HEWLETT PACKARD

Digital Modulation in Communications Systems – An Introduction

Application Note 1298

Introduction

This application note introduces the concepts of digital modulation used in many communications systems today. Emphasis is placed on explaining the tradeoffs that are made to optimize efficiencies in system design.

Most communications systems fall into one of three categories: bandwidth efficient, power efficient, or cost efficient. Bandwidth efficiency describes the ability of a modulation scheme to accommodate data within a limited bandwidth. Power efficiency describes the ability of the system to reliably send information at the lowest practical power level. In most systems, there is a high priority on bandwidth efficiency. The parameter to be optimized depends on the demands of the particular system, as can be seen in the following two examples.

For designers of digital terrestrial microwave radios, their highest priority is good bandwidth efficiency with low bit-error-rate. They have plenty of power available and are not concerned with power efficiency. They are not especially concerned with receiver cost or complexity because they do not have to build large numbers of them.

On the other hand, designers of hand-held cellular phones put a high priority on power efficiency because these phones need to run on a battery. Cost is also a high priority because cellular phones must be low-cost to encourage more users. Accordingly, these systems sacrifice some bandwidth efficiency to get power and cost efficiency.

Every time one of these efficiency parameters (bandwidth, power or cost) is increased, another one decreases, or becomes more complex or does not perform well in a poor environment. Cost is a dominant system priority. Low-cost radios will always be in demand. In the past, it was possible to make a radio low-cost by sacrificing power and bandwidth efficiency. This is no longer possible. The radio spectrum is very valuable and operators who do not use the spectrum efficiently could lose their existing licenses or lose out in the competition for new ones. These are the tradeoffs that must be considered in digital RF communications design.

This application note covers

- the reasons for the move to digital modulation;
- how information is modulated onto in-phase (*I*) and quadrature (*Q*) signals;
- different types of digital modulation;
- filtering techniques to conserve bandwidth;
- ways of looking at digitally modulated signals;
- multiplexing techniques used to share the transmission channel;
- how a digital transmitter and receiver work;
- measurements on digital RF communications systems;
- an overview table with key specifications for the major digital communications systems; and
- a glossary of terms used in digital RF communications.

These concepts form the building blocks of any communications system. If you understand the building blocks, then you will be able to understand how any communications system, present or future, works.

Table of contents

Table of contents

1. Why digital modulation?

The move to digital modulation provides more information capacity, compatibility with digital data services, higher data security, better quality communications, and quicker system availability. Developers of communications systems face these constraints:

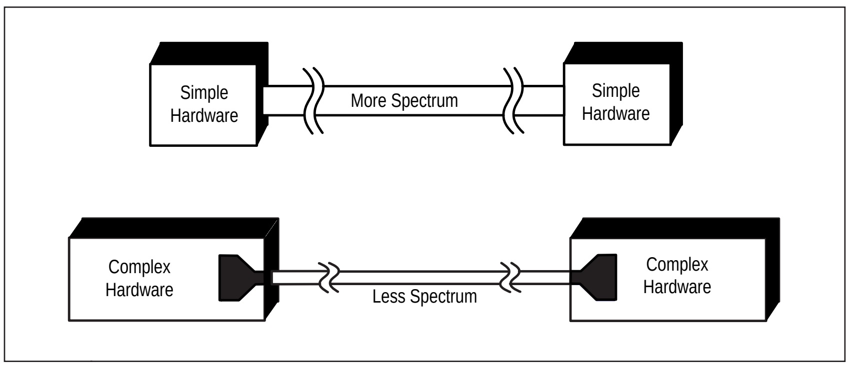
- available bandwidth
- permissible power
- inherent noise level of the system

The RF spectrum must be shared, yet every day there are more users for that spectrum as demand for communications services increases. Digital modulation schemes have greater capacity to convey large amounts of information than analog modulation schemes.

1.1 Trading off simplicity and bandwidth

There is a fundamental tradeoff in communication systems. Simple hardware can be used in transmitters and receivers to communicate information. However, this uses a lot of spectrum which limits the number of users. Alternatively, more complex transmitters and receivers can be used to transmit the same information over less bandwidth. The transition to more and more spectrally efficient transmission techniques requires more and more complex hardware. Complex hardware is difficult to design, test, and build. This tradeoff exists whether communication is over air or wire, analog or digital.

Figure 1.
The Fundamental Trade-off



1.2 Industry trends

Over the past few years a major transition has occurred from simple analog Amplitude Modulation (AM) and Frequency/Phase Modulation (FM/PM) to new digital modulation techniques. Examples of digital modulation include

- QPSK (Quadrature Phase Shift Keying)
- FSK (Frequency Shift Keying)
- MSK (Minimum Shift Keying)
- QAM (Quadrature Amplitude Modulation)

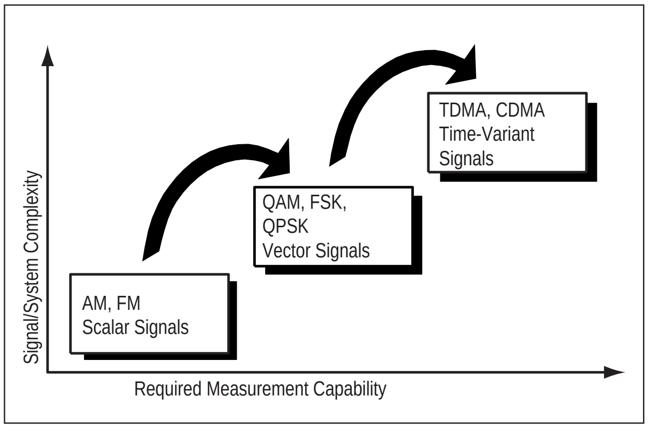


Figure 2.
Trends in the Industry

Another layer of complexity in many new systems is multiplexing. Two principal types of multiplexing (or “multiple access”) are TDMA (Time Division Multiple Access) and CDMA (Code Division Multiple Access). These are two different ways to add diversity to signals allowing different signals to be separated from one another.

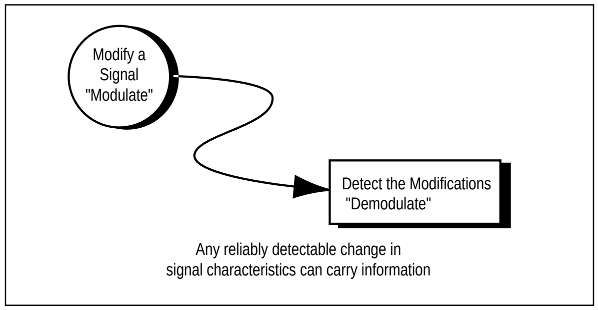
2. Using *I/Q* modulation to convey information.

2.1 Transmitting information

To transmit a signal over the air, there are three main steps:

1. A pure carrier is generated at the transmitter.
2. The carrier is modulated with the information to be transmitted. Any reliably detectable change in signal characteristics can carry information.
3. At the receiver the signal modifications or changes are detected and demodulated.

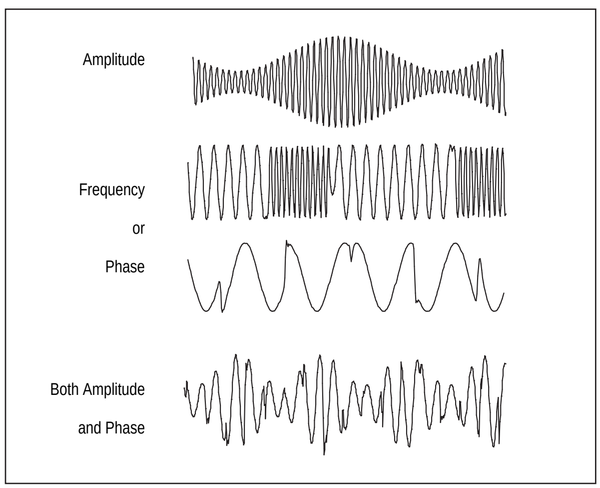
Figure 3. Transmitting Information... (Analog or Digital)



2.2 Signal characteristics that can be modified

There are only three characteristics of a signal that can be changed over time: amplitude, phase or frequency. However, phase and frequency are just different ways to view or measure the same signal change.

Figure 4. Signal Characteristics to Modify



In AM, the amplitude of a high-frequency carrier signal is varied in proportion to the instantaneous amplitude of the modulating message signal.

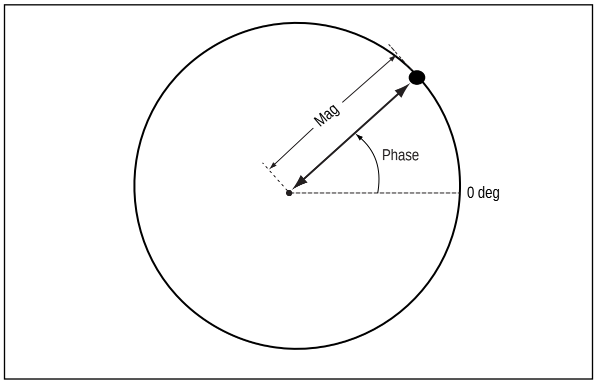
Frequency Modulation (FM) is the most popular analog modulation technique used in mobile communications systems. In FM, the amplitude of the modulating carrier is kept constant while its frequency is varied by the modulating message signal.

Amplitude and phase can be modulated simultaneously and separately, but this is difficult to generate, and especially difficult to detect. Instead, in practical systems the signal is separated into another set of independent components: *I* (In-phase) and *Q* (Quadrature). These components are orthogonal and do not interfere with each other.

2.3 Polar display - magnitude and phase represented together

A simple way to view amplitude and phase is with the polar diagram. The carrier becomes a frequency and phase reference and the signal is interpreted relative to the carrier. The signal can be expressed in polar form as a magnitude and a phase. The phase is relative to a reference signal, the carrier in most communication systems. The magnitude is either an absolute or relative value. Both are used in digital communication systems. Polar diagrams are the basis of many displays used in digital communications, although it is common to describe the signal vector by its rectangular coordinates of *I* (In-phase) and *Q* (Quadrature).

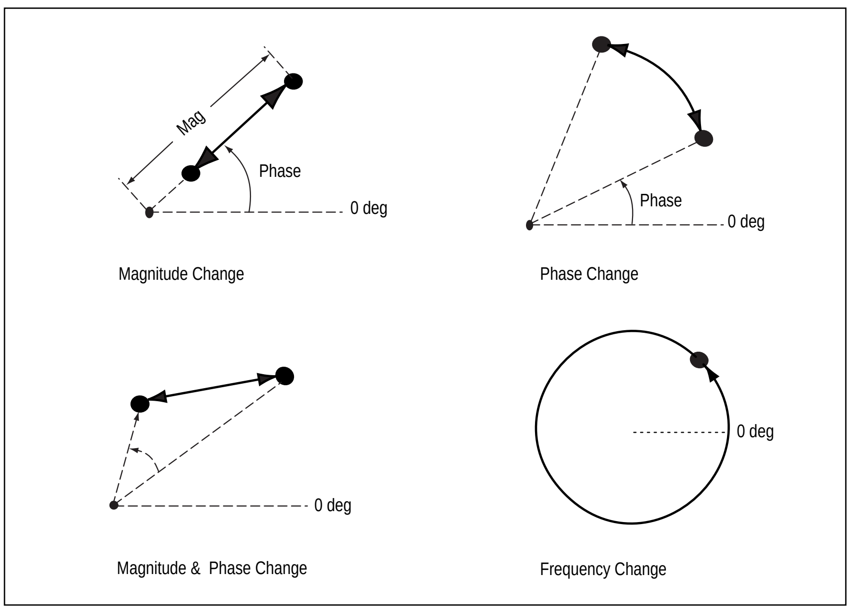
Figure 5.
Polar Display - Magnitude and Phase Represented Together



2.4 Signal changes or modifications in polar form

This figure shows different forms of modulation in polar form. Magnitude is represented as the distance from the center and phase is represented as the angle.

Figure 6.
Signal Changes or Modifications



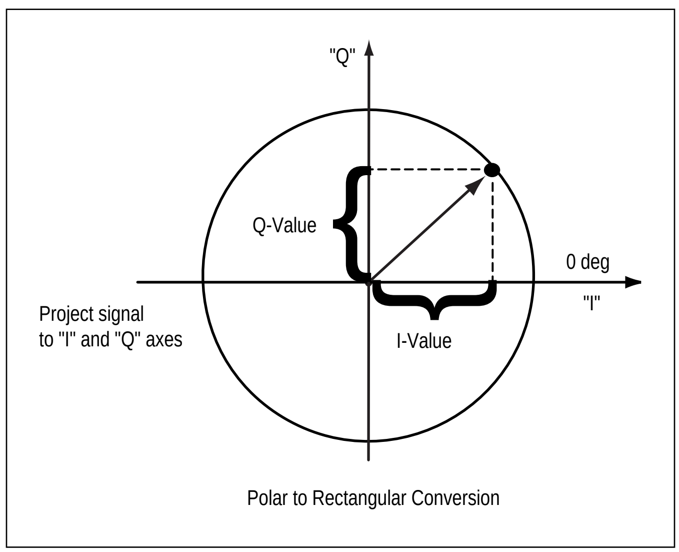
Amplitude modulation (AM) changes only the magnitude of the signal. Phase modulation (PM) changes only the phase of the signal. Amplitude and phase modulation can be used together. Frequency modulation (FM) looks similar to phase modulation, though frequency is the controlled parameter, rather than relative phase.

One example of the difficulties in RF design can be illustrated with simple amplitude modulation. Generating AM with no associated angular modulation should result in a straight line on a polar display. This line should run from the origin to some peak radius or amplitude value. In practice, however, the line is not straight. The amplitude modulation itself often can cause a small amount of unwanted phase modulation. The result is a curved line. It could also be a loop if there is any hysteresis in the system transfer function. Some amount of this distortion is inevitable in any system where modulation causes amplitude changes. Therefore, the degree of effective amplitude modulation in a system will affect some distortion parameters.

