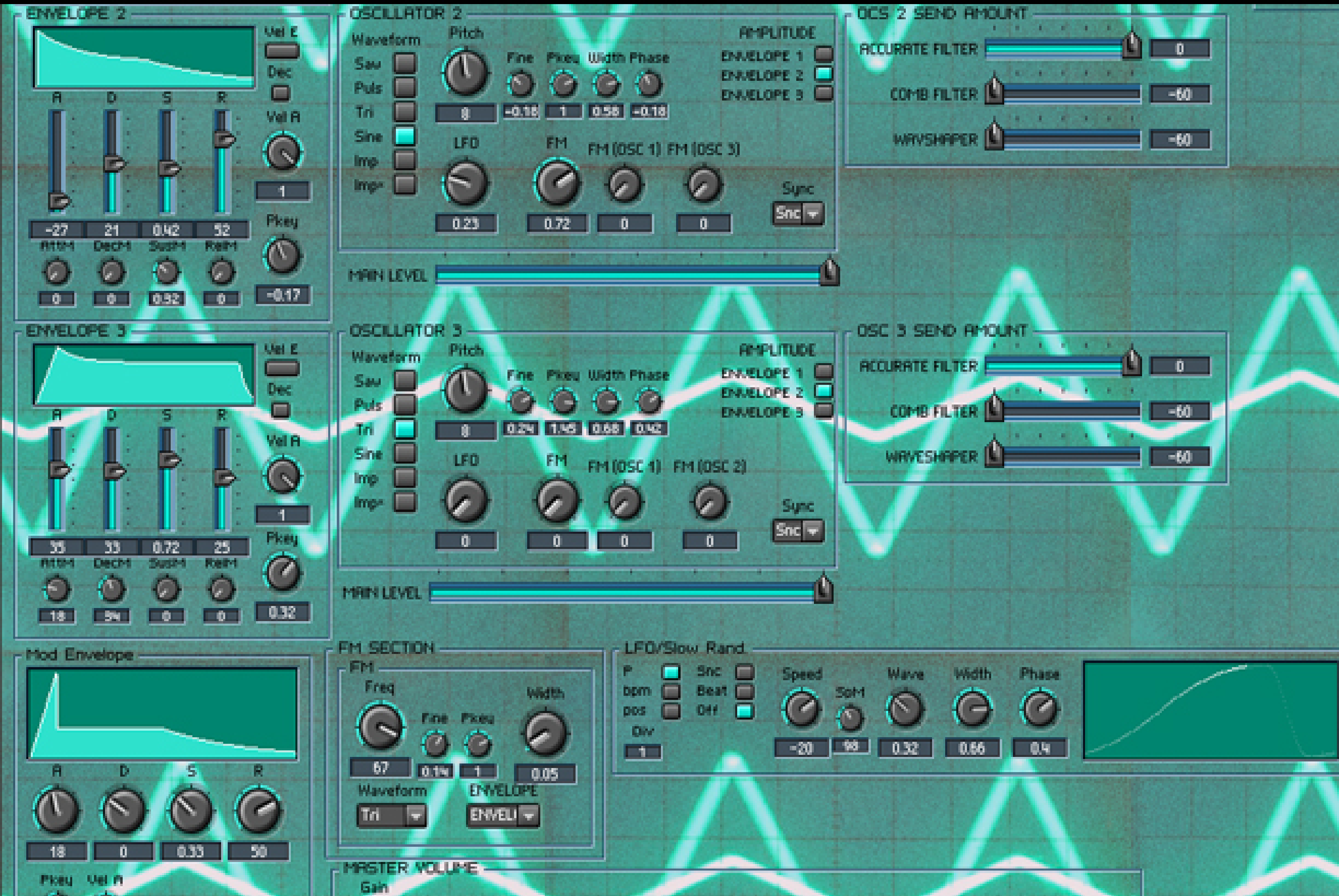
Analog Modulation
and
Digital Modulation



Steve Stevenson

First Edition, 2012

ISBN 978-81-323-3508-5

Published by:
University Publications
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter- 1

Amplitude Modulation

Amplitude modulation (**AM**) is a technique used in electronic communication, most commonly for transmitting information via a radio carrier wave. AM works by varying the strength of the transmitted signal in relation to the information being sent. For example, changes in the signal strength can be used to specify the sounds to be reproduced by a loudspeaker, or the light intensity of television pixels. (Contrast this with frequency modulation, also commonly used for sound transmissions, in which the frequency is varied; and phase modulation, often used in remote controls, in which the phase is varied)

In the mid-1870s, a form of amplitude modulation—initially called "undulatory currents"—was the first method to successfully produce quality audio over telephone lines. Beginning with Reginald Fessenden's audio demonstrations in 1906, it was also the original method used for audio radio transmissions, and remains in use today by many forms of communication—"AM" is often used to refer to the mediumwave broadcast band.

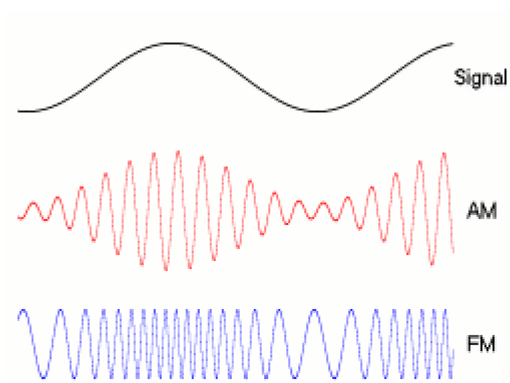


Fig 1: An audio signal (top) may be carried by an AM or FM radio wave.

Forms of amplitude modulation

In radio communication, a continuous wave radio-frequency signal (a sinusoidal carrier wave) has its amplitude modulated by an audio waveform before being transmitted.

In the frequency domain, amplitude modulation produces a signal with power concentrated at the carrier frequency and in two adjacent sidebands. Each sideband is equal in bandwidth to that of the modulating signal and is a mirror image of the other. Amplitude modulation that results in two sidebands and a carrier is often called *double-sideband amplitude modulation* (DSB-AM). Amplitude modulation is inefficient in terms of power usage. At least two-thirds of the power is concentrated in the carrier signal, which carries no useful information (beyond the fact that a signal is present).

To increase transmitter efficiency, the carrier can be removed (suppressed) from the AM signal. This produces a reduced-carrier transmission or *double-sideband suppressed-carrier* (DSBSC) signal. A suppressed-carrier amplitude modulation scheme is three times more power-efficient than traditional DSB-AM. If the carrier is only partially suppressed, a *double-sideband reduced-carrier* (DSBRC) signal results. DSBSC and DSBRC signals need their carrier to be regenerated (by a beat frequency oscillator, for instance) to be demodulated using conventional techniques.

Improved bandwidth efficiency is achieved—at the expense of increased transmitter and receiver complexity—by completely suppressing both the carrier and one of the sidebands. This is single-sideband modulation, widely used in amateur radio due to its efficient use of both power and bandwidth.

A simple form of AM often used for digital communications is *on-off keying*, a type of *amplitude-shift keying* by which binary data is represented as the presence or absence of a carrier wave. This is commonly used at radio frequencies to transmit Morse code, referred to as continuous wave (CW) operation.

ITU designations

In 1982, the International Telecommunication Union (ITU) designated the various types of amplitude modulation as follows:

Designation	Description
A3E	double-sideband full-carrier - the basic AM modulation scheme
R3E	single-sideband reduced-carrier
H3E	single-sideband full-carrier
J3E	single-sideband suppressed-carrier
B8E	independent-sideband emission
C3F	vestigial-sideband

Lincompex   linked compressor and expander

Example: double-sideband AM

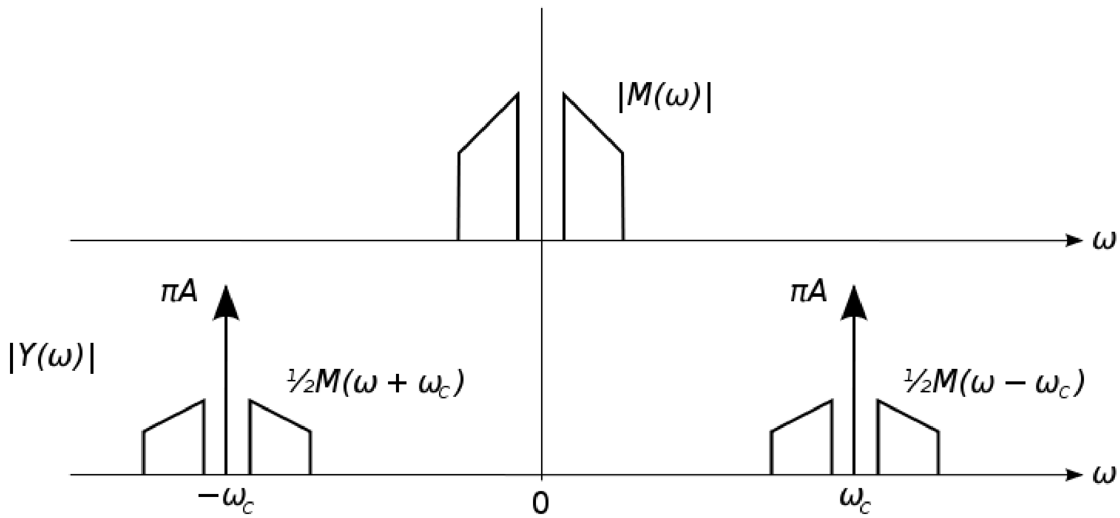


Fig 2: The (2-sided) spectrum of an AM signal.

A carrier wave is modeled as a simple sine wave, such as:

$$c(t) = C \cdot \sin(\omega_c t + \phi_c),$$

where the radio frequency (in Hz) is given by:

$$\omega_c/(2\pi).$$

The constants C and ϕ_c represent the carrier amplitude and initial phase, and are introduced for generality. For simplicity however, their respective values can be set to 1 and 0.

Let *m*(*t*) represent an arbitrary waveform that is the message to be transmitted. And let the constant *M* represent its largest magnitude. For instance:

$$m(t) = M \cdot \cos(\omega_m t + \phi).$$

Thus, the message might be just a simple audio tone of frequency

$$\omega_m/(2\pi).$$

It is generally assumed that $\omega_m \ll \omega_c$ and that $\min[m(t)] = -M.$

Then amplitude modulation is created by forming the product:

$$y(t) = [A + m(t)] \cdot c(t),$$
$$= [A + M \cdot \cos(\omega_m t + \phi)] \cdot \sin(\omega_c t).$$

A represents the carrier amplitude which is a constant that we would choose to demonstrate the modulation index. The values A=1, and M=0.5, produce a $y(t)$ depicted by the graph labelled "50% Modulation" in Figure 4.

For this simple example, $y(t)$ can be trigonometrically manipulated into the following equivalent form:

$$y(t) = A \cdot \sin(\omega_c t) + \frac{M}{2}\left[\sin((\omega_c + \omega_m)t + \phi) + \sin((\omega_c - \omega_m)t - \phi)\right].$$

Therefore, the modulated signal has three components, a carrier wave and two sinusoidal waves (known as sidebands) whose frequencies are slightly above and below ω_c.

Also notice that the choice A=0 eliminates the carrier component, but leaves the sidebands. That is the DSBSC transmission mode. To generate double-sideband full carrier (A3E), we must choose:

$$A \geq M.$$

Spectrum

For more general forms of $m(t)$, trigonometry is not sufficient. But if the top trace of Figure 2 depicts the frequency spectrum, of $m(t)$, then the bottom trace depicts the modulated carrier. It has two groups of components: one at positive frequencies (centered on $+\omega_c$) and one at negative frequencies (centered on $-\omega_c$). Each group contains the two sidebands and a narrow component in between that represents the energy at the carrier frequency. We need only be concerned with the positive frequencies. The negative ones are a mathematical artifact that contains no additional information. Therefore, we see that an AM signal's spectrum consists basically of its original (2-sided) spectrum shifted up to the carrier frequency.

Figure 2 is a result of computing the Fourier transform of:
$[A + m(t)] \cdot \sin(\omega_c t)$, using the following transform pairs:

$$
\begin{aligned}
m(t) &\overset{\mathcal{F}}{\Longleftrightarrow} M(\omega) \\
\sin(\omega_c t) &\overset{\mathcal{F}}{\Longleftrightarrow} i\pi \cdot [\delta(\omega+\omega_c) - \delta(\omega-\omega_c)] \\
A \cdot \sin(\omega_c t) &\overset{\mathcal{F}}{\Longleftrightarrow} i\pi A \cdot [\delta(\omega+\omega_c) - \delta(\omega-\omega_c)] \\
m(t) \cdot \sin(\omega_c t) &\overset{\mathcal{F}}{\Longleftrightarrow} \frac{1}{2\pi} \cdot \{M(\omega)\} * \{i\pi \cdot [\delta(\omega+\omega_c) - \delta(\omega-\omega_c)]\} \\
&= \frac{i}{2} \cdot [M(\omega+\omega_c) - M(\omega-\omega_c)]
\end{aligned}
$$

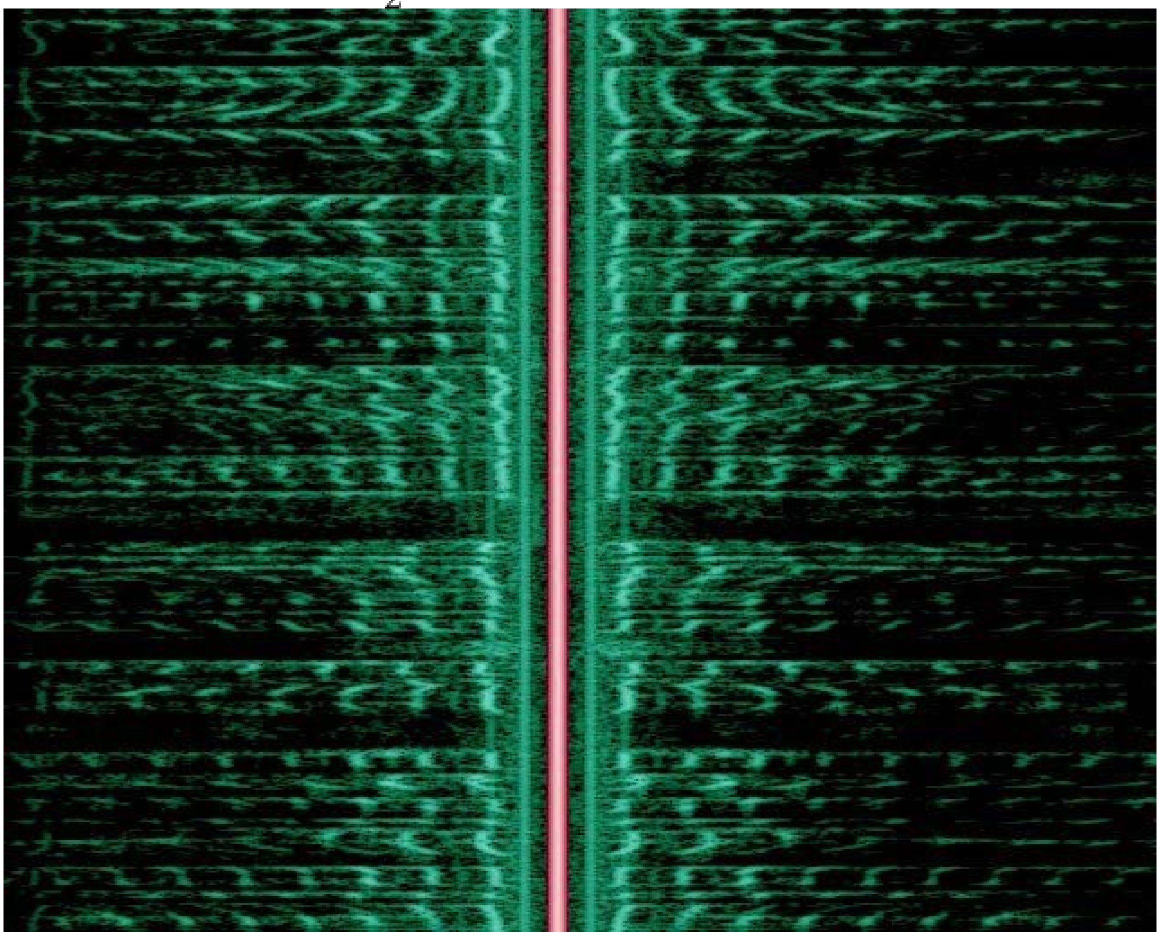


Fig 3: The spectrogram of an AM broadcast shows its two sidebands (green) separated by the carrier signal (red).

Power and spectrum efficiency

In terms of the positive frequencies, the transmission bandwidth of AM is twice the signal's original (baseband) bandwidth—since both the positive and negative sidebands are shifted up to the carrier frequency. Thus, double-sideband AM (DSB-AM) is spectrally inefficient, meaning that fewer radio stations can be accommodated in a given broadcast band. The various suppression methods in Forms of AM can be readily understood in terms of the diagram in Figure 2. With the carrier suppressed there would be no energy at the center of a group. And with a sideband suppressed, the "group" would have the same bandwidth as the positive frequencies of $M(\omega)$. The transmitter power efficiency of DSB-AM is relatively poor (about 33%). The benefit of this system is that receivers are cheaper to produce. The forms of AM with suppressed carriers are found to be 100% power efficient, since no power is wasted on the carrier signal which conveys no information.

Modulation index

It can be defined as the measure of extent of amplitude variation about an unmodulated maximum carrier. As with other modulation indices, in AM, this quantity, also called *modulation depth*, indicates by how much the modulated variable varies around its 'original' level. For AM, it relates to the variations in the carrier amplitude and is defined as:

$$h = \frac{\text{peak value of } m(t)}{A} = \frac{M}{A},$$ where M and A were introduced above.

So if $h = 0.5$, the carrier amplitude varies by 50% above and below its unmodulated level, and for $h = 1.0$ it varies by 100%. To avoid distortion in the A3E transmission mode, modulation depth greater than 100% must be avoided. Practical transmitter systems will usually incorporate some kind of limiter circuit, such as a VOGAD, to ensure this. However, AM demodulators can be designed to detect the inversion (or 180 degree phase reversal) that occurs when modulation exceeds 100% and automatically correct for this effect.

Variations of modulated signal with percentage modulation are shown below. In each image, the maximum amplitude is higher than in the previous image. Note that the scale changes from one image to the next.

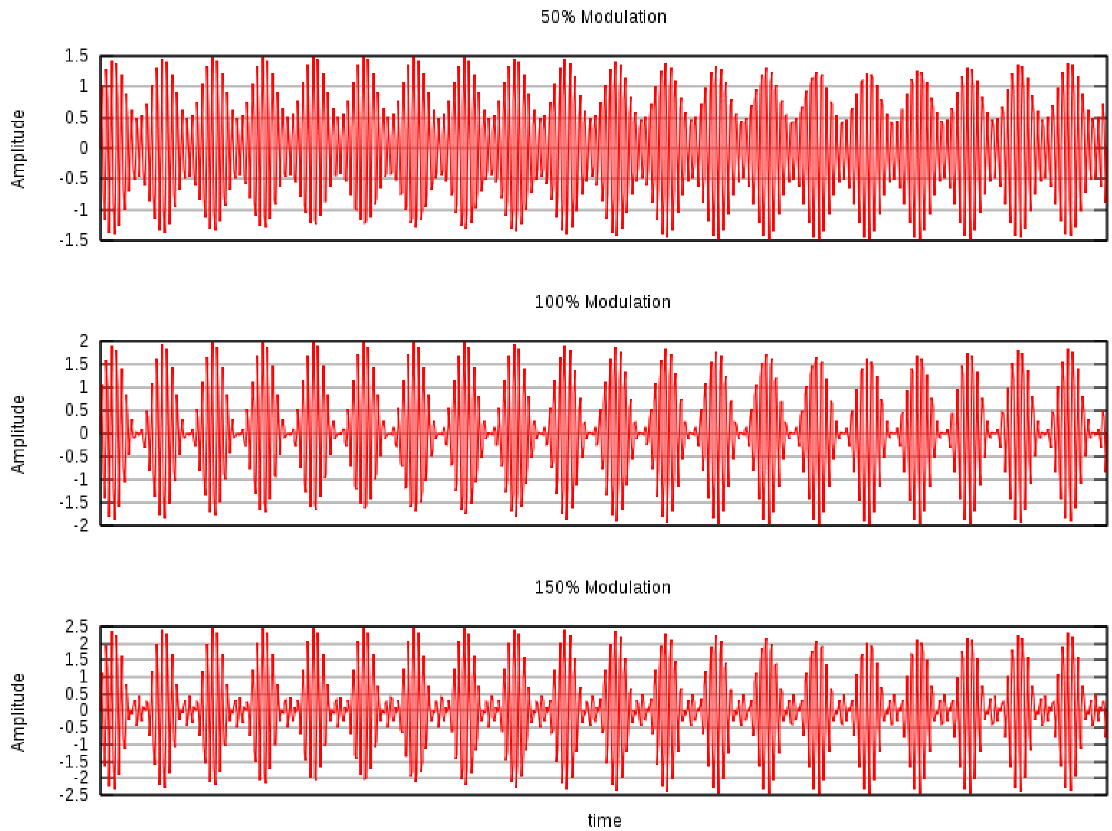
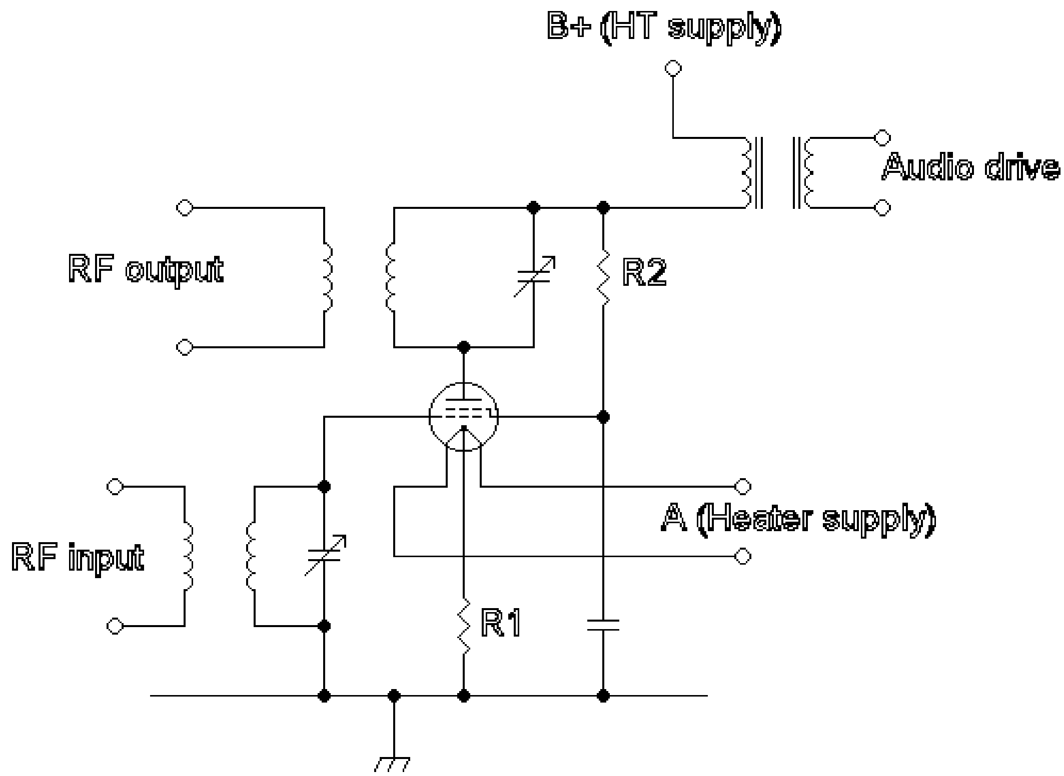


Fig 4: Modulation depth

Amplitude modulator designs

Circuits

A wide range of different circuits have been used for AM, but one of the simplest circuits uses anode or collector modulation applied via a transformer. While it is perfectly possible to create good designs using solid-state electronics, valved (vacuum tube) circuits are shown here. In general, valves are able to more easily yield RF powers, in excess of what can be easily achieved using solid-state transistors. Many high-power broadcast stations still use valves.



Anode modulation using a transformer. The tetrode is supplied with an anode supply (and screen grid supply) which is modulated via the transformer. The resistor R1 sets the grid bias; both the input and outputs are tuned LC circuits which are tapped into by inductive coupling

Modulation circuit designs can be broadly divided into low and high level.

Low level

Here a small audio stage is used to modulate a low power stage; the output of this stage is then amplified using a linear RF amplifier. Wideband power amplifiers are used to preserve the sidebands of the modulated waves. In this arrangement, modulation is done at low power. To amplify it we use a wideband power amplifier at the output.

Advantages

The advantage of using a linear RF amplifier is that the smaller early stages can be modulated, which only requires a small audio amplifier to drive the modulator.

Disadvantages

The great disadvantage of this system is that the amplifier chain is less efficient, because it has to be linear to preserve the modulation. Hence Class C amplifiers cannot be employed.

An approach which marries the advantages of low-level modulation with the efficiency of a Class C power amplifier chain is to arrange a feedback system to compensate for the substantial distortion of the AM envelope. A simple detector at the transmitter output (which can be little more than a loosely coupled diode) recovers the audio signal, and this is used as negative feedback to the audio modulator stage. The overall chain then acts as a linear amplifier as far as the actual modulation is concerned, though the RF amplifier itself still retains the Class C efficiency. This approach is widely used in practical medium power transmitters, such as AM radiotelephones.

High level

With high level modulation, the modulation takes place at the final amplifier stage where the carrier signal is at its maximum

Advantages

One advantage of using class C amplifiers in a broadcast AM transmitter is that only the final stage needs to be modulated, and that all the earlier stages can be driven at a constant level. These class C stages will be able to generate the drive for the final stage for a smaller DC power input. However, in many designs in order to obtain better quality AM the penultimate RF stages will need to be subject to modulation as well as the final stage.

