Intuitive Guide to Principles of Communications
www.complextoreal.com

All About Modulation - Part I

Basic Concepts, Signal Space, Constellations and Phase Shift Keying modulations (PSK, QPSK, OQPSK, M-PSK, π/4-QPSK, MSK, and GMSK)

Basic Concepts of modulation

Three kinds of modulations

Modulation is the process of facilitating the transfer of information over a medium. Voice can not be sent be very far by screaming. To extend the range of sound, we need to transmit it through a medium other than air, such as a phone line or radio. The process of converting information (voice in this case) so that it can be successfully sent through a medium (wire or radio waves) is called modulation.

We begin our discussion of digital modulation by starting with the three basic types of digital modulation techniques. These are;

Amplitude-Shift Keying (ASK)
Frequency-Shift Keying (FSK)
Phase-Shift Keying (PSK)

All of these techniques vary a parameter of a sinusoid to represent the information which we wish to send. A sinusoid has three different parameters than can be varied. These are its amplitude, phase and frequency.

In ASK, the amplitude of the signal is changed in response to information and all else is kept fixed. Bit 1 is transmitted by a signal of one particular amplitude. To transmit 0, we change the amplitude keeping the frequency constant. On-Off Keying (OOK) is a special form of ASK, where one of the amplitudes is zero as shown below.

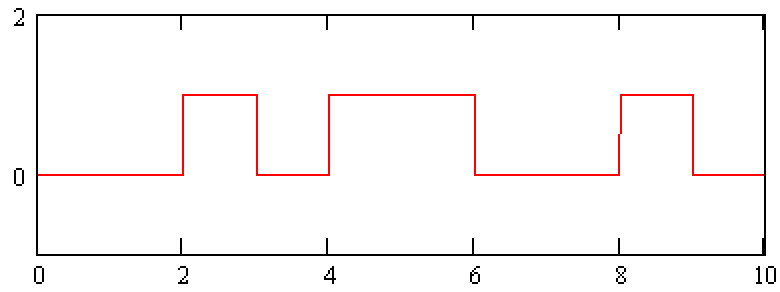


Figure 1 - Baseband information sequence – 0010110010

$$ASK(t) = s(t)\sin(2\pi f t)$$

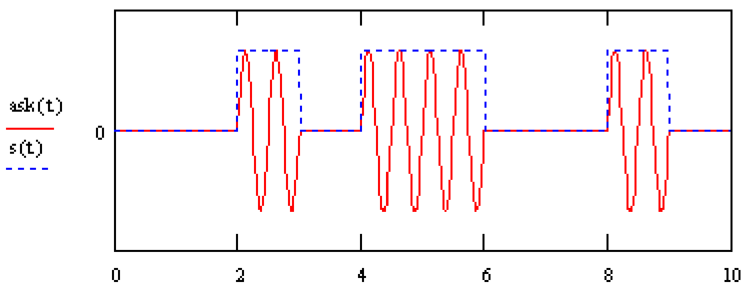


Figure 2 - Binary ASK (OOK) Carrier

In FSK, we change the frequency in response to information, one particular frequency for a 1 and another frequency for a 0 as shown below for the same bit sequence as above. In the example below, frequency f1 for bit 1 is higher than f2 used for the 0 bit.

$$FSK(t) = \begin{cases} \sin(2\pi f_1 t) & \text{for bit } 1 \\ \sin(2\pi f_2 t) & \text{for bit } 0 \end{cases}$$

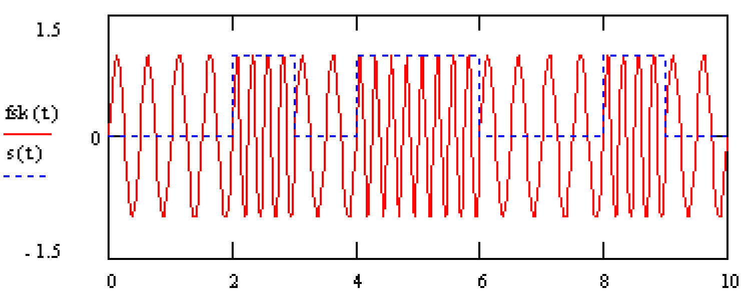


Figure 3 - Binary FSK Carrier

In PSK, we change the phase of the sinusoidal carrier to indicate information. Phase in this context is the starting angle at which the sinusoid starts. To transmit

0, we shift the phase of the sinusoid by 180°. Phase shift represents the change in the state of the information in this case.

$$PSK(t) = \begin{cases} \sin(2\pi f t) & \mathit{for\ bit}\ 1 \\ \sin(2\pi f t + \pi) & \mathit{for\ bit}\ 0 \end{cases}$$

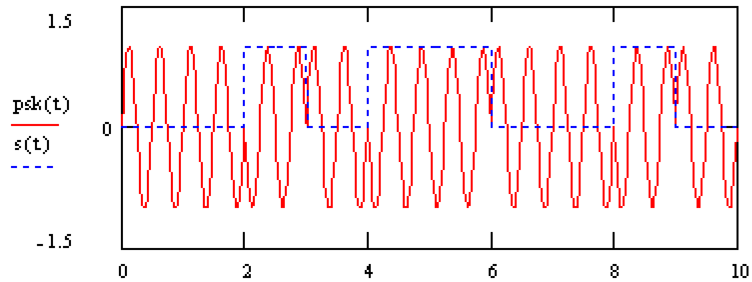


Figure 4 - Binary PSK Carrier (Note the 180° phase shifts at bit edges)

ASK techniques are most susceptible to the effects of non-linear devices which compress and distort signal amplitude. To avoid such distortion, the system must be operated in the linear range, away from the point of maximum power where most of the non-linear behavior occurs. Despite this problem in high frequency carrier systems, Amplitude Shift Keying is often used in wire-based radio signaling both with or without a carrier.

ASK is also combined with PSK to create hybrid systems such as Quadrature Amplitude Modulation (QAM) where both the amplitude and the phase are changed at the same time.

What is digital, what is analog?

There are three parts to a communications system.

1. The information, also called the baseband
2. The medium
3. The carrier

Information can be defined in two forms, digital or analog. Analog signal is considered continuous. Its signal amplitude can take on any number of values between the signal maximum and minimum. Voice is analog and can take any number of volume levels between its "dynamic-range" which is the range of volumes your vocal cords can produce. Digital devices convert analog voice to a digital signal by process of **sampling and quantization**. The analog signal is first sampled and then quantized in levels and then each level is converted to a binary number. For example, we may quantize your voice in 16 levels. Each of these levels can be represented by four bits.

Perhaps you remember when your telephone system went to the "tone" dialing. It went from being a pure analog system to a digital system based on sampling and quantization. Other examples of analog information are music and voice transmitted via FM and AM radio transmissions. Nearly everything else nowadays is digital.

The medium is *thing* the signal travels through. It can be air, space or wires of all sorts. Each of these mediums offers its own unique set of advantages and distortions that determine what is used as a carrier. A short wire in a chip for example may not need a carrier at all. A signal through space such as for satellite transmission may need a very high frequency carrier that can overcome space loss and other losses.

If medium is the road taken, then carrier is the truck that carries the information hence we call it *Carrier*. Depending on the medium, it can be light as in optical communications or a microwave signal as for mobile communications. An electromagnetic carrier can be of any frequency depending on the medium and the communication needs. Most mediums dictate what type of carrier (its frequency, amplitude) can propagate through it and the type of distortions it will suffer while traveling through it.

Anything that is wireless is analog – always. Wired signals can be digital or analog. Communications inside a computer are examples of pure digital communications, digital data over digital medium. LAN communications are digital data over analog medium. The radios are examples of analog data over analog medium.

In general when we talk about a digital system, we are usually talking about digital information over an analog medium. However, there are exceptions. Pulse Coded modulation (PCM) is a form of modulation where there is no carrier, so that makes it a pure digital system.

The "Shift Keying" the second two terms in the name of these modulations imply that they are digital modulations, i.e. the information is digital.

Modulation is the process by which we map information on to a carrier. How many ways can you fit your bags in car? Can you repack your stuff to make it fit better? These are the conceptual questions we tackle when we talk about modulation.

Signal Spaces and basis functions

Study of signal spaces provides us with a geometric method of conceptualizing the modulation process. In a physical space when we describe a vector by its coordinates (x, y); the vector is being described by a linear combination of two functions (1, 0) and (0, 1). Any vector can be written as a linear combination of these two functions which are called **basis functions** and are orthogonal to each other.

Another example of such a family of functions is the unit width pulses separated in time shown below. Each of these is independent of others and clearly we can use these functions to create any random data sequence consisting of square pulses. Each one of these single pulses is a basis function. However, this is not a very efficient set of basis functions as it takes a large number of these functions to create a random signal.

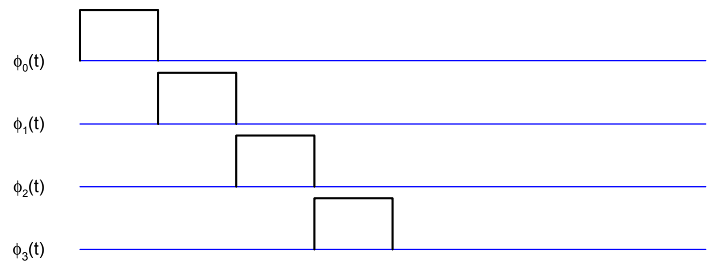


Figure 5 - Ortho-normal basis set

Ideally we want as few basis functions as possible which when combined can create a large number of independent signals, both digital and analog. In addition, basis functions should

- Have unit energy, such as the (1, 0) and the (0, 1) vectors and the above unit pulses.
- They should be orthogonal to every other function in the set, represented mathematically by

$$\int_{-\infty}^{+\infty} \phi_i(t)\phi_j(t) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Another important example is the pair of sine and cosine functions of unit amplitude. This special **basis set** is used as carriers in all real communications systems.

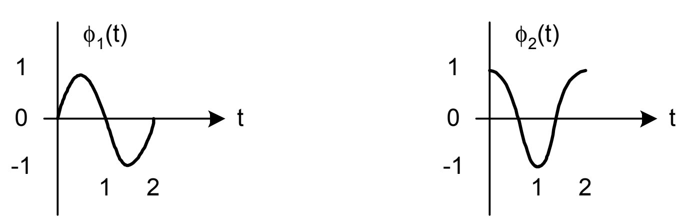


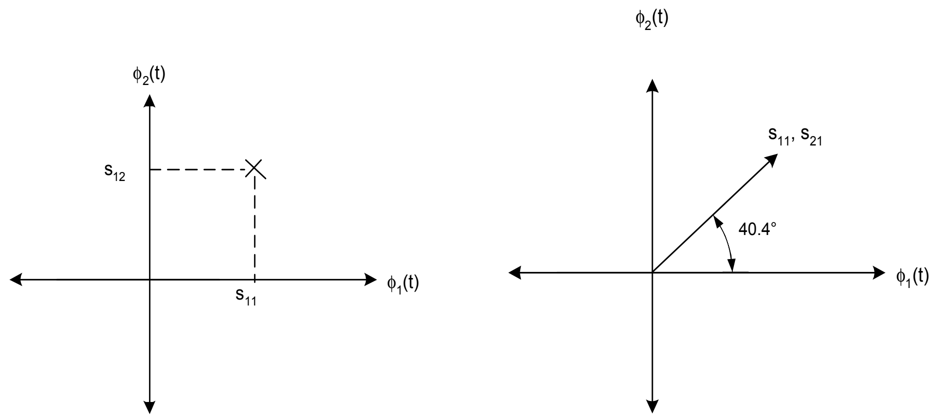
Figure 6 - Sine and cosine, two orthogonal functions are the basis set for all modern communications

The concept of I and Q Channels

We can write real signals as a vector sum of two signals in quadrature called I and Q. You can think of I and Q as the x, y axis projection of a signal.

$S = [I \;\; Q]$

Below you see two views of a signal space. One shows a signal in I-Q space and the other in its polar form.



(a) I and Q projections **(b) Polar form**

Figure 7 - Signal vector plotted on signal space

In Figure 7(a) the x and y-axis are I and Q projections of the signal. Quantity s_{11} is I channel projection and s_{12} is the Q projection of the signal. Figure 7 (b) shows the same signal in polar form, with its length equal to its amplitude and the angle is equal to its phase.