2.5 *I/Q* formats

In digital communications, modulation is often expressed in terms of *I* and *Q*. This is a rectangular representation of the polar diagram. On a polar diagram, the *I* axis lies on the zero degree phase reference, and the *Q* axis is rotated by 90 degrees. The signal vector's projection onto the *I* axis is its "I" component and the projection onto the *Q* axis is its "Q" component.

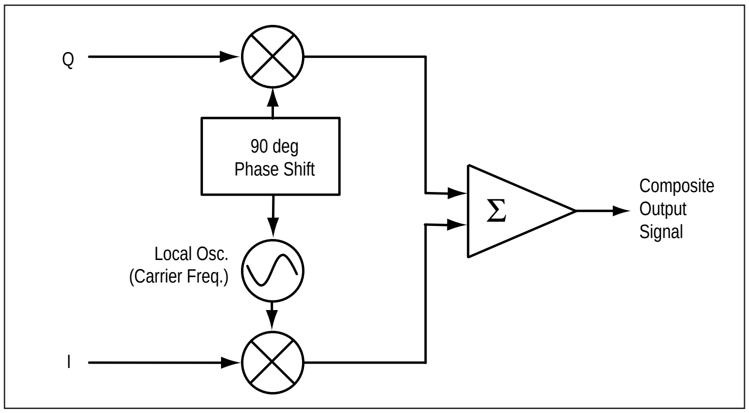
Figure 7.
"*I-Q*" Format



2.6 *I* and *Q* in a radio transmitter

I/Q diagrams are particularly useful because they mirror the way most digital communications signals are created using an *I/Q* modulator. In the transmitter, *I* and *Q* signals are mixed with the same local oscillator (LO). A 90 degree phase shifter is placed in one of the LO paths. Signals that are separated by 90 degrees are also known as being orthogonal to each other or in quadrature. Signals that are in quadrature do not interfere with each other. They are two independent components of the signal. When recombined, they are summed to a composite output signal. There are two independent signals in *I* and *Q* that can be sent and received with simple circuits. This simplifies the design of digital radios. The main advantage of *I/Q* modulation is the symmetric ease of combining independent signal components into a single composite signal and later splitting such a composite signal into its independent component parts.

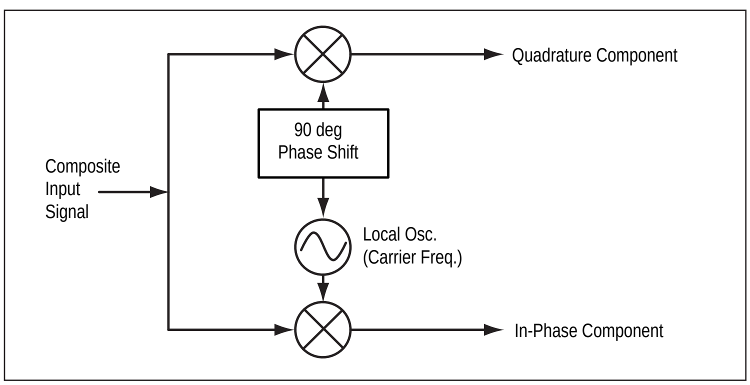
Figure 8.
***I* and *Q* in a Practical Radio Transmitter**



2.7 *I* and *Q* in a radio receiver

The composite signal with magnitude and phase (or *I* and *Q*) information arrives at the receiver input. The input signal is mixed with the local oscillator signal at the carrier frequency in two forms. One is at an arbitrary zero phase. The other has a 90 degree phase shift. The composite input signal (in terms of magnitude and phase) is thus broken into an in-phase, *I*, and a quadrature, *Q*, component. These two components of the signal are independent and orthogonal. One can be changed without affecting the other. Normally, information cannot be plotted in a polar format and reinterpreted as rectangular values without doing a polar-to-rectangular conversion. This conversion is exactly what is done by the in-phase and quadrature mixing processes in a digital radio. A local oscillator, phase shifter, and two mixers can perform the conversion accurately and efficiently.

Figure 9.
***I* and *Q* in a Radio Receiver**



2.8 Why use I and Q?

Digital modulation is easy to accomplish with I/Q modulators. Most digital modulation maps the data to a number of discrete points on the I/Q plane. These are known as constellation points. As the signal moves from one point to another, simultaneous amplitude and phase modulation usually results. To accomplish this with an amplitude modulator and a phase modulator is difficult and complex. It is also impossible with a conventional phase modulator. The signal may, in principal, circle the origin in one direction forever, necessitating infinite phase shifting capability. Alternatively, simultaneous AM and Phase Modulation is easy with an I/Q modulator. The I and Q control signals are bounded, but infinite phase wrap is possible by properly phasing the I and Q signals.

3. Digital modulation types and relative efficiencies

This section covers the main digital modulation formats, their main applications, relative spectral efficiencies and some variations of the main modulation types as used in practical systems. Fortunately, there are a limited number of modulation types which form the building blocks of any system.

3.1 Applications

This table covers the applications for different modulation formats in both wireless communications and video.

Modulation format	Application
MSK, GMSK	GSM, CDPD
BPSK	Deep space telemetry, cable modems
QPSK, $^{\pi}/_{4}$ DQPSK	Satellite, CDMA, NADC, TETRA, PHS, PDC, LMDS, DVB-S, cable (return path), cable modems, TFTS
OQPSK	CDMA, satellite
FSK, GFSK	DECT, paging, RAM mobile data, AMPS, CT2, ERMES, land mobile, public safety
8, 16 VSB	North American digital TV (ATV), broadcast, cable
8PSK	Satellite, aircraft, telemetry pilots for monitoring broadband video systems
16 QAM	Microwave digital radio, modems, DVB-C, DVB-T
32 QAM	Terrestrial microwave, DVB-T
64 QAM	DVB-C, modems, broadband set top boxes, MMDS
256 QAM	Modems, DVB-C (Europe), Digital Video (US)

Although this note focuses on wireless communications, video applications have also been included in the table for completeness and because of their similarity to other wireless communications.

3.1.1 Bit rate and symbol rate

To understand and compare different modulation format efficiencies, it is important to first understand the difference between bit rate and symbol rate. The signal bandwidth for the communications channel needed depends on the symbol rate, not on the bit rate.

$$\text{Symbol rate} = \frac{\text{bit rate}}{\text{the number of bits transmitted with each symbol}}$$

Bit rate is the frequency of a system bit stream. Take, for example, a radio with an 8 bit sampler, sampling at 10 kHz for voice. The bit rate, the basic bit stream rate in the radio, would be eight bits multiplied by 10K samples per second, or 80 Kbits per second. (For the moment we will ignore the extra bits required for synchronization, error correction, etc.).

Figure 10.
Bit Rate and Symbol Rate

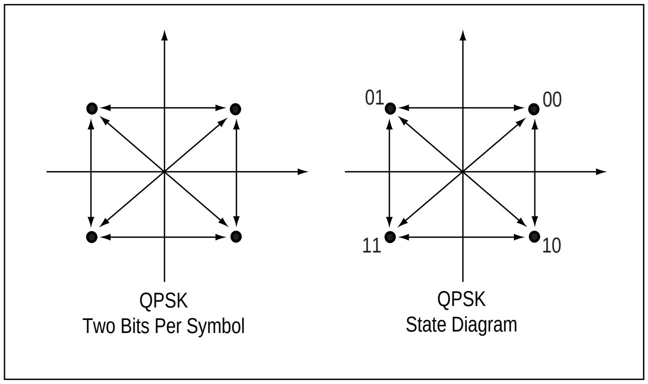


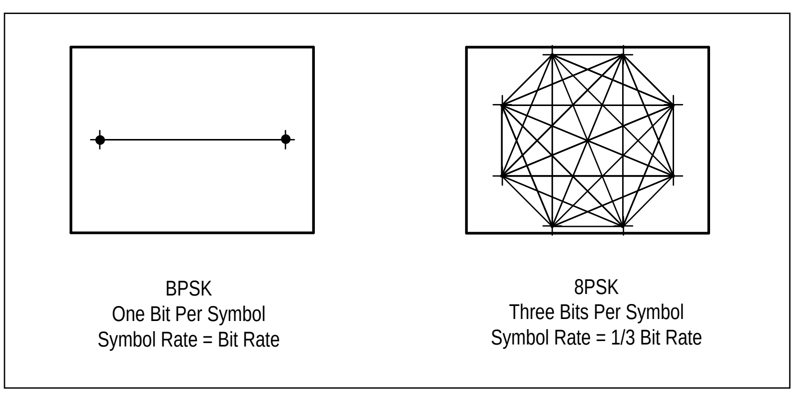
Figure 10 is an example of a state diagram of a Quadrature Phase Shift Keying (QPSK) signal. The states can be mapped to zeros and ones. This is a common mapping, but it is not the only one. Any mapping can be used.

The symbol rate is the bit rate divided by the number of bits that can be transmitted with each symbol. If one bit is transmitted per symbol, as with BPSK, then the symbol rate would be the same as the bit rate of 80 Kbits per second. If two bits are transmitted per symbol, as in QPSK, then the symbol rate would be half of the bit rate or 40 Kbits per second. Symbol rate is sometimes called baud rate. Note that baud rate is not the same as bit rate. These terms are often confused. If more bits can be sent with each symbol, then the same amount of data can be sent in a narrower spectrum. This is why modulation formats that are more complex and use a higher number of states can send the same information over a narrower piece of the RF spectrum.

3.1.2 Spectrum (bandwidth) requirements

An example of how symbol rate influences spectrum requirements can be seen in eight-state Phase Shift Keying (8PSK). It is a variation of PSK. There are eight possible states that the signal can transition to at any time. The phase of the signal can take any of eight values at any symbol time. Since $2^3 = 8$, there are three bits per symbol. This means the symbol rate is one third of the bit rate. This is relatively easy to decode.

Figure 11.
Spectrum Requirements



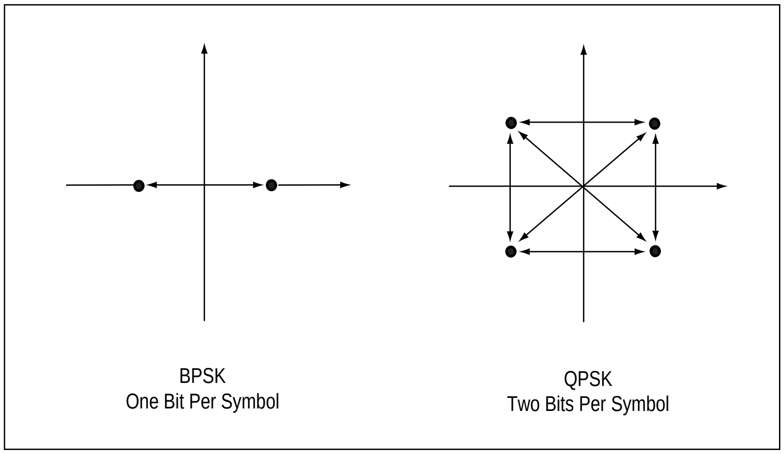
3.1.3 Symbol clock

The symbol clock represents the frequency and exact timing of the transmission of the individual symbols. At the symbol clock transitions, the transmitted carrier is at the correct *I/Q* (or magnitude/phase) value to represent a specific symbol (a specific point in the constellation).

3.2 Phase Shift Keying

One of the simplest forms of digital modulation is binary or Bi-Phase Shift Keying (BPSK). One application where this is used is for deep space telemetry. The phase of a constant amplitude carrier signal moves between zero and 180 degrees. On an *I* and *Q* diagram, the *I* state has two different values. There are two possible locations in the state diagram, so a binary one or zero can be sent. The symbol rate is one bit per symbol.

Figure 12.
Phase Shift Keying

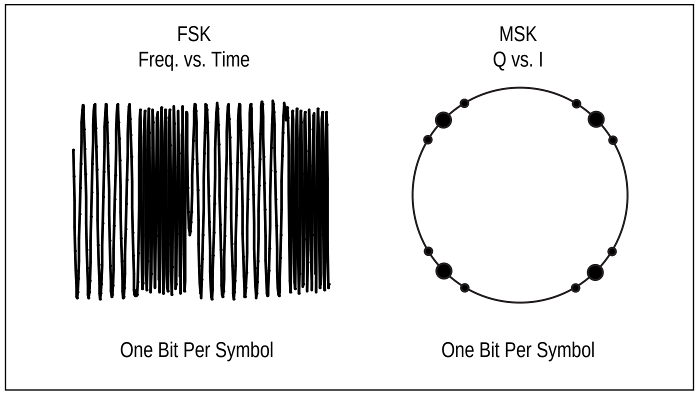


A more common type of phase modulation is Quadrature Phase Shift Keying (QPSK). It is used extensively in applications including CDMA (Code Division Multiple Access) cellular service, wireless local loop, Iridium (a voice/data satellite system) and DVB-S (Digital Video Broadcasting - Satellite). Quadrature means that the signal shifts between phase states which are separated by 90 degrees. The signal shifts in increments of 90 degrees from 45 to 135, –45, or –135 degrees. These points are chosen as they can be easily implemented using an *I/Q* modulator. Only two *I* values and two *Q* values are needed and this gives two bits per symbol. There are four states because $2^2 = 4$. It is therefore a more bandwidth-efficient type of modulation than BPSK, potentially twice as efficient.

3.3 Frequency Shift Keying

Frequency modulation and phase modulation are closely related. A static frequency shift of +1 Hz means that the phase is constantly advancing at the rate of 360 degrees per second (2 π rad/sec), relative to the phase of the unshifted signal.

Figure 13. Frequency Shift Keying



FSK (Frequency Shift Keying) is used in many applications including cordless and paging systems. Some of the cordless systems include DECT (Digital Enhanced Cordless Telephone) and CT2 (Cordless Telephone 2).

In FSK, the frequency of the carrier is changed as a function of the modulating signal (data) being transmitted. Amplitude remains unchanged. In binary FSK (BFSK or 2FSK), a “1” is represented by one frequency and a “0” is represented by another frequency.

