# Introduction

bandpass digital modulation

This paper introduces the concepts of digital modulation used in many communications systems today. Emphasis is places 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 a 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 radio, 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 if them.

On the other hand, designers of hand-held cellular phones put 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, 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 competition for new ones. These are the tradeoffs that must be considered in digital RF communications design.

This paper addresses

* 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 of 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. It is essential to understand each of these building blocks to understand how communications systems, present or future, works.

# Digital Modulation

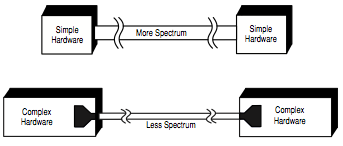
The move to digital modulation provides more information capacity, compatibility with digital data services, higher data security, better quality communications, and quicker system availability. Developers of communications systems face these constraints:

* 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.

## Trading off simplicity and bandwidth

There is a fundamental tradeoff in communications systems. Simple hardware can be used in transmitters and receivers to communicate information. However, this uses a lost 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.

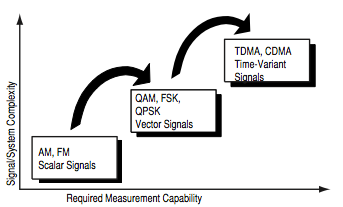


The fundamental tradeoff

# Industry trends

Over the past few years a major transition has occurred from simple analog Amplitude Modulation (AM) and Frequency/Phase Modulation (AM/PM) top new digital modulation including

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



Trends in the industry

Another layer of complexity in many new systems is multiplexing. Two principle 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.

# Using *I/Q* Modulation to Convey Information

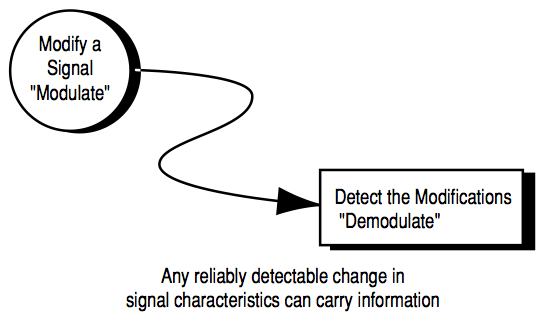
## Transmitting information

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 reliability detectable change in signal characteristics can carry information.
3. At the receiver the signal modifications or changes are detected and demodulated.

## 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.

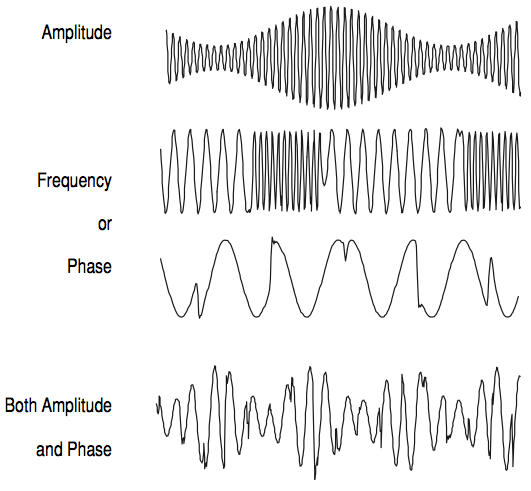


Transmitting Information (Analog or Digital)

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.

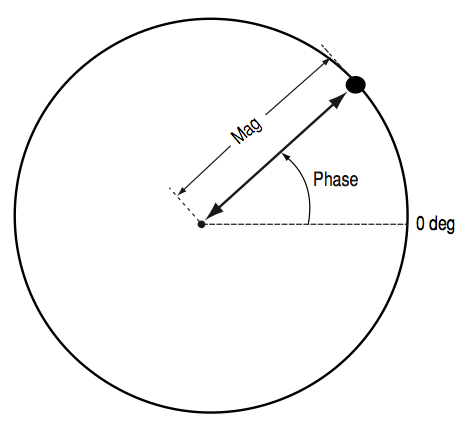
Amplitude and phase can be modulated simultaneously and separated, but this is difficult to generate, and especially difficult to detect. Instead, in practical systems the signal is separated into another set of independent components: *I* (In-phase) and *Q* (Quadrature). These components are orthogonal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



Signal Characteristics to Modify

## 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 communications systems. The magnitude is either an absolute or relative value. Both are used in digital communications 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).

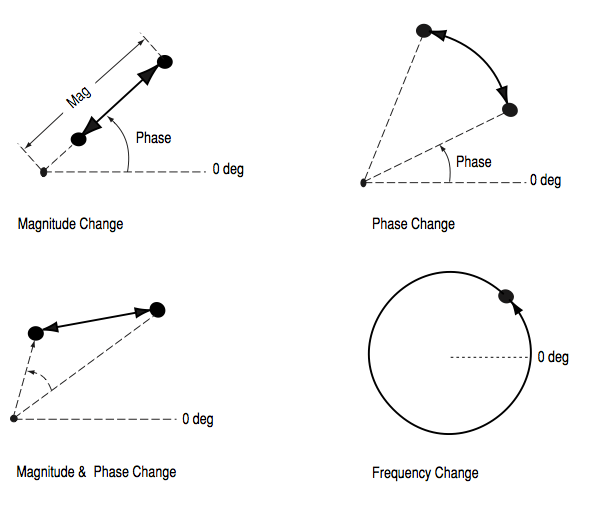


Polar Display – Magnitude and Phase Represented Together

## Signal changes or modifications in polar form

Figure 6 shows different forms of modulation in polar form. Magnitude is represented as the distance form the center and phase is represented as the angle.

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.

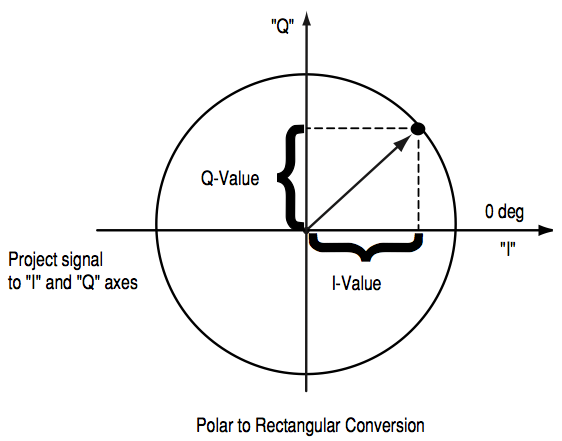


Signal changes or modifications

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.