The coefficients s_{11} represent the amplitude of I signal and s_{12} the amplitude of the Q signal. These amplitudes when plotted on the x and y axis respectively, give the signal vector. The angle the signal vector makes with the x-axis is the phase of this signal.

$$\text{Magnitude of signal S} = \sqrt{I^2 + Q^2}$$

$$\text{Phase of the signal} = \tan^{-1}\frac{I}{Q}$$

Constellation diagrams

In a spectrum analyzer, we obtain the constellation diagram by sampling both I and Q channels at the same instant and then plotting I value against Q value on a x-y diagram. The time axis can be imagined as coming out of the analyzer so new points lie on top of the old ones at the transmitter. But when the constellation diagram is created for a signal after it has gone through the medium, it shows some offset from the transmitted points because of noise and other effects.

The tip of the signal vector is called the constellation point. The length of this vector or the distance from the origin is the energy of that signal.

$$E_{s_i} = [S_i]^2 = s_{i1}^{\ 2} + s_{i2}^{\ 2}$$

From above equation we can compute the energy of each signal shown below by adding the squared I and Q coefficients. The points that lie further from the origin have higher energy than those that lies close in. In a hybrid modulation with the following constellation pattern, the outlying signal sig2, has higher energy (i.e. higher amplitude, higher E_s) than inner signal, sig1. Another way to think is to realize that we are really plotting the amplitude. So a longer vector has a larger amplitude and hence higher energy.

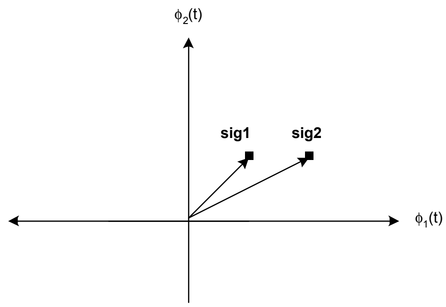


Figure 8 - The energy of a signal is proportional to its length

The distance between two signals (distance between the tips of the vectors) can be written as the dot product between the two signals.

$$E_{s_1 - s_2} = s_1 \cdot s_2 = |S_1| \times |S_2| \times \cos\theta$$

The distance between the tips is the difference in the energy of two signals. The angle between the two vectors is the correlation of the two signals. So if two signals are 90° to each other than, then cos θ is 0 and the two signals are un-correlated or are said to be orthogonal. In this example, the angle is not 90°, so a correlation exists between these two signals and this makes it harder to resolve the differences between the signals when received.

All carrier signals other than BPSK and QPSK have some level of correlation between their signals. This is the reason why BPSK and QPSK form the limit of the lowest bit error rate possible.

Symbols, bits and bauds

A symbol is quite apart from a bit in concept although both can be represented by sinusoidal or wave functions. Where bit is the unit of information, the symbol is a unit of transmission energy. It is the representation of the bit that the medium transmits to convey the information. Imagine bits as widgets, and symbols as boxes in which the widgets travel on a truck. We can have one widget per box or we can have more. Packing of widgets (bits) per box (symbols) is what modulation is all about.

In communications, the analog signal shape, by pre-agreed convention, stands for a certain number of bits and is called a symbol.

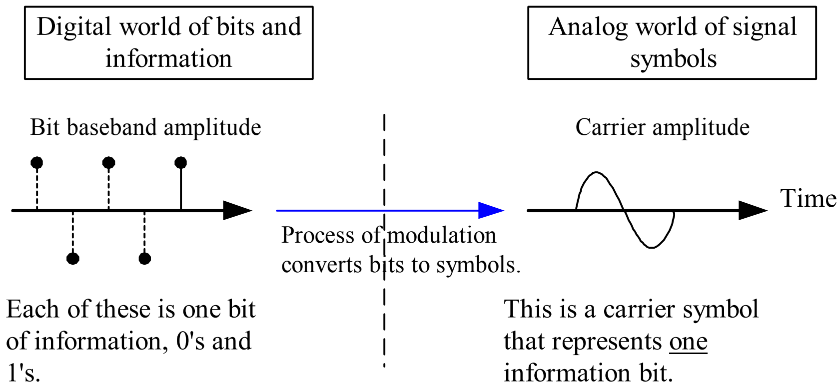


Figure 9 – Digital information travels on analog carrier

A symbol is just a symbol. It can stand for any number of bits, not just one bit. The bits that it stands for are not being transmitted, what is transmitted is the symbol or actually the little signal packet shown above. The frequency of this packet is usually quite high. The 1 Hz signal shown above is just an abstraction.

A baud is same as the symbol rate of a communication system.

PSK modulations

BPSK

Let's imagine a ship lost at sea with no communication system. It sees an airplanes flying overhead and wants to communicate its plight to the airplane while it is overhead. The captain marks two spots on each side of the mast as shown below.

Now he holds a bright light and runs back and forth between the marked spots to signal a message. Spot to the right means a 1 and spot to the left means a 0. We assume that all airplanes seeing this know that what each light stands.

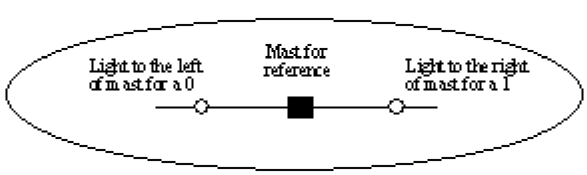


Figure 10 – Two signaling spots, a simple modulation system

This is a one dimensional signal, because the captain uses only one dimension (running from left to right) to indicate a symbol change.

The shining of the light is a symbol. Simplest thing is to have the symbol stand for just one bit and we call this BPSK modulation. We utilize just one sinusoid as the basis function. We vary the phase of this signal to transmit information which is identical in concept to the example of shining the light from the deck. Each symbol is signaled by a change in phase or position of the light as in this example. In BPSK we define two little packets of the cosine wave, one with zero phase and second one with a 180 degree different phase.

Table 1 lists the two symbols and the signals used to represent them. (The carrier signal shown is for f = 1.) The I and Q amplitudes are the x and y projections computed as follows.

$$I_{amplitude} = (symbol\ \exp ression)\ x\ \cos(phase)$$

$$Q_{amplitude} = (symbol\ \exp ression)\ x\ \sin(phase)$$

From this we get, I = 1 for the first symbol and -1 for the second symbol. Q amplitude is zero for both symbols because sin of both 0° and 180° is zero.

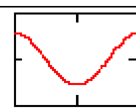
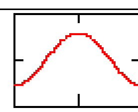
Symbol	Bit	Expression	Carrier Signal	I	Q
S1	0	$\sqrt{\frac{2E_s}{T}}\cos(\omega t)$		1	0
S2	1	$\sqrt{\frac{2E_s}{T}}\cos(\omega t + \pi)$		-1	0

Table 1 – Mapping rules for BPSK which uses two signals, both a variation of a cosine signal

Look at the carrier signals in the above table. This is what is transmitted in response to the bits not the bits themselves.

What are those funny coefficients in front of the expressions above?

Recall from Tutorial 1 that energy of a signal is equal to

$$E_s = \frac{A^2 T}{2}$$

So instead of writing a amplitude term to make the expression general, we write the equation in terms of energy, where A = $\sqrt{2E_s / T}$. When referring to carrier signals, we typically talk in terms of signal or bit energy, so it makes sense to write the equations in terms of energy, which is what this scaling factor is. Now we can scale the carrier signal for the power with which it is transmitted. Later when we talk about bit energy, and E_b/N_0 this will become clear.

Creating a BPSK carrier

How would we send a bit sequence 0111 0101 0010 1011 using BPSK signaling technique? To transmit this sequence, we need 16 symbols since each BPSK symbol stands for one bit. These are

s1 s2 s2 s2   s1 s2 s1 s2   s1 s1 s2 s1   s2 s1 s2 s2

Now string together the appropriate symbol signal packets from Table 1 in the right order. Figure below is the modulated carrier that would be transmitted for this sequence if we use the mapping in Table I.

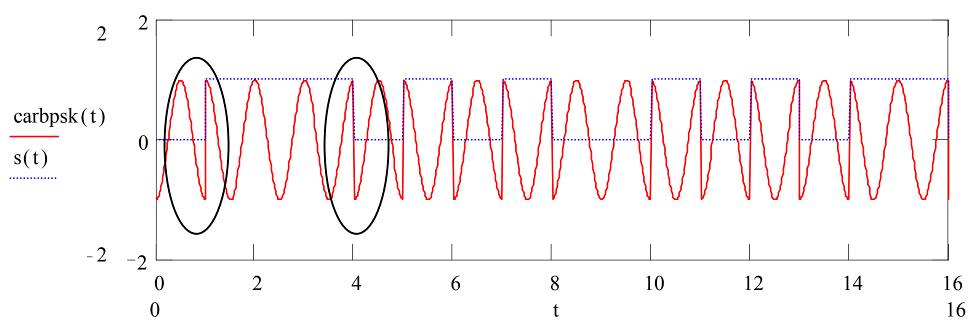


Figure 11 - A BPSK signal for bit sequence 0111 0101 0010 1011

If you could catch the modulated carrier and look at it on a network analyzer, you would see the above. However, the above picture is at a carrier frequency of 1 Hz, which is not realistic. In real systems, the carrier frequency is very high and we would see a signal that covers a lot of cycles between each transition.

What is a transition? A transition is the time at which we switch from one symbol to the next. What happens at the transition boundary is different for various modulations and is quite an important thing. In the case of BPSK, at every bit transition the signal does a 180 degree phase shift.

We worry about what the signal does at transitions because of amplifier non-linearities. Amplifiers used in communications have a very hard time with sudden changes in signal amplitudes and introduce distortions. Since this makes it harder to decode the symbol, we try to control these transitions.

Doing modulation in hardware

Notice that in the above description of the modulated carrier, there are no I or Q channels. There is just one signal. Well here is an important bit of information. We don't actually transmit I and Q. We transmit just one modulated signal over the airwaves but this signal is the composite of I and Q channels.

So and why do we need I and Q channels? Not only I and Q representations help us to understand and create consistent modulation schemes but they are also used to simplify the hardware design.

Now let's see how in hardware we would create the composite modulated signal. The figure below shows a bit stream that we would like to send using BPSK modulation.

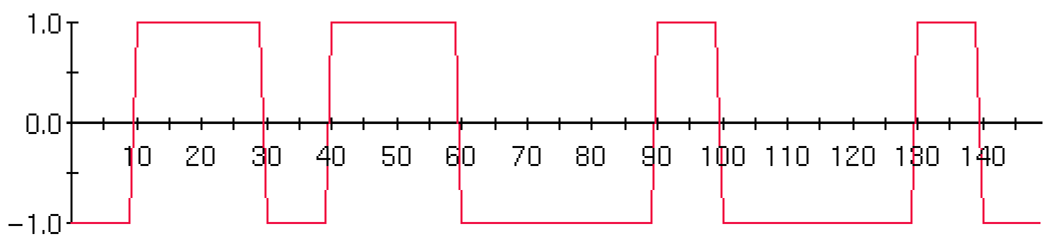


Figure 12 – A random bit stream, 01101100010001

Now the modulator takes a look at the above bit stream and makes a decision about which symbol to send. Below are first few symbols from this bit stream. The Q channel for BPSK is zero so just ignore it. (BPSK is one dimensional scheme.)The modulator assigns a certain symbol (and there are just two in case of BPSK) to the I channel depending on the bit to be sent.

Bit	Symbol	I	Q
0	S1	-1	0
1	S2	1	0
1	S2	1	0
0	S1	-1	0
1	S2	1	0
1	S2	1	0
0	S1	-1	0

Table 2 – BPSK Symbol to I channel mapping for the bit stream in Figure 12

Now in the next step, a carrier of free-running frequency f_c, is used to create the packet of signal. All we have to do to create the correct modulated signal is to multiply the carrier with the I channel amplitude values. Here is how the mapping would take place.

$$\begin{pmatrix}0\\1\\1\\0\\1\\1\\0\end{pmatrix}\rightarrow\begin{pmatrix}-1\\1\\1\\-1\\1\\1\\-1\end{pmatrix}\rightarrow\begin{pmatrix}-1\\1\\1\\-1\\1\\1\\-1\end{pmatrix}\times\cos(\omega t)=\begin{pmatrix}s1\\s2\\s2\\s1\\s2\\s2\\s1\end{pmatrix}$$

The first column contains the bits. These are mapped to amplitudes of the I channels by the BPSK mapping rules in Table 1. These are then multiplied by a cosine wave which is held for symbol time T. This results in a packet of analog signal called the symbol. The frequency ω of the cosine wave is called the **carrier frequency**.

