

Let N_S be the 1-sided PSD of noise waveform $s_N(t)$.

From the above summary and figs 1.7c to 1.7e:

$$N_S = \tfrac{1}{2}N_0$$

$$\therefore \; \frac{P_0}{N_0} = \frac{2P_S}{2N_S} = \frac{P_S}{N_S}$$

Hence SNR is unaffected by conversion from real signals to phasors.

Real receiver amplifiers are characterised by their noise figure F, which is the ratio of the actual output noise PSD (due to amplifier noise and input resistance noise, but not noise from the antenna) to the ideal output noise PSD (assuming the amplifier is ideal and the only source of noise is the input resistance). The noise PSD from a resistor (of any value) at absolute temperature T is kT (where k is Boltzmann's constant), so the 1-sided noise PSD out of a real receiver amplifier (ignoring antenna noise which is usually small) is given by:

$$N_S = GN_{in} = GkTF \quad \text{(joule or watt/Hz)}$$

where G is the amplifier power gain and N_{in} is the effective noise PSD at the amplifier input. Note that here F should be expressed as a power ratio, and not be in dB.

The signal power out of the amplifier is:

$$P_S = GP_{in} \quad \text{(watt)}$$

where P_{in} is the receiver input signal power.

$$\therefore \; \frac{P_0}{N_0} = \frac{P_S}{N_S} = \frac{GP_{in}}{GkTF} = \frac{P_{in}}{kTF} \quad \text{(Hz)}$$

These formulae apply even when there is frequency translation and bandlimiting in the receiver, as long as the bandlimiting does not remove any of the wanted signal components.

2 Digital Modulation Summary

2.1 Why Digital (and not Analogue) ?

Main reasons:

- **Less susceptible to cumulative degradations.**
  Effects of additive noise can be eliminated at regular intervals in a digital comms link by threshold detection and error correction coding.
- Ultimate noise levels are determined by the analogue/digital conversions and not by the channel, giving **much better dynamic range** (e.g. audio CDs vs tape cassettes).
- Can be made **more secure** (encryption) and **less detectable** (spread spectrum).
- Can be multiplexed using frequency-division, time-division or spread-spectrum multiple access **(FDMA, TDMA or SSMA)**; whereas FDMA is the only sensible method for analogue signals.
- Digital links can handle a **wide range of source material** (multi-media):
  E.g. audio, video, still images and data, when used with appropriate source coding / compression techniques.
  (Note the variety of material now available on CDs and DVDs.)

Disadvantages:

- **More bandwidth** is needed unless source compression is used.
- **More complex processing** is required (OK with today's chips).

2.2 Some digital communications systems:

- Telegraph Communications, Morse code (1830s onwards).
  (Bell did not invent the analogue telephone until the 1870s!)
- Telex and teleprinters, 50 to 110 b/s (1930s?).
- Computer Modems, up to 2.4 kb/s over telephone network (1960s).
- Encrypted digital speech at 2.4 kb/s for military / strategic use (1960s).
- Satellite trunk telephone circuits, at 56 or 64 kb/s (1970s).
- Fax machines, binary images at 2.4 to 9.6 kb/s (1980s).
- Teletext, digital text on analogue TV transmissions (1980s).
- Digital Telephone Exchanges (system X in UK) (1980s).
- Nicam stereo digital sound on analogue TV (late 1980s).
- Fibre-optic internet links (e.g. Granta Backbone Network) (late 1980s)
- Internet modems, up to 56 kbit/s over telephone network (1990s).
- Digital cellular phones - GSM (Groupe Speciale Mobile) (1990s).
- Video phones / video conferencing (1990s).
- Digital Audio Broadcasting - DAB (late 1990s).
  See section 6 of this course.
- Digital TV (late 1990s).
  See section 6 of this course.
- Asymmetric internet modems (asymmetric digital subscriber loop, ADSL),
  $\sim 2 - 8$ Mbit/s one-way (early 2000s).
- Internet (WAP and 3rd Generation) mobile phones (early 2000s).
- WiFi comms for laptops etc (early 2000s).
- Digital High Definition TV (HDTV) (mid 2000s).
- Analogue TV switch-off in UK, to release more bandwidth / capacity for digital TV and radio services (2008 onwards, depending on region)

(See Couch Chapter 1 for a more complete list.)

2.3 The Major Digital Modulation Techniques

Figure 2.1 shows the modulated waveforms for various simple forms of digital modulation, with the input data and the carrier wave shown at the top. We shall now look at how these modulation methods are defined, and how they are represented by phasor waveforms.

Binary Modulation:

- **Binary Amplitude Shift Keying (ASK)** – fig 2.2
  2 amplitude levels (usually 0 and a_0).
  Phase constant ($= \phi_0$).
- **Binary Phase-Shift Keying (BPSK)** – fig 2.3
  Amplitude constant ($= a_0$).
  2 phase values (usually ϕ_0 and $\phi_0 + \pi$).
- **Binary Frequency-Shift Keying (BFSK)** – fig 2.4
  Amplitude constant ($= a_0$).
  2 frequency values ($\Delta\omega = \pm\omega_D$, the frequency deviation).

Multi-level (M-ary) Modulation:

Involves grouping data bits into m-bit symbols, and each symbol is transmitted using one of $M = 2^m$ modulation levels.

- **4-level Amplitude Shift Keying (4ASK)** – fig 2.5
  4 amplitude levels (usually $-3a_0$, $-a_0$, $+a_0$ and $+3a_0$).
  Phase constant ($= \phi_0$).
  $m = 2$ bits per symbol.
- **Quadrature Phase-Shift Keying (QPSK)** – fig 2.6
  Amplitude constant ($= a_0$).
  4 phase values (usually $\phi_0 \pm \pi/4$ and $\phi_0 \pm 3\pi/4$).
  Equivalent to BPSK on two quadrature carriers at ϕ_0 and $\phi_0 + \pi/2$.
  $m = 2$ bits per symbol.
- **M-ary PSK** –
  Amplitude constant ($= a_0$).
  M phase values (usually $\phi_0 + \dfrac{2\pi k}{M}$, $\quad k = 0 \rightarrow M-1$).
- **Quadrature Amplitude Modulation (QAM)** –
  M-level ASK on each of two quadrature carriers.
  This gives M^2 states, representing $2m$ bits.
- **M-ary FSK** –
  Amplitude constant ($= a_0$).
  M frequency values (usually $\Delta\omega = (k - \frac{M-1}{2})\omega_S$,
  where $k = 0 \rightarrow M-1$ and $\omega_S = 2\pi(\text{symbol rate}) = \dfrac{2\pi(\text{bit rate})}{m}$).

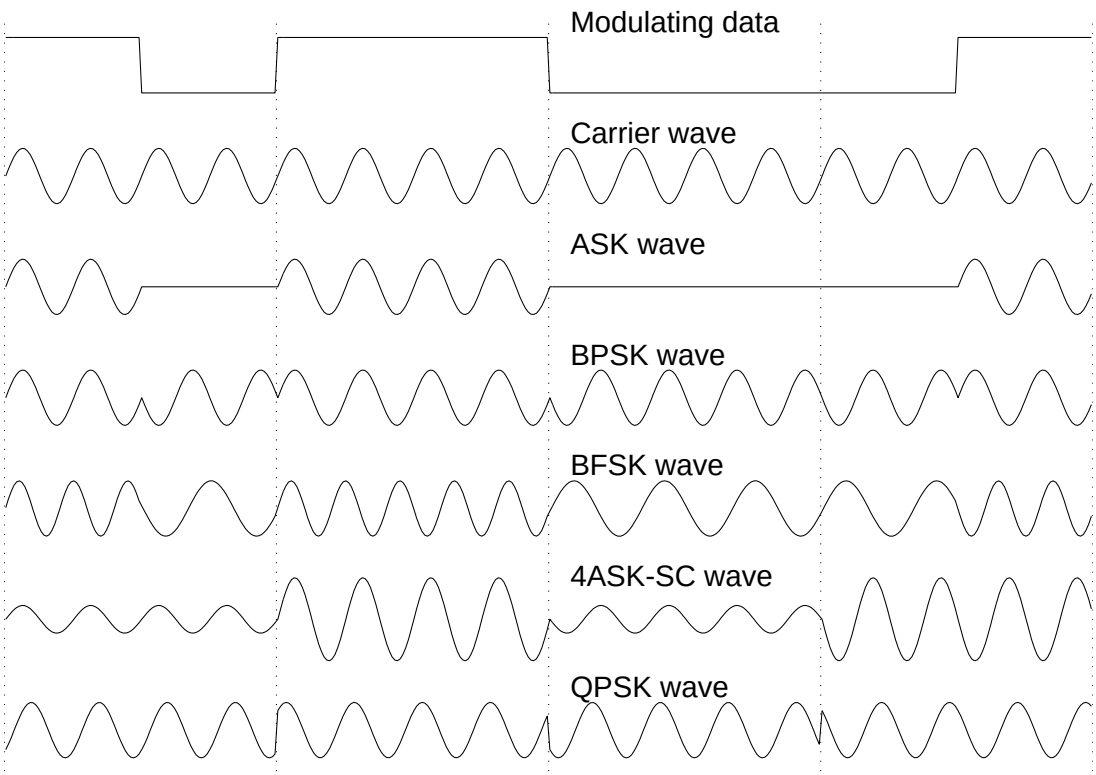

Fig 2.1: Basic digital modulation schemes

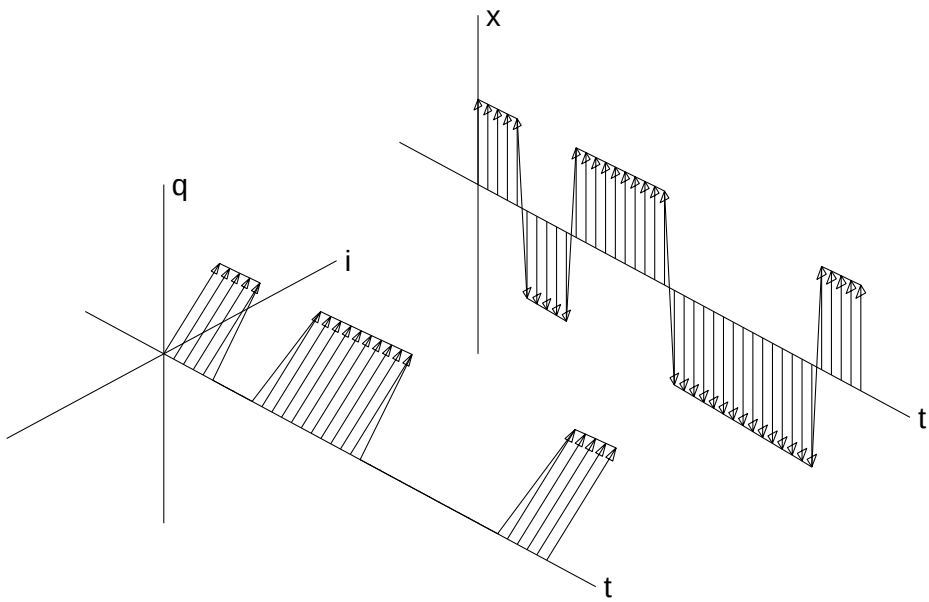

Fig 2.2: Binary amplitude shift keying (ASK) phasor waveform

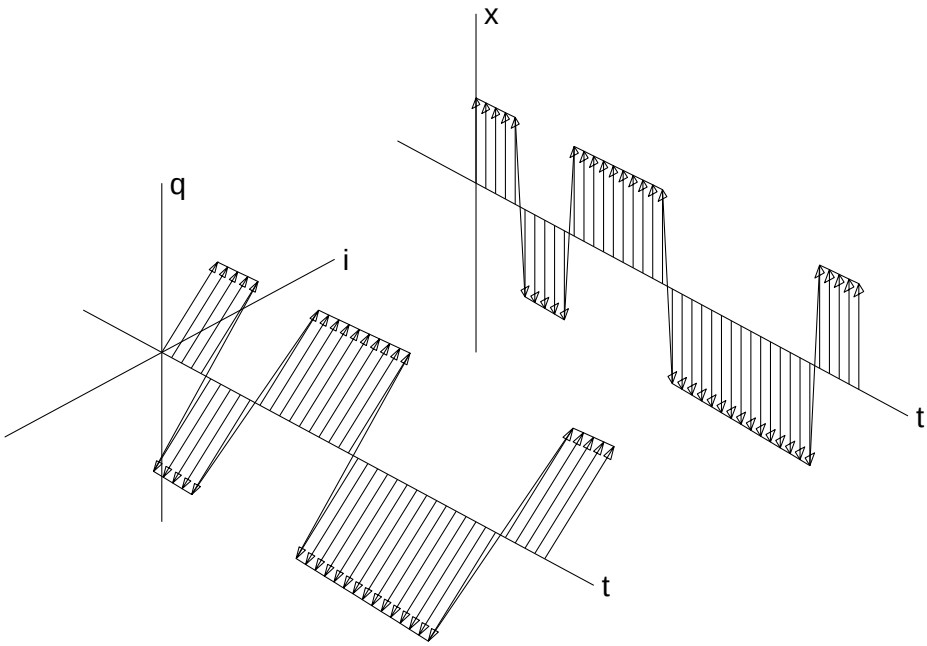

Fig 2.3: Binary phase shift keying (BPSK) phasor waveform

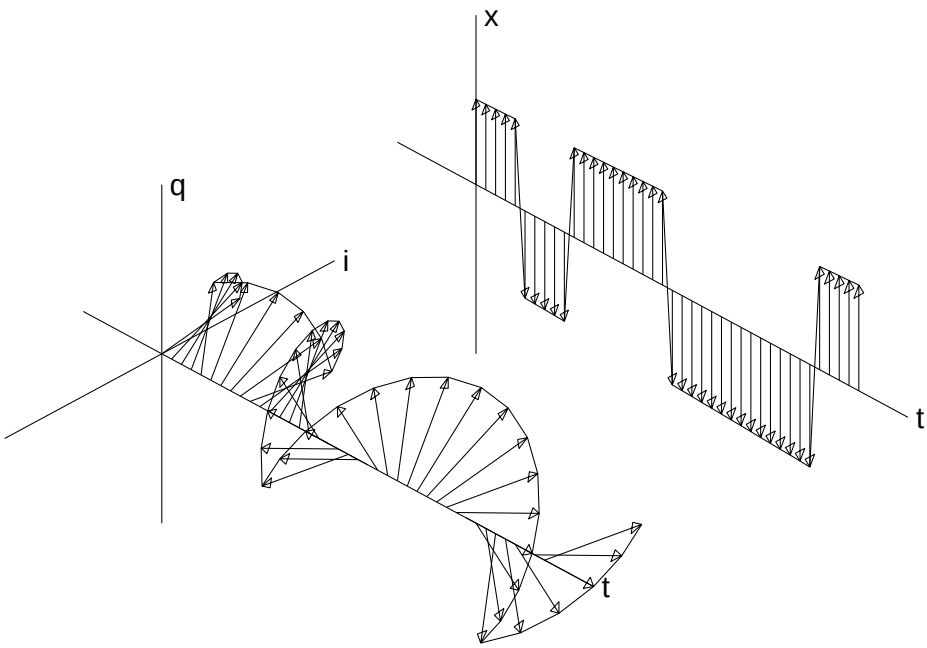

Fig 2.4: Binary frequency shift keying (BFSK) phasor waveform

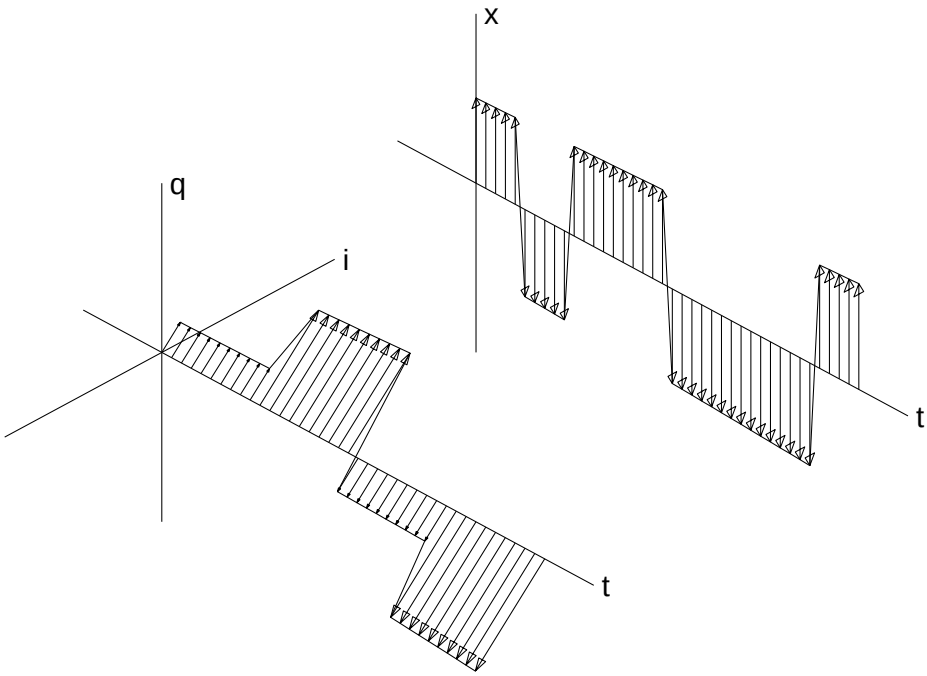

Fig 2.5: 4-level ASK - suppressed carrier (4ASK-SC) phasor waveform

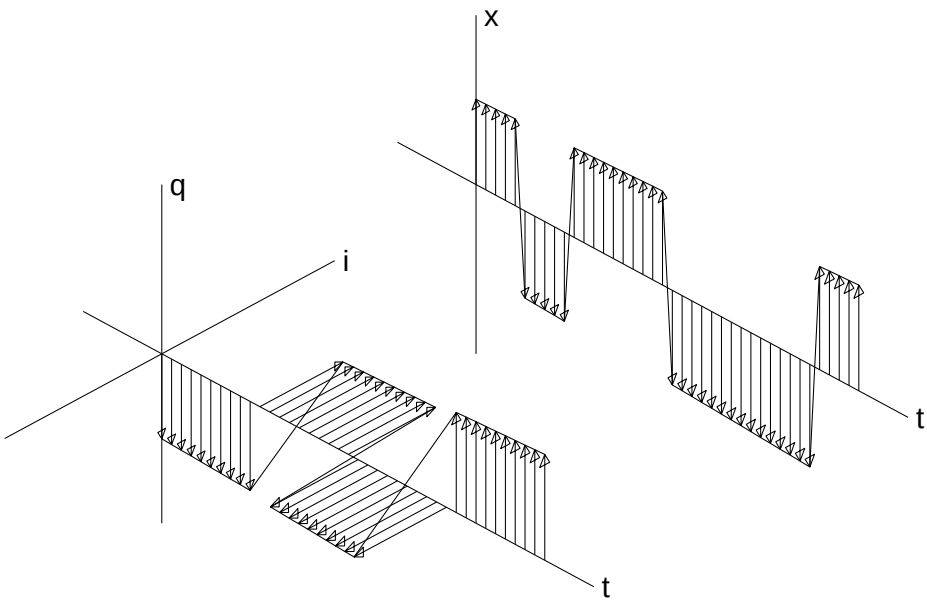

Fig 2.6: Quadrature phase shift keying (QPSK) phasor waveform

3 Binary Phase-Shift Keying (BPSK)

BPSK is perhaps the simplest modulation scheme in common use and we shall analyse it in some detail. The results for other schemes can often be derived from those for BPSK.