## *I*/*Q* formats

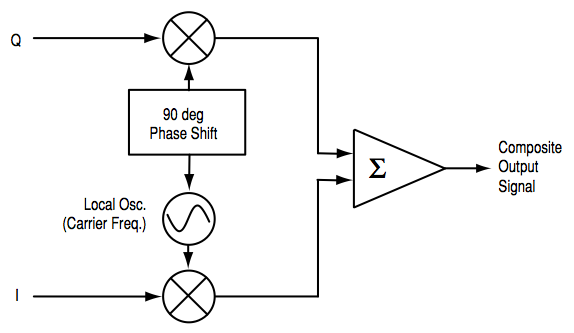
In digital communications, modulation is often expressed in terms of *I* and *Q*. This is a rectangular representation of the polar diagram. On a polar diagram, the I-axis lies on the zero degree phase reference, and the Q axis is rotated by 90 degrees. The signal vector’s projection onto the *I* axis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



“I-Q” Format

## *I* and *Q* in a Radio Transmitter

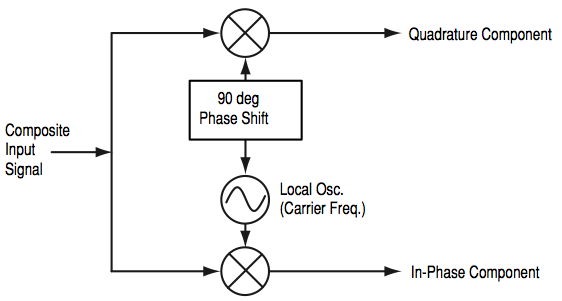
*I*/*Q* diagrams are particularly useful because they mirror the way most digital communications signals are created using an *I*/*Q* modulator. In the transmitter, I and Q signals are mixed with the same local oscillator (LO). A 90-degree phase shifter is placed in one of the LO paths. Signals that are separated by 90 degrees are also known as 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 component parts.



*I* and *Q* in a Radio Transmitter

*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 missed 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 process in a digital radio. A local oscillator, phase shifter, and two mixers can perform the conversion accurately and efficiently.



I and Q in a Radio Receiver

## Why use *I* and *Q*

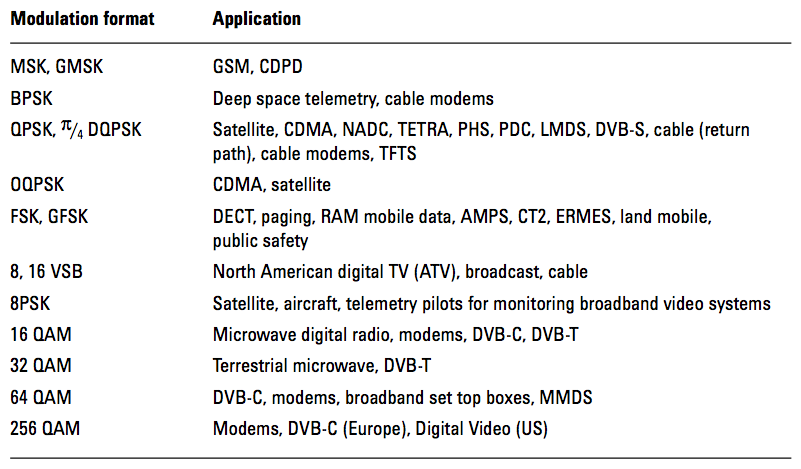
Digital modulation is easy to accomplish with *I*/*Q* modulators. Most digital modulation maps the data to a number of discrete points on the *I*/*Q* plane. These are known as constellation points. As the signal moves from one point to another, simultaneous amplitude and phase modulation usually results. 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 principle, 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.

# Digital Modulation Types and Relative Efficiencies

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

## Applications

The table below covers the applications for different modulation formats in both audio and video wireless communications.



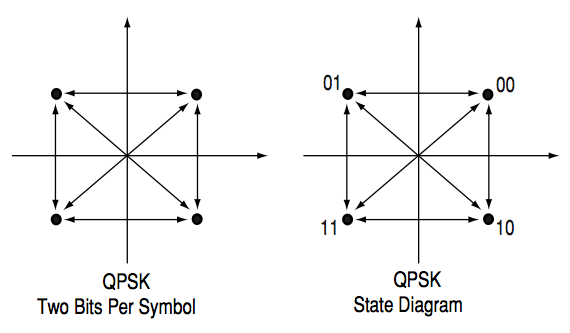
Although this paper focuses on wireless communications, video applications have also been included in the table for completeness and because of their similarity to other wireless communications.

### 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 a communications channel needed depends on the symbol rate, not on the bit rate.

Bit rate is the frequency of a system bit stream. 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 ignore the extra bits required overhead including synchronization, error correction, and so forth.

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 associated with binary transmission. This is a common mapping, but there are others. Any mapping can be used.

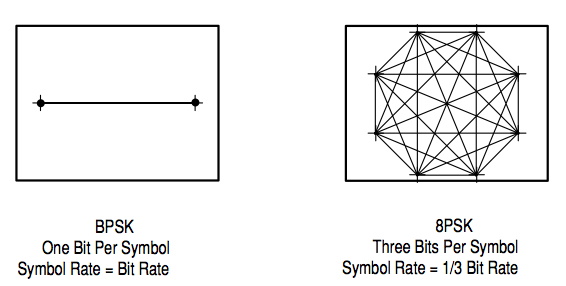


Bit rate and symbol rate

The symbol rate is the bit rate divided by the number of bits that can be transmitted with each symbol. If one bit were 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 were 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.

### 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 the eight values at any symbol time. Since 23 = 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.



