

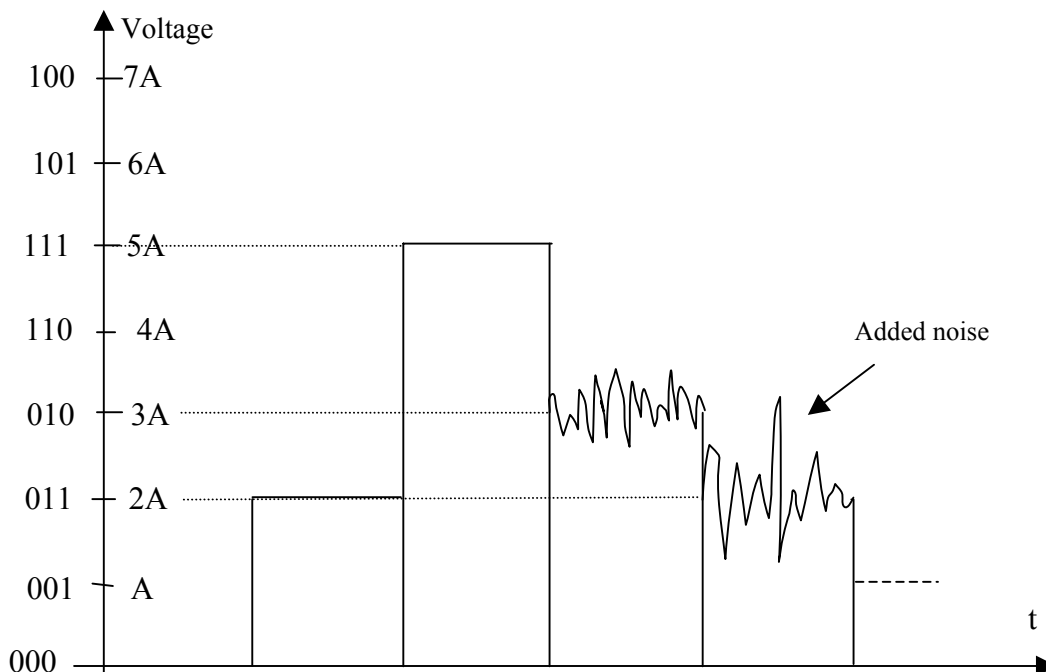
University of Manchester
School of Computer Science
CS3282: Digital Communications '06
Section 9: Multi-level digital modulation & demodulation.

9.1. Introduction:

So far, in looking at carrier modulated data transmission, we have concentrated mainly on binary signalling using simple forms of ASK, FSK and PSK. With a suitable pulse-shape, binary signalling can achieve a bandwidth efficiency of up to 2 bits/second per Hz using real unipolar pulses (base-band equivalent of ASK) or real bipolar pulses (base-band equivalent of PSK) at base-band. Since multiplying the base-band signal by a sinusoidal carrier doubles its bandwidth, the maximum band-width efficiency of binary ASK and binary PSK becomes one bit/second/Hz. With this band-width efficiency, a computer modem could achieve a maximum of about 3.1 kb/s over a 300-3400 kHz domestic telephone link. This is less than one tenth of what we know to be achievable. A similar calculation for the bandwidth-efficiency of binary FSK is a little more complicated, but if we take the frequency spacing used by MSK, i.e. $1/(2T)$ Hz, which is the minimum spacing compatible with the desirable highly property of 'orthogonality' between the 2 symbols, we may expect to achieve a maximum of slightly less than 2 bit/second per Hz. To increase the band-width efficiency over what can be achieved with binary signalling we must use multi-level modulation schemes where each symbol represents more than one bit.

9.2. Multi-level ASK and Gray coding:

Perhaps the most obvious multi-level modulation technique, though rarely used, is 'M-ary ASK' which modulates the amplitude of a sinusoidal carrier with M different amplitudes. To encode N bits per symbol, rather than just one as with binary signalling, choose $M=2^N$ rather than $M=2$. For example, to encode 3 bits per symbol, we could have 8 rectangular symbols of heights 0, A, 2A, 3A, 4A, 5A, 6A and 7A volts. Consider what happens to the 'bit-error probability', P_B , when we adopt this new signalling strategy. We can hope to keep P_B approximately the same as for binary signalling by using the same voltage separation (A in our example) between the levels of the rectangular symbols. However, a noise level which would previously have given only a single bit-error with binary signalling could, for example, change 011 to 100 producing three bit errors. This unfortunate complication is easily avoided by using a 3-bit "Gray-code" to allocate sequences of bits to signals as illustrated below.



Assuming $s_0, s_1, s_2, \dots, s_7$ remain at 0, A, 2A, ..., 7A volts respectively, Gray code allocates 000 to s_0 , 001 to s_1 , 011 to s_2 , 010 to s_3 , 110 to s_4 , 111 to s_5 , 101 to s_6 and 100 to s_7 . Now a noise level changing s_n to s_{n-1} or s_{n+1} will result in only one bit-error as the 'distance' between the 3-bit pattern for adjacent symbol levels is always one. If we neglect the possibility of such a large noise sample coming along that say s_3 becomes s_5 or s_1 (this is a fairly safe assumption at reasonably low bit-error probabilities) we can assume that, with Gray-coding, the number of bit-errors per second will be very close to the number of symbol errors per second. This means that the 'bit-error probability' will be equal to the 'symbol error probability' divided by M, the number of bits per symbol. Remember that the bit-error probability P_B measures the number of bit-errors in a quantity of bits (e.g. 1 in 1000) rather than the number of bit-errors per second. Note that if we use M-ary signalling to reduce the bandwidth required for a given bit-rate, the noise power reduces as the bandwidth reduces.