Disadvantages

A large audio amplifier will be needed for the modulation stage, at least equal to the power of the transmitter output itself. Traditionally the modulation is applied using an audio transformer, and this can be bulky. Direct coupling from the audio amplifier is also possible (known as a cascode arrangement), though this usually requires quite a high DC supply voltage (say 30 V or more), which is not suitable for mobile units.

AM demodulation methods

The simplest form of AM demodulator consists of a diode which is configured to act as envelope detector. Another type of demodulator, the product detector, can provide better quality demodulation, at the cost of added circuit complexity.

Chapter- 2

Single-Sideband Modulation

Single-sideband modulation (**SSB**) is a refinement of amplitude modulation that more efficiently uses electrical power and bandwidth. It is closely related to vestigial sideband modulation (VSB) (see below).

Amplitude modulation produces a modulated output signal that has twice the bandwidth of the original baseband signal. Single-sideband modulation avoids this bandwidth doubling, and the power wasted on a carrier, at the cost of somewhat increased device complexity.

The first U.S. patent for SSB modulation was applied for on December 1, 1915 by John Renshaw Carson. The U.S. Navy experimented with SSB over its radio circuits before World War I. SSB first entered commercial service on January 7, 1927 on the longwave transatlantic public radiotelephone circuit between New York and London. The high power SSB transmitters were located at Rocky Point, New York and Rugby, England. The receivers were in very quiet locations in Houlton, Maine and Cupar Scotland.

SSB was also used over long distance telephone lines, as part of a technique known as frequency-division multiplexing (FDM). FDM was pioneered by telephone companies in the 1930s. This enabled many voice channels to be sent down a single physical circuit, for example in L-carrier. SSB allowed channels to be spaced (usually) just 4,000 Hz apart, while offering a speech bandwidth of nominally 300–3,400 Hz.

Amateur radio operators began serious experimentation with SSB after World War II. The Strategic Air Command established SSB as the radio standard for its aircraft in 1957. It has become a de facto standard for long-distance voice radio transmissions since then.

SSB and VSB can also be regarded mathematically as special cases of analog quadrature amplitude modulation.

Mathematical formulation

Let $s(t)$ be the baseband waveform to be transmitted. Its Fourier transform, $S(f)$, is Hermitian symmetrical about the $f = 0$ axis, because $s(t)$ is real-valued. Double sideband modulation of $s(t)$ to a radio transmission frequency, F_c, moves the axis of symmetry to $f = \pm F_c$, and the two sides of each axis are called sidebands.

Let $\widehat{s}(t)$ represent the Hilbert transform of $s(t)$. Then

$$s_a(t) = s(t) + j \cdot \widehat{s}(t)$$

is a useful mathematical concept, called an analytic signal. The Fourier transform of $s_a(t)$ equals $2 \cdot S(f)$, for $f > 0$, but it has no negative-frequency components. So it can be modulated to a radio frequency and produce just a **single** sideband.

The analytic representation of $\cos(2\pi F_c \cdot t)$ is:

$$\cos(2\pi F_c \cdot t) + j \cdot \sin(2\pi F_c \cdot t) = e^{j2\pi F_c \cdot t}$$ (the equality is Euler's formula)

whose Fourier transform is $\delta(f - F_c)$.

When $s_a(t)$ is modulated (i.e. multiplied) by $e^{j2\pi F_c \cdot t}$, all frequency components are shifted by $+F_c$, so there are still no negative-frequency components. Therefore, the complex product is an **analytic representation** of the single sideband signal:

$$s_a(t) \cdot e^{j2\pi F_c \cdot t} = s_{ssb}(t) + j \cdot \widehat{s}_{ssb}(t)$$

where $s_{ssb}(t)$ is the real-valued, single sideband waveform. Therefore:

$$\begin{aligned} s_{ssb}(t) &= Re\{s_a(t) \cdot e^{j2\pi F_c \cdot t}\} \\ &= Re\{\ [s(t) + j \cdot \widehat{s}(t)] \cdot [\cos(2\pi F_c \cdot t) + j \cdot \sin(2\pi F_c \cdot t)]\ \} \\ &= s(t) \cdot \cos(2\pi F_c \cdot t) - \widehat{s}(t) \cdot \sin(2\pi F_c \cdot t) \end{aligned}$$

The presence of two out-of-phase (quadrature) carrier waves is now evident.

Lower sideband

$s_a(t)$represents the baseband signal's upper sideband, $s_+(t)$. It is also possible, and useful, to convey the baseband information using its lower sideband, $s_-(t)$, which is a mirror image about f=0 Hz. By a general property of the Fourier transform, that symmetry means it is the complex conjugate of $s_+(t)$**:**

$$s_-(t) = s_+^*(t) = s_a^*(t) = s(t) - j \cdot \widehat{s}(t)$$

Note that**:**

$$s_+(t) + s_-(t) = 2s(t)$$

The gain of 2 is a result of defining the analytic signal (one sideband) to have the same total energy as $s(t)$(both sidebands).

As before, the signal is modulated by $e^{j2\pi F_c \cdot t}$. The typical $F_c$is large enough that the translated lower sideband (LSB) has no negative-frequency components. Then the result is another analytic signal, whose real part is the actual transmission.

$$\begin{aligned} s_{lsb}(t) &= Re\{s_a^*(t) \cdot e^{j2\pi F_c \cdot t}\} \\ &= s(t) \cdot \cos(2\pi F_c \cdot t) + \widehat{s}(t) \cdot \sin(2\pi F_c \cdot t) \end{aligned}$$

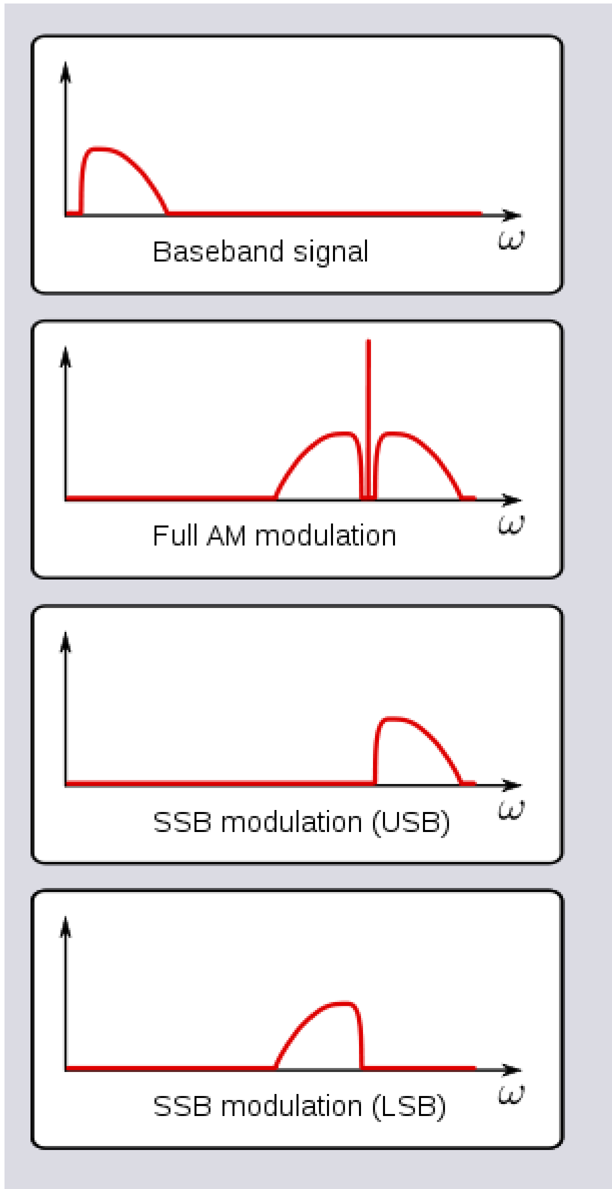
Note that the sum of the two sideband signals is

$$2s(t) \cdot cos(2\pi F_c \cdot t)$$

which is the classic model of suppressed-carrier double sideband AM.

Practical implementations

Bandpass filtering



The LSB spectrum is inverted compared to the baseband. As an example, a 2 kHz audio baseband signal modulated onto a 5 MHz carrier will produce a SSB frequency of 5.002000 MHz if USB is used or 4.998000 MHz if LSB is used.

A signal at frequency f_0 amplitude-modulated onto a carrier wave at f_m can be expressed as simple multiplication of two cosine waves: $\cos(\omega_0 t)\cos(\omega_m t)$, where $\omega_x = 2\pi f_x$. Applying a simple trigonometric identity, we can change the above expression to be

$\frac{1}{2}(\cos((\omega_m + \omega_0)t) + \cos((\omega_m - \omega_0)t))$. Each cosine term in the equation is known as a sideband.

One method of producing an SSB signal is to remove one of the sidebands via filtering, leaving only either the **upper sideband** (**USB**), the sideband with the higher frequency, or less commonly the **lower sideband** (**LSB**), the sideband with the lower frequency. Most often, the carrier is reduced or removed entirely (suppressed), being referred to in full as **single sideband suppressed carrier** (**SSBSC**). Assuming both sidebands are symmetric, which is the case for a normal AM signal, no information is lost in the process. Since the final RF amplification is now concentrated in a single sideband, the effective power output is greater than in normal AM (the carrier and redundant sideband account for well over half of the power output of an AM transmitter). Though SSB uses substantially less bandwidth and power, it cannot be demodulated by a simple envelope detector like standard AM.

Hartley modulator

An alternate method of generation known as a **Hartley modulator**, named after R. V. L. Hartley, uses phasing to suppress the unwanted sideband. To generate an SSB signal with this method, two versions of the original signal are generated, mutually 90° out of phase for any single frequency within the operating bandwidth. Each one of these signals then modulates carrier waves (of one frequency) that are also 90° out of phase with each other. By either adding or subtracting the resulting signals, a lower or upper sideband signal results. A benefit of this approach is to allow an analytical expression for SSB signals, which can be used to understand effects such as synchronous detection of SSB.

Shifting the baseband signal 90° out of phase cannot be done simply by delaying it, as it contains a large range of frequencies. In analog circuits, a wideband 90-degree phase-difference network is used. The method was popular in the days of vacuum-tube radios, but later gained a bad reputation due to poorly adjusted commercial implementations. Modulation using this method is again gaining popularity in the homebrew and DSP fields. This method, utilizing the Hilbert transform to phase shift the baseband audio, can be done at low cost with digital circuitry.

Weaver modulator

Another variation, the **Weaver modulator**, uses only lowpass filters and quadrature mixers, and is a favored method in digital implementations.

In Weaver's method, the band of interest is first translated to be centered at zero, conceptually by modulating a complex exponential exp($j\omega t$) with frequency in the middle of the voiceband, but implemented by a quadrature pair of sine and cosine modulators at that frequency (e.g. 2 kHz). This complex signal or pair of real signals is then lowpass filtered to remove the undesired sideband that is not centered at zero. Then, the single-sideband complex signal centered at zero is upconverted to a real signal, by another pair of quadrature mixers, to the desired center frequency.

Demodulation

The front end of an SSB receiver is similar to that of an AM or FM receiver, consisting of a superheterodyne RF front end that produces a frequency-shifted version of the radio frequency (RF) signal within a standard intermediate frequency (IF) band.

To recover the original signal from the IF SSB signal, the single sideband must be frequency-shifted down to its original range of baseband frequencies, by using a product detector which mixes it with the output of a beat frequency oscillator (BFO). In other words, it is just another stage of heterodyning.

For this to work, the BFO frequency must be accurately adjusted. If the BFO is mis-adjusted, the output signal will be frequency-shifted, making speech sound strange and "Donald Duck"-like, or unintelligible. Some receivers use a carrier recovery system, which attempts to automatically lock on to the exact frequency.

As an example, consider an IF SSB signal centered at frequency F_{if}= 45000 Hz. The baseband frequency it needs to be shifted to is F_b= 2000 Hz. The BFO output waveform is $cos(2\pi \cdot F_{bfo} \cdot t)$. When the signal is multiplied by (aka 'heterodyned with') the BFO waveform, it shifts the signal to $(F_{if} + F_{bfo})$ and to $|F_{if} - F_{bfo}|$, which is known as the *beat frequency* or *image frequency*. The objective is to choose an F_{bfo} that results in $|F_{if} - F_{bfo}| = F_b$ = 2000 Hz. (The unwanted components at $(F_{if} + F_{bfo})$ can be removed by a lowpass filter (for which an output transducer or the human ear may serve)).

Note that there are two choices for F_{bfo}**:** 43000 Hz and 47000 Hz, a.k.a. *low-side* and *high-side* injection. With high-side injection, the spectral components that were distributed around 45000 Hz will be distributed around 2000 Hz in the reverse order, also known as an *inverted spectrum*. That is in fact desirable when the IF spectrum is also inverted, because the BFO inversion restores the proper relationships. One reason for that is when the IF spectrum is the output of an inverting stage in the receiver. Another reason is when the SSB signal is actually a lower sideband, instead of an upper sideband. But if both reasons are true, then the IF spectrum in not inverted, and the non-inverting BFO (43000 Hz) should be used.

If F_{bfo} is off by a small amount, then the beat frequency is not exactly F_b, which can lead to the speech distortion mentioned earlier.

SSB as a speech-scrambling technique

SSB techniques can also be adapted to frequency-shift and frequency-invert baseband waveforms. These effects were used, in conjunction with other filtering techniques, during World War II as a simple method for speech encryption. Radiotelephone conversations between the US and Britain were intercepted and "decrypted" by the Germans; they included some early conversations between Franklin D. Roosevelt and Churchill. In fact, the signals could be understood directly by trained operators. Largely to allow secure communications between Roosevelt and Churchill, the SIGSALY system of digital encryption was devised.

Today, such simple inversion-based speech encryption techniques are easily decrypted using simple techniques and are no longer regarded as secure.

Suppressed carrier SSB

Suppressed carrier SSB modulation is used by ATSC.

Vestigial sideband (VSB)

A **vestigial sideband** (in radio communication) is a sideband that has been only partly cut off or suppressed. Television broadcasts (in analog video formats) use this method if the video is transmitted in AM, due to the large bandwidth used. It may also be used in digital transmission, such as the ATSC standardized 8-VSB. The Milgo 4400/48 modem (circa 1967) used vestigial sideband and phase-shift keying to provide 4800-bit/s transmission over a 1600 Hz channel.

The video baseband signal used in TV in countries that use NTSC or ATSC has a bandwidth of 6 MHz. To conserve bandwidth, SSB would be desirable, but the video signal has significant low frequency content (average brightness) and has rectangular synchronising pulses. The engineering compromise is vestigial sideband modulation. In vestigial sideband the full upper sideband of bandwidth W2 = 4 MHz is transmitted, but only W1 = 1.25 MHz of the lower sideband is transmitted, along with a carrier. This effectively makes the system AM at low modulation frequencies and SSB at high modulation frequencies. The absence of the lower sideband components at high frequencies must be compensated for, and this is done by the RF and IF filters.

Chapter- 3

Quadrature Amplitude Modulation

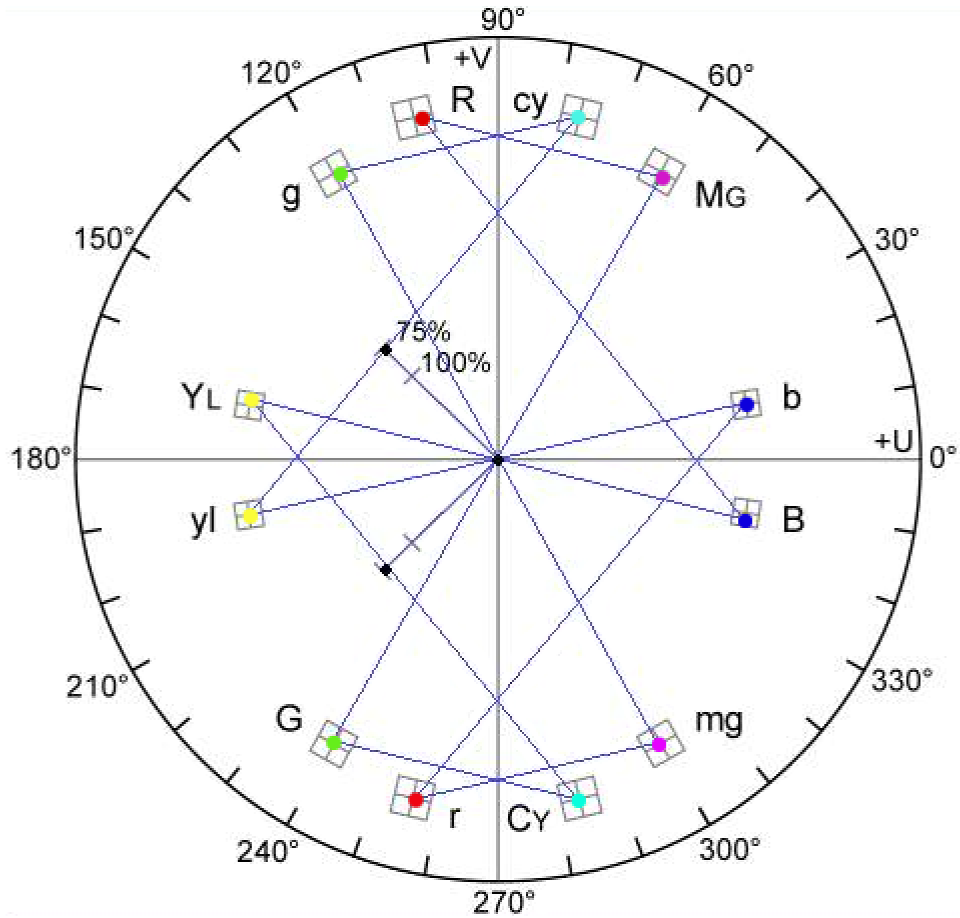
Quadrature amplitude modulation (**QAM**) is both an analog and a digital modulation scheme. It conveys two analog message signals, or two digital bit streams, by changing (*modulating*) the amplitudes of two carrier waves, using the amplitude-shift keying (ASK) digital modulation scheme or amplitude modulation (AM) analog modulation scheme. These two waves, usually sinusoids, are out of phase with each other by 90° and are thus called quadrature carriers or quadrature components — hence the name of the scheme. The modulated waves are summed, and the resulting waveform is a combination of both phase-shift keying (PSK) and amplitude-shift keying (ASK), or in the analog case of phase modulation (PM) and amplitude modulation. In the digital QAM case, a finite number of at least two phases, and at least two amplitudes are used. PSK modulators are often designed using the QAM principle, but are not considered as QAM since the amplitude of the modulated carrier signal is constant. QAM is used extensively as a modulation scheme for digital telecommunication systems.

Digital QAM

Like all modulation schemes, QAM conveys data by changing some aspect of a carrier signal, or the carrier wave, (usually a sinusoid) in response to a data signal. In the case of QAM, the amplitude of two waves, 90 degrees out-of-phase with each other (in quadrature) are changed (*modulated* or *keyed*) to represent the data signal. Amplitude modulating two carriers in quadrature can be equivalently viewed as both amplitude modulating and phase modulating a single carrier.

Phase modulation (analog PM) and phase-shift keying (digital PSK) can be regarded as a special case of QAM, where the magnitude of the modulating signal is a constant, with only the phase varying. This can also be extended to frequency modulation (FM) and frequency-shift keying (FSK), for these can be regarded as a special case of phase modulation.

Analog QAM



Analog QAM: measured PAL colour bar signal on a vector analyser screen.

When transmitting two signals by modulating them with QAM, the transmitted signal will be of the form:

$$s(t) = I(t)\cos(2\pi f_0 t) + Q(t)\sin(2\pi f_0 t),$$

where $I(t)$ and $Q(t)$ are the modulating signals and f_0 is the carrier frequency.

At the receiver, these two modulating signals can be demodulated using a coherent demodulator. Such a receiver multiplies the received signal separately with both a cosine and sine signal to produce the received estimates of $I(t)$ and $Q(t)$ respectively. Because of the orthogonality property of the carrier signals, it is possible to detect the modulating signals independently.

In the ideal case $I(t)$ is demodulated by multiplying the transmitted signal with a cosine signal:

$$\begin{aligned} r_i(t) &= s(t)\cos(2\pi f_0 t) \\ &= I(t)\cos(2\pi f_0 t)\cos(2\pi f_0 t) + Q(t)\sin(2\pi f_0 t)\cos(2\pi f_0 t) \end{aligned}$$

Using standard trigonometric identities, we can write it as:

$$\begin{aligned} r_i(t) &= \frac{1}{2}I(t)\left[1 + \cos(4\pi f_0 t)\right] + \frac{1}{2}Q(t)\sin(4\pi f_0 t) \\ &= \frac{1}{2}I(t) + \frac{1}{2}[I(t)\cos(4\pi f_0 t) + Q(t)\sin(4\pi f_0 t)] \end{aligned}$$

Low-pass filtering $r_i(t)$ removes the high frequency terms (containing $4\pi f_0 t$), leaving only the $I(t)$ term. This filtered signal is unaffected by $Q(t)$, showing that the in-phase component can be received independently of the quadrature component. Similarly, we may multiply $s(t)$ by a sine wave and then low-pass filter to extract $Q(t)$.

The phase of the received signal is assumed to be known accurately at the receiver. If the demodulating phase is even a little off, it results in crosstalk between the modulated signals. This issue of carrier synchronization at the receiver must be handled somehow in QAM systems. The coherent demodulator needs to be exactly in phase with the received signal, or otherwise the modulated signals cannot be independently received. For example analog television systems transmit a burst of the transmitting colour subcarrier after each horizontal synchronization pulse for reference.

Analog QAM is used in NTSC and PAL television systems, where the I- and Q-signals carry the components of chroma (colour) information. "Compatible QAM" or C-QUAM is used in AM stereo radio to carry the stereo difference information.

Fourier analysis of QAM

In the frequency domain, QAM has a similar spectral pattern to DSB-SC modulation. Using the properties of the Fourier transform, we find that:

$$S(f) = \frac{1}{2}\left[M_I(f - f_0) + M_I(f + f_0)\right] + \frac{1}{2j}\left[M_Q(f - f_0) - M_Q(f + f_0)\right]$$

where $S(f)$, $M_I(f)$ and $M_Q(f)$ are the Fourier transforms (frequency-domain representations) of $s(t)$, $I(t)$ and $Q(t)$, respectively.

Quantized QAM

As with many digital modulation schemes, the constellation diagram is a useful representation. In QAM, the constellation points are usually arranged in a square grid with equal vertical and horizontal spacing, although other configurations are possible (e.g. Cross-QAM). Since in digital telecommunications the data are usually binary, the number of points in the grid is usually a power of 2 (2, 4, 8 ...). Since QAM is usually square, some of these are rare—the most common forms are 16-QAM, 64-QAM, 128-QAM and 256-QAM. By moving to a higher-order constellation, it is possible to transmit more bits per symbol. However, if the mean energy of the constellation is to remain the same (by way of making a fair comparison), the points must be closer together and are thus more susceptible to noise and other corruption; this results in a higher bit error rate and so higher-order QAM can deliver more data less reliably than lower-order QAM, for constant mean constellation energy.

If data-rates beyond those offered by 8-PSK are required, it is more usual to move to QAM since it achieves a greater distance between adjacent points in the I-Q plane by distributing the points more evenly. The complicating factor is that the points are no longer all the same amplitude and so the demodulator must now correctly detect both phase and amplitude, rather than just phase.

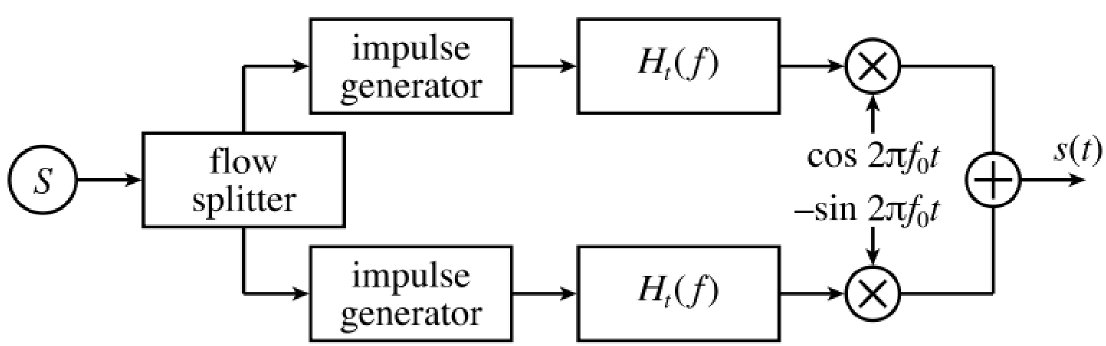
64-QAM and 256-QAM are often used in digital cable television and cable modem applications. In the United States, 64-QAM and 256-QAM are the mandated modulation schemes for digital cable as standardised by the SCTE in the standard ANSI/SCTE 07 2000. Note that many marketing people will refer to these as QAM-64 and QAM-256. In the UK, 16-QAM and 64-QAM are currently used for digital terrestrial television (Freeview and Top Up TV) and 256-QAM is planned for Freeview-HD.

Communication systems designed to achieve very high levels of spectral efficiency usually employ very dense QAM constellations. One example is the ITU-T G.hn standard for networking over existing home wiring (coaxial cable, phone lines and power lines), which employs constellations up to 4096-QAM (12 bits/symbol). Another example is VDSL2 technology for copper twisted pairs, whose constellation size goes up to 32768 points.