Spectrum requirements

### 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.

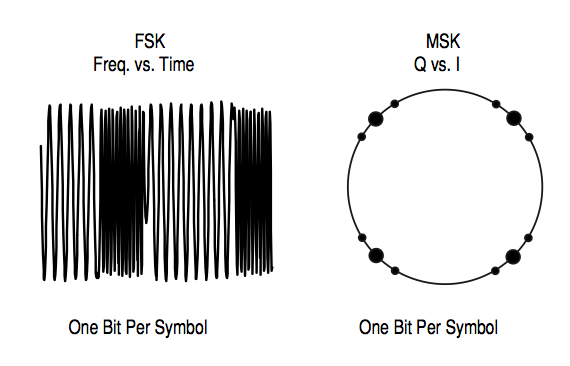
## Phase Shift Keying

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 tow 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.

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 shift 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 22 = 4. It is therefore a more bandwidth-efficient type of modulation than BPSK, potentially twice as efficient.

## 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 rate 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 ± π/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.



Frequency Shift Keying

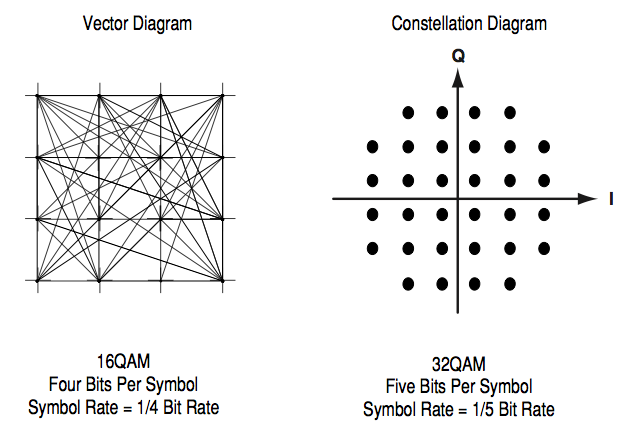
FSK and MSK produce constant envelope carrier signals, which have not amplitude variations. This is a desirable characteristic for improving the power efficiency of transmitters. Amplitude variations can be exercise nonlinearities in an amplifier’s amplitude-transfer function, generating spectral regrowth, 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).

# 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.

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 form one state to any other state at every symbol time. Since 24 = 16, 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.



Quadrature Amplitude Modulation

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 25 = 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 28 = 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 compare 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 Ksamples-per-second sending 1 bit per symbol. A system using 256QAM sends eight bits per symbol so the 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 become 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 as symbol as being n 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.

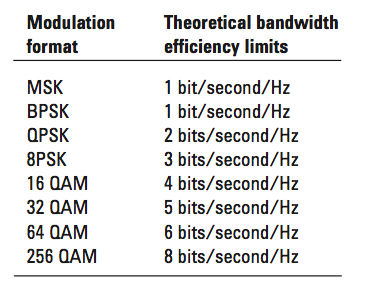
## Theoretical bandwidth efficiency limits

Bandwidth efficiency describes how efficiently the allocated bandwidth is utilized or the ability of a modulation scheme to accommodate data, within a limited bandwidth. The table below 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.

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 are used in NADC, PDC, and PHS, give the best spectral efficiency
* Restrict the types of transitions.



*Effects of Going Through the Origin*

## 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 of 1.6 bits per second per Hz. It is a π/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 Mb/s over a very wide bandwidth of 52.5 MHz. The spectral efficiency is 2.7 bits 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

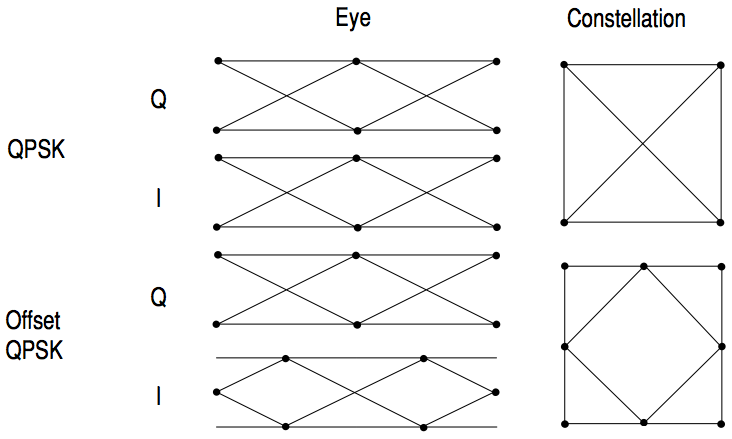
The modulation types outlined in Sections 3.2 to 3.4 form the building blocks for many systems. There are three main variations on these basic building blocks that are used in communications systems:

* *I*/*Q* offset modulation
* Differential modulation
* Constant envelope modulation

## *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.

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.



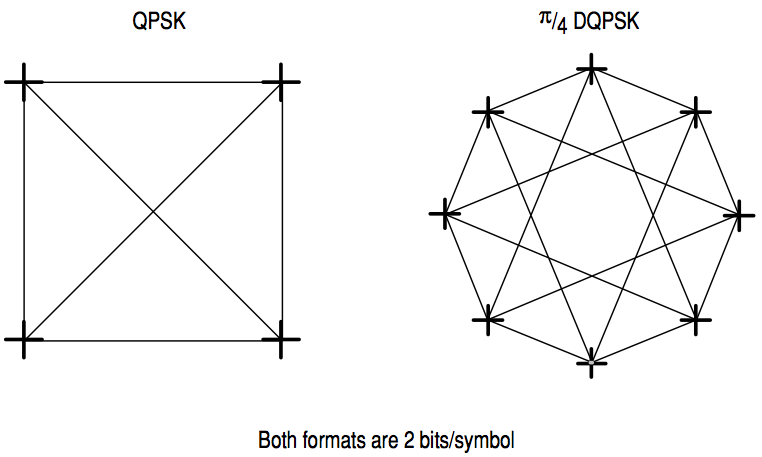
*I-Q* Offset Modulation

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 constellation, even though there are still just two *I*/*Q*. 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 the two *I* and two *Q* states. The reduced amplitude variations (about 3 dB for OQPSK, versus 30 to 40 dB for QPSK) allow a more power-efficient, less linear RF power amplifier to be used.