Keeping the same value of A for 8-ary ASK as is used for binary ASK will normally not be practical as the power and amplitude of the transmission will be increased. To keep the average power the same we must reduce A and this will then increase the bit-error probability as the voltages between decision thresholds will be reduced.

Exercise: A binary ASK transmitter & receiver communicate at 1000 bits/second with a bit-error probability of 10^{-9} . There is no matched filter. To increase the bit-rate to 3000 bits/second over the same channel, the binary ASK equipment is replaced by an 8-ary transmitter & receiver with the same transmitter power. Estimate the new bit-error probability.

Solution

With 0 and A rectangular pulses without a MF, if $P_B = 10^{-9} = Q(A/(2\sigma))$, $A/(2\sigma) = 6$

Av power = $E_B \times (1/T) = (A^2/2)T/T = A^2/2$

With 8-ary ASK (0, B, 2B, ... 7B), av energy per symbol is:

$$A_v \{0, B^2T, (2B)^2T, \dots, (7B)^2T\} = (140/8)B^2 = 17.5B^2T$$

So av power = $17.5B^2$ Same av power, so $17.5B^2 = A^2/2$ & $B = A / 5.9$

Same noise power, so Prob of a symbol error = $(14/8)Q(B/(2\sigma)) \approx 2Q((A/35) / (2\sigma))$
 $= 2Q(6/5.9) = 2Q(1.02) = 0.24$

Note that Prob of symbol error $\approx 2Q(B/(2\sigma))$ because a noise spike $> B/2$ or $< -B/2$ will cause a symbol error for 6 out of 8 symbols. It would be more accurate to replace 2 by (14/8) since the minimum voltage gives a symbol error only when a noise spike $> B/2$ and the max voltage only for -ve noise spike $< -B/2$.

Assuming Grey coding used, $P_B \approx (\text{Symbol error prob}) / 3 = 0.08$

Exercise: A binary ASK transmitter & receiver communicate at 1000 bits/second with a bit-error probability of 10^{-9} . To reduce the channel bandwidth required for the same 1000 bits/second transmission, the binary ASK equipment is replaced by an 8-ary transmitter & receiver with the same transmitter power. Estimate the new bit-error probability.

The advantage of multi-level signalling over binary is an increase in bandwidth efficiency (b/s per Hz). The disadvantage with multi-level ASK is an increase in the bit-error probability for a given transmission power. Multi-level ASK could be an effective method where the noise is relatively low so that we can safely reduce the voltage between detection thresholds without incurring large increases in bit-errors. This is normally the case with modem transmissions over telephone lines.

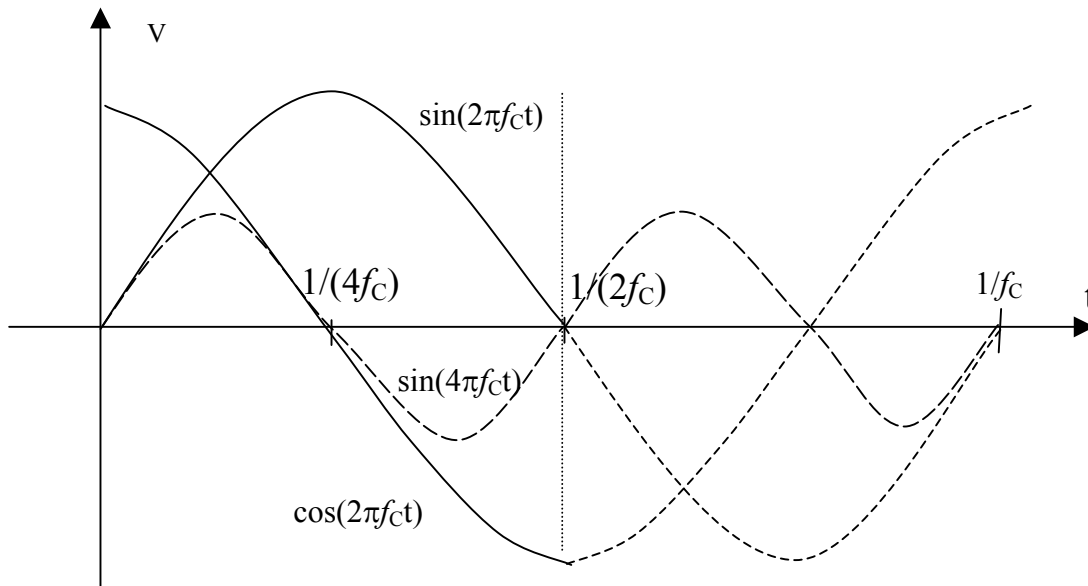
9.3. Orthogonality:

The increase in bit-error probability incurred by changing from binary to multi-level signalling can be avoided if the additional bits are conveyed by signalling which is 'orthogonal' to the original binary signalling. This can be achieved with '4-phase' PSK (known as QPSK) which is a multi-level signalling technique. It can also be achieved with multi-level FSK. Orthogonality is definitely not present with multi-level ASK as presented above.