QPSK

Now imagine a different ship. Its captain thinks up a different signaling arrangement. Here the he has marked out four spots on the deck, to the East and West and North and South. He assigns four different combinations to each of the spots as shown below. He can send two bits, with each flash of the light. If he can do it in the same time period as the first ship, then this man would be able to communicate twice as fast.

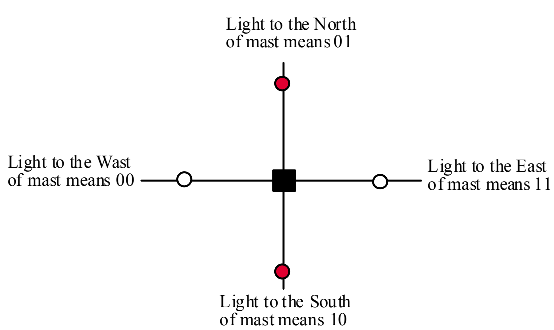


Figure 13 – Two dimensional signaling system

By creating four signaling spots, he has added another dimension. This gives two basis functions, the East-West and the North-South movements. Now there are four different symbol positions possible and we can assign 2 bits to each unique symbol.

QPSK uses two basis functions, a sine and a cosine whereas BPSK uses just one. By varying the phase of each of these carriers (in the ship example, the position) we can send two bits per each signal. The dimensionality of a modulation is defined by the number of basis functions used. That makes QPSK a two-dimensional signal. Not because it sends two bits per symbol, but because it uses **two** independent signals (a sine and a cosine) to create the symbols. All PSK modulations we will discuss here are two-dimensional.

Here are the four symbol mapping definitions for QPSK. Each packet is defined in terms of a sine or a cosine but with a different phase. (Note that phase is the angle at which the signal starts)

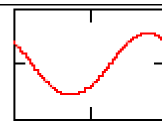
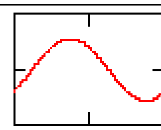
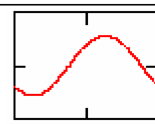
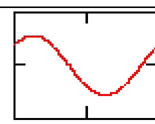
Symbol	**Bits**	**Expression**	**Phase, (Deg.)**	**Carrier Signal**	**I**	**Q**
S1	00	$\sqrt{\frac{2E_s}{T}}\cos(\omega t+\pi/4)$	45		$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$
S2	01	$\sqrt{\frac{2E_s}{T}}\sin(\omega t+3\pi/4)$	135		$-\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$
S3	11	$\sqrt{\frac{2E_s}{T}}\cos(\omega t+3\pi/4)$	225		$-\frac{1}{\sqrt{2}}$	$-\frac{1}{\sqrt{2}}$
S4	10	$\sqrt{\frac{2E_s}{T}}\sin(\omega t+\pi/4)$	310		$\frac{1}{\sqrt{2}}$	$-\frac{1}{\sqrt{2}}$

Table 3 – Mapping rules for QPSK

In QPSK the four symbol definitions, written in sine or a cosine, can be decomposed further so we can compute the I and Q channel amplitudes. Let's take for example the symbol S1.

Using this trigonometric identity,

$$\cos(x+y)=\cos x\cos y-\sin x\sin y$$

We can write this equivalent expression

$$\sqrt{\frac{2E_s}{T}}\cos(\omega t+\pi/4)=\sqrt{\frac{2E_s}{T}}\left(\cos\omega t\cos\pi/4-\sin\omega t\sin\pi/4\right)$$
$$=\sqrt{\frac{E_s}{T}}(\cos\omega t-\sin\omega t)$$

We see that the little packet of carrier signal representing a particular symbol can be created by a free running sine and a cosine wave of certain amplitudes. This makes hardware realization possible.

Although we created the BPSK modulated signal by stringing together the appropriate packets of signals, in real systems, we can not create a modulated carrier this way. What we have at our disposal are oscillators that produce sines and cosines. We can not just use a certain part of the signal as if they are sitting on a shelf for us to grab a piece. We need a way to create a signal packet of a particular phase when needed out of a free-running sine or cosine. This is where **Quadrature Modulation** and I and Q channels come into play. I and Q channels are not just concepts but also how modulators are designed.

Now take a moment to talk about the constellation of QPSK.

Constellation of QPSK

A constellation is a plot of the I channel *amplitude* against the Q channel *amplitude* when sampled at the symbol rate. If the symbol rate is .1 seconds, then, we would sample the time-domain signal every .1 second at the best possible moment (more about this later) and then plot the measured I and Q values. The figure below shows the QPSK constellation diagram. It is created by plotting the values of I and Q amplitudes shown in Table 3. Each point is a pair of (I, Q) values representing a modulated signal or symbol. These I and Q values are computed by multiplying the signal expression (its amplitude) by sine or cosine of the phase angle.

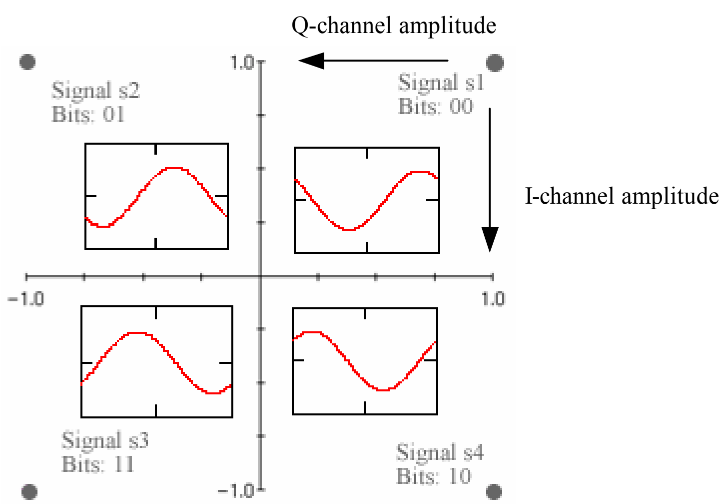


Figure 14 – Constellation points are the tips of the modulating signal

This constellation diagram is created just prior to combining both I and Q into a composite signal. Constellation diagrams are always at baseband, that is at carrier frequency equal to zero. Eye diagrams similarly are also at baseband and show the same information but in time-domain.

The constellation diagram of Figure 14 is a perfect square because the channel has not suffered any degradation due to the amplifier or the medium. When a constellation diagram is created at the receiver, the picture is not so perfect looking and can tell us the type of distortions experienced by the signal. The pulse shaping also changes this perfect square constellation.

Shaping the pulse to reduce bandwidth

The square pulses shown here are not practical to send. They are hard to create and require a lot of bandwidth. So in their lieu we send shaped pulses that convey the same information but use smaller bandwidths and have other good properties such as intersymbol interference rejection. One of the most common pulse shaping is called "root raised cosine". This pulse shaping has a parameter called the roll-off which controls the shape and the bandwidth of the signal.

Some common pulse shaping methods are

- Root Raised cosine (used with QPSK)
- Half-sinusoid (used with MSK)
- Gaussian (used with GMSK)
- Quadrature partial response (used with QPR)

In Figure 17 we see the time domain traces of a QPSK I and Q channel where the signal has been shaped by root raised cosine pulses. The general bit shape can be seen easily.

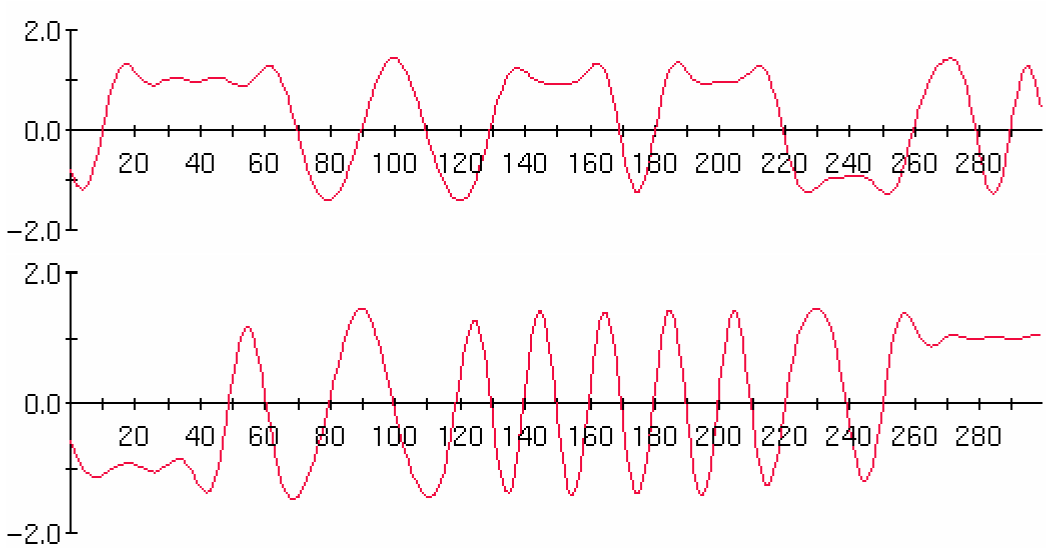


Figure 15 – Root-raised cosine shaped pulses

We can draw a constellation diagram of this signal by sampling this signal every 20 samples, and then plotting the measured I values against Q values. We get the following constellation diagram.

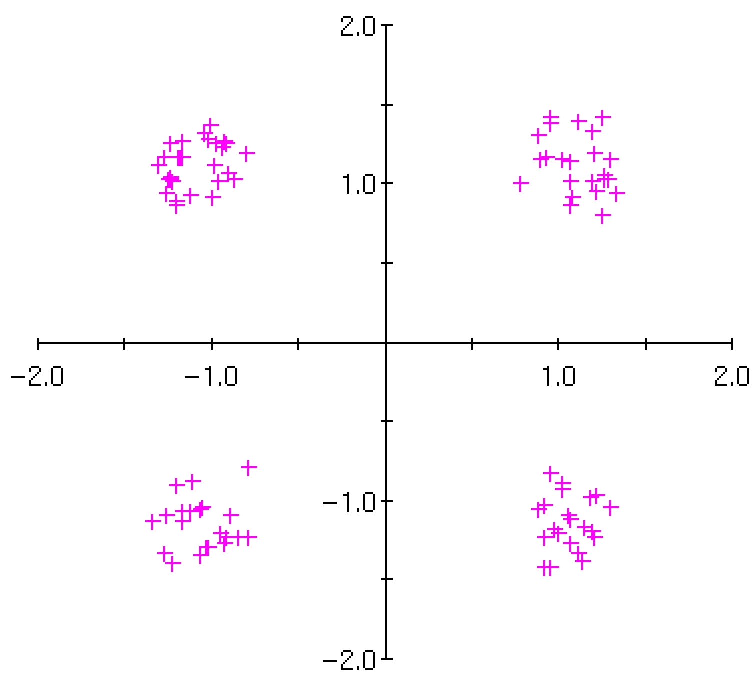


Figure 16 – Constellation diagram of a root-raised cosine shaped signal

Note that the RRC constellation diagram has a scatter around the ideal points. This is an inevitable consequence of any pulse shaping and relates to increased bit error rate for the signal.

A Gaussian pulse shape replaces the square pulses with a Gaussian pulse. Here is a time domain trace for a Gaussian pulse shaped signal.

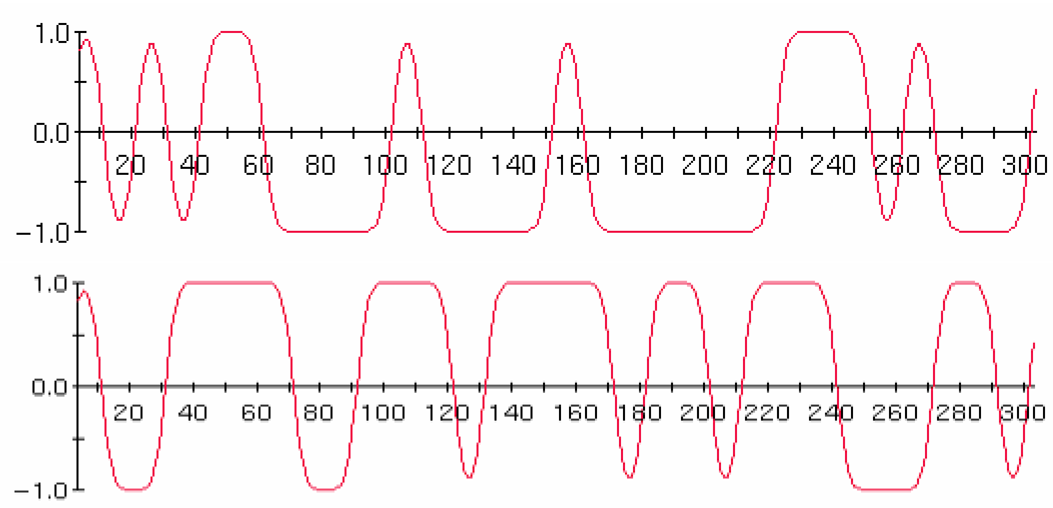


Figure 17 – The time domain signal shaped by Gaussian pulses.