3.1 Definition of BPSK:

Let k^{th} data bit $b_k = +1$ or -1 and bit period $= T_b$.

The modulated phasor during the k^{th} bit period is:

$$p_k(t) = b_k \; a_0 \; e^{j\phi_0} \qquad \text{for } kT_b \le t < (k+1)T_b$$

To apply this for all t, we introduce a time limited pulse:

$$g(t) = \begin{cases} a_0 & \text{for } 0 \le t < T_b \\ 0 & \text{elsewhere.} \end{cases}$$

$$\therefore \; p(t) = e^{j\phi_0} \sum_k b_k \; g(t - kT_b) \tag{3.1}$$

Note that $g(t)$ is normally a rectangular pulse, but modified forms of BPSK can use other shapes such as half-sine or raised cosine.

3.2 Power Spectrum for Random Data:

We observe that $p(t)$ is just a constant phasor $e^{j\phi_0}$ multiplied by a polar binary data stream, in which the data impulses have been filtered (convolved) with an impulse response $g(t)$.

Hence we use techniques similar to those introduced in the 3F4 Baseband Transmission course for power spectra for line codes.

The discrete autocorrelation function (ACF) of the random data stream b_k, in which b_k is uncorrelated with b_{k-L} for any integer $L \neq 0$, is:

$$R_{bb}(L) = E\{b_k \; b_{k-L}\} = \begin{cases} 1 & \text{for } L = 0 \\ 0 & \text{for } L \neq 0 \end{cases}$$

The power spectrum (power spectral density) of the stream of random data impulses, $b(t) = \sum_k b_k \delta(t - kT_b)$, is then given by:

$$E\{|B(\omega)|^2\} = \lim_{T\to\infty} \frac{E\{|B_T(\omega)|^2\}}{T} = \frac{1}{T_b} \sum_L R_{bb}(L) \; e^{jL\omega T_b} = \frac{1}{T_b}$$

where $B_T(\omega)$ is the Fourier transform of $b(t)$ limited to a time interval of T.

Since the stream of data impulses $b(t) = \sum_k b_k \delta(t - kT_b)$ are convolved with $g(t)$ and phase-rotated by ϕ_0 in equation (3.1), the Fourier transform of $p(t)$ is given by:

$$P(\omega) = e^{j\phi_0} B(\omega)\ G(\omega)$$

So the power spectrum of $p(t)$ is:

$$E\{|P(\omega)|^2\} = |e^{j\phi_0}|^2\ E\{|B(\omega)|^2\}\ |G(\omega)|^2 = \frac{1}{T_b}\ |G(\omega)|^2$$

If $g(t)$ is a rectangular pulse of amplitude a_0 and duration T_b:

$$|G(\omega)|^2 = \left[a_0\, T_b\ \text{sinc}\left(\frac{\omega T_b}{2}\right)\right]^2 \quad \text{and so} \quad E\{|P(\omega)|^2\} = a_0^2\, T_b\ \text{sinc}^2\left(\frac{\omega T_b}{2}\right)$$

For alternative $g(t)$ pulse shapes, the power spectrum will be proportional to $|G(\omega)|^2$, so this gives a good way of controlling the spectrum of the modulated signal.

The power spectrum of the modulated signal $|S(\omega)|^2$ is just a frequency shifted version of $|P(\omega)|^2$ as derived in equation (1.9):

$$S(\omega) = \tfrac{1}{2}[P(\omega - \omega_C) + P^*(-(\omega + \omega_C))]$$

Hence, assuming no overlap between $P(\omega - \omega_C)$ and $P(-(\omega + \omega_C))$:

$$E\{|S(\omega)|^2\} = \tfrac{1}{4}[\ E\{|P(\omega - \omega_C)|^2\} + E\{|P(-(\omega + \omega_C))|^2\}\ ]$$

Fig 3.1 shows these spectra for rectangular data pulses.

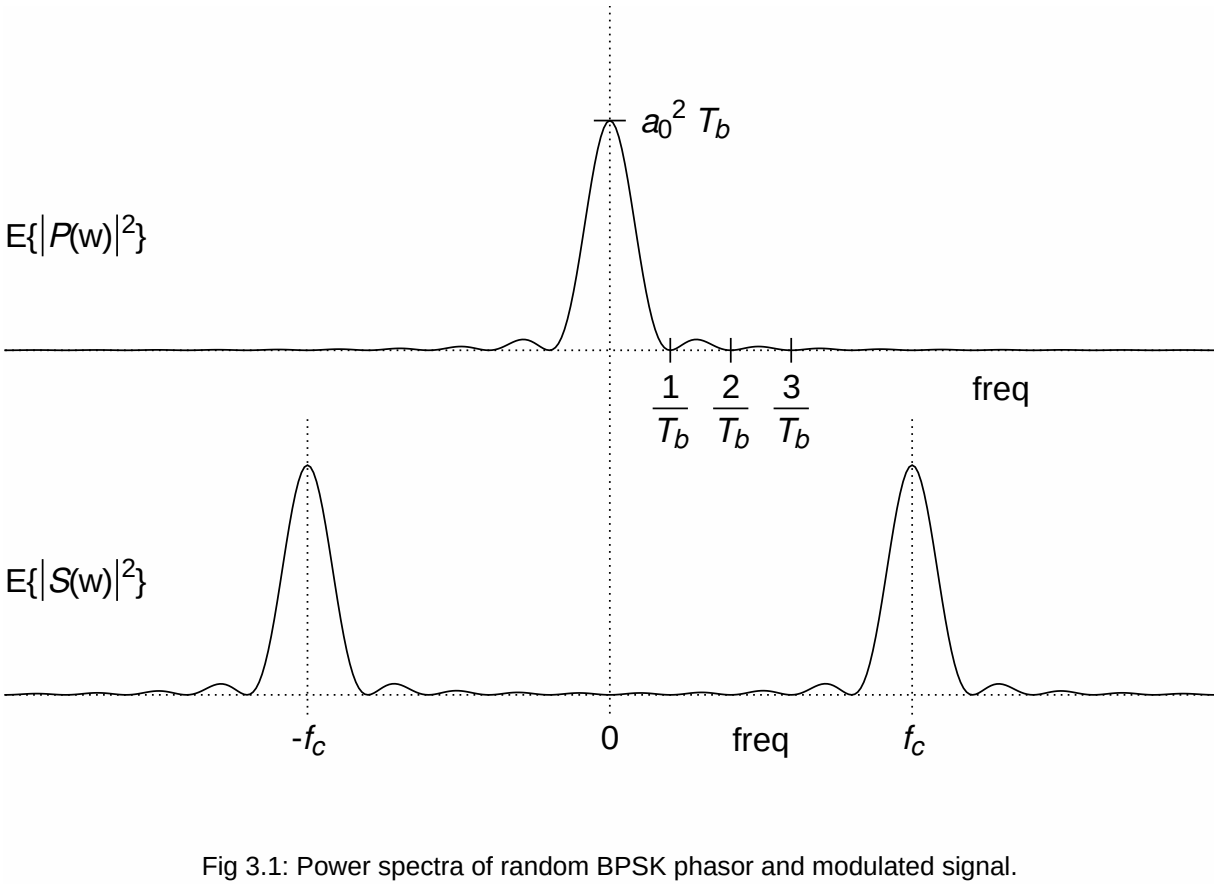

Fig 3.1: Power spectra of random BPSK phasor and modulated signal.

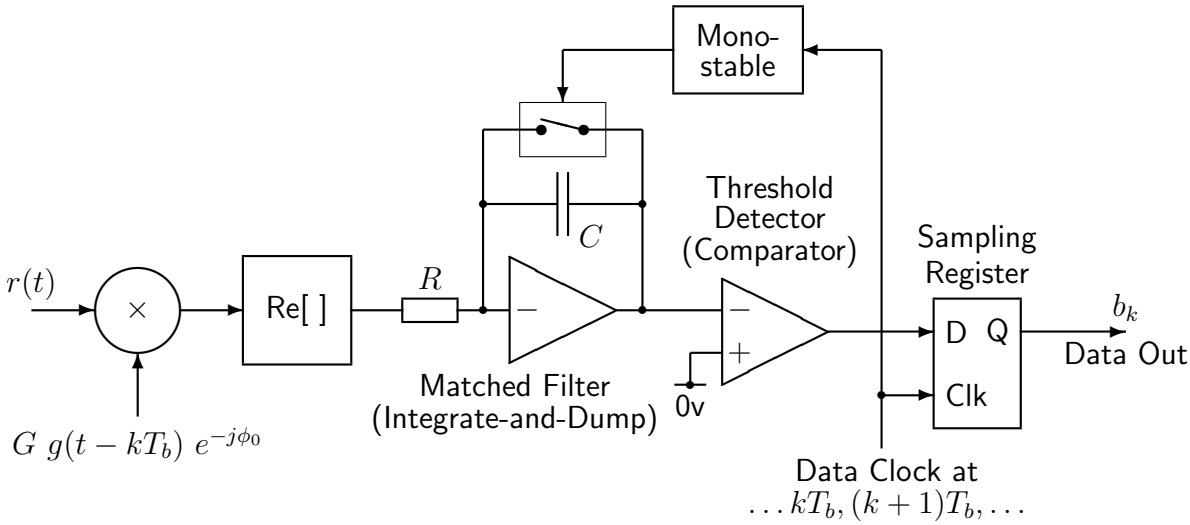

Fig 3.2: Matched correlator demodulator for BPSK.

3.3 Optimum Demodulator:

An optimum demodulator for equiprobable signals in gaussian noise is one which selects the data value that results in the **minimum mean squared error (MSE)** between the received signal plus noise and the signal corresponding to that data value. (Bayesian analysis can be used to prove this.)

If the received signal+noise phasor is

$$r(t) = p(t) + p_N(t)$$

the optimum demodulator for binary data during the k^{th} bit period measures the MSEs:

$$\int_{kT_b}^{(k+1)T_b} |r(t) - \underbrace{g(t-kT_b)e^{j\phi_0}}_{+1 \text{ phasor}}|^2 dt \quad \text{and} \quad \int_{kT_b}^{(k+1)T_b} |r(t) - \underbrace{(-g(t-kT_b)e^{j\phi_0})}_{-1 \text{ phasor}}|^2 dt$$

and selects whichever gives the **smaller** result.

But we can simplify this.

$$|r-g|^2 = (r-g)(r^*-g^*) = |r|^2 - 2Re[rg^*] + |g|^2$$

If $g = \pm g(t-kT_b)e^{j\phi_0}$, $|r|^2$ and $|g|^2$ are the same for both integrals, so we may ignore these terms and simply calculate

$$y(b,k) = \int_{kT_b}^{(k+1)T_b} Re[r(t)\ b\ g(t-kT_b)e^{-j\phi_0}]dt$$

for $b = +1$ and -1, and select the value of b according to which b gives the **larger** result.

Since b can be taken out of the integral, we find that $y(-1,k) = -y(1,k)$; and so, detecting the larger of $y(1,k)$ and $y(-1,k)$ is equivalent to detecting the polarity of just $y(1,k)$.

Thus b_k = polarity of $y(k)$, where

$$y(k) = G \int_{kT_b}^{(k+1)T_b} Re[r(t)\ g(t - kT_b)e^{-j\phi_0}]dt \tag{3.2}$$

and G is an arbitrary positive gain constant.

This is known as a **matched correlation demodulator**, because it correlates $r(t)$ with a waveform $g(t - kT_b)e^{j\phi_0}$ that is matched to the data pulse being detected.

To work correctly, the demodulator must know the reference phase ϕ_0 of the received phasors. This is normally extracted by a phase-locked loop (see section 3.6).

A demodulator using **phasors** for the input and reference signals is shown in fig 3.2. It implements equation 3.2. The correct limits for the integration are achieved by discharging the capacitor at the start of each bit period (at $t = kT_b$ using a short pulse from the monostable), and by sampling the polarity of the integrator output at the end of each bit period (at $t = (k + 1)T_b$) just before the monostable discharges the integrator for the next bit.

Note that if the data pulses $g(t)$ are rectangular and constant over the period of integration, we may replace $g(t - kT_b)$ by a_0 in equation 3.2, and the input multiplier in fig 3.2 simply produces a phase shift ϕ_0 and an arbitrary gain Ga_0.

A more practical implementation of an optimum demodulator, using real bandpass signals rather than phasors, is described in section 3.6.

3.4 Bit Error Performance in Noise:

The signal+noise phasor received during the k^{th} bit period is

$$r(t) = p(t) + p_N(t) = b_k\ g(t - kT_b)\ e^{j\phi_0} + [n_1(t) + jn_2(t)]\ e^{j\phi_N}$$

where $n_1(t)$ and $n_2(t)$ are real uncorrelated noise waveforms of equal PSDs.

If we choose the arbitrary noise reference phase ϕ_N to equal ϕ_0 (as discussed in section 1.7), the output of the matched correlator is:

$$\begin{aligned} y(k) &= G\int_{kT_b}^{(k+1)T_b} Re[\{b_k g(t-kT_b) + n_1(t) + jn_2(t)\}\, e^{j\phi_0}\ g(t-kT_b)\, e^{-j\phi_0}]\ dt \\ &= \underbrace{Gb_k\int_0^{T_b} g^2(t)\ dt}_{\text{signal}} + \underbrace{G\int_{kT_b}^{(k+1)T_b} n_1(t)\ g(t-kT_b)\ dt}_{\text{noise}} \end{aligned}$$

Now the noise integral is equivalent to convolving the noise with a filter whose impulse response $h(t)$ equals $g(T_b - t)$ from 0 to T_b and is zero elsewhere. This is in fact the matched filter for the data.

i.e.

$$\int_{kT_b}^{(k+1)T_b} n_1(t)\ g(t-kT_b)\ dt = \int_{-\infty}^{\infty} n_1(t)\ h((k+1)T_b - t)\ dt$$

Let the 2-sided PSD of $p_N(t)$ be N_0 watt/Hz.

Hence n_1 and n_2 will need to be real uncorrelated white noise waveforms with 2-sided PSDs of $N_0/2$ watt/Hz $= N_0/(4\pi)$ watt s/rad (see section 1.7).

If $h(t) \rightleftharpoons H(\omega)$:

$$\text{Mean noise power from filter} = G^2\int_{-\infty}^{\infty} E\{|N_1(\omega)\ H(\omega)|^2\}\ d\omega$$

$$= G^2\int_{-\infty}^{\infty}\frac{N_0}{4\pi}\,|H(\omega)|^2\ d\omega = \frac{G^2N_0}{2}\int_{-\infty}^{\infty}|h(t)|^2\ dt \quad \text{(by Parseval's Theorem)}$$

$$= \frac{G^2N_0}{2}\int_0^{T_b} g(t)^2\ dt = \frac{G^2N_0E_b}{2}$$

where $E_b = P_0T_b$ is the energy of the $g(t)$ pulse, or the energy per bit of the signal phasor.

$$\text{Hence rms noise voltage } \sigma \text{ in } y(k) = G\sqrt{\frac{N_0E_b}{2}}$$

$$\text{Signal voltage amplitude } v_s \text{ in } y(k) = \left|Gb_k\int_0^{T_b} g^2(t)\ dt\right| = GE_b$$

The detector threshold = 0 since the signal voltage $= \pm GE_b$, which is symmetric about zero.

$$\therefore\ \text{Voltage SNR at threshold detector} = \frac{v_s - 0}{\sigma} = \frac{GE_b}{G\sqrt{N_0E_b/2}} = \sqrt{\frac{2E_b}{N_0}}$$

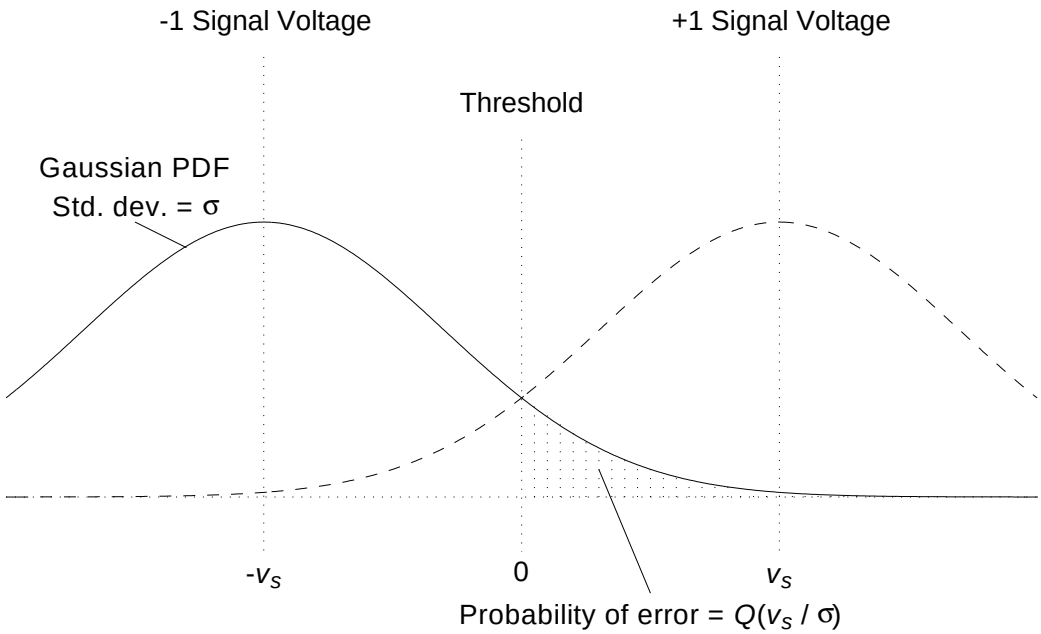

Fig 3.3: Probability density functions of detected signal+noise.