Signals $a(t)$ and $b(t)$ are orthogonal over a period R seconds if:

$$\int_0^R a(t)b(t)dt = 0$$

It is easily shown that $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$ are orthogonal signals when R is an integer multiple of $1/(2f_c)$. When $R=1/(2f_c)$ then R is the duration of one half-cycle of the carrier. One simple way to visualise this result is to remember that $\cos(2\pi f_c t) \sin(2\pi f_c t) = 0.5\sin(4\pi f_c t)$ and examine the following graph noting the equal area of $0.5\sin(4\pi f_c t)$ above and below the t axis in the range $0 < t < 1/(2f_c)$. This also applies for the range $0 < t < 1/f_c$.



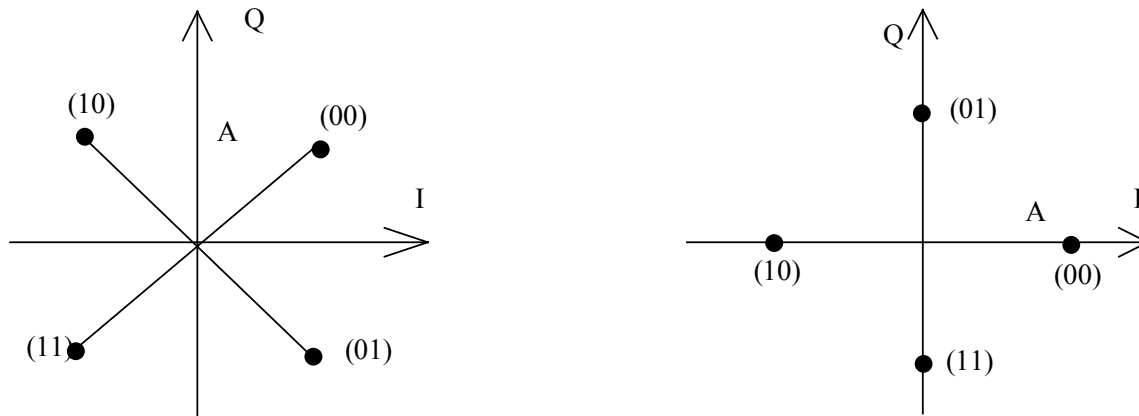
QPSK can transmit two bits per symbol without increasing the bit-error probability over what is obtained, for a given E_B/N_0 , with binary PSK transmitting one bit per symbol. This is made possible by the orthogonality of the 'in-phase' and 'quadrature' components of a sinusoidal carrier which allows us to introduce two extra symbols into a binary PSK scheme in such a way that the extra symbols are invisible to the coherent detector used for the two original symbols. Similarly a coherent detector for the two extra symbols can be designed in such a way that it is 'blind' to the two original symbols. Such orthogonality requires a suitable choice of symbol interval T equal to an integer multiple of $1/(2f_c)$ and a detector that performs an integration over one or more of these symbol intervals. A vector-demodulator with low-pass filters removing all spectral power above $2f_c - B/2$ where B is the bandwidth of the QPSK signal will also recover the complex base-band QPSK signal with real & imaginary parts completely invisible to each other. QPSK will still work even if exact orthogonality is not achieved, but the lowest possible bit-error probability requires T and the detection process to be chosen to meet the orthogonality conditions outlined above.

Binary FSK can be extended to multi-level FSK of any level without increasing the bit-error probability if the FSK frequencies and symbol interval are chosen such that all the FSK symbols are orthogonal. It may

be shown (Sklar pp 152-154) that $\cos(2\pi f_0 t + \phi)$ and $\cos(2\pi f_1 t)$ are orthogonal over R seconds for any value of ϕ when R is an integer multiple of $1/|f_1 - f_0|$. Also $\cos(2\pi f_0 t)$ and $\cos(2\pi f_1 t)$ are orthogonal over R seconds when R is an integer multiple of $1/(2|f_1 - f_0|)$. Multi-level MSK is a form of FSK which achieves this orthogonality as will be discussed later.