QPSK step-by-step

Now let's follow the process of creating a QPSK modulated carrier step by step.

We wish to transmit the bit sequence: 11 00 11 10 00 10 00 01 00 … using QPSK.

Step 1 – Convert bit sequence to a polar signal (convert 0 to -1)

1 1 -1-1 1 1 1-1 -1-1 1-1 -1-1 -1 1 -1-1

Step 2 – Send alternating bits to I and Q channels

I 1 -1 1 1 -1 1 -1 -1 -1 -1 -1 1 -1 -1 1

Q 1 -1 1 -1 -1 -1 -1 1 -1 -1 -1 1 1 -1 -1

Set the amplitude of I and Q based on Table 3. (The following have been scaled to 1.0 instead of .707)

Step 2 – Normally we would shape the pulses by a RRC filter at this point. Skip this step for this example.

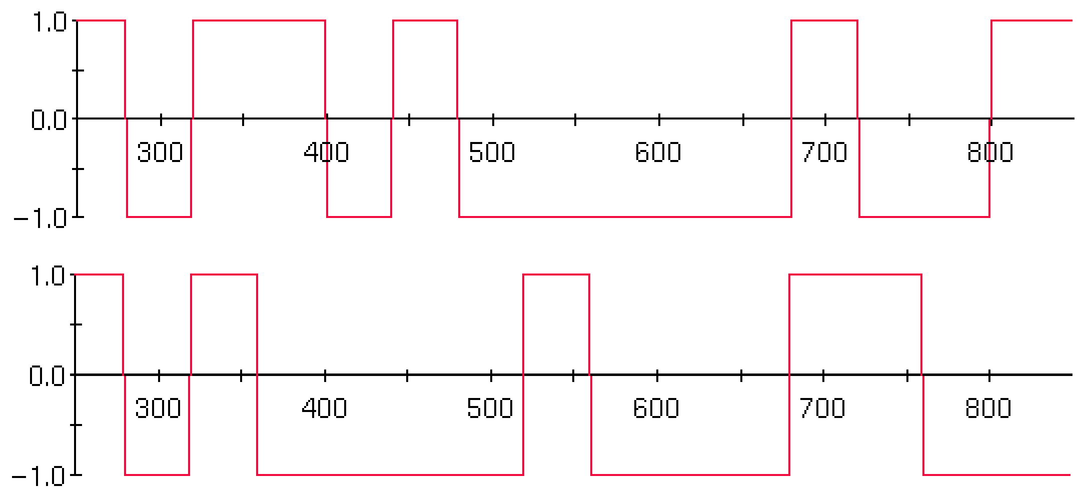


Figure 18 – The bit stream mapped to I and Q channel levels

Step 3 – Multiply I channel with a cosine of frequency f_c ($f_c = 1$ in this example) Note that the signal transitions at symbol bit boundary solely because of I channel amplitude changes.

Multiply Q channel with the same free running cosine but shift it first by 90° so it is a sine wave. This way we can use just one piece of hardware to get both a sine and a cosine signal. We get the following two modulated carrier signals.

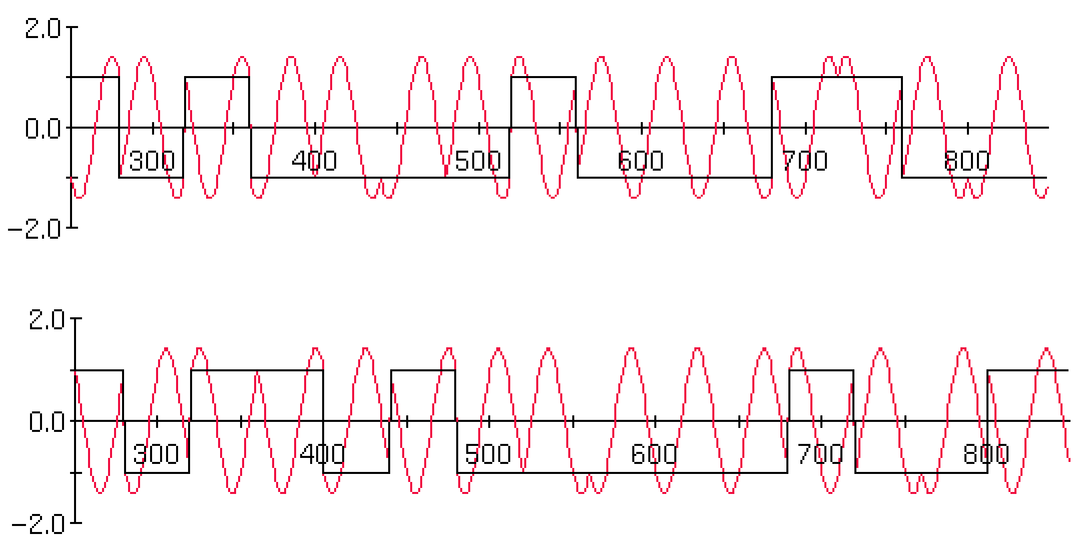


Figure 19 – The I channel multiplied by a cosine of frequency =1 and the Q channel multiplied by a sine of frequency =1

Step 4 – Add I and Q channels and we get the transmitted carrier.

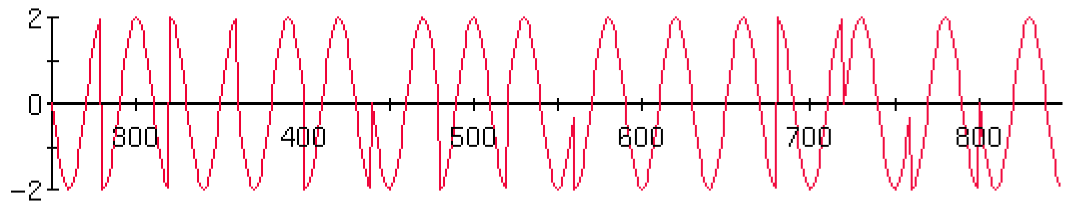


Figure 20 – Sum I and Q channels to create the composite transmitted signal

This is how we do it in hardware.

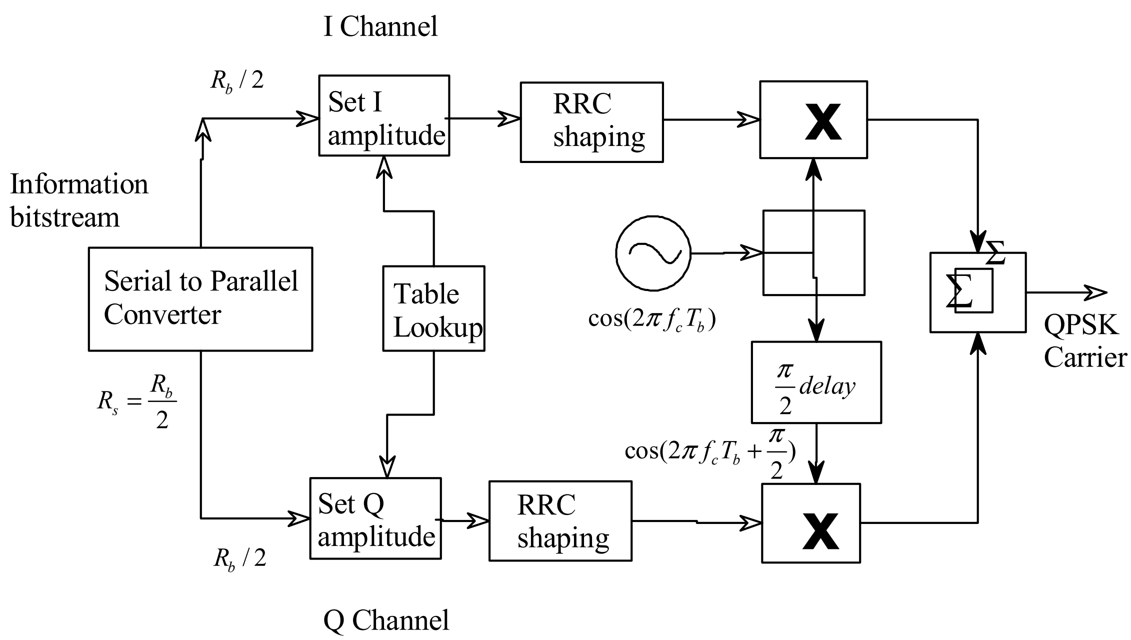


Figure 21 – Hardware implementation of the QPSK modulator

Summary: The Serial to parallel converter takes the bit stream coming in at a bit rate of Rb and splits it into two streams, each of half the bit rate. Depending on the dual bit pattern coming in, I and Q amplitudes are set from a table lookup function. Each of these is then individually modulated by a sine or a cosine wave of carrier frequency ω after being shaped into a root raised cosine pulse. These are added together to get the transmitted signal.

How QPSK differs from BPSK: QPSK has two basis signals. It can be considered to be composed of two BPSK signals each of one half the bit rate.

Constant Envelope modulation

QPSK is part of a class of signals called **constant-envelope** signals. There is no rigorous definition of a constant envelope signal. One definition is; when sampled at the symbol rate, the sampled value of the amplitude is constant. Another is that there are no discontinuous phase changes. Yet another is that the maximum and minimum amplitude attained by the signal over one period is constant. The sine wave is an ideal constant envelope signal.

Constant envelope signals suffer less distortion in high power amplifiers and are preferred for wireless applications. The reason is that amplifiers work by changing a signal's amplitude, either increasing or decreasing it. To increase a signal’s power is to increase its amplitude. A non-linear amplifier changes the signal amplitude by differing amount depending on the instantaneous amplitude of the signal. The more the amplitude of a signal varies, the more non-linear amplification occurs and this results in a distorted signal. QPSK is not technically a constant envelope because of its discontinuous phase shifts but is considered nearly so.

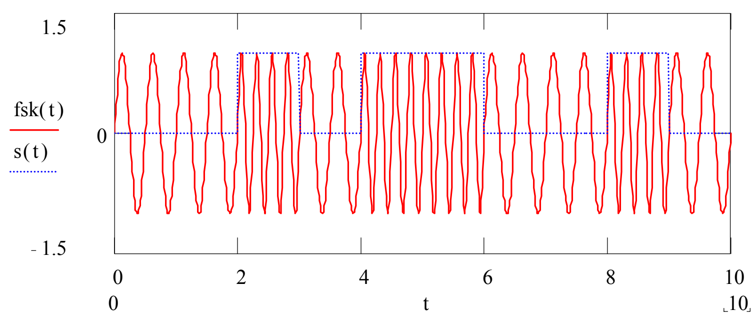


Figure 22 – FSK is definitely a constant envelope modulation.

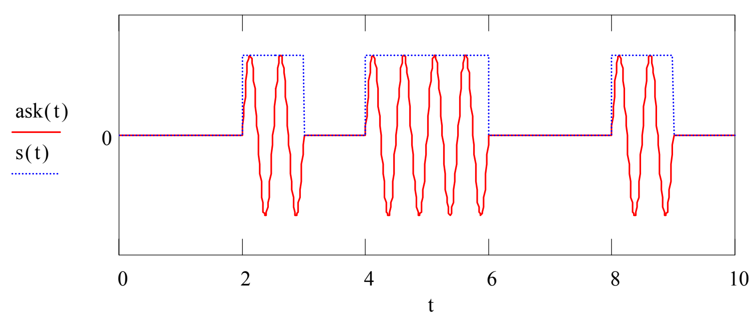


Figure 23 - ASK is definitely not a constant envelope modulation

Offset QPSK

Offset QPSK is a minor but important variation on QPSK. In Offset QPSK, the Q channel is shifted by half a symbol time so that I and Q channel signals do not transition at the same time. The result of this simple change is that phase shifts at any one time are limited and hence offset QPSK is more "constant-envelope" than straight QPSK. In high power amplifiers and for certain satellite applications, Offset QPSK offers better performance. Although in a linear channel its bit error rate is the same as QPSK, in non-linear applications, its BER is lower when operating close to the saturation point of the transmitting amplifier. Offset QPSK (OQPSK) is also called staggered QPSK (SQPSK).

