MODULATION IN DIGITAL TRANSMISSION

In modulation, a baseband signal is translated in frequency domain to certain center frequency.

Topics:

1. Linear Digital Modulation

- Complex Quadrature (I/Q) Modulation
- Constellations: QAM, QPSK
- Special Methods: CAP, VSB, Differential Modulation
- BER vs. Noise Analysis

2. Nonlinear Modulation Methods

- FSK Type (Nonlinear) Modulation methods: MSK, GMSK

3. Some Performance Comparisons

Source: Lee&Messerschmitt, Chapter 6.

I/Q-Modulation

Traditional linear modulation methods have the following problems:

- DSB and AM consumes too much frequency band, AM also transmit power.
- SSB has difficulties from implementation point of view.

Most of the important modulation methods in digital transmission systems are based on

- complex alphabets
- complex quadrature modulation (I/Q-modulation)

There are many possible complex alphabets, like QAM or PSK. Here, in the beginning, we consider the situation in a generic way, so the used alphabet is not so important.

Quadrature modulation is based on the fact that sine and cosine waveforms can be DSB-modulated and detected independently using different information signals.

I/Q-Modulation Principle

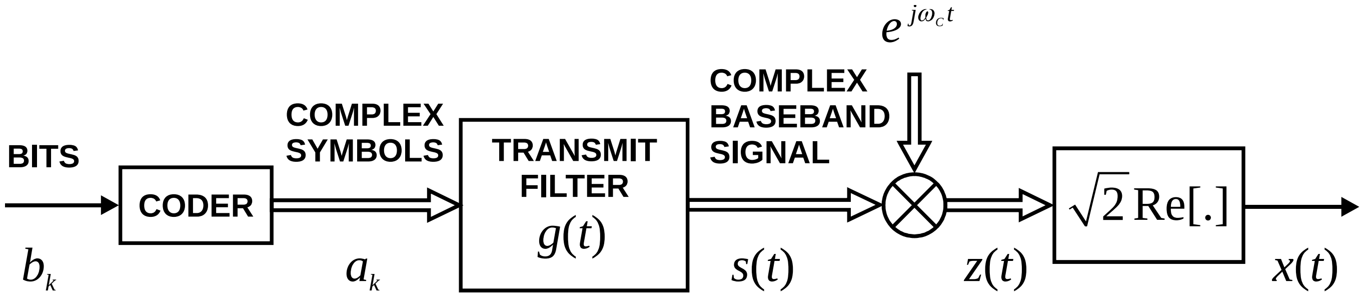
- Two independent real baseband signals (I and Q, In-phase and quadrature) are transmitted by modulating them into cosine and sine waveforms of the carrier frequency.

- For I- and Q-components, Nyquist pulse shaping principle (Overlapping pulses with zero-intersymbol interference, 0-ISI) is utilized in order to achieve high spectral efficiency.

  - The transmitter and receiver filters contribute to the pulse shaping in ideal case.

  - The receiver filter has also the task of suppressing neighbouring frequency channel signals; this is in practice done in several stages of the receiver.

Complex Quadrature Modulation Model



Here

$$s(t) = \sum_{m=-\infty}^{\infty} a_m g(t - mT)$$

$$x(t) = \sqrt{2}\,\mathrm{Re}\left[e^{j\omega_c t} \sum_{m=-\infty}^{\infty} a_m g(t - mT) \right]$$

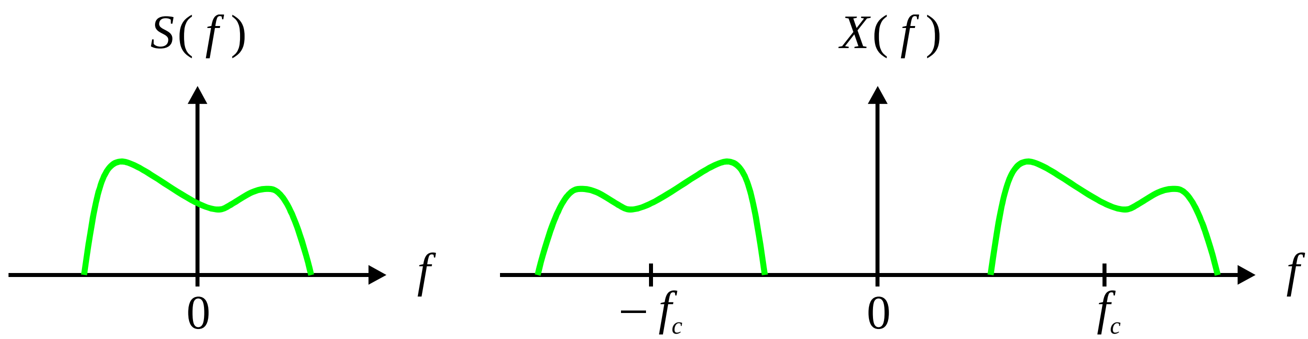
and

ω_c is the carrier angular frequency. The symbol sequence $\{a_m\}$ is assumed to be a random process, and for mathematical reasons time-limited,

$a_m = 0 \quad \text{when} \quad |m| > M$.

It can be proven that if the energy of $s(t)$ is finite, it is the same as the energy of $x(t)$.

Modulating signal spectrum — Modulated signal spectrum

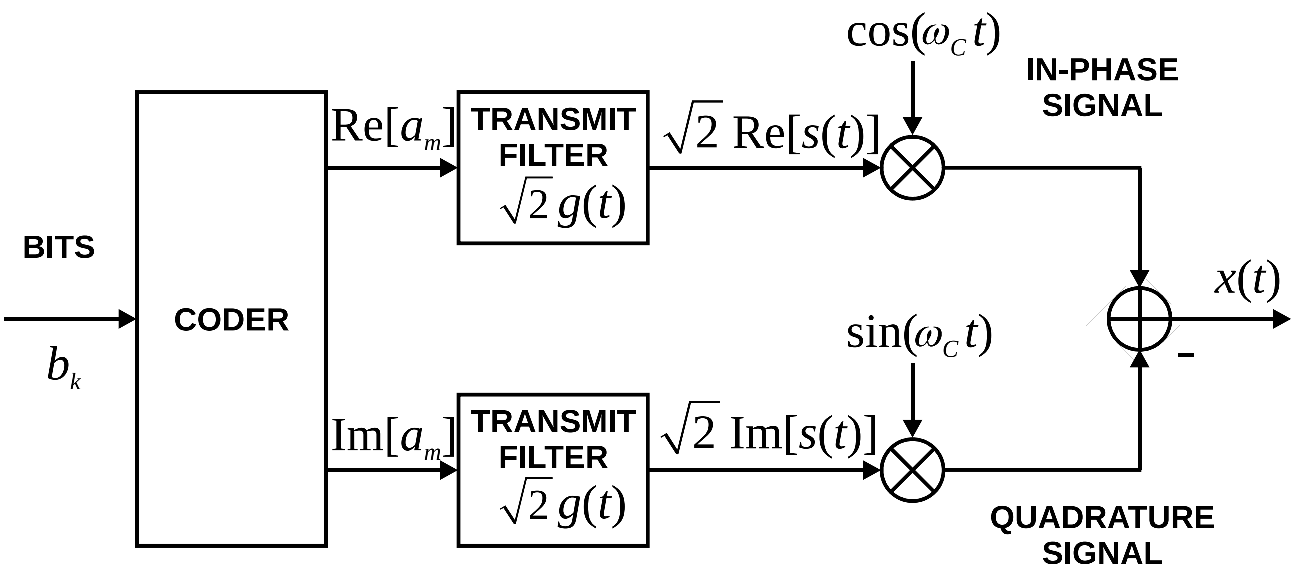


I/Q-Modulation

In practice, almost always $g(t)$ is a real pulse and then

$$x(t) = \sqrt{2}\cos(\omega_c t) \sum_{m=-\infty}^{\infty} \mathrm{Re}[a_m] g(t-mT)$$
$$-\sqrt{2}\sin(\omega_c t) \sum_{m=-\infty}^{\infty} \mathrm{Im}[a_m] g(t-mT)$$

Here sine and cosine components are modulated by the real and imaginary parts of the baseband signal.



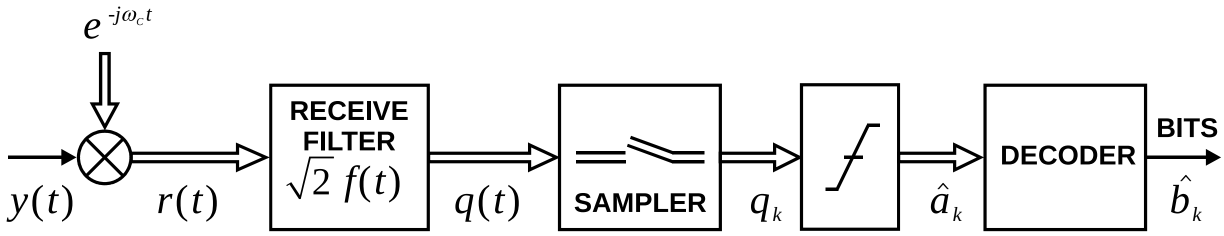
This is the practical way of doing I/Q modulation.

Complex notation is often used in this context mainly because of the simple representation.

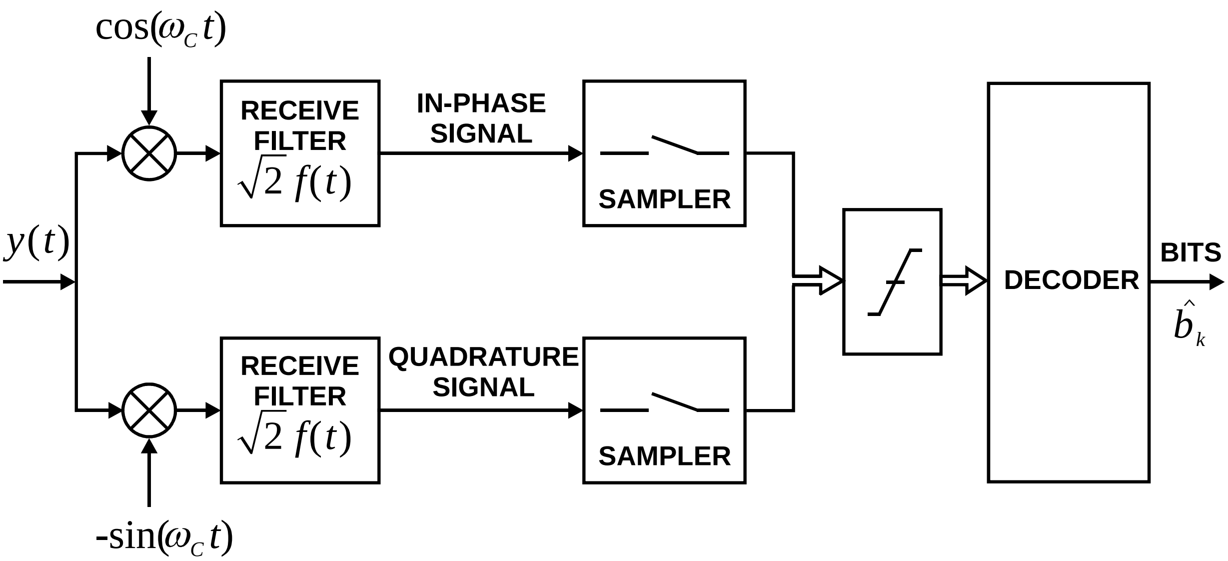
Passband I/Q Receiver

Two equivalent structures:

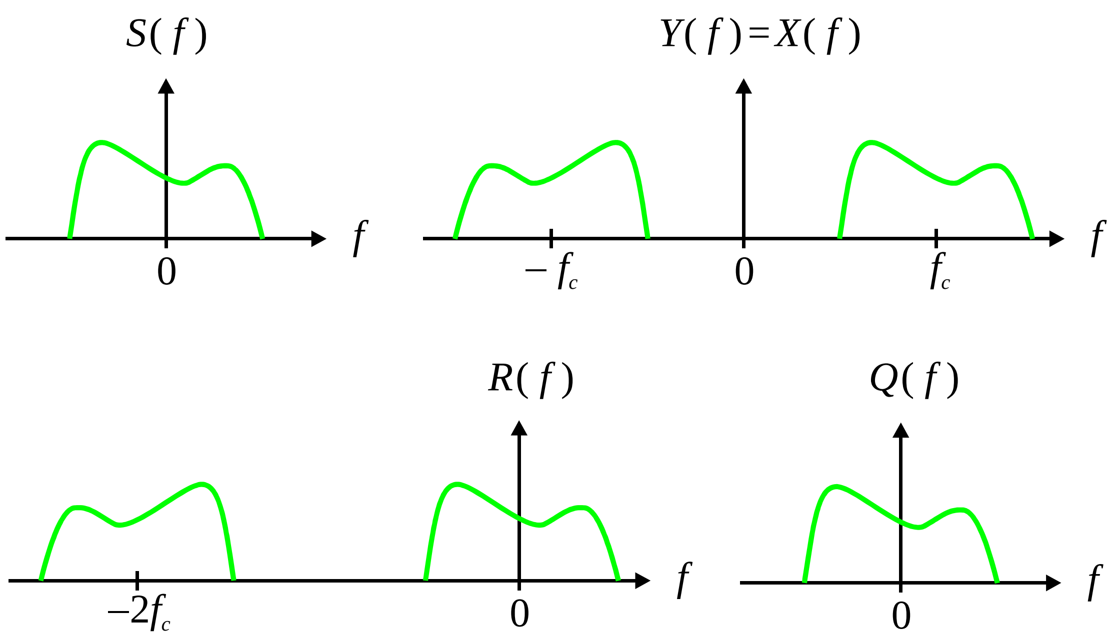
(a) Using complex notation:



(b) Using real signals:



Spectra:



Passband I/Q Receiver

Assumptions:

- $x(t)$ is deterministic and has Fourier transform
- ideal channel, noise free $\Rightarrow y(t) = x(t)$
- $f(t)$ is a lowpass filter that has unit gain in passband and removes the spectral components around the frequency $\pm 2 f_c$, and also noise and other interferences outside the information band.
- symbol rate sampling, ideally synchronized, no intersymbol interference

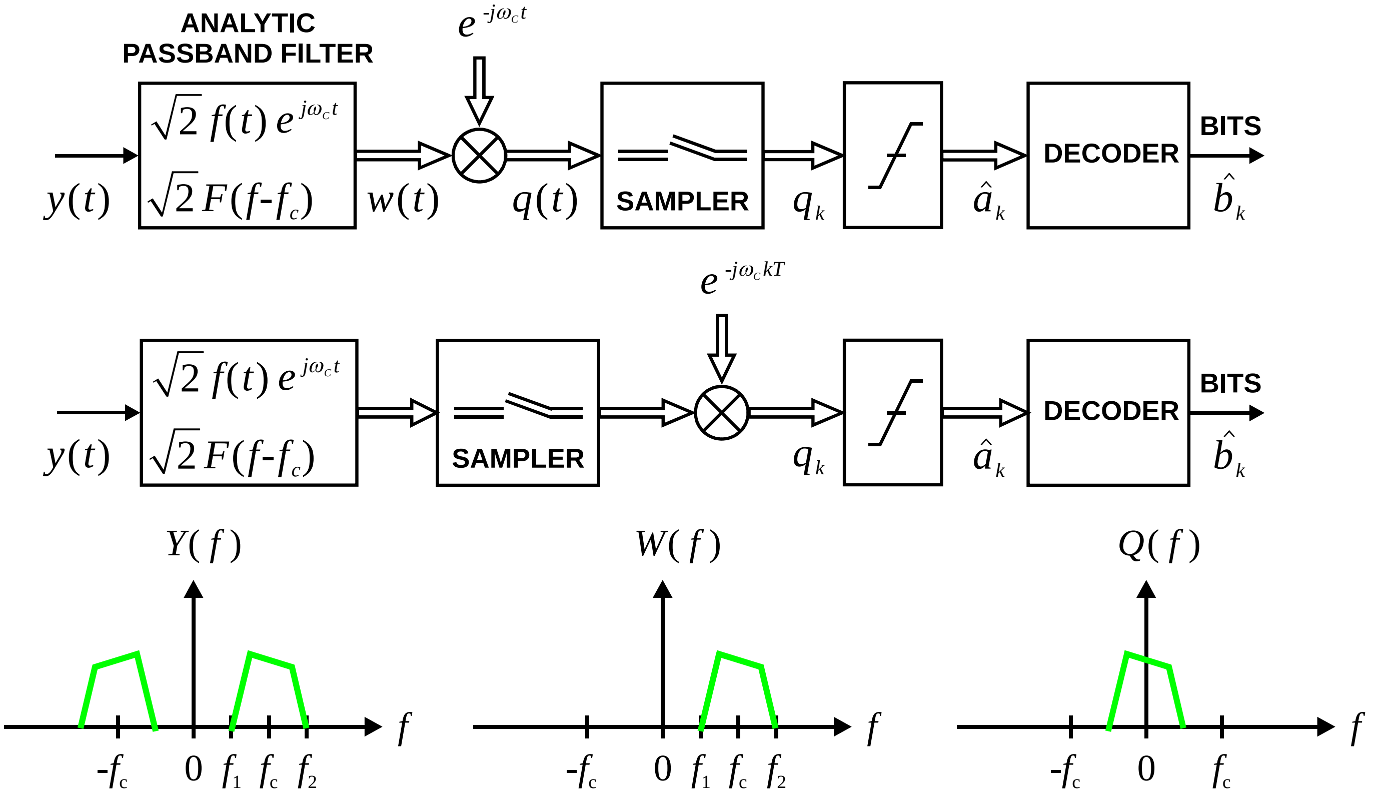
$\Rightarrow$ The received symbol values are equal to the original ones, $\hat{a}_k = a_k$.

The factors $\sqrt{2}$ ensure that the transmitted and received signal powers are the same.

It should be noted that the input of the slicer is complex valued and it can not be replaced by two real valued slicers, except in special cases (also important ones, like QAM).

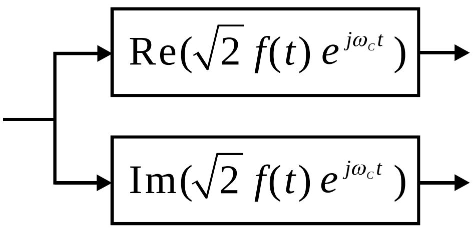
Practical receivers include mixing, fixed intermediate frequencies (e.g., the superheterodyne principle), etc. But, here we use a simplified model.

Alternative Receiver Structures



In the model of (a), the filter is placed before down-conversion. From the spectral figures, it is easy to see that this can be done.

The resulting *analytic bandpass* signal is a complex signal. Two real filters are needed for implementing the analytic bandpass filtering:

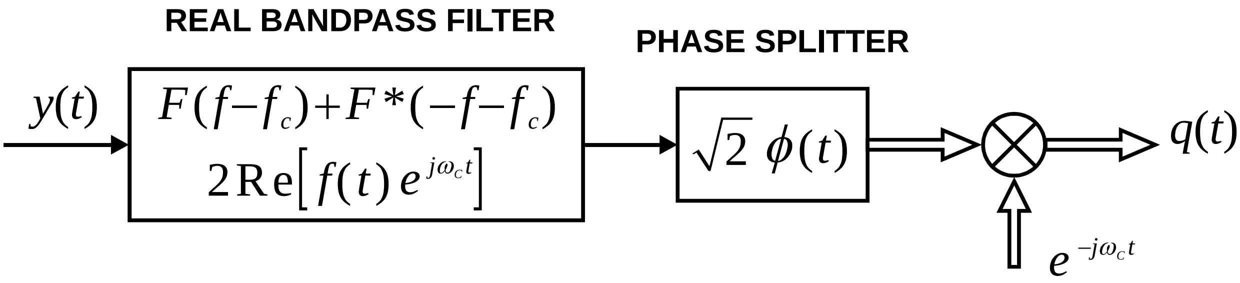


In figure (b), also the sampling is placed before the frequency shift. In this case bandpass sampling is utilized. In principle, the required sampling rate is proportional to the bandwidth of the analytic bandpass signal and doesn't depend on the center frequency.

In theory, if proper receiver filtering (matched filter) is implemented on the analog side and if the sampling clock is properly synchronized, symbol rate sampling can be used also in this bandpass sampling model. This may not be a practical model for implementation, but it is useful in some analytical considerations.

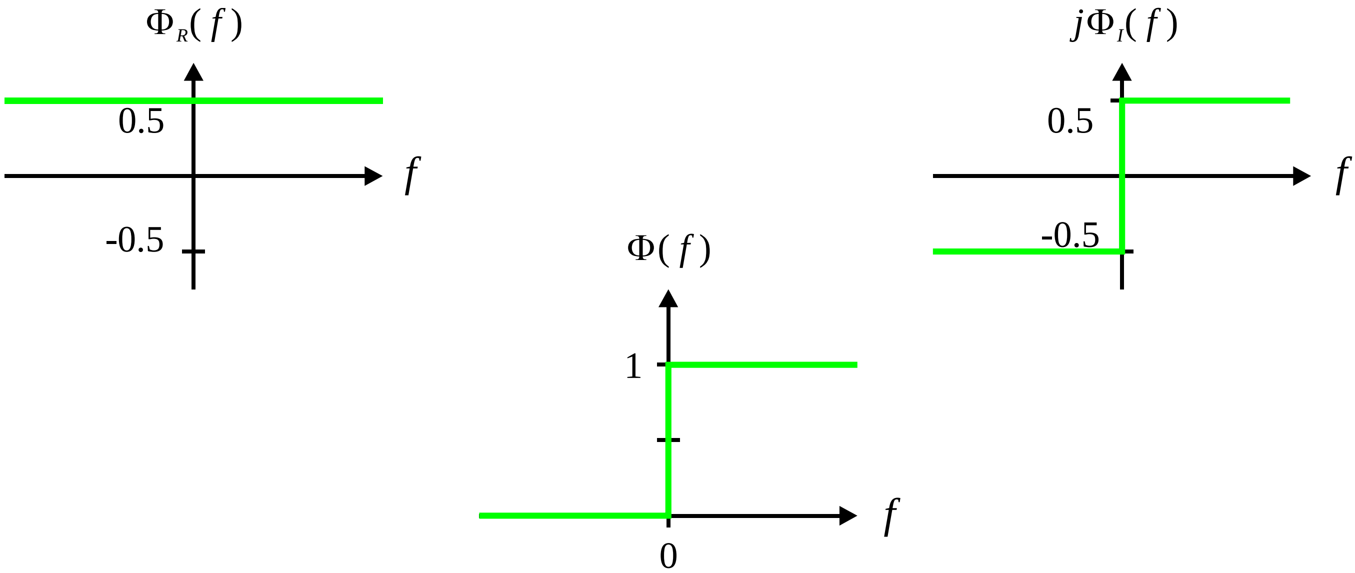
About Analytic Bandpass Filter Implementation

The analytic bandpass filter can be implemented as a cascade of a real bandpass filter and a phase splitter:



An ideal phase splitter has the transfer function:

$$\Phi(f) = \Phi_R(f) + j\Phi_I(f) = \frac{1}{2}\left[1 + j\hat{H}(f)\right]$$



The imaginary part can be implemeted with a (bandpass) Hilbert-transformer ($\hat{H}(f)$). Basically, what needs to be implemented is a 90-degree phase shift within the desired signal band. The frequency response of the phase splitter/Hilbert transformer in the stopband region of the real bandpass filter is insignificant.

Signal at the Sampler of the Receiver

Without noise and other disturbances the received signal is

$$y(t) = x(t) = \sqrt{2}\,\mathrm{Re}\left[e^{j\omega_c t} \sum_{k=-\infty}^{\infty} a_k h(t-kT) \right]$$

$$h(t) = b_E(t) * g(t) = \left\{ e^{-j\omega_c t} b(t) \right\} * g(t)$$

$$H(f) = B\left(f + f_c\right) G(f)$$

Here *h(t)* is the equivalent baseband pulse shape, which is in practise always complex in passband transmission.

The signal coming to the sampler is

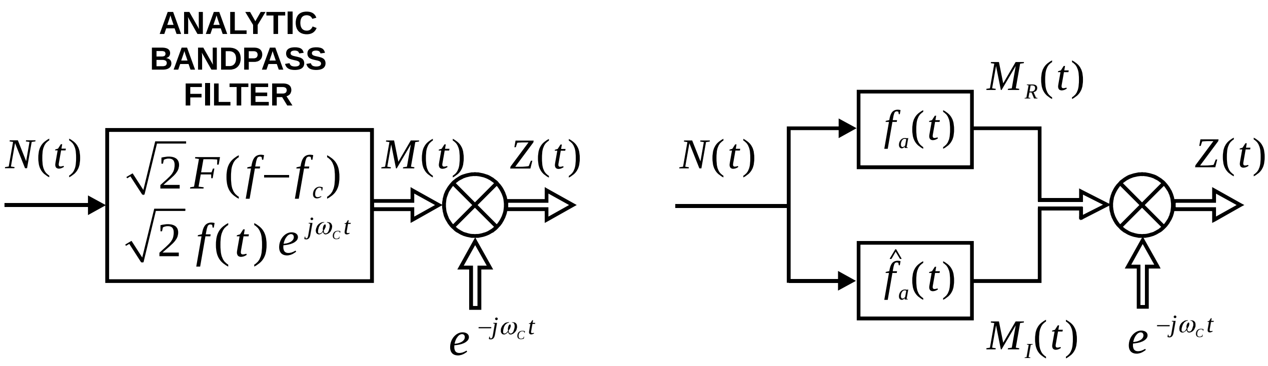
$$q(t) = \left[\sum_{k=-\infty}^{\infty} a_k h(t-kT) \right] * f(t) = \sum_{k=-\infty}^{\infty} a_k p(t-kT)$$

$$p(t) = h(t) * f(t)$$

This gives the overall pulse shape in case of passband transmission as a function of the transmit and receive filters and channel response. This is also complex in practise.

Receiver Noise Considerations

In the following, we examine the properties of the noise that appears at the sampler and slicer/detector blocks. We use the previously developed receiver structure based on the analytic bandpass filtering:



- We assume the AWGN-model for the channel noise $N(t)$.
- We also assume that the receiver filter $f(t)$ removes all the out-of-band signal components, as described earlier.
- We also consider a truly carrier-modulated system. (The following considereations do not apply in the baseband case.)

With these assumptions, noise after the receiver filter, $M(t)$, is Gaussian (remember that filtered Gaussian noise is still Gaussian) but not necessarily white.