We can now calculate the probability of bit error – see fig 3.3. We assume the noise PDF is gaussian, since it is heavily bandlimited and assume equal probability of 1 and -1 in b_k.

$$\text{PDF of noise with rms voltage } \sigma = \frac{1}{\sigma}\, f\left(\frac{v}{\sigma}\right)$$

where $f(x) = \dfrac{e^{-x^2/2}}{\sqrt{2\pi}}$ is the PDF of a unit variance zero mean gaussian process and $x = \dfrac{v}{\sigma}$.

$$\begin{aligned}\text{Probability of error} &= \int_0^\infty \frac{1}{\sigma} f\left(\frac{v+v_s}{\sigma}\right) dv \;=\; \int_{v_s}^\infty \frac{1}{\sigma} f\left(\frac{v}{\sigma}\right) dv \\ &= \int_{v_s/\sigma}^\infty f(x)\, dx \;=\; Q\left(\frac{v_s}{\sigma}\right)\end{aligned}$$

where $Q\left(\dfrac{v_s}{\sigma}\right)$ is the gaussian integral function (see section 3.7 on Approximation Formulae for $Q(x)$) and $\dfrac{v_s}{\sigma}$ is the voltage SNR at the threshold detector.

$$\therefore\ \text{Probability of a bit error } P_E = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$

This function is plotted in fig 3.5, as the BPSK (and QPSK) curve.

Note that this result is independent of the shape of the signal pulse $g(t)$. Only the energy E_b of the pulse and the PSD N_0 of the noise are important.

For a rectangular pulse of amplitude a_0 and duration T_b, $E_b = a_0^2\, T_b$.

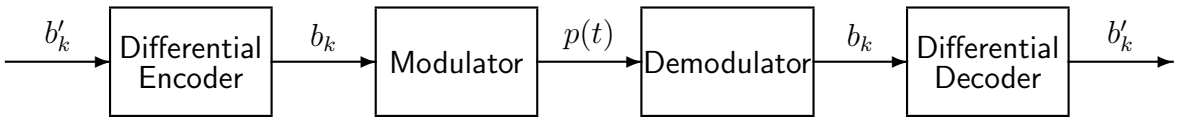

Fig 3.4: Differential encoding and decoding.

3.5 Differential Coding:

A real BPSK receiver cannot detect which of the two received carrier phases is 0° and which is 180°, since there is nothing to distinguish them in the presence of an arbitrary phase offset ϕ_0 (see fig 2.3). Hence the receiver can lock to the **wrong signal phase**, and thus generate inverted data.

$\therefore$ Differential Coding is used (see fig. 3.4).

At the Diff. Encoder:

$$b_k = b_{k-1}\ b'_k$$

Hence b_k changes state if $b'_k = -1$, and remains in its previous state if $b'_k = 1$.

At the Diff. Decoder:

$$b'_k = b_k / b_{k-1} = b_k\ b_{k-1} \qquad \text{since } b_{k-1} = \pm 1$$

Hence $b'_k = -1$ if b_k and b_{k-1} differ, and $b'_k = 1$ if they are the same.

If b_k and b_{k-1} are both inverted, b'_k will still be decoded correctly.

$\therefore$ 180° phase ambiguity does not matter.

BUT differential coding **increases the output bit error rate**.

If the probability of error in b_k is P_E for all k, then b'_k will be in error if:

b_k is in error and b_{k-1} is correct,

or

b_{k-1} is in error and b_k is correct.

Each event has probability $P_E(1 - P_E)$ and they are mutually exclusive,

$$\therefore \text{ Probability of error in } b'_k = 2P_E(1 - P_E) \approx 2P_E \quad \text{if } P_E \text{ is small.}$$

Hence the error rate is approximately doubled by differential coding and errors tend to occur in pairs. However only a small increase in SNR ($\sim$ 0.5dB) is needed to compensate for this – see fig 3.5, curves for BPSK with and without differential coding.

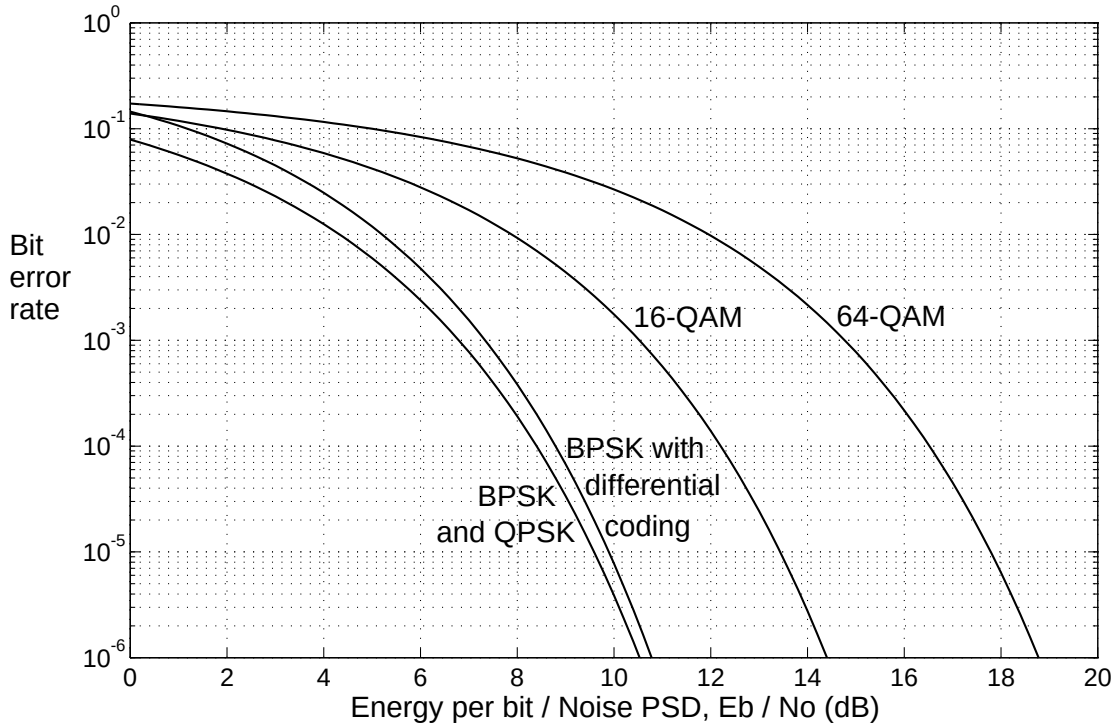

Fig 3.5: Bit error rate curves for PSK and QAM.

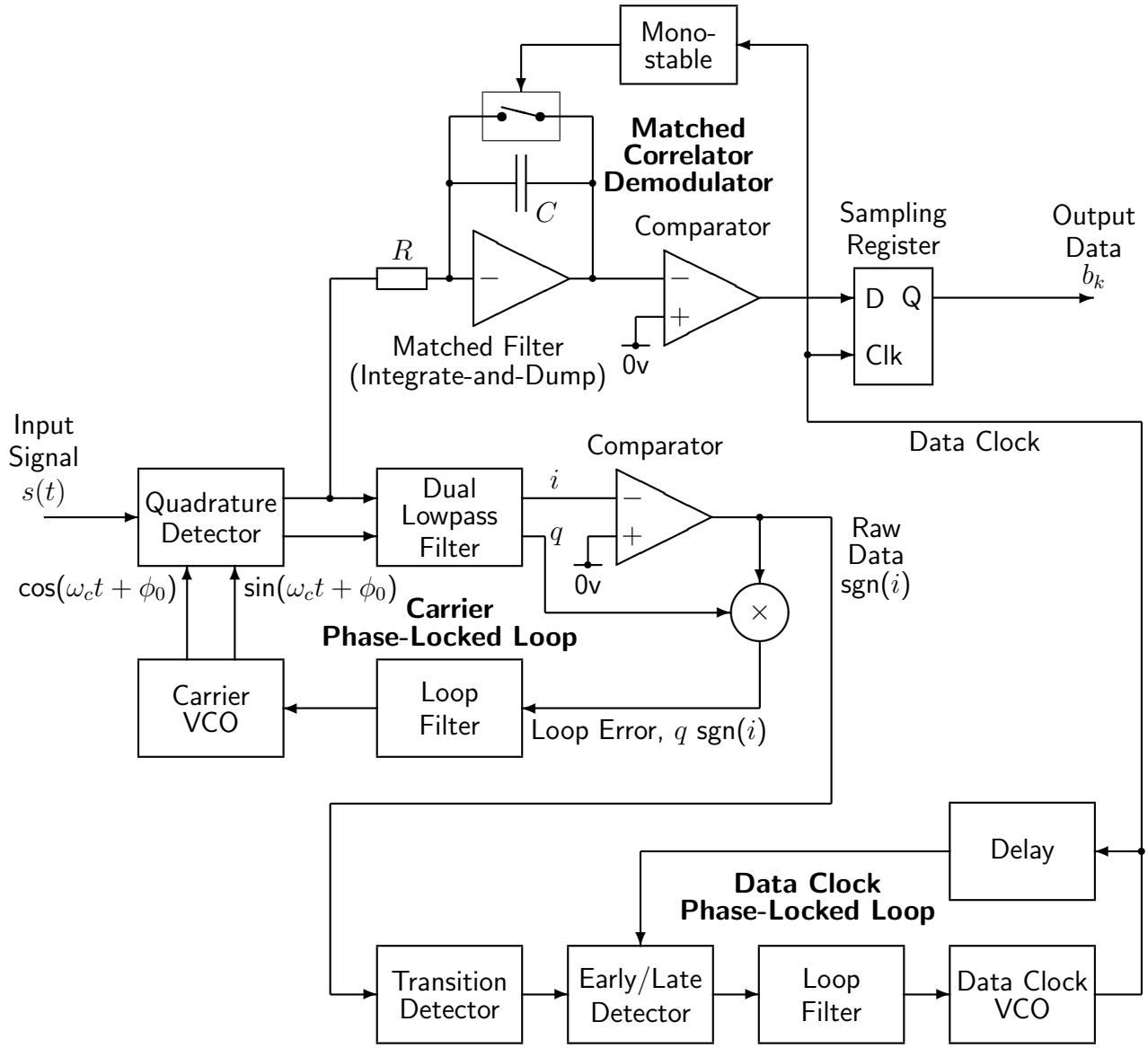

Fig 3.6: Practical implementation of an optimum demodulator for BPSK.

3.6 Practical Implementation of an Optimum BPSK Demodulator:

Fig 3.6 shows the block diagram of a practical demodulator.

The received bandpass signal $s(t)$ is demodulated into real and imag parts by the quadrature detector and is bandlimited by the dual lowpass filters to a bandwidth approximately equal to the bit rate. This generates the signals $i(t)$ and $q(t)$.

The carrier phase-locked loop (PLL) error signal is q sgn(i) (see fig 3.7), in order to produce a characteristic which repeats at multiples of π. This allows the loop to lock up at the positive zero-crossings of the error characteristic such that the phase error between $s(t)$ and the Carrier VCO (voltage controlled oscillator) is either 0 or π. Thus the Carrier VCO will either be in

phase or in antiphase with the carrier of $s(t)$, and the $i(t)$ signal will be a smoothed version of the original data or its complement. Differential decoding (not shown) must be used to eliminate this ambiguity.

The dual lowpass filters provide continuous outputs to allow the Carrier and Data Clock PLLs to operate correctly, but it is difficult to obtain the ideal rectangular impulse response (required for optimum matched-correlation detection) from real continuous filters. Therefore the data is detected via an alternative discrete-time filter which is an Integrate-and-Dump circuit. This circuit takes its input from the inphase (real) component of the quadrature detector, and provides optimum matched-correlation filtering for data detection.

The clock for the sampling register and the reset pulse for the Integrate-and-Dump is provided from the Data Clock PLL. This loop is controlled by an Early/Late detector, which generates a positive pulse if a data transition occurs early with respect to the nearest delayed clock edge, and a negative pulse if a transition occurs late. These pulses are integrated by the loop filter and used to increase or decrease the clock rate very slightly, so as gradually to lock the delayed clock edges to the data transitions.

The Dual LPF introduces delay to the timing of the transitions, so a compensating delay is included between the Data Clock VCO and the Early/Late detector. The clock timing will then be correct for the Integrate-and-Dump and sampling register. The Monostable generates pulses which are short compared with T_b, but long enough to discharge the integrator capacitor fully at the start of each bit period.

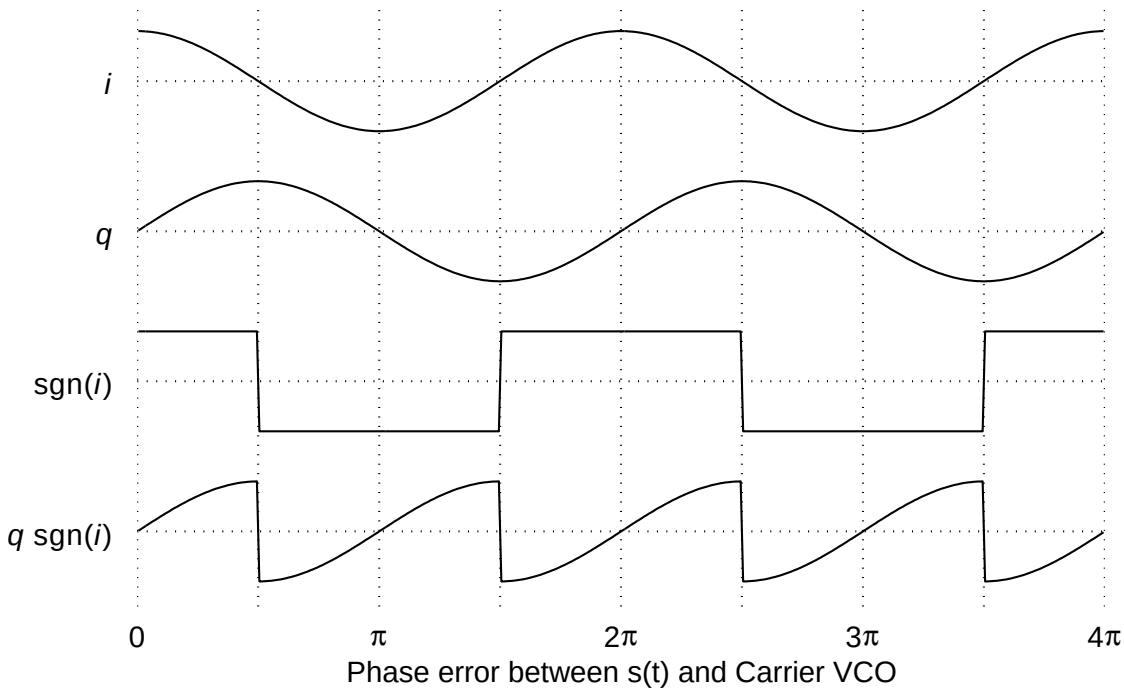

Fig 3.7: Phase detector characteristic of the Carrier Phase-Locked Loop.

3.7 Approximation Formulae for the Gaussian Error Integral, Q(x)

A Gaussian probability density function (PDF) with unit variance is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$$

The probability that a signal with a PDF given by $f(x)$ lies above a given threshold x is given by the Gaussian Error Integral or Q function:

$$Q(x) = \int_x^\infty f(u)\, du$$

There is no analytical solution to this integral, but it has a simple relationship to the error function, erf(x), or its complement, erfc(x), which are tabulated in many books of mathematical tables. $Q(x)$ is tabulated in Appendix A-10 of Couch, "Digital and Analog Communication Systems", 3rd edition, and in Appendix D of Shanmugam, same title.

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2}\, du \qquad \text{and} \qquad \text{erfc}(x) = 1 - \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-u^2}\, du$$

$$\therefore\ Q(x) = \tfrac{1}{2}\text{erfc}\left(\frac{x}{\sqrt{2}}\right) = \tfrac{1}{2}\left[1 - \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$$

Note that erf(0) = 0 and erf(∞) = 1, and therefore $Q(0) = 0.5$ and $Q(x) \to 0$ very rapidly as x becomes large.

It is useful to derive simple approximations to $Q(x)$ which can be used on a calculator and avoid the need for tables.