3.4 Minimum Shift Keying

Since a frequency shift produces an advancing or retarding phase, frequency shifts can be detected by sampling phase at each symbol period. Phase shifts of $(2N + 1)\,\pi/2$ radians are easily detected with an *I/Q* demodulator. At even numbered symbols, the polarity of the *I* channel conveys the transmitted data, while at odd numbered symbols the polarity of the *Q* channel conveys the data. This orthogonality between *I* and *Q* simplifies detection algorithms and hence reduces power consumption in a mobile receiver. The minimum frequency shift which yields orthogonality of *I* and *Q* is that which results in a phase shift of $\pm\pi/2$ radians per symbol (90 degrees per symbol). FSK with this deviation is called MSK (Minimum Shift Keying). The deviation must be accurate in order to generate repeatable 90 degree phase shifts. MSK is used in the GSM (Global System for Mobile Communications) cellular standard. A phase shift of +90 degrees represents a data bit equal to “1”, while –90 degrees represents a “0”. The peak-to-peak frequency shift of an MSK signal is equal to one-half of the bit rate.

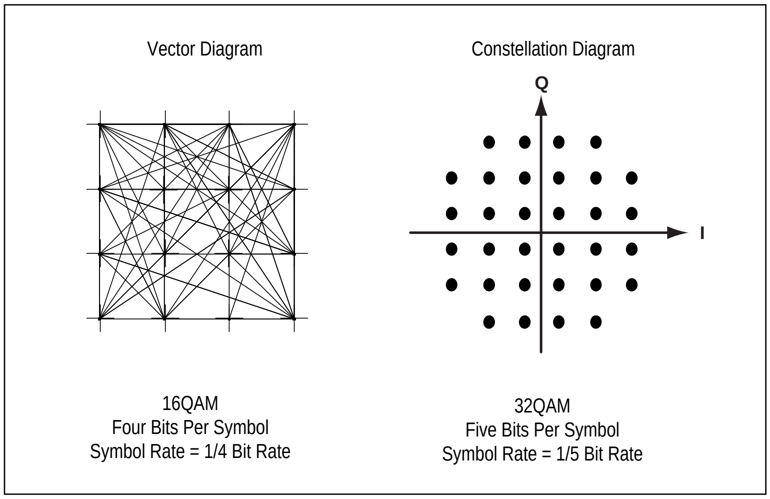
FSK and MSK produce constant envelope carrier signals, which have no amplitude variations. This is a desirable characteristic for improving the power efficiency of transmitters. Amplitude variations can exercise nonlinearities in an amplifier’s amplitude-transfer function, generating spectral regrowth, a component of adjacent channel power. Therefore, more efficient amplifiers (which tend to be less linear) can be used with constant-envelope signals, reducing power consumption.

MSK has a narrower spectrum than wider deviation forms of FSK. The width of the spectrum is also influenced by the waveforms causing the frequency shift. If those waveforms have fast transitions or a high slew rate, then the spectrum of the transmitter will be broad. In practice, the waveforms are filtered with a Gaussian filter, resulting in a narrow spectrum. In addition, the Gaussian filter has no time-domain overshoot, which would broaden the spectrum by increasing the peak deviation. MSK with a Gaussian filter is termed GMSK (Gaussian MSK).

3.5 Quadrature Amplitude Modulation

Another member of the digital modulation family is Quadrature Amplitude Modulation (QAM). QAM is used in applications including microwave digital radio, DVB-C (Digital Video Broadcasting - Cable) and modems.

Figure 14. Quadrature Amplitude Modulation



In 16-state Quadrature Amplitude Modulation (16QAM), there are four *I* values and four *Q* values. This results in a total of 16 possible states for the signal. It can transition from any state to any other state at every symbol time. Since $16 = 2^4$, four bits per symbol can be sent. This consists of two bits for *I* and two bits for *Q*. The symbol rate is one fourth of the bit rate. So this modulation format produces a more spectrally efficient transmission. It is more efficient than BPSK, QPSK or 8PSK. Note that QPSK is the same as 4QAM.

Another variation is 32QAM. In this case there are six *I* values and six *Q* values resulting in a total of 36 possible states (6x6=36). This is too many states for a power of two (the closest power of two is 32). So the four corner symbol states, which take the most power to transmit, are omitted. This reduces the amount of peak power the transmitter has to generate. Since $2^5 = 32$, there are five bits per symbol and the symbol rate is one fifth of the bit rate.

The current practical limits are approximately 256QAM, though work is underway to extend the limits to 512 or 1024 QAM. A 256QAM system uses 16 *I*-values and 16 *Q*-values giving 256 possible states. Since $2^8 = 256$, each symbol can represent eight bits. A 256QAM signal that can send eight bits per symbol is very spectrally efficient. However, the symbols are very close together and are thus more subject to errors due to noise and distortion. Such a signal may have to be transmitted with extra power (to effectively spread the symbols out more) and this reduces power efficiency as compared to simpler schemes.

Compare the bandwidth efficiency when using 256QAM versus BPSK modulation in the radio example in section 3.1.1 (which uses an eight-bit sampler sampling at 10 kHz for voice). BPSK uses 80 Ksymbols-per-second sending 1 bit per symbol. A system using 256QAM sends eight bits per symbol so the symbol rate would be 10 Ksymbols per second. A 256QAM system enables the same amount of information to be sent as BPSK using only one eighth of the bandwidth. It is eight times more bandwidth efficient. However, there is a tradeoff. The radio becomes more complex and is more susceptible to errors caused by noise and distortion. Error rates of higher-order QAM systems such as this degrade more rapidly than QPSK as noise or interference is introduced. A measure of this degradation would be a higher Bit Error Rate (BER).

In any digital modulation system, if the input signal is distorted or severely attenuated the receiver will eventually lose symbol lock completely. If the receiver can no longer recover the symbol clock, it cannot demodulate the signal or recover any information. With less degradation, the symbol clock can be recovered, but it is noisy, and the symbol locations themselves are noisy. In some cases, a symbol will fall far enough away from its intended position that it will cross over to an adjacent position. The *I* and *Q* level detectors used in the demodulator would misinterpret such a symbol as being in the wrong location, causing bit errors. QPSK is not as efficient, but the states are much farther apart and the system can tolerate a lot more noise before suffering symbol errors. QPSK has no intermediate states between the four corner-symbol locations so there is less opportunity for the demodulator to misinterpret symbols. QPSK requires less transmitter power than QAM to achieve the same bit error rate.

3.6 Theoretical bandwidth efficiency limits

Bandwidth efficiency describes how efficiently the allocated bandwidth is utilized or the ability of a modulation scheme to accommodate data, within a limited bandwidth. This table shows the theoretical bandwidth efficiency limits for the main modulation types. Note that these figures cannot actually be achieved in practical radios since they require perfect modulators, demodulators, filter and transmission paths.

Modulation format	Theoretical bandwidth efficiency limits
MSK	1 bit/second/Hz
BPSK	1 bit/second/Hz
QPSK	2 bits/second/Hz
8PSK	3 bits/second/Hz
16 QAM	4 bits/second/Hz
32 QAM	5 bits/second/Hz
64 QAM	6 bits/second/Hz
256 QAM	8 bits/second/Hz

If the radio had a perfect (rectangular in the frequency domain) filter, then the occupied bandwidth could be made equal to the symbol rate.

Techniques for maximizing spectral efficiency include the following:

- Relate the data rate to the frequency shift (as in GSM).
- Use premodulation filtering to reduce the occupied bandwidth. Raised cosine filters, as used in NADC, PDC, and PHS give the best spectral efficiency.
- Restrict the types of transitions.

Effects of going through the origin

Take, for example, a QPSK signal where the normalized value changes from 1, 1 to –1, –1. When changing simultaneously from I and Q values of +1 to I and Q values of –1, the signal trajectory goes through the origin (the I/Q value of 0,0). The origin represents 0 carrier magnitude. A value of 0 magnitude indicates that the carrier amplitude is 0 for a moment.

Not all transitions in QPSK result in a trajectory that goes through the origin. If I changes value but Q does not (or vice-versa) the carrier amplitude changes a little, but it does not go through zero. Therefore some symbol transitions will result in a small amplitude variation, while others will result in a very large amplitude variation. The clock-recovery circuit in the receiver must deal with this amplitude variation uncertainty if it uses amplitude variations to align the receiver clock with the transmitter clock.

Spectral regrowth does not automatically result from these trajectories that pass through or near the origin. If the amplifier and associated circuits are perfectly linear, the spectrum (spectral occupancy or occupied bandwidth) will be unchanged. The problem lies in nonlinearities in the circuits.

A signal which changes amplitude over a very large range will exercise these nonlinearities to the fullest extent. These nonlinearities will cause distortion products. In continuously-modulated systems they will cause "spectral regrowth" or wider modulation sidebands (a phenomenon related to intermodulation distortion). Another term which is sometimes used in this context is "spectral splatter". However this is a term that is more correctly used in association with the increase in the bandwidth of a signal caused by pulsing on and off.

3.7 Spectral efficiency examples in practical radios

The following examples indicate spectral efficiencies that are achieved in some practical radio systems.

The TDMA version of the North American Digital Cellular (NADC) system, achieves a 48 Kbits-per-second data rate over a 30 kHz bandwidth or 1.6 bits per second per Hz. It is a $\pi/4$ DQPSK based system and transmits two bits per symbol. The theoretical efficiency would be two bits per second per Hz and in practice it is 1.6 bits per second per Hz.

Another example is a microwave digital radio using 16QAM. This kind of signal is more susceptible to noise and distortion than something simpler such as QPSK. This type of signal is usually sent over a direct line-of-sight microwave link or over a wire where there is very little noise and interference. In this microwave-digital-radio example the bit rate is 140 Mbits per second over a very wide bandwidth of 52.5 MHz. The spectral efficiency is 2.7 bits per second per Hz. To implement this, it takes a very clear line-of-sight transmission path and a precise and optimized high-power transceiver.

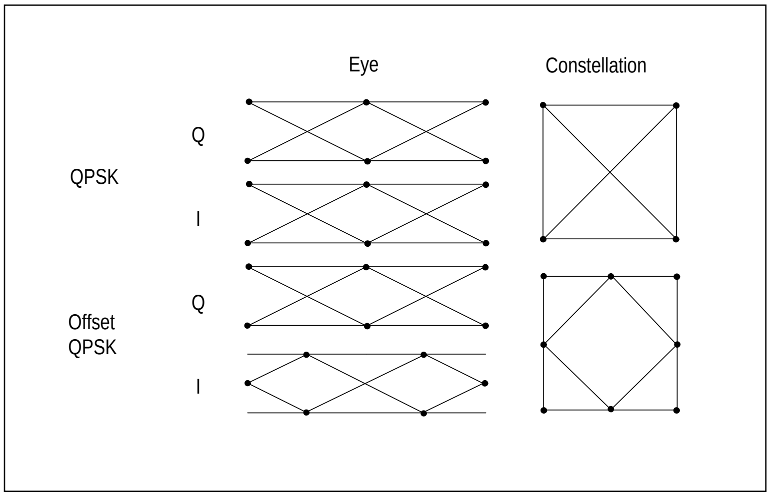
Digital modulation types - variations

The modulation types outlined in sections 3.2 to 3.4 form the building blocks for many systems. There are three main variations on these basic building blocks that are used in communications systems: *I/Q* offset modulation, differential modulation, and constant envelope modulation.

3.8 *I/Q* offset modulation

The first variation is offset modulation. One example of this is Offset QPSK (OQPSK). This is used in the cellular CDMA (Code Division Multiple Access) system for the reverse (mobile to base) link.

Figure 15.
***I-Q* "Offset" Modulation**



In QPSK, the *I* and *Q* bit streams are switched at the same time. The symbol clocks, or the *I* and *Q* digital signal clocks, are synchronized. In Offset QPSK (OQPSK), the *I* and *Q* bit streams are offset in their relative alignment by one bit period (one half of a symbol period). This is shown in the diagram. Since the transitions of *I* and *Q* are offset, at any given time only one of the two bit streams can change values. This creates a dramatically different constellation, even though there are still just two *I/Q* values. This has power efficiency advantages. In OQPSK the signal trajectories are modified by the symbol clock offset so that the carrier amplitude does not go through or near zero (the center of the constellation). The spectral efficiency is the same with two *I* states and two *Q* states. The reduced amplitude variations (perhaps 3 dB for OQPSK, versus 30 to 40 dB for QPSK) allow a more power-efficient, less linear RF power amplifier to be used.

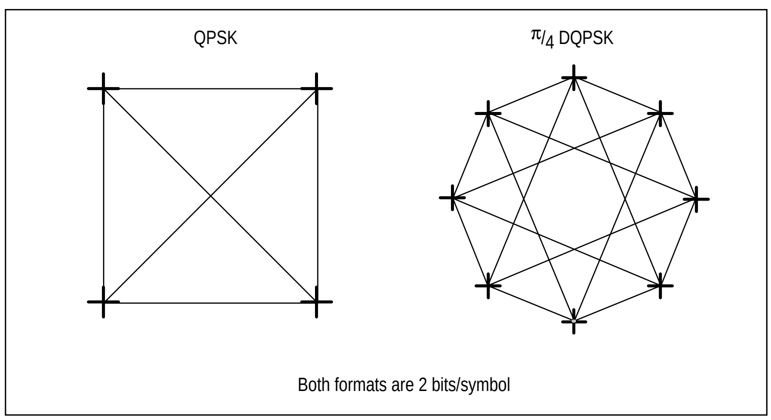
3.9 Differential modulation

The second variation is differential modulation as used in differential QPSK (DQPSK) and differential 16QAM (D16QAM). Differential means that the information is not carried by the absolute state, it is carried by the transition between states. In some cases there are also restrictions on allowable transitions. This occurs in $^{\pi}/_{4}$ DQPSK where the carrier trajectory does not go through the origin. A DQPSK transmission system can transition from any symbol position to any other symbol position. The $^{\pi}/_{4}$ DQPSK modulation format is widely used in many applications including

- cellular
  -NADC- IS-54 (North American digital cellular)
  -PDC (Pacific Digital Cellular)
- cordless
  -PHS (personal handyphone system)
- trunked radio
  -TETRA (Trans European Trunked Radio)

The $^{\pi}/_{4}$ DQPSK modulation format uses two QPSK constellations offset by 45 degrees ($^{\pi}/_{4}$ radians). Transitions must occur from one constellation to the other. This guarantees that there is always a change in phase at each symbol, making clock recovery easier. The data is encoded in the magnitude and direction of the phase shift, not in the absolute position on the constellation. One advantage of $^{\pi}/_{4}$ DQPSK is that the signal trajectory does not pass through the origin, thus simplifying transmitter design. Another is that $^{\pi}/_{4}$ DQPSK, with root raised cosine filtering, has better spectral efficiency than GMSK, the other common cellular modulation type.