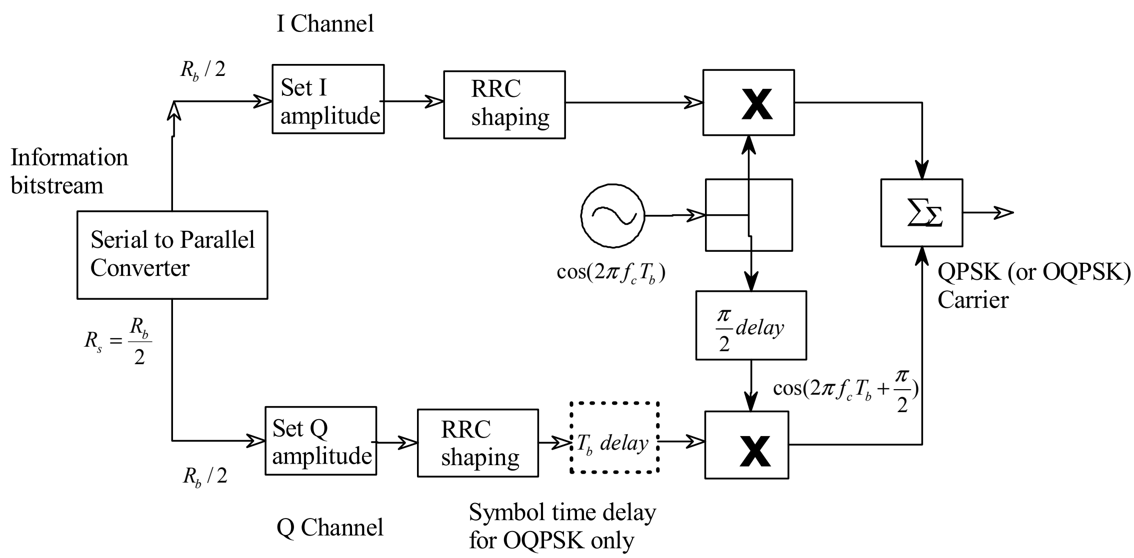


Figure 24 – QPSK modified to become OQPSK

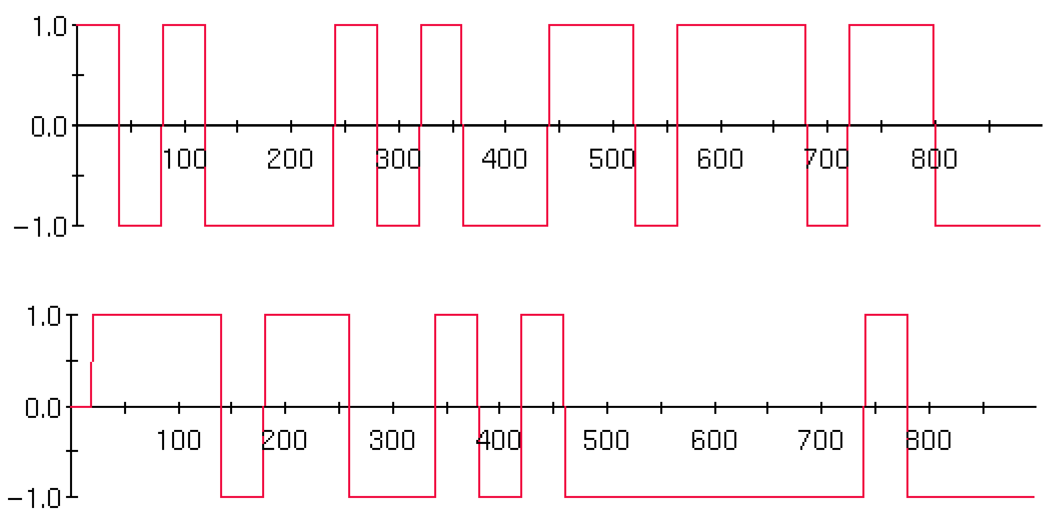


Figure 25 – I and Q channel mappings of an Offset QPSK signal, the symbol transitions do not occur at the same time.

Unlike QPSK, I and Q channels of an OQPSK signal do not transition at the same time. One consequence of this is that when we look at the constellation diagram of the OQPSK, the symbol transitions occur only to neighbors. This means that the transitions are never more than 90°. At any symbol change, for either I or Q channel, only one axis can change at a time, either I or the Q but not both. (At any transition, only I or the Q changes but not both.) In constellation-speak, if the signal was in the right upper quadrant, the next signal can only go to either the lower right quadrant or to upper left quadrant but not across. Note how this is different from QPSK, where all transitions can occur.

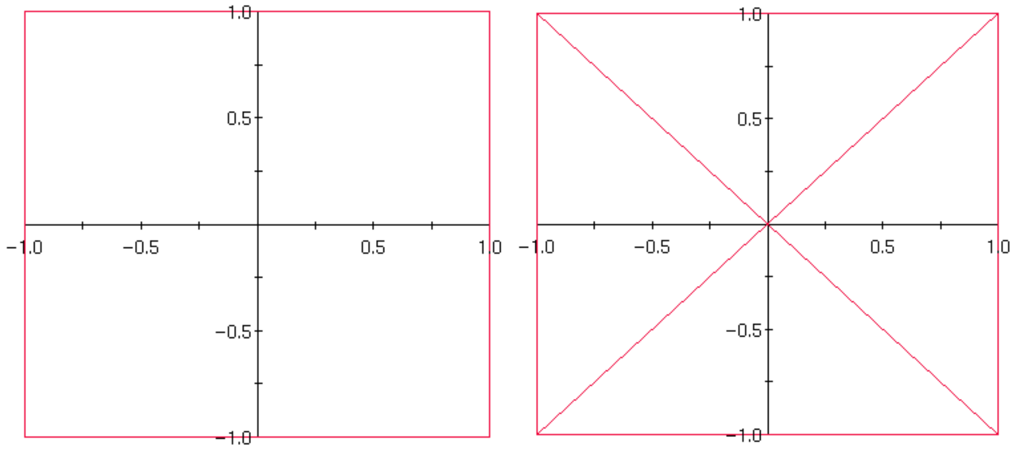
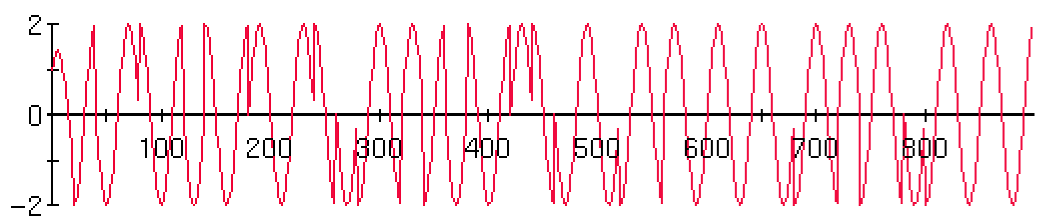
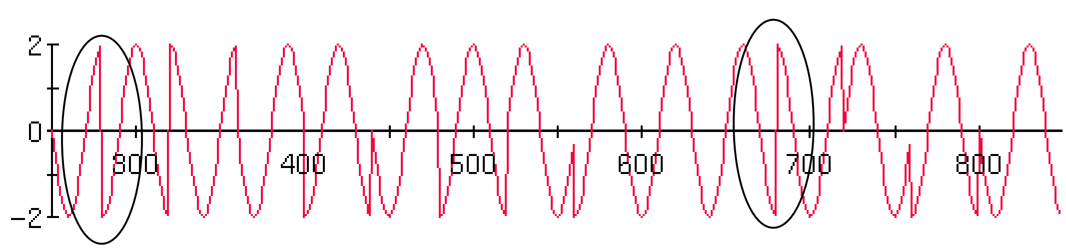


Figure 26 – OQPSK and QPSK constellation diagrams, (1) OQPSK, (2) QPSK



(a) OQPSK – All phase shifts are 90°.



(b) QPSK - Note the 180° phase shift.

Figure 27 – The phase jumps at the symbol transition for OQPSK are smaller. (Note that the figures above are not of the same scale in time.)

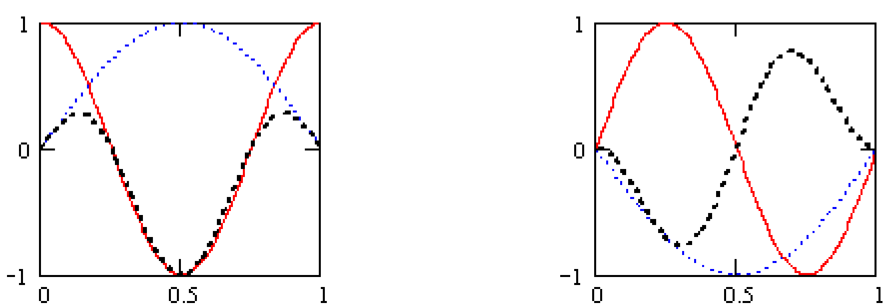
Figure 27 compares the OQPSK signal with a QPSK signal. Note that the OQPSK signal never transitions more than 90°. QPSK on the other hand goes through phase change of 180° for some transitions. The larger transitions are a source of trouble for amplifiers and to be avoided if possible. In satellite transmission, QPSK reigns supreme, it is easy to build and operate. Military often uses OQPSK because of its need to use low power radios and minimum adjacent channel interference issues.

How OQPSK differs from QPSK: The Q channel of OQPSK is delayed by a half a symbol time, staggering the two quadrature channels.

Minimum Shift Keying (MSK)

Although MSK is often classified as FM modulation, it is also related to offset-QPSK owing to the dual nature of FSK and PSK modulations. OQPSK is created from QPSK by delaying Q channel by half a symbol from I channel. This delay reduces the phase shifts the signal goes through at any one time and results in an amplifier-friendly signal.

MSK can be derived from OQPSK by making one further change - OQPSK I and Q channels use square root-raised cosine pulses. For MSK, change the pulse shape to a half-cycle sinusoid. Figure 28 shows a MSK pulse signal and then multiplication by the carrier. The red curve is the carrier signal, and the blue the MSK pulse shape and the black the multiplication of the pulse shape and the carrier giving the modulated carrier.



(a) MSK pulse and carrier for a 1 bit (b) MSK pulse and carrier for a 0 bit

Figure 28 – MSK pulse shaping is a half-sine wave shown in blue, positive for a 1 and negative for a 0.

The carrier signal expression for MSK is

$$c(t) = a(t)\sin\left(\frac{\pi}{2T}t\right)\cos\left(\frac{\pi}{T}t\right) + a(t)\sin\left(\frac{\pi}{2T}t\right)\sin\left(\frac{\pi}{T}t\right)$$

with the underlined portion, the half-sinusoid pulse shape. Figure 29 shows how MSK pulses look compared to QPSK square pulses.

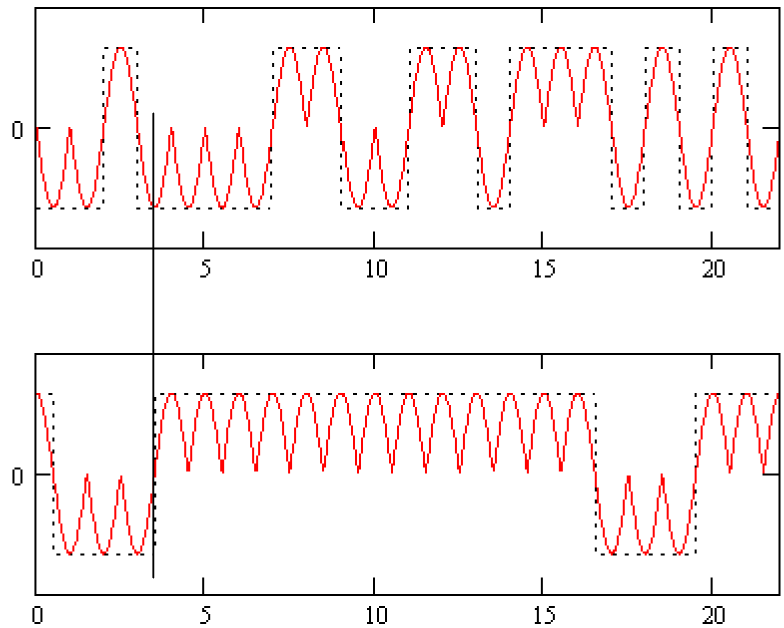


Figure 29 – MSK pulse, each pulse is a half cycle sine wave.