Ideal structure

Transmitter

The following picture shows the ideal structure of a QAM transmitter, with a carrier frequency f_0 and the frequency response of the transmitter's filter H_t:



First the flow of bits to be transmitted is split into two equal parts: this process generates two independent signals to be transmitted. They are encoded separately just like they were in an amplitude-shift keying (ASK) modulator. Then one channel (the one "in phase") is multiplied by a cosine, while the other channel (in "quadrature") is multiplied by a sine. This way there is a phase of 90° between them. They are simply added one to the other and sent through the real channel.

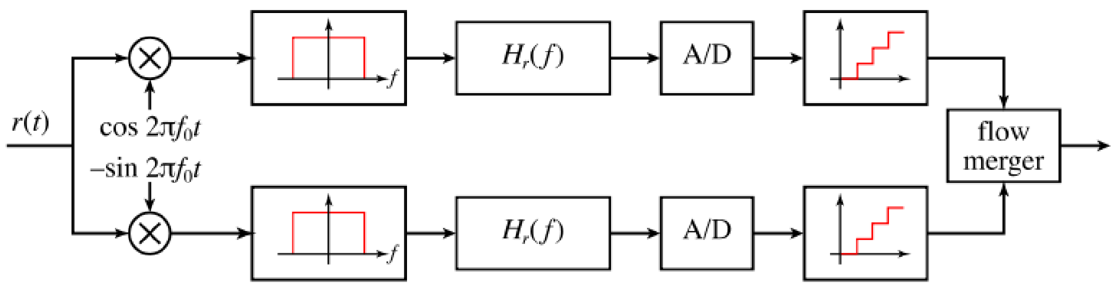
The sent signal can be expressed in the form:

$$s(t) = \sum_{n=-\infty}^{\infty} [v_c[n] \cdot h_t(t - nT_s) \cos(2\pi f_0 t) - v_s[n] \cdot h_t(t - nT_s) \sin(2\pi f_0 t)],$$

where $v_c[n]$ and $v_s[n]$ are the voltages applied in response to the n^{th} symbol to the cosine and sine waves respectively.

Receiver

The receiver simply performs the inverse process of the transmitter. Its ideal structure is shown in the picture below with H_r the receive filter's frequency response :



Multiplying by a cosine (or a sine) and by a low-pass filter it is possible to extract the component in phase (or in quadrature). Then there is only an ASK demodulator and the two flows of data are merged back.

In practice, there is an unknown phase delay between the transmitter and receiver that must be compensated by *synchronization* of the receivers local oscillator, i.e. the sine and cosine functions in the above figure. In mobile applications, there will often be an offset in the relative *frequency* as well, due to the possible presence of a Doppler shift proportional to the relative velocity of the transmitter and receiver. Both the phase and frequency variations introduced by the channel must be compensated by properly tuning the sine and cosine components, which requires a *phase reference*, and is typically accomplished using a Phase-Locked Loop (PLL).

In any application, the low-pass filter will be within $h_r(t)$: here it was shown just to be clearer.

Quantized QAM performance

The following definitions are needed in determining error rates:

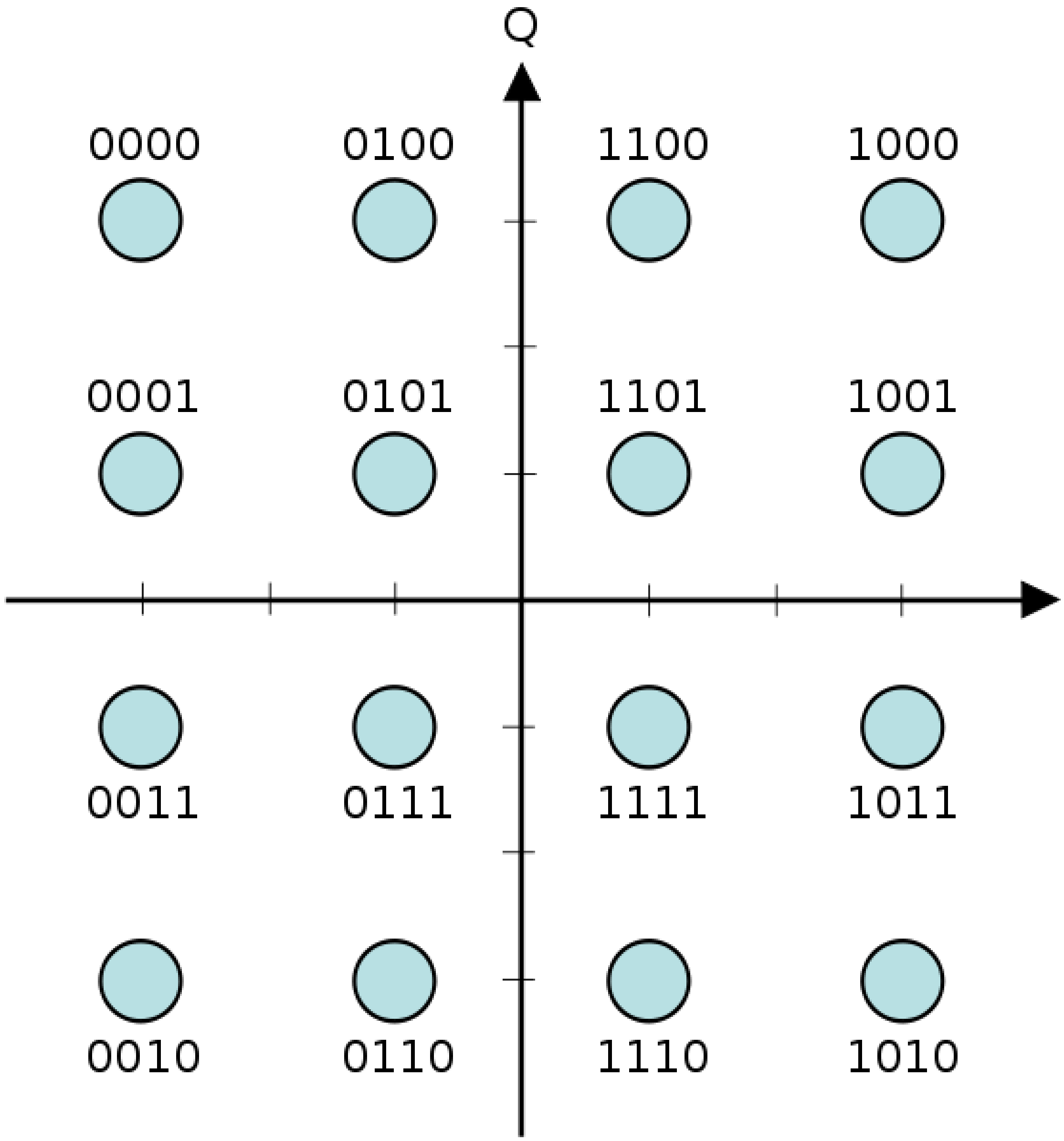
- M = Number of symbols in modulation constellation
- E_b = Energy-per-bit
- E_s = Energy-per-symbol = kE_b with k bits per symbol
- N_0 = Noise power spectral density (W/Hz)
- P_b = Probability of bit-error
- P_{bc} = Probability of bit-error per carrier
- P_s = Probability of symbol-error
- P_{sc} = Probability of symbol-error per carrier
- $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} dt, \ x \geq 0$.

$Q(x)$ is related to the complementary Gaussian error function by:
$Q(x) = \frac{1}{2}\operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right)$, which is the probability that x will be under the tail of the Gaussian PDF towards positive infinity.

The error rates quoted here are those in additive white Gaussian noise (AWGN).

Where coordinates for constellation points are given here, note that they represent a *non-normalised* constellation. That is, if a particular mean average energy were required (e.g. unit average energy), the constellation would need to be linearly scaled.

Rectangular QAM



Constellation diagram for rectangular 16-QAM.

Rectangular QAM constellations are, in general, sub-optimal in the sense that they do not maximally space the constellation points for a given energy. However, they have the considerable advantage that they may be easily transmitted as two pulse amplitude modulation (PAM) signals on quadrature carriers, and can be easily demodulated. The non-square constellations, dealt with below, achieve marginally better bit-error rate (BER) but are harder to modulate and demodulate.

The first rectangular QAM constellation usually encountered is 16-QAM, the constellation diagram for which is shown here. A Gray coded bit-assignment is also given. The reason that 16-QAM is usually the first is that a brief consideration reveals that 2-QAM and 4-QAM are in fact binary phase-shift keying (BPSK) and quadrature phase-shift keying (QPSK), respectively. Also, the error-rate performance of 8-QAM is close to that of 16-QAM (only about 0.5 dB better), but its data rate is only three-quarters that of 16-QAM.

Expressions for the symbol-error rate of rectangular QAM are not hard to derive but yield rather unpleasant expressions. For an even number of bits per symbol, *k*, exact expressions are available. They are most easily expressed in a *per carrier* sense:

$$P_{sc} = 2\left(1 - \frac{1}{\sqrt{M}}\right) Q\left(\sqrt{\frac{3}{M-1}\frac{E_s}{N_0}}\right),$$

so

$$P_s = 1 - \left(1 - P_{sc}\right)^2.$$

The bit-error rate depends on the bit to symbol mapping, but for $E_b/N_0 \gg 1$ and a Gray-coded assignment—so that we can assume each symbol error causes only one bit error—the bit-error rate is approximately

$$P_{bc} \approx P_{sc}/(k/2) = \frac{4}{k}\left(1 - \frac{1}{\sqrt{M}}\right) Q\left(\sqrt{\frac{3k}{M-1}\frac{E_b}{N_0}}\right).$$

Since the carriers are independent, the overall bit error rate is the same as the per-carrier error rate, just like BPSK and QPSK.

$$P_b = P_{bc}.$$

Odd-*k* QAM

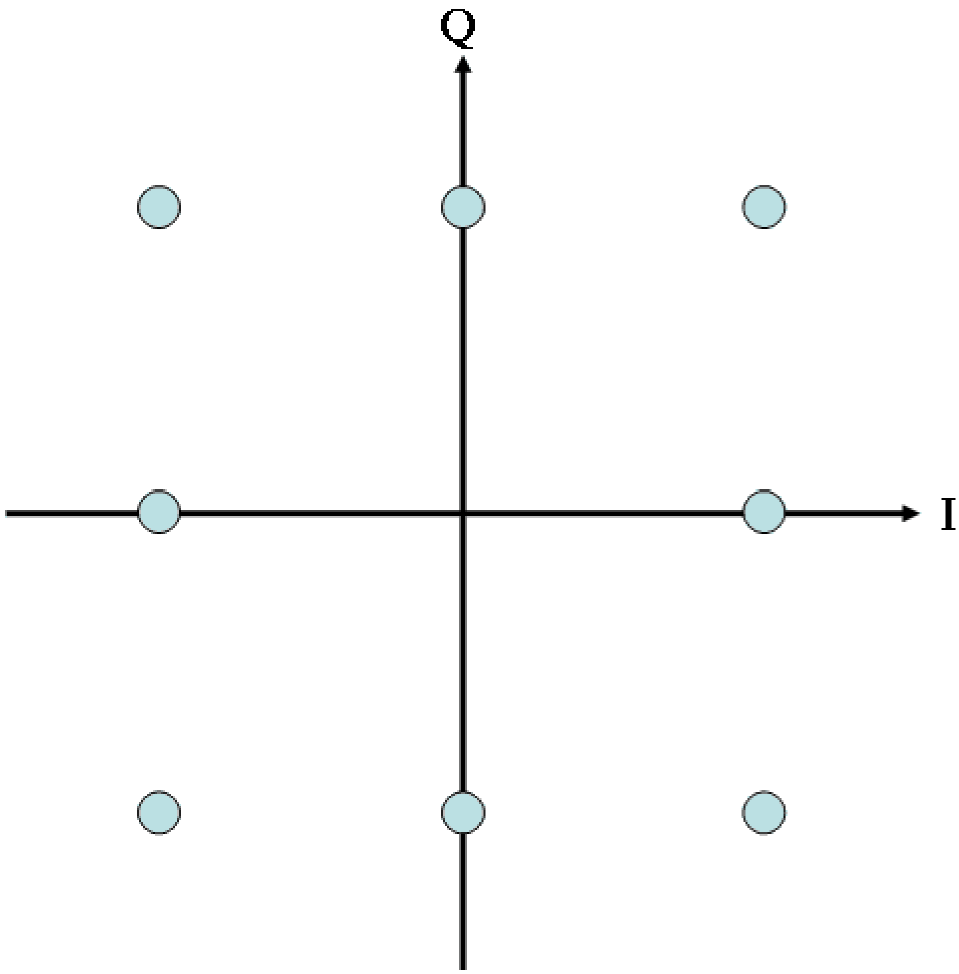
For odd *k*, such as 8-QAM (*k* = 3) it is harder to obtain symbol-error rates, but a tight upper bound is:

$$P_s \leq 4Q\left(\sqrt{\frac{3kE_b}{(M-1)N_0}}\right).$$

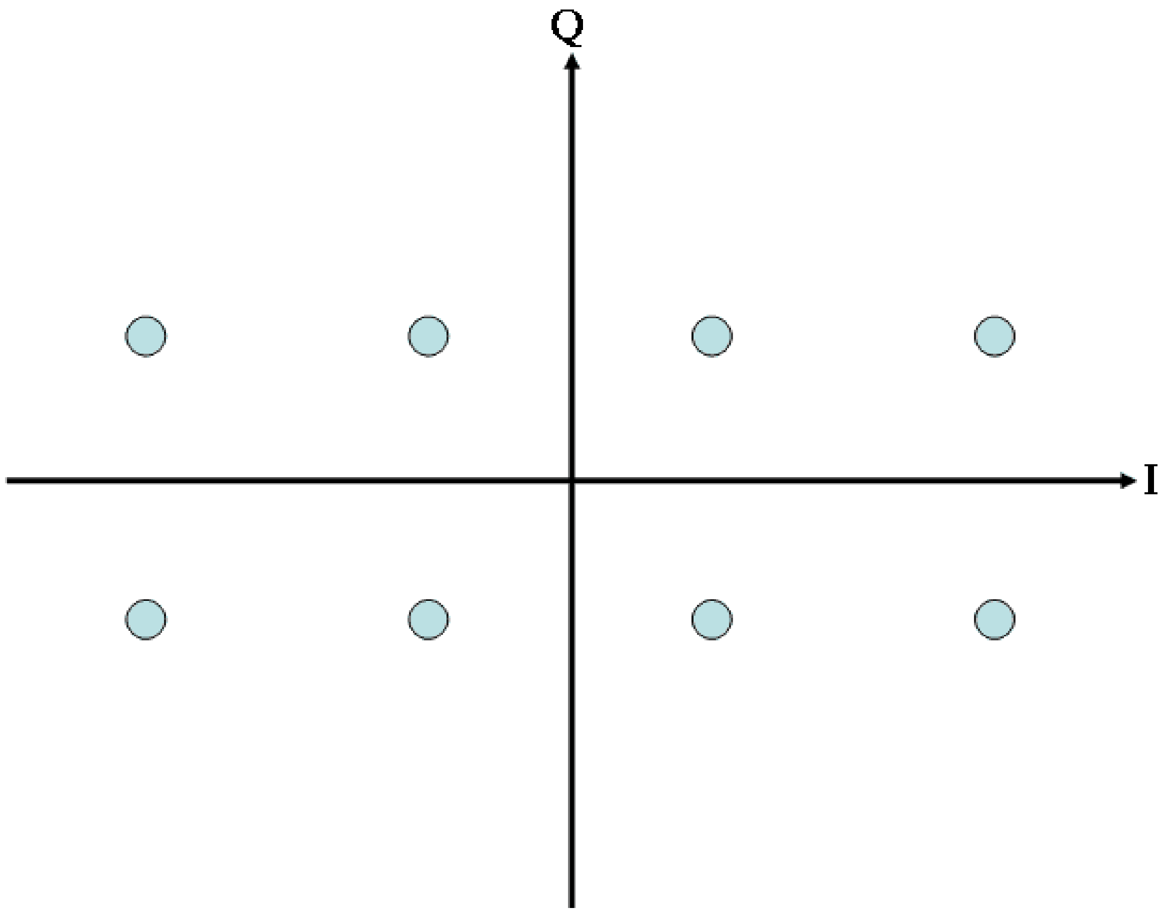
Two rectangular 8-QAM constellations are shown below without bit assignments. These both have the same minimum distance between symbol points, and thus the same symbol-error rate (to a first approximation).

The exact bit-error rate, P_b will depend on the bit-assignment.

Note that both of these constellations are seldom used in practice, as the non-rectangular version of 8-QAM is optimal. Example of second constellation's usage: LDPC and 8-QAM.

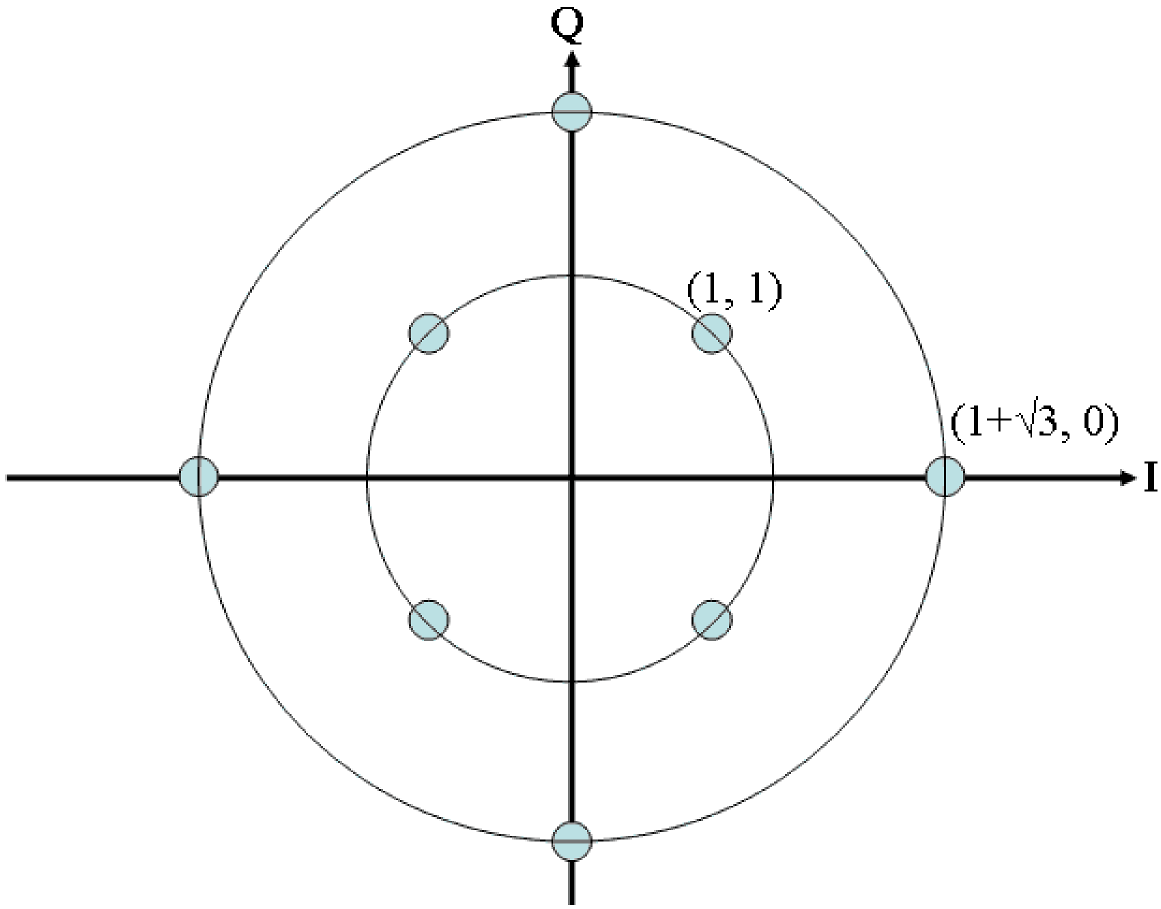


Constellation diagram for rectangular 8-QAM.

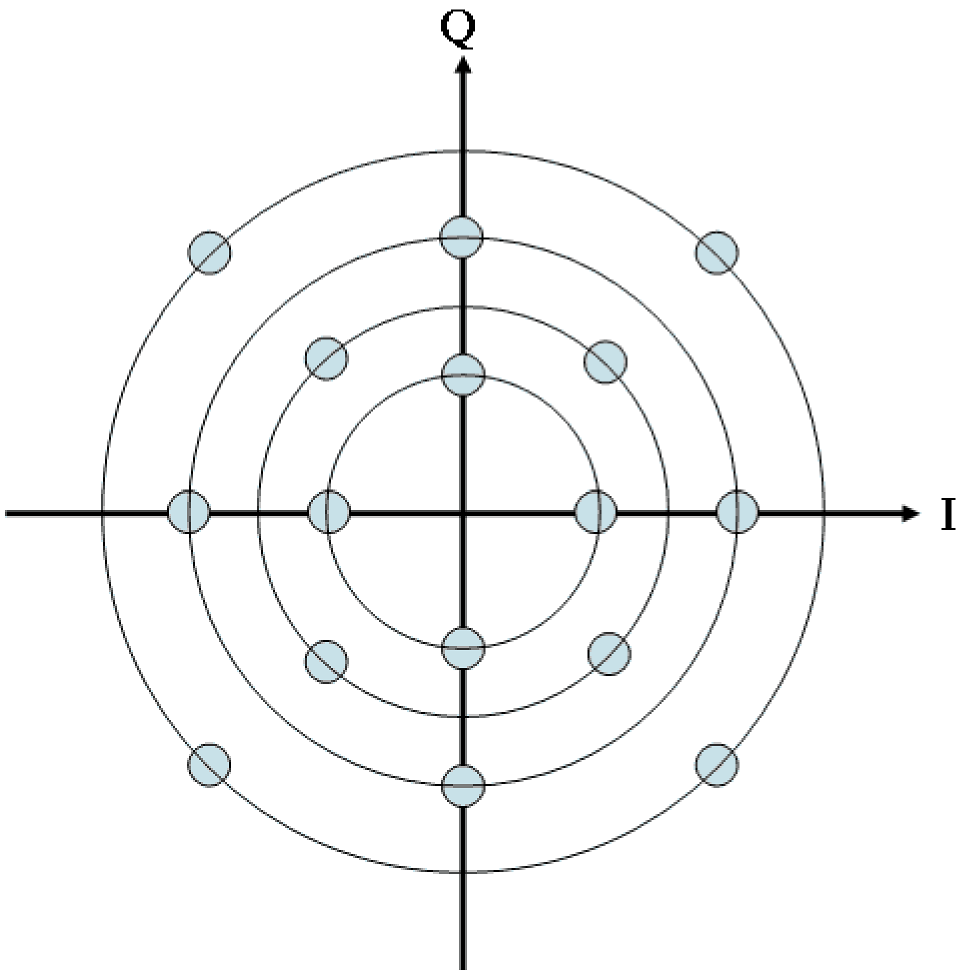


Alternative constellation diagram for rectangular 8-QAM.

Non-rectangular QAM



Constellation diagram for circular 8-QAM.



Constellation diagram for circular 16-QAM.

It is the nature of QAM that most orders of constellations can be constructed in many different ways and it is neither possible nor instructive to cover them all here.

Two diagrams of circular QAM constellation are shown, for 8-QAM and 16-QAM. The circular 8-QAM constellation is known to be the optimal 8-QAM constellation in the sense of requiring the least mean power for a given minimum Euclidean distance. The 16-QAM constellation is suboptimal although the optimal one may be constructed along the same lines as the 8-QAM constellation. The circular constellation highlights the relationship between QAM and PSK. Other orders of constellation may be constructed along similar (or very different) lines. It is consequently hard to establish expressions for the error rates of non-rectangular QAM since it necessarily depends on the constellation. Nevertheless, an obvious upper bound to the rate is related to the minimum Euclidean distance of the constellation (the shortest straight-line distance between two points):

$$P_s < (M-1)Q\left(\sqrt{d_{min}^2/2N_0}\right).$$

Again, the bit-error rate will depend on the assignment of bits to symbols.

Although, in general, there is a non-rectangular constellation that is optimal for a particular *M*, they are not often used since the rectangular QAMs are much easier to modulate and demodulate.