Noise Before Demodulation

We write the analytic bandpass filter impulse response as:

$$\sqrt{2}\, f(t) e^{j\omega_c t} = f_a(t) + j\hat{f}_a(t)$$

It satisfies:

$$2\int_{-\infty}^{\infty} |f(t)|^2 dt = \int_{-\infty}^{\infty} f_a^2(t)dt + \int_{-\infty}^{\infty} \hat{f}_a^2(t)dt$$

$$\int_{-\infty}^{\infty} f_a^2(t)dt = \int_{-\infty}^{\infty} \hat{f}_a^2(t)dt$$

$$\Rightarrow \int_{-\infty}^{\infty} f_a^2(t)dt = \int_{-\infty}^{\infty} \hat{f}_a^2(t)dt = \int_{-\infty}^{\infty} |f(t)|^2 dt$$

The power (variance) of the complex bandpass noise can be calculated as:

$$E\left[M_R^2(t)\right] = E\left[M_I^2(t)\right] = \sigma^2 \qquad E\left[|M(t)|^2\right] = 2\sigma^2$$

$$\sigma^2 = N_0 \int_{-\infty}^{\infty} |f(t)|^2 dt = N_0 \int_{-\infty}^{\infty} |F(f)|^2 df$$

We can see that the real and imaginary parts of the noise have the same variance.

They are also uncorrelated, because

$$E[M_R(t)M_I(t)] = N_0 \int_{-\infty}^{\infty} f_a(t)\hat{f}_a(t)dt$$

$$= N_0 \int_{-\infty}^{\infty} |F_a(j\omega)|^2 (j\,\mathrm{sgn}\, f)df = 0$$

Noise after Demodulation

Demodulation doesn’t effect the noise statistics, so the same properties are satisfied:

$$E\left[\left\{\mathrm{Re}\left[Z(t)\right]\right\}^2\right] = E\left[\left\{\mathrm{Im}\left[Z(t)\right]\right\}^2\right] = \sigma^2$$

$$E\left[\mathrm{Re}\left[Z(t)\right]\mathrm{Im}\left[Z(t)\right]\right] = 0$$

The power spectrum of the noise is:

$$S_M(f) = 2N_0\left|F(f - f_c)\right|^2$$

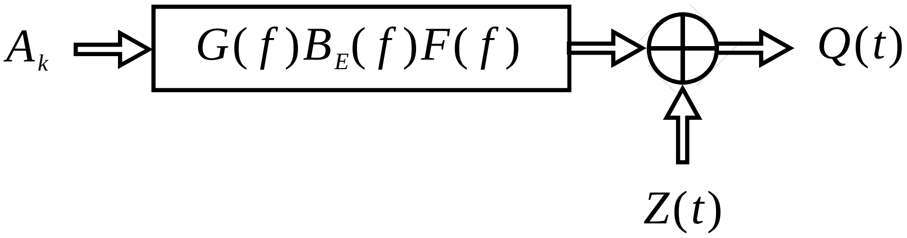
$$S_Z(f) = 2N_0\left|F(f)\right|^2$$

After sampling, at the slicer output, the discrete-time noise, $Z_k = Z(kT)$, is Gaussian and has independent real and imaginary parts whose variance is σ^2.

The noise power spectrum is the aliased version of the continuous-time one:

$$S_Z(e^{j2\pi fT}) = \frac{2N_0}{T}\sum_{m=-\infty}^{\infty}\left|F\left(f - \frac{m}{T}\right)\right|^2$$

Equivalent Baseband System Model



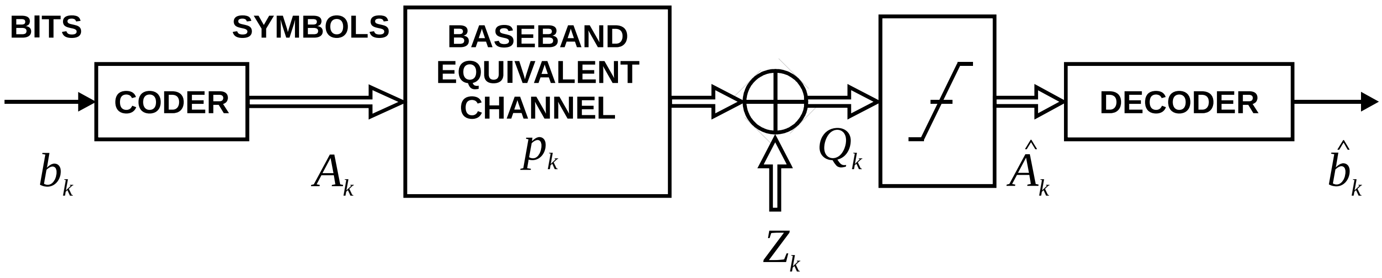
$$S_Z(f)=2N_0|F(f)|^2$$

Here

$G(f)$	Fourier transform of the transmitted pulse.
$F(f)$	Transfer function of the baseband receiver filter.
$B_E(f)=B(f+f_c)$	Channel transfer function around the carrier frequency translated to baseband.

The real and imaginary parts of the noise are Gaussian, independent, and they have the same variance.

Equivalent Discrete-Time Baseband System Model



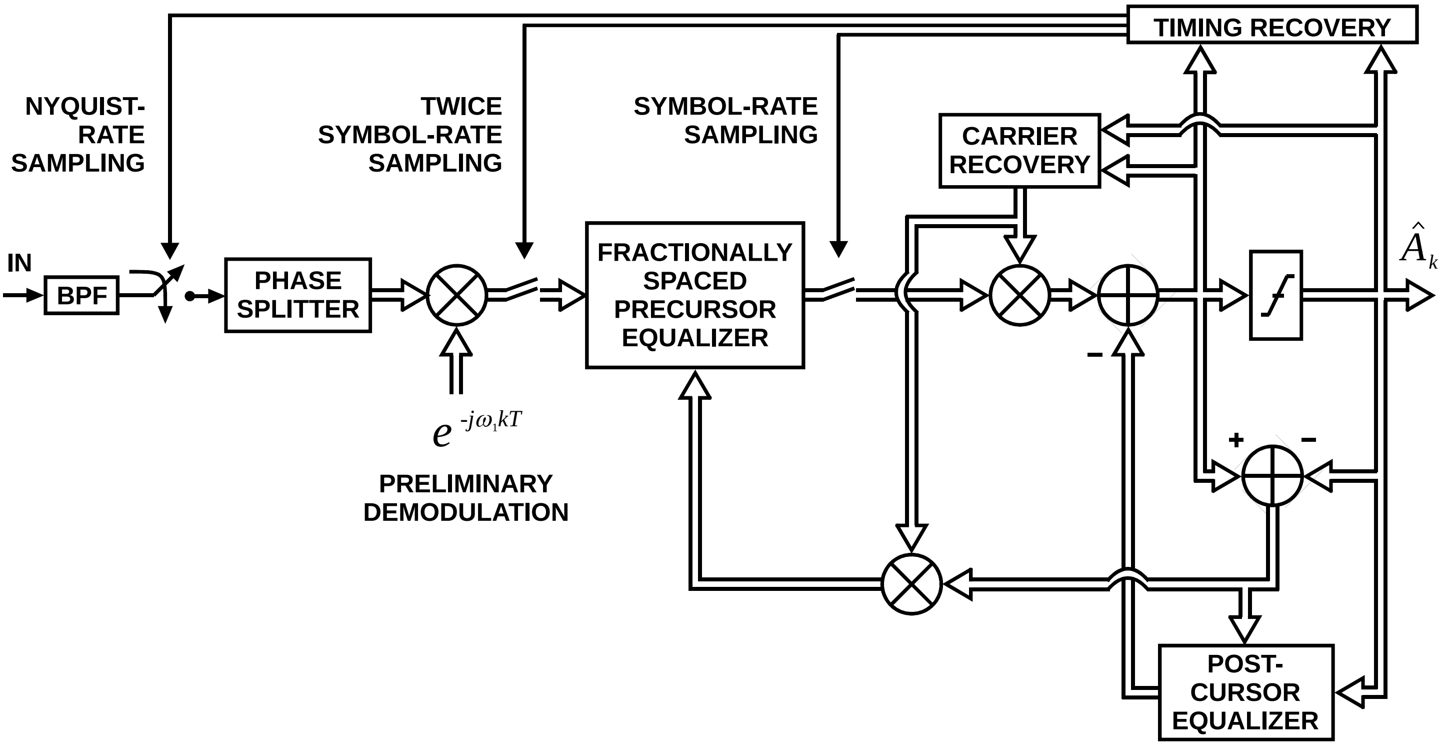
Here the transmit filter, channel, receive filter, and sampler are modelled with a single discrete-time filter, the impulse response of which is $p_k=p(kT)$, transfer function $P(z)$, and frequency response

$$P(e^{j2\pi fT}) = \frac{1}{T}\sum_{m=-\infty}^{\infty} G\left(f-\frac{m}{T}\right)B_E\left(f-\frac{m}{T}\right)F\left(f-\frac{m}{T}\right)$$

The discrete-time noise model $Z_k=Z(kT)$ has independent real and imaginary parts whose variance is σ^2. The noise power spectrum is

$$S_Z(e^{j2\pi fT}) = \frac{2N_0}{T}\sum_{m=-\infty}^{\infty} \left|F\left(f-\frac{m}{T}\right)\right|^2$$

A Complete I/Q Receiver



This is just one example of many alternatives. In addition to the previously discussed blocks, it includes:

- Adaptive channel equalizer (fractionally-spaced decision feedback equalizer, DFE)
- Carrier recovery (synchronization) loop.
- Timing recovery loop to adjust the sampling clock.

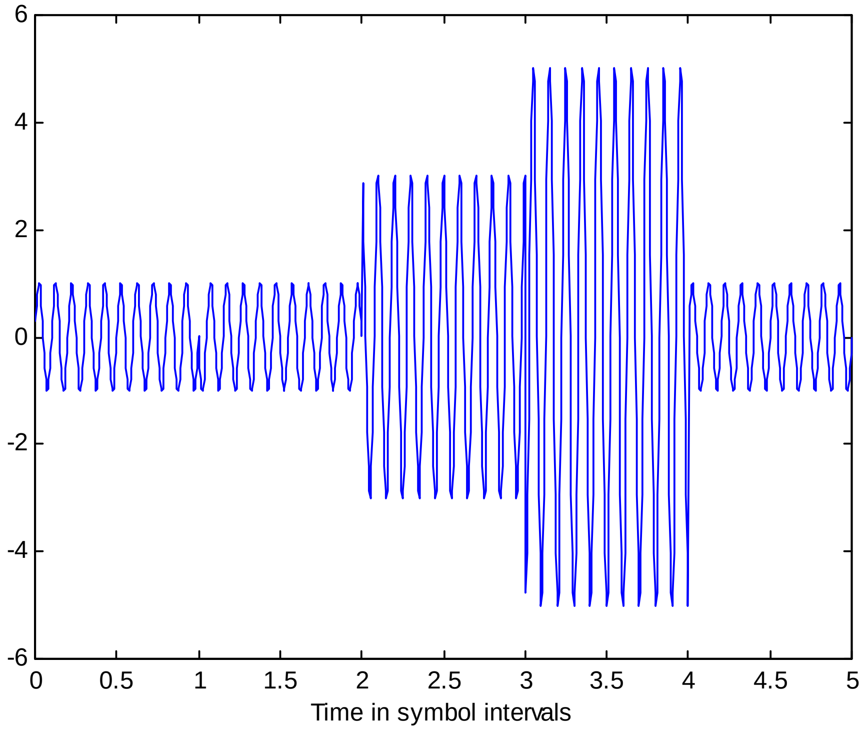
These topics will be discussed later in the course.

CONSTELLATIONS, NOISE, AND ERROR RATES

Symbol values are complex in I/Q modulation. They determine the amplitude and phase of the modulated carrier at the sampling instants. The transitions between adjacent symbols depend on the used pulse shaping.

Example: Square pulse shaping,

symbol sequence 1 -1 3i -5i 1



The complex symbol values (alphabet) can be represented by constellation:

- QAM: points are situated on a regular rectangular grid
- PSK: points are situated on a circle
- Also other constellations are available, but not in wide use.

Constellations Examples

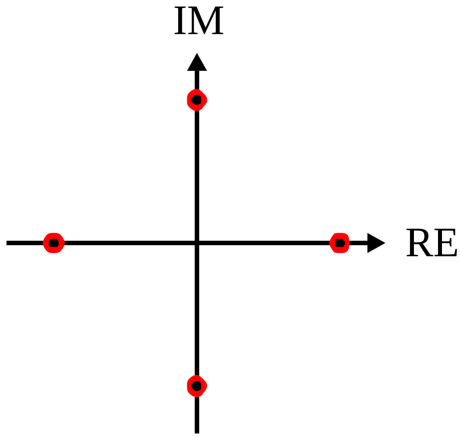
(a) 4-PSK (QPSK)

Alphabet size $2^B = 4$, two bits per symbol, B=2.