## 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 states. In some cases there are also restriction on allowable transitions. This occurs in π/4 DQPSK where the carrier trajectory does not go through the origin. A DQPSK transmission can transition from any symbol position to any other symbol position. The π/4 DQPSK modulation format is widely used in many applications including

* NADC-IS-54 (North American digital cellular)
* PDC (Pacific Digital Cellular)
* Cordless
  + PHS (Personal Handyphone System
* Trunked radio
  + TETRA (Trans European Trunked Radio)



Differential Modulation

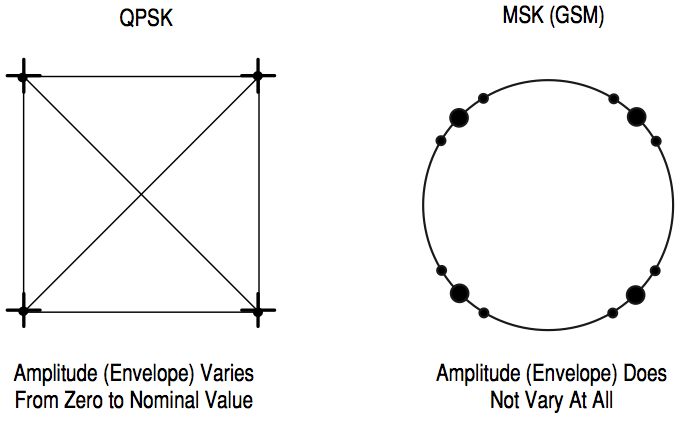
The π/4 DQPSK modulation format uses two QPSK constellations offset by 45 degrees (π/4 radians). Transitions must occur from one constellation to another. This guarantees that there is always a change in phase at each symbol, making clock recovery easier. The data is encoded in the magnitude and direction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

## 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).

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 that 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 that power efficiency, constant envelope modulation is not as well suited.

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



Constant Amplitude Modulation

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.

# 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

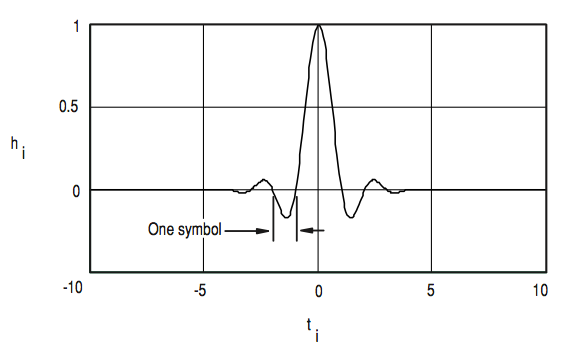
Any fast transition in the signal, whether it is amplitude, phase, or frequency, will require a wide occupied bandwidth. Any technique that helps to slow down these transitions will narrow the occupied bandwidth. Filtering serves to smooth these transitions (in *I* and *Q*). Filtering reduces interference because it reduces the tendency of one signal or one transmitter to interfere with another in a Frequency Division Multiple Access 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 to the filter.

## Nyquist or raised cosine filter

Figure 18 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 chosen to ring, or have impulse response of the filter cross through zero at the symbol clock frequency.



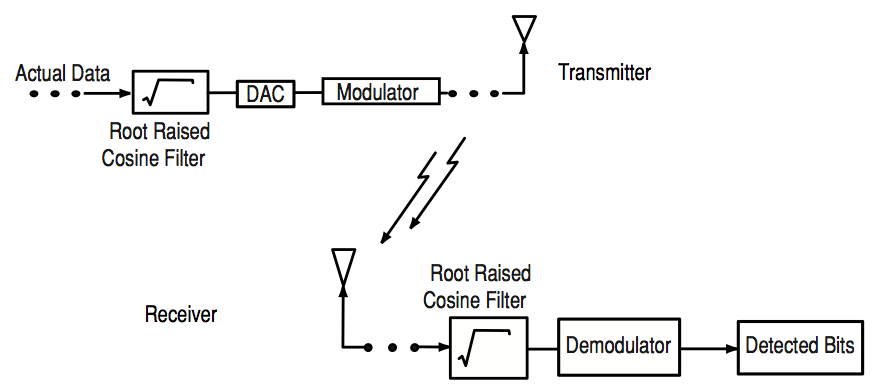
Nyquist or Raised Cosine Filter

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 ISI. 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.

## 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, therefore its potential for interfering with other transmitters.

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



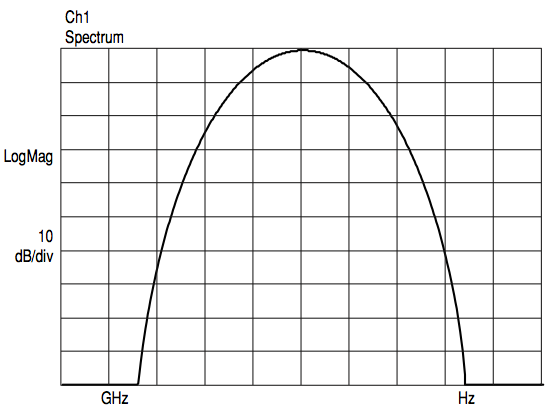
Transmitter-Receiver Matched Filters

To get zero 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 different companies often manufacture the transmitter and receiver. 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 . Matched filters are not used in Gaussian filtering.

## Gaussian filter

By 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.

Gaussian filters are used in GSM because of their advantage 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 INI) 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.



Gaussian Filter

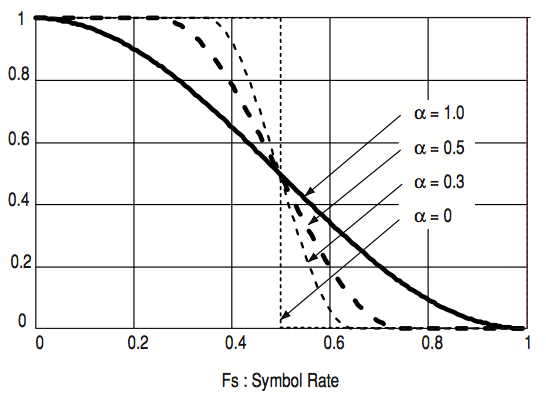
## Filter bandwidth parameter alpha

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

.

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

.



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.