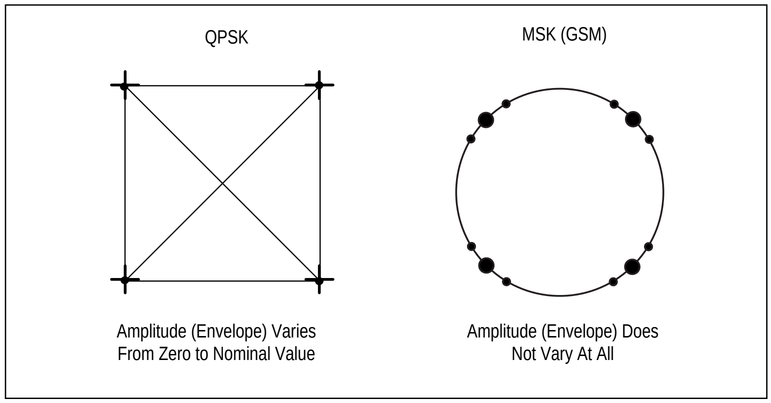
Figure 16. "Differential" Modulation



3.10 Constant amplitude modulation

The third variation is constant-envelope modulation. GSM uses a variation of constant amplitude modulation format called 0.3 GMSK (Gaussian Minimum Shift Keying).

Figure 17.
Constant Amplitude Modulation



In constant-envelope modulation the amplitude of the carrier is constant, regardless of the variation in the modulating signal. It is a power-efficient scheme that allows efficient class-C amplifiers to be used without introducing degradation in the spectral occupancy of the transmitted signal. However, constant-envelope modulation techniques occupy a larger bandwidth than schemes which are linear. In linear schemes, the amplitude of the transmitted signal varies with the modulating digital signal as in BPSK or QPSK. In systems where bandwidth efficiency is more important than power efficiency, constant envelope modulation is not as well suited.

MSK (covered in section 3.4) is a special type of FSK where the peak-to-peak frequency deviation is equal to half the bit rate.

GMSK is a derivative of MSK where the bandwidth required is further reduced by passing the modulating waveform through a Gaussian filter. The Gaussian filter minimizes the instantaneous frequency variations over time. GMSK is a spectrally efficient modulation scheme and is particularly useful in mobile radio systems. It has a constant envelope, spectral efficiency, good BER performance and is self-synchronizing.

4. Filtering

Filtering allows the transmitted bandwidth to be significantly reduced without losing the content of the digital data. This improves the spectral efficiency of the signal.

There are many different varieties of filtering. The most common are

- raised cosine
- square-root raised cosine
- Gaussian filters

Any fast transition in a signal, whether it be amplitude, phase or frequency will require a wide occupied bandwidth. Any technique that helps to slow down these transitions will narrow the occupied bandwidth. Filtering serves to smooth these transitions (in *I* and *Q*). Filtering reduces interference because it reduces the tendency of one signal or one transmitter to interfere with another in a Frequency-Division-Multiple-Access (FDMA) system. On the receiver end, reduced bandwidth improves sensitivity because more noise and interference are rejected.

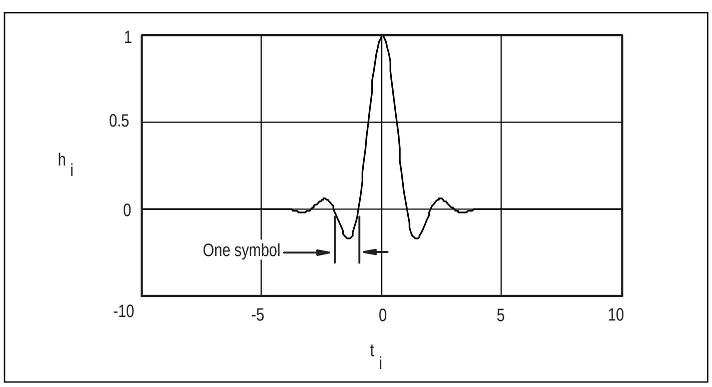
Some tradeoffs must be made. One is that some types of filtering cause the trajectory of the signal (the path of transitions between the states) to overshoot in many cases. This overshoot can occur in certain types of filters such as Nyquist. This overshoot path represents carrier power and phase. For the carrier to take on these values it requires more output power from the transmitter amplifiers. It requires more power than would be necessary to transmit the actual symbol itself. Carrier power cannot be clipped or limited (to reduce or eliminate the overshoot) without causing the spectrum to spread out again. Since narrowing the spectral occupancy was the reason the filtering was inserted in the first place, it becomes a very fine balancing act.

Other tradeoffs are that filtering makes the radios more complex and can make them larger, especially if performed in an analog fashion. Filtering can also create Inter-Symbol Interference (ISI). This occurs when the signal is filtered enough so that the symbols blur together and each symbol affects those around it. This is determined by the time-domain response, or impulse response of the filter.

4.1 Nyquist or raised cosine filter

This graph shows the impulse or time-domain response of a raised cosine filter, one class of Nyquist filter. Nyquist filters have the property that their impulse response rings at the symbol rate. The filter is chosen to ring, or have the impulse response of the filter cross through zero, at the symbol clock frequency.

Figure 18. Nyquit or Raised Cosine Filter

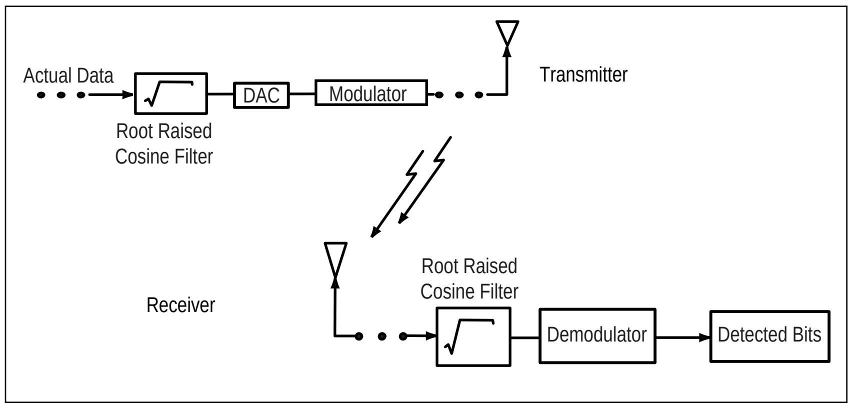


The time response of the filter goes through zero with a period that exactly corresponds to the symbol spacing. Adjacent symbols do not interfere with each other at the symbol times because the response equals zero at all symbol times except the center (desired) one. Nyquist filters heavily filter the signal without blurring the symbols together at the symbol times. This is important for transmitting information without errors caused by Inter-Symbol Interference. Note that Inter-Symbol Interference does exist at all times except the symbol (decision) times. Usually the filter is split, half being in the transmit path and half in the receiver path. In this case root Nyquist filters (commonly called root raised cosine) are used in each part, so that their combined response is that of a Nyquist filter.

4.2 Transmitter-receiver matched filters

Sometimes filtering is desired at both the transmitter and receiver. Filtering in the transmitter reduces the adjacent-channel-power radiation of the transmitter, and thus its potential for interfering with other transmitters.

Figure 19. Transmitter-Receiver Matched Filters



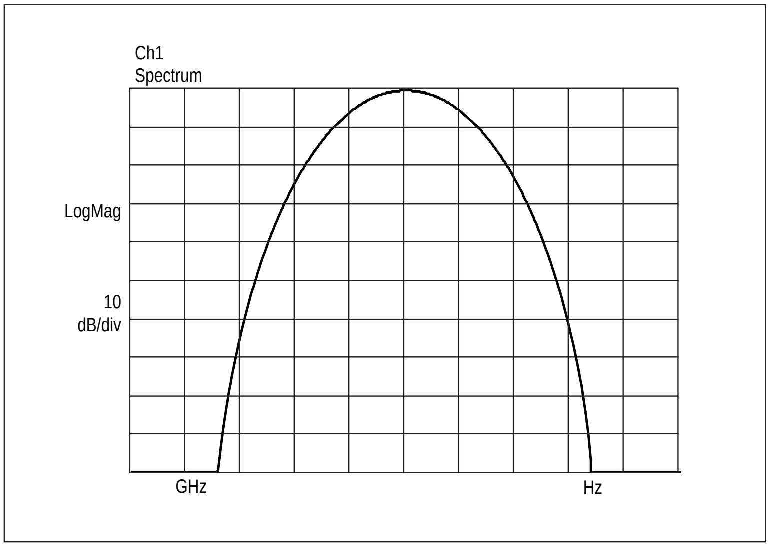
Filtering at the receiver reduces the effects of broadband noise and also interference from other transmitters in nearby channels.

To get zero Inter-Symbol Interference (ISI), both filters are designed until the combined result of the filters and the rest of the system is a full Nyquist filter. Potential differences can cause problems in manufacturing because the transmitter and receiver are often manufactured by different companies. The receiver may be a small hand-held model and the transmitter may be a large cellular base station. If the design is performed correctly the results are the best data rate, the most efficient radio, and reduced effects of interference and noise. This is why root-Nyquist filters are used in receivers and transmitters as $\sqrt{Nyquist} \ x \ \sqrt{Nyquist} = Nyquist$. Matched filters are not used in Gaussian filtering.

4.3 Gaussian filter

In contrast, a GSM signal will have a small blurring of symbols on each of the four states because the Gaussian filter used in GSM does not have zero Inter-Symbol Interference. The phase states vary somewhat causing a blurring of the symbols as shown in figure 17. Wireless system architects must decide just how much of the Inter-Symbol Interference can be tolerated in a system and combine that with noise and interference.

Figure 20.
Gaussian Filter



Gaussian filters are used in GSM because of their advantages in carrier power, occupied bandwidth and symbol-clock recovery. The Gaussian filter is a Gaussian shape in both the time and frequency domains, and it does not ring like the raised cosine filters do. Its effects in the time domain are relatively short and each symbol interacts significantly (or causes ISI) with only the preceding and succeeding symbols. This reduces the tendency for particular sequences of symbols to interact which makes amplifiers easier to build and more efficient.

4.4 Filter bandwidth parameter alpha

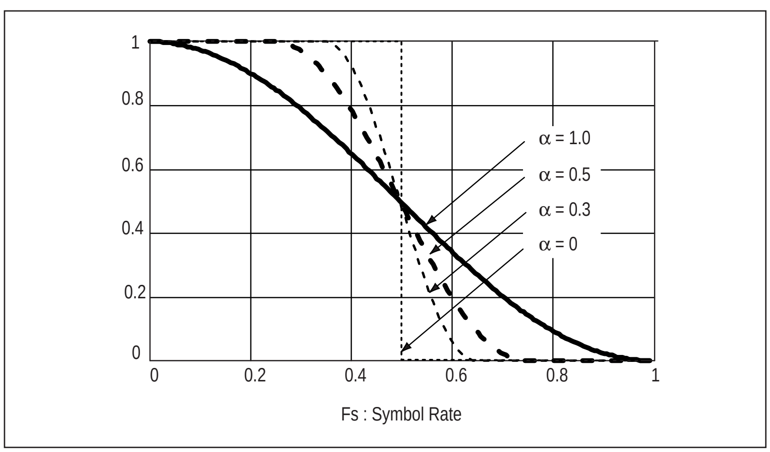
The sharpness of a raised cosine filter is described by alpha (α). Alpha gives a direct measure of the occupied bandwidth of the system and is calculated as

occupied bandwidth = symbol rate X *(1 +* α*).*

If the filter had a perfect (brick wall) characteristic with sharp transitions and an alpha of zero, the occupied bandwidth would be

for α *= 0, occupied bandwidth = symbol rate* X *(1 + 0) = symbol rate.*

Figure 21.
Filter Bandwidth
Parameters "α"



In a perfect world, the occupied bandwidth would be the same as the symbol rate, but this is not practical. An alpha of zero is impossible to implement.

Alpha is sometimes called the "excess bandwidth factor" as it indicates the amount of occupied bandwidth that will be required in excess of the ideal occupied bandwidth (which would be the same as the symbol rate).

At the other extreme, take a broader filter with an alpha of one, which is easier to implement. The occupied bandwidth will be

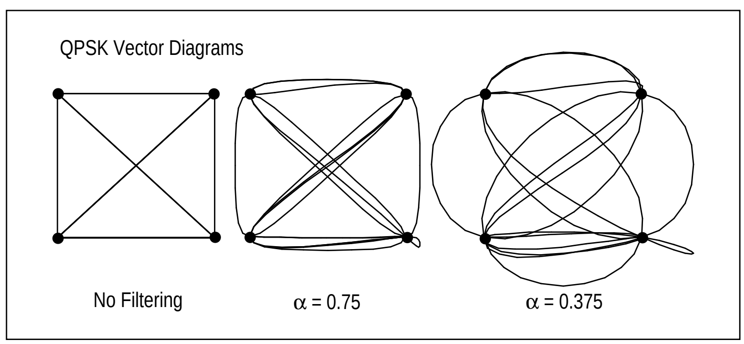
for $\alpha = 1$, *occupied bandwidth = symbol rate* X *(1 + 1) = 2* X *symbol rate.*

An alpha of one uses twice as much bandwidth as an alpha of zero. In practice, it is possible to implement an alpha below 0.2 and make good, compact, practical radios. Typical values range from 0.35 to 0.5, though some video systems use an alpha as low as 0.11. The corresponding term for a Gaussian filter is BT (bandwidth time product). Occupied bandwidth cannot be stated in terms of BT because a Gaussian filter's frequency response does not go identically to zero, as does a raised cosine. Common values for BT are 0.3 to 0.5.