Symbols: $A_m = b\, e^{j\phi_m}$, $\quad \phi_m \in \{0, \pi/2, \pi, 3\pi/2\}$

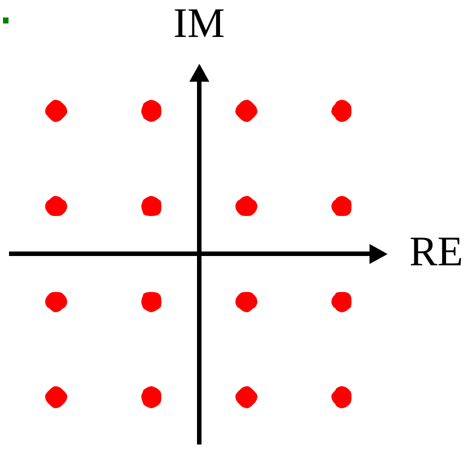
Transmitted signal: $x(t) = b\sqrt{2} \sum_{m=-\infty}^{\infty} \cos(\omega_c t + \phi_m) g(t - mT)$

Information is carried on the carrier phase.



(b) 16-QAM

Alphabet size $2^B = 16$, 4 bits per symbol B=4.
Rectangular constellation.



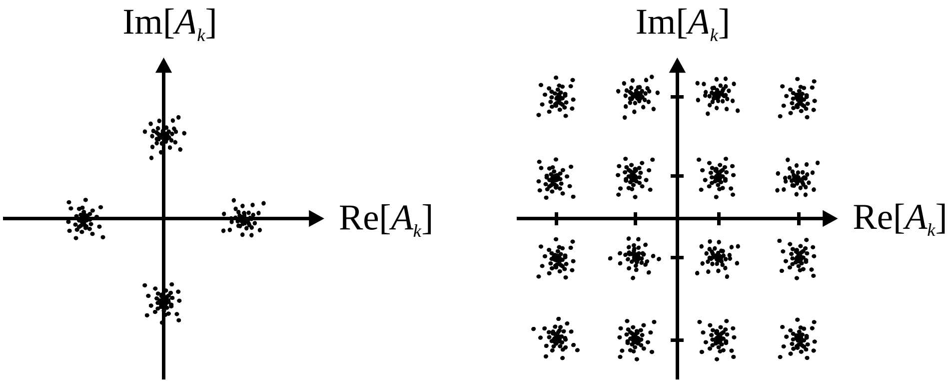
More generally:

M-PSK: $2^B = M$ constellation points on circle at regular intervals.

M-QAM: $2^B = M$ points in regular rectangular shape.

Noisy Constellation

Because of ISI, noise, and other type of distortions, the received samples will not correspond exactly to points in the signal constellation. E.g., in the cases of 4-PSK and 16-QAM, the following constellations could be formed:

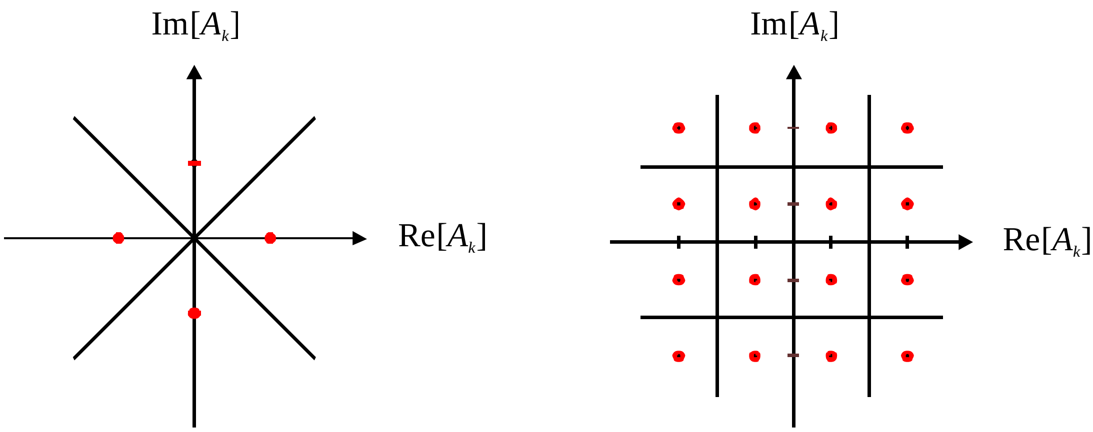


In fact, in case of Gaussian noise as the only distortion, the results of p. 77 show that the received samples will form a Gaussian cloud around the points in the constellation.

Next we consider making the decisions. The problem is to decide which symbol has been transmitted, based on the help of received complex sample value Q_k.

Intuitively, the slicer selects the $\hat{A}_k$ in the alphabet that minimizes the distance $\left|Q_k - \hat{A}_k\right|$. This is the optimal choice in certain conditions (in ISI-free case), as will be discussed later in detail.

The complex plane can therefore be divided into decision regions where each decision region is the set of points that is closest to some symbol. Here are the decision regions for 4-PSK and 16-QAM:



Gaussian Distributed Random Variables

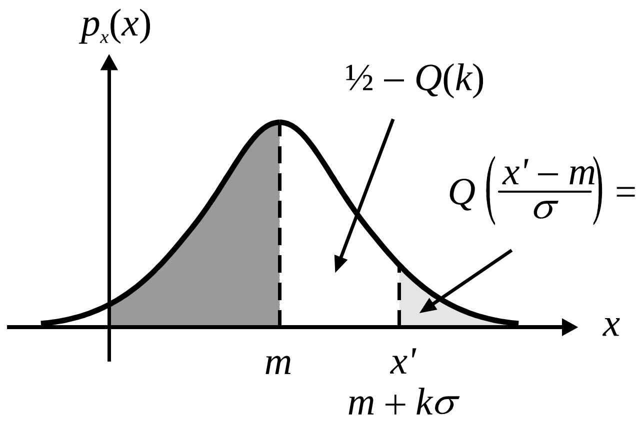
The function $Q(k)$ is commonly used when analyzing the bit or symbol error rate performance of digital modulation methods. It is defined for a normalized Gaussian distribution with variance $\sigma^2 = 1$ and mean $m = 0$ as follows:

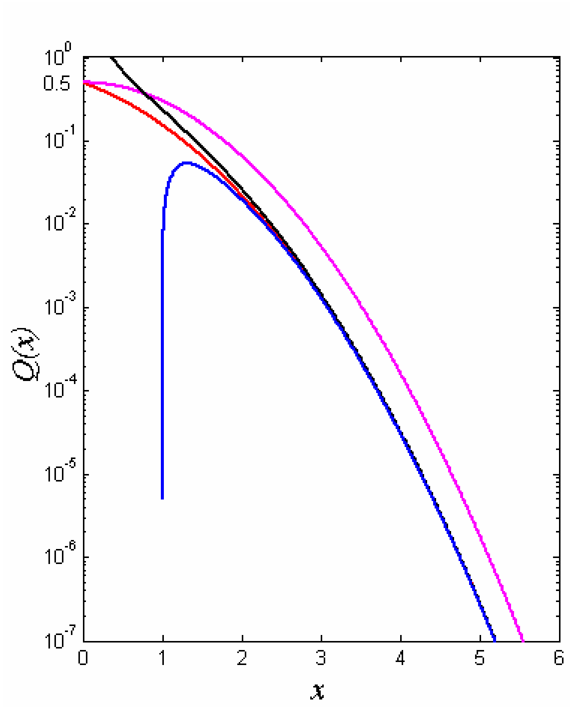
$$Q(k) = \Pr\left(X > k\right) = \int_k^{\infty} p_X(x)dx = \frac{1}{\sqrt{2\pi}} \int_k^{\infty} e^{-x^2/2} dx$$

The values of this function have been tabulated, the Matlab function ERFC can be used for calculating it, or one of the approximations shown below can be used for evaluating it for high values of k (k>3), that are usually the interesting cases.

For arbitrary mean m and variance σ^2, a change of variables gives

$$\Pr(X > x') = Q\left(\frac{x'-m}{\sigma}\right)$$





$$\frac{1}{2} e^{-x^2/2}$$

$$\frac{1}{\sqrt{2\pi}x} e^{-x^2/2}$$

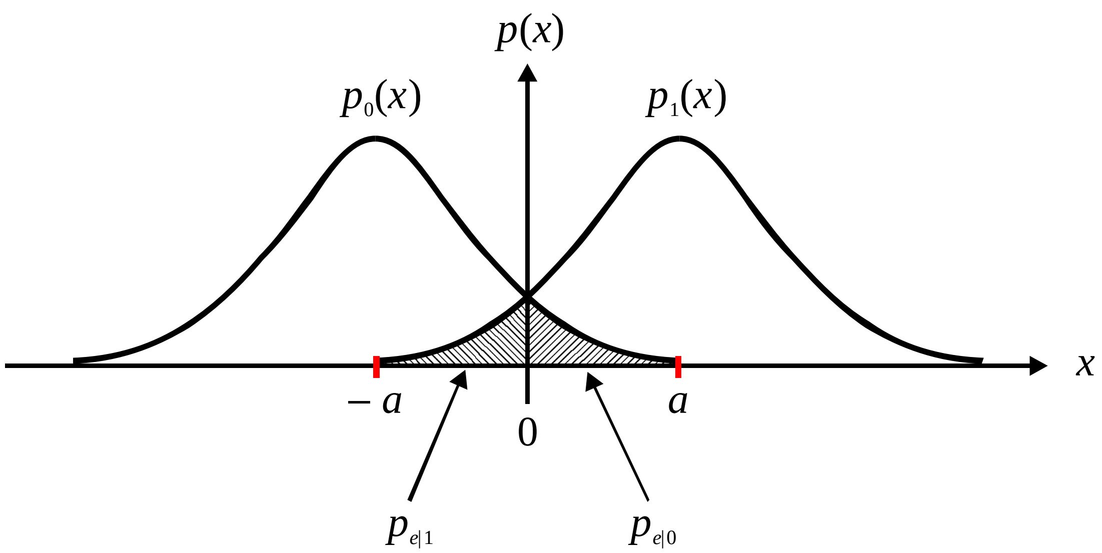
$$Q(x)$$

$$\frac{1}{\sqrt{2\pi}x}\left(1 - \frac{1}{x^2}\right) e^{-x^2/2}$$

Symbol Error Probability in Case of Binary PSK

In case of binary signaling (M=2) the alphabet is usually real. We consider the case of 2-PSK, or binary antipodal signaling, that uses the symbol values $-a$ and a.

With the earlier assumptions, the probablity distribution has the Gaussian shape around each of the symbol values:



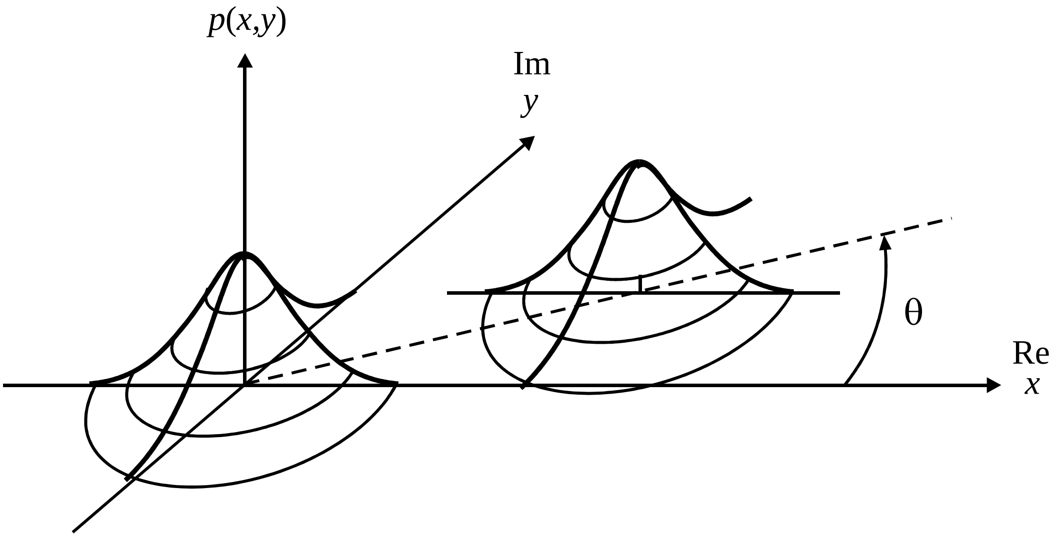
Naturally, the decision level is chosen to be 0 (more about this later on).