Let $v = u - x$:

$$\therefore\ Q(x) = \int_0^\infty f(v+x)\, dv = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-(v^2+2vx+x^2)/2}\, dv = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \int_0^\infty e^{-vx}\, e^{-v^2/2}\, dv$$

Now if $x \gg 1$, we may obtain an approximate solution by replacing the $e^{-v^2/2}$ term in the integral by unity, since it will initially decay much slower than the e^{-vx} term.

$$\therefore\ Q(x) < \frac{e^{-x^2/2}}{\sqrt{2\pi}} \int_0^\infty e^{-vx}\, dv = \frac{e^{-x^2/2}}{\sqrt{2\pi}\, x}$$

This approximation is an upper bound, and its ratio to the true value of $Q(x)$ becomes less than 1.1 only when $x > 3$, as shown in fig 3.8. We may obtain a much better approximation to $Q(x)$ by altering the denominator above from $(\sqrt{2\pi}\, x)$ to $(1.64x + \sqrt{0.76x^2 + 4})$ to give:

$$Q(x) \approx \frac{e^{-x^2/2}}{1.64x + \sqrt{0.76x^2 + 4}}$$

This improved approximation (developed originally by Borjesson and Sundberg, IEEE Trans. on Communications, March 1979, p 639) gives a curve indistinguishable from $Q(x)$ in fig 3.8 and its ratio to the true $Q(x)$ is now within $\pm 0.3\%$ of unity for all $x \geq 0$ as shown in fig 3.9. This accuracy is sufficient for nearly all practical problems.

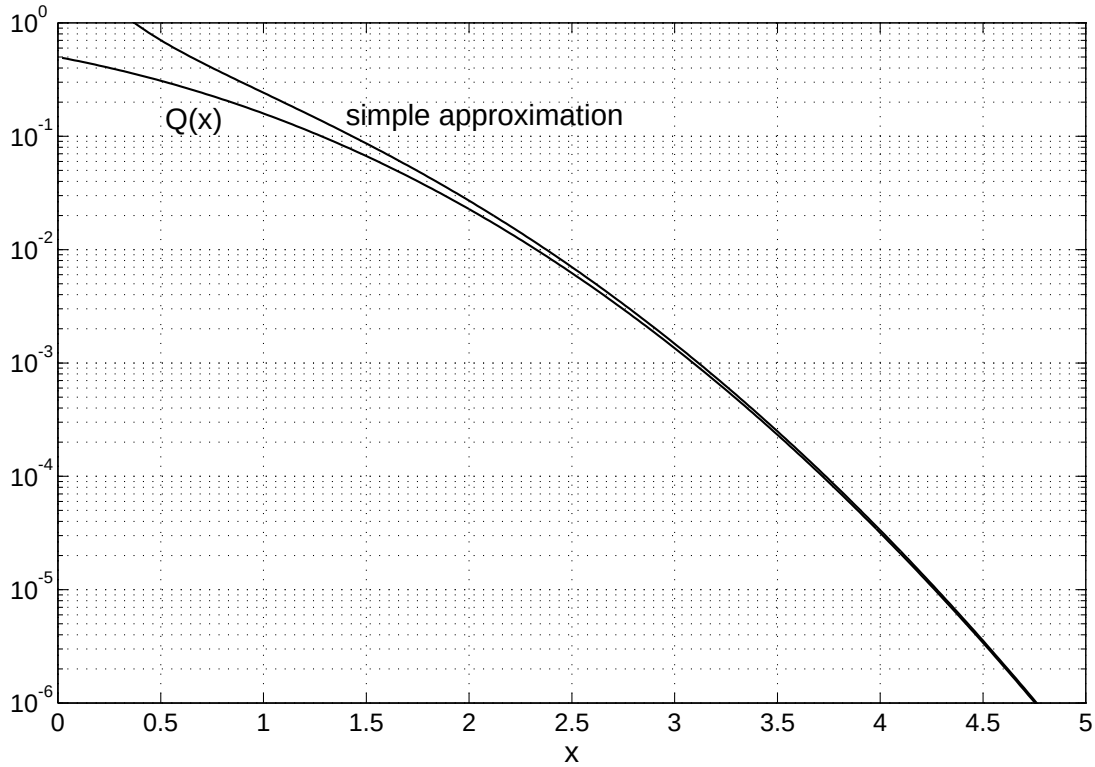

Fig 3.8: Q(x) and the simple approximation exp(-x*x/2) / sqrt(2*pi)*x

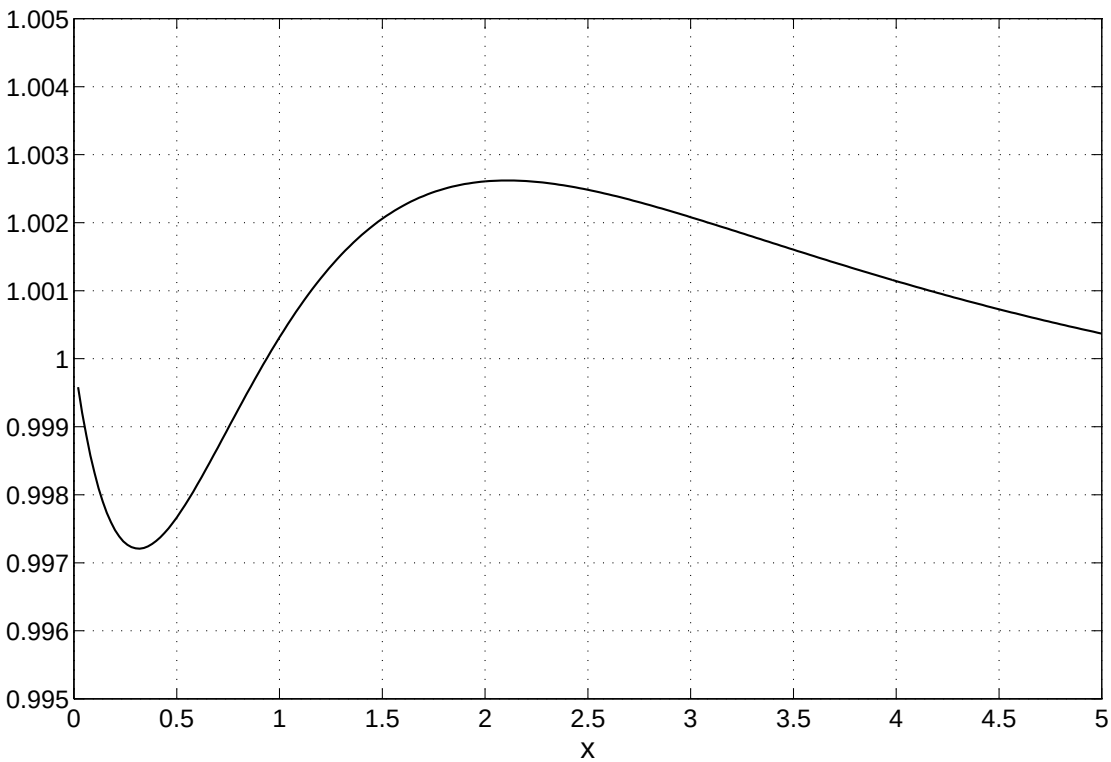

Fig 3.9: The ratio of the improved approximation of Q(x) to its true value

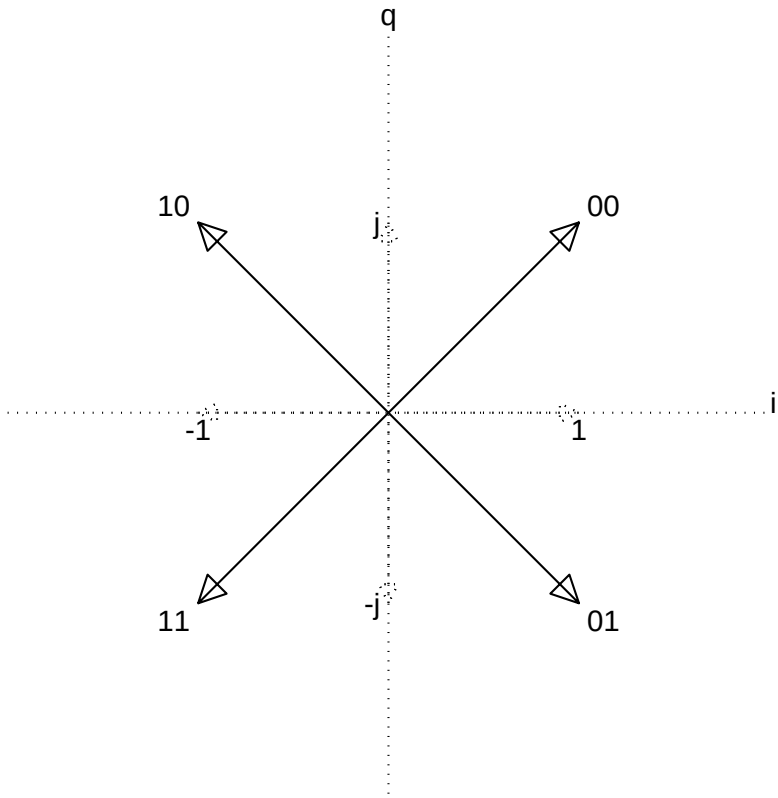

Fig 4.1: QPSK phasor diagram.

4 Other Binary Schemes

4.1 Quadrature PSK (QPSK):

QPSK is equivalent to **BPSK on two quadrature carriers**.

Even bits b_{2k} modulate the **inphase** carrier.

Odd bits b_{2k+1} modulate the **quadrature** carrier.

$$\therefore \; p_k(t) = [\; b_{2k} + j\; b_{2k+1} \;]\; g(t - kT_s)\; e^{j\phi_0}$$

where $g(t)$ is as for BPSK except that it is now non-zero from $t = 0$ to T_s, the 2-bit symbol period ($T_s = 2T_b$).

Hence $p(t)$ can have one of 4 values:

$$(\pm 1 \pm j)\; g(t - kT_s)\; e^{j\phi_0}$$

See fig 4.1 for QPSK phasor diagram.

QPSK can be regarded as 4-level modulation, but it is usually easier to treat it as two independent 2-level (binary) processes.

See Fig. 4.3 for symbol timing.

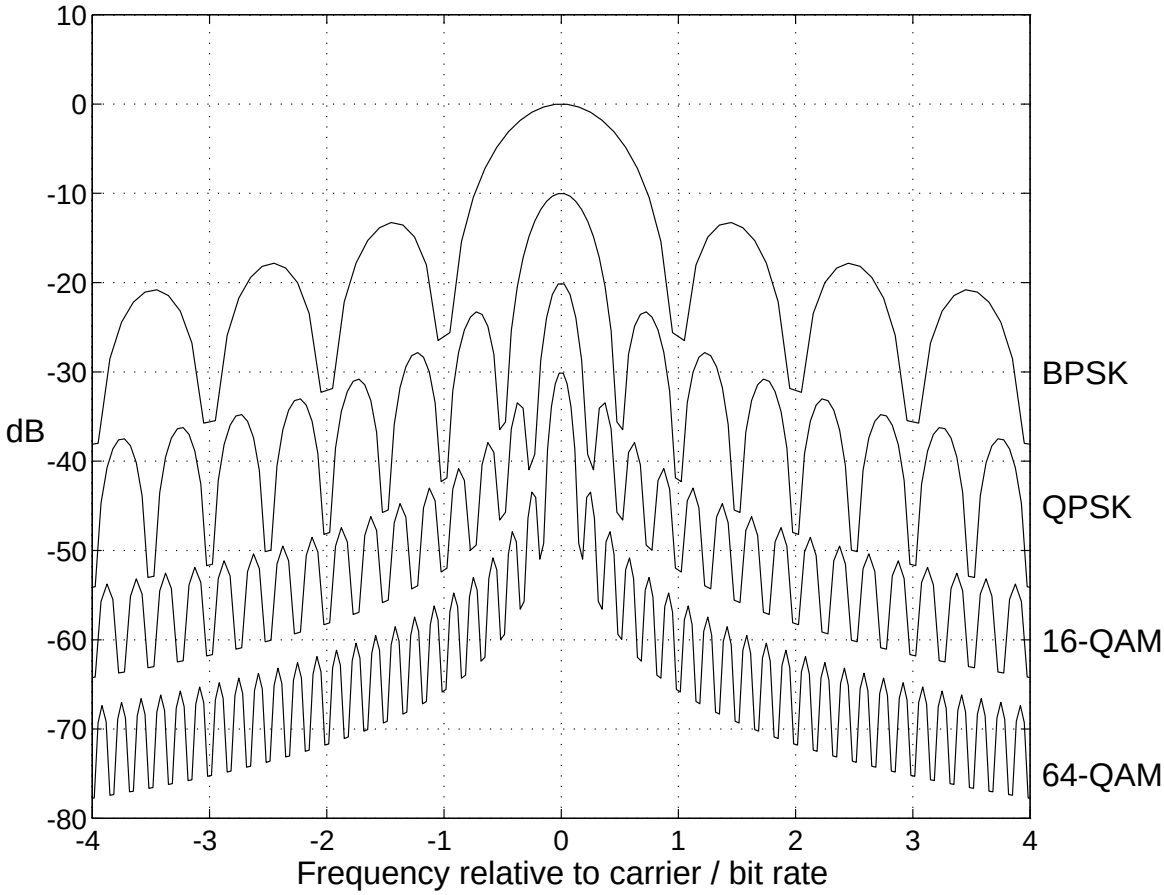

Fig 4.2: Power spectra of BPSK, QPSK, 16-QAM and 64-QAM for a given bit rate.

Power Spectrum of QPSK:

If each quadrature carrier of amplitude a_0 is BPSK modulated by rectangular data pulses at a symbol rate of $1/T_s$, then, using the BPSK result, the spectrum of each carrier is given by:

$$E\{|P_I(\omega)|^2\} = E\{|P_Q(\omega)|^2\} = a_0^2\ T_s\ \mathrm{sinc}^2(\omega T_s/2)$$

The data on the two carriers are uncorrelated, so the power spectra add to give a total spectrum of:

$$E\{|P(\omega)|^2\} = 2\ a_0^2\ T_s\ \mathrm{sinc}^2(\omega T_s/2) = 4\ a_0^2\ T_b\ \mathrm{sinc}^2(\omega T_b) \quad \text{since } T_b = T_s/2$$

$\therefore$ the QPSK spectrum is **half as wide** as the BPSK spectrum for a given data rate - a big advantage!

See fig 4.2 for spectra.

Bit Error Performance of QPSK:

Since the two carriers are in quadrature, they may be demodulated independently as two BPSK phasor components.

Power in each component $= \frac{1}{2}$ total power.

Bit rate for each component $= \frac{1}{2}$ total bit rate.

But E_b = (signal power) × (bit period) $= \dfrac{\text{signal power}}{\text{bit rate}}$

so E_b is the same for each component as it is for the total QPSK signal.

N_0 is unaffected by the modulation method.

$\therefore$ applying the BPSK result of section 3.4 to each component:

$$\text{Probability of bit error } P_E = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \quad - \quad \textbf{the same as BPSK.}$$

Differential Coding is normally used with QPSK to overcome carrier phase ambiguity, as for BPSK, again approx. doubling P_E.

Offset QPSK (OQPSK) is a variant of QPSK (see fig 4.3):

1. The keying of the two carriers is staggered by $T_b = T_s/2$.
2. The performance and spectrum are the same as QPSK.
3. OQPSK has less amplitude fluctuation after bandlimiting than QPSK. It is therefore better suited to non-linear bandlimited channels, such as those involving satellite on-board transmitters.

QPSK Demodulator Design:

This may be based on the BPSK design of fig 3.6 with the following additions:

- Data is determined from the polarities of **both the i and q outputs** of the Quadrature Detector, using two Matched Correlator Demodulators to give the odd and even data bits respectively.
- The Carrier PLL error signal is

  $$q\ \text{sgn}(i) - i\ \text{sgn}(q)$$

  in order to give a phase error characteristic which repeats at multiples of $\pi/2$ and has lock points at odd multiples of $\pi/4$.

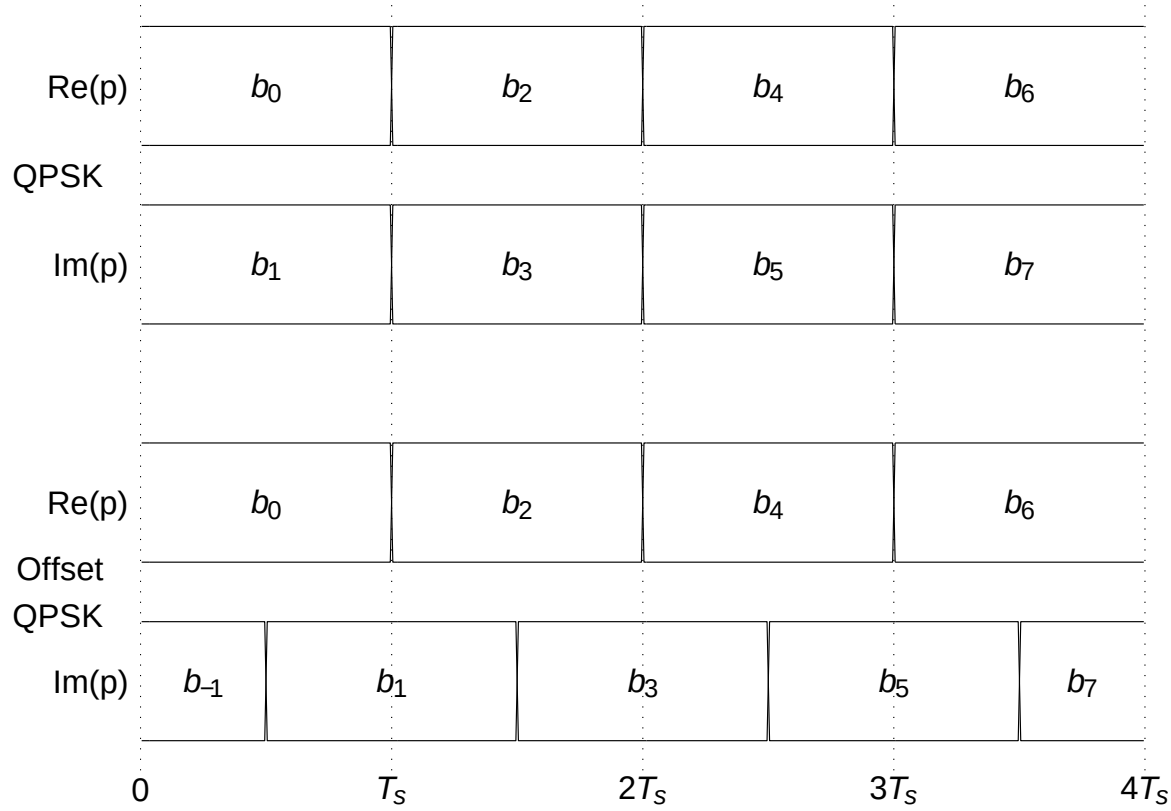

Fig 4.3: QPSK and Offset QPSK symbol timing.