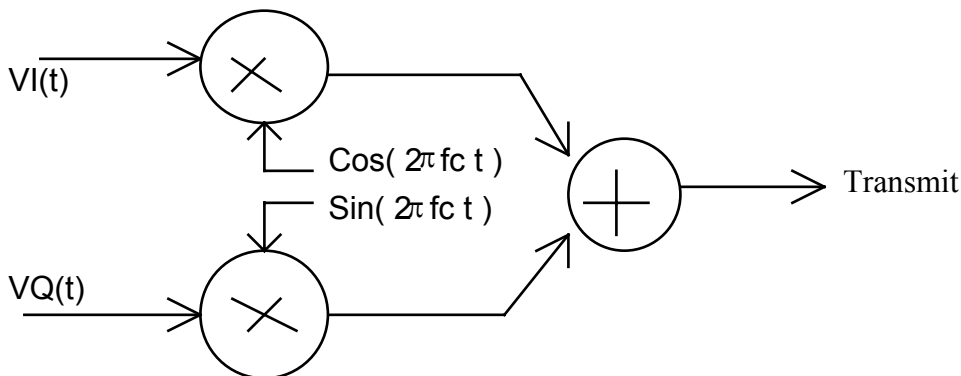
9.4. QPSK and orthogonality

A constellation diagram as shown below (left) describes a QPSK system with 2-bits per symbol, each symbol being a T second segment of sine-wave, of constant amplitude A say, with the same frequency f_c Hz = $2\pi\Omega_c$ as the carrier and with phase lead of respectively 45, 135, -135 and -45 degrees with respect to the carrier. The required modulation may be achieved by applying pairs of bits to a symbol allocation unit producing a pair of voltages as described by the symbol allocation table shown below, and applying the two voltages obtained to a vector modulator (shown below). The constellation on the right is an alternative form for QPSK with phase leads of 0, 90, 180, and 270 degrees. We will call this the "St George's cross" QPSK constellation to distinguish it from the "St Andrew's" or "St David's" cross QPSK constellation on the left. Pulse-shaping, to minimise ISI when the symbols are band-limited, is applied to the voltages $v_I(t)$ and $v_Q(t)$ before they are multiplied by the carrier; i.e. instead of being constant they are actually shaped waveforms. N.B. In the diagram on the left, the bits are reversed as compared with the QPSK scheme in Section 8.

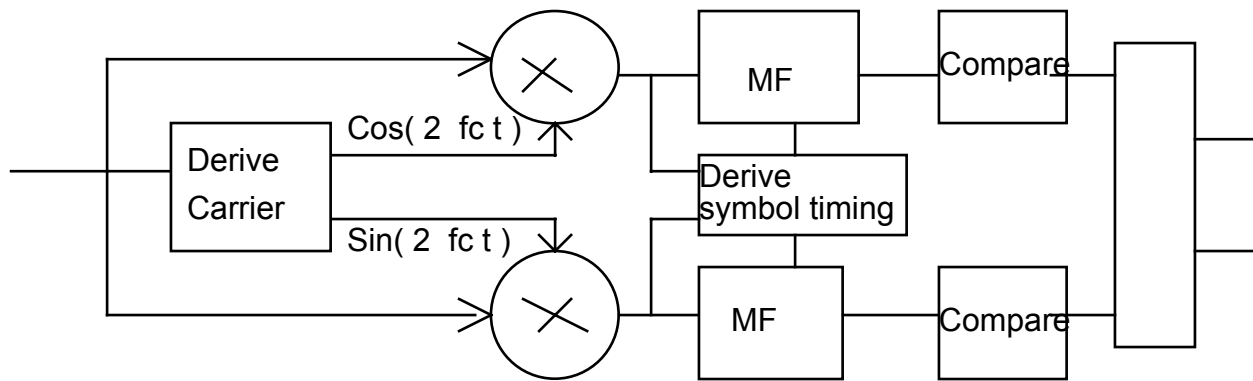


Bit 1	Bit 2	VI	VQ
0	0	A	A
0	1	A	-A
1	0	-A	A
1	1	-A	-A

Bit 1	Bit 2	VI	VQ
0	0	A	0
0	1	0	A
1	0	-A	0
1	1	0	-A



A coherent QPSK detector for the left (St. Andrew's) constellation is as shown below. This is, essentially, two independent PSK detectors:



A coherent QPSK detector for the right (St. George) constellation is a little different and illustrates a more general principle: i.e. a detection strategy that minimises the 'Euclidean' distance as seen on the constellation diagram between the received symbol and each of the four possibilities. This requires us to examine the 'In Phase' and 'Quadrature' components (real and imag) together, and we no longer have two independent PSK detectors. There are several efficient ways of implementing the 'minimum distance' detection strategy which involve dividing the constellation diagram into regions, each region being represented by one symbol. This is particularly easy with the 'St.George' QPSK constellation, but is also adaptable to more complicated constellations with many more than four symbols.

With binary PSK, the two symbols are not orthogonal. They are 'anti-podal' or 'bipolar' which means of identical shape but of opposite signs. To achieve QPSK, we add a quadrature component to each binary PSK symbol. To achieve exact orthogonality, we must use a symbol interval, T , equal to an integer multiple of half the carrier period and use an integrator as a detector. With QPSK, each of the four possible symbols can be made orthogonal to two others and anti-podal to one other. It is not essential to have orthogonality with QPSK, but if we do have it, the bit-error probability for a given E_B/N_0 will be exactly as for binary PSK at half the bit-rate. Doubling the bit-rate will have been achieved without increasing the bit-error probability from what is obtained with binary PSK transmitting at the same power level.

The coherent detectors must use matched filters or correlation detectors (e.g. integrate & dump if pulse shaping not present) for the shaped "I" and "Q" symbols. This makes the "Quadrature-channel" and the "In phase-channel" truly independent and effectively we get two channels for the price of one.

Other forms of QPSK: $\pi/4$ -QPSK, DQPSK, etc. See text-book.