The dashed line is the QPSK I and Q channel symbols and the solid lines show how these have been shaped by the half sine wave. The I and Q channels are computed by

$$MSKI(t) = QPSKI(t)\left|\sin\left(\frac{\pi t}{2T}\right)\right|$$

$$MSKQ(t) = QPSKQ(t+.5T)\left|\sin\left(\frac{\pi(t+.5T)}{2T}\right)\right|$$

The I and Q channels are then multiplied by the carrier, cosine for the I channel and sine for the Q channel. Note that the period of pulse shape is twice that of the symbol rate.

$$MSKcarrI(t) = QPSKI(t)\left|\sin\left(\frac{\pi t}{2T}\right)\right|\cos\left(\frac{\pi t}{T}\right)$$

$$MSKcarrQ(t) = QPSKQ(t)\left|\sin\left(\frac{\pi t}{2T}\right)\right|\sin\left(\frac{\pi(t+.5T)}{T}\right)$$

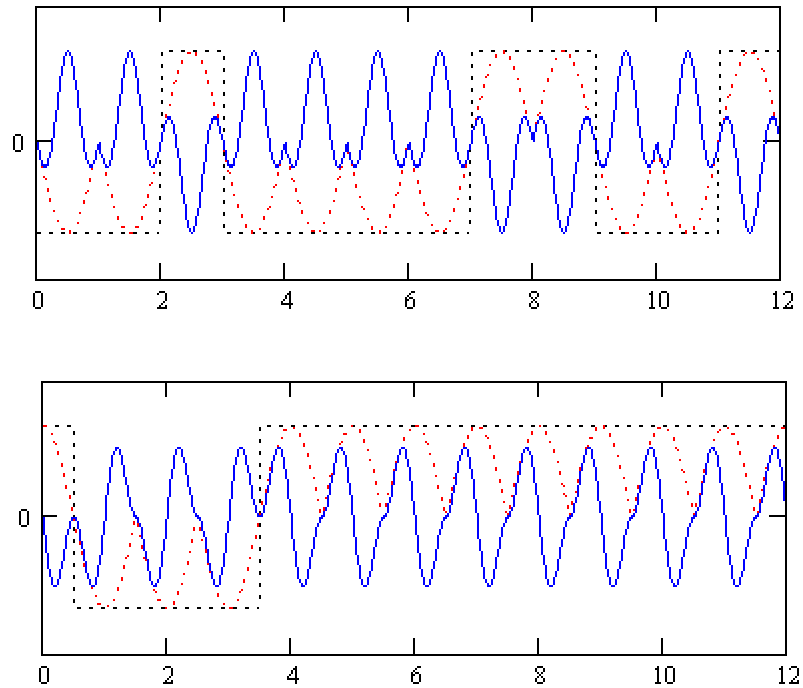


Figure 30 – MSK I and Q modulated carriers.

Adding the I and Q components gives the MSK carrier of Figure 31. Compare this carrier to a QPSK carrier. This one has much smoother phase shifts at the symbol boundaries. This results in lower side lobes which is an advantageous property for wireless signals since it results in less adjacent signal interference.

Minimum Shift Keying (MSK) is also called continuous phase (CP) Frequency Shift Keying (FSK). FSK is the digital version of analog Frequency Modulation (FM) and MSK is a form of FSK, where modulation index is equal to .5 which results in a minimum frequency separation such that the modulation frequencies are still orthogonal. (See FM tutorial)

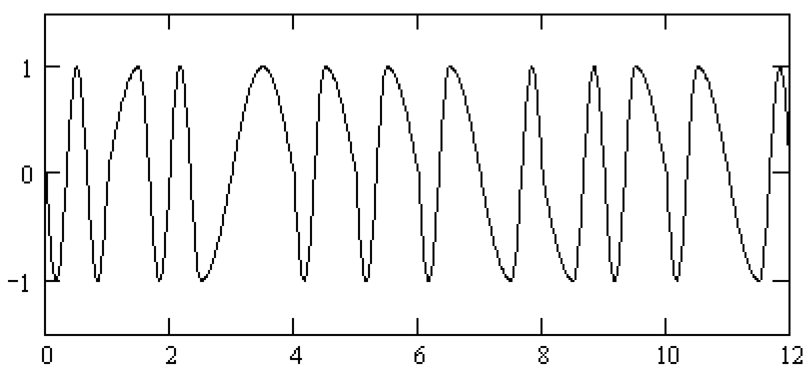


Figure 31 – MSK modulated carrier

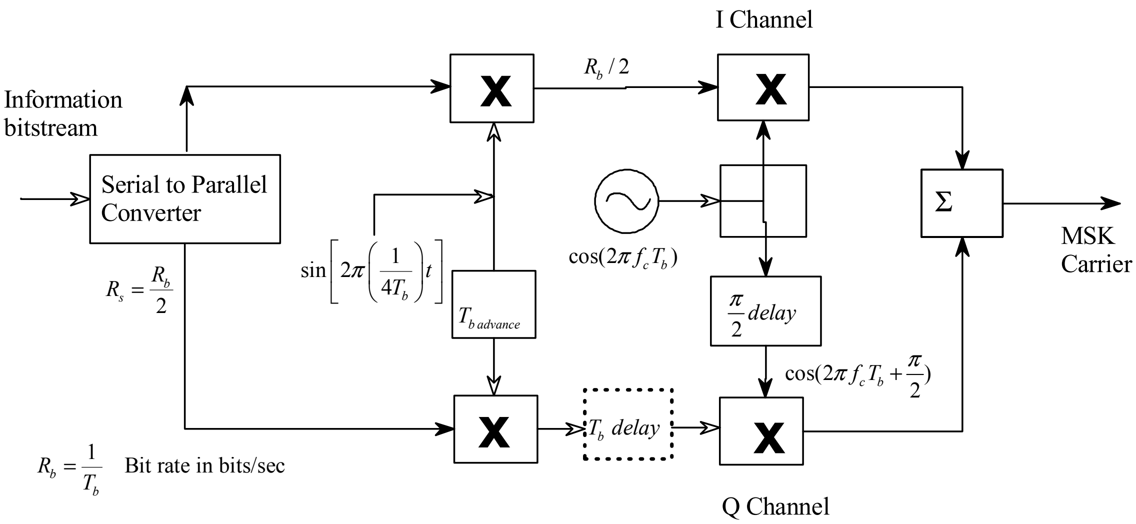


Figure 32 – MSK modulator block diagram

Figure 32 shows the modification made to the QPSK modulator to create the MSK signal. Only the pulse shaping has been changed. The half cycle time shift of the OQPSK stays.

How MSK differs from QPSK: MSK is generally considered a FSK modulation but it is exactly the same as OQPSK except that it uses a half-sinusoid for pulse shaping instead of root-raised cosine pulses.

Gaussian MSK (GMSK)

We created MSK by applying a half sinusoid to the square pulse. By using a Gaussian pulse shape, the result can be improved even further. The modulation obtained this way is called GMSK.

GMSK is used in several mobile systems around the world. Global Speciale Mobile (GSM), Digital European Cordless Telephone (DECT), Cellular Digital Packet Data (CDPD), DCS1800 (Digital Communications System in the 1800 MHz band) in Europe, and GSM-based PCS1900 (Personal Communications Services in the 1900 MHz band) in the U.S. uses GMSK.

Recall that the root-raised cosine pulse has a roll off factor, α. The roll-off factor determines how sharply the pulse rolls off to zero energy. A Gaussian pulse similarly has a BT factor that determines how sharply it rolls off. A BT of .3 is used commonly.

MSK and GMSK, both being related to FM modulation, can both be created two ways, 1. as a PSK signal and 2. as a FSK signal. Both are most commonly implemented as a FSK technique.

The Gaussian pulse shape used instead of the half-sinusoid or the root raised cosine is given by

$$g(t) = \frac{1}{2T}\left(Q\left(2\pi B_b \left(\frac{t-.5T}{\sqrt{\ln 2}}\right)\right) - Q\left(2\pi B_b \left(\frac{t+.5T}{\sqrt{\ln 2}}\right)\right)\right)$$

where

$$Q(t) = \int_t \frac{1}{\sqrt{2}} e^{-(x^2/2)} dx$$

Quantity B_b is signal bandwidth. BT factor is equal to this number times the symbol time, T.

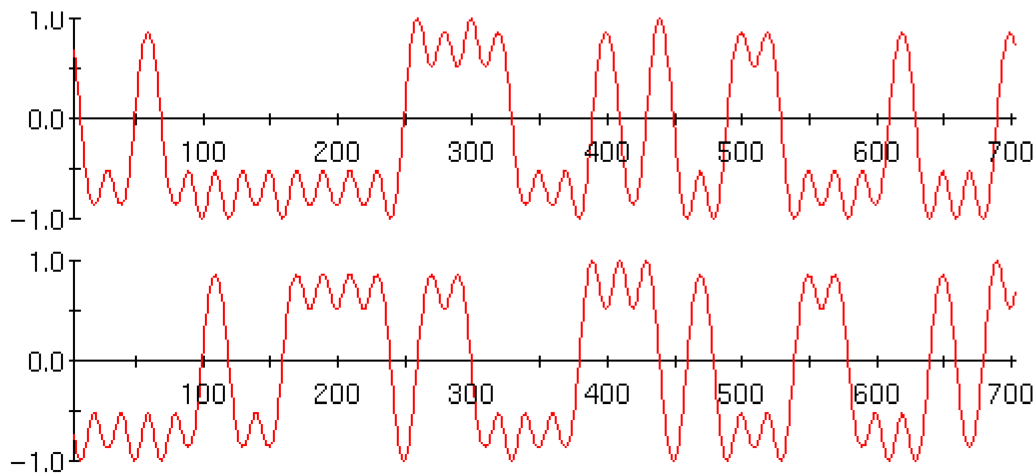


Figure 33 – Modulated I and Q GMSK carriers

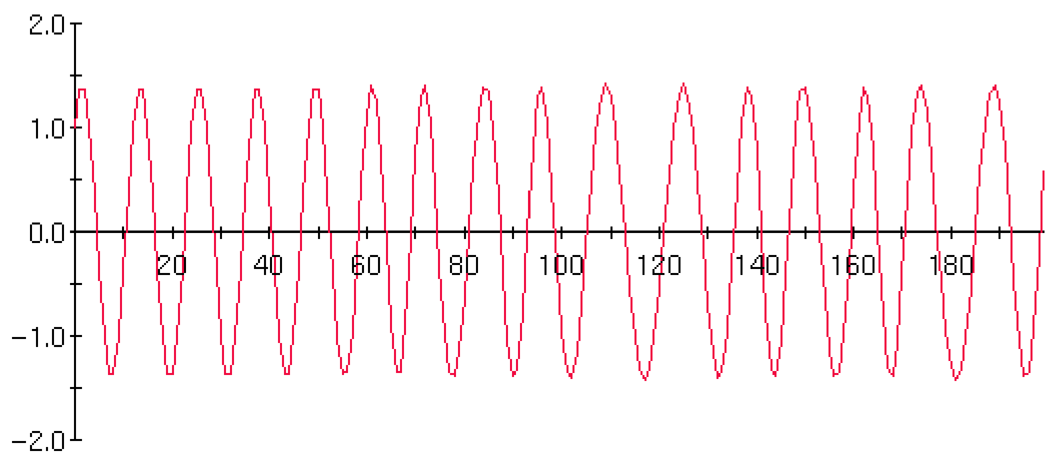


Figure 34 – Add I and Q GMSK carriers to obtain the composite carrier

The GMSK modulated carrier is even better at transitions than MSK and this is the main reason it is used as a standard in some cellular systems.

Figure 35 - GMSK block diagram

How GMSK differs from MSK: GMSK is nearly always implemented as a FM modulation. However conceptually it is same as MSK except instead of half-sinusoid as a pulse shape a Gaussian pulse shape is used instead.

8-PSK

Imagine once more the man on the ship, he figures he's got the space on the deck, so why not add more signaling positions. He marks out a circle and doubles the number of places where he will stand to flash up the signal. We can see the problem right away, how is the airplane going to make out where he is standing. But never mind, he goes ahead with his plan. Here is how his new signaling positions look.