The error probability becomes:

$$\begin{aligned}\Pr(symbol\ error) &= \Pr(symbol\ error\ when\ a\ transmitted)\\ &= \Pr(symbol\ error\ when\ -a\ transmitted)\\ &= Q\left(\frac{a}{\sigma}\right)\end{aligned}$$

Symbol Error Probability in Case of Binary Modulations in General

More generally, for *M=2,* but with arbitrary complex valued symbols a_i and a_j, the situation looks like this (in this example one of the symbols is 0):



We assume that the additive noise is Gaussian with equal and independent real and imaginary parts, each with variance σ^2. Then the probability distribution along the line joining the two symbols is similar to the previous case of 2-PSK. We denote the Euclidean distance of the symbols by $d=\left|a_i - a_j\right|$. The error probability becomes

$$\begin{aligned}\Pr(symbol\ error) &= \Pr(symbol\ error\ when\ a_1\ transmitted)\\ &= \Pr(symbol\ error\ when\ a_2\ transmitted)\\ &= Q\left(\frac{d}{2\sigma}\right)\end{aligned}$$

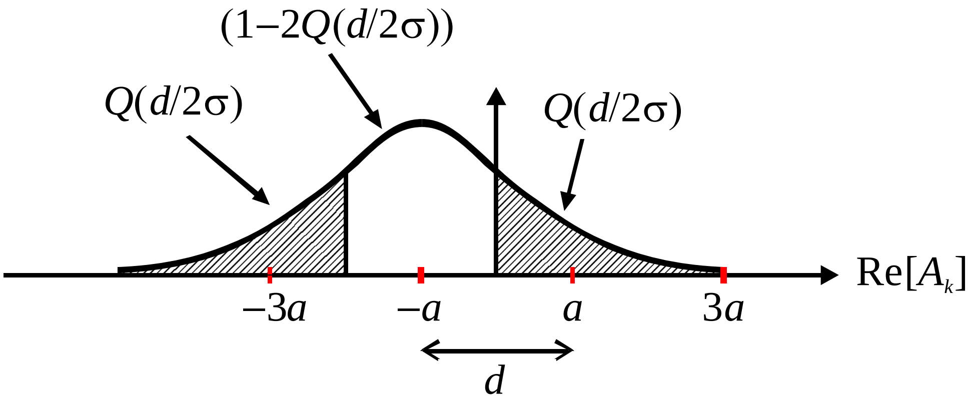
This is a fundamental result that will be applied in the following to calculate the symbol error probabilities also for other constellations, based on partial error probabilities of the form:

$$\Pr(slicer\ \ prefers\ a_i\ \ when\ a_j\ \ transmitted) = Q\left(\frac{d}{2\sigma}\right)$$

Example: 4-PAM

The symbol values are: $\pm a$ and $\pm 3a$.

When, for example, A_k=-a is transmitted, the distribution of the received sample, Q_k, is as follows:



The error probabilities are obtained as:

$$\Pr(symbol\ error\ at\ time\ k \mid A_k = \pm a) = 2Q\left(\frac{d}{2\sigma}\right)$$

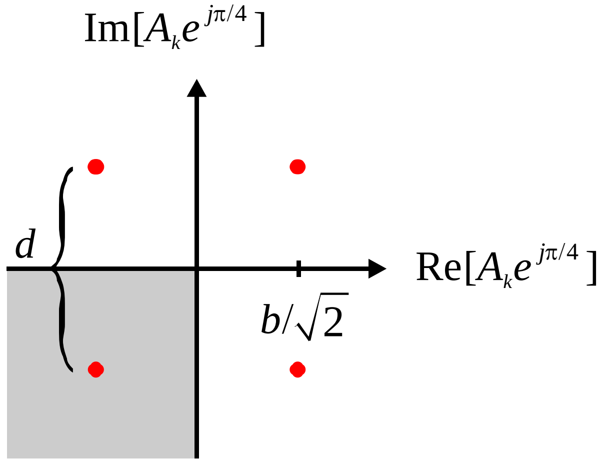
$$\Pr(symbol\ error\ at\ time\ k \mid A_k = \pm 3a) = Q\left(\frac{d}{2\sigma}\right)$$

If the four symbols are equally probable, then:

$$\boxed{\Pr(symbol\ error) = 1.5Q\left(\frac{d}{2\sigma}\right)}$$

Symbol Error Probability of QPSK

Assuming optimum decision regions, rotation of the constellation doesn't affect the error probabilities in any way. Therefore, we can consider the following QPSK constellation rotated by 45°:



The noise samples are now also rotated, $M_k = Z_k e^{j\pi/4}$. With the earlier assumptions, the noise statistics are not affected, so the real and imaginary parts of the noise are still independent and they have variance σ^2.

In case of $A_k = -b$, the correct decision corresponds to the shaded region of the above figure, i.e.,

$$\begin{aligned}
&\Pr(\textit{correct decision at time } k \mid A_k = -b) \\
&= \Pr\left(\mathrm{Re}[M_k] < d/2,\ \mathrm{Im}[M_k] < d/2\right) \\
&= \Pr\left(\mathrm{Re}[M_k] < d/2\right) \cdot \Pr\left(\mathrm{Im}[M_k] < d/2\right) \\
&= \left(1 - Q\left(\frac{d}{2\sigma}\right)\right)^2 \\
&= 1 - 2Q\left(\frac{d}{2\sigma}\right) + Q^2\left(\frac{d}{2\sigma}\right)
\end{aligned}$$

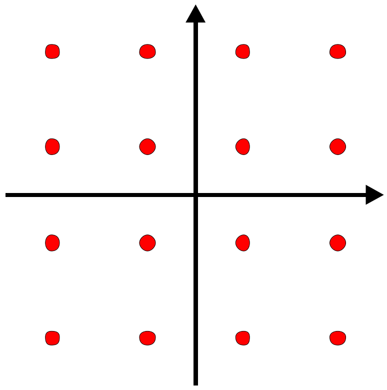
Symbol Error Probability of QPSK *(continued)*

All the symbols have the same error probabilities, so

$$
\begin{aligned}
\Pr(\textit{symbol error at time } k) &= \Pr(\textit{symbol error at time } k \mid A_k = -b) \\
&= 1 - \Pr(\textit{correct decision at time } k \mid A_k = -b) \\
&= 2Q\left(\frac{d}{2\sigma}\right) - Q^2\left(\frac{d}{2\sigma}\right) \\
&\cong 2Q\left(\frac{d}{2\sigma}\right)
\end{aligned}
$$

The latter form applies when $Q\left(\dfrac{d}{2\sigma}\right)$ is small. In practical cases, it can be assumed to be less than 0.1, so the approximation error is no more than 5 %.

Symbol Error Probability of 16QAM



There are three different types of points in the constellation.

(a) In the four inside points:

$$\Pr(\mathit{correct\ decision} | A_k \mathit{\ on\ the\ inside}) = \left(1 - 2Q\left(\frac{d}{2\sigma}\right)\right)^2$$

$$\Pr(\mathit{symbol\ error} | A_k \mathit{\ on\ the\ inside}) = 4Q\left(\frac{d}{2\sigma}\right) - 4Q^2\left(\frac{d}{2\sigma}\right)$$

$$\cong 4Q\left(\frac{d}{2\sigma}\right)$$

(b) In the four corner points we have:

$$\Pr(\mathit{symbol\ error} | A_k \mathit{\ in\ the\ corner}) = 2Q\left(\frac{d}{2\sigma}\right) - 2Q^2\left(\frac{d}{2\sigma}\right)$$

$$\cong 2Q\left(\frac{d}{2\sigma}\right)$$

(c) In the eight other points:

$$\Pr(\mathit{symbol\ error} | A_k \mathit{\ not\ inside\ or\ in\ the\ corner}) \cong 3Q\left(\frac{d}{2\sigma}\right)$$

Symbol Error Probability of 16QAM *(continued)*

Assuming equal probabilities of the symbols and weighting each of the above error probabilities by the number of cases, we obtain for the average symbol error probability:

$$\boxed{\Pr(symbol\ error) \cong 3Q\left(\frac{d}{2\sigma}\right)}$$

From the above cases we can notice that, generally, the coefficient of $Q(\cdot)$ is equal to the number of nearest neighbours of the corresponding constellation point.

As a consequence, when the size of the QAM constellation is increased, the error probability approaches

$$\Pr(symbol\ error) \to 4Q\left(\frac{d}{2\sigma}\right) - 4Q^2\left(\frac{d}{2\sigma}\right)$$

assuming that all the symbols are equally probable.

We have seen that the symbol error probability depends essentially on the ratio of the minimum distance between constellation points, d, and standard deviation of the noise.

Union Bound

Above we have derived exact symbol error probability expressions, and some simplified approximative expressions, for certain constellations. However, this analytic approach is not feasible for all constellations, and one has to be satisfied with more crude approximations. One generic idea for deriving such approximations / limits is the union bound principle.

Let us consider linear digital modulation with alphabet

$$\Omega_A = \{a_i ; 0 < i \le M\}.$$

Let us denote by E_{ij} the case that the received sample is closer to a_j than to the transmitted a_i. We can write

$$\begin{aligned} &\Pr(symbol\ error | A_k = a_i) \\ &= \Pr\left(E_{i1} \cup E_{i2} \cup \cdots \cup E_{i,i-1} \cup E_{i,i+1} \cup \cdots \cup E_{iM}\right) \end{aligned}$$

The union bound gives an upper limit to the error probability:

$$\Pr(symbol\ error | A_k = a_i) \le \sum_{\substack{j=1 \\ j \ne i}}^{M} \Pr\left(E_{ij}\right)$$

For example, in the QPSK case, the union bound gives the upper limit:

$$\Pr(symbol\ error | A_k) \le 2Q\left(\frac{d}{2\sigma}\right) + Q\left(\frac{d}{\sqrt{2}\sigma}\right)$$

which is a good approximation to the earlier exact expression with reasonable noise levels.

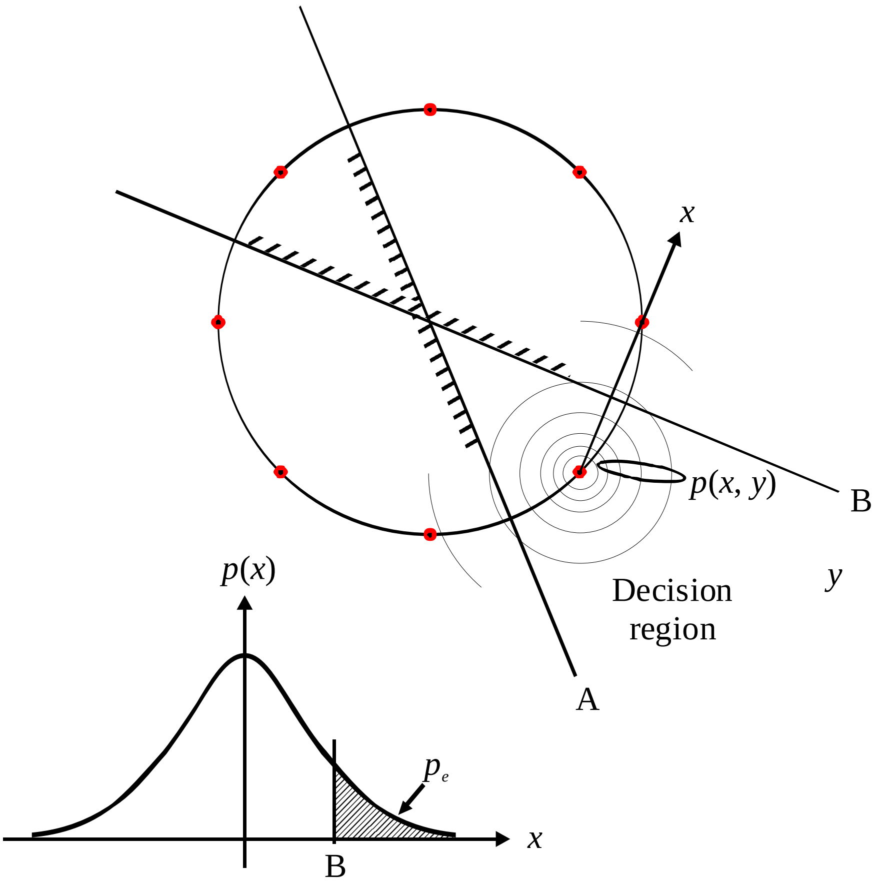
Union Bound *(continued)*

In practice, it is enough to take the closest neighbours from the constellation to the union bound –summation.

For example, the union bound can be used for obtaining an approximation for the error probability of M-PSK with any M.

In the example case below, the two error cases corresponding to the closest neighbours are indicated. The union bound –related approximation error corresponds to ignoring the overlapping of the two types of error cases, i.e., dublicating some error events in the probability calculation. However, their effect on the overall result is very small.

(Please note: Microsoft wants this line to be here above, I cannot help!)



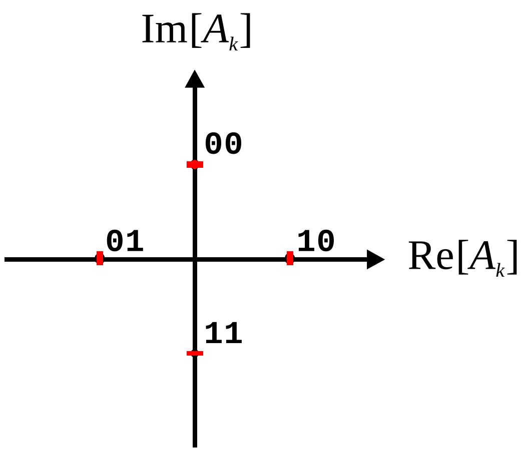
Bit Error Probability and Gray Codes

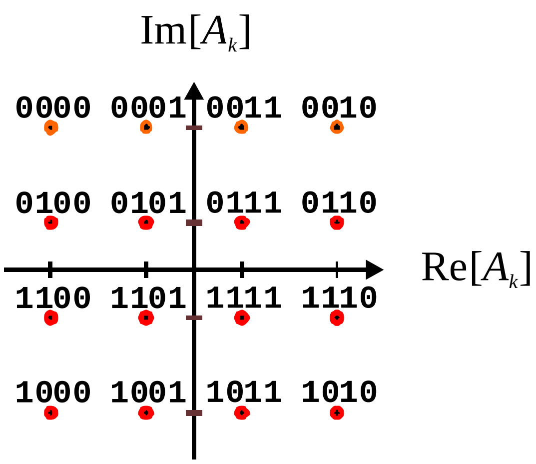
The quality of digital transmission systems is usually measured in terms of bit error probability, or bit error rate (BER), instead of symbol error rate (SER).

When the SER is known, the BER depends essentially on the used bit mapping.