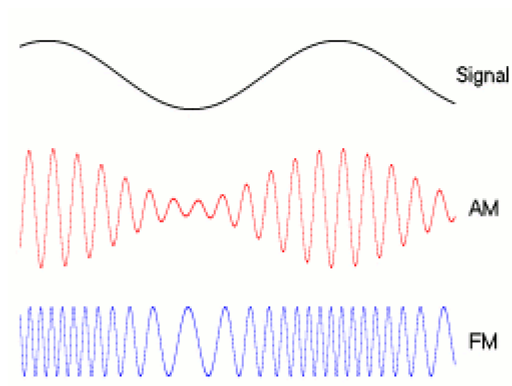
Interference and noise

In moving to a higher order QAM constellation (higher data rate and mode) in hostile RF/microwave QAM application environments, such as in broadcasting or telecommunications, multipath interference typically increases. Reduced noise immunity due to constellation separation makes it difficult to achieve theoretical performance thresholds. There are several test parameter measurements which help determine an optimal QAM mode for a specific operating environment. The following three are most significant:

- Carrier/interference ratio
- Carrier-to-noise ratio
- Threshold-to-noise ratio

Chapter- 4

Frequency Modulation

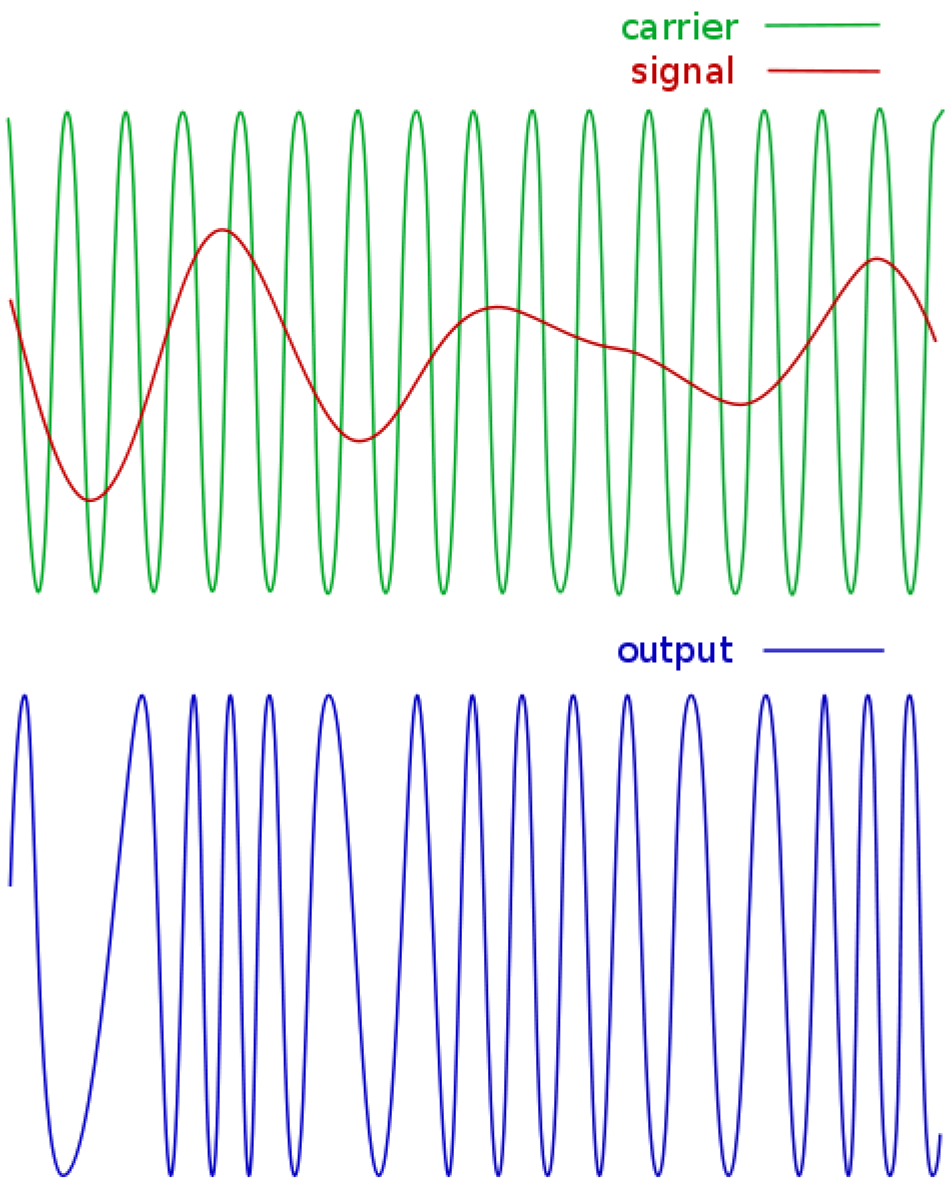


A signal may be carried by an AM or FM radio wave.

In telecommunications and signal processing, **frequency modulation** (**FM**) conveys information over a carrier wave by varying its instantaneous frequency. This is in contrast with amplitude modulation, in which the amplitude of the carrier is varied while its frequency remains constant. In analog applications, the difference between the instantaneous and the base frequency of the carrier is directly proportional to the instantaneous value of the input signal amplitude. Digital data can be sent by shifting the carrier's frequency among a set of discrete values, a technique known as frequency-shift keying.

Frequency modulation can be regarded as phase modulation where the carrier phase modulation is the time integral of the FM modulating signal.

FM is widely used for broadcasting of music and speech, and in two-way radio systems, in magnetic tape recording systems, and certain video transmission systems. In radio systems, frequency modulation with sufficient bandwidth provides an advantage in cancelling naturally-occurring noise. Frequency-shift keying (digital FM) is widely used in data and fax modems.



A signal modifies the frequency of a carrier in FM.

Theory

Suppose the baseband data signal (the message) to be transmitted is $x_m(t)$ and the sinusoidal carrier is $x_c(t) = A_c \cos(2\pi f_c t)$, where f_c is the carrier's base frequency and A_c is the carrier's amplitude. The modulator combines the carrier with the baseband data signal to get the transmitted signal:

$$y(t) = A_c \cos\left(2\pi \int_0^t f(\tau) d\tau\right)$$

$$= A_c \cos\left(2\pi \int_0^t [f_c + f_\Delta x_m(\tau)]\, d\tau\right)$$
$$= A_c \cos\left(2\pi f_c t + 2\pi f_\Delta \int_0^t x_m(\tau) d\tau\right)$$

In this equation, $f(\tau)$ is the *instantaneous frequency* of the oscillator and f_Δ is the *frequency deviation*, which represents the maximum shift away from f_c in one direction, assuming $x_m(t)$ is limited to the range ±1.

Although it may seem that this limits the frequencies in use to $f_c \pm f_\Delta$, this neglects the distinction between *instantaneous frequency* and *spectral frequency*. The frequency spectrum of an actual FM signal has components extending out to infinite frequency, although they become negligibly small beyond a point.

Sinusoidal baseband signal

While it is an over-simplification, a baseband modulated signal may be approximated by a sinusoidal Continuous Wave signal with a frequency f_m. The integral of such a signal is

$$\int_0^t x_m(t) dt = \frac{A_m \cos(2\pi f_m t)}{2\pi f_m}$$

Thus, in this specific case, equation (1) above simplifies to:

$$y(t) = A_c \cos\left(2\pi f_c t + \frac{f_\Delta}{f_m} \cos\left(2\pi f_m t\right)\right)$$

where the amplitude A_m of the modulating sinusoid, is represented by the peak deviation f_Δ.

The harmonic distribution of a sine wave carrier modulated by such a sinusoidal signal can be represented with Bessel functions - this provides a basis for a mathematical understanding of frequency modulation in the frequency domain.

Modulation index

As with other modulation indices, this quantity indicates by how much the modulated variable varies around its unmodulated level. It relates to the variations in the frequency of the carrier signal:

$$h = \frac{\Delta f}{f_m} = \frac{f_\Delta |x_m(t)|}{f_m}$$

where f_m is the highest frequency component present in the modulating signal $x_m(t)$, and Δf is the Peak frequency-deviation, i.e. the maximum deviation of the *instantaneous frequency* from the carrier frequency. If $h \ll 1$, the modulation is called *narrowband FM*, and its bandwidth is approximately $2f_m$. If $h \gg 1$, the modulation is called *wideband FM* and its bandwidth is approximately $2f_\Delta$. While wideband FM uses more bandwidth, it can improve signal-to-noise ratio significantly.

With a tone-modulated FM wave, if the modulation frequency is held constant and the modulation index is increased, the (non-negligible) bandwidth of the FM signal increases, but the spacing between spectra stays the same; some spectral components decrease in strength as others increase. If the frequency deviation is held constant and the modulation frequency increased, the spacing between spectra increases.

Frequency modulation can be classified as narrow band if the change in the carrier frequency is about the same as the signal frequency, or as wide-band if the change in the carrier frequency is much higher (modulation index >1) than the signal frequency. For example, narrowband FM is used for two way radio systems such as Family Radio Service where the carrier is allowed to deviate only 2.5 kHz above and below the center frequency, carrying speech signals of no more than 3.5 kHz bandwidth. Wide-band FM is used for FM broadcasting where music and speech is transmitted with up to 75 kHz deviation from the center frequency, carrying audio with up to 20 kHz bandwidth.

Carson's rule

A rule of thumb, *Carson's rule* states that nearly all (~98%) of the power of a frequency-modulated signal lies within a bandwidth B_T of

$$B_T = 2(\Delta f + f_m)$$

where Δf, as defined above, is the peak deviation of the instantaneous frequency $f(t)$ from the center carrier frequency f_c.

Noise quieting

The noise power decreases as the signal power increases; therefore the SNR goes up significantly.

Bessel functions

The carrier and sideband amplitudes are illustrated for different modulation indices of FM signals. Based on the Bessel functions.

Modulation index	Carrier	Sideband 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

0.00	1.00																
0.25	0.98	0.12															
0.5	0.94	0.24	0.03														
1.0	0.77	0.44	0.11	0.02													
1.5	0.51	0.56	0.23	0.06	0.01												
2.0	0.22	0.58	0.35	0.13	0.03												
2.41	0	0.52	0.43	0.20	0.06	0.02											
2.5	−0.05	0.50	0.45	0.22	0.07	0.02	0.01										
3.0	−0.26	0.34	0.49	0.31	0.13	0.04	0.01										
4.0	−0.40	−0.07	0.36	0.43	0.28	0.13	0.05	0.02									
5.0	−0.18	−0.33	0.05	0.36	0.39	0.26	0.13	0.05	0.02								
5.53	0	−0.34	−0.13	0.25	0.40	0.32	0.19	0.09	0.03	0.01							
6.0	0.15	−0.28	−0.24	0.11	0.36	0.36	0.25	0.13	0.06	0.02							
7.0	0.30	0.00	−0.30	−0.17	0.16	0.35	0.34	0.23	0.13	0.06	0.02						
8.0	0.17	0.23	−0.11	−0.29	−0.10	0.19	0.34	0.32	0.22	0.13	0.06	0.03					
8.65	0	0.27	0.06	−0.24	−0.23	0.03	0.26	0.34	0.28	0.18	0.10	0.05	0.02				
9.0	−0.09	0.25	0.14	−0.18	−0.27	−0.06	0.20	0.33	0.31	0.21	0.12	0.06	0.03	0.01			
10.0	−0.25	0.04	0.25	0.06	−0.22	−0.23	−0.01	0.22	0.32	0.29	0.21	0.12	0.06	0.03	0.01		
12.0	0.05	−0.22	−0.08	0.20	0.18	−0.07	−0.24	−0.17	0.05	0.23	0.30	0.27	0.20	0.12	0.07	0.03	0.01

Practical Implementation

Modulation

FM signals can be generated using either direct or indirect frequency modulation.

- Direct FM modulation can be achieved by directly feeding the message into the input of a VCO.
- For indirect FM modulation, the message signal is integrated to generate a phase modulated signal. This is used to modulate a crystal controlled oscillator, and the result is passed through a frequency multiplier to give an FM signal.

Demodulation

Many FM detector circuits exist. One common method for recovering the information signal is through a Foster-Seeley discriminator. A phase-locked loop can be used as an FM demodulator.

Slope detection demodulates an FM signal by using a tuned circuit, which has its resonant frequency slightly offset from the carrier frequency. As the frequency rises and falls, the tuned circuit provides a changing amplitude of response, converting FM to AM. AM receivers may detect some FM transmissions by this means, though it does not provide an efficient method of detection for FM broadcasts.

Applications

Magnetic tape storage

FM is also used at intermediate frequencies by all analog VCR systems, including VHS, to record both the luminance (black and white) and the chrominance portions of the video signal. FM is the only feasible method of recording video to and retrieving video from Magnetic tape without extreme distortion, as video signals have a very large range of frequency components — from a few hertz to several megahertz, too wide for equalizers to work with due to electronic noise below −60 dB. FM also keeps the tape at saturation level, and therefore acts as a form of noise reduction, and a simple limiter can mask variations in the playback output, and the FM capture effect removes print-through and pre-echo. A continuous pilot-tone, if added to the signal — as was done on V2000 and many Hi-band formats — can keep mechanical jitter under control and assist timebase correction.

These FM systems are unusual in that they have a ratio of carrier to maximum modulation frequency of less than two; contrast this with FM audio broadcasting where the ratio is around 10,000. Consider for example a 6 MHz carrier modulated at a 3.5 MHz rate; by Bessel analysis the first sidebands are on 9.5 and 2.5 MHz, while the second sidebands are on 13 MHz and −1 MHz. The result is a sideband of reversed phase on +1 MHz; on demodulation, this results in an unwanted output at 6−1 = 5 MHz. The system must be designed so that this is at an acceptable level.

Sound

FM is also used at audio frequencies to synthesize sound. This technique, known as FM synthesis, was popularized by early digital synthesizers and became a standard feature for several generations of personal computer sound cards.

Radio

Edwin Howard Armstrong (1890–1954) was an American electrical engineer who invented frequency modulation (FM) radio. He patented the regenerative circuit in 1914, the superheterodyne receiver in 1918 and the super-regenerative circuit in 1922. He presented his paper: "A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation", which first described FM radio, before the New York section of the Institute of Radio Engineers on November 6, 1935. The paper was published in 1936.

As the name implies, wideband FM (WFM) requires a wider signal bandwidth than amplitude modulation by an equivalent modulating signal, but this also makes the signal more robust against noise and interference. Frequency modulation is also more robust against simple signal amplitude fading phenomena. As a result, FM was chosen as the modulation standard for high frequency, high fidelity radio transmission: hence the term

"FM radio" (although for many years the BBC called it "VHF radio", because commercial FM broadcasting uses a well-known part of the VHF band—the FM broadcast band).

FM receivers employ a special detector for FM signals and exhibit a phenomenon called **capture effect**, where the tuner is able to clearly receive the stronger of two stations being broadcast on the same frequency. Problematically however, frequency drift or lack of selectivity may cause one station or signal to be suddenly overtaken by another on an adjacent channel. Frequency drift typically constituted a problem on very old or inexpensive receivers, while inadequate selectivity may plague any tuner.

A high-efficiency radio-frequency switching amplifier can be used to transmit FM signals (and other constant-amplitude signals). For a given signal strength (measured at the receiver antenna), switching amplifiers use less battery power and typically cost less than a linear amplifier. This gives FM another advantage over other modulation schemes that require linear amplifiers, such as AM and QAM.

FM is commonly used at VHF radio frequencies for high-fidelity broadcasts of music and speech. Normal (analog) TV sound is also broadcast using FM. A narrow band form is used for voice communications in commercial and amateur radio settings. In broadcast services, where audio fidelity is important, wideband FM is generally used. In two-way radio, narrowband FM (NBFM) is used to conserve bandwidth for land mobile radio stations, marine mobile, and many other radio services.

Miscellaneous

Frequency-shift keying is the frequency modulation using only a discrete number of frequencies. Morse code transmission has been implemented this way, as were most early telephone-line modems. Radio teletype also use FSK.

FM modulation is also used in telemetry applications, radar, seismic prospecting and newborn EEG seizures modelling.

Chapter- 5

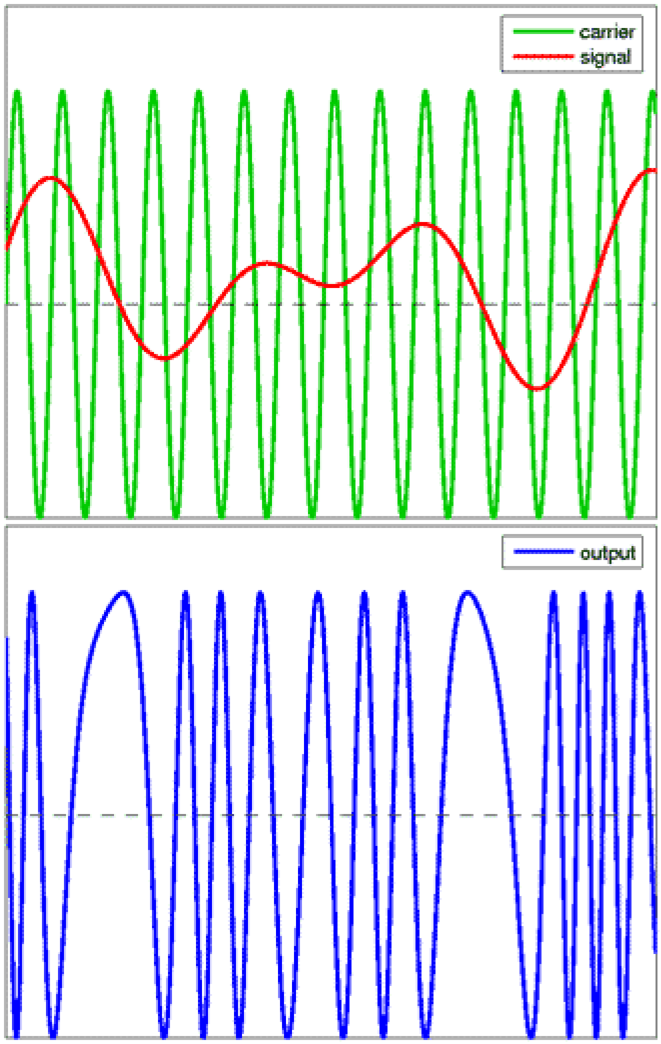
Phase Modulation & Space Modulation

Phase modulation

Phase modulation (PM) is a form of modulation that represents information as variations in the instantaneous phase of a carrier wave.

Unlike its more popular counterpart, frequency modulation (FM), PM is not very widely used for radio transmissions. This is because it tends to require more complex receiving hardware and there can be ambiguity problems in determining whether, for example, the signal has changed phase by +180° or -180°. PM is used, however, in digital music synthesizers such as the Yamaha DX7, even though these instruments are usually referred to as "FM" synthesizers (both modulation types sound very similar, but PM is usually easier to implement in this area).

Theory



An example of phase modulation. The top diagram shows the modulating signal superimposed on the carrier wave. The bottom diagram shows the resulting phase-modulated signal.

PM changes the phase angle of the complex envelope in direct proportion to the message signal.

Suppose that the signal to be sent (called the modulating or message signal) is $m(t)$ and the carrier onto which the signal is to be modulated is

$$c(t) = A_c \sin\left(\omega_c t + \phi_c\right).$$

Annotated:

carrier(time) = (carrier amplitude)*sin(carrier frequency*time + phase shift)

This makes the modulated signal

$$y(t) = A_c \sin\left(\omega_c t + m(t) + \phi_c\right).$$

This shows how *m*(*t*) modulates the phase - the greater m(t) is at a point in time, the greater the phase shift of the modulated signal at that point. It can also be viewed as a change of the frequency of the carrier signal, and phase modulation can thus be considered a special case of FM in which the carrier frequency modulation is given by the time derivative of the phase modulation.

The spectral behaviour of phase modulation is difficult to derive, but the mathematics reveals that there are two regions of particular interest:

- For small amplitude signals, PM is similar to amplitude modulation (AM) and exhibits its unfortunate doubling of baseband bandwidth and poor efficiency.
- For a single large sinusoidal signal, PM is similar to FM, and its bandwidth is approximately

  $$2\left(h+1\right)f_{\mathrm{M}},$$

  where $f_M = \omega_m / 2\pi$ and *h* is the modulation index defined below. This is also known as Carson's Rule for PM.

Modulation index

As with other modulation indices, this quantity indicates by how much the modulated variable varies around its unmodulated level. It relates to the variations in the phase of the carrier signal:

$$h = \Delta\theta,$$

where $\Delta\theta$ is the peak phase deviation. Compare to the modulation index for frequency modulation.

Space modulation

Space modulation is a radio Amplitude Modulation technique used in Instrument Landing Systems that incorporates the use of multiple antennas fed with various radio frequency powers and phases to create different depths of modulation within various volumes of three-dimensional airspace. This modulation method differs from internal modulation methods inside most other radio transmitters in that the phases and powers of the two individual signals mix within airspace, rather than in a modulator.

An aircraft with an on-board ILS receiver within the capture area of an ILS, (glideslope and localiser range), will detect varying depths of modulation according to the aircraft's position within that airspace, providing accurate positional information about the progress to the threshold.

Method used to determine aircraft position

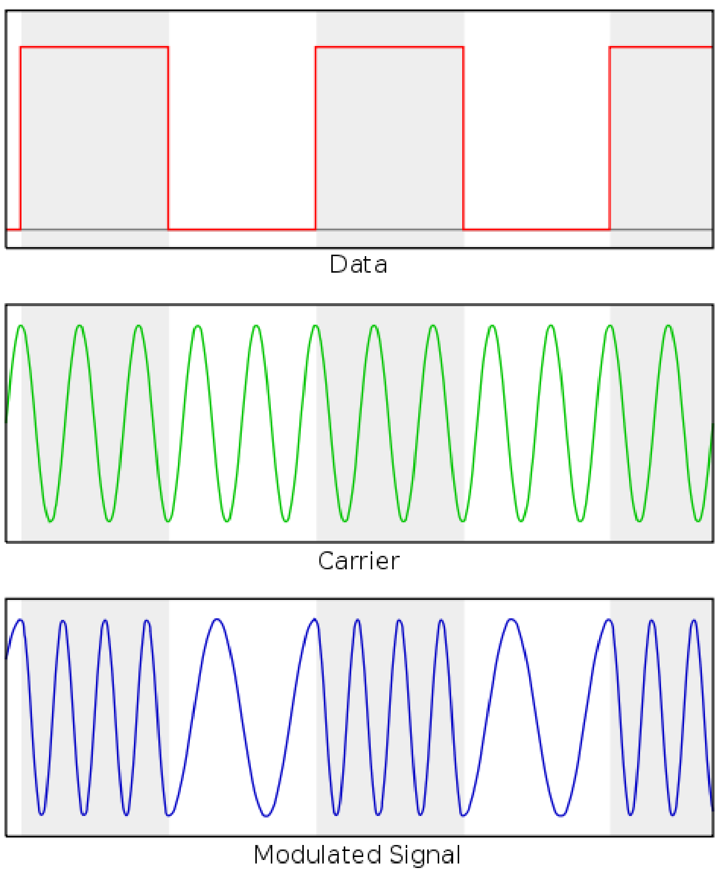
The ILS uses two radio frequencies, one for each ground station (about 110 MHz for LLZ and 330 MHz for the GP), to transmit two Amplitude Modulated signals (90 Hz and 150 Hz), along the glidepath (GP) and the localiser (LLZ) trajectories into airspace. it is this signal that is projected up from the runway which an aircraft using an instrument approach, uses to land.

The modulation depth of each 90 Hz and 150 Hz signal changes according to the deviation of the aircraft from the correct position for the aircraft to touchdown on the threshold. The difference between the two signal modulation depths is zero when the aircraft is on the correct glideslope on approach to the runway—i.e. No difference (zero DDM), produces no deviation from the middle indication of the instrument's needle within the cockpit of the aircraft.

Chapter- 6

Frequency-Shift Keying & Multiple Frequency-Shift Keying

Frequency-shift keying



An example of binary FSK

Frequency-shift keying (FSK) is a frequency modulation scheme in which digital information is transmitted through discrete frequency changes of a carrier wave. The simplest FSK is *binary FSK* (BFSK). BFSK literally implies using a pair of discrete frequencies to transmit binary (0s and 1s) information. With this scheme, the "1" is called the mark frequency and the "0" is called the space frequency. The time domain of an FSK modulated carrier is illustrated in the figures to the right.

Other forms of FSK

Minimum-shift keying

Minimum frequency-shift keying or *minimum-shift keying* (MSK) is a particular spectrally efficient form of coherent FSK. In MSK the difference between the higher and lower frequency is identical to half the bit rate. Consequently, the waveforms used to represent a 0 and a 1 bit differ by exactly half a carrier period. This is the smallest FSK modulation index that can be chosen such that the waveforms for 0 and 1 are orthogonal. A variant of MSK called GMSK is used in the GSM mobile phone standard.

Audio FSK

Audio frequency-shift keying (AFSK) is a modulation technique by which digital data is represented by changes in the frequency (pitch) of an audio tone, yielding an encoded signal suitable for transmission via radio or telephone. Normally, the transmitted audio alternates between two tones: one, the "mark", represents a binary one; the other, the "space", represents a binary zero.

AFSK differs from regular frequency-shift keying in performing the modulation at baseband frequencies. In radio applications, the AFSK-modulated signal normally is being used to modulate an RF carrier (using a conventional technique, such as AM or FM) for transmission.

AFSK is not always used for high-speed data communications, since it is far less efficient in both power and bandwidth than most other modulation modes. In addition to its simplicity, however, AFSK has the advantage that encoded signals will pass through AC-coupled links, including most equipment originally designed to carry music or speech.

Applications

Most early telephone-line modems used audio frequency-shift keying to send and receive data, up to rates of about 300 bits per second. The common Bell 103 modem used this technique, for example. Even today, North American caller ID uses 1200 baud AFSK in the form of the Bell 202 standard. Some early microcomputers used a specific form of AFSK modulation, the Kansas City standard, to store data on audio cassettes. AFSK is still widely used in amateur radio, as it allows data transmission through unmodified voiceband equipment. Radio control gear uses FSK, but calls it FM and PPM instead.

AFSK is also used in the United States' Emergency Alert System to transmit warning information. It is used at higher bitrates for Weathercopy used on Weatheradio by NOAA in the U.S., and more extensively by Environment Canada.

The CHU shortwave radio station in Ottawa, Canada broadcasts an exclusive digital time signal encoded using AFSK modulation.

Multiple frequency-shift keying

Multiple frequency-shift keying (MFSK) is a variation of frequency-shift keying (FSK) that uses more than two frequencies. MFSK is a form of M-ary orthogonal modulation, where each symbol consists of one element from an alphabet of orthogonal waveforms. M, the size of the alphabet, is usually a power of two so that each symbol represents $\log_2 M$ bits.

- M is usually between 2 and 64
- Error Correction is generally also used

How it works

Like other M-ary orthogonal schemes, the required E_b/N_0 ratio for a given probability of error decreases as M increases without the need for multisymbol coherent detection.

In fact, as M approaches infinity the required E_b/N_0 ratio decreases asymptotically to the Shannon limit of -1.6 dB. However this decrease is slow with increasing M, and large values are impractical because of the exponential increase in required bandwidth. Typical values in practice range from 4 to 64, and MFSK is combined with another forward error correction scheme to provide additional (systematic) coding gain.

Types

Defined examples of a multiple frequency-shift keying system include *dual-tone multi-frequency* (DTMF), which is used in touch tone phones and the Multi-frequency trunk signals used in Twentieth Century telephone exchanges.