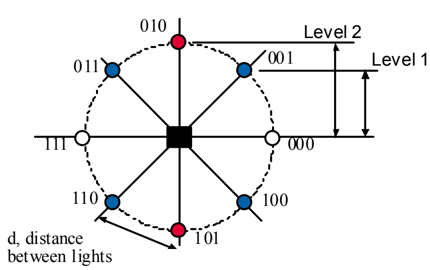
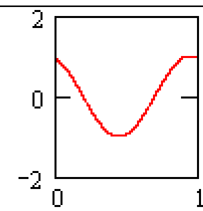
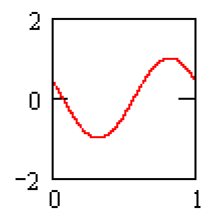
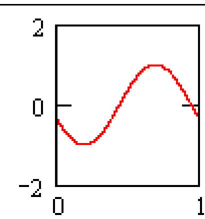
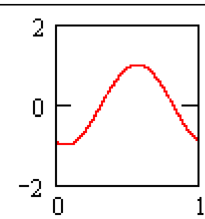


Figure 36 – 8-PSK uses eight different unique signals

He assigns bit values to each of the eight positions as shown. Note that each set of bits is just one bit different from its neighbor. So if the airplane does make an error in reading his position, most likely this will result in only one bit being misinterpreted. The eight positions are created with x and y distances or by phases of sines and cosines in communications. We have two basis functions again, a sine and a cosine and each configuration has a different phase to indicate a specific bit pattern. We use four different phase values, namely $\pi/8,\ 3\pi/8,\ 5\pi/8\ \textit{and}\ \ 7\pi/8$. Each of these phase shifts is 45 degrees apart. Each of these is applied to the sine and the cosine to give us a total of eight values.

(Assume $\sqrt{\frac{2E_s}{T}} = 2$)

Symbol	**Expression**	**Phase**	**I**	**Q**	**Signal**
S1 000	$s_1(t) = \sqrt{\frac{2E_s}{T}}\cos(\omega t + \pi/8)$	22.5°	1.414	.707	
S2 001	$s_2(t) = \sqrt{\frac{2E_s}{T}}\cos(\omega t + 3\pi/8)$	67.5°	.707	1.414	
S3 011	$s_3(t) = \sqrt{\frac{2E_s}{T}}\cos(\omega t + 5\pi/8)$	112.5°	-.707	1.414	
S4 010	$s_4(t) = \sqrt{\frac{2E_s}{T}}\cos(\omega t + 7\pi/8)$	157.5°	-1.414	.707	

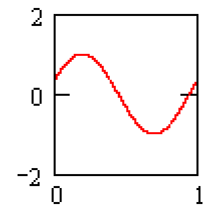
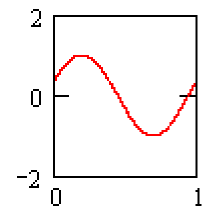
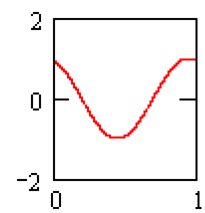
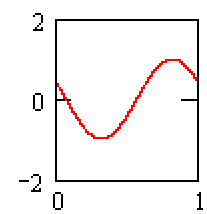
S5 110	$s_5(t) = \sqrt{\frac{2E_s}{T}} \sin(\omega t + \pi/8)$	202.5°	-1.414	-.707	
S6 111	$s_6(t) = \sqrt{\frac{2E_s}{T}} \sin(\omega t + 3\pi/8)$	247.5°	-.707	-1.414	
S7 101	$s_7(t) = \sqrt{\frac{2E_s}{T}} \sin(\omega t + 5\pi/8)$	292.5°	.707	-1.414	
S8 100	$s_8(t) = \sqrt{\frac{2E_s}{T}} \sin(\omega t + 7\pi/8)$	337.5°	1.414	-.707	

Table 4 – 8-PSK signals

8-PSK step-by-step

Step 1 - The bits stream to send is: 100 111 111 111 111 001 ….

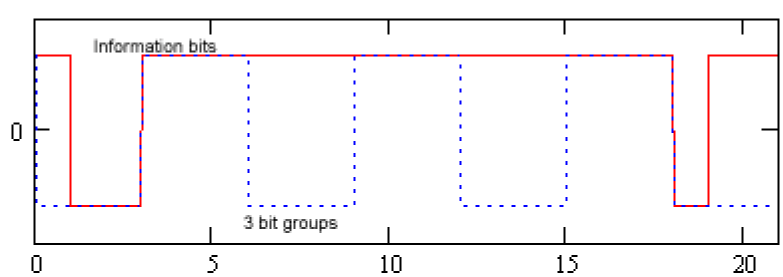


Figure 37 – Bit stream to be modulated using 8-PSK, 100 111 111 111 111 001

Let's name these bit packets for convenience.

Symbol sequence: s8 s6 s6 s6 s6 s2 ….

Map each symbol to I and Q using the amplitudes in Table 4.

	s8	s6	s6	s6	s6	s2
I	-0.707	0.707	0.707	0.707	0.707	1.414
Q	-1.414	-1.414	-1.414	-1.414	-1.414	0.707

Each of these three-bit packets are mapped to levels of I and Q channel from the table above.

After multiplication with cosine and sine wave of the carrier frequency we get the signals shown in Figure 39. When I and Q are added together, you get the composite signal in Figure 40.

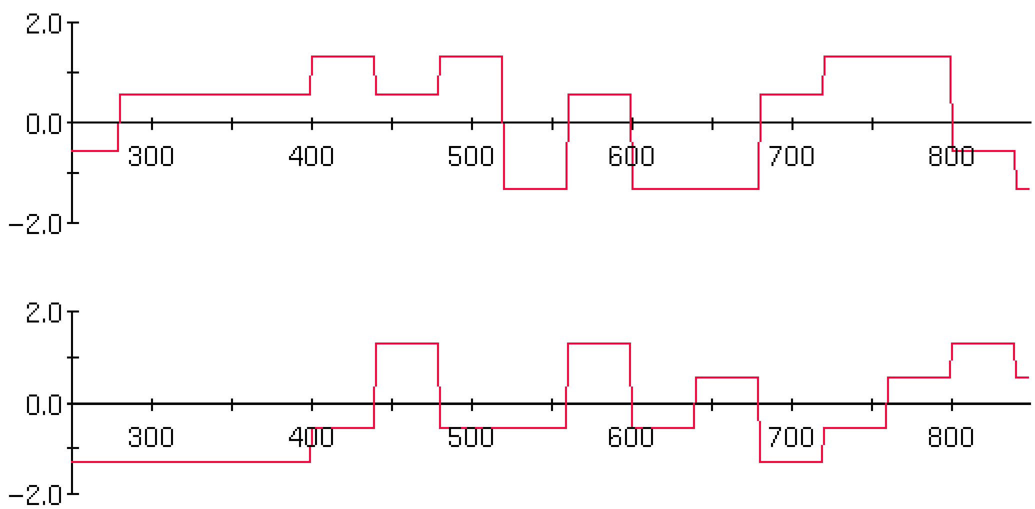
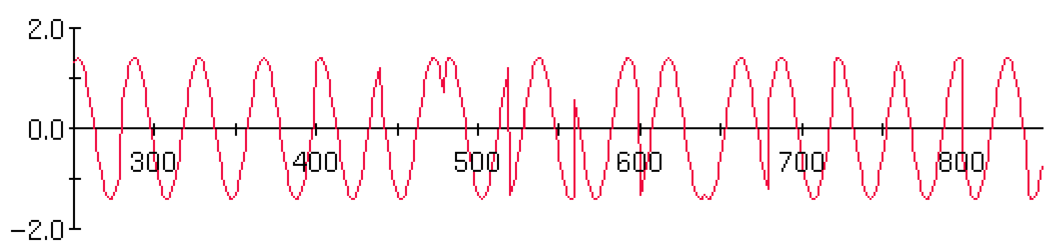


Figure 38 – 8-PSK mapping of I and Q for sequence in Figure 36. Note that both I and Q can take on 4 different values.



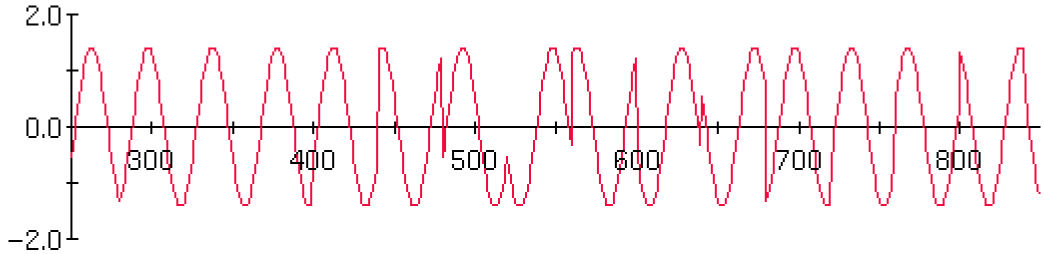


Figure 39 –I and Q of sequence in Figure 38 multiplied by the carrier of frequency 1 Hz.

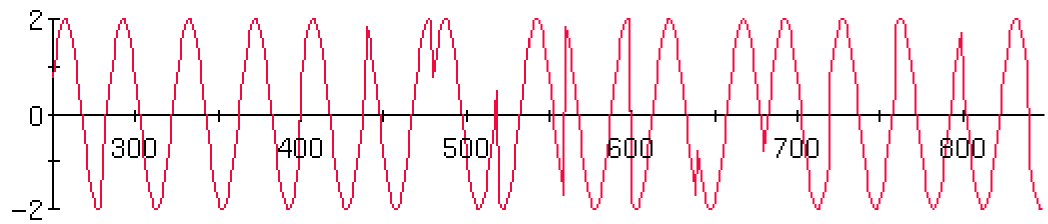


Figure 40 –Adding I and Q to obtain the composite carrier

8-PSK transmitted signal shows smaller phase transitions (on the average) then QPSK which is a good thing but since the signals are also less distinctly different from each other, makes 8-PSK prone to higher bit errors. Why then would we want to use 8-PSK? Because, we can pack more bits per symbol, with each symbol transmitted, we can convey three bits. The throughput of 8-PSK is 50% better than QPSK which can transmit just 2 bits per symbol as compared to 3 for 8-PSK. 8-PSK is the first of the bandwidth-efficient modulations.

How 8-PSK differs from QPSK: 8-PSK uses 8 different symbols (signals) to convey three bits per symbol. The I and Q channels have 4 levels each. Its implementation is identical to QPSK.

π/4-QPSK – a variation on both QPSK and 8-PSK

This a variation of QPSK that mimics 8-PSK. Like QPSK, π/4-QPSK transmits two bits per symbol. So only four carrier signals are needed but this is where the twist comes in. In QPSK we have four signals that are used to send the four two-bit symbols. In π/4-QPSK we have eight signals, every alternate symbol is transmitted using a π/4 shifted pattern of the QPSP constellation. Symbol A uses a signal on Path A as shown below and the next symbol, B, even if it is exactly the same bit pattern uses a signal on Path B. So we always get a phase shift even when the adjacent symbols are exactly the same.

The constellation diagram looks similar to 8-PSK. Note that a 8-PSK constellation can be broken into two QPSK constellations as show below. In π/4-QPSK, one symbol is transmitted on the A constellation and the next one is transmitted using the B constellation. Even though on a network analyzer, the constellation looks like 8-PSK, this modulation is strictly a form of QPSK with same BER and bandwidth. Although the symbols move around, they always convey just 2 bits per symbol.

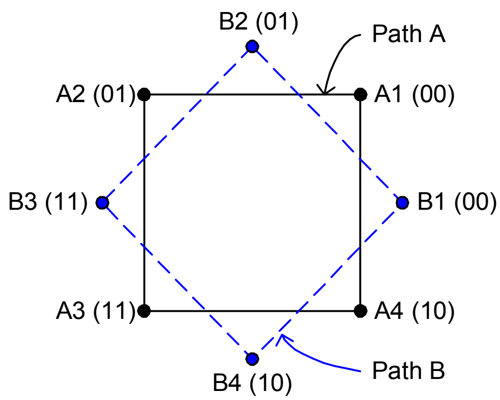


Figure 41 - π/4-QPSK constellation mimics 8-PSK but it is two QPSK constellations that are phase shifted.

Step-by-step π/4-QPSK

We wish to transmit the following bit sequence. We divide the bit sequence into 2-bit pieces just as we would do for QPSK

Bit sequence: 00 00 10 00 01 11 11 00 01 00

Transmit the first symbol using the A constellation shown in Figure 41 and the next symbol uses the B constellation. For each 2-bit, the I and Q values are the signal coordinates as shown below.