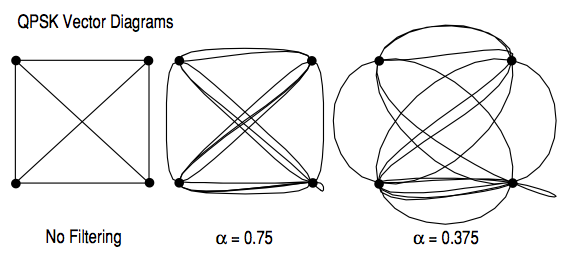
Alpha is sometimes called the *excess bandwidth factor* as it indicates the amount of occupied bandwidth that will be required in excess of the ideal occupied (which would be the same as the symbol rate).

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

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.

## Filter bandwidth effects

Different filter bandwidths show different effects. For example, a QPSK signal manifests different vector behaviors based on the alpha value. If the radio has no transmitter filter as shown on the left figure, the transitions between states are instantaneous. No filtering means an alpha of infinity.



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 filtering

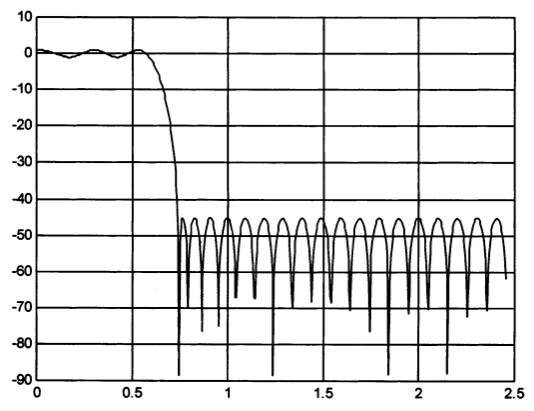
Consider an example of a π/4 DQPSK signal as used in NADC (IS-54). If an alpha of 1.0 is used, the transitions between the states are more gradual that for an alpha of infinity. Less power is needed to handle those transitions. Using an alpha of 0.5, the transmitted bandwidth decreases from 2 times the symbol rate of 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 overshoots 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 5 dB. 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 filter shape does not affect the power of these signals.

## Chebyshev equiripple FIR

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 there is a “rippled” magnitude frequency response 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.



Chebyshev Equiripple FIR Filter

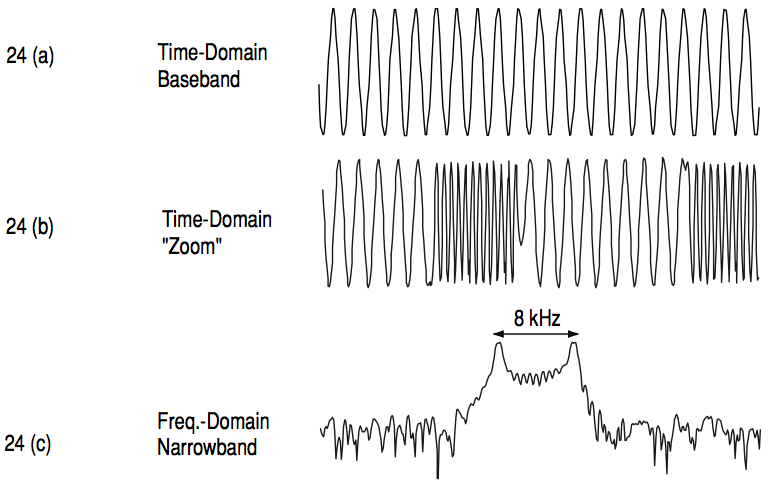
## Spectral efficiency versus power consumption

As with any natural resource, it makes no sense to waste the RF spectrum by using channel bands that 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 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. Spectral efficiency is highly desirable, but there are penalties in cost, size, weight complexity, talk time, and reliability.

# Digitally Modulated Signals in the Time and Frequency Domains

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. The pager uses a two-level FSK and the carrier shifts back and forth between two frequencies that are 8 kHz (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 24a. 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.

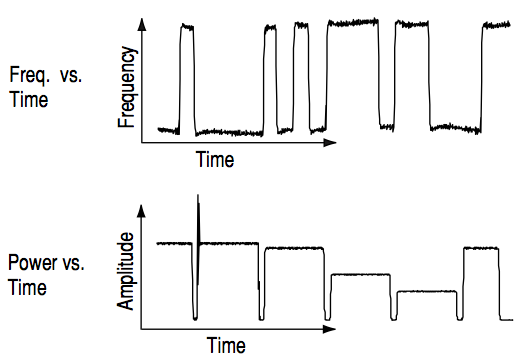
In a pager receiver the signals are first down converted to an IF or baseband frequency. In this example, the 930.004 MHz FSK-modulated signal is mixed with another signal at 930.002 MHz. The FSK modulation causes the transmitted signal to switch between 930.000 MHz and 930.008 MHz.



The result is a baseband signal that alternates between two frequencies, -2 kHz and +6 kHz. The 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 is how the signal looks in the IF portion of a receiver. This is how the IF of the radio detects the difference bits that are present. The frequency domain representation is shown in Figure 24c. 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.

## 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, sine 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.

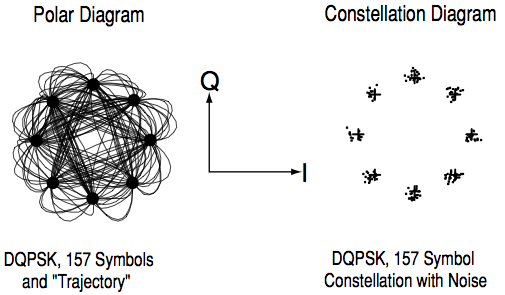


Power and Frequency View

## 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 for *I* and an independent set for *Q*).

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



Constellation Diagram

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 shift values.

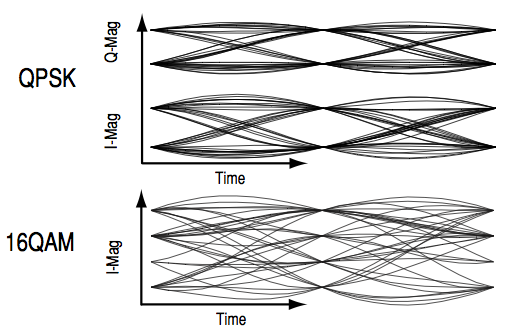
The constellation diagram shows a repetitive snapshot of that same burst, with values shown only at the decision points. The constellation diagram displays phase errors, as well as amplitude errors, at the decision points. The transitions between the decision points affects transmitted bandwidth. This 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 phenomenon such as Inter-Symbol Interference.