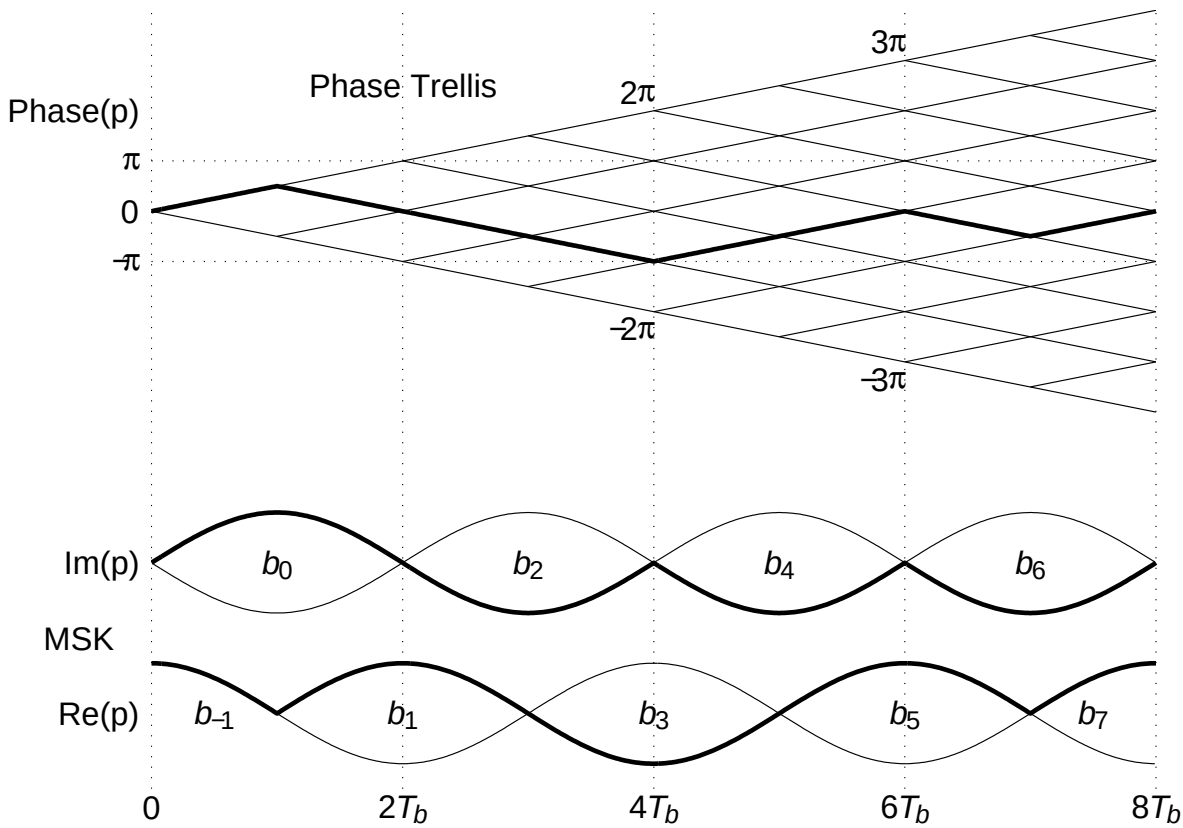

Fig 4.4: MSK phase trellis and symbol timing.

4.2 Binary Frequency-Shift Keying (BFSK):

Data causes a shift between 2 frequencies, $-\omega_D$ and $+\omega_D$ w.r.t. the carrier.

$$f_D = \omega_D/2\pi \quad \text{is the } \textbf{frequency deviation} \text{ (in Hz).}$$

Hence the phasor during the k^{th} bit period is:

$$p_k(t) = a_0 \; e^{j(b_k\,\omega_D(t-kT_b) \; + \; \phi_k)}$$

ϕ_k is the initial phase value at the start of each bit period, and it is updated at each bit boundary so as to prevent abrupt phase changes in $p(t)$. This has the advantage of reducing spectral sidelobes and it can also improve performance against noise.

FSK with this constraint is known as Continuous Phase FSK (CPFSK). See fig 2.4.

The FSK mod. index is given by

$$m_{FSK} = \frac{\text{difference between the 2 transmit freqs.}}{\text{bit rate}} = 2f_DT_b$$

As with Analogue FM, FSK is difficult to analyse in general. Its bandwidth is equal to the peak-to-peak frequency deviation, $2f_D$, plus an amount that is approximately equal to the rate of signalling (ie the symbol rate), $1/T_b$. Hence

$$\text{Bandwidth required } \approx 2f_D + \frac{1}{T_b} = \frac{m_{FSK}+1}{T_b}$$

(similar to Carson's rule for FM).

Minimum (Freq.) Shift Keying (MSK) – a special case:

$$m_{FSK} = 0.5 \qquad \therefore \; f_D = \frac{m_{FSK}}{2T_b} = \frac{1}{4T_b}$$

so the phase of p increases or decreases by 90° in each bit period, as in fig 2.4.

This is the minimum m_F which gives an error performance that is optimum in some sense and is the reason for the name MSK.

Fig 4.4 shows the phase plot (phase trellis) for MSK and below this are the i and q components of the phasor waveform. Note the similarity to to OQPSK except that the $g(t)$ pulses for MSK are half-sine pulses instead of rectangles. The broad line shows a typical phase trajectory and its components.

The power spectrum of MSK is proportional to the square of the Fourier transform of the half-sine pulse shape, rather the than the rectangular pulse shape of QPSK. The righthand halves of these two spectra are shown in fig. 4.5 and labelled MSK and QPSK respectively. Note that MSK has a **50% wider main lobe** than QPSK, but this is compensated for by MSK's **significantly lower sidelobe levels** (due to MSK being a continuous phase process).

If the data determines the slope of the phase trajectory, then the polarities of the half-sine pulses are obtained by differentially encoding the data, after inverting alternate bits. Hence the error performance for MSK is equivalent to OQPSK, QPSK and BPSK, each with differential coding.

Gaussian-filtered MSK (GMSK)

In many practical systems, such as in the mobile phone system GSM, substantially lower levels of spectral sidelobes are required than even MSK is able to give. In these cases it is common to apply smoothing to the binary data pulses before they are applied to the MSK modulator.

For GMSK, the smoothing lowpass filter has a Gaussian frequency response, whose -3 dB bandwidth is typically 0.3 times the bit rate. This filter bandwidth is chosen to give a good tradeoff between narrow transmitted bandwidth and low intersymbol interference. This scheme is known as $0.3R$ GMSK.

Figures 4.5 and 4.6 illustrate this tradeoff. Fig. 4.5 shows one side of the power spectrum of a QPSK signal, a pure MSK signal, and three GMSK signals with gaussian filter bandwidths of $0.5R_b$, $0.3R_b$ and $0.2R_b$ respectively. Fig. 4.5 shows the eye diagrams at the threshold detector of an ideal MSK demodulator for the three filter bandwidths. We can see that the $0.3R_b$ filter represents a good tradeoff between a well contained spectrum and a good eye opening. In practise an equaliser would probably be used to improve the eye opening further.

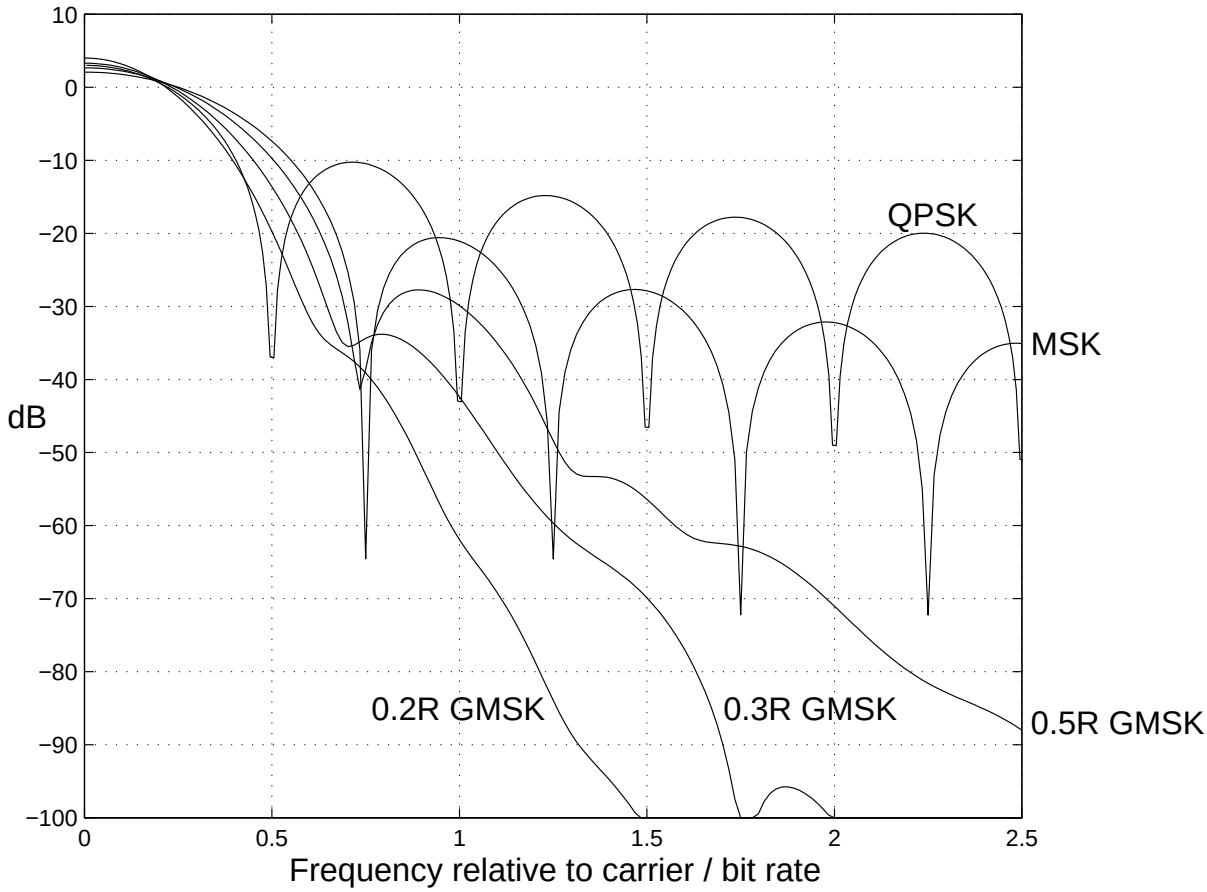

Fig 4.5: Power spectra of QPSK, MSK, 0.5R GMSK, 0.3R GMSK and 0.2R GMSK.

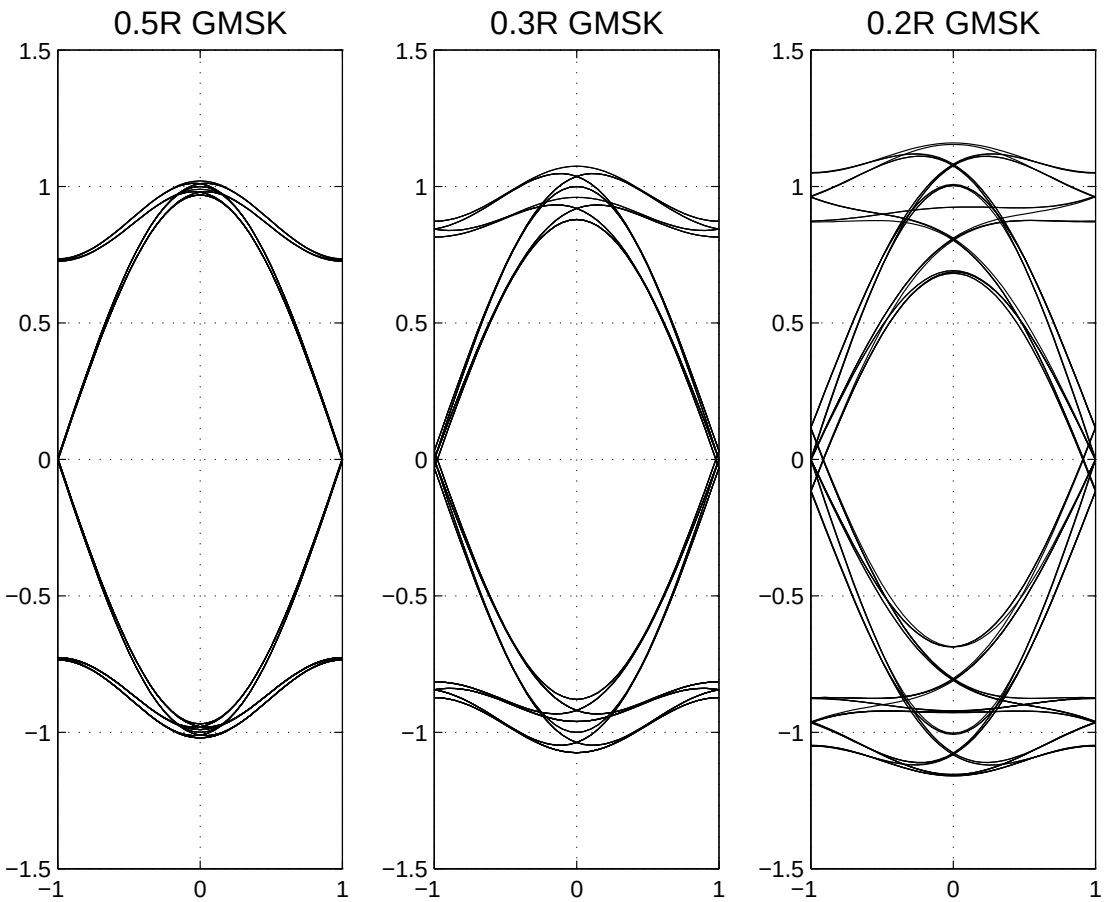

Fig 4.6: Eye diagrams for GMSK at the detector of an ideal MSK demodulator.

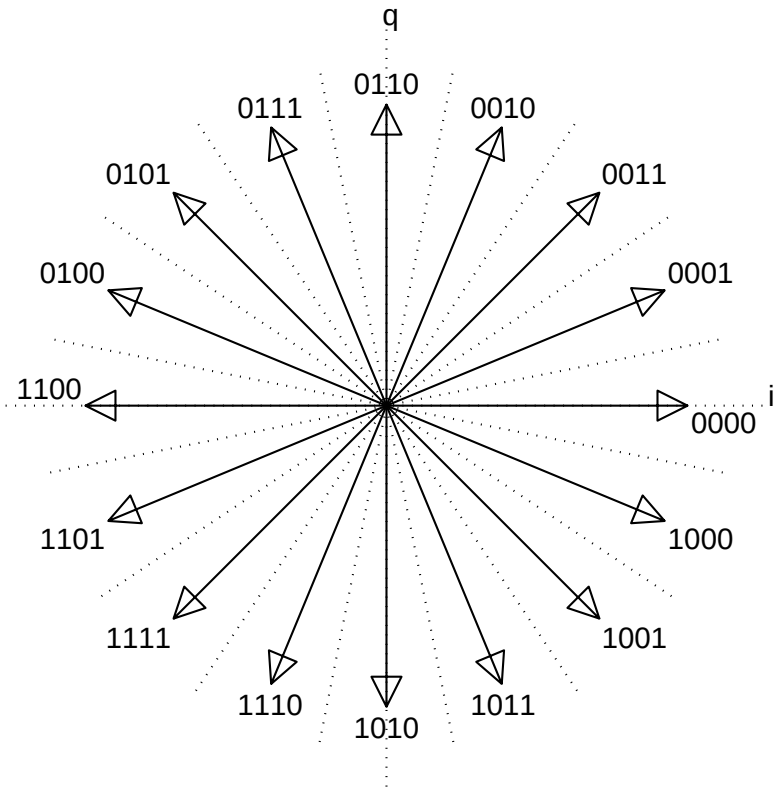

Fig 5.1: 16-level PSK phasor diagram

5 Multi-level Modulation

M-ary modulation uses one of M signals during each symbol interval T_s, thereby transmitting $m = \log_2 M$ bits of information per T_s ($M = 2^m$). **By varying M we may trade bandwidth for performance in noise.**

5.1 M-ary PSK (MPSK):

Transmits one of M phases, usually $2\pi i/M$ for $i = 0 \ldots M-1$. For example:

BPSK uses 2 phases ($m = 1$)

QPSK uses 4 phases ($m = 2$)

8-PSK uses 8 phases ($m = 3$) etc.

$$\text{Bandwidth} \propto \frac{1}{T_s} = \frac{1}{mT_b}$$

Noise immunity $\propto$ distance between adjacent points $= 2a_0 \sin(\pi/M)$, if amplitude $= a_0$.

$\therefore$ Increasing m reduces bandwidth;
But it rapidly worsens the performance in noise.

Fig 5.1 shows the phasor diagram of 16-PSK.