4.5 Filter bandwidth effects

Different filter bandwidths show different effects. For example, look at a QPSK signal and examine how different values of alpha effect the vector diagram. If the radio has no transmitter filter as shown on the left of the graph, the transitions between states are instantaneous. No filtering means an alpha of infinity.

Figure 22.
Effect of Different Filter Bandwidth



Transmitting this signal would require infinite bandwidth. The center figure is an example of a signal at an alpha of 0.75. The figure on the right shows the signal at an alpha of 0.375. The filters with alphas of 0.75 and 0.375 smooth the transitions and narrow the frequency spectrum required.

Different filter alphas also affect transmitted power. In the case of the unfiltered signal, with an alpha of infinity, the maximum or peak power of the carrier is the same as the nominal power at the symbol states. No extra power is required due to the filtering.

Take an example of a $^{\pi}/_{4}$ DQPSK signal as used in NADC (IS-54). If an alpha of 1.0 is used, the transitions between the states are more gradual than for an alpha of infinity. Less power is needed to handle those transitions. Using an alpha of 0.5, the transmitted bandwidth decreases from 2 times the symbol rate to 1.5 times the symbol rate. This results in a 25% improvement in occupied bandwidth. The smaller alpha takes more peak power because of the overshoot in the filter's step response. This produces trajectories which loop beyond the outer limits of the constellation.

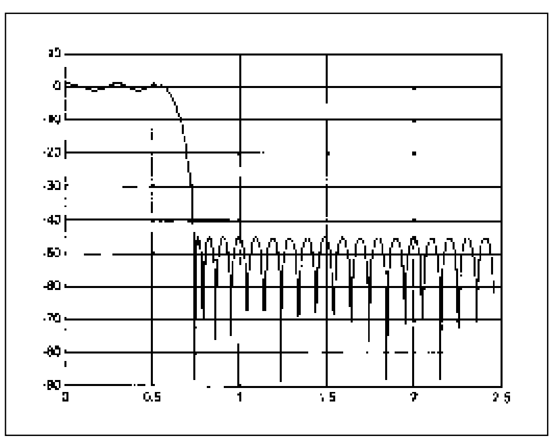
At an alpha of 0.2, about the minimum of most radios today, there is a need for significant excess power beyond that needed to transmit the symbol values themselves. A typical value of excess power needed at an alpha of 0.2 for QPSK with Nyquist filtering would be approximately 5dB. This is more than three times as much peak power because of the filter used to limit the occupied bandwidth.

These principles apply to QPSK, offset QPSK, DQPSK, and the varieties of QAM such as 16QAM, 32QAM, 64QAM, and 256QAM. Not all signals will behave in exactly the same way, and exceptions include FSK, MSK and any others with constant-envelope modulation. The power of these signals is not affected by the filter shape.

4.6 Chebyshev equiripple FIR (finite impulse respone) filter

A Chebyshev equiripple FIR (finite impulse response) filter is used for baseband filtering in IS-95 CDMA. With a channel spacing of 1.25 MHz and a symbol rate of 1.2288 MHz in IS-95 CDMA, it is vital to reduce leakage to adjacent RF channels. This is accomplished by using a filter with a very sharp shape factor using an alpha value of only 0.113. A FIR filter means that the filter's impulse response exists for only a finite number of samples. Equiripple means that there is a "rippled" magnitude frequency-respone envelope of equal maxima and minima in the pass- and stopbands. This FIR filter uses a much lower order than a Nyquist filter to implement the required shape factor. The IS-95 FIR filter does not have zero Inter Symbol Interference (ISI). However, ISI in CDMA is not as important as in other formats since the correlation of 64 chips at a time is used to make a symbol decision. This "coding gain" tends to average out the ISI and minimize its effect.

Figure 23. Chebyshev Equiripple FIR Filter



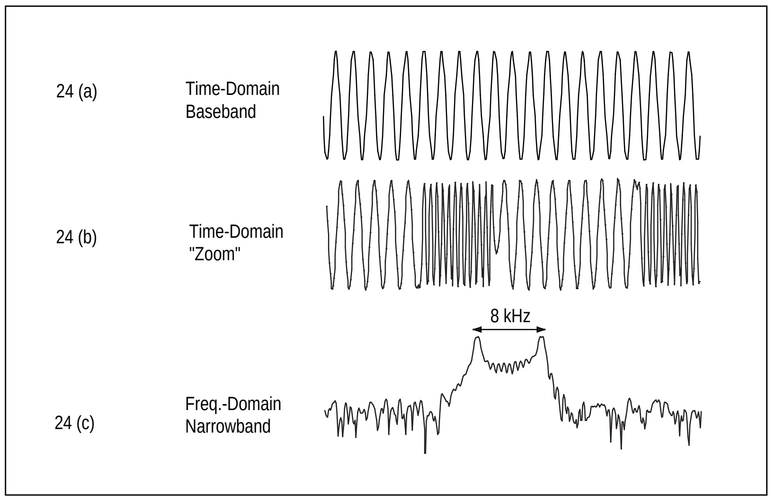
4.7 Competing goals of spectral efficiency and power consumption

As with any natural resource, it makes no sense to waste the RF spectrum by using channel bands that are too wide. Therefore narrower filters are used to reduce the occupied bandwidth of the transmission. Narrower filters with sufficient accuracy and repeatability are more difficult to build. Smaller values of alpha increase ISI because more symbols can contribute. This tightens the requirements on clock accuracy. These narrower filters also result in more overshoot and therefore more peak carrier power. The power amplifier must then accommodate the higher peak power without distortion. The bigger amplifier causes more heat and electrical interference to be produced since the RF current in the power amplifier will interfere with other circuits. Larger, heavier batteries will be required. The alternative is to have shorter talk time and smaller batteries. Constant envelope modulation, as used in GMSK, can use class-C amplifiers which are the most efficient. In summary, spectral efficiency is highly desirable, but there are penalties in cost, size, weight, complexity, talk time, and reliability.

5. Different ways of looking at a digitally-modulated signal time and frequency domain view

There are a number of different ways to view a signal. This simplified example is an RF pager signal at a center frequency of 930.004 MHz. This pager uses two-level FSK and the carrier shifts back and forth between two frequencies that are 8 kHz apart (930.000 MHz and 930.008 MHz). This frequency spacing is small in proportion to the center frequency of 930.004 MHz. This is shown in figure 24 (a). The difference in period between a signal at 930 MHz and one at 930 MHz plus 8 kHz is very small. Even with a high performance oscilloscope, using the latest in high-speed digital techniques, the change in period cannot be observed or measured.

Figure 24.
Time and Frequency Domain View



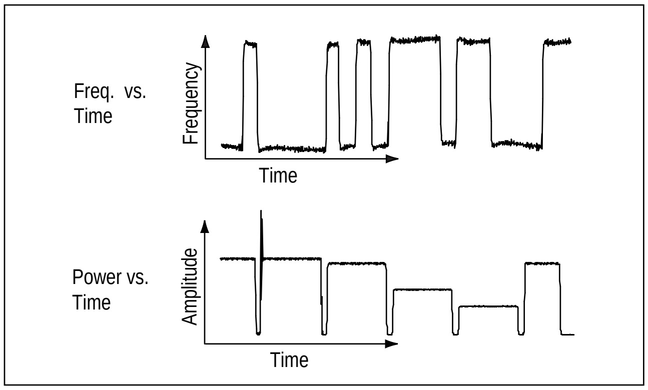
In a pager receiver the signals are first downconverted to an IF or baseband frequency. In this example, the 930.004 MHz FSK-modulated signal is mixed with another signal at 930.002 MHz. The FSK modulation causes the transmitted signal to switch between 930.000 MHz and 930.008 MHz. The result is a baseband signal that alternates between two frequencies, –2 kHZ and +6 kHz. The demodulated signal shifts between –2 kHz and +6 kHz. The difference can be easily detected.

This is sometimes referred to as “zoom” time or IF time. To be more specific, it is a band-converted signal at IF or baseband. IF time is important as it is how the signal looks in the IF portion of a receiver. This is how the IF of the radio detects the different bits that are present. The frequency domain representation is shown in figure 24 (c). Most pagers use a two-level, Frequency-Shift-Keying (FSK) scheme. FSK is used in this instance because it is less affected by multipath propagation, attenuation and interference, common in urban environments. It is possible to demodulate it even deep inside modern steel/concrete buildings, where attenuation, noise and interference would otherwise make reliable demodulation difficult.

5.1 Power and frequency view

There are many different ways of looking at a digitally-modulated signal. To examine how transmitters turn on and off, a power-versus-time measurement is very useful for examining the power level changes involved in pulsed or bursted carriers. For example, very fast power changes will result in frequency spreading or spectral regrowth. This is also known as frequency "splatter". Very slow power changes waste valuable transmit time, as the transmitter cannot send data when it is not fully on. Turning on too slowly can also cause high bit error rates at the beginning of the burst. In addition, peak and average power levels must be well understood, since asking for excessive power from an amplifier can lead to compression or clipping. These phenomena distort the modulated signal and usually lead to spectral regrowth as well.

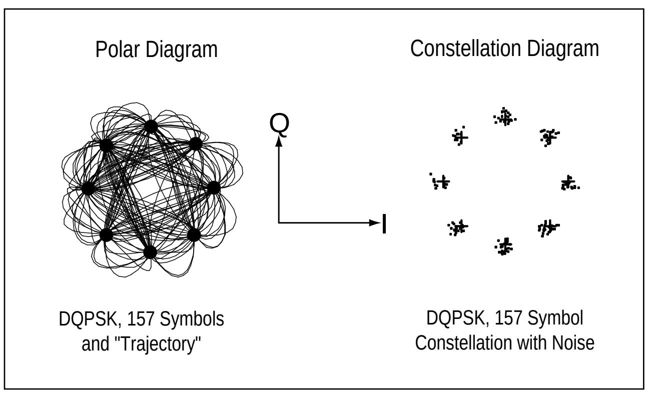
Figure 25.
Power and Frequency View



5.2 Constellation diagrams

As discussed, the rectangular *I/Q* diagram is a polar diagram of magnitude and phase. A two-dimensional diagram of the carrier magnitude and phase (a standard polar plot) can be represented differently by superimposing rectangular axes on the same data and interpreting the carrier in terms of in-phase (*I*) and quadrature-phase (*Q*) components. It would be possible to perform AM and PM on a carrier at the same time and send data this way; it is easier for circuit design and signal processing to generate and detect a rectangular, linear set of values (one set for *I* and an independent set for *Q*).

Figure 26.
Constellation Diagram



The example shown is a $\pi/4$ Differential Quadrature Phase Shift Keying ($\pi/4$ DQPSK) signal as described in the North American Digital Cellular (NADC) TDMA standard. This example is a 157-symbol DQPSK burst.

The polar diagram shows several symbols at a time. That is, it shows the instantaneous value of the carrier at any point on the continuous line between and including symbol times, represented as I/Q or magnitude/phase values.

The constellation diagram shows a repetitive "snapshot" of that same burst, with values shown only at the decision points. The constellation diagram displays phase errors, as well as amplitude errors, at the decision points. The transitions between the decision points affects transmitted bandwidth. This display shows the path the carrier is taking but does not explicitly show errors at the decision points. Constellation diagrams provide insight into varying power levels,the effects of filtering, and phenomena such as Inter-Symbol Interference.

The relationship between constellation points and bits per symbol is

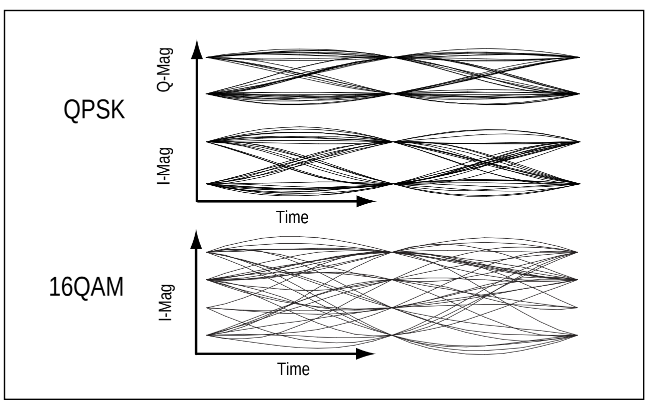
$$M=2^n \text{ where } M = \text{number of constellation points}$$
$$n = \text{bits/symbol}$$
$$\text{or } n = \log_2(M)$$

This holds when transitions are allowed from any constellation point to any other.

5.3 Eye diagrams

Another way to view a digitally modulated signal is with an eye diagram. Separate eye diagrams can be generated, one for the *I*-channel data and another for the *Q*-channel data. Eye diagrams display *I* and *Q* magnitude versus time in an infinite persistence mode, with retraces. The *I* and *Q* transitions are shown separately and an "eye" (or eyes) is formed at the symbol decision times. QPSK has four distinct I/Q states, one in each quadrant. There are only two levels for *I* and two levels for *Q*. This forms a single eye for each *I* and *Q*. Other schemes use more levels and create more nodes in time through which the traces pass. The lower example is a 16QAM signal which has four levels forming three distinct "eyes". The eye is open at each symbol. A "good" signal has wide open eyes with compact crossover points.

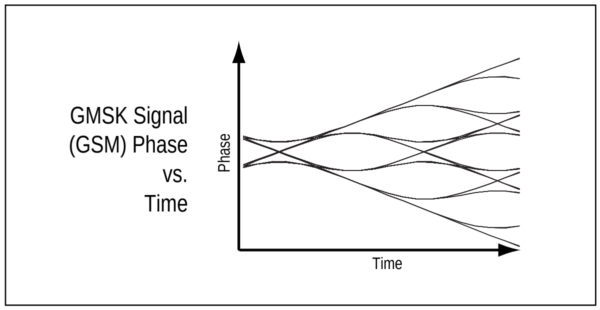
Figure 27.
***I* and *Q* Eye Diagrams**



5.4 Trellis diagrams

This figure is called a "trellis" diagram, because it resembles a garden trellis. The trellis diagram shows time on the X-axis and phase on the Y-axis. This allows the examination of the phase transitions with different symbols. In this case it is for a GSM system. If a long series of binary ones were sent, the result would be a series of positive phase transitions of, in the example of GSM, 90 degrees per symbol. If a long series of binary zeros were sent, there would be a constant declining phase of 90 degrees per symbol. Typically there would be intermediate transmissions with random data. When troubleshooting, trellis diagrams are useful in isolating missing transitions, missing codes, or a blind spot in the *I/Q* modulator or mapping algorithm.