9.5. Carrier derivation:

A reasonable method of deriving the carrier for binary psk is to square the incoming signal. If the phases are zero & 180 degrees; i.e if we transmit $A\cos(2\pi f_c t)$ or $A\cos(2\pi f_c t + \pi) = -A\cos(2\pi f_c t)$, this gives $A^2 \cos^2(2\pi f_c t) = 0.5A^2(1 + \cos(2\pi 2f_c t))$.

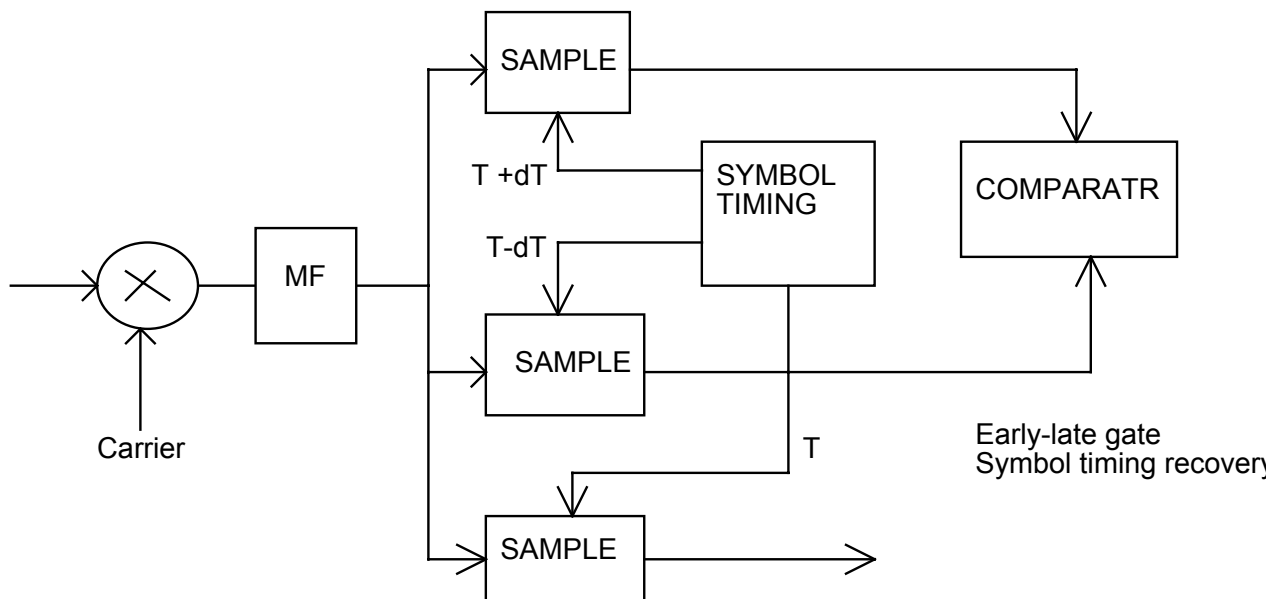
Removing the constant component gives us a cosine wave of twice the frequency needed. A phase locked loop can now be locked to this frequency and, by various methods a frequency halving may then be carried out. For example, the signal can be made into a square-wave and then into a 10101010 logic sequence which may be passed through a "divide by two counter". The resulting digital signal may then be low-pass filtered to obtain a sine-wave of the correct halved frequency. Alternatively, the phase-locked loop itself can be made to square its reference before the multiplication by the incoming signal so that the reference will have half the required frequency. The halving in frequency, being somewhat like a

square rooting process, will usually leave us with a sign ambiguity. A known training signal is often sent from time to time to allow this ambiguity to be resolved.

For QPSK the carrier derivation process may be as above except that the fourth power of the incoming signal (rather than the square) must be calculated to eliminate the modulation, and a frequency divide by four must be implemented.

9.6. Symbol timing recovery:

Even when the carrier has been derived, there remains the problem of determining when one symbol ends and the next one begins. Of course, the value of T , the symbol duration, will be known exactly, and in some cases it may be possible to assume that the symbols are locked to the carrier, so that say one symbol is 1.5 carrier cycles for example. This is not always possible however since very slight inaccuracy in any previous demodulation process, say from 900 MHz to some intermediate frequency, or Doppler shifts due to receiving signals from moving vehicles, and/or other effects can remove this synchronism. A popular method for symbol timing recovery is the “early-late gate” method. This aims to adjust its sampling time to make the samples obtained a maximum. There are two detectors, one fed with a slightly early timing reference $T + dT$ and the other with a slightly late timing reference $T - dT$ from the “symbol timing” module shown in the diagram below. The outputs of the two detectors are periodically compared (by the comparator module) to see which is producing the larger output. The timing is then advanced or retarded slightly according to whether the output is larger at $T + dT$ or $T - dT$. This should increase the larger output, and as this process is repeated it will be maximised. Further adjustments will cause the two samplers to settle around a situation where both produce approximately the same output, the true sampling point being equidistant between the early one $T + dT$ and the late one $T - dT$. This true sampling point is used for a third accurate sampler.



We have now seen most of the principles used in digital modulation, and can now discuss some of the more advanced multi-level digital modulation schemes.