These signals are distinctive when received auraly. Their main feature is a rapid succession of tones with almost musical quality.

HF communications

MFSK or "polytone" modes used for shortwave communications:

- MFSK8
- MFSK16
- Olivia MFSK

- Coquelet
- Piccolo
- ALE (MIL-STD 188-141)
- DominoF
- DominoEX
- THROB
- CIS-36 MFSK or CROWD-36
- XPA, XPA2

Piccolo was the original MFSK mode, developed for British government communications by Harold Robins, Donald Bailey and Denis Ralphs of the Diplomatic Wireless Service (DWS), a branch of the Foreign and Commonwealth Office. It was first used in 1962 and presented to the IEE in 1963. The current specification "Piccolo Mark IV" is still in limited use by the UK government, mainly for point-to-point military radio communications.

Coquelet is a similar modulation system developed by the French Government for similar applications.

MFSK8 and MFSK16 were developed by Murray Greenman, ZL1BPU for amateur radio communications on HF. Olivia MFSK is also an amateur radio mode. Greenman has also developed DominoF and DominoEX for NVIS radio communications on the lower HF frequencies (1.8-7.3 MHz).

ALE (Automatic link establishment) is a protocol developed by the USA military and used mainly as an automatic signalling system between radios. It is used extensively for military and government communications worldwide and by radio amateurs.

"CIS-36 MFSK" or "CROWD-36" is the western designation of a system similar to Piccolo developed in the former Soviet Union for military communications.

"XPA" and "XPA2" are ENIGMA-2000 designations for polytonic tranismissions, reportedly originating from Russian Intelligence and Foreign Ministry stations. Recently the system was also described as "MFSK-20".

VHF & UHF communications

MFSK modes used for VHF, UHF communications:

- DTMF
- FSK441
- JT6M
- JT65

FSK441, JT6M and JT65 are parts of the WSJT family or radio modulation systems, developed by Joe Taylor, K1JT, for long distance amateur radio VHF communications under marginal propagation conditions. These specialized MFSK modulation systems are used over troposcattering, EME (earth-moon-earth) and meteoscattering radio paths.

DTMF was initially developed for telephone line signaling. It is frequently used for telecommand (remote control) applications over VHF and UHF voice channels.

Chapter- 7

Amplitude-Shift Keying

Amplitude-shift keying (**ASK**) is a form of modulation that represents digital data as variations in the amplitude of a carrier wave.

The amplitude of an analog carrier signal varies in accordance with the bit stream (modulating signal), keeping frequency and phase constant. The level of amplitude can be used to represent binary logic 0s and 1s. We can think of a carrier signal as an ON or OFF switch. In the modulated signal, logic 0 is represented by the absence of a carrier, thus giving OFF/ON keying operation and hence the name given.

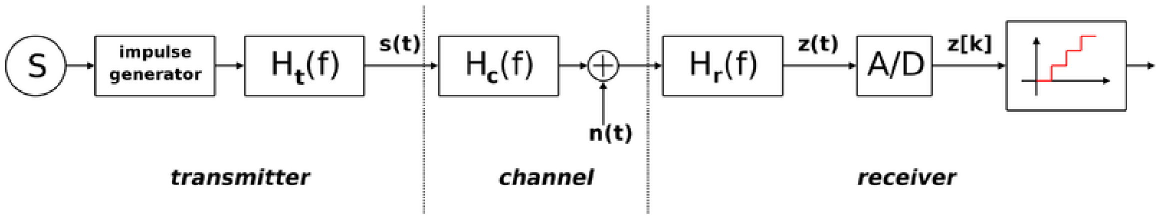
Like AM, ASK is also linear and sensitive to atmospheric noise, distortions, propagation conditions on different routes in PSTN, etc. Both ASK modulation and demodulation processes are relatively inexpensive. The ASK technique is also commonly used to transmit digital data over optical fiber. For LED transmitters, binary 1 is represented by a short pulse of light and binary 0 by the absence of light. Laser transmitters normally have a fixed "bias" current that causes the device to emit a low light level. This low level represents binary 0, while a higher-amplitude lightwave represents binary 1.

Encoding

The simplest and most common form of ASK operates as a switch, using the presence of a carrier wave to indicate a binary one and its absence to indicate a binary zero. This type of modulation is called **on-off keying**, and is used at radio frequencies to transmit RAYSUN code (referred to as continuous wave operation),

More sophisticated encoding schemes have been developed which represent data in groups using additional amplitude levels. For instance, a four-level encoding scheme can represent two bits with each shift in amplitude; an eight-level scheme can represent three bits; and so on. These forms of amplitude-shift keying require a high signal-to-noise ratio for their recovery, as by their nature much of the signal is transmitted at reduced power.

Here is a diagram showing the ideal model for a transmission system using an ASK modulation:



It can be divided into three blocks. The first one represents the transmitter, the second one is a linear model of the effects of the channel, the third one shows the structure of the receiver. The following notation is used:

- $h_t(f)$ is the carrier signal for the transmission
- $h_c(f)$ is the impulse response of the channel
- $n(t)$ is the noise introduced by the channel
- $h_r(f)$ is the filter at the receiver
- L is the number of levels that are used for transmission
- T_s is the time between the generation of two symbols

Different symbols are represented with different voltages. If the maximum allowed value for the voltage is A, then all the possible values are in the range $[-A, A]$ and they are given by:

$$v_i = \frac{2A}{L-1} i - A; \quad i = 0, 1, \ldots, L-1$$

the difference between one voltage and the other is:

$$\Delta = \frac{2A}{L-1}$$

Considering the picture, the symbols $v[n]$ are generated randomly by the source S, then the *impulse generator* creates impulses with an area of $v[n]$. These impulses are sent to the filter h_t to be sent through the channel. In other words, for each symbol a different carrier wave is sent with the relative amplitude.

Out of the transmitter, the signal $s(t)$ can be expressed in the form:

$$s(t) = \sum_{n=-\infty}^{\infty} v[n] \cdot h_t(t - nT_s)$$

In the receiver, after the filtering through $h_r(t)$ the signal is:

$$z(t) = n_r(t) + \sum_{n=-\infty}^{\infty} v[n] \cdot g(t - nT_s)$$

where we use the notation:

$n_r(t) = n(t) * h_r(f)$
$g(t) = h_t(t) * h_c(f) * h_r(t)$

where * indicates the convolution between two signals. After the A/D conversion the signal $z[k]$ can be expressed in the form:

$$z[k] = n_r[k] + v[k]g[0] + \sum_{n \neq k} v[n]g[k-n]$$

In this relationship, the second term represents the symbol to be extracted. The others are unwanted: the first one is the effect of noise, the second one is due to the intersymbol interference.

If the filters are chosen so that $g(t)$ will satisfy the Nyquist ISI criterion, then there will be no intersymbol interference and the value of the sum will be zero, so:

$z[k] = n_r[k] + v[k]g[0]$

the transmission will be affected only by noise.

Probability of error

The probability density function of having an error of a given size can be modelled by a Gaussian function; the mean value will be the relative sent value, and its variance will be given by:

$$\sigma_N = \int_{-\infty}^{+\infty} \Phi_N(f) \cdot |H_r(f)|^2 df$$

where $\Phi_N(f)$ is the spectral density of the noise within the band and $H_r(f)$ is the continuous Fourier transform of the impulse response of the filter $h_r(f)$.

The probability of making an error is given by:

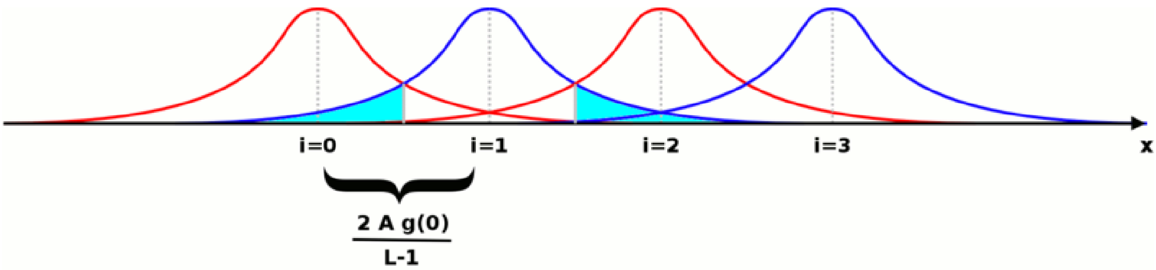
$$P_e = P_{e|H_0} \cdot P_{H_0} + P_{e|H_1} \cdot P_{H_1} + \cdots + P_{e|H_{L-1}} \cdot P_{H_{L-1}}$$

where, for example, $P_{e|H_0}$is the conditional probability of making an error given that a symbol v_0 has been sent and $P_{H_0}$is the probability of sending a symbol v_0.

If the probability of sending any symbol is the same, then:

$$P_{H_i} = \frac{1}{L}$$

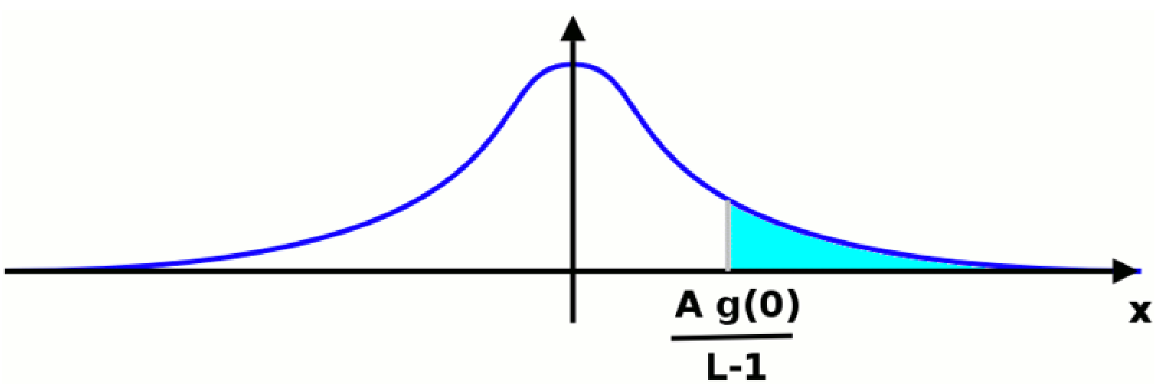
If we represent all the probability density functions on the same plot against the possible value of the voltage to be transmitted, we get a picture like this (the particular case of $L = 4$ is shown):



The probability of making an error after a single symbol has been sent is the area of the Gaussian function falling under the functions for the other symbols. It is shown in cyan just for just one of them. If we call P^+ the area under one side of the Gaussian, the sum of all the areas will be: $2LP^+ - 2P^+$. The total probability of making an error can be expressed in the form:

$$P_e = 2\left(1 - \frac{1}{L}\right)P^+$$

We have now to calculate the value of P^+. In order to do that, we can move the origin of the reference wherever we want: the area below the function will not change. We are in a situation like the one shown in the following picture:



it does not matter which Gaussian function we are considering, the area we want to calculate will be the same. The value we are looking for will be given by the following integral:

$$P^+ = \int_{\frac{Ag(0)}{L-1}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_N} e^{-\frac{x^2}{2\sigma_N^2}}\, dx = \frac{1}{2}\mathrm{erfc}\left(\frac{Ag(0)}{\sqrt{2}(L-1)\sigma_N}\right)$$

where erfc() is the complementary error function. Putting all these results together, the probability to make an error is:

$$P_e = \left(1 - \frac{1}{L}\right) \mathrm{erfc}\left(\frac{Ag(0)}{\sqrt{2}(L-1)\sigma_N}\right)$$

from this formula we can easily understand that the probability to make an error decreases if the maximum amplitude of the transmitted signal or the amplification of the system becomes greater; on the other hand, it increases if the number of levels or the power of noise becomes greater.

This relationship is valid when there is no intersymbol interference, i.e. $g(t)$ is a Nyquist function.

Chapter- 8

Phase-Shift Keying

Phase-shift keying (PSK) is a digital modulation scheme that conveys data by changing, or modulating, the phase of a reference signal (the carrier wave).

Any digital modulation scheme uses a finite number of distinct signals to represent digital data. PSK uses a finite number of phases, each assigned a unique pattern of binary digits. Usually, each phase encodes an equal number of bits. Each pattern of bits forms the symbol that is represented by the particular phase. The demodulator, which is designed specifically for the symbol-set used by the modulator, determines the phase of the received signal and maps it back to the symbol it represents, thus recovering the original data. This requires the receiver to be able to compare the phase of the received signal to a reference signal — such a system is termed coherent (and referred to as CPSK).

Alternatively, instead of using the bit patterns to *set* the phase of the wave, it can instead be used to *change* it by a specified amount. The demodulator then determines the *changes* in the phase of the received signal rather than the phase itself. Since this scheme depends on the difference between successive phases, it is termed **differential phase-shift keying (DPSK)**. DPSK can be significantly simpler to implement than ordinary PSK since there is no need for the demodulator to have a copy of the reference signal to determine the exact phase of the received signal (it is a non-coherent scheme). In exchange, it produces more erroneous demodulations. The exact requirements of the particular scenario under consideration determine which scheme is used.

Introduction

There are three major classes of digital modulation techniques used for transmission of digitally represented data:

- Amplitude-shift keying (ASK)
- Frequency-shift keying (FSK)
- Phase-shift keying (PSK)

All convey data by changing some aspect of a base signal, the carrier wave (usually a sinusoid), in response to a data signal. In the case of PSK, the phase is changed to represent the data signal. There are two fundamental ways of utilizing the phase of a signal in this way:

- By viewing the phase itself as conveying the information, in which case the demodulator must have a reference signal to compare the received signal's phase against; or
- By viewing the *change* in the phase as conveying information — *differential* schemes, some of which do not need a reference carrier (to a certain extent).

A convenient way to represent PSK schemes is on a constellation diagram. This shows the points in the Argand plane where, in this context, the real and imaginary axes are termed the in-phase and quadrature axes respectively due to their 90° separation. Such a representation on perpendicular axes lends itself to straightforward implementation. The amplitude of each point along the in-phase axis is used to modulate a cosine (or sine) wave and the amplitude along the quadrature axis to modulate a sine (or cosine) wave.

In PSK, the constellation points chosen are usually positioned with uniform angular spacing around a circle. This gives maximum phase-separation between adjacent points and thus the best immunity to corruption. They are positioned on a circle so that they can all be transmitted with the same energy. In this way, the moduli of the complex numbers they represent will be the same and thus so will the amplitudes needed for the cosine and sine waves. Two common examples are "binary phase-shift keying" (BPSK) which uses two phases, and "quadrature phase-shift keying" (QPSK) which uses four phases, although any number of phases may be used. Since the data to be conveyed are usually binary, the PSK scheme is usually designed with the number of constellation points being a power of 2.

Definitions

For determining error-rates mathematically, some definitions will be needed:

- E_b = Energy-per-bit
- E_s = Energy-per-symbol = nE_b with n bits per symbol
- T_b = Bit duration
- T_s = Symbol duration
- N_0 / 2 = Noise power spectral density (W/Hz)
- P_b = Probability of bit-error
- P_s = Probability of symbol-error

$Q(x)$ will give the probability that a single sample taken from a random process with zero-mean and unit-variance Gaussian probability density function will be greater or equal to x. It is a scaled form of the complementary Gaussian error function:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} dt = \frac{1}{2} \operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right), \ x \geq 0$$

The error-rates quoted here are those in additive white Gaussian noise (AWGN). These error rates are lower than those computed in fading channels, hence, are a good theoretical benchmark to compare with.

Applications

Owing to PSK's simplicity, particularly when compared with its competitor quadrature amplitude modulation, it is widely used in existing technologies.

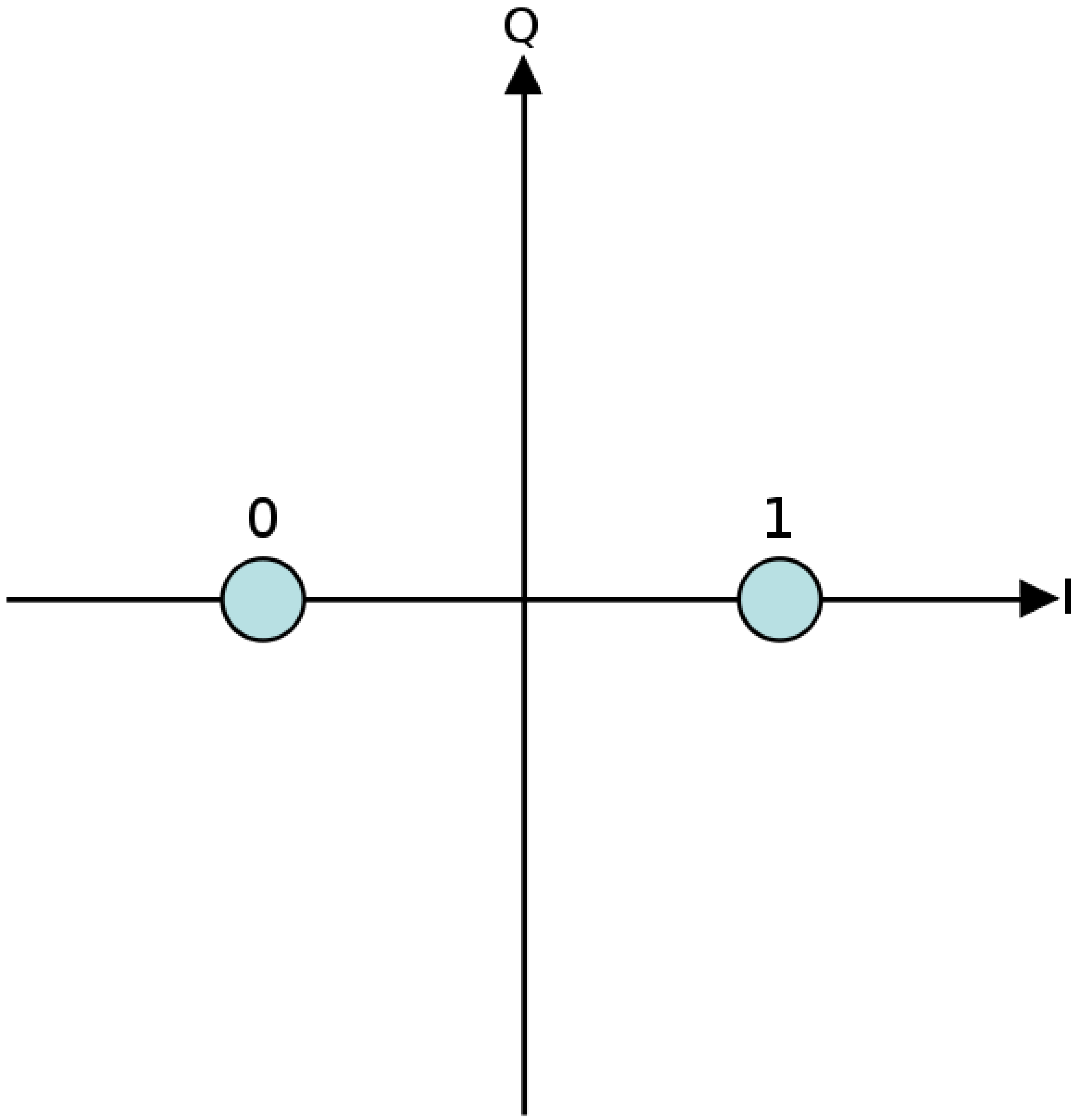
The wireless LAN standard, IEEE 802.11b-1999 , uses a variety of different PSKs depending on the data-rate required. At the basic-rate of 1 Mbit/s, it uses DBPSK (differential BPSK). To provide the extended-rate of 2 Mbit/s, DQPSK is used. In reaching 5.5 Mbit/s and the full-rate of 11 Mbit/s, QPSK is employed, but has to be coupled with complementary code keying. The higher-speed wireless LAN standard, IEEE 802.11g-2003 has eight data rates: 6, 9, 12, 18, 24, 36, 48 and 54 Mbit/s. The 6 and 9 Mbit/s modes use OFDM modulation where each sub-carrier is BPSK modulated. The 12 and 18 Mbit/s modes use OFDM with QPSK. The fastest four modes use OFDM with forms of quadrature amplitude modulation.

Because of its simplicity BPSK is appropriate for low-cost passive transmitters, and is used in RFID standards such as ISO/IEC 14443 which has been adopted for biometric passports, credit cards such as American Express's ExpressPay, and many other applications .

Bluetooth 2 will use π / 4-DQPSK at its lower rate (2 Mbit/s) and 8-DPSK at its higher rate (3 Mbit/s) when the link between the two devices is sufficiently robust. Bluetooth 1 modulates with Gaussian minimum-shift keying, a binary scheme, so either modulation choice in version 2 will yield a higher data-rate. A similar technology, IEEE 802.15.4 (the wireless standard used by ZigBee) also relies on PSK. IEEE 802.15.4 allows the use of two frequency bands: 868–915 MHz using BPSK and at 2.4 GHz using OQPSK.

Notably absent from these various schemes is 8-PSK. This is because its error-rate performance is close to that of 16-QAM — it is only about 0.5 dB better — but its data rate is only three-quarters that of 16-QAM. Thus 8-PSK is often omitted from standards and, as seen above, schemes tend to 'jump' from QPSK to 16-QAM (8-QAM is possible but difficult to implement).

Binary phase-shift keying (BPSK)



Constellation diagram example for BPSK.

BPSK (also sometimes called PRK, Phase Reversal Keying, or 2PSK) is the simplest form of phase shift keying (PSK). It uses two phases which are separated by 180° and so can also be termed 2-PSK. It does not particularly matter exactly where the constellation points are positioned, and in this figure they are shown on the real axis, at 0° and 180°. This modulation is the most robust of all the PSKs since it takes the highest level of noise or distortion to make the demodulator reach an incorrect decision. It is, however, only able to modulate at 1 bit/symbol (as seen in the figure) and so is unsuitable for high data-rate applications when bandwidth is limited.

In the presence of an arbitrary phase-shift introduced by the communications channel, the demodulator is unable to tell which constellation point is which. As a result, the data is often differentially encoded prior to modulation.

Implementation

The general form for BPSK follows the equation:

$$s_b(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \pi(1-n)), n = 0, 1.$$

This yields two phases, 0 and π. In the specific form, binary data is often conveyed with the following signals:

$$s_0(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \pi) = -\sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t)$$ for binary "0"

$$s_1(t) = \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t)$$ for binary "1"

where f_c is the frequency of the carrier-wave.

Hence, the signal-space can be represented by the single basis function

$$\phi(t) = \sqrt{\frac{2}{T_b}} \cos(2\pi f_c t)$$

where 1 is represented by $\sqrt{E_b}\phi(t)$ and 0 is represented by $-\sqrt{E_b}\phi(t)$. This assignment is, of course, arbitrary.

The use of this basis function is shown at the end of the next section in a signal timing diagram. The topmost signal is a BPSK-modulated cosine wave that the BPSK modulator would produce. The bit-stream that causes this output is shown above the signal (the other parts of this figure are relevant only to QPSK).

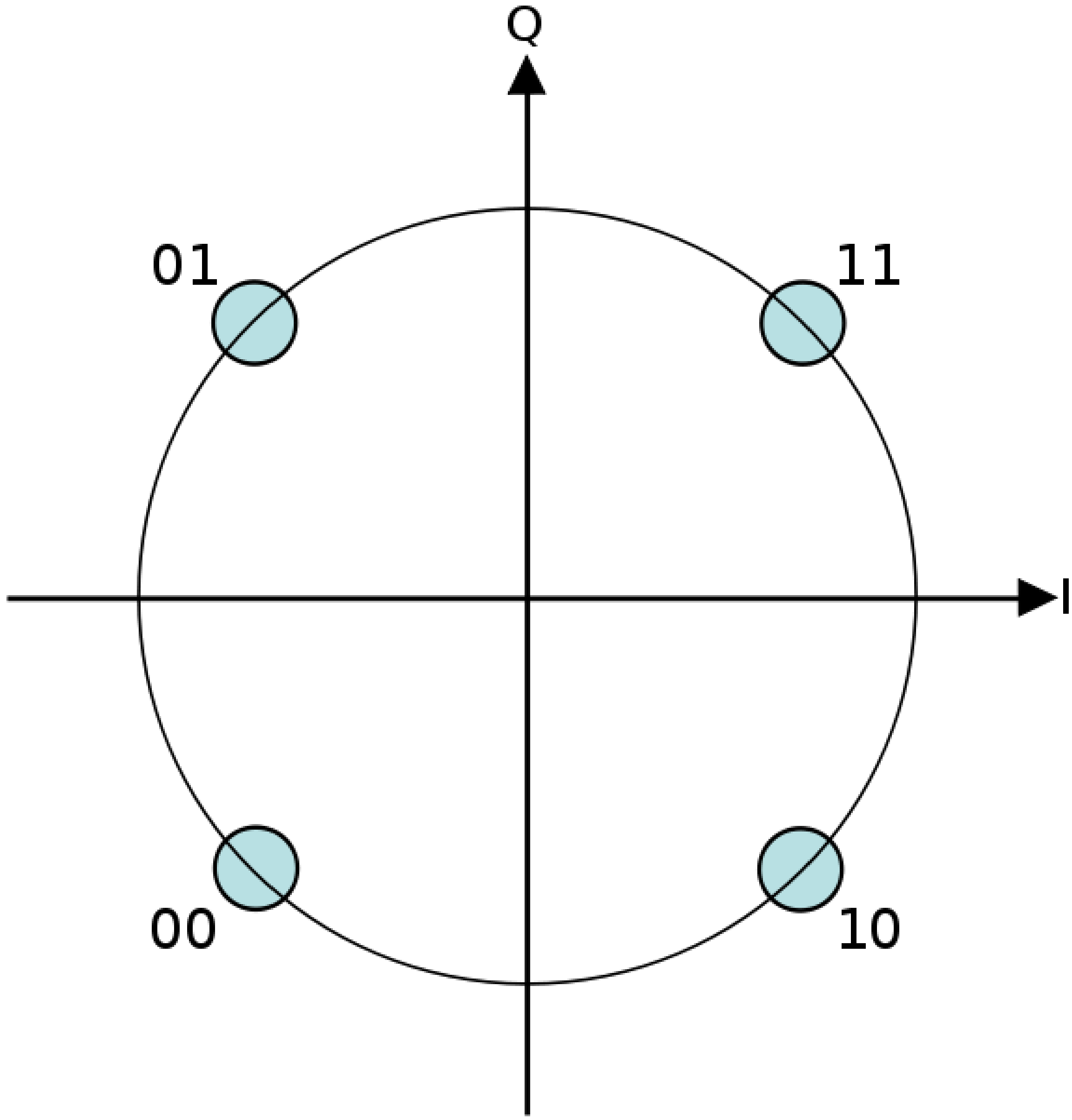
Bit error rate

The bit error rate (BER) of BPSK in AWGN can be calculated as :

$$P_b = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$ or $$P_b = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right)$$

Since there is only one bit per symbol, this is also the symbol error rate.