Figure 28.
Trellis Diagram



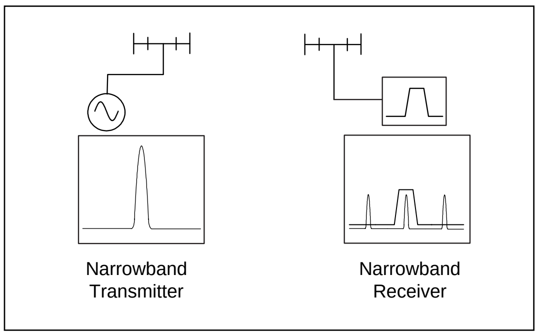
6. Sharing the channel

The RF spectrum is a finite resource and is shared between users using multiplexing (sometimes called channelization). Multiplexing is used to separate different users of the spectrum. This section covers multiplexing frequency, time, code, and geography. Most communications systems use a combination of these multiplexing methods.

6.1 Multiplexing - frequency

Frequency Division Multiple-Access (FDMA) splits the available frequency band into smaller fixed frequency channels. Each transmitter or receiver uses a separate frequency. This technique has been used since around 1900 and is still in use today. Transmitters are narrowband or frequency-limited. A narrowband transmitter is used along with a receiver that has a narrow-band filter so that it can demodulate the desired signal and reject unwanted signals, such as interfering signals from adjacent radios.

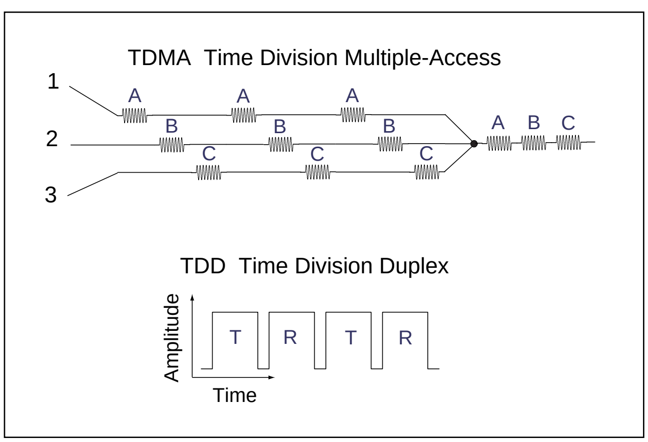
Figure 29. Multiplexing - Frequency



6.2 Multiplexing - time

Time-division multiplexing involves separating the transmitters in time so that they can share the same frequency. The simplest type is Time Division Duplex (TDD). This multiplexes the transmitter and receiver on the same frequency. TDD is used, for example, in a simple two-way radio where a button is pressed to talk and released to listen. This kind of time division duplex, however, is very slow. Modern digital radios like CT2 and DECT use Time Division Duplex but they multiplex hundreds of times per second. TDMA (Time Division Multiple Access) multiplexes several transmitters or receivers on the same frequency. TDMA is used in the GSM digital cellular system and also in the US NADC-TDMA system.

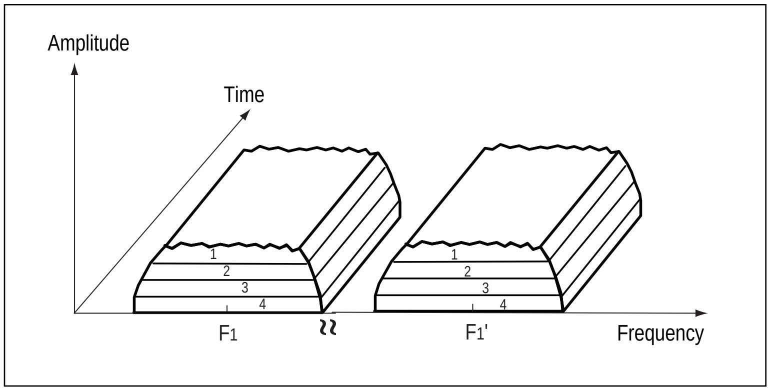
Figure 30. Multiplexing - Time



6.3 Multiplexing - code

CDMA is an access method where multiple users are permitted to transmit simultaneously on the same frequency. Frequency division multiplexing is still performed but the channel is 1.23 MHz wide. In the case of US CDMA telephones, an additional type of channelization is added, in the form of coding.

Figure 31. Multiplexing - Code

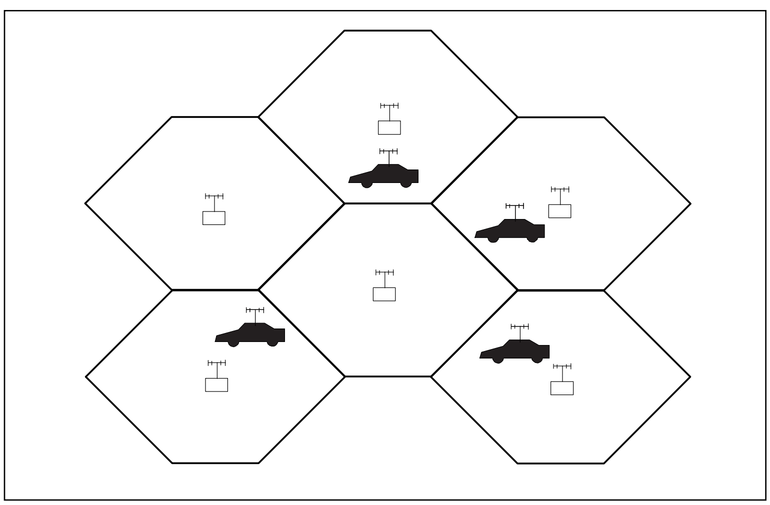


In CDMA systems, users timeshare a higher-rate digital channel by overlaying a higher-rate digital sequence on their transmission. A different sequence is assigned to each terminal so that the signals can be discerned from one another by correlating them with the overlaid sequence. This is based on codes that are shared between the base and mobile stations. Because of the choice of coding used, there is a limit of 64 code channels on the forward link. The reverse link has no practical limit to the number of codes available.

6.4 Multiplexing - geography

Another kind of multiplexing is geographical or cellular. If two transmitter/receiver pairs are far enough apart, they can operate on the same frequency and not interfere with each other. There are only a few kinds of systems that do not use some sort of geographic multiplexing. Clear-channel international broadcast stations, amateur stations, and some military low frequency radios are about the only systems that have no geographic boundaries and they broadcast around the world.

Figure 32. Multiplexing - Geography



6.5 Combining multiplexing modes

In most of these common communications systems, different forms of multiplexing are generally combined. For example, GSM uses FDMA, TDMA, FDD and geographic. DECT uses FDMA, TDD and geographic multiplexing. For a full listing see the table in section ten.

6.6 Penetration versus efficiency

Penetration means the ability of a signal to be used in environments where there is a lot of attenuation or noise or interference. One very common example is the use of pagers versus cellular phones. In many cases, pagers can receive signals even if the user is inside a metal building or a steel-reinforced concrete structure like a modern skyscraper. Most pagers use a two-level FSK signal where the frequency deviation is large and the modulation rate (symbol rate) is quite slow. This makes it easy for the receiver to detect and demodulate the signal since the frequency difference is large (the symbol locations are widely separated) and these different frequencies persist for a long time (a slow symbol rate).

However, the factors causing good pager signal penetration also cause inefficient information transmission. There are typically only two symbol locations. They are widely separated (approximately 8 kHz), and a small number of symbols (500 to 1200) are sent each second. Compare this with a cellular system such as GSM which sends 270,833 symbols each second. This is not a big problem for the pager since all it needs to receive is its unique address and perhaps a short ASCII text message.

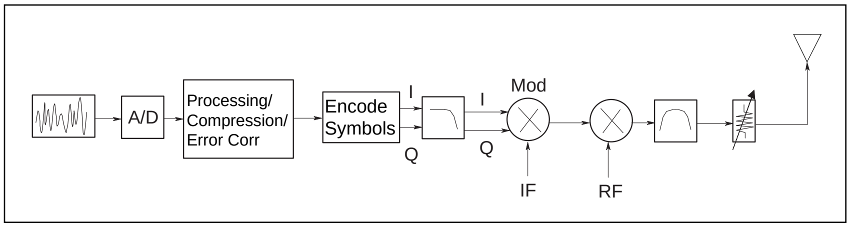
A cellular phone signal, however, must transmit live duplex voice. This requires a much higher bit rate and a much more efficient modulation technique. Cellular phones use more complex modulation formats (such as $\pi/4$ DQPSK and 0.3 GMSK) and faster symbol rates. Unfortunately, this greatly reduces penetration and one way to compensate is to use more power. More power brings in a host of other problems, as described previously.

7. How digital transmitters and receivers work

7.1 A digital communications transmitter

Here is a simplified block diagram of a digital communications transmitter. It begins and ends with an analog signal. The first step is to convert a continuous analog signal to a discrete digital bit stream. This is called digitization.

Figure 33.
A Digital Transmitter



The next step is to add voice coding for data compression. Then some channel coding is added. Channel coding encodes the data in such a way as to minimize the effects of noise and interference in the communications channel. Channel coding adds extra bits to the input data stream and removes redundant ones. Those extra bits are used for error correction or sometimes to send training sequences for identification or equalization. This can make synchronization (or finding the symbol clock) easier for the receiver. The symbol clock represents the frequency and exact timing of the transmission of the individual symbols. At the symbol clock transitions, the transmitted carrier is at the correct I/Q (or magnitude/phase) value to represent a specific symbol (a specific point in the constellation). Then the values (I/Q or magnitude/ phase) of the transmitted carrier are changed to represent another symbol. The interval between these two times is the symbol clock period. The reciprocal of this is the symbol clock frequency. The symbol clock phase is correct when the symbol clock is aligned with the optimum instant(s) to detect the symbols.

The next step in the transmitter is filtering. Filtering is essential for good bandwidth efficiency. Without filtering, signals would have very fast transitions between states and therefore very wide frequency spectra — much wider than is needed for the purpose of sending information. A single filter is shown for simplicity, but in reality there are two filters; one each for the *I* and *Q* channels. This creates a compact and spectrally efficient signal that can be placed on a carrier.

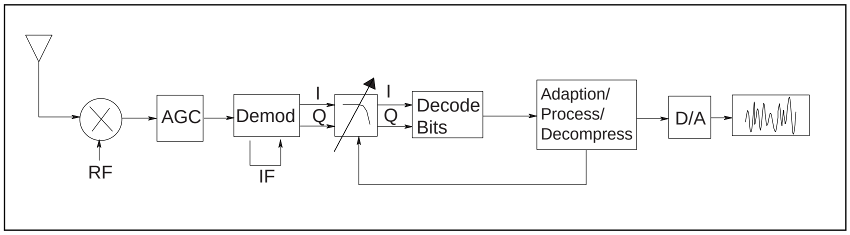
The output from the channel coder is then fed into the modulator. Since there are independent *I* and *Q* components in the radio, half of the information can be sent on *I* and the other half on *Q*. This is one reason digital radios work well with this type of digital signal. The *I* and *Q* components are separate.

The rest of the transmitter looks similar to a typical RF transmitter or microwave transmitter/receiver pair. The signal is converted up to a higher intermediate frequency (IF), and then further upconverted to a higher radio frequency (RF). Any undesirable signals that were produced by the upconversion are then filtered out.

7.2 A digital communications receiver

The receiver is similar to the transmitter but in reverse. It is more complex to design. The incoming (RF) signal is first downconverted to (IF) and demodulated. The ability to demodulate the signal is hampered by factors including atmospheric noise, competing signals, and multipath or fading.

Figure 34.
A Digital Receiver



Generally, demodulation involves the following stages:

1. carrier frequency recovery (carrier lock)
2. symbol clock recovery (symbol lock)
3. signal decomposition to *I* and *Q* components
4. determining *I* and *Q* values for each symbol ("slicing")
5. decoding and de-interleaving
6. expansion to original bit stream
7. digital-to-analog conversion, if required

In more and more systems, however, the signal starts out digital and stays digital. It is never analog in the sense of a continuous analog signal like audio. The main difference between the transmitter and receiver is the issue of carrier and clock (or symbol) recovery.

Both the symbol-clock frequency and phase (or timing) must be correct in the receiver in order to demodulate the bits successfully and recover the transmitted information. A symbol clock could be at the right frequency but at the wrong phase. If the symbol clock was aligned with the transitions between symbols rather than the symbols themselves, demodulation would be unsuccessful.

Symbol clocks are usually fixed in frequency and this frequency is accurately known by both the transmitter and receiver. The difficulty is to get them both aligned in phase or timing. There are a variety of techniques and most systems employ two or more. If the signal amplitude varies during modulation, a receiver can measure the variations. The transmitter can send a specific synchronization signal or a predetermined bit sequence such as *10101010101010* to "train" the receiver's clock. In systems with a pulsed carrier, the symbol clock can be aligned with the power turn-on of the carrier.

In the transmitter, it is known where the RF carrier and digital data clock are because they are being generated inside the transmitter itself. In the receiver there is not this luxury. The receiver can approximate where the carrier is but has no phase or timing symbol clock information. A difficult task in receiver design is to create carrier and symbol-clock recovery algorithms. That task can be made easier by the channel coding performed in the transmitter.