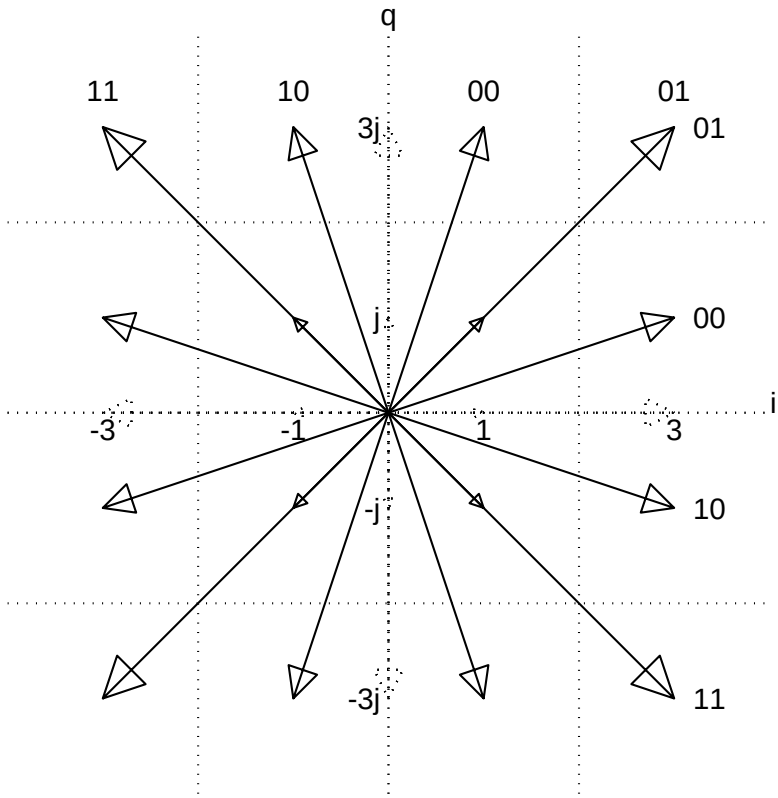

Fig 5.2: 16-QAM phasor diagram

5.2 Quadrature Amplitude Modulation (QAM):

QAM achieves a greater distance between adjacent points by **filling the p-plane more uniformly** than MPSK. However the constant amplitude property is then lost.

QAM is the basic modulation method of choice for almost all **bandwidth critical** systems of today, such as internet modems and digital TV broadcasting.

QAM is similar to QPSK except that it uses multilevel ASK-SC (with zero mean) on the two quadrature carriers. See figs 5.2 and 5.3 for 16-QAM and 64-QAM.

8-QAM and 32-QAM are possible, but are more complicated due to the problem of dividing an odd number of bits between the two carriers.

Generalising the QPSK case, the QAM phasor during the k^{th} symbol period is:

$$p_k(t) = [s_{2k} + j\ s_{2k+1}]\ g(t - kT_s)\ e^{j\phi_0} \tag{5.1}$$

where s_{2k} or $s_{2k+1} = 2i + 1 - M$, and the state i is chosen from $0 \ldots M-1$ according to the data.

To analyse QAM noise performance, we consider each carrier separately. We shall analyse the s_{2k} component and then assume the same performance for the other component.

For M^2-QAM, the number of levels for each carrier $M = 2^m$, conveying m bits per symbol **on each carrier**. Hence the total capacity is $2m$ **bits per symbol**.

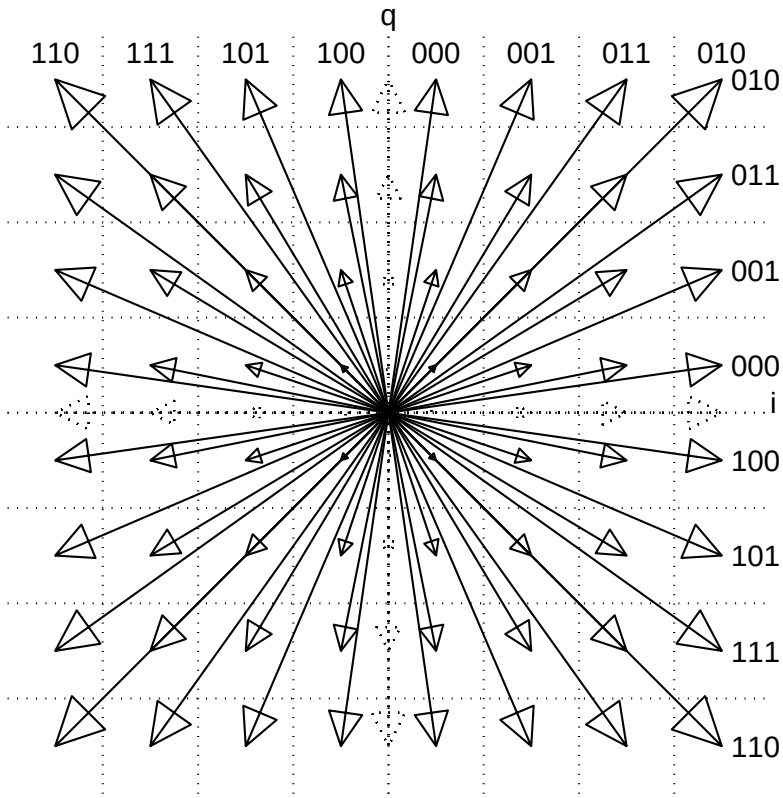

Fig 5.3: 64-QAM phasor diagram

Noise performance of QAM

Let the signal level expected at the receiver inphase threshold detector for the k^{th} symbol be $s_{2k}v_s$ where:

$$s_{2k} = (2i + 1 - M) \qquad \text{for } i = 0 \ldots M-1.$$

e.g. if $M = 8, s_{2k} = -7, -5, -3, -1, 1, 3, 5,$ or 7.

Hence the signal levels are separated by $2v_s$ and the optimum threshold levels are v_s above and below each expected signal level, assuming equiprobable data symbols.

Let the rms noise volts at the threshold detector be σ.

Consider a signal component in state i, where $0 < i < M-1$ (so i is not an outer state):

Probability of state i becoming $i+1 = Q(\frac{v_s}{\sigma})$

Probability of state i becoming $i-1 = Q(\frac{v_s}{\sigma})$

$$\therefore \text{ Probability of error from state } i = 2Q(\frac{v_s}{\sigma})$$

There are $M-2$ states in this category.

For the two outer states, where i = 0 or $M-1$, there is only one error direction:

$$\therefore \text{ Probability of error from state 0 or } M-1 = Q(\frac{v_s}{\sigma})$$

There are 2 states in this category. Hence the mean probability of symbol error is

$$P_{SE} = \frac{M-2}{M}\ 2Q(\frac{v_s}{\sigma}) + \frac{2}{M}\ Q(\frac{v_s}{\sigma}) = 2(1-\frac{1}{M})\ Q(\frac{v_s}{\sigma}) \tag{5.2}$$

If we use m-bit Gray (unit distance) coding for the M levels, each symbol error to an adjacent state will only cause a single bit error in each m-bit word. We ignore errors to non-adjacent states (unlikely except at poor SNR). Since there are m bits for every symbol:

$$\therefore \text{ Mean probability of bit error, } P_{BE} = \frac{P_{SE}}{m} = \frac{2}{m}(1-\frac{1}{M})\ Q(\frac{v_s}{\sigma}) \tag{5.3}$$

Now we evaluate v_s/σ in terms of E_b/N_0.

The optimum receiver is as for BPSK (i.e. it correlates the input with $g(t-kT_s)e^{-j\phi_0}$), except that $M-1$ threshold detectors are used to detect the M levels of each quadrature component. Modifying the BPSK result from section 3.4 to get the waveform at the inphase threshold detectors gives:

$$\begin{aligned} y(k) &= Gs_{2k}\int_0^{T_s} g^2(t)\ dt + G\int_{kT_s}^{(k+1)T_s} n_1(t)\ g(t-kT_s)\ dt \\ &= Gs_{2k}E_g + \left(\text{noise of mean power } \frac{G^2N_0E_g}{2}\right) \end{aligned} \tag{5.4}$$

where the energy of the $g(t)$ pulse is:

$$E_g = \int_0^{T_s} g^2(t)\ dt$$

The signal levels $Gs_{2k}E_g$ are the same as those we assumed above to be $s_{2k}v_s$.

$$\therefore\ v_s = GE_g$$

The rms noise volts were assumed to be σ, so from (5.4):

$$\sigma = G\sqrt{\frac{N_0E_g}{2}}$$

$$\therefore\ \frac{v_s}{\sigma} = \frac{GE_g}{G\sqrt{N_0E_g/2}} = \sqrt{\frac{2E_g}{N_0}} \tag{5.5}$$

Finally we need to express E_g in terms of the energy per bit E_b (these are no longer the same with multilevel modulation).

Using (5.1), the average symbol energy on the s_{2k} carrier is:

$$E_s = \frac{1}{M} \sum_{i=0}^{M-1} (2i + 1 - M)^2 \int_0^{T_s} g^2(t)\, dt$$

But

$$\sum_{i=0}^{M-1} (2i + 1 - M)^2 = \frac{M(M^2 - 1)}{3} \qquad \text{(Prove by induction)}$$

$$\therefore\ E_s = \frac{M^2 - 1}{3}\, E_g$$

Since there are m bits per symbol, $E_s = mE_b$.

$$\therefore\ E_b = \frac{E_s}{m} = \frac{M^2 - 1}{3m}\, E_g \tag{5.6}$$

Substituting (5.6) into (5.5):

$$\frac{v_s}{\sigma} = \sqrt{\frac{3m}{M^2 - 1}\,\frac{2E_b}{N_0}}$$

$\therefore$ from (5.3) the mean probability of bit error is given by:

$$P_{BE} = \frac{2}{m}(1 - \frac{1}{M})\, Q\left(\sqrt{\frac{3m}{M^2 - 1}\,\frac{2E_b}{N_0}}\right) \tag{5.7}$$

Fig 3.5 (repeated on next page) shows the error performance curves of 16-QAM and 64-QAM, compared with BPSK & QPSK. We see that **16-QAM is approximately 4 dB worse than BPSK and QPSK**, and that **64-QAM is another 4 dB worse than 16-QAM**.

This difference in performance is almost entirely due to the factor $\sqrt{\frac{3m}{M^2 - 1}}$ in the argument of the Q function in the above equation, because Q is so steep on a log scale in the region of interest that the scaling of Q by $\frac{2}{m}(1 - \frac{1}{M})$ has minimal effect. The values of the former factor, as a voltage ratio (and in dBs), are given below for various m and $M = 2^m$:

m	1	2	3	4
M	2 (BPSK/QPSK)	4 (16-QAM)	8 (64-QAM)	16 (256-QAM)
$\sqrt{\frac{3m}{M^2-1}}$	1 (0dB)	$\sqrt{\frac{6}{15}}$ (−3.98dB)	$\sqrt{\frac{9}{63}}$ (−8.45dB)	$\sqrt{\frac{12}{255}}$ (−13.27dB)

We see that the predicted degradations by 3.98 dB and 8.45 dB agree very closely with the spacing of the curves in fig 3.5.

Fig 5.4 shows this result graphically for QPSK, 16-QAM and 64-QAM, whose constellations are all plotted to the same scale of E_b (energy per bit). This is achieved by scaling the units of the axes in proportion to $\sqrt{E_g}$ in eq(5.6), as E_b is held constant. In fig 5.4, the separation of the constellation points directly shows the resilience of each modulation scheme to noise.

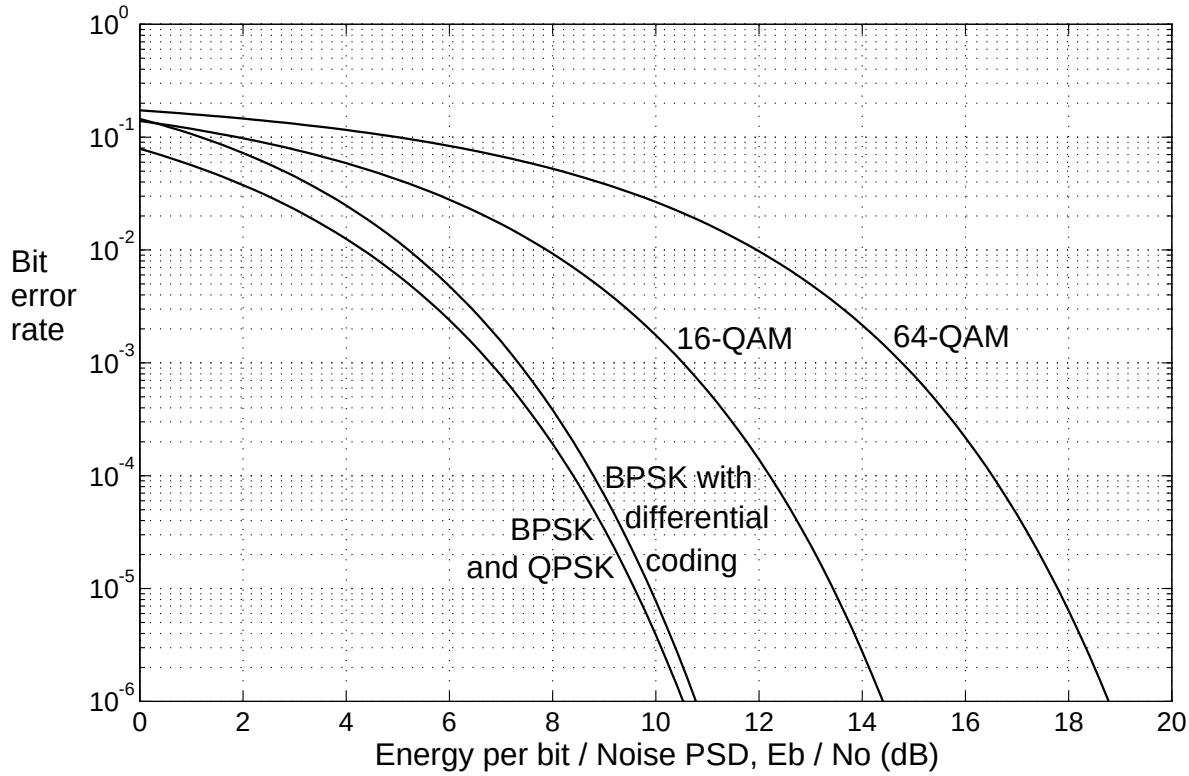

Fig 3.5: Bit error rate curves for PSK and QAM.

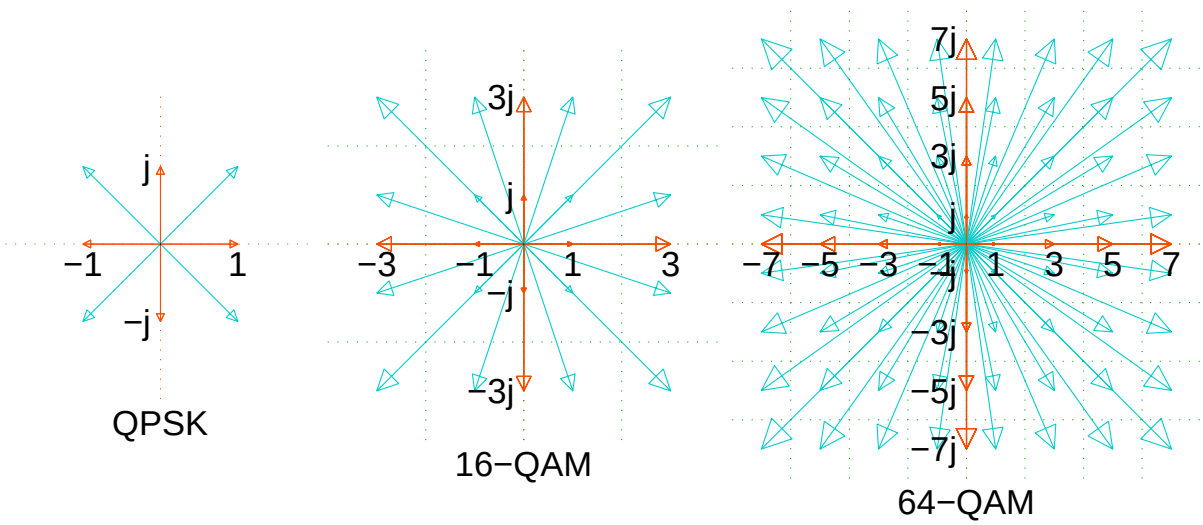

Fig 5.4: Constellations, drawn to scale, with equal mean energy per bit, to allow visual assessment of their relative performance in noise.

Power spectrum of QAM

The power spectrum of QAM is given by the squared magnitude of the Fourier transform of the basic signal pulse $g(t)$, since the pulses for consecutive symbols are still uncorrelated with each other and have zero mean. Hence the autocorrelation function (ACF) of the QAM phasors is still proportional to the ACF of $g(t)$.

Therefore for rectangular pulses of rms ampl a_0 and duration $T_s = 2mT_b$ (there are 2m bits per symbol when both carriers are included):

$$E\{|P(\omega)|^2\} = 2\,a_0^2\,T_s\,\mathrm{sinc}^2(\omega T_s/2) = 4\,m\,a_0^2\,T_b\,\mathrm{sinc}^2(m\omega T_b)$$

Hence **the bandwidth is reduced by $2m$ relative to BPSK (or m relative to QPSK)** – this is the main advantage of QAM over QPSK (see fig 4.2).

BUT the error performance is worse due mainly to the $\sqrt{\frac{3m}{M^2-1}}$ term in the expression for P_{BE} becoming small as m increases ($M^2 = 2^{2m}$ and increases much faster than m).

5.3 M-ary FSK (MFSK):

MFSK can provide **better error performance** than QPSK at the expense of **wider bandwidth**.

MFSK is similar to BFSK except that M different frequencies are used instead of just two. These are usually spaced by $1/T_s = 1/mT_b$ to give an orthogonal set of waveforms (similar to an FFT) for optimum performance.