In most cases, the SNR is reasonably good, so that symbol errors take place mostly between neighbouring constellation points. In that case the BER can be minimized by using a symbol mapping where the codes of the neighbouring constellation points differ from each other with the minimum number of bits.

In the optimum case, all the pairs of neighbouring points differ only by one bit. Such a symbol coding is known as Gray code. Some examples:





In this case, with alphabet size 2^B and with resonably good SNR, BER can be evaluated as:

$$\Pr(bit\ error) = \frac{1}{B}\Pr(symbol\ error)$$

About Bit Mappings

When considering the mapping of a bit-stream into symbols, several things have to be taken into account:

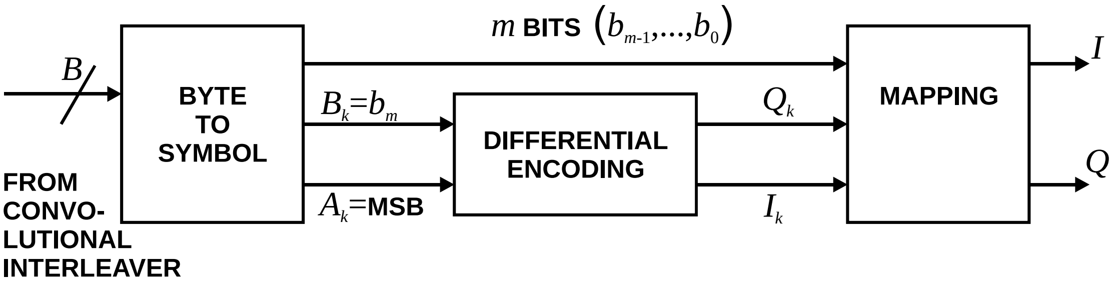
1. Gray-coding minimizes the BER for a given SER, i.e., for a given S/N-ratio in an AWGN channel.

   Gray coding can be realized completely for QPSK, 16-QAM, 64-QAM, etc., but, e.g., for 32 QAM complete Gray code cannot be found (i.e, some neighbouring symbols will differ in more than one bit)

2. Using a system shown on the following page, it is possible to achieve a situation where the system is invariant to carrier phase shifts which are integer multiples of 90°. In this example (taken from European digital TV transmission systems, DVB), $\boldsymbol{m}$ most significat bits are mapped on the symbols in such way that the constellation is invariant to n*90° degree rotations. The two least significant bits are coded differentially.

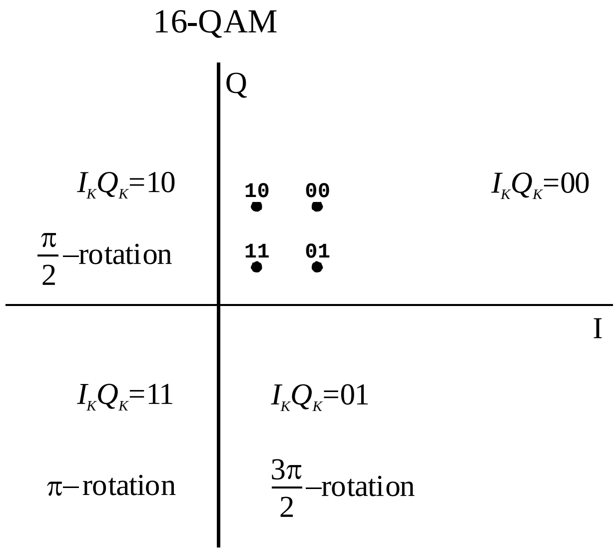
   This principle is important, because it is difficult to make the carrier recovery to find the right multiple of 90 degrees.

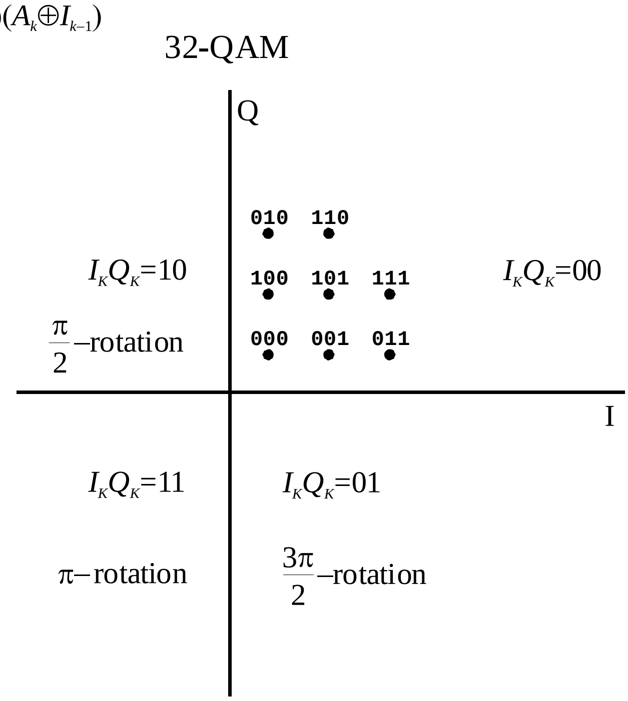
Bit-Mappings which are Invariant under n*90° Phase Shifts

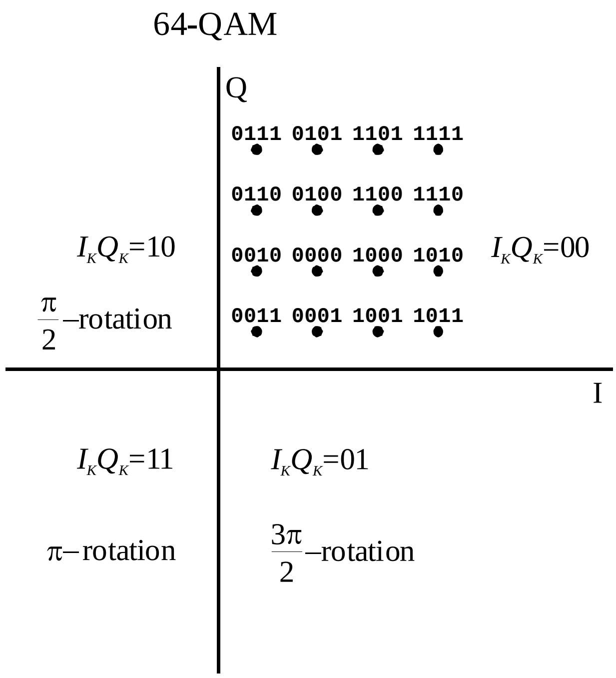


m = 2 FOR 16-QAM, 3 FOR 32-QAM, 4 FOR 64-QAM

$$I_k = (\overline{A_k \oplus B_k})(A_k \oplus I_{k-1}) + (A_k \oplus B_k)(B_k \oplus Q_{k-1})$$
$$Q_k = (\overline{A_k \oplus B_k})(B_k \oplus Q_{k-1}) + (A_k \oplus B_k)(A_k \oplus I_{k-1})$$







Block Error Probability

In many cases the data is transmitted as blocks or packets. The system may include error detection (e.g., Cyclic Redundancy Check, CRC) and retransmission schemes. In such systems, block error probability, BLER, is an important performance measure.

If the block includes L bits, then it corresponds to L/B symbols, so the BLER becomes

$$\begin{aligned}\Pr(block\ error) &= 1-(1-\Pr(symbol\ error))^{L/B} \\ &\approx \frac{L}{B}\Pr(symbol\ error)\end{aligned}$$

The latter form is valid when the SER is low.

Example

In a typical case

$$L{=}1000,\ B{=}4,\ \Pr(symbol\ error) = 10^{-6}.$$

and

$$\Pr(block\ error) = 2.5\times 10^{-4}.$$

Transmission Power

In principle, the BER/SER could be improved by just increasing SNR. However, in practise, the transmission power is usually limited by system standards or technical reasons, or the electrical power consumption should be minimized.

For these reasons, to get a meaningful performance measure, the SER/BER performance should be related to received SNR.

Transmission Power

Assumptions:

- the symbol sequence is white: $S_A(e^{j2\pi fT}) = \beta$
- the symbol sequence has zero mean
- the transmit filter is scaled such that $\int_{-\infty}^{\infty} |G(f)|^2 df = T$

Then the power of the symbol sequence is $E\left[|A_k|^2\right] = \beta$

The power spectrum of the complex baseband signal is

$$S_s(f) = \frac{1}{T}|G(f)|^2 S_A(e^{j2\pi fT}) = \frac{E\left[|A_k|^2\right]}{T}|G(f)|^2$$

and its power is

$$P_s = \int_{-\infty}^{\infty} S_s(f)df = E\left[|A_k|^2\right].$$

The transmitted passband signal is

$$x(t) = \sqrt{2}\,\mathrm{Re}\left[e^{j\omega_c t}s(t)\right]$$

and its power is equal to that of the baseband signal,

$$P_x = E\left[|A_k|^2\right].$$

So, when the transmission power is limited, there is an upper limit for the minimum distance, d, for each constellation. This, in turn, determines the SER according to the earlier discussions.

Symbol Error Rates for Different Constellations

With suitable scaling, the SNR can be defined as

$$\mathrm{SNR} = \frac{E\left[\left|A_k\right|^2\right]}{2\sigma^2}.$$

So the signal power is equal to the power of the signal sequence and the noise power is equal to the sum of the real and imaginary noise powers.

The following slide shows the signal powers and SER vs. SNR expressions for the earlier analyzed constellations.

Three different formulas are given for the error probabilities: *exact*, *good approximation*, and *crude approximation*.

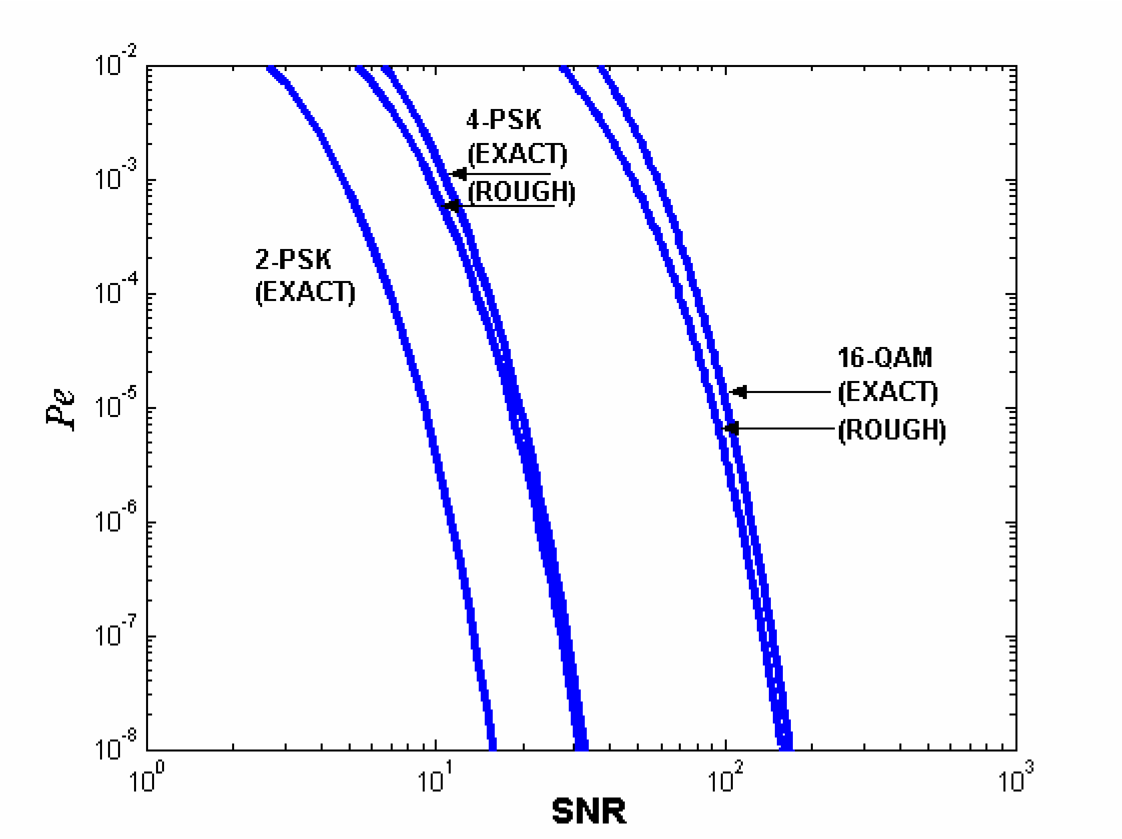
- The good approximation ignores the higher order terms, and the approximation errors are insignificant.
- Crude approximation ignores also the coefficient of the Q-function (i.e., the number of closest neighbours), and the approximation errors are no more than 1 dB.

In general, it can be observed that,

- Larger constellations need higher signal power to achive the same SER as smaller constellations.