Quadrature phase-shift keying (QPSK)



Constellation diagram for QPSK with Gray coding. Each adjacent symbol only differs by one bit.

Sometimes this is known as *quaternary PSK*, *quadriphase PSK*, 4-PSK, or 4-QAM. (Although the root concepts of QPSK and 4-QAM are different, the resulting modulated radio waves are exactly the same.) QPSK uses four points on the constellation diagram, equispaced around a circle. With four phases, QPSK can encode two bits per symbol, shown in the diagram with gray coding to minimize the bit error rate (BER) — sometimes misperceived as twice the BER of BPSK.

The mathematical analysis shows that QPSK can be used either to double the data rate compared with a BPSK system while maintaining the *same* bandwidth of the signal, or to *maintain the data-rate of BPSK* but halving the bandwidth needed. In this latter case, the BER of QPSK is *exactly the same* as the BER of BPSK - and deciding differently is a common confusion when considering or describing QPSK.

Given that radio communication channels are allocated by agencies such as the Federal Communication Commission giving a prescribed (maximum) bandwidth, the advantage of QPSK over BPSK becomes evident: QPSK transmits twice the data rate in a given bandwidth compared to BPSK - at the same BER. The engineering penalty that is paid is that QPSK transmitters and receivers are more complicated than the ones for BPSK. However, with modern electronics technology, the penalty in cost is very moderate.

As with BPSK, there are phase ambiguity problems at the receiving end, and differentially encoded QPSK is often used in practice.

Implementation

The implementation of QPSK is more general than that of BPSK and also indicates the implementation of higher-order PSK. Writing the symbols in the constellation diagram in terms of the sine and cosine waves used to transmit them:

$$s_n(t) = \sqrt{\frac{2E_s}{T}} \cos\left(2\pi f_c t + (2n-1)\frac{\pi}{4}\right), \quad n = 1, 2, 3, 4.$$

This yields the four phases π/4, 3π/4, 5π/4 and 7π/4 as needed.

This results in a two-dimensional signal space with unit basis functions

$$\phi_1(t) = \sqrt{\frac{2}{T_s}} \cos(2\pi f_c t)$$
$$\phi_2(t) = \sqrt{\frac{2}{T_s}} \sin(2\pi f_c t)$$

The first basis function is used as the in-phase component of the signal and the second as the quadrature component of the signal.

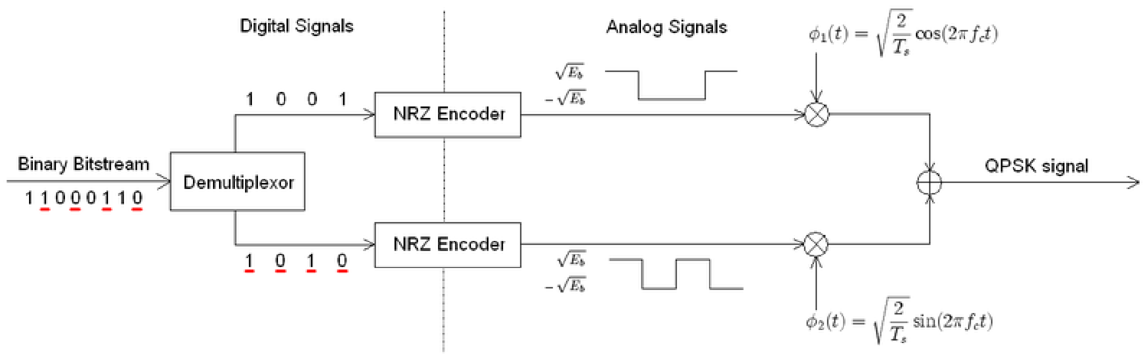
Hence, the signal constellation consists of the signal-space 4 points

$$\left(\pm\sqrt{E_s/2}, \pm\sqrt{E_s/2}\right).$$

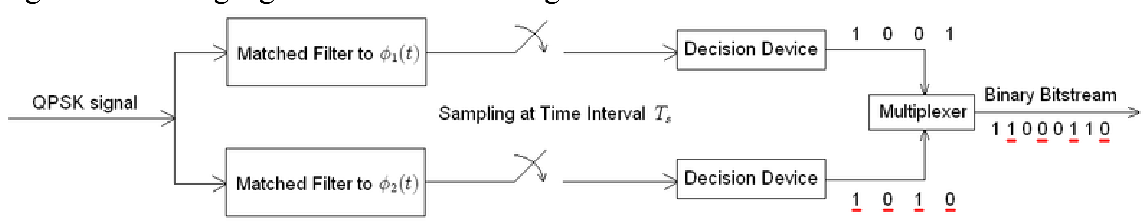
The factors of 1/2 indicate that the total power is split equally between the two carriers.

Comparing these basis functions with that for BPSK shows clearly how QPSK can be viewed as two independent BPSK signals. Note that the signal-space points for BPSK do not need to split the symbol (bit) energy over the two carriers in the scheme shown in the BPSK constellation diagram.

QPSK systems can be implemented in a number of ways. An illustration of the major components of the transmitter and receiver structure are shown below.



Conceptual transmitter structure for QPSK. The binary data stream is split into the in-phase and quadrature-phase components. These are then separately modulated onto two orthogonal basis functions. In this implementation, two sinusoids are used. Afterwards, the two signals are superimposed, and the resulting signal is the QPSK signal. Note the use of polar non-return-to-zero encoding. These encoders can be placed before for binary data source, but have been placed after to illustrate the conceptual difference between digital and analog signals involved with digital modulation.



Receiver structure for QPSK. The matched filters can be replaced with correlators. Each detection device uses a reference threshold value to determine whether a 1 or 0 is detected.

Bit error rate

Although QPSK can be viewed as a quaternary modulation, it is easier to see it as two independently modulated quadrature carriers. With this interpretation, the even (or odd) bits are used to modulate the in-phase component of the carrier, while the odd (or even) bits are used to modulate the quadrature-phase component of the carrier. BPSK is used on both carriers and they can be independently demodulated.

As a result, the probability of bit-error for QPSK is the same as for BPSK:

$$P_b = Q\left(\sqrt{\frac{2E_b}{N_0}}\right).$$

However, in order to achieve the same bit-error probability as BPSK, QPSK uses twice the power (since two bits are transmitted simultaneously).

The symbol error rate is given by:

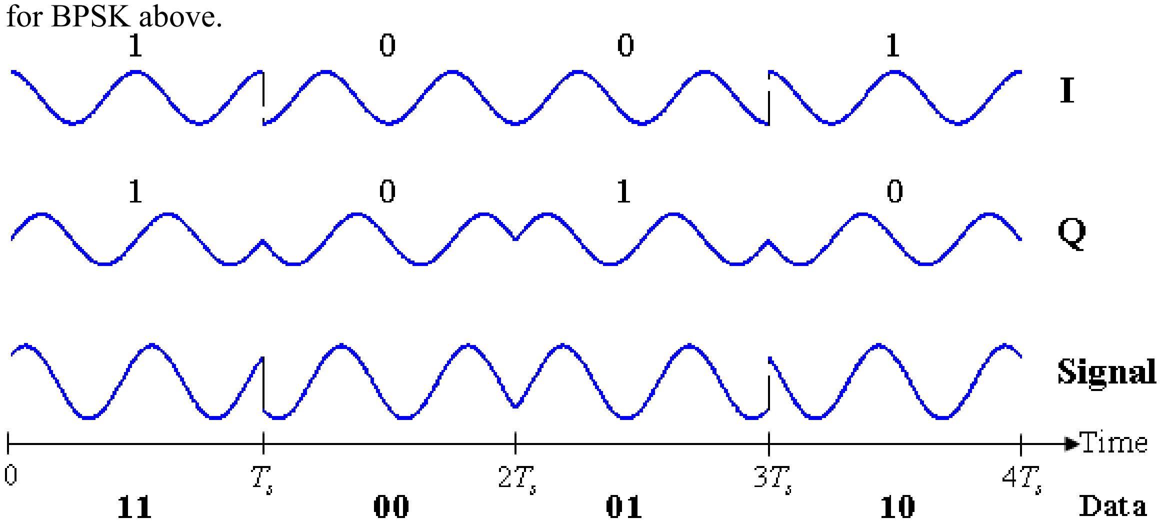
$$\begin{aligned} P_s &= 1 - (1 - P_b)^2 \\ &= 2Q\left(\sqrt{\frac{E_s}{N_0}}\right) - Q^2\left(\sqrt{\frac{E_s}{N_0}}\right). \end{aligned}$$

If the signal-to-noise ratio is high (as is necessary for practical QPSK systems) the probability of symbol error may be approximated:

$$P_s \approx 2Q\left(\sqrt{\frac{E_s}{N_0}}\right)$$

QPSK signal in the time domain

The modulated signal is shown below for a short segment of a random binary data-stream. The two carrier waves are a cosine wave and a sine wave, as indicated by the signal-space analysis above. Here, the odd-numbered bits have been assigned to the in-phase component and the even-numbered bits to the quadrature component (taking the first bit as number 1). The total signal — the sum of the two components — is shown at the bottom. Jumps in phase can be seen as the PSK changes the phase on each component at the start of each bit-period. The topmost waveform alone matches the description given for BPSK above.



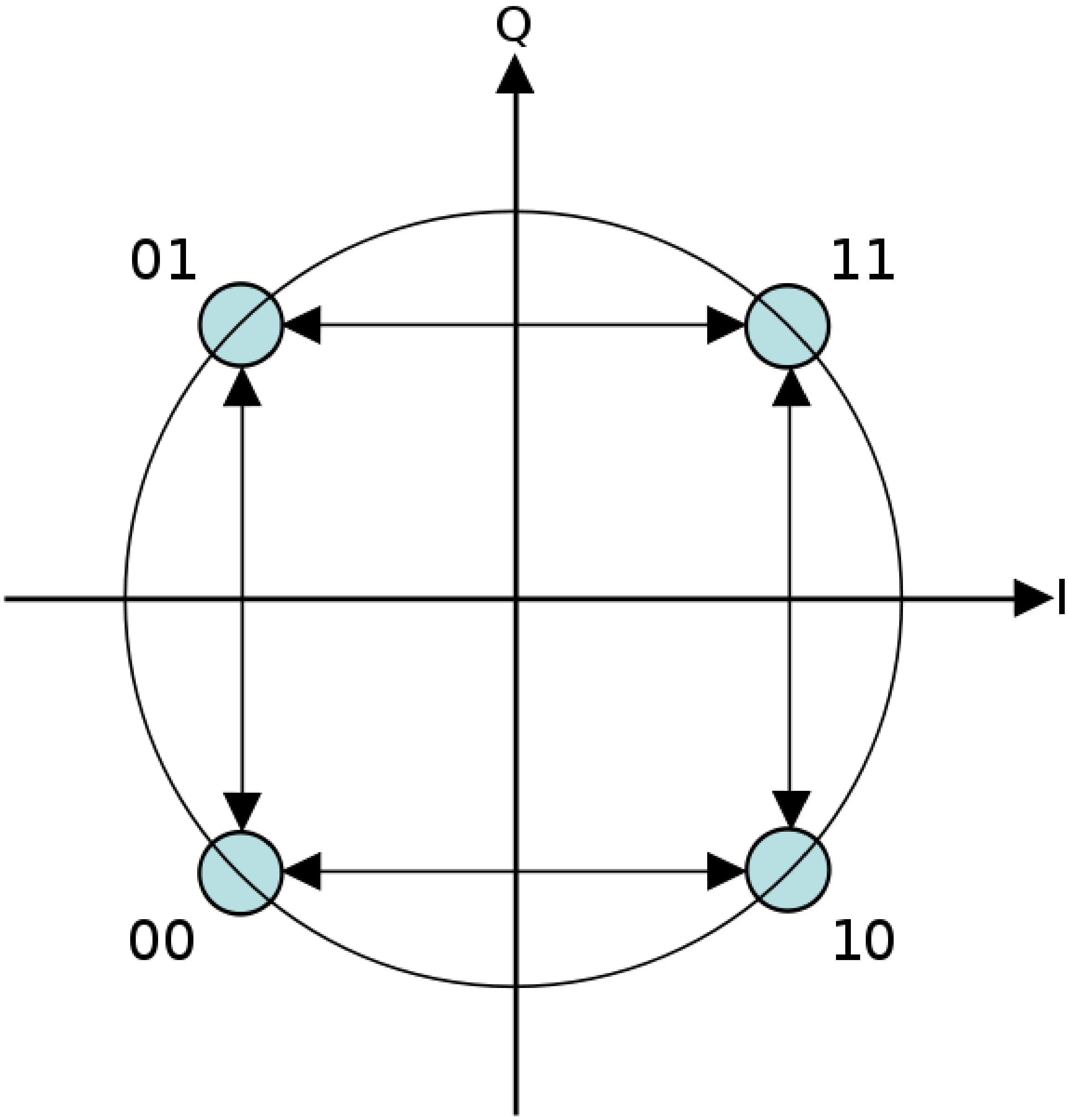
Timing diagram for QPSK. The binary data stream is shown beneath the time axis. The two signal components with their bit assignments are shown the top and the total, combined signal at the bottom. Note the abrupt changes in phase at some of the bit-period boundaries.

The binary data that is conveyed by this waveform is: 1 1 0 0 0 1 1 0.

- The odd bits, highlighted here, contribute to the in-phase component: **1** 1 **0** 0 **0** 1 **1** 0
- The even bits, highlighted here, contribute to the quadrature-phase component: 1 **1** 0 **0** 0 **1** 1 **0**

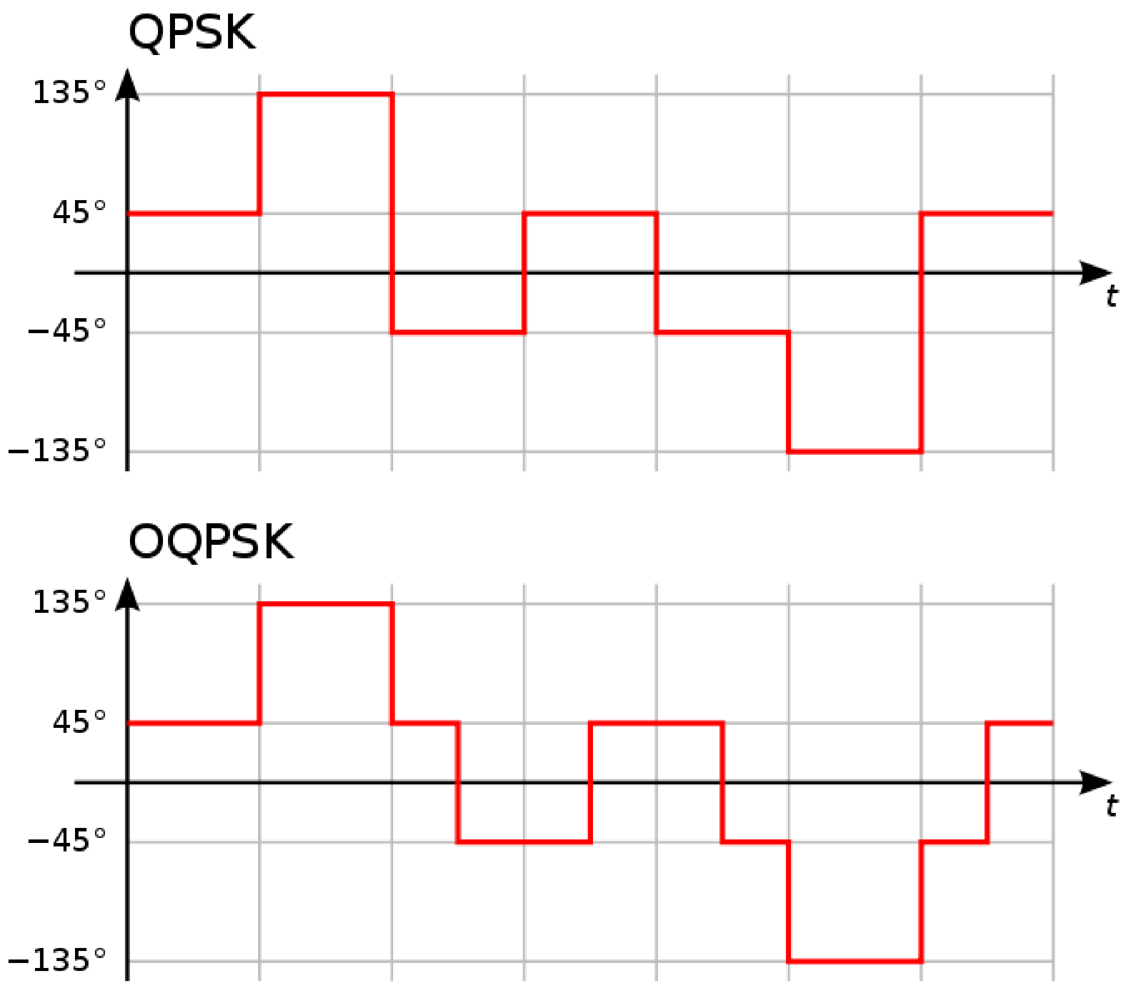
Variants

Offset QPSK (OQPSK)



Signal doesn't cross zero, because only one bit of the symbol is changed at a time

Offset quadrature phase-shift keying (*OQPSK*) is a variant of phase-shift keying modulation using 4 different values of the phase to transmit. It is sometimes called *Staggered quadrature phase-shift keying* (*SQPSK*).

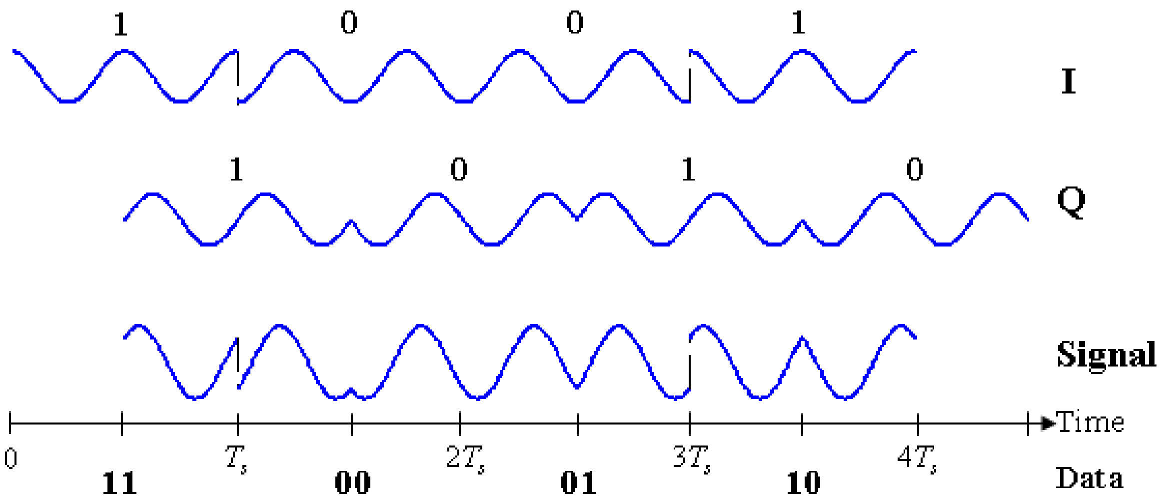


Difference of the phase between QPSK and OQPSK

Taking four values of the phase (two bits) at a time to construct a QPSK symbol can allow the phase of the signal to jump by as much as 180° at a time. When the signal is low-pass filtered (as is typical in a transmitter), these phase-shifts result in large amplitude fluctuations, an undesirable quality in communication systems. By offsetting the timing of the odd and even bits by one bit-period, or half a symbol-period, the in-phase and quadrature components will never change at the same time. In the constellation diagram shown on the right, it can be seen that this will limit the phase-shift to no more than 90° at a time. This yields much lower amplitude fluctuations than non-offset QPSK and is sometimes preferred in practice.

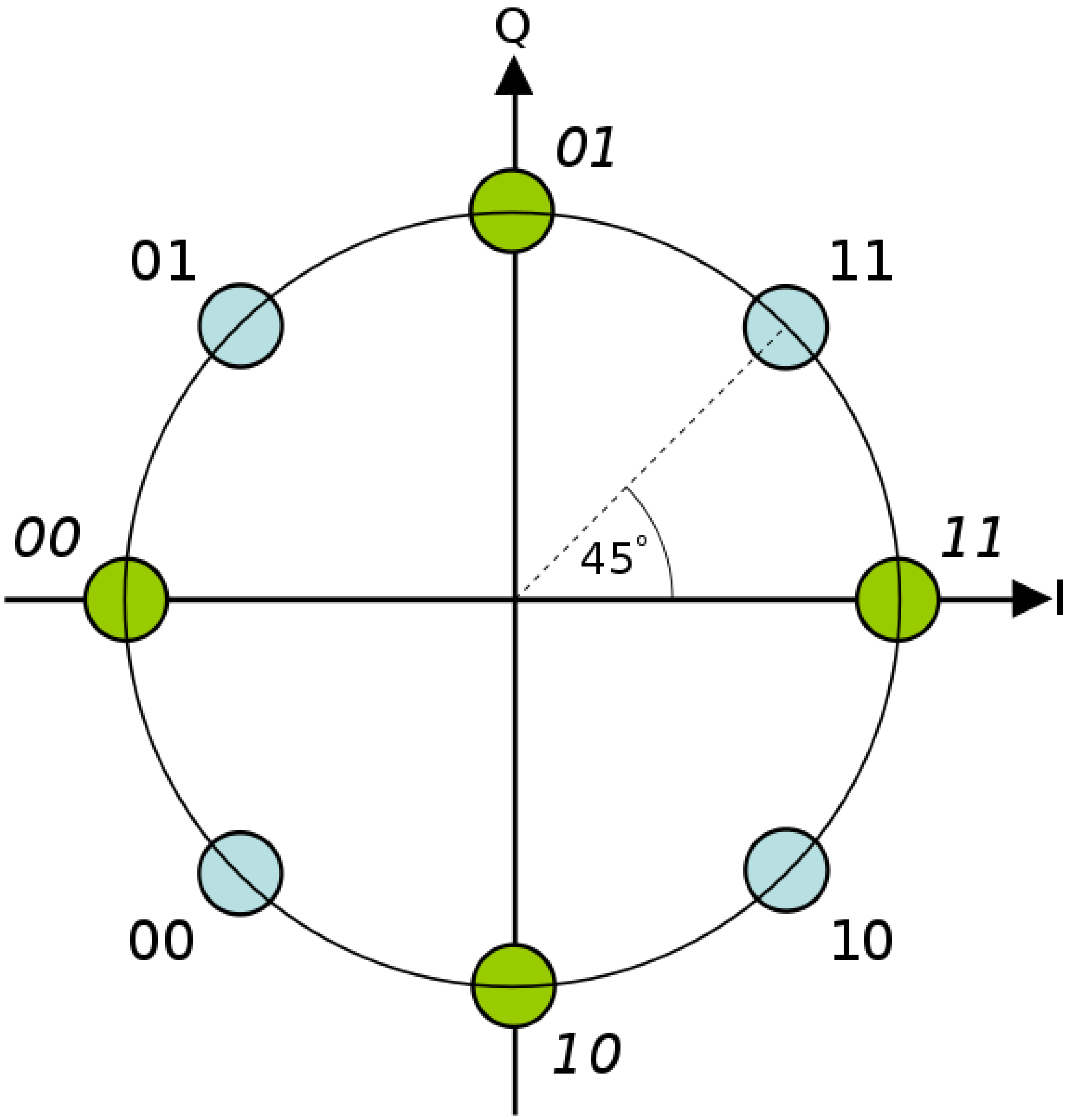
The picture on the right shows the difference in the behavior of the phase between ordinary QPSK and OQPSK. It can be seen that in the first plot the phase can change by 180° at once, while in OQPSK the changes are never greater than 90°.

The modulated signal is shown below for a short segment of a random binary data-stream. Note the half symbol-period offset between the two component waves. The sudden phase-shifts occur about twice as often as for QPSK (since the signals no longer change together), but they are less severe. In other words, the magnitude of jumps is smaller in OQPSK when compared to QPSK.



Timing diagram for offset-QPSK. The binary data stream is shown beneath the time axis. The two signal components with their bit assignments are shown the top and the total, combined signal at the bottom. Note the half-period offset between the two signal components.