The bandwidth required for MFSK with a spacing of $1/T_s$ is approximately:

$$(M-1)\ (\text{tone spacing}) + (\text{tone bandwidth}) = \frac{M-1}{T_s} + \frac{2}{T_s} = \frac{M+1}{mT_b}$$

Hence increasing M above 4 tends to increase the bandwidth.

It is difficult to derive exact expressions for the performance of MFSK demodulators, and the performance depends on whether detection is **coherent** (relying on locking to the carrier phase over many symbol periods) or **non-coherent** (allowing arbitrary initial carrier phase for each symbol period). MFSK is most frequently used when non-coherent detection is necessary, such as when frequency hopping (a form of spread spectrum) is employed.

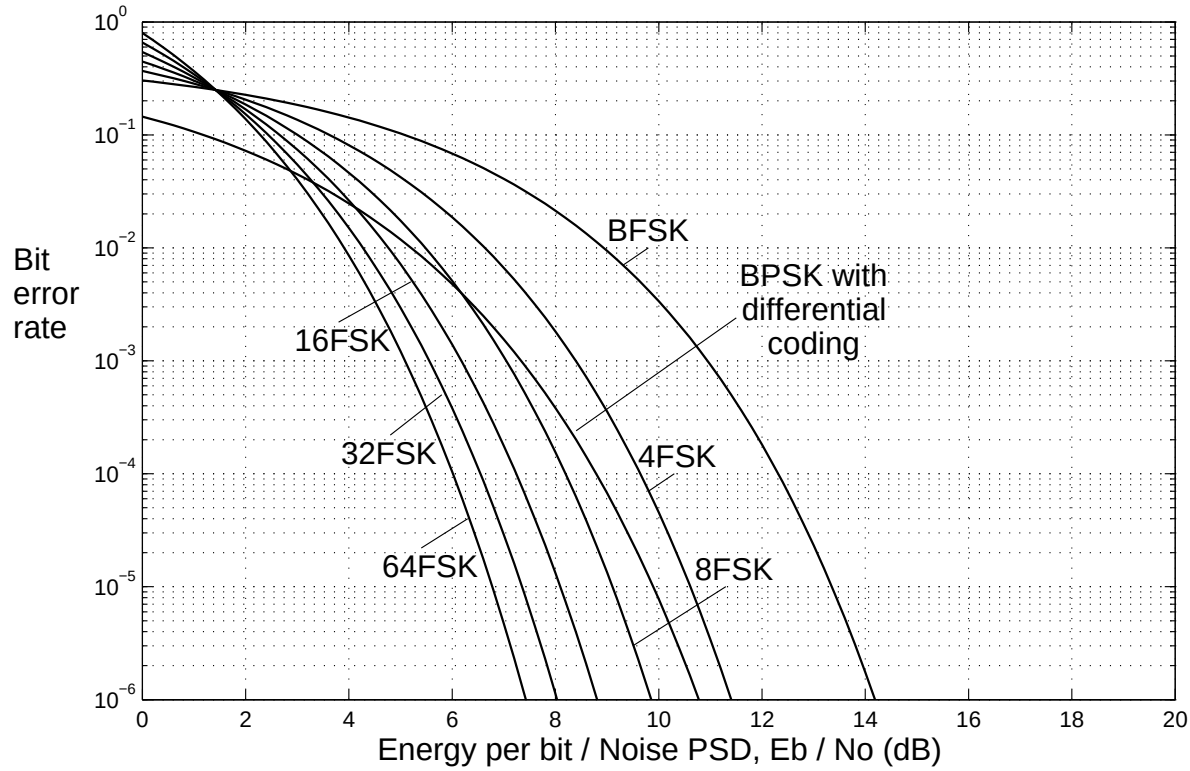

Fig 5.5: Approximate bit error rate curves for MFSK.

In the non-coherent case, it can be shown that the approximate bit error probability of MFSK at good SNR is given by:

$$P_{BE} \approx \frac{M}{4} e^{(-mE_b/2N_0)} = \tfrac{1}{4}\left(2e^{(-E_b/2N_0)}\right)^m$$

This is plotted in fig 5.5 and the differentially decoded BPSK curve from fig 3.5 is included for comparison. We see that MFSK gives improved performance with increasing M and tends to outperform BPSK (which uses coherent detection) when $M \geq 8$.

Comparing fig 5.5 with fig 3.5 (shown 2 pages back), we see the very different tradeoffs available with MFSK compared with QAM schemes. With MFSK, performance improves with increasing M while bandwidth worsens, whereas with QAM, performance degrades with increasing M while bandwidth improves! This is because MFSK increases the number of modulation states by occupying more bandwidth and without reducing the spacing between states, whereas QAM reduces the signalling rate (and hence bandwidth) but also substantially reduces the spacing between states.

6 Digital Audio and TV Broadcasting

In this final section of the course, we shall show that many of the ideas and techniques, described so far, are used in two very important modern communications applications that have revolutionised broadcasting over the last 10 years – Digital Audio Broadcasting (digital radio), and Digital Video Broadcasting (digital TV) over terrestrial channels (freeview).

6.1 Digital Audio Broadcasting (DAB)

DAB has been designed to meet the following **requirements**:

- Audio quality comparable with CD
- Solid reliable reception (even in cars)
- Simple selection and identification of stations
- Avoidance of frequent retuning

However to do this, a designer encounters the following **problems**:

- CD audio needs 1.5 Mb/s. Transmitting this would either occupy too much bandwidth for a single programme or would require QAM with many levels and hence be very sensitive to noise and interference.
- Good reception in cars and on personal radios requires immunity to multipath effects, typical of dense urban surroundings. This requires a low rate of modulation so that different multipath delays do not cause excessive ISI. Typical delay differences can be up to 10μs (approx 3 km path difference) but will usually be much less than this.
- Multipath with a delay difference between paths of Δt tends to cause frequency selective fading at intervals of $1/\Delta t$ across the frequency band, so that a small (but unpredictable) proportion of the frequency band may be unusable.

In order to overcome these problems, the following **techniques** have been developed:

- Audio compression (MUSICAM = MPEG layer 2, similar to MP3). 20 kHz audio is converted to 128 kb/s (stereo = 192, 224, or 256 kb/s). Masking properties of the human hearing system are used to achieve this.
- Orthogonal frequency-division multiplexing (OFDM) modulation uses many carriers in parallel to reduce the signaling rate on each carrier. Guard intervals provide minimal degradation due to multipath delay differences.
- QPSK or QAM is used on each carrier to achieve good tradeoffs between spectral efficiency and noise resilience.
- Error Correction Coding is used to add some redundancy so that carriers which are subject to frequency selective fading may be ignored and all the data can still be correctly recovered.

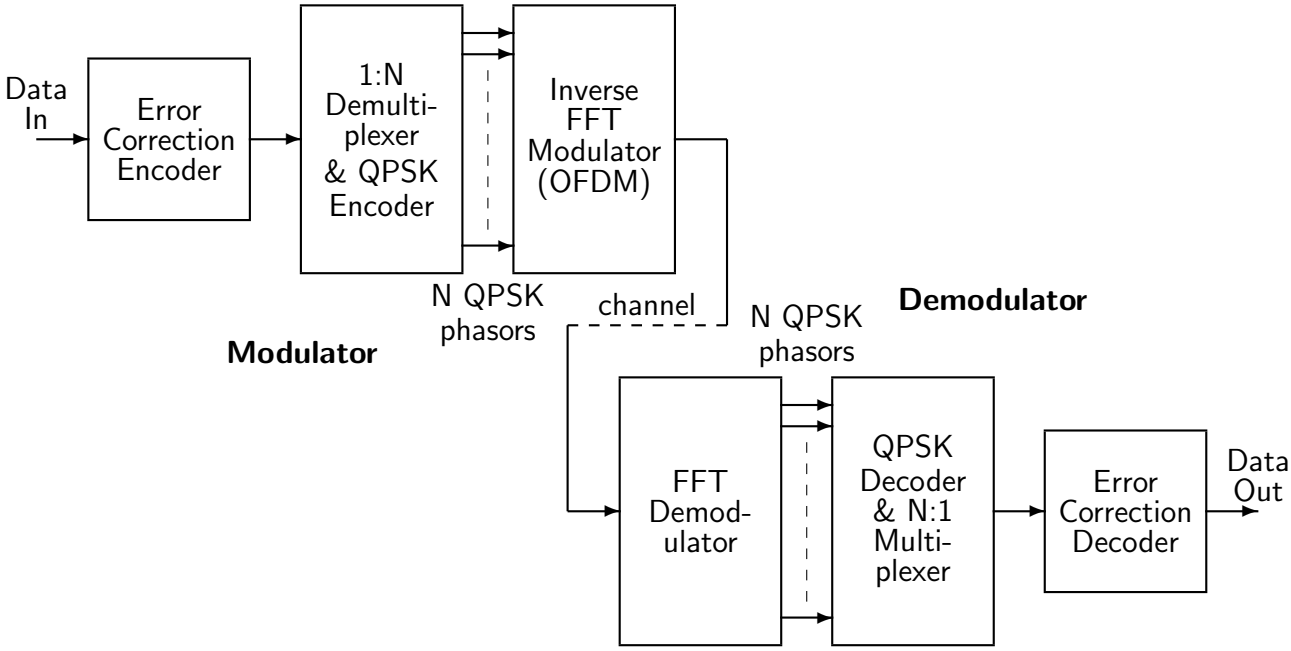

Fig 6.1: Block diagram of a basic Coded Orthogonal Frequency Division Multiplexing (COFDM) system.

6.2 Coded Orthogonal Frequency Division Multiplexing (COFDM)

COFDM comprises two parts, the error correction coding (C) and the modulation (OFDM). Fig. 6.1 shows a system block diagram.

Coding

Error Correction Coding (ECC) is provided for DAB by a **convolutional code**. This is similar to a block code, except that the data is passed through a shift register, and new parity bits are calculated from the contents of the register after each new bit is shifted in. Hence one or more parity polynomials are convolved with the input data, to produce a continuous stream of parity bits (similar to one or more FIR digital filters, except that modulo-2 arithmetic is used).

The most powerful convolutional codes are **non-systematic**, which means that the raw data bits are not transmitted; only the various parity bit streams are transmitted. DAB specifies a basic code that is rate 1/4; i.e. 4 parity bits are generated after each input bit is shifted into the encoder shift register. However a range of code rates may be obtained by a technique, known as **puncturing**, which simply removes parity bits at predefined regular intervals from the encoder output bit stream. By deleting (puncturing) alternate bits from the rate 1/4 code, we obtain a rate 1/2 code, which is the code rate used by current DAB signals in the UK. Different code rates can provide a range of tradeoffs between error correction performance, bandwidth and user data rates.

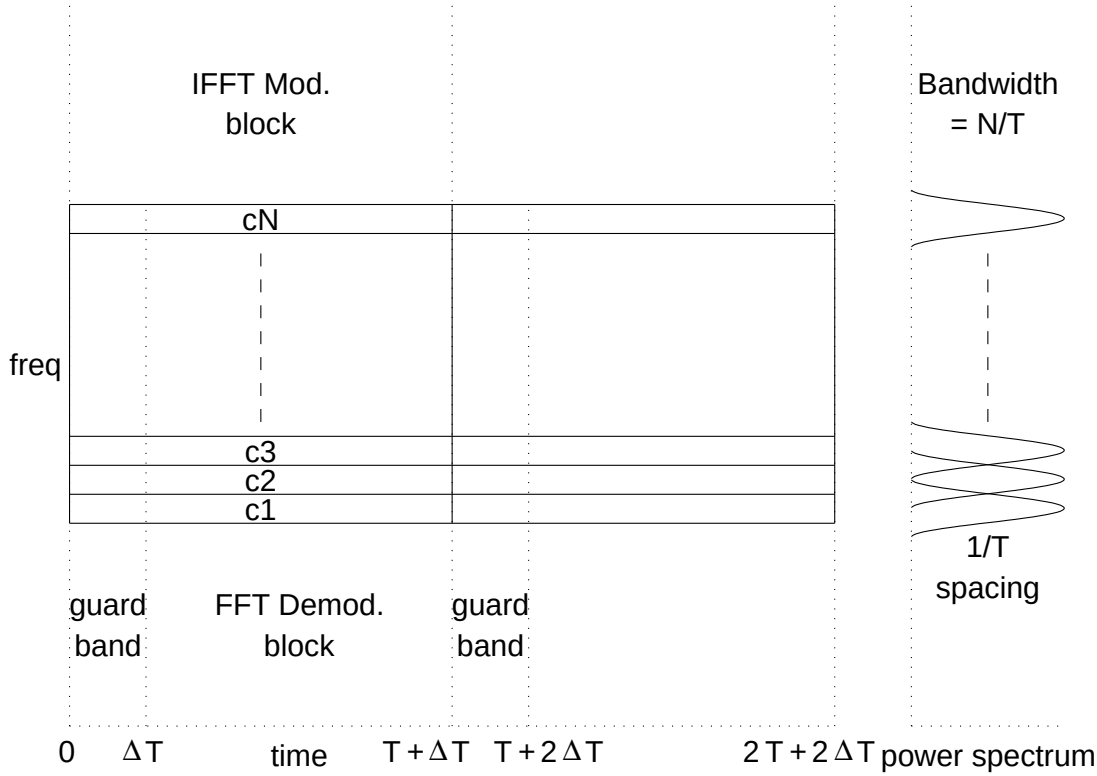

Fig 6.2: Orthogonal Frequency Division Multiplexing (OFDM) with N carriers.

Orthogonal Frequency Division Multiplexing

The aim of OFDM is to demultiplex the high-speed bit stream into N **streams**, each at $1/N$ of the original rate, which are then modulated onto N separate carrier waves, as shown in fig. 6.2. These have much improved resilience to typical multipath delays because of their **lower modulation rates**. Typically $N \approx 1000$ to 2000.

The inverse FFT may be used to put QPSK (or QAM) data on each of N carriers, spaced by $1/T$ Hz, where T is the IFFT block period. Each carrier is an IFFT basis function which is multiplied by the modulation phasor ($\pm 1 \pm j$ in the case of QPSK). In this way the **carriers are orthogonal** to each other and may be demodulated by an equivalent FFT process without mutual interference at the receiver. The mutual orthogonality of the IFFT basis functions, means that there should be no interference between each modulated carrier and its neighbours. Orthogonality is not affected by the modulation process, because the modulation rate is no faster than once per FFT block period, so each modulated carrier is a pure tone for the duration of the block period T.

Figures 6.3 to 6.10 show the 2 input data bits and the modulated IFFT basis functions for QPSK modulation of frequency slots 1 to 8. Figure 6.11 shows the result of combining all 8 slots together into an OFDM signal. Since all the 8 slots are at positive frequencies, there is an overall positive bias to the rotation of the composite phasor waveform in fig 6.11 . In practise negative frequency slots would be used as well, and fig 6.12 shows the result for 32 carriers occupying slots -16 to $+15$. The resulting phasor waveform looks much more random and noiselike.

Multipath delays tend to vary across the band of N carriers, and this could upset the orthogonality property because some modulation transitions would need to occur during the

demodulator FFT analysis block if each block followed its predecessor immediately. To avoid this problem, a **guard period** ΔT is inserted between consecutive blocks in the modulator and demodulator. For optimum demodulator performance, the modulator should extend (by periodic extension) the inverse FFT output waveform into the guard period before each block, so that the transmitted waveform is continuous from the point where the modulation transitions occur at the start of each guard period. The FFT demodulator analyses the interval from ΔT to $T + \Delta T$. In this way multipath delays varying from 0 to ΔT can be tolerated without any modulation transitions intruding into the FFT analysis interval and spoiling the orthogonality of the carrier waves. Unfortunately the guard periods either reduce the throughput of the system or increase its bandwidth in the ratio $T : (T + \Delta T)$.

The outputs of the FFT demodulator are N complex Fourier coefficients, each of which is a phasor representing the demodulated amplitude and phase of the carrier in that FFT slot. For each carrier, conventional demodulation methods may then be used to recover the binary data from the sequence of phasors from consecutive FFT analysis intervals.

Finally the data from the N QPSK decoders is multiplexed back into a single serial data stream which is passed on to the error correction decoder. This can correct errors which typically occur when multipath causes selective fading of some carriers. Improved error correction performance can often be achieved if soft decision information is passed to the error decoder from each QPSK decoder.