9.7. Non-coherently detected M-ary FSK:

We have seen how binary FSK may be non-coherently detected and how it is advantageous to have the difference between the two signalling frequencies equal to $1/T$ Hz. This places a null in the magnitude spectrum created by transmissions at one frequency at the other frequency. It is easily shown (Sklar pp 152-154) that this means that the non-phase synchronised binary FSK symbols $A \cos(2\pi f_0 t + \phi)$ and

$A \cos(2\pi f_1 t)$ are orthogonal over the symbol period T seconds when $|f_1 - f_0| = 1/T$ regardless of the phase difference ϕ between the symbols.

Exercise: Check this out.

It may be argued that if the symbol rate is $1/T$, the minimum frequency separation between tones for non-coherent orthogonal FSK signalling is $1/T$ Hz.

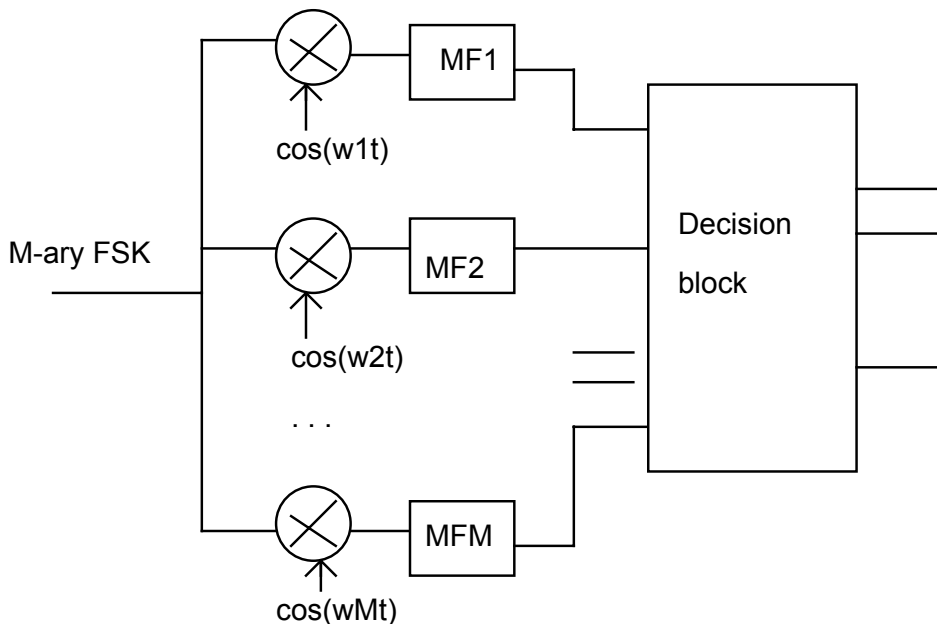
Of course non-orthogonal non-coherent (or coherent) FSK signalling could be used, but the effect of noise and inter-symbol interference would be much greater.

The idea of binary FSK can be extended to M-ary FSK simply by introducing more tones. An 8-ary FSK system at 1800 bits/sec (\therefore 600 baud) with minimum frequency spacing for orthogonal non-coherent detection could have tones at 1000Hz (for 000) 1600Hz (001), 2200 Hz (011), 2800Hz (010), 3400Hz (100), 4000Hz (101), 4600Hz (111), 5400Hz (110). Because of the orthogonality of the symbols, the BER for this scheme would be about the same as for binary FSK given the same noise PSD, though the bandwidth is wider.

Exercise: Why are the symbols Gray-coded?

9.8. Coherently detected orthogonal M-ary FSK

Where locally generated tones locked in frequency and phase to the received tones can be generated at the receiver, greater efficiency can be obtained.



It is easily shown (Sklar pp 152-154) that the phase synchronised binary FSK symbols $A \cos(2\pi f_0 t)$ and $A \cos(2\pi f_1 t)$ are orthogonal over the symbol period T seconds when $|f_1 - f_0| = 0.5/T$.

Exercise: Check this.

This means that with coherent detection, the minimum frequency spacing for orthogonal M-ary FSK is not $1/T$ but $0.5/T$ Hz. This is a very significant saving in bandwidth with no increase in BER.

The signalling technique is now known as M-ary minimum shift keying or MSK.

It is argued (Sklar) that for a given signal-to-noise ratio and required bit-rate below the Hartley-Shannon limit, M-ary MSK with increasing M and decreasing symbol rate (to maintain a constant bit-rate) may be used to approach any arbitrarily low BER. The lower the BER, the higher M and the longer the averaging

time T which will reduce the effect of the noise, but the more complicated will be the coder & decoder . This is not necessarily the best practical approach to achieve a given BER but it shows us what can be done in theory. Remember that if E_b/N_0 is less than -1.6 dB (where E_b is the energy per bit and N_0 is the noise PSD) there is no bit-rate for which we can say that any given BER can be achieved by this or any other method.