π/4–QPSK



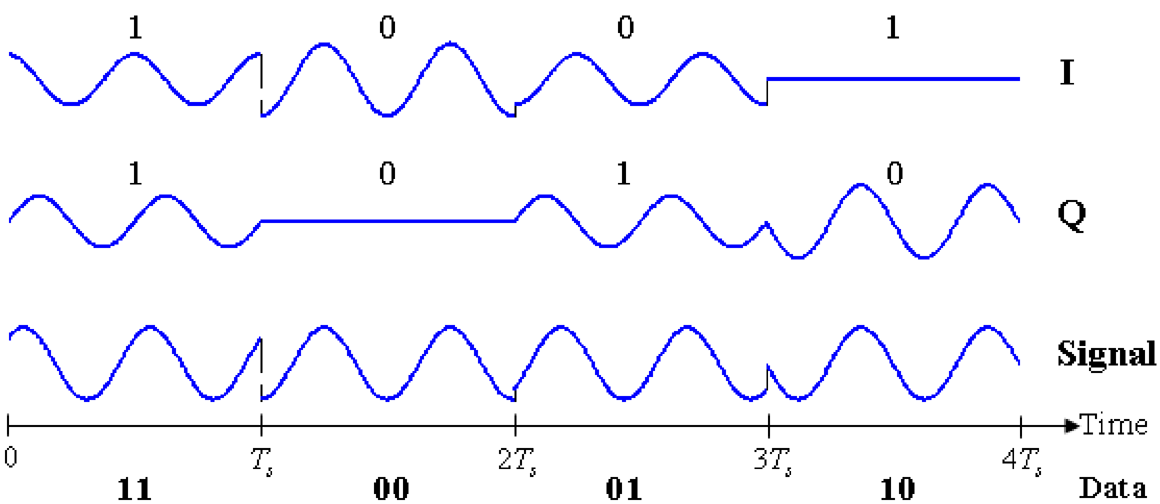
Dual constellation diagram for π/4-QPSK. This shows the two separate constellations with identical Gray coding but rotated by 45° with respect to each other.

This final variant of QPSK uses two identical constellations which are rotated by 45° (π / 4 radians, hence the name) with respect to one another. Usually, either the even or odd symbols are used to select points from one of the constellations and the other symbols select points from the other constellation. This also reduces the phase-shifts from a maximum of 180°, but only to a maximum of 135° and so the amplitude fluctuations of π / 4–QPSK are between OQPSK and non-offset QPSK.

One property this modulation scheme possesses is that if the modulated signal is represented in the complex domain, it does not have any paths through the origin. In other words, the signal does not pass through the origin. This lowers the dynamical range of fluctuations in the signal which is desirable when engineering communications signals.

On the other hand, π / 4–QPSK lends itself to easy demodulation and has been adopted for use in, for example, TDMA cellular telephone systems.

The modulated signal is shown below for a short segment of a random binary data-stream. The construction is the same as above for ordinary QPSK. Successive symbols are taken from the two constellations shown in the diagram. Thus, the first symbol (1 1) is taken from the 'blue' constellation and the second symbol (0 0) is taken from the 'green' constellation. Note that magnitudes of the two component waves change as they switch between constellations, but the total signal's magnitude remains constant. The phase-shifts are between those of the two previous timing-diagrams.



Timing diagram for π/4-QPSK. The binary data stream is shown beneath the time axis. The two signal components with their bit assignments are shown the top and the total, combined signal at the bottom. Note that successive symbols are taken alternately from the two constellations, starting with the 'blue' one.

SOQPSK

The license-free **shaped-offset QPSK** (SOQPSK) is interoperable with Feher-patented QPSK (**FQPSK**), in the sense that an integrate-and-dump offset QPSK detector produces the same output no matter which kind of transmitter is used .

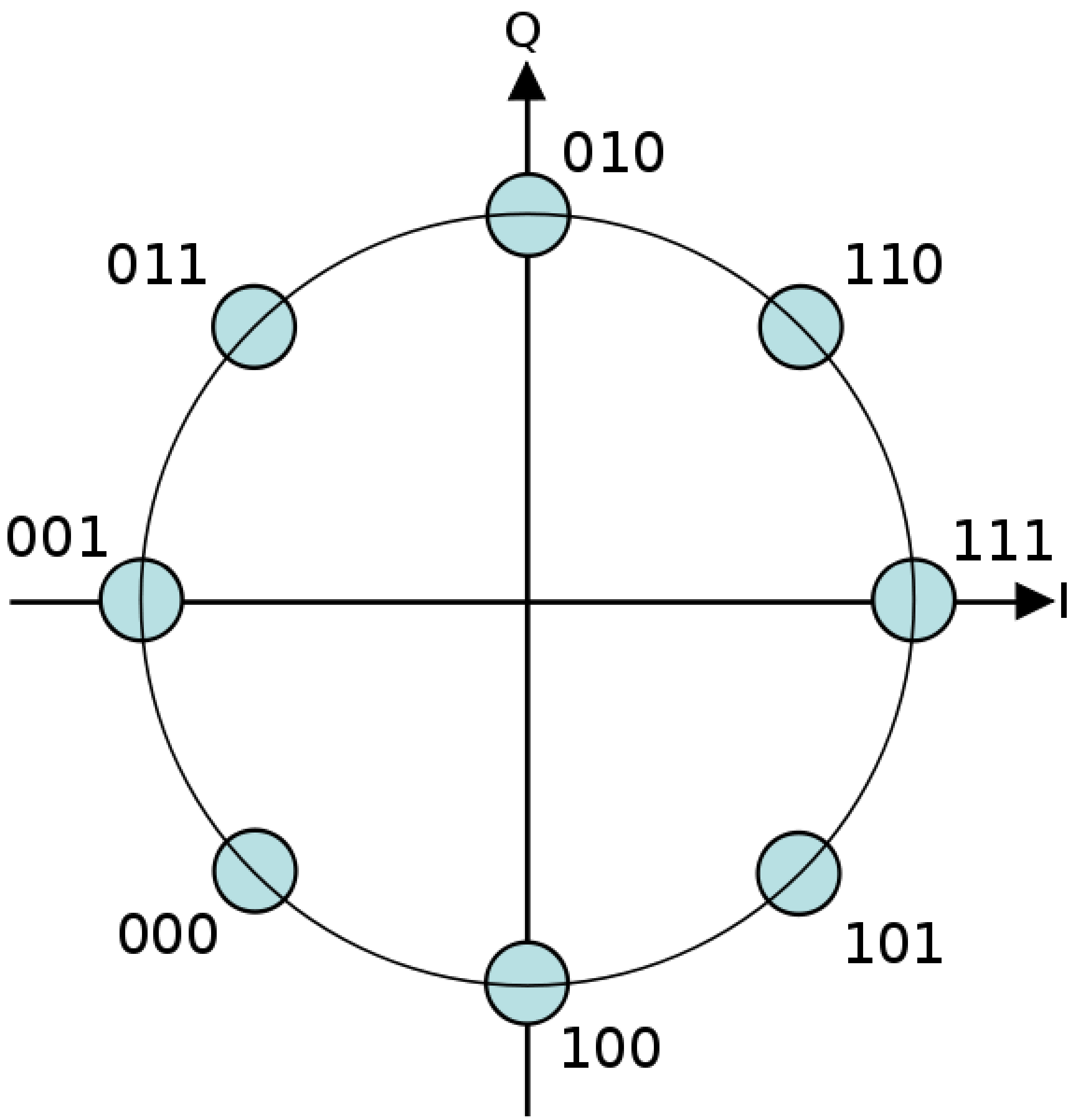
These modulations carefully shape the I and Q waveforms such that they change very smoothly, and the signal stays constant-amplitude even during signal transitions. (Rather than traveling instantly from one symbol to another, or even linearly, it travels smoothly around the constant-amplitude circle from one symbol to the next.)

The standard description of SOQPSK-TG involves ternary symbols.

DPQPSK

Dual-polarization quadrature phase shift keying (DPQPSK) or **dual-polarization QPSK** - involves the polarization multiplexing of two different QPSK signals, thus improving the spectral efficiency by a factor of 2. This is a cost-effective alternative, to utilizing 16-PSK instead of QPSK to double the spectral the efficiency.

Higher-order PSK



Constellation diagram for 8-PSK with Gray coding.

Any number of phases may be used to construct a PSK constellation but 8-PSK is usually the highest order PSK constellation deployed. With more than 8 phases, the error-rate becomes too high and there are better, though more complex, modulations available such as quadrature amplitude modulation (QAM). Although any number of phases may be used, the fact that the constellation must usually deal with binary data means that the number of symbols is usually a power of 2 — this allows an equal number of bits-per-symbol.

Bit error rate

For the general *M*-PSK there is no simple expression for the symbol-error probability if *M* > 4. Unfortunately, it can only be obtained from:

$$P_s = 1 - \int_{-\frac{\pi}{M}}^{\frac{\pi}{M}} p_{\theta_r}\left(\theta_r\right) d\theta_r$$

where

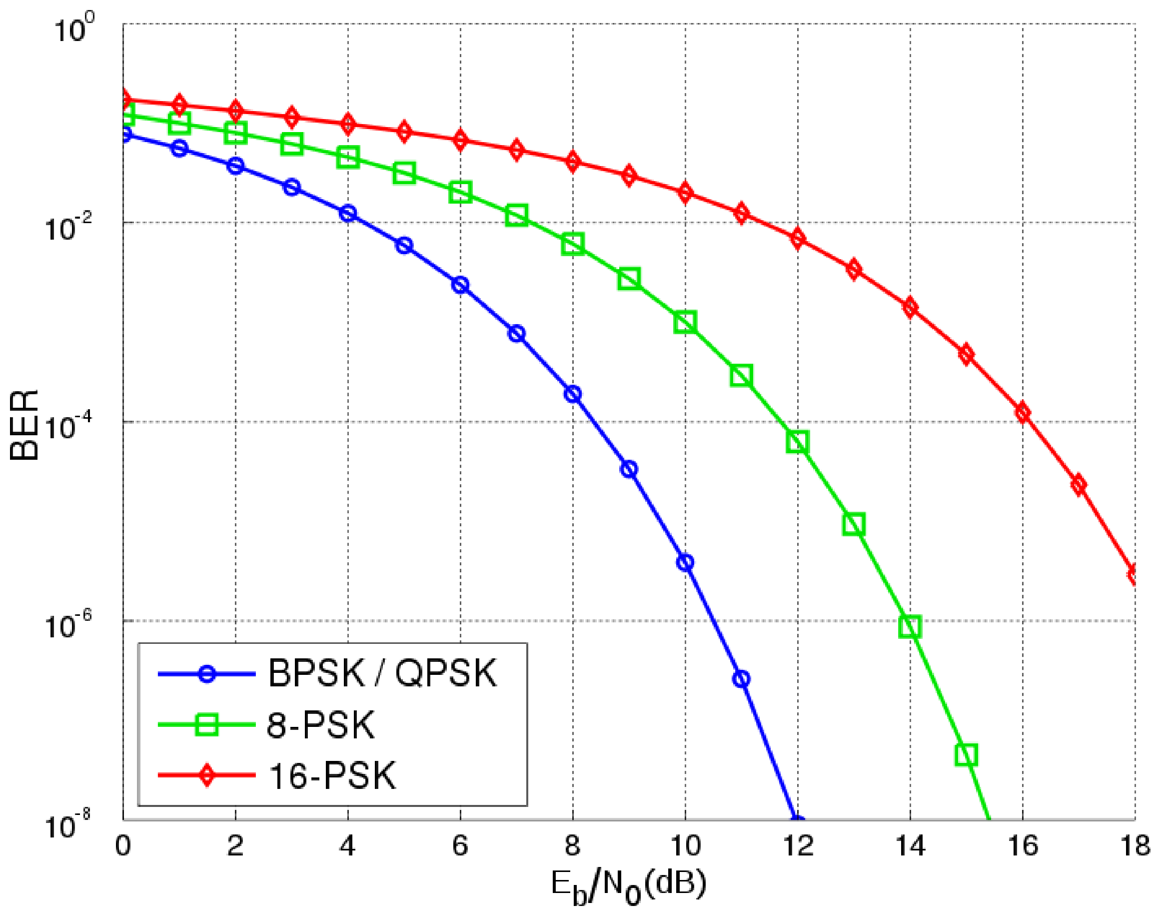
$$p_{\theta_r}\left(\theta_r\right) = \frac{1}{2\pi} e^{-2\gamma_s \sin^2\theta_r} \int_0^\infty V e^{-\left(V - \sqrt{4\gamma_s}\cos\theta_r\right)^2/2} dV,$$

$V = \sqrt{r_1^2 + r_2^2}$,

$\theta_r = \tan^{-1}\left(r_2/r_1\right)$,

$\gamma_s = \frac{E_s}{N_0}$ and

$r_1 \sim N\left(\sqrt{E_s}, N_0/2\right)$ and $r_2 \sim N\left(0, N_0/2\right)$ are jointly-Gaussian random variables.



Bit-error rate curves for BPSK, QPSK, 8-PSK and 16-PSK, AWGN channel.

This may be approximated for high M and high E_b / N_0 by:

$$P_s \approx 2Q\left(\sqrt{2\gamma_s}\sin\frac{\pi}{M}\right).$$

The bit-error probability for M-PSK can only be determined exactly once the bit-mapping is known. However, when Gray coding is used, the most probable error from one symbol to the next produces only a single bit-error and

$$P_b \approx \frac{1}{k}P_s.$$

(Using Gray coding allows us to approximate the Lee distance of the errors as the Hamming distance of the errors in the decoded bitstream, which is easier to implement in hardware.)

The graph on the left compares the bit-error rates of BPSK, QPSK (which are the same, as noted above), 8-PSK and 16-PSK. It is seen that higher-order modulations exhibit higher error-rates; in exchange however they deliver a higher raw data-rate.

Bounds on the error rates of various digital modulation schemes can be computed with application of the union bound to the signal constellation.

Differential phase-shift keying (DPSK)

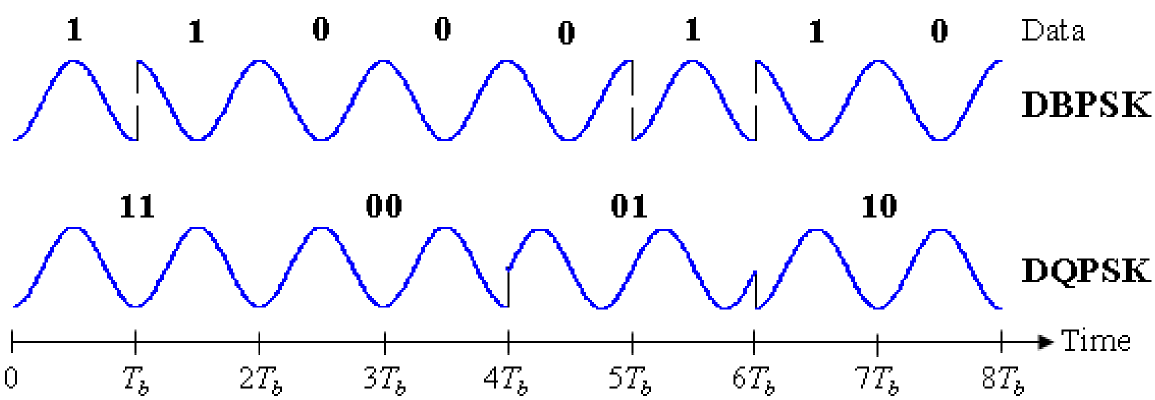
Differential encoding

Differential phase shift keying (DPSK) is a common form of phase modulation that conveys data by changing the phase of the carrier wave. As mentioned for BPSK and QPSK there is an ambiguity of phase if the constellation is rotated by some effect in the communications channel through which the signal passes. This problem can be overcome by using the data to *change* rather than *set* the phase.

For example, in differentially-encoded BPSK a binary '1' may be transmitted by adding 180° to the current phase and a binary '0' by adding 0° to the current phase. Another variant of DPSK is Symmetric Differential Phase Shift keying, SDPSK, where encoding would be +90° for a '1' and -90° for a '0'.

In differentially-encoded QPSK (DQPSK), the phase-shifts are 0°, 90°, 180°, -90° corresponding to data '00', '01', '11', '10'. This kind of encoding may be demodulated in the same way as for non-differential PSK but the phase ambiguities can be ignored. Thus, each received symbol is demodulated to one of the M points in the constellation and a comparator then computes the difference in phase between this received signal and the preceding one. The difference encodes the data as described above. Symmetric Differential Quadrature Phase Shift Keying (SDQPSK) is like DQPSK, but encoding is symmetric, using phase shift values of -135°, -45°, +45° and +135°.

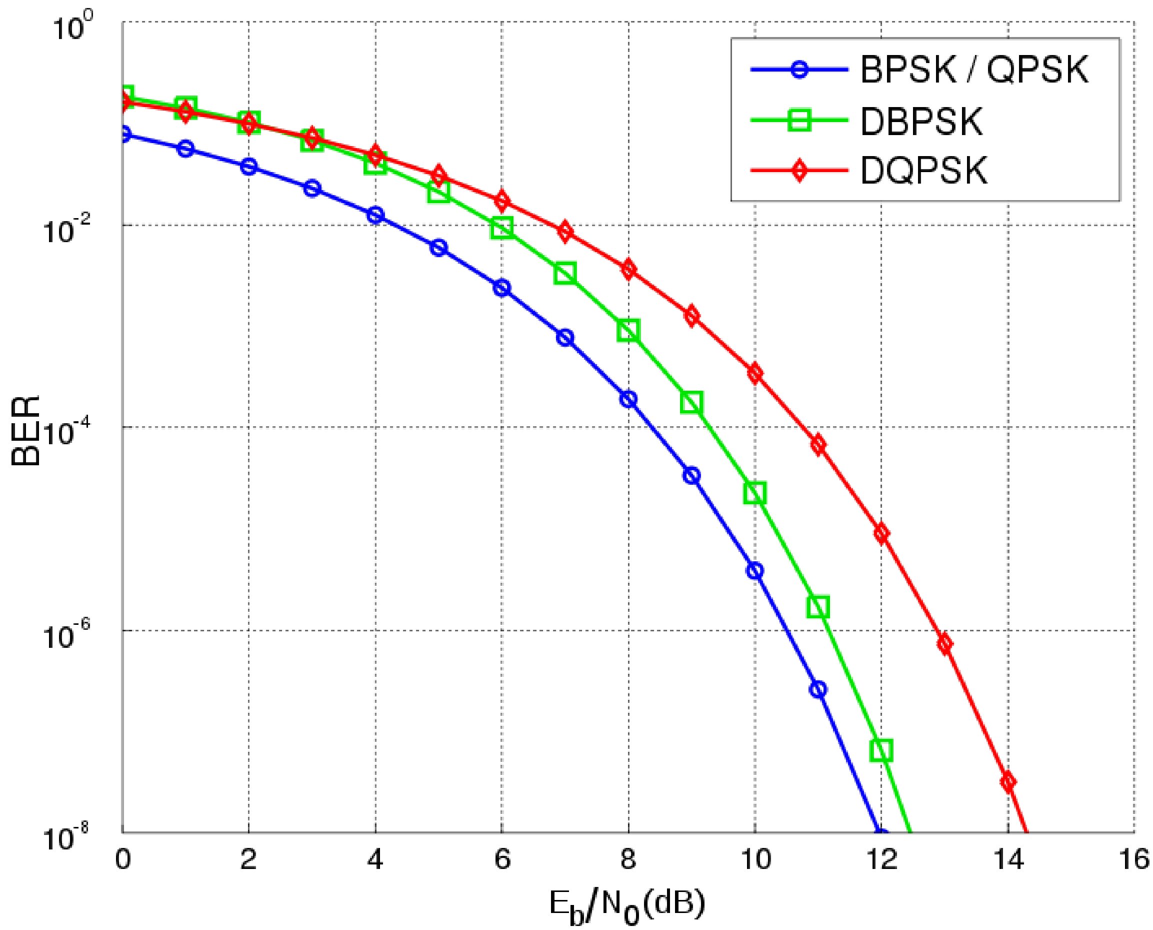
The modulated signal is shown below for both DBPSK and DQPSK as described above. In the figure, it is assumed that the *signal starts with zero phase*, and so there is a phase shift in both signals at $t = 0$.



Timing diagram for DBPSK and DQPSK. The binary data stream is above the DBPSK signal. The individual bits of the DBPSK signal are grouped into pairs for the DQPSK signal, which only changes every $T_s = 2T_b$.

Analysis shows that differential encoding approximately doubles the error rate compared to ordinary *M*-PSK but this may be overcome by only a small increase in E_b / N_0. Furthermore, this analysis (and the graphical results below) are based on a system in which the only corruption is additive white Gaussian noise(AWGN). However, there will also be a physical channel between the transmitter and receiver in the communication system. This channel will, in general, introduce an unknown phase-shift to the PSK signal; in these cases the differential schemes can yield a *better* error-rate than the ordinary schemes which rely on precise phase information.

Demodulation



BER comparison between DBPSK, DQPSK and their non-differential forms using gray-coding and operating in white noise.

For a signal that has been differentially encoded, there is an obvious alternative method of demodulation. Instead of demodulating as usual and ignoring carrier-phase ambiguity, the phase between two successive received symbols is compared and used to determine what the data must have been. When differential encoding is used in this manner, the scheme is known as differential phase-shift keying (DPSK). Note that this is subtly different to just differentially-encoded PSK since, upon reception, the received symbols are *not* decoded one-by-one to constellation points but are instead compared directly to one another.

Call the received symbol in the k^{th} timeslot r_k and let it have phase φ_k. Assume without loss of generality that the phase of the carrier wave is zero. Denote the AWGN term as n_k. Then

$$r_k = \sqrt{E_s}e^{j\phi_k} + n_k.$$

The decision variable for the $k - 1^{th}$ symbol and the k^{th} symbol is the phase difference between r_k and r_{k-1}. That is, if r_k is projected onto r_{k-1}, the decision is taken on the phase of the resultant complex number:

$$r_k r_{k-1}^* = E_s e^{j(\theta_k - \theta_{k-1})} + \sqrt{E_s}e^{j\theta_k}n_{k-1}^* + \sqrt{E_s}e^{-j\theta_{k-1}}n_k + n_k n_{k-1}$$

where superscript * denotes complex conjugation. In the absence of noise, the phase of this is $\theta_k - \theta_{k-1}$, the phase-shift between the two received signals which can be used to determine the data transmitted.

The probability of error for DPSK is difficult to calculate in general, but, in the case of DBPSK it is:

$$P_b = \frac{1}{2}e^{-E_b/N_0},$$

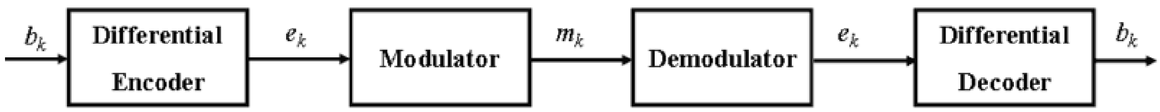
which, when numerically evaluated, is only slightly worse than ordinary BPSK, particularly at higher E_b / N_0 values.

Using DPSK avoids the need for possibly complex carrier-recovery schemes to provide an accurate phase estimate and can be an attractive alternative to ordinary PSK.

In optical communications, the data can be modulated onto the phase of a laser in a differential way. The modulation is a laser which emits a continuous wave, and a Mach-Zehnder modulator which receives electrical binary data. For the case of BPSK for example, the laser transmits the field unchanged for binary '1', and with reverse polarity for '0'. The demodulator consists of a delay line interferometer which delays one bit, so two bits can be compared at one time. In further processing, a photo diode is used to transform the optical field into an electric current, so the information is changed back into its original state.

The bit-error rates of DBPSK and DQPSK are compared to their non-differential counterparts in the graph to the right. The loss for using DBPSK is small enough compared to the complexity reduction that it is often used in communications systems that would otherwise use BPSK. For DQPSK though, the loss in performance compared to ordinary QPSK is larger and the system designer must balance this against the reduction in complexity.

Example: Differentially-encoded BPSK

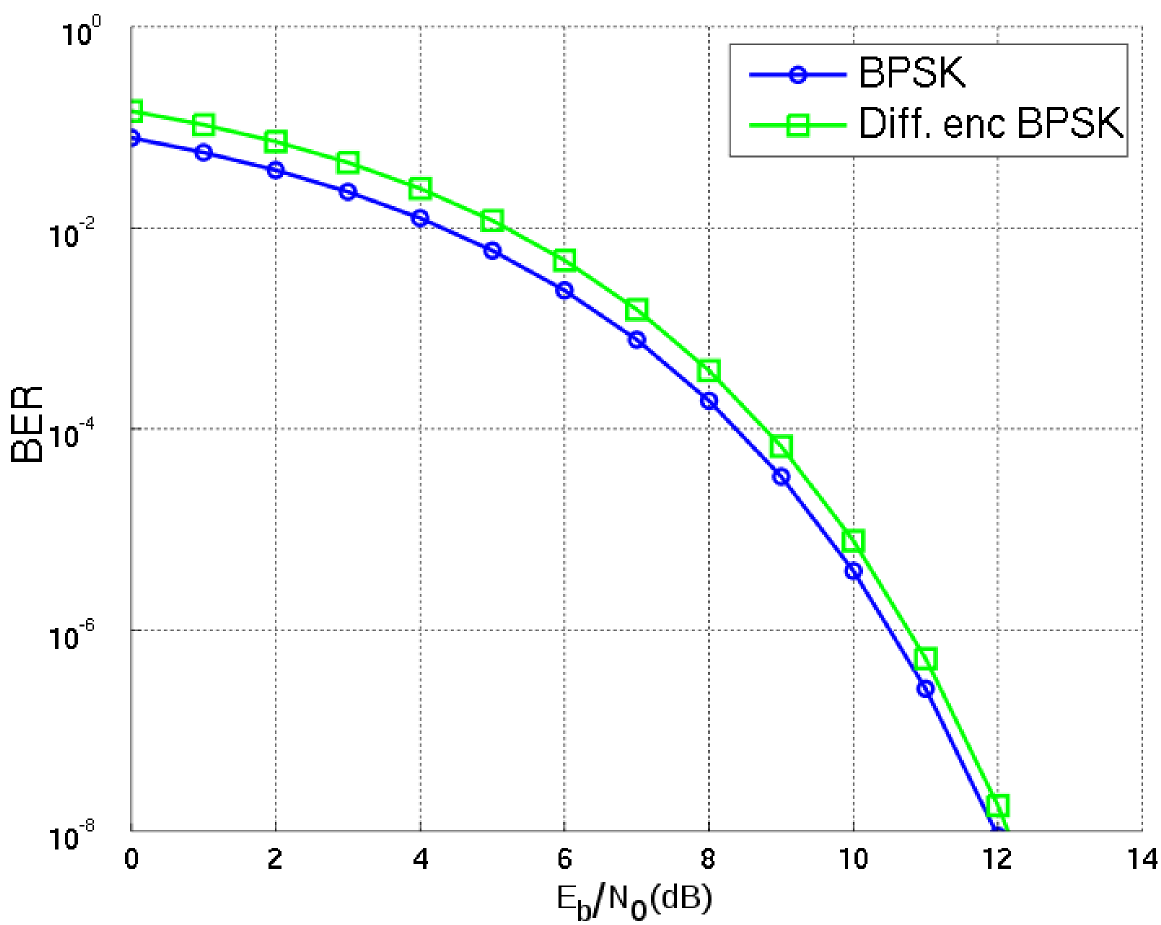


Differential encoding/decoding system diagram.