Symbol	**Bits**	**Symbol ID**	**I coordinate**	**Q coordinate**
1	00	A1	.707	.707
2	00	B1	0	1
3	10	A4	.707	-.707
4	00	B1	0	1
5	01	A2	-.707	.707
6	11	B2	-1	0
7	11	A3	-.707	-.707
8	00	B1	0	1
9	01	A2	-.707	.707
10	00	B1	0	1

Table 5 - π/4-QPSK symbols mapping to I and Q

The I and Q channels for a **π/4-QPSK signal** are shown below in Figure 42. Note that there are five possible levels (1, .707. 0 -.707, -1) and I and the Q channel show this variation in response to the symbols.

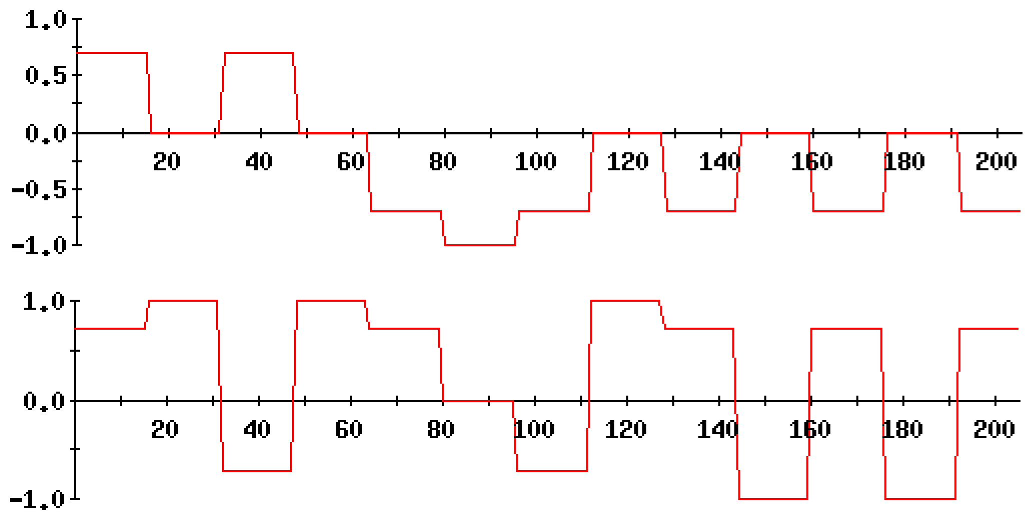


Figure 42 – I and Q mapping of π/4-QPSK symbols

Step 1 – Map bits to symbols

Bits	00	00	10	00	01	11	11	00	01	00
Symbols	A1	B1	A4	B1	A2	B2	A3	B1	A2	B3

Step 2 - Multiply the I and Q with a carrier (in the example below, the carrier frequency is 1 Hz.) and you get an 8-PSK signal constellation.

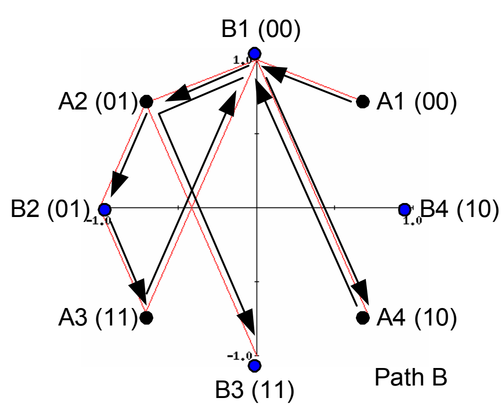


Figure 43 – of π/4-QPSK symbols traverse over a 8-PSK constellation

The constellation diagram is a path that the symbols have traced in time as we can see in the above diagram of just the symbols of this signal. The path stars with symbol A1, then goes to B1 which is on path B. From here, the next symbol A2 is back on Path A. Each transition, we see above goes back and forth between Path A and B.

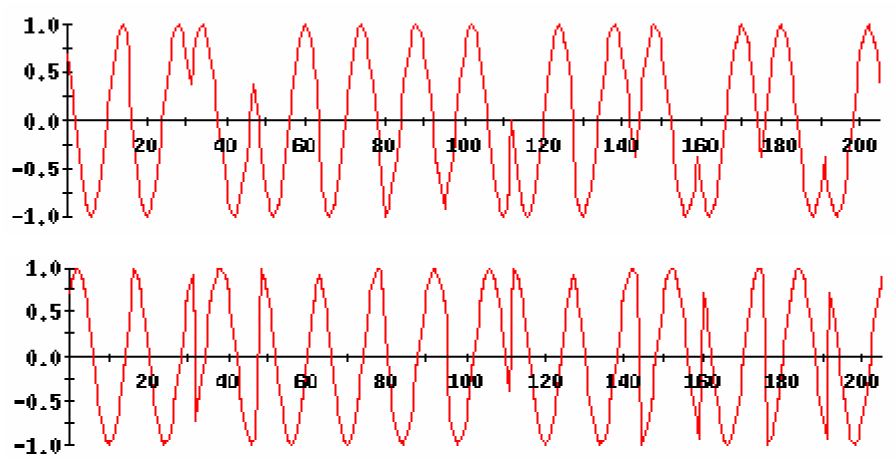


Figure 44 – π/4-QPSK modulated I and Q Channels

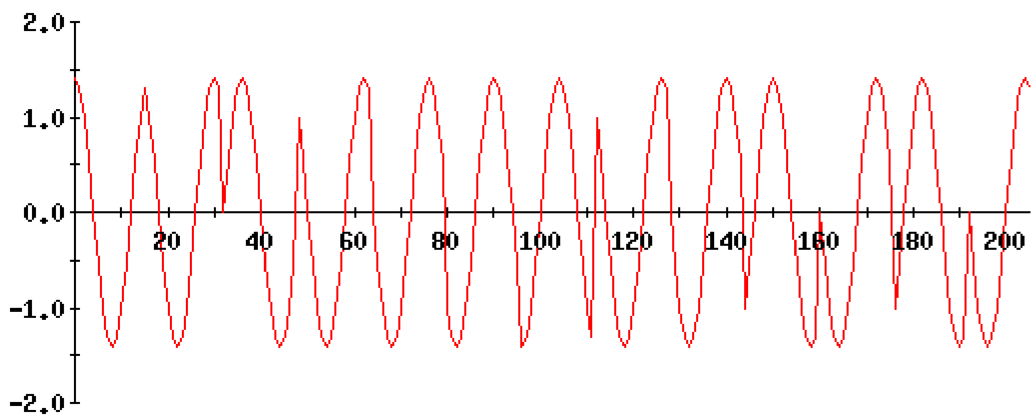


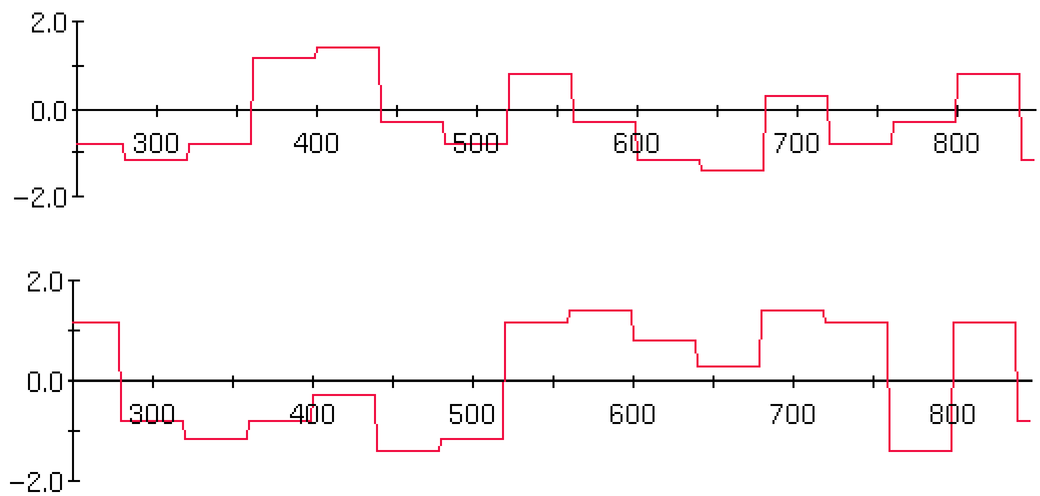
Figure 45 – π/4-QPSK modulated carrier

What is the advantage of doing this? On the average, the phase transitions are somewhat less than a straight QPSK and this does two things, one is that the side lobes are smaller so less adjacent carrier interference. Secondly the response to Class C amplifiers is better. This modulation is used in many mobile systems.

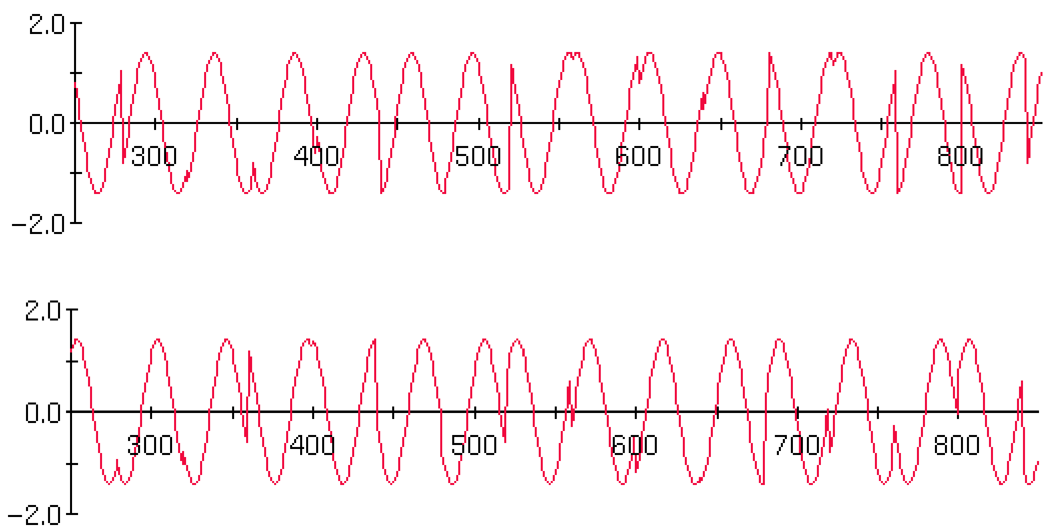
There is also a modification to this modulation where a differential encoding is added to the bits prior to modulation. (More about differential encoding in Tutorial 2.) When differential coding is added, the modulation is referred to as π/4-DQPSK.

16-PSK

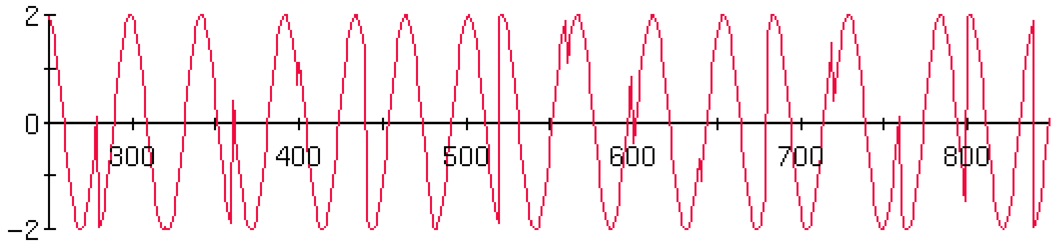
We can keep on subdividing the signal space into smaller regions. Doing so one more time for 8-PSK so that each signal is now only 22.5° apart, gives us 16-PSK. This will give 16 signals or symbols, so each symbol can convey 4 bits. Bit rate is now four times that of BPSK for the same symbol rate. The following figures show the 16-PSK signal at various stages during modulation.



(a) 16-PSK symbol mapping to I and Q channels. (Now the signal has 8 levels.)



(b) 16-PSK modulated I and Q Channels



(c) 16-PSK modulated carrier

Figure 46 – 16-PSK modulated signal

We can see where this is going. We can keep on increasing bits per symbol this way. However, 16-PSK is rarely used. Despite the fact that 16-PSK is bandwidth efficient is that it has higher bit error rate than a common modulation from the class of **Quadrature Amplitude Modulation** called 16-QAM which has the same bit efficiency.

This will be covered in the next part.

Thanks to Rolando Menendez for his help and corrections.

I can be contacted at
mntcastle@earthlink.net

Other tutorials at
www.complextoreal.com