The relationship between constellation points and bits per symbol is

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

## Eye Diagrams

Another way to view 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* transition are shown separately and “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.



*I* and *Q* Diagrams

## Trellis Diagrams

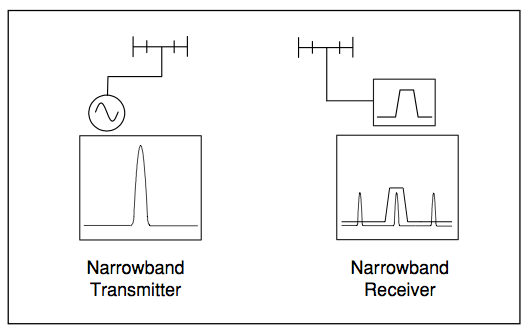
Figure 28 is called a *trellis* diagram, because it resembles a garden trellis. The trellis diagram shows time on the X-axis and phase on the Y-axis. This allows the examination of the phase transitions with different symbols. In this case it is for a GSM system. If a long series of binary ones were 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 where 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.

# 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.

## Multiplexing – Frequency

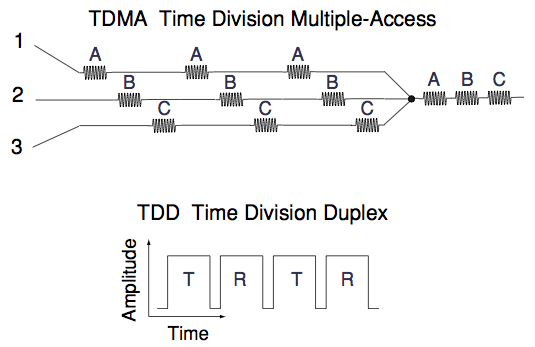
Frequency Division Multiple Access (FDMA) splits the available frequency band into smaller, fixed frequency channels. Each transmitter or receiver uses a separate frequency. Transmitters are narrowband or frequency-limited. A narrowband transmitter is used along with a receiver that has a narrowband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



Frequency Multiplexing

## Time Multiplexing

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.



Time Multiplexing

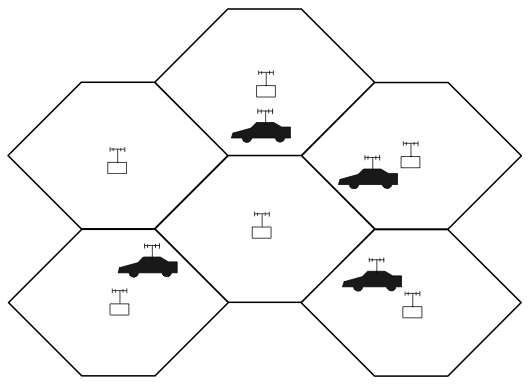
## Code Multiplexing

Code Division Multiple Access (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 1.25 MHz wide. In the case of US CDMA telephones, an additional type of channelization is added, in the form of coding.

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.

## Geographic Multiplexing

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.



Multiplexing – Geography

## 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.

## Penetration versus efficiency

Penetration means the ability of a signal to be used in environments where there is a lot of attenuation, 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 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.

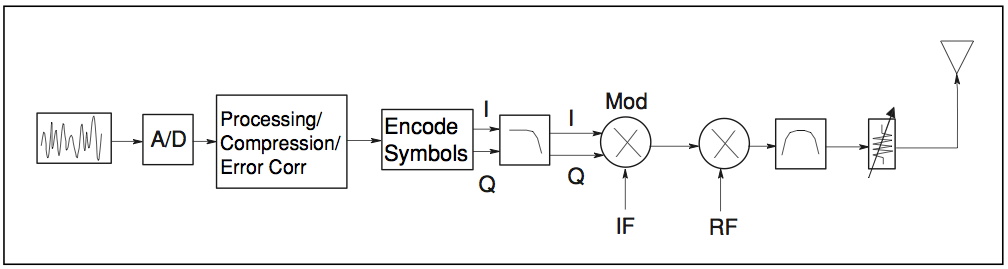
A cellular phone signals, however, must transmit live duplex voice. This requires a much higher bit rate and a much more efficient modulation technique. Cellular phones use more complex modulation formats (such as π/4 DQPSK and 0.3 GMSK) and faster symbol rates. Unfortunately, this greatly reduces penetration and one way to compensate is to use more power.

# How Digital Transmitters and Receivers Work

## A digital communications transmitter

Figure 33 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.

The next step is to add voice coding for data compression. Then some channel coding is added. Channel coding encodes the data QQQQQQQQQQQQQ 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 changes 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 that 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 intro 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.

## 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.

Generally, demodulation involves the following stages:

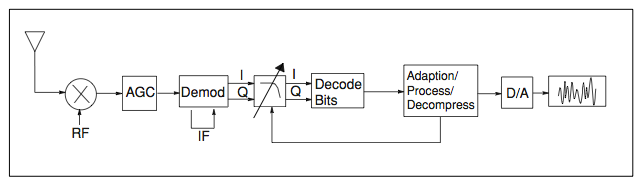
1. Carrier frequency recovery (carrier lock)
2. Symbol clock recovery (symbol lock)
3. Signal decomposition to *I* and *Q* components
4. Determining *I* and *Q* values for each symbol (“slicing”)
5. Decoding and de-interleaving
6. Expansion to original bit stream
7. Digital-to-analog conversion, if required

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 order to demodulate the bits successfully and recover the transmitted information. A symbol clock could be at the correct frequency but at the wrong phase. If the symbol clock were aligned with the transitions between symbols rather than the symbols themselves, demodulation would be unsuccessful.

Symbol clocks are usually fixed in frequency, which is accurately known by both the transmitter and receiver. The challenge 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 1010101010101010 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 advantage. The receiver can approximate where the carrier is but has no phase of timing symbol clock information to confirm. A difficult task in receiver design is t create carrier and symbol clock recover algorithms. That task can be made easier by the channel coding performed in the transmitter.