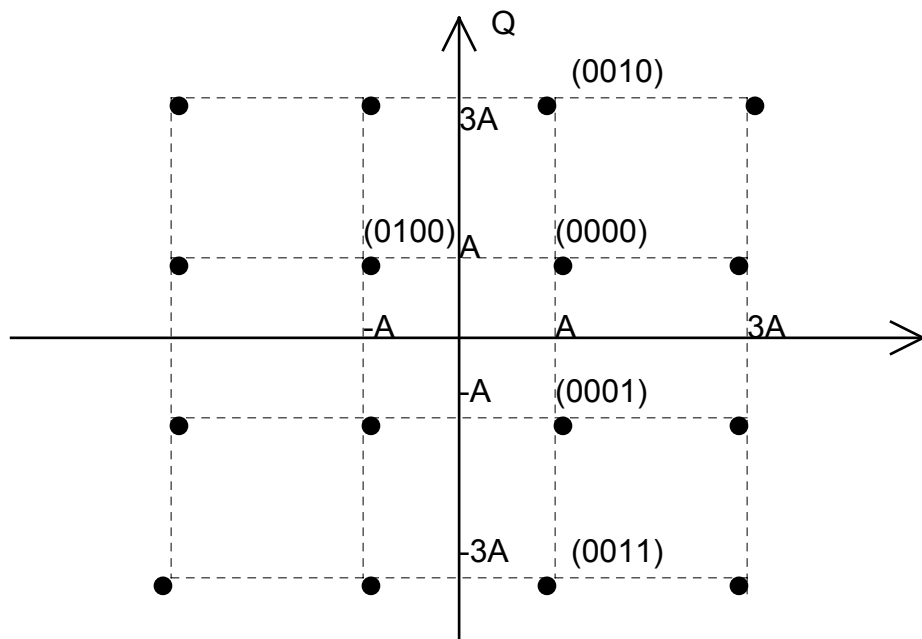
Binary MSK is used with Gaussian symbol shaping prior to the modulation (similar to the raised cosine shaping we have seen) is used in GSM mobile radio systems.

9.9. Combined amplitude and phase shift keying (QAM and APK)

To approach the bandwidth efficiency of modern computer modems, a combination of ASK, FSK and PSK may be used. The most popular combination is amplitude and phase shift keying referred to as M-ary “amplitude-phase keying” (APK) or sometimes as “quadrature amplitude modulation” (QAM). This may be considered an extension of QPSK where different amplitudes (apart from $+A, 0$ and $-A$ as used up to now) are used to modulate the in-phase and quadrature carrier. Consider the 16-QAM system below where 4-bits are used to produce the following base-band symbols:-

Bit 1	bit 2	bit 3	bit 4	VI	VQ	Bit 1	bit 2	bit 3	bit 4	VI	VQ
0	0	0	0	A	A	1	0	0	0	3A	A
0	0	0	1	A	-A	1	0	0	1	3A	-A
0	0	1	0	A	3A	1	0	1	0	3A	3A
0	0	1	1	A	-3A	1	0	1	1	3A	-3A
0	1	0	0	-A	A	1	1	0	0	-3A	A
0	1	0	1	-A	-A	1	1	0	1	-3A	-A
0	1	1	0	-A	3A	1	1	1	0	-3A	3A
0	1	1	1	-A	-3A	1	1	1	1	-3A	-3A

Applying VI and VQ (normally after shaping) to a vector modulator as used to generate QPSK gives a signal whose constellation diagram is shown below:



Exercise: Is this constellation Gray coded? If not, how would you recode the symbols.

Exercise: Specify the 16 symbols in terms of amplitude and phase.

Exercise: If the bit-rate is 1200 bit/s and the carrier is 600 Hz, sketch the waveform produced (without band-limiting or shaping) when 0000 0101 1010 1011 is transmitted.

Exercise: Calculate the average power (normalised to 1 Ohm) of the “square 16-QAM” modulator when all symbols are equally likely. What is the peak voltage produced?

Solution: Since a sine-wave of amplitude A has power $A^2/2$: the average power of the 16-QAM modulator is: $4(2A^2/2 + 18A^2/2 + 10A^2/2 + 10A^2/2)/16 = 5A^2$ Watts.

Peak voltage: $\sqrt{18}$ A Volts

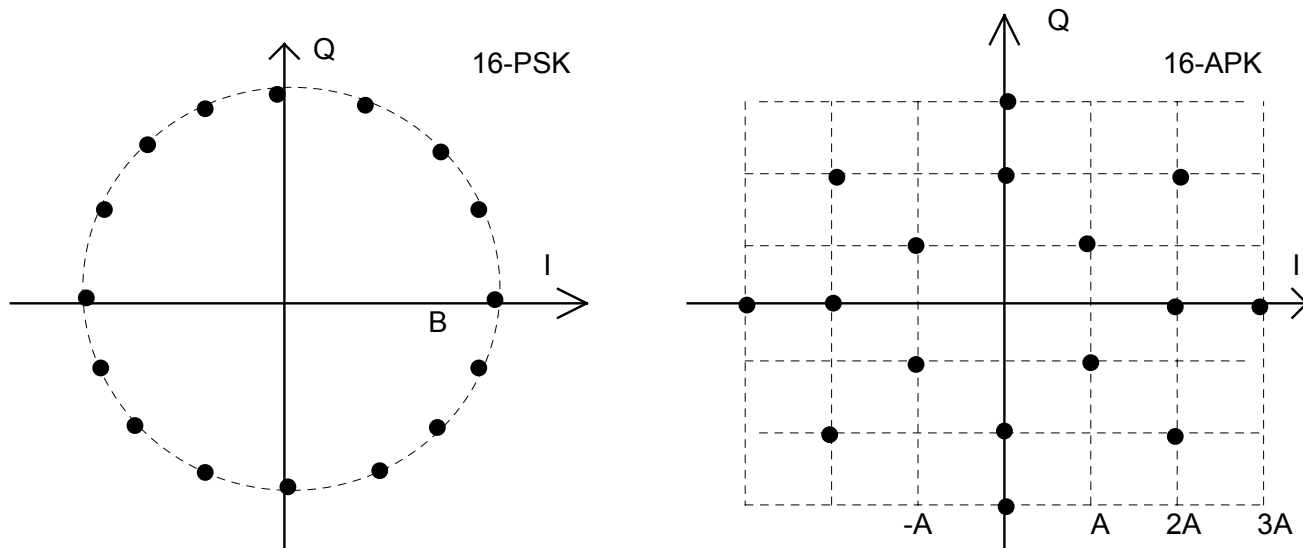
It is interesting to compare “square 16-QAM” with other possible schemes. Consider the 16-PSK scheme and the 16-APK schemes whose constellations are shown below.