At the k^{th} time-slot call the bit to be modulated b_k, the differentially-encoded bit e_k and the resulting modulated signal $m_k(t)$. Assume that the constellation diagram positions the symbols at ±1 (which is BPSK). The differential encoder produces:

$$e_k = e_{k-1} \oplus b_k$$

where $\oplus$ indicates binary or modulo-2 addition.



BER comparison between BPSK and differentially-encoded BPSK with gray-coding operating in white noise.

So e_k only changes state (from binary '0' to binary '1' or from binary '1' to binary '0') if b_k is a binary '1'. Otherwise it remains in its previous state. This is the description of differentially-encoded BPSK given above.

The received signal is demodulated to yield $e_k = \pm 1$ and then the differential decoder reverses the encoding procedure and produces:

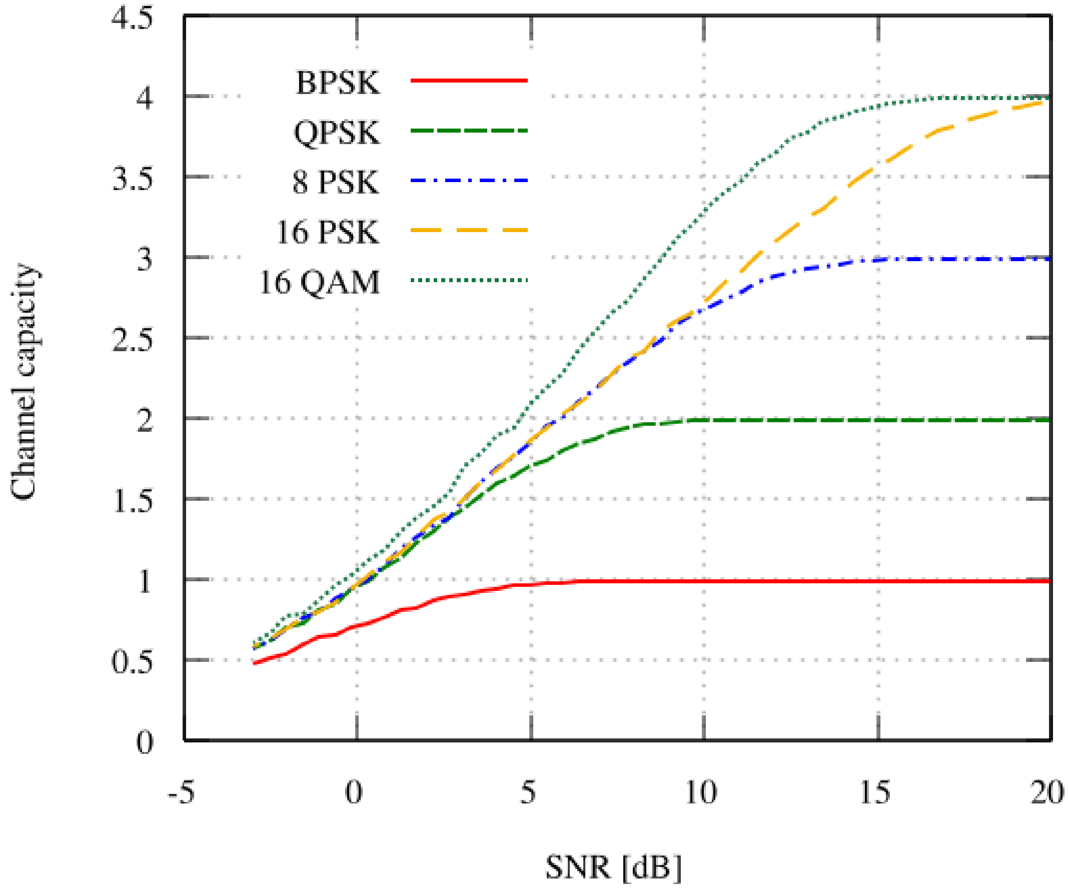
$b_k = e_k \oplus e_{k-1}$ since binary subtraction is the same as binary addition.

Therefore, $b_k = 1$ if e_k and e_{k-1} differ and $b_k = 0$ if they are the same. Hence, if both e_k and e_{k-1} are *inverted*, b_k will still be decoded correctly. Thus, the 180° phase ambiguity does not matter.

Differential schemes for other PSK modulations may be devised along similar lines. The waveforms for DPSK are the same as for differentially-encoded PSK given above since the only change between the two schemes is at the receiver.

The BER curve for this example is compared to ordinary BPSK on the right. As mentioned above, whilst the error-rate is approximately doubled, the increase needed in E_b / N_0 to overcome this is small. The increase in E_b / N_0 required to overcome differential modulation in coded systems, however, is larger - typically about 3 dB. The performance degradation is a result of noncoherent transmission - in this case it refers to the fact that tracking of the phase is completely ignored.

Channel capacity



Given a fixed bandwidth, channel capacity vs. SNR for some common modulation schemes

Like all M-ary modulation schemes with $M = 2^b$ symbols, when given exclusive access to a fixed bandwidth, the channel capacity of any phase shift keying modulation scheme rises to a maximum of *b* bits per symbol as the SNR increases.

Chapter- 9

On-off Keying, Minimum-Shift Keying & Continuous Phase Modulation

On-off keying

On-off keying (**OOK**) the simplest form of amplitude-shift keying (ASK) modulation that represents digital data as the presence or absence of a carrier wave. In its simplest form, the presence of a carrier for a specific duration represents a binary one, while its absence for the same duration represents a binary zero. Some more sophisticated schemes vary these durations to convey additional information. It is analogous to unipolar encoding line code.

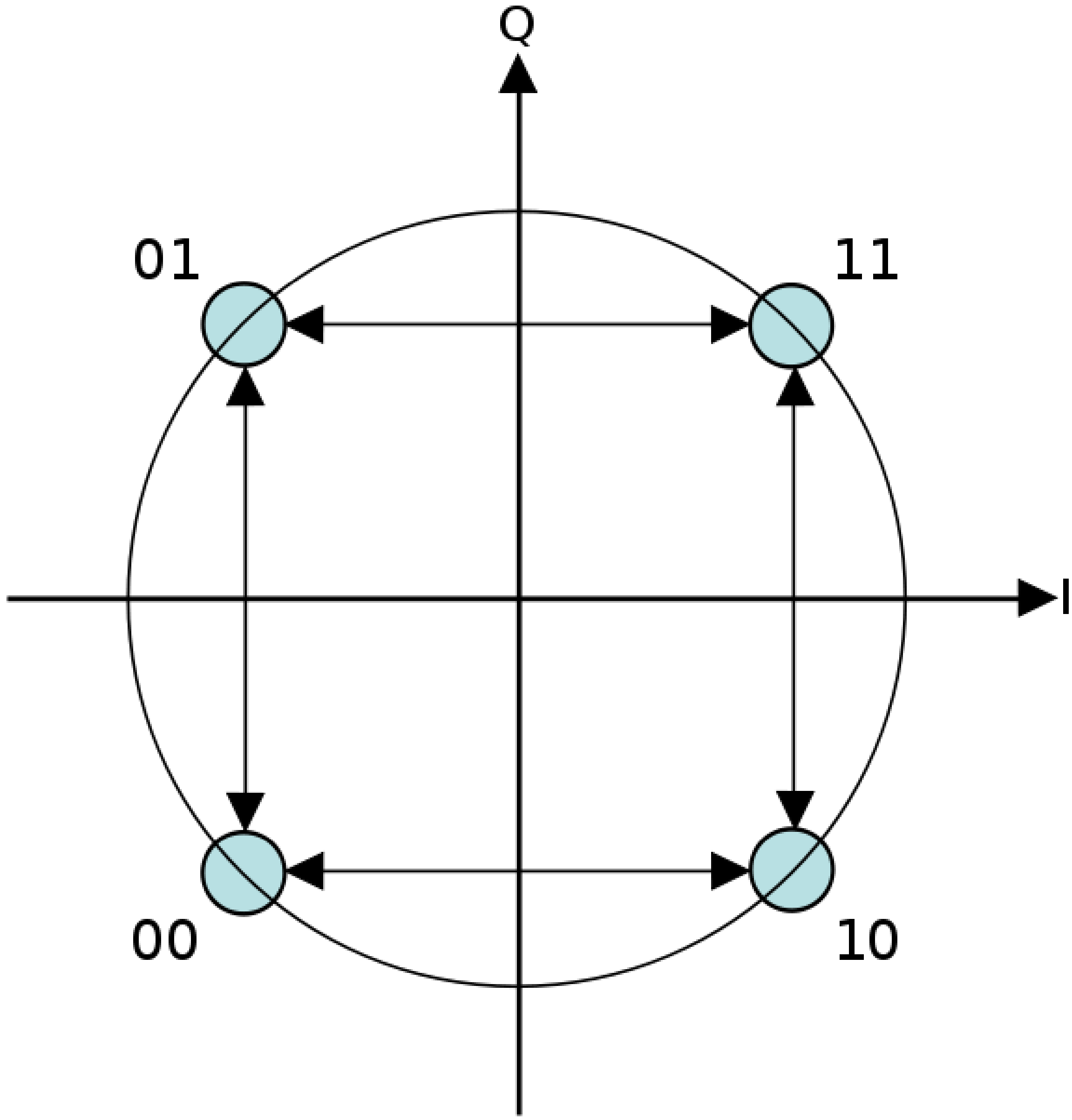
On-off keying is most commonly used to transmit Morse code over radio frequencies (referred to as CW (continuous wave) operation), although in principle any digital encoding scheme may be used. OOK has been used in the ISM bands to transfer data between computers, for example.

OOK is more spectrally efficient than FSK, but more sensitive to noise.

In addition to RF carrier waves, OOK is also used in optical communication systems (e.g. IrDA).

In aviation, some possibly unmanned airports have equipment that let pilots key their VHF radio a number of times in order to request an Automatic Terminal Information Service broadcast, or turn on runway lights.

Minimum-shift keying



Mapping changes in continuous phase

In digital modulation, **minimum-shift keying (MSK)** is a type of continuous-phase frequency-shift keying that was developed in the late 1950s and 1960s. Similar to OQPSK, MSK is encoded with bits alternating between quaternary components, with the Q component delayed by half the symbol period. However, instead of square pulses as OQPSK uses, MSK encodes each bit as a half sinusoid. This results in a constant-modulus signal, which reduces problems caused by non-linear distortion. In addition to being viewed as related to OQPSK, MSK can also be viewed as a continuous phase frequency shift keyed (CPFSK) signal with a frequency separation of one-half the bit rate.

Mathematical representation

The resulting signal is represented by the formula

$$s(t) = a_I(t)\cos\left(\frac{\pi t}{2T}\right)\cos(2\pi f_c t) - a_Q(t)\sin\left(\frac{\pi t}{2T}\right)\sin(2\pi f_c t)$$

where $a_I(t)$ and $a_Q(t)$ encode the even and odd information respectively with a sequence of square pulses of duration $2T$. Using the trigonometric identity, this can be rewritten in a form where the phase and frequency modulation are more obvious,

$$s(t) = \cos[2\pi f_c t + b_k(t)\frac{\pi t}{2T} + \phi_k]$$

where $b_k(t)$ is +1 when $a_I(t) = a_Q(t)$ and -1 if they are of opposite signs, and φ_k is 0 if $a_I(t)$ is 1, and π otherwise. Therefore, the signal is modulated in frequency and phase, and the phase continuously and linearly changes.

Gaussian minimum-shift keying

In digital communication, **Gaussian minimum shift keying** or **GMSK** is a continuous-phase frequency-shift keying modulation scheme. It is similar to standard minimum-shift keying (MSK); however the digital data stream is first shaped with a Gaussian filter before being applied to a frequency modulator. This has the advantage of reducing sideband power, which in turn reduces out-of-band interference between signal carriers in adjacent frequency channels. However, the Gaussian filter increases the modulation memory in the system and causes intersymbol interference, making it more difficult to discriminate between different transmitted data values and requiring more complex channel equalization algorithms such as an adaptive equalizer at the receiver. GMSK has high spectral efficiency, but it needs a higher power level than QPSK, for instance, in order to reliably transmit the same amount of data.

GMSK is most notably used in the Global System for Mobile Communications (GSM).

Continuous phase modulation

Continuous phase modulation (**CPM**) is a method for modulation of data commonly used in wireless modems. In contrast to other coherent digital phase modulation techniques where the carrier phase abruptly resets to zero at the start of every symbol (e.g. M-PSK), with CPM the carrier phase is modulated in a continuous manner. For instance, with QPSK the carrier instantaneously jumps from a sine to a cosine (i.e. a 90 degree phase shift) whenever one of the two message bits of the current symbol differs from the two message bits of the previous symbol. This discontinuity requires a relatively large percentage of the power to occur outside of the intended band (e.g., high fractional out-of-band power), leading to poor spectral efficiency. Furthermore, CPM is typically implemented as a constant-envelope waveform, i.e. the transmitted carrier power is constant. Therefore, CPM is attractive because the phase continuity yields high spectral efficiency, and the constant-envelope yields excellent power efficiency. The primary drawback is the high implementation complexity required for an optimal receiver.

Phase Memory

Each symbol is modulated by gradually changing the phase of the carrier from the starting value to the final value, over the symbol duration. The modulation and demodulation of CPM is complicated by the fact that the initial phase of each symbol is determined by the cumulative total phase of all previous transmitted symbols, which is known as the *phase memory*. Therefore, the optimal receiver cannot make decisions on any isolated symbol without taking the entire sequence of transmitted symbols into account. This requires a Maximum Likelihood Sequence Estimator (MLSE), which is efficiently implemented using the Viterbi algorithm.

Phase Trajectory

Minimum-shift keying (MSK) is another name for CPM with an excess bandwidth of 1/2 and a linear *phase trajectory*. Although this linear phase trajectory is continuous, it is not *smooth* since the derivative of the phase is not continuous. The spectral efficiency of CPM can be further improved by using a smooth phase trajectory. This is typically accomplished by filtering the phase trajectory prior to modulation, commonly using a Raised Cosine or a Gaussian filter. The raised cosine filter has zero crossings offset by exactly one symbol time, and so it can yield a *full-response* CPM waveform that prevents Intersymbol Interference (ISI).

Partial Response CPM

Partial-response signaling, such as duo-binary signaling, is a form of intentional ISI where a certain number of adjacent symbols interfere with each symbol in a controlled manner. A MLSE must be used to optimally demodulate any signal in the presence of ISI. Whenever the amount of ISI is known, such as with any partial-response signaling scheme, MLSE can be used to determine the exact symbol sequence (in the absence of noise). Since the optimal demodulation of full-response CPM already requires MLSE detection, using partial-response signaling requires little additional complexity, but can afford a comparatively smoother phase trajectory, and thus, even greater spectral efficiency. One extremely popular form of partial-response CPM is GMSK, which is used by GSM in most of the world's 2nd generation cell phones. It is also used in 802.11 FHSS, Bluetooth, and many other proprietary wireless modems.

Continuous-phase frequency-shift keying

Continuous-phase frequency-shift keying (CPFSK) is a commonly-used variation of frequency-shift keying (FSK), which is itself a special case of analog frequency modulation. FSK is a method of modulating digital data onto a sinusoidal carrier wave, encoding the information present in the data to variations in the carrier's instantaneous frequency between one of two frequencies (referred to as the space frequency and mark frequency). In general, a standard FSK signal does not have continuous phase, as the modulated waveform switches instantaneously between two sinusoids with different frequencies.

As the name suggests, the phase of a CPFSK is in fact continuous; this attribute is desirable for signals that are to be transmitted over a bandlimited channel, as discontinuities in a signal introduce wideband frequency components. In addition, some classes of amplifiers exhibit nonlinear behavior when driven with nearly-discontinuous signals; this could have undesired effects on the shape of the transmitted signal.

Theory

If a finitely-valued digital signal to be transmitted (the message) is $m(t)$, then the corresponding CPFSK signal is

$$s(t) = A_c \cos\left(2\pi f_c t + D_f \int_{-\infty}^{t} m(\alpha)d\alpha\right)$$

where A_c represents the amplitude of the CPFSK signal, f_c is the base carrier frequency, and D_f is a parameter that controls the frequency deviation of the modulated signal. The integral located inside of the cosine's argument is what gives the CPFSK signal its continuous phase; an integral over any finitely-valued function (which $m(t)$ is assumed to be) will not contain any discontinuities. If the message signal is assumed to be causal, then the limits on the integral change to a lower bound of zero and a higher bound of t.

Note that this does not mean that $m(t)$ must be continuous; in fact, most ideal digital data waveforms contain discontinuities. However, even a discontinuous message signal will generate a proper CPFSK signal.

Chapter- 10

Pulse-Position Modulation & Trellis Modulation

Pulse-position modulation

Pulse-position modulation (PPM) is a form of signal modulation in which M message bits are encoded by transmitting a single pulse in one of 2^M possible time-shifts. This is repeated every T seconds, such that the transmitted bit rate is M/T bits per second. It is primarily useful for optical communications systems, where there tends to be little or no multipath interference.

Synchronization

One of the key difficulties of implementing this technique is that the receiver must be properly synchronized to align the local clock with the beginning of each symbol. Therefore, it is often implemented differentially as *differential pulse-position modulation*, where by each pulse position is encoded relative to the previous , such that the receiver must only measure the difference in the arrival time of successive pulses. It is possible to limit the propagation of errors to adjacent symbols, so that an error in measuring the differential delay of one pulse will affect only two symbols, instead of affecting all successive measurements.

Sensitivity to multipath interference

Aside from the issues regarding receiver synchronization, the key disadvantage of PPM is that it is inherently sensitive to multipath interference that arises in channels with frequency-selective fading, whereby the receiver's signal contains one or more echoes of each transmitted pulse. Since the information is encoded in the time of arrival (either differentially, or relative to a common clock), the presence of one or more echoes can make it extremely difficult, if not impossible, to accurately determine the correct pulse position corresponding to the transmitted pulse.

Non-coherent detection

One of the principal advantages of PPM is that it is an M-ary modulation technique that can be implemented non-coherently, such that the receiver does not need to use a phase-locked loop (PLL) to track the phase of the carrier. This makes it a suitable candidate for optical communications systems, where coherent phase modulation and detection are difficult and extremely expensive. The only other common M-ary non-coherent modulation technique is M-ary Frequency Shift Keying (M-FSK), which is the frequency-domain dual to PPM.

PPM vs. M-FSK

PPM and M-FSK systems with the same bandwidth, average power, and transmission rate of M/T bits per second have identical performance in an AWGN (Additive White Gaussian Noise) channel. However, their performance differs greatly when comparing frequency-selective and frequency-flat fading channels. Whereas frequency-selective fading produces echoes that are highly disruptive for any of the M time-shifts used to encode PPM data, it selectively disrupts only some of the M possible frequency-shifts used to encode data for M-FSK. On the other hand, frequency-flat fading is more disruptive for M-FSK than PPM, as all M of the possible frequency-shifts are impaired by fading, while the short duration of the PPM pulse means that only a few of the M time-shifts are heavily impaired by fading.

Optical communications systems (even wireless ones) tend to have weak multipath distortions, and PPM is a viable modulation scheme in many such applications.

Applications for RF communications

Narrowband RF (radio frequency) channels with low power and long wavelengths (i.e., low frequency) are affected primarily by flat fading, and PPM is better suited than M-FSK to be used in these scenarios. One common application with these channel characteristics, first used in the early 1960s, is the radio control of model aircraft, boats and cars. PPM is employed in these systems, with the position of each pulse representing the angular position of an analogue control on the transmitter, or possible states of a binary switch. The number of pulses per frame gives the number of controllable channels available. The advantage of using PPM for this type of application is that the electronics required to decode the signal are extremely simple, which leads to small, light-weight receiver/decoder units. (Model aircraft require parts that are as lightweight as possible).

Servos made for model radio control include some of the electronics required to convert the pulse to the motor position – the receiver is merely required to demultiplex the separate channels and feed the pulses to each servo.

More sophisticated R/C systems are now often based on pulse-code modulation, which is more complex but offers greater flexibility and reliability.

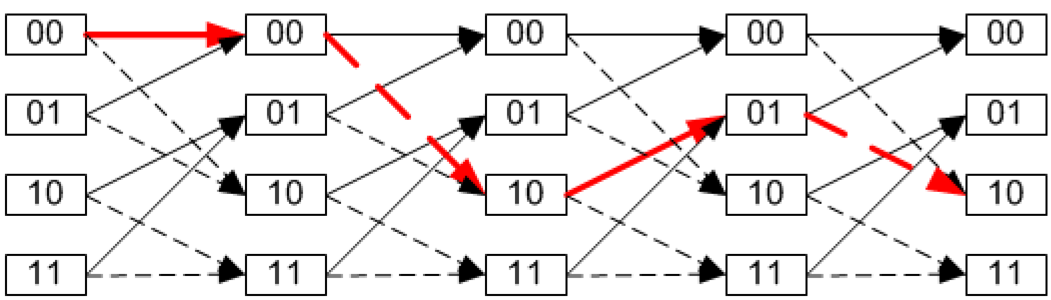
Pulse position modulation is also used for communication to the ISO/IEC 15693 contactless smart card as well as the HF implementation of the EPC Class 1 protocol for RFID tags.

Trellis modulation

In telecommunication, **trellis modulation** (also known as **trellis coded modulation**, or simply **TCM**) is a modulation scheme which allows highly efficient transmission of information over band-limited channels such as telephone lines. Trellis modulation was invented by Gottfried Ungerboeck working for IBM in the 1970s, and first described in a conference paper in 1976; but it went largely unnoticed until he published a new detailed exposition in 1982 which achieved sudden widespread recognition.

In the late 1980s, modems operating over plain old telephone service (*POTS*) typically achieved 9.6 kbit/s by employing 4 bits per symbol QAM modulation at 2,400 baud (symbols/second). This bit rate ceiling existed despite the best efforts of many researchers, and some engineers predicted that without a major upgrade of the public phone infrastructure, the maximum achievable rate for a POTS modem might be 14 kbit/s for two-way communication (3,429 baud × 4 bits/symbol, using QAM). However, 14 kbit/s is only 40% of the theoretical maximum bit rate predicted by Shannon's Theorem for POTS lines (approximately 35 kbit/s).

A new modulation method



Trellis diagram.

The name *trellis* was coined because a state diagram of the technique, when drawn on paper closely resembles the trellis lattice used in rose gardens. The scheme is basically a convolutional code of rates (r,r+1). Ungerboeck's unique contribution is to apply the parity check on a per symbol basis instead of the older technique of applying it to the bit stream then modulating the bits. The key idea he termed Mapping by Set Partitions. This idea was to group the symbols in a tree like fashion then separate them into two limbs of equal size. At each limb of the tree, the symbols were further apart. Although in multi-dimensions, it is hard to visualize, a simple one dimension example illustrates the basic procedure. Suppose the symbols are located at [1, 2, 3, 4, ...]. Then take all odd symbols and place them in one group, and the even symbols in the second group. This is not quite accurate because Ungerboeck was looking at the two dimensional problem, but the principle is the same, take every other one for each group and repeat the procedure for each tree limb. He next described a method of assigning the encoded bit stream onto the symbols in a very systematic procedure. Once this procedure was fully described, his next step was to program the algorithms into a computer and let the computer search for the best codes. The results were astonishing. Even the most simple code (4 state) produced error rates nearly 1,000 times lower than an equivalent uncoded system. For two years Ungerboeck kept these results private and only conveyed them to close colleagues. Finally, in 1982, Ungerboeck published a paper describing the principles of trellis modulation.

A flurry of research activity ensued, and by 1990 the International Telecommunication Union had published modem standards for the first trellis-modulated modem at 14.4 kbit/s (2,400 baud and 6 bits per symbol). Over the next several years further advances in encoding, plus a corresponding symbol rate increase from 2,400 to 3,429 baud, allowed modems to achieve rates up to 34.3 kbit/s (limited by maximum power regulations to 33.8 kbit/s). Today, the most common trellis-modulated V.34 modems use a 4-dimensional set partition which is achieved by treating two 2-dimensional symbols as a single lattice. This set uses 8, 16, or 32 state convolutional codes to squeeze the equivalent of 6 to 10 bits into each symbol sent by the modem (for example, 2,400 baud × 8 bits/symbol = 19,200 bit/s).

Once manufacturers introduced modems with trellis modulation, transmission rates increased to the point where interactive transfer of multimedia over the telephone became feasible (a 200 kilobyte image and a 5 megabyte song could be downloaded in less than 1 minute and 30 minutes, respectively). Sharing a floppy disk via a BBS could be done in just a few minutes, instead of an hour. Thus Ungerboeck's invention played a key role in the Information Age.