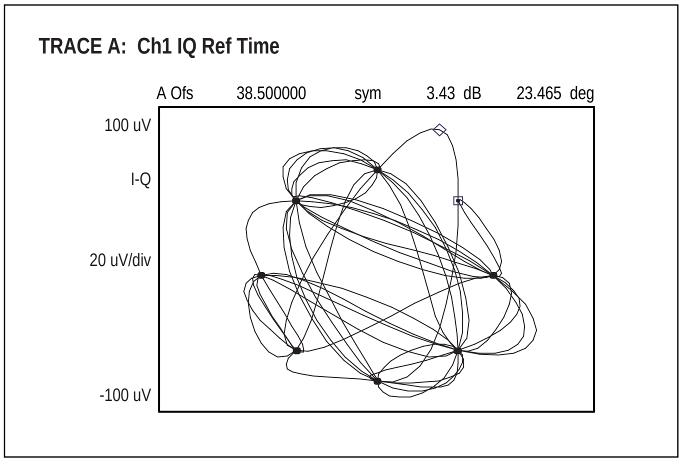
8. Measurements on digital RF communications systems

Complex tradeoffs in frequency, phase, timing, and modulation are made for interference-free, multiple-user communications systems. It is necessary to accurately measure parameters in digital RF communications systems to make the right tradeoffs. Measurements include analyzing the modulator and demodulator, characterizing the transmitted signal quality, locating causes of high Bit-Error-Rate and investigating new modulation types. Measurements on digital RF communications systems generally fall into four categories: power, frequency, timing, and modulation accuracy.

8.1 Power measurements

Power measurements include carrier power and associated measurements of gain of amplifiers and insertion loss of filters and attenuators. Signals used in digital modulation are noise-like. Band-power measurements (power integrated over a certain band of frequencies) or power spectral density (PSD) measurements are often made. PSD measurements normalize power to a certain bandwidth, usually 1 Hz.

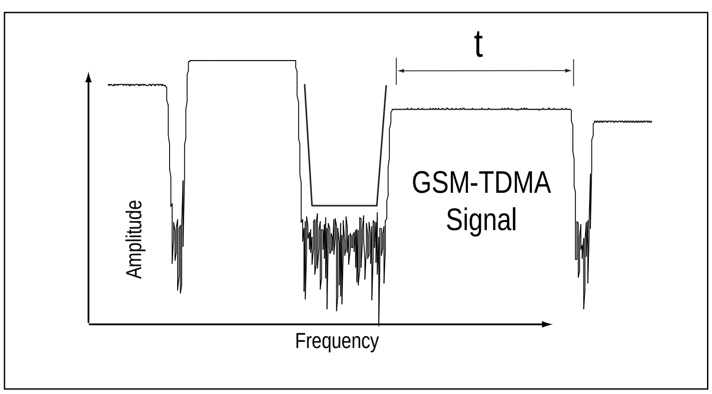
Figure 35. Power Measurement



8.1.1 Adjacent channel power

Adjacent channel power is a measure of interference created by one user that effects other users in nearby channels. This test quantifies the energy of a digitally-modulated RF signal that spills from the intended communication channel into an adjacent channel. The measurement result is the ratio (in dB) of the power measured in the adjacent channel to the total transmitted power. A similar measurement is alternate channel power which looks at the same ratio two channels away from the intended communication channel.

Figure 36. Power and Timing Measurements

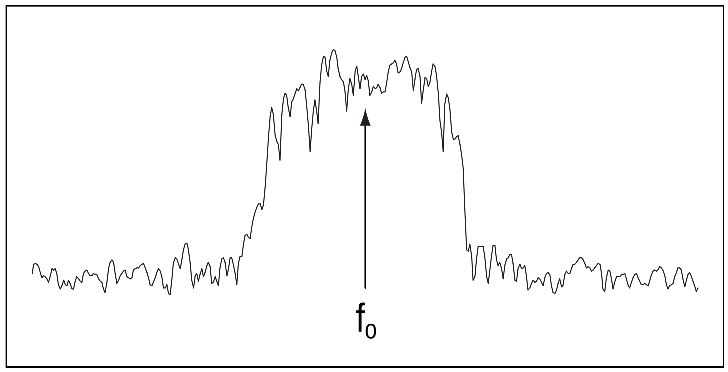


For pulsed systems (such as TDMA), power measurements have a time component and may have a frequency component, also. Burst power profile (power versus time) or turn-on and turn-off times may be measured. Another measurement is average power when the carrier is on or averaged over many on/off cycles.

8.2 Frequency measurements

Frequency measurements are often more complex in digital systems since factors other than pure tones must be considered. Occupied bandwidth is an important measurement. It ensures that operators are staying within the bandwidth that they have been allocated. Adjacent channel power is also used to detect the effects one user has on other users in nearby channels.

Figure 37. Frequency Measurements



8.2.1 Occupied bandwidth

Occupied bandwidth (BW) is a measure of how much frequency spectrum is covered by the signal in question. The units are in Hz, and measurement of occupied BW generally implies a power percentage or ratio. Typically, a portion of the total power in a signal to be measured is specified. A common percentage used is 99%. A measurement of power versus frequency (such as integrated band power) is used to add up the power to reach the specified percentage. For example, one would say “99% of the power in this signal is contained in a bandwidth of 30 kHz.” One could also say “The occupied bandwidth of this signal is 30 kHz” if the desired power ratio of 99% was known.

Typical occupied bandwidth numbers vary widely, depending on symbol rate and filtering. The figure is about 30 kHz for the NADC $\pi/4$ DQPSK signal and about 350 kHz for a GSM 0.3 GMSK signal. For digital video signals occupied bandwidth is typically 6 to 8 MHz.

Simple frequency-counter-measurement techniques are often not accurate or sufficient to measure center frequency. A carrier “centroid” can be calculated which is the center of the distribution of frequency versus PSD for a modulated signal.

8.3 Timing measurements

Timing measurements are made most often in pulsed or burst systems. Measurements include pulse repetition intervals, on-time, off-time, duty cycle, and time between bit errors. Turn-on and turn-off times also involve power measurements.

8.4 Modulation accuracy

Modulation accuracy measurements involve measuring how close either the constellation states or the signal trajectory is relative to a reference (ideal) signal trajectory. The received signal is demodulated and compared with a reference signal. The main signal is subtracted and what is left is the difference or residual. Modulation accuracy is a residual measurement.

Modulation accuracy measurements usually involve precision demodulation of a signal and comparison of this demodulated signal with a (mathematically-generated) ideal or "reference" signal. The difference between the two is the modulation error, and it can be expressed in a variety of ways including Error Vector Magnitude (EVM), magnitude error, phase error, *I*-error and *Q*-error. The reference signal is subtracted from the demodulated signal, leaving a residual error signal. Residual measurements such as this are very powerful for troubleshooting. Once the reference signal has been subtracted, it is easier to see small errors that may have been swamped or obscured by the modulation itself. The error signal itself can be examined in many ways; in the time domain or (since it is a vector quantity) in terms of its I/Q or magnitude/phase components. A frequency transformation can also be performed and the spectral composition of the error signal alone can be viewed.

8.5 Understanding Error Vector Magnitude

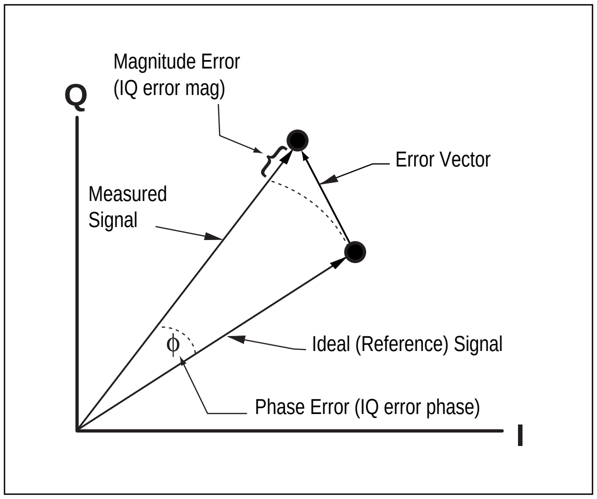
Recall first the basics of vector modulation: Digital bits are transferred onto an RF carrier by varying the carrier's magnitude and phase. At each symbol-clock transition, the carrier occupies any one of several unique locations on the *I* versus *Q* plane. Each location encodes a specific data symbol, which consists of one or more data bits. A constellation diagram shows the valid locations (i.e., the magnitude and phase relative to the carrier) for all permitted symbols of which there must be 2^n, given *n* bits transmitted per symbol. To demodulate the incoming data, the exact magnitude and phase of the received signal for each clock transition must be accurately determined.

The layout of the constellation diagram and its ideal symbol locations is determined generically by the modulation format chosen (BPSK, 16QAM, $\pi/4$ DQPSK, etc.). The trajectory taken by the signal from one symbol location to another is a function of the specific system implementation, but is readily calculated nonetheless.

At any moment, the signal's magnitude and phase can be measured. These values define the actual or "measured" phasor. At the same time, a corresponding ideal or "reference" phasor can be calculated, given knowledge of the transmitted data stream, the symbol-clock timing, baseband filtering parameters, etc. The differences between these two phasors form the basis for the EVM measurements.

Figure 38 defines EVM and several related terms. As shown, EVM is the scalar distance between the two phasor end points, i.e. it is the magnitude of the difference vector. Expressed another way, it is the residual noise and distortion remaining after an ideal version of the signal has been stripped away.

Figure 38.
EVM and Related Quantities



In the NADC-TDMA (IS-54) standard, EVM is defined as a percentage of the signal voltage at the symbols. In the $\pi/4$ DQPSK modulation format, these symbols all have the same voltage level, though this is not true of all formats. IS-54 is currently the only standard that explicitly defines EVM, so EVM could be defined differently for other modulation formats.

In a format such as 64QAM, for example, the symbols represent a variety of voltage levels. EVM could be defined by the average voltage level of all the symbols (a value close to the average signal level) or by the voltage of the outermost (highest voltage) four symbols. While the error vector has a phase value associated with it, this angle generally turns out to be random because it is a function of both the error itself (which may or may not be random) and the position of the data symbol on the constellation (which, for all practical purposes, is random). A more useful angle is measured between the actual and ideal phasors (*I/Q* phase error), which contains information useful in troubleshooting signal problems. Likewise, *I-Q* magnitude error shows the magnitude difference between the actual and ideal signals. EVM, as specified in the standard, is the root-mean-square (RMS) value of the error values at the instant of the symbol-clock transition. Trajectory errors between symbols are ignored.

8.6 Troubleshooting with error vector measurements

Measurements of error vector magnitude and related quantities can, when properly applied, provide much insight into the quality of a digitally modulated signal. They can also pinpoint the causes for any problems uncovered by identifying exactly the type of degradation present in a signal and even help identify their sources. For more detail on using error-vector-magnitude measurements to analyze and troubleshoot vector-modulated signals, see product note 89400-14. The Hewlett-Packard literature number is 5965-2898E.

EVM measurements are growing rapidly in acceptance, having already been written into such important system standards as NADC and PHS, and they are poised to appear in several upcoming standards including those for digital video transmission.

8.7 Magnitude versus phase error

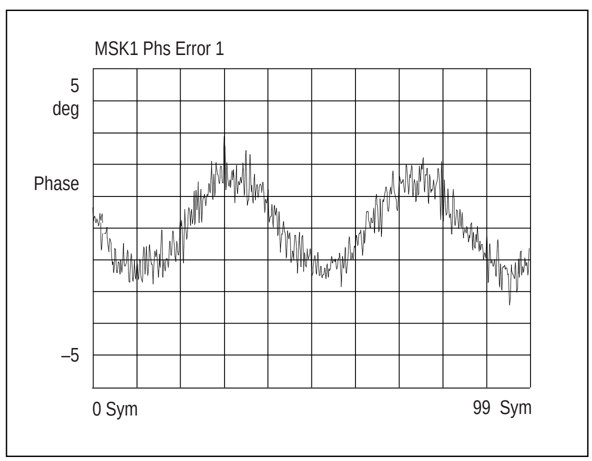
Different error mechanisms affect signals in different ways: in magnitude only, phase only, or both simultaneously. Knowing the relative amounts of each type of error can quickly confirm or rule out certain types of problems. Thus, the first diagnostic step is to resolve EVM into its magnitude and phase error components (see figure 38) and compare their relative sizes.

When the average phase error (in degrees) is substantially larger than the average magnitude error (in percent), some sort of unwanted phase modulation is the dominant error mode. This could be caused by noise, spurious or cross-coupling problems in the frequency reference, phase-locked loops, or other frequency-generating stages. Residual AM is evidenced by magnitude errors that are significantly larger than the phase angle errors.

8.8 I/Q phase error versus time

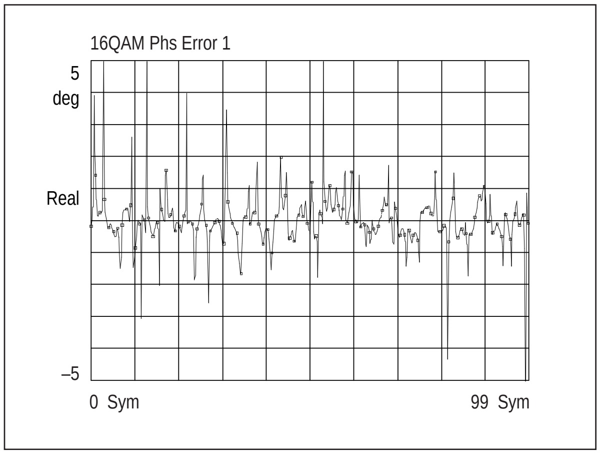
Phase error is the instantaneous angle difference between the measured signal and the ideal reference signal. When viewed as a function of time (or symbol), it shows the modulating waveform of any residual or interfering PM signal. Sinewaves or other regular waveforms indicate an interfering signal. Uniform noise is a sign of some form of phase noise (random jitter, residual PM/FM, etc.).

Figure 39. Incidental (inband) PM sinewave is clearly visible even at only three degrees peak-to-peak.



A perfect signal will have a uniform constellation that is perfectly symmetric about the origin. I/Q imbalance is indicated when the constellation is not "square", i.e. when the Q-axis height does not equal the I-axis width. Quadrature error is seen in any "tilt" to the constellation. Quadrature error is caused when the phase relationship between the I and Q vectors is not exactly 90 degrees.