A Digital Receiver

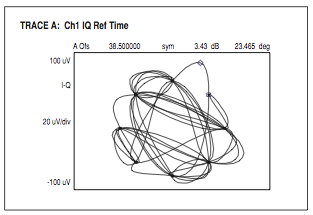
# Measurements on Digital RF Communications Systems

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 including 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
* Modulation accuracy

## Power measurements

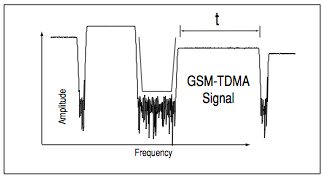
Power measurements include carrier power and associated measurements of gain or amplifiers and insertion loss of filters and attenuators. Signals used in digital modulation are noise-like. Band-power measurements (power integrated over a certain band of frequencies) or power spectral density (PSD) measurements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



Power Measurement

### Adjacent channel power

Adjacent channel power is a measure of interference created by one user that affects 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.



Power and Timing Measurements

For pulsed systems (such as TDMA), power measurements have a time component and have a frequency component, also. Burst power (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.

## Frequency measurements

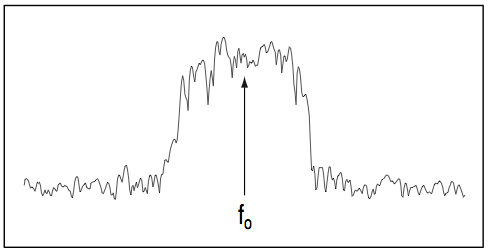
Frequency measurements are often more complex is digital systems since factors other than pure tones must be considered. Occupied bandwidth is an important measurement. It ensures that operators are staying within the bandwidth that they have been allocated. Adjacent channel power is also used to detect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

### 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 π/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.



Frequency measurements

## Timing measurements

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

## 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 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 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.

## Understanding Error Vector Magnitude

In view of 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 signals of which there must be , 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, π/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 and so forth. The differences between these tow phasors form the basis for the EVM measurements.

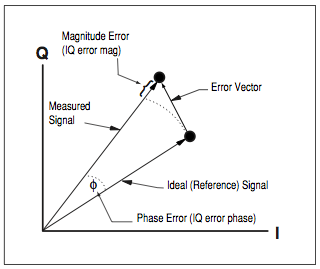
Figure defines EVM and several related terms. As shown EVM is the scalar difference between two phasor end points – the magnitude of the difference vector. Expressed another way, it is the residual noise and distortion remaining after an ideal version of the signal has been stripped away.

In the NADC-TDMA (IS-54) standard, EVM is defined as percentage of the signal voltage at the symbols. In the π/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 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 al the symbols (a value close to the average signal level) or by the voltage of the outermost (highest voltage) for 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.

## Troubleshooting with Error Vector Measurements

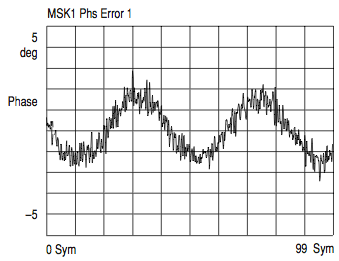
Measurements of error vector magnitude and related qualities can, when properly applied, provide much insight into the quality of a digitally modulated signal. They can pinpoint the causes for any problems uncovered by identifying exactly the type of degradation present in a signal and even help identify their sources.



EVM and Related Qualities

## Magnitude Versus Phase Error

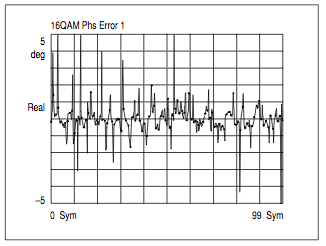
Different error mechanisms affect signals in different ways: in magnitude only, phase only, or both simultaneously. Knowing the relative amounts of each type of error can quickly confirm or rule out certain types of problems. Thus, the first diagnostic step is to resolve EVM into its magnitude and phase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

## *I*/*Q* Phase-Error versus Time

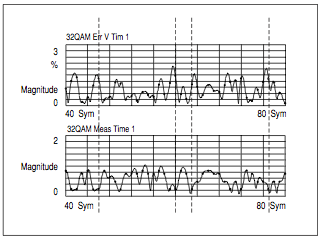
Phase error is the instantaneous angle difference between the measured signal and the ideal reference signal. Uniform noise is a sign of some form of phase noise (random jitter, residual PM/FM, and so forth).



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,” that is, 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.

## Error Vector Magnitude versus Time

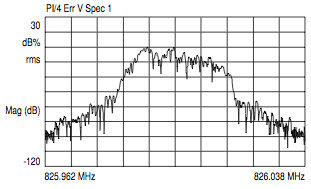
EVM is the difference between the input signal and the internally generated ideal reference. When viewed as a function of symbol or time, errors may be correlated to specific points on the input waveform, such as peaks or zero crossings. EVM is a scalar (magnitude-only) value. Error peaks occurring with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

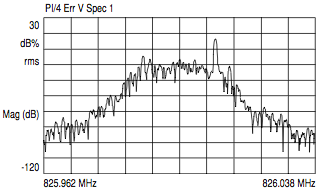


In Figure, EVM peaks on this signal (upper trace) occur every time the signal magnitude (lower trace) approaches zero. This is probably a zero-crossing error is amplification.

## 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 detains not visible in the time domain. In most digital systems, non-uniform noise distribution or discrete signal peaks indicate the presence of externally coupled interference.





# Summary

Communications system design requires the simultaneous conservation of bandwidth, power, and cost. This paper has presented the building blocks of any communications system. With these concepts in mind, it is possible to make informed decisions regarding the tradeoffs required to optimize a communications system (McClellan, Schafer, & Yoder, 2003) (Sklar, 2001) (Kronenburger & Sebeson, 2008) (Agilent Technologies, 2001). (Oppenheim, Willsky, & Nawab, 1996)

## Digital Modulation – an Overview

Modern mobile communication systems use digital modulation techniques. Advancement in very large-scale integration (VLSI) and digital signal processing (DSP) technology has made digital modulation more cost effective than analog transmission systems. Digital modulation offers many advantages over analog modulation. Some advantages include greater noise immunity and robustness to channel impairments, easier multiplexing of various forms of information (including voice, data, and video) and greater security. Furthermore, digital transmissions accommodate digital error-control codes, which detect and/or correct transmission errors, and support complex signal conditioning and processing techniques such as source coding, encryption, and equalization to improve the performance of the overall communications link. New multipurpose programmable digital signal processors have made it possible to implement digital modulators and demodulators completely in software. Instead of having a particular modem design permanently frozen as hardware, embedded software implementations now allow alterations and improvements without having to redesign or replace the modem (Rappaport, 2002).