The average power and peak voltage of 16-PSK is clearly $B^2/2$ Watts (normalised to 1 Ohm) and B volts respectively. If $B^2 = 10A^2$, i.e. if $B \approx 3.16A$, the power for the square 16-QAM and the 16-PSK will be the same. The peak voltage for 16-PSK will then be $B=3.16A$ Volts as opposed to 4.24A Volts for square 16-QAM, so 16-PSK is better in this respect. What about the sensitivity to noise when the average power is equalised by making $B=3.16A$?

The minimum distance between any two symbols for square-16 QAM is $2A$, whereas it is approx.

$B \cdot 2\pi/16 = 6.4A\pi/16 = 1.24A$. (it is a bit less than this, actually) The symbols are closer for 16-PSK, hence the effect of noise will be greater.

Exercise: Check these figures, and repeat for the 16-APK scheme shown below.



Solution: Let’s compare all three, i.e. ‘square 16-QAM, 16-PSK and our 16-APK scheme.

Replace A in the 16-APK diagram by C. The average power is:

$$4((2C)^2/2 + (3C)^2/2 + 2C^2/2 + 8C^2/2) /16 = 2.875C^2 . \text{ Therefore:}$$

	<u>Av power</u>	<u>Pk voltage</u>	<u>Min distance</u>
Square 16-QAM	$5A^2$	4.243A	2A
16-PSK	$B^2/2$	B	$\pi B/8$
16-APK scheme	$2.875C^2$	3C	C

For same average power (i.e. same E_B/N_0):

$$C = 1.32A \text{ and } B = 3.16A$$

For same peak voltage:

$$C = 1.414A \text{ and } B = 4.243A$$

Minimum distance on constellation diagram (taken as a measure of noise immunity):

	<u>Min distce for same av. power</u>	<u>Min distce for same pk voltage</u>
Square 16-QAM	2A	2A
16-PSK	1.24A	1.667A
16-APK scheme	1.32A	1.414A

Square QAM appears ‘best’ in both circumstances. It has the greatest minimum distance in both circumstances. I am slightly surprised by the ‘peak voltage’ result and expected 16-PSK to be better than square-16-QAM in this case. Check result please.

CCITT modem standards:-

Version	Bits per second	Modulation	Protocol
Bell 103	300	FSK	Async
Bell 202	1200	FSK	Async
V22	1200/600	QPSK/FSK	Async/sync
V26bis	2400	QPSK	Sync
V27	4800/2400	8-DQPSK	sync
V29	9,600	16-APK	sync
V32	9,600	32-QAM/16-QAM	sync
V33	14,400	32-QAM	sync
V34	33,600	>1024-QAM	sync
V90	56,000	>1024-QAM	sync

9.10. Problems:

P9.1: A digital system transmits a 0Hz to 2MHz video signal sampled at 8MHz with a 16 bit analogue to digital converter. The signal is transmitted by 16-QAM with raised cosine spectrum pulse shaping with roll-off factor α (or r) equal to 0.5. What is the transmission bandwidth needed.

Solution: Bit-rate = 128Mb/s. Four bits/symbol. This would give 4 bits/sec/Hz if $\alpha=0$, but as $\alpha = 0.5$ the bandwidth efficiency is reduced by $(1+\alpha) = 1.5$. This is because of the spectrum of the shaped pulse. See previous notes. It becomes 2.66 bits/sec per Hz. Therefore required bandwidth is 48 MHz. This is considerably greater than the analogue bandwidth.

P9.2: An 8-ary ASK scheme uses a root-raised cosine spectrum filter in both transmitter and receiver, with $\alpha=0.33$. What bandwidth is required for 64 kb/s?

P9.3: What are the advantages and disadvantages of coherent detection as opposed to non-coherent detection for 8-FSK?

P9.4: Devise a suitable constellation and symbol assignment for (a) 8-APK, (b) 32-QAM.

P9.5: Compare the following ‘circular-16’ QAM scheme with the three schemes mentioned in the notes.