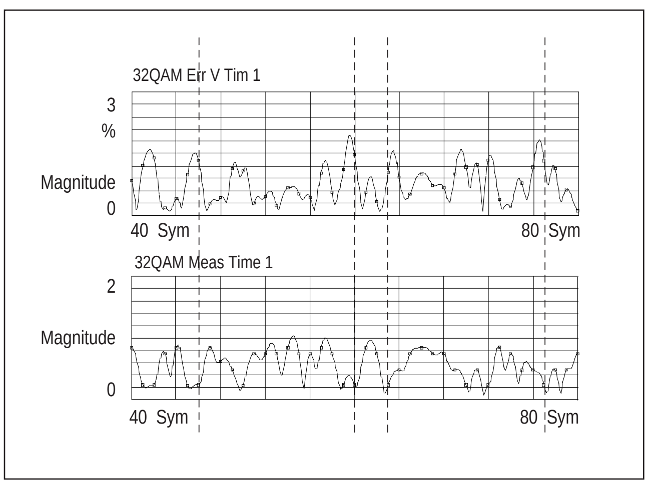
Figure 40.
Phase noise appears random in the time domain.



8.9 Error Vector Magnitude versus time

EVM is the difference between the input signal and the internally-generated ideal reference. When viewed as a function of symbol or time, errors may be correlated to specific points on the input waveform, such as peaks or zero crossings. EVM is a scalar (magnitude-only) value. Error peaks occurring with signal peaks indicate compression or clipping. Error peaks that correlate to signal minima suggest zero-crossing nonlinearities.

Figure 41.
EVM peaks on this signal (upper trace) occur every time the signal magnitude (lower trace) approaches zero. This is probably a zero-crossing error in an amplification stage.



An example of zero-crossing nonlinearities is in a push-pull amplifier, where the positive and negative halves of the signal are handled by separate transistors. It can be quite a challenge (especially in high-power amplifiers) to precisely bias and stabilize the amplifiers such that one set is turning off exactly as the other set is turning on, with no discontinuities. The critical moment is zero crossing, a well-known effect in analog design. It is also known as zero-crossing errors, distortion, or nonlinearities.

8.10 Error spectrum (EVM versus frequency)

The error spectrum is calculated from the Fast Fourier Transform (FFT) of the EVM waveform and results in a frequency-domain display that can show details not visible in the time domain. In most digital systems, nonuniform noise distribution or discrete signal peaks indicate the presence of externally-coupled interference.

Figure 42. Interference from adjacent (lower) channel causes uneven EVM spectral distribution.

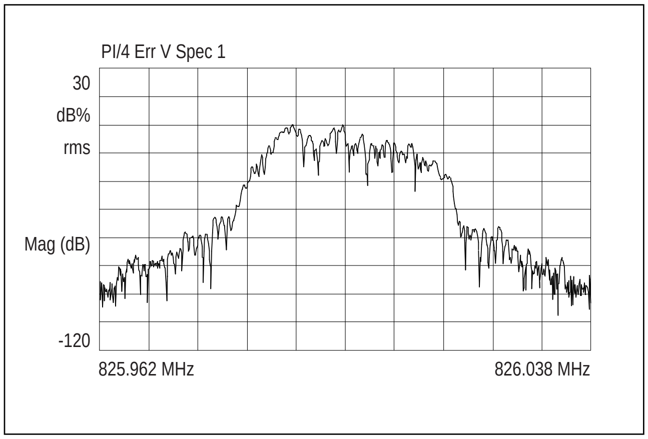
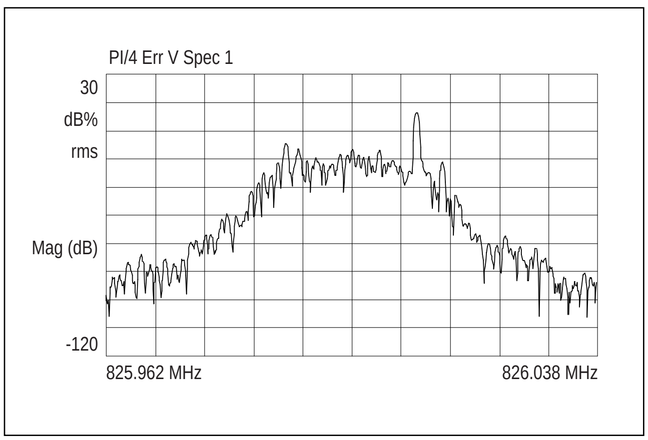


Figure 43. Switching-power-supply interference appears as EVM spur, offset from carrier by 10kHz.



For more detail on EVM measurements, see product note 89400-14 “Using Error-Vector-Magnitude Measurements to Analyze and Troubleshoot Vector-Modulated Signals.” The Hewlett-Packard literature number is 5965-2898E.

9. Summary

Communication system design requires the simultaneous conservation of bandwidth, power, and cost. In the past, it was possible to make a radio low cost by sacrificing parameters such as power and bandwidth efficiency.

This application note has presented the building blocks of any communications system. With these concepts, you will be able to understand how communications systems work, and make more informed decisions regarding the tradeoffs required to optimize your system.

10. Overview of communications systems

	GSM900	NADC	PDC	CDMA
Geography	Europe	North America	Japan	North America, Korea, Japan
Introduction	1992	1992	1993-1994	1995-1997
Frequency Range	935-960 MHz down 890-915 MHz up EGSM 925-960 MHz 880-915 MHz	869-894 MHz down 824-849 MHz up	810-826 MHz down 940-956 MHz up 1777-1801 MHz down 1429-1453 MHz up	824-849 MHz (US) 869-894 MHz (US) 832-834, 843-846, 860-870 MHz (Japan) 887-889, 898-901, 915-925 MHz (Japan)
Data Structure	TDMA	TDMA	TDMA	CDMA
Channel per Frequency	8-16	3-6	3-6	32-64 (Dyn. adapt)
Modulation	0.3 GMSK (1 bit/symbol)	$\pi/4$ DQPSK (2 bits/symbol)	$\pi/4$ DQPSK (2 bits/symbol)	Mobile: QPSK Base: OQPSK (1 bit/symbol)
Speech CODEC	RELP-LTP 13 Kbits/s	VSELP 8 Kbits/s EFR	VSELP 8 Kbits/s	8 Kbits/s var rate CELP 13 kbit/s var rate CELP
Mobile Output Power	3.7mW to 20W	2.2mW to 6W	.3W to 3W	10nW to 1W
Modulation Data Rate	270.833 Kbits/s (1 bit/symbol)	48.6 Kbits/s (2 bits/symbol)	42 Kbits/s (2 bits/symbol)	9600/14,400 bps data; 1.2288 Mb/s spreading
Filter	0.3 Gaussian	SQRT raised cosine $\alpha = .35$	SQRT raised cosine $\alpha = .50$	Chebychev low pass (FIR)
Channel Spacing	200 kHz	30 kHz	50 kHz 25 kHz interleave	1.23 MHz
Number of Channels	124 frequency ch. w/8 timeslots per ch. (1000)	832 frequency ch. w/3 users per ch. (2496)	1600 frequency ch. w/3 users per ch. (4800)	19-20 frequencies
Est # of Subscribers by year 2000	15-20 million	35-40 million (8.9 million 9/92)	5 million	
Source	GSM Standard	IS-54	RCR Spec Std 27B	IS-95 spec
Service	Public Cellular	Public Cellular	Public Cellular	Public Cellular

10. Overview of communications systems

	DCS1800	PHS	DECT	TETRA Trans European Trunked Radio
Geography	Europe	Japan/China	Europe/China	Europe
Introduction	1993	1993 Private office 1995 Public	1993	1995
Frequency Range	1.7-1.9 GHz 1710-1785 MHz down 1805-1880 MHz up	1895-1918 MHz up/down 1.9, 1.93 GHz (China)	1.897-1.913 GHz 1.9, 1.93 GHz (China)	450 MHz < 1 GHz
Data Structure	TDMA	TDMA/TDD	TDMA/TDD	TDMA
Channel per Frequency	8-16	4-8	12	4
Modulation	0.3 GMSK (1 bit/symbol)	π/4 DQPSK (2 bits/symbol)	0.5 GFSK ±202-403 kHz dev (1 bit/symbol)	π/4 DQPSK
Speech CODEC	RELP-LTP 13 Kbits/s	ADPCM 32 Kbits/s	ADPCM 32 Kbits/s	Includes channel & speech coding 7.2 Kbits/s
Mobile Output Power	250mW to 2W	10mW	250mW	
Modulation Data Rate	270.833 Kbits/s	384 Kbits/s	1.152 Mbit/s	19.2 Kb/s
Filter	0.3 Gaussian	SQRT raised cosine α = .50	0.5 Gaussian	α = 0.4 SQRT raised cosine
Channel Spacing	200 kHz	300 kHz	1.728 MHz	25 kHz
Number of Channels	3000-6000		10 carrier frequencies w/12 users per frequency (120)	
Est # of Subscribers by year 2000	4-13 million	6.5-13 million		
Source	prI-ETS 30 176 prETS 300 175-2	RCR spec Std 28 China-First News Release 8/15/96	CI Spec., Part 1, Rev 05.2e China-First News Release 8/15/96	Mobile Europe Magazine 1/92
Service	Personal Communications	Cordless Telephone Personal Communications	Wireless PBX	Trunked system Adj. ch. sel > 60 dB

11. Glossary of terms

ACP	Adjacent Channel Power
ADPCM	Adaptive Digital Pulse Code Modulation
AM	Amplitude Modulation
AMPS	Advanced Mobile Phone System
B-CDMA	Broadband Code Division Multiple Access
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
BFSK	Binary Frequency Shift Keying
BW	Bandwidth
CDMA	Code Division Multiple Access
CDPD	Cellular Digital Packet Data
COFDM	Coded Orthogonal Frequency Division Multiplexing
CRC	Cyclic Redundancy Check
CT2	Cordless Telephone - 2
DAB	Digital Audio Broadcast
DCS 1800	Digital Communication System - 1800 MHz
DECT	Digital Enhanced Cordless Telephone
DMCA	Digital MultiChannel Access, similar to iDEN
DQPSK	Differential Quadrature Phase Shift Keying
DSP	Digital Signal Processing
DVB-C	Digital Video Broadcast - Cable
DVB-S	Digital Video Broadcast - Satellite
DVB-T	Digital Video Broadcast - Terrestrial
EGSM	Extended Frequency GSM
ERMES	European Radio Message System
ETSI	European Telecommunications Standards Institute
EVM	Error Vector Magnitude
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
FER	Frame Error Rate
FFSK	Fast Frequency Shift Keying
FFT	Fast Fourier Transform
FLEX	4-level FSK-based paging standard developed by Motorola
FM	Frequency Modulation
FSK	Frequency Shift Keying
GFSK	Gaussian Frequency Shift Keying
Globalstar	Satellite system using 48 low-earth orbiting satellites
GSM	Global System for Mobile Communication
GMSK	Gaussian Minimum Shift Keying
HDTV	High Definition Television
iDEN	integrated Dispatch Enhanced Network (Motorola designed system for dispatch, cellular and conference calling)

11. Glossary of terms (cont'd)

IF	Intermediate Frequency
I/Q	In phase / Quadrature
Iridium	Motorola voice/data 66-satellite system worldwide
ISI	Intersymbol Interference
IS-54	Interim Standard for US Digital Cellular (NADC)
IS-95	Interim Standard for US Code Division Multiple Access
IS-136	Interim Standard for NADC with Digital Control Channels
LMDS	Local Multipoint Distribution System
MFSK	Minimum Frequency Shift Keying
MMDS	Multichannel Multipoint Distribution System
MPSK	Minimum Phase Shift Keying
MSK	Minimum Shift Keying
NADC	North American Digital Cellular system
OFDM	Orthogonal Frequency Division Multiplexing
OQPSK	Offset Quadrature Phase Shift Keying
PACS	Personal Access Communications Service
PCS	Personal Communications System
PCM	Pulse Code Modulation
PDC	Pacific Digital Cellular System (formerly JDC)
PHS	Personal Handyphone System (formerly PHP)
PRBS	Pseudo-Random Bit Sequence
PSD	Power Spectral Density
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RAM	Wireless data network
RF	Radio Frequency
RMS	Root Mean Square
SQRT	Square Root
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TETRA	Trans European Trunked Radio
TFTS	Terrestrial Flight Telephone System
VSB	Vestigal Side Band
WLL	Wireless Local Loop

For more information about Hewlett-Packard test and measurement products, applications, services, and for a current sales office listing, visit our web site, http://www.hp.com/go/tmdir. You can also contact one of the following centers and ask for a test and measurement sales representative.

United States:
Hewlett-Packard Company
Test and Measurement Call Center
P.O. Box 4026
Englewood, CO 80155-4026
1 800 452 4844

Canada:
Hewlett-Packard Canada Ltd.
5150 Spectrum Way
Mississauga, Ontario L4W 5G1
(905) 206 4725

Europe:
Hewlett-Packard
European Marketing Centre
P.O. Box 999
1180 AZ Amstelveen
The Netherlands
(31 20) 547 9900

Japan:
Hewlett-Packard Japan Ltd.
Measurement Assistance Center
9-1, Takakura-Cho, Hachioji-Shi,
Tokyo 192, Japan
Tel: (81) 426-56-7832
Fax: (81) 426-56-7840

Latin America:
Hewlett-Packard
Latin American Region Headquarters
5200 Blue Lagoon Drive, 9th Floor
Miami, Florida 33126, U.S.A.
(305) 267 4245/4220

Australia/New Zealand:
Hewlett-Packard Australia Ltd.
31-41 Joseph Street
Blackburn, Victoria 3130, Australia
1 800 629 485

Asia Pacific:
Hewlett-Packard Asia Pacific Ltd.
17-21/F Shell Tower, Times Square,
1 Matheson Street, Causeway Bay,
Hong Kong
Tel: (852) 2599 7777
Fax: (852) 2506 9285

Data Subject to Change

Printed in U.S.A. 7/97
5965-7160E