Symbol Error Rates for Different Constellations

constellation	2-PSK	4-PSK	16-QAM
alphabet	$\{\pm a\}$	$\{\pm b, \pm jb\}$	$\{(\pm c \text{ or } \pm 3c) + j(\pm c \text{ or } \pm 3c)\}$
minimum dist. d	$2a$	$b\sqrt{2}$	$2c$
power $E[\lvert A_k\rvert^2]$	a^2	b^2	$10c^2$
exact P_e	$Q\left(\sqrt{2 \cdot SNR}\right)$	$2Q\left(\sqrt{SNR}\right) - Q^2\left(\sqrt{SNR}\right)$	$3Q\left(\sqrt{SNR/5}\right) - 2.25Q^2\left(\sqrt{SNR/5}\right)$
mild approx. P_e	$Q\left(\sqrt{2 \cdot SNR}\right)$	$2Q\left(\sqrt{SNR}\right)$	$3Q\left(\sqrt{SNR/5}\right)$
rough approx. P_e	$Q\left(\sqrt{2 \cdot SNR}\right)$	$Q\left(\sqrt{SNR}\right)$	$Q\left(\sqrt{SNR/5}\right)$



Energy per Bit

Instead of SNR, the following 'normalized' performance measure is commonly used in the literature:

$$\boxed{\frac{E_b}{N_0}}$$

Here E_b is the received signal energy per bit and $N_0/2$ is the two-sided noise power spectral density.

The usual assumption that the channel noise is additive white Gaussian applies. When the receiver filter is (or is modeled as) an ideal lowpass with bandwidth W, the noise power is

$$N = WN_0.$$

(Here the noise power of $WN_0/2$ is contributed both from positive and negative frequency portions of the spectrum.)

The signal power depends on the bit energy and bit rate R:

$$S = RE_b$$

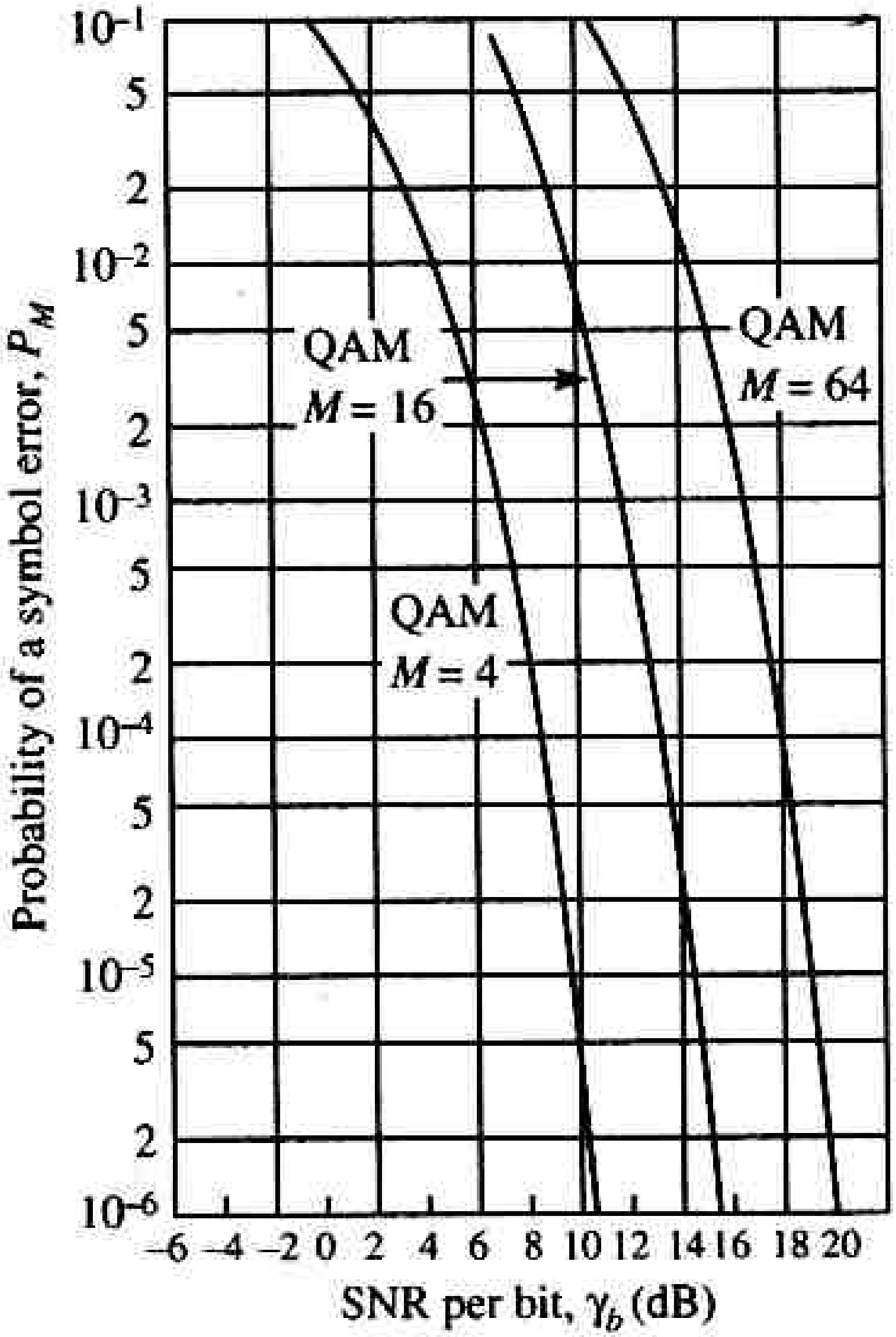
We can see that the SNR and the E_b / N_0 -ratio are related through

$$\boxed{\frac{S}{N} = \frac{E_b}{N_0}\frac{R}{W}}$$

The E_b / N_0 -ratio takes into account the bit rates of different constellations. In a sense, SER/BER figure for a given E_b / N_0 -ratio is a more fair perfromance measure than for given SNR. In addition, this ratio is independent of the implementation details, like receiver filter design which has a considerable effect on the SNR at the detector.

Energy per Bit *(continued)*

The following figure shows the SER performance of some constellations as a function of E_b/N_0-ratio (essentially, even though the x-axis is labelled differently) (from Proakis):



Optimizing the Constellations

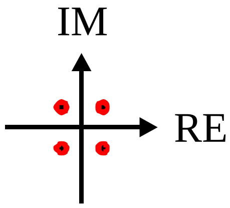
When optimizing the constellations, the goal is to maximize the minimum distance for a given power limit (either peak power or average power).

One criterion is that the mean value (DC) of the symbol set should be zero. The DC component corresponds to a translation of the constellation with respect to the origin, and zero DC corresponds to minimum power for symmetric constellations.

It is not easy to find such fully optimized constellations, but systematic methods have been developed. However, the improvements have been some fractions of dBs when compared to the simple basic constellations, and the implementation would be much more difficult.

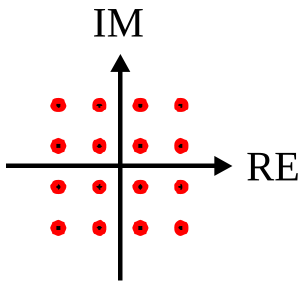
Basic Constellations

QAM, even *B*



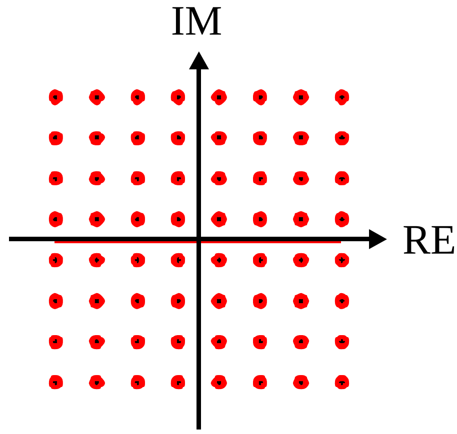
$B=2$

$M=2^2=4$



$B=4$

$M=2^4=16$



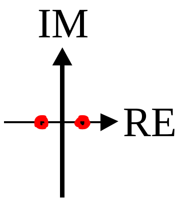
$B=6$

$M=2^6=64$

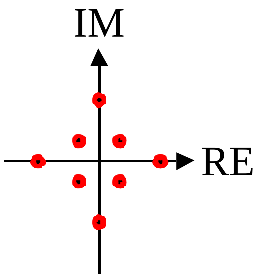
The I and Q components are independent $\sqrt{M}=2^{B/2}$ level PAM signals.

- The coder and detector/decoder are easy to implement using two independent branches and slicers.

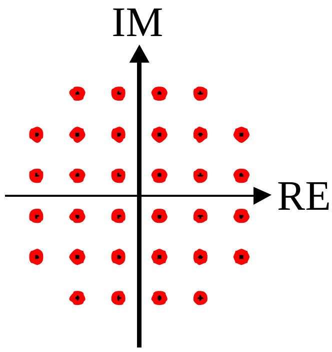
QAM, odd *B*



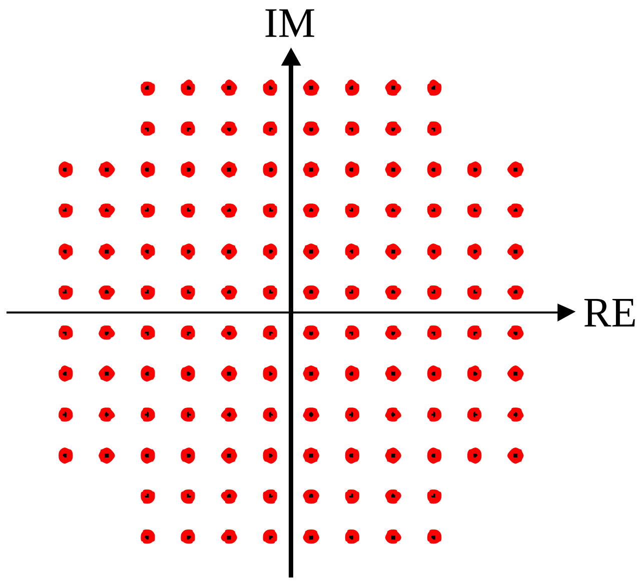
$M=2$



$M=8$



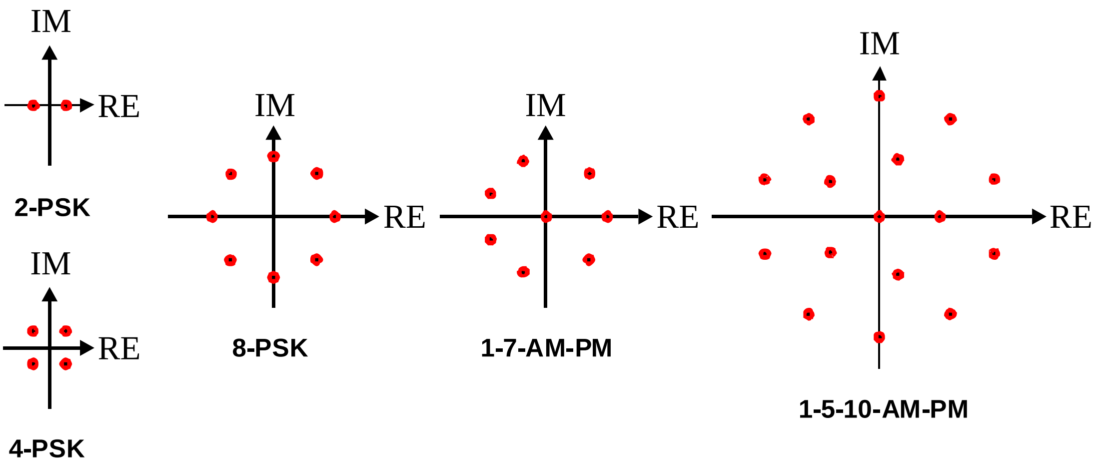
$M=32$



$M=128$

More Constellations

PSK Type

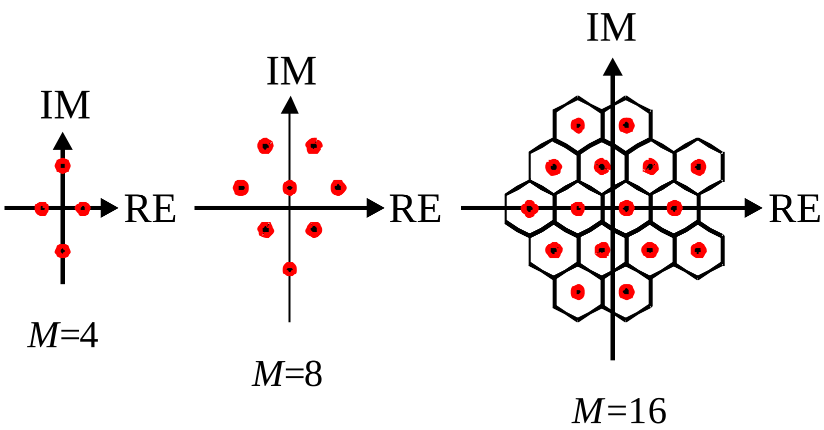


In the basic PSK, the points are equally spaced on a circle. In this case, there is no information in the amplitude of the modulated signal, so the amplitude variation is relatively small. (However, linear PSK is not a constant enevelope modulation, except for one very special case, that of MSK.) Such modulations are easier to implement from the transmitter power amplifier point of view than, e.g., QAM modulations.

In certain variants (that are not in common use), the constellation points are on a number of circles with different radii.

Hexagonal Constellations

Here the points are packed more densely than in the rectangular grid of QAM. However, the improvement in performance is relatively small (typically fractions of dB) and the implementation becomes considerably more difficult because the basic slicer principle cannot be used.



Special Linear Digital Modulation Methods
1. Offset QAM, Offset PSK

In offset modulations (or staggered modulations) the idea is to reduce the envelope variation, in order to releave the transmitter power amplifier linearity requirements, by having a half symbol interval offset between the transmitted I and Q branch signals.