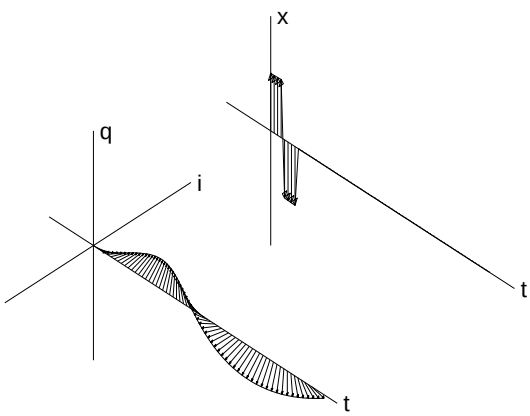

Fig 6.3: OFDM slot 1

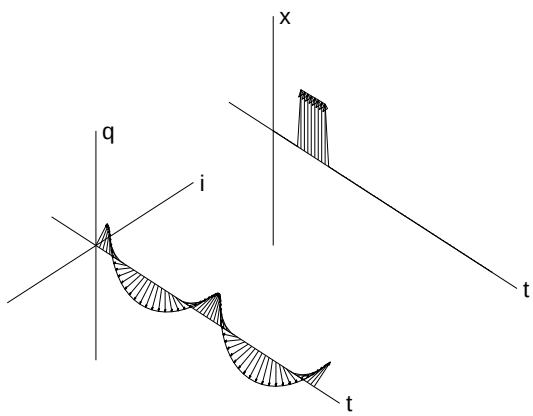

Fig 6.4: OFDM slot 2

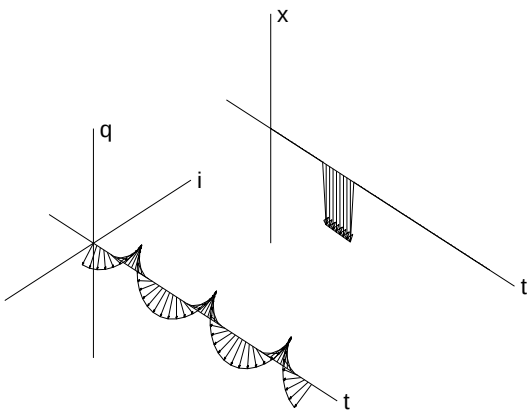

Fig 6.5: OFDM slot 3

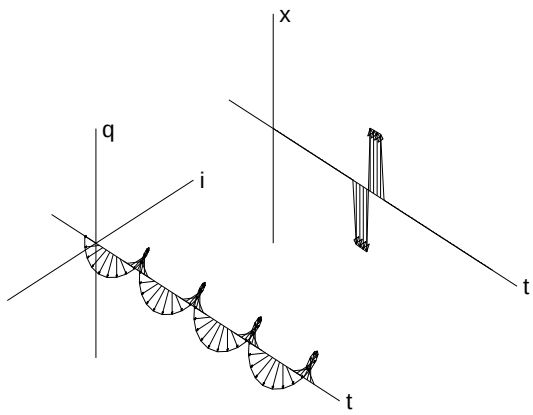

Fig 6.6: OFDM slot 4

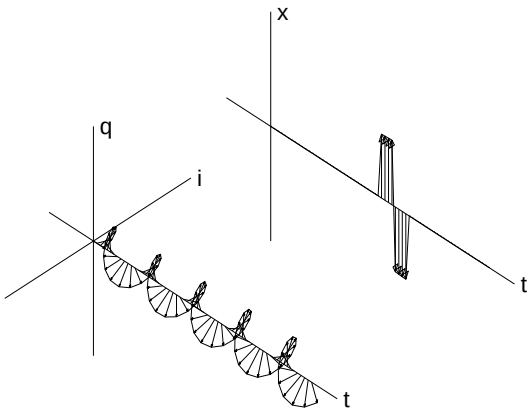

Fig 6.7: OFDM slot 5

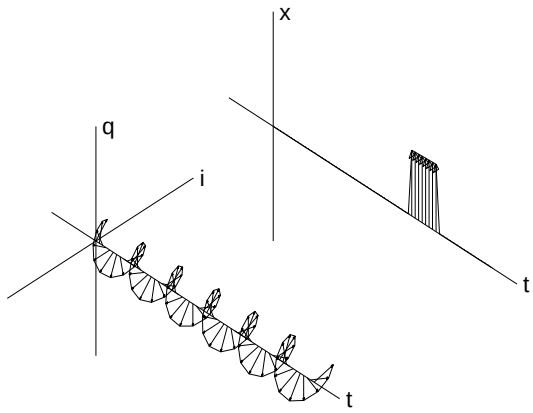

Fig 6.8: OFDM slot 6

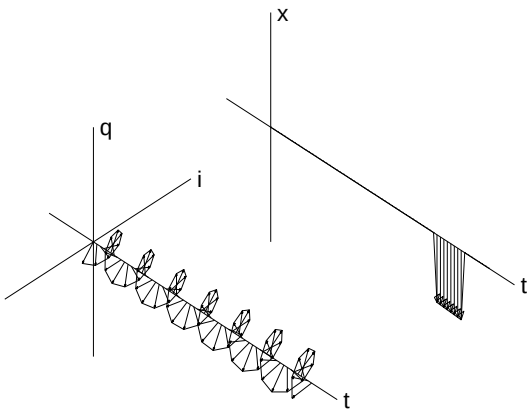

Fig 6.9: OFDM slot 7

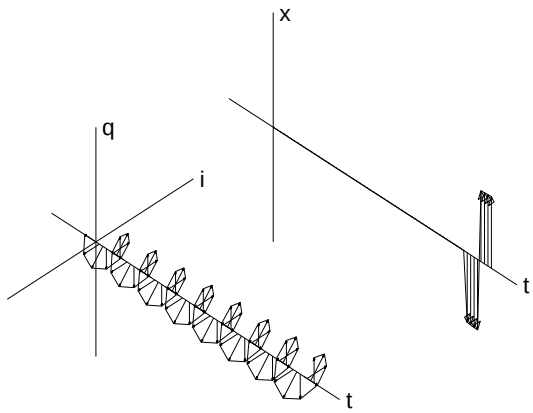

Fig 6.10: OFDM slot 8

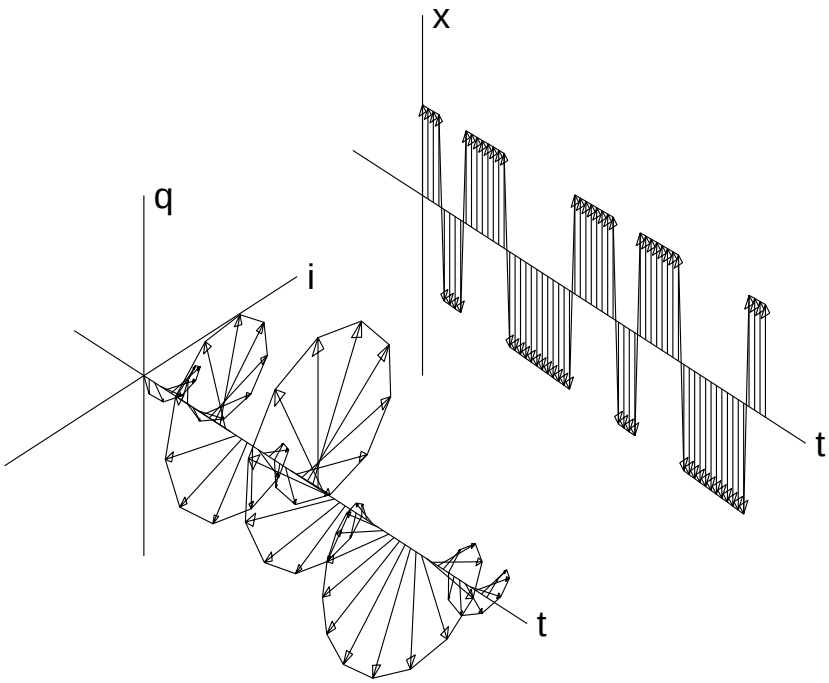

Fig 6.11: OFDM slots 1 to 8 combined

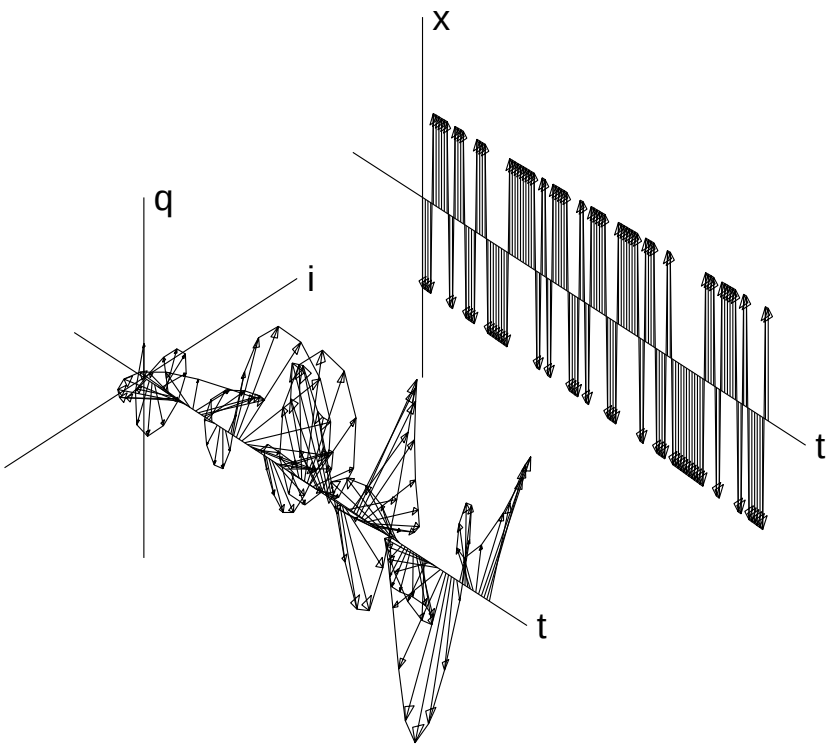

Fig 6.12: OFDM slots −16 to 15 combined

Format of DAB broadcasts in the UK:

- Input data rate of 1 DAB block = 1.23 Mb/s. (**Up to 6 audio channels per block**, 1 stereo channel = 256, 224 or 192 kb/s; 1 mono channel = 128 or 64 kb/s.)
- Error correction code rate = 1/2, yielding coded data at 2.46 Mb/s.
- Number of carriers, $N = 1536$.
- FFT block period, $T = 1$ ms. (Hence the carrier frequency spacing is 1 kHz.)
- Guard period, $\Delta T = 0.246$ ms. This allows path delays to differ by up to 73 km before any ISI occurs. Hence **single frequency operation** is possible across the entire UK, if transmitters are spaced less than 73 km apart.
- Modulation method = QPSK at a rate of $1/(T + \Delta T) = 1/0.001246 = 802.5$ symbol/s = 1605 bit/s on each carrier. (Hence 1536 carriers give a total bit rate of $1536 \times 1605 = 2.465$ Mb/s.) Data is coded as the **change in phase** of each FFT carrier from one block to the next, so that phase uncertainties due to unknown path delays can be cancelled out by differential decoding.
- Bandwidth required = $(1536 + 1)/T = 1.537$ MHz. (1 FFT slot is left vacant at the centre of the band to aid frequency acquisition.) In practise the $\mathrm{sinc}(x)$ sidelobes decay rather slowly so an additional gap of 200 kHz is left between DAB blocks and filtering of the transmitter output is used to suppress sidelobes beyond this gap. The allocated band is 217.5 to 230 MHz.

Spectral efficiency of DAB vs analogue FM

Compare the 1.537 MHz + 200 kHz = 1.737 MHz for a DAB block of 6 stereo signals, with 6 analogue FM stereo channels at 300 kHz spacing = 1.8 MHz (each channel is 180 kHz bandwidth plus a 120 kHz guard region) and we see they have **approximately the same efficiency**.

BUT **single frequency operation** across the UK allows national radio on DAB to consume a tiny fraction of the total bandwidth needed by analogue FM. This is because, for analogue FM, each station needs separate frequency allocations for each adjacent local region, since analogue FM cannot tolerate two transmissions from adjacent regions being received on the same frequency at a receiver which is close to the boundary between two regions. This would cause unpleasant distortion and echo effects on the demodulated audio signal. On the other hand, DAB using OFDM can tolerate multipath delays that include the worst case delays from transmitters in adjacent cells to a user near the boundary of the cells, and so the adjacent cells may use **the same frequency for all cells**. This saves bandwidth by a factor n_f equal to the frequency re-use factor of the analogue system.

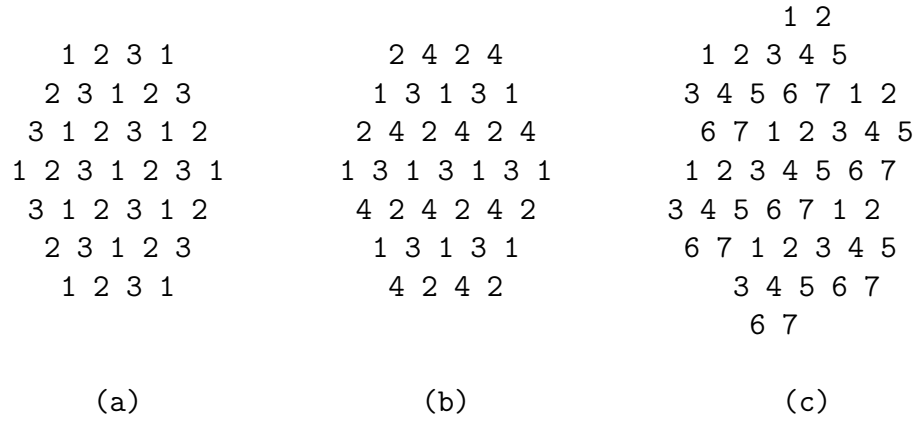

Fig 6.13: Frequency re-use patterns for multiple cells on an approximately hexagonal (close-packed) grid of cells: (a) $n_f = 3$ frequency re-use, $D = \sqrt{3}X$; (b) $n_f = 4$ frequency re-use, $D = 2X$; (c) $n_f = 7$ frequency re-use, $D = \sqrt{7}X$. X is the spacing between cell centres, and D is the distance between the centres of cells using the same frequencies.

Fig 6.13 shows frequency re-use patterns for systems with $n_f = 3$, 4 and 7 frequency slots per channel. We see that as n_f is increased, the distance between pairs of cells with the same frequency allocation increases in proportion to $\sqrt{n_f}$. Typical analogue FM systems, such as that in the UK, need a reuse pattern with $n_f = 7$ in order to keep mutual interference down to acceptable levels. Hence the bandwidth saving, by adopting DAB instead of FM for national radio stations such as BBC R1, R2, R3 and R4, is at least 7:1.

Local radio is different, because we need separate content for different local regions. To accommodate this, local DAB needs $n_f = 4$ DAB blocks of 1.737 MHz each for the whole UK so that adjacent regions can use different frequencies for transmissions with differing content. Digital transmissions are much more tolerant than analogue ones to low level interference from other stations sharing the same frequency, so a frequency reuse pattern with $n_f = 4$ is acceptable for digital local radio. This is significantly better than the $n_f = 7$ pattern needed for analogue FM and still results in a useful saving of bandwidth by 'going digital'.

6.3 Digital TV

Digital video broadcasting (DVB) uses different modulation techniques for the 3 main transmission methods: cable, satellite, and terrestrial radio waves. We shall concentrate on the terrestrial radio-wave system (FreeView in the UK) because it is more interesting.

For all 3 transmission methods, **capacity** is a key parameter so MPEG-2 or MPEG-4 compression is used, and achieves a bit rate for the composite video and audio of approximately 3 Mbit/s for a single normal definition broadcast programme, and of around 6 to 8 Mbit/s for a high-definition (HD) programme. (Video compression methods are discussed at the end of the 4F8 Image Coding course and audio ones are similar to those for DAB.) Like DAB, typically 6 or more programmes are time-division multiplexed into a composite bit stream, but in DVB the composite bit rate is set at 24.13 Mbit/s.

The main problem then becomes:

How best to transmit 24.13 Mb/s over terrestrial radio links ?

(Satellite and cable are much less of a problem because there are no multipath effects and the channels are much better defined.)

Once again COFDM is the chosen technique. However **bandwidth efficiency** is much more of an issue than for DAB because of the higher bit rates, so shorter frames and a much shorter guard period are used. The fact that most DVB receivers will use fixed antennas with strongly directional beams, means that the multipath delay spread is likely to be very much less than for DAB, so short guard periods are acceptable. To achieve good bandwidth efficiency, **64-QAM** is used instead of QPSK. This works for DVB, whereas it would not work for DAB, because the path of a DVB signal, from fixed base station to fixed antennas, is relatively stable. Hence we can tolerate the 8dB loss of performance with 64-QAM and also not be concerned about the fact that QAM receivers, with their amplitude-sensitive comparators, have difficulties with rapidly fluctuating path losses when the path is unstable (such as a DAB path to a moving vehicle).

The key parameters for DVB in the UK have therefore been selected to be:

- Input data rate of 1 DVB block = 24.13 Mb/s, with currently **6 or more video channels per block** at typical rates from 2 to 8 Mb/s each. Higher bit rates are allowed on selected channels if required; e.g. for fast-moving sports programmes in HD. The channel bit rates can be **adaptively reallocated** within the 24.13 Mb/s total as the programme content changes **if** the broadcasters are clever.

- Error correction coding: (204,188) Reed-Solomon outer code with rate = 188/204; and a convolutional inner code with rate = 2/3, yielding coded data at $(204/188) \times (3/2) \times 24.13 = 39.27$ Mb/s.

- Number of carriers, $N = 1705$; but 193 of these carriers are used as pilot tones, so the number of data carrying carriers is $1705 - 193 = 1512$. 45 of the 193 pilot tones are fixed at regular intervals across the band, and 131 are moved around in a pseudo-random way. These are designed to provide phase and amplitude references to aid the receiver in decoding the data tones. A further 17 pilot tones carry transmitter parameter signalling information.

- FFT block period, $T = 0.224$ ms. (Hence the carrier frequency spacing is 4.464 kHz.)
- Guard period, $\Delta T = T/32 = 0.007$ ms. This allows **path delay differences of up to 2.1 km** before any ISI occurs, and copes with likely multipath spreads to fixed antennas. Single frequency operation is not possible in this case, but a four-frequency reuse pattern for the whole UK is feasible and allows for local as well as national stations.
- Modulation method = 64-QAM at a rate of $1/(T + \Delta T) = 1/0.000231 = 4329$ symbol/s. This provides $6 \times 4329 = 25974$ bit/s on each carrier. Hence 1512 carriers give a total bit rate of $1512 \times 25974 = 39.27$ Mb/s. The pilot tones provide phase and amplitude references, so differential coding (and its associated performance penalty) can be avoided.
- Bandwidth required = $1705/T =$ **7.612 MHz**. This is very similar to a single programme of analogue TV with an FM audio sub-carrier. However there are now **6 or more programmes in this bandwidth**, and a **more efficient frequency reuse pattern** is feasible ($n_f = 4$ instead of $n_f = 7$).

A full description of the DAB options are given in the European Telecommunications Standards Institute (ETSI) standard ETS 300 401, and of the DVB options in ETS 300 744, both available from

http://www.etsi.org/

but these are not light reading. In the DVB standard, there are over 1000 possible different combinations of parameters that can be used!