In digital wireless communications systems, the modulating signal (the message) may be represented as a time sequence of symbols or pulses, where each symbol has finite states. Each symbol represents bits of information, where  
 bits/symbol. Many digital modulation schemes are used in modern wireless communications systems, and many more are sure to be introduced. Some of these techniques have subtle differences between one another, and each technique belongs to a family of related modulation methods. For example, phase shift keying (PSK) may be either coherently or differentially detected; and may have two, four eight or more possible levels ( bits) per symbol, depending on the manner in which information transmitted within a single symbol.

### Factors that Influence the Choice of Digital Modulation

Several factors influence the choice of a digital modulation scheme. A desirable modulation scheme provides low bit error rates at low received signal-to-noise ratios, performs well in multi-path and fading conditions, occupies a minimum of bandwidth, and is easy and cost-effective to implement. Existing modulation schemes do not simultaneously satisfy all of these requirements. Some modulation schemes are better in terms of the bit error rate performance, while others are better in terms of bandwidth efficiency. Depending on the demands of the particular application, tradeoffs are mad when selecting a digital modulation.

The performance of a modulation scheme is often measured in terms of its *power efficiency* and *bandwidth efficiency*. Power efficiency describes the ability of a modulation technique to preserve the fidelity of the digital message at low power levels. In a digital communications system, in order to increase noise immunity, it is necessary to increase the signal power. However, the amount by which the signal power should be increased to obtain a certain level of fidelity (an acceptable bit error probability) depends on the particular type of modulation employed. The *power efficiency*, (sometimes called energy efficiency)) of a digital modulation scheme is a measure of how favorably this tradeoff between fidelity and signal power is made, and is often expressed as the ratio of the *signal energy per bit* to *noise power spectral density* () required at the receiver input for a certain probability of error (say ).

*Bandwidth efficiency* describes the ability of a modulation scheme to accommodate data within a limited bandwidth. In general, increasing the data rate implies decreasing the pulse width of digital symbol, which increases the bandwidth of the signal. Thus, there is an unavoidable relationship between data rate and bandwidth occupancy. However, some modulation schemes perform better than others in making this tradeoff. Bandwidth efficiency reflects of efficiently the allocated bandwidth is utilized and is defined as the ratio of *throughput data rate per Hertz* in a given bandwidth. If is the data rate in bits per second, and is the bandwidth occupied by the modulated RF signal, then bandwidth efficiency, is expressed as  
 bps/Hz.

The system capacity of a digital mobile communication system is directly related to the bandwidth efficiency of the modulation scheme, since a modulation with a greater value of will transmit more data in a given spectrum allocation.

There is a fundamental upper bound on the achievable bandwidth efficiency. Shannon’s channel coding theorem states that for an arbitrarily small probability of error, the maximum possible bandwidth efficiency is limited by the noise in the channel, and is given by the channels capacity formula. Note that Shannon’s bound applies for AWGN non-fading channels.

where *C* is the channel capacity (in bps), *B* is the RF bandwidth, and is the signal-to-noise ratio.

In the design of a digital communication system, very often there is a tradeoff between control coding to a message increases the bandwidth occupancy (and this, in turn, reduces the bandwidth efficiency. For example, adding error control coding to a message increases the bandwidth increases the bandwidth occupancy (and this, in turn, reduces the bandwidth efficiency), but at the same time reduces the required received power for a particular bit error rate, and hence trades bandwidth efficiency for power efficiency. On the other hand, higher-level modulation schemes (M-ary keying) decrease bandwidth occupancy but increase the required received power, and hence trade power efficiency for bandwidth efficiency.

While power and bandwidth efficiency considerations are very important, other factors also affect the choice of a digital modulation scheme. For example, for all personal communications systems that serve a large user community, the cost and complexity of the subscriber receiver must be minimized, and a modulation, which is simple to detect, is most attractive. The performance of the modulation scheme under various types of channel impairments such as Raleigh and Ricean fading and multipath time dispersion, given a particular demodulator implementation, is another key factor in selecting a modulation. In cellular systems where interference is major issue, the performance of a modulation scheme in an interference environment is extremely important. Sensitivity to detection of timing jitter, caused by time-varying channels, is also an important consideration in choosing a particular modulation scheme. In general, the modulation, interference, and implementation of the time-varying effects of the channel as well as the performance of the specific demodulator are analyzed as a complete system using simulation to determine relative performance and ultimate selection.

### Bandwidth and Power Spectral Density of Digital Signals

The definition of signal *bandwidth* varies with context, and there is no single definition that suits all applications. All definitions, however, are based on some measure on the *power spectral density* (PSD) of the signal. The power spectral density of a random signal is defined as



# Works Cited

Agilent Technologies. (2001). *Digital Modulation in Communications Systems: An Introduction.* Santa Clara, CA, USA: Agilent Technologies.

Kronenburger, J., & Sebeson, J. (2008). *Analog and Digital Signal Processing: An Integrated Computational Approach with MATLAB.* Clifton Park, NY, USA: Thompson Delmar Learning.

McClellan, J. H., Schafer, R. W., & Yoder, M. A. (2003). *Signal Processing First.* Upper Sadle River, NJ, USA: Pearson Education, Inc.

Oppenheim, A. V., Willsky, A. S., & Nawab, S. H. (1996). *Signals & Systems.* Upper Saddle River, NJ, USA: Prentice Hall.

Sklar, B. (2001). *Digital Communications: Fundamentals and Applications.* Upper Sadle River, NJ, USA: Pearson Education, Inc.

Rappaport, T. S. (2002). *Wireless Communications: Principles and Practice.* Upper Saddle River, NJ, USA: Prentice Hall PTR.

(Agilent Technologies, 2001) (Agilent Technologies, 2001) (Agilent Technologies, 2001)