Also, there is a half interval offset in the receiver when sampling the I and Q branch signals.

Since the I and Q signals are transmitted independently of each other, this doesn’t have any effects on the system SER/BER performance with average signal power constraints. However, the peak signal power of the transmitter is reduced considerably.

One way to illustrate this is to notice that the amplitude of usual QAM or PSK signals may have zero crossings between symbols, but in offset modulations this doesn’t happen.

Special Linear Digital Modulation Methods
2. CAP

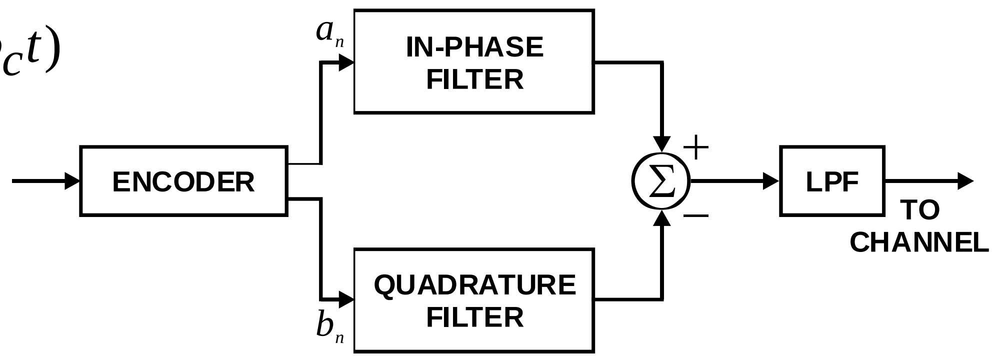
CAP (Carrierless amplitude/phase modulation) is a special case of QAM where

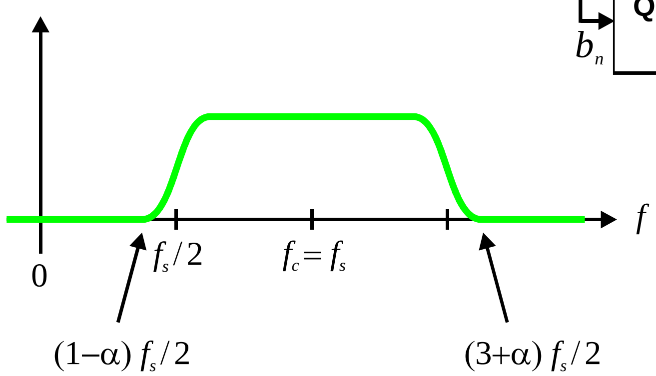
carrier frequency = *r**symbol rate

and the Nyquist pulse-shaping filtering and carrier-modulation are combined. The ratio, r, of carrier frequency and symbol rate is fixed and usually an integer or simple fraction. The case $r=1$ is a good an practical example. Effectively, in CAP, carrier modulation is implemented by using bandpass pulse shaping filters instead of lowpass filters:

$$g_I(t) = g(t)\cos(\omega_c t)$$

$$g_Q(t) = g(t)\sin(\omega_c t)$$





Here $g(t)$ is the used lowpass pulse shape, like root raised cosine impulse response.

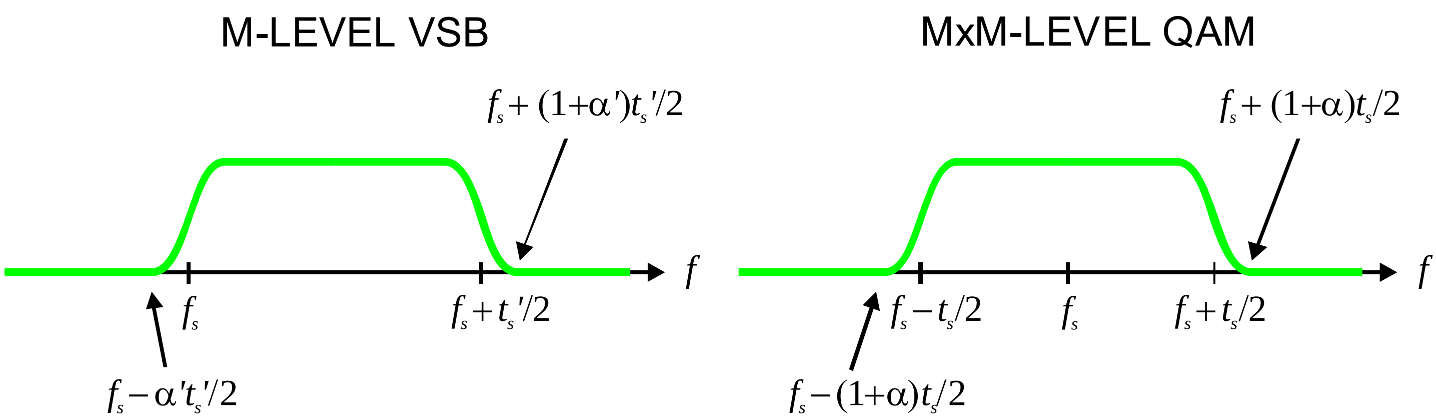
Compared to QAM, the implementation of transmitter and receiver signal processing is simplified, because no carrier local oscillator or mixer is needed and the receiver synchronization is simplified.

CAP is mostly considered for wireline applications where the transmitted signal is close to baseband.

Special Linear Digital Modulation Methods
3. VSB

VSB has also gained some popularity as digital modulation method, especially, in the context of US terrestrial TV transmission system.

VSB signal is obtained from real PAM signal by filtering out the other sideband according to the vestigial sideband principle.



In case of QAM, both transition bands are shaped due to Nyquist pulse shaping

In case of VSB, one transition band is shaped by Nyquist pulse shaping, and the other is shaped due to the VSB principle. However, this leads essentially to the same result. With suitable choice of the excess bandwidth parameters, the two systems have the same spectral efficiency.

VSB is claimed to have some benefits with respect to synchronization. In addition, VSB has a stong heritage from analog TV.

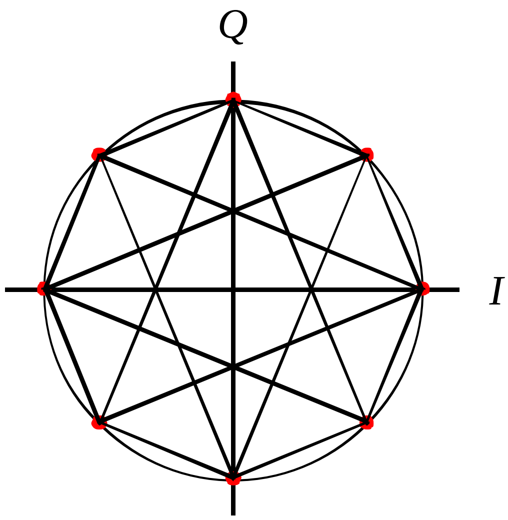
Special Linear Digital Modulation Methods
4. π/4 QPSK

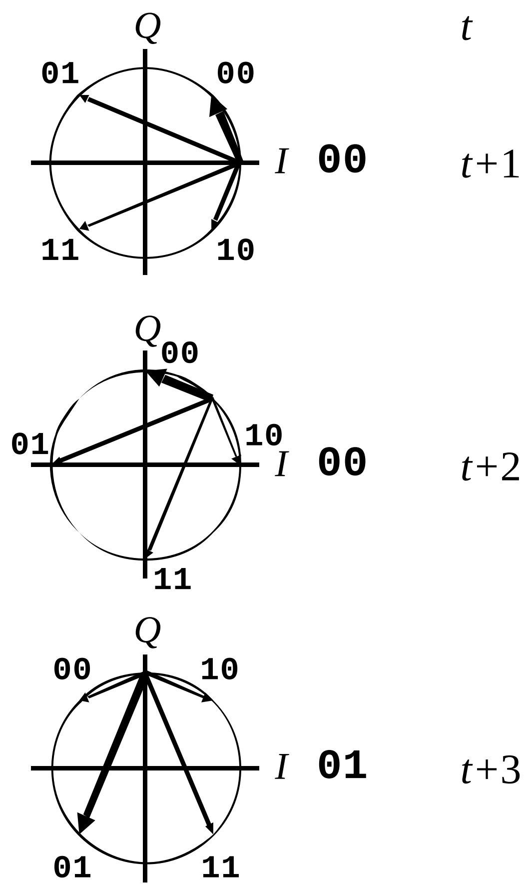
π/4 QPSK modulation uses 8 constellation points (those of 8PSK) for transmitting 2 bits per symbol. During each symbol interval, the constellation is a rotated version of QPSK, and the constellation is rotated by 45° between consequtive symbols. So the phase change between symbols is $\pm 45°$ or $\pm 135°$.

In this way the zero crossings of envelope are avoided and the peak-to-average power ratio is reduced.

The SER/BER performance of π/4 QPSK is the same as for QPSK.

π/4 QPSK is a commonly used modulation method in digital mobile communication systems (DAMPS, etc.)





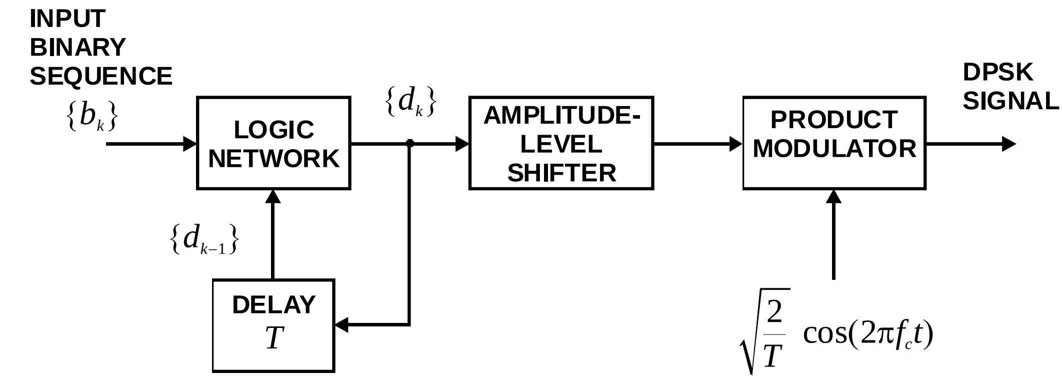
Special Linear Digital Modulation Methods
4. Differential PSK

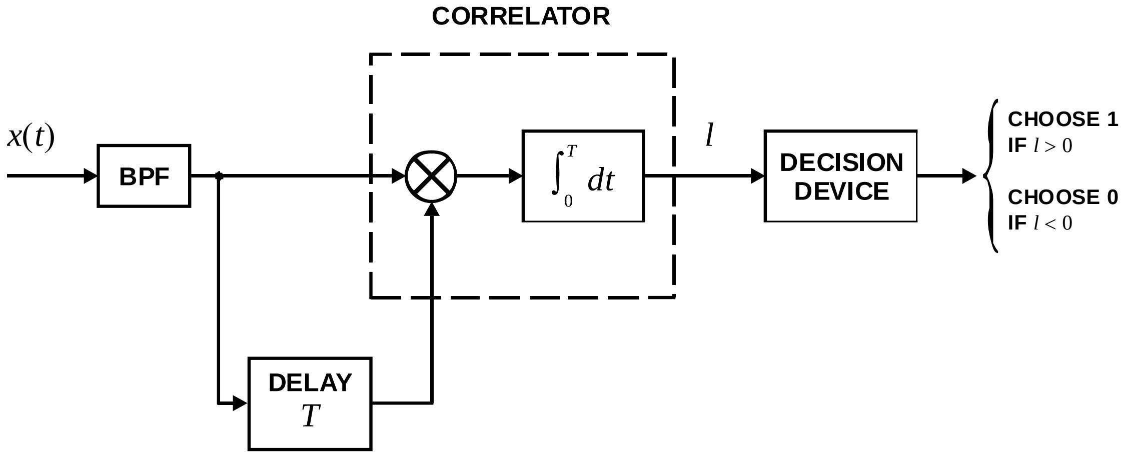
This is another variant of PSK modulation methods. The idea is that transmitted symbol information is carried in the phase change, not in the absolute phase.

For example, in the binary case, the mapping rule could be:

bit 0 → phase remains the same
bit 1 → phase changed by 180º

Block diagrams for DPSK transmitter and receiver:





Special Linear Digital Modulation Methods
4. Differential PSK *(continued)*

The receiver utilizes differentially coherent detection. This is based on the correlation between two consecutive symbol intervals, which can be used for determing the possible phase change between the symbols.

The main benefit of DPSK is simple receiver signal processing, since no absolute carrier phase reference is needed. This means that the DPSK performs rather well also in fast fading channels. However, channel equalization methods cannot be easily combined with the DQPSK idea, so it is mostly applicable to frequency non-selective cases.

The problem of DPSK is that symbol errors appear pairwise: each received symbol is used twice, and a noise/ interference spike in a symbol would cause two consequtive symbols to be errorneous. So, as a good approximation, the noise analysis of PSK applies also for DPSK, but the error rates are doubled.

The DPSK idea can also be applied for higher order PSK modulations. In case of DQPSK, some commonly used mappings are shown in the following:

	Phase change (degrees)	
Bits	**V.26**	**V.22**
00	0	90
01	90	0
11	180	270
10	270	180

Spectral Efficiency

In many applications, it is important to maximize the information transmission rate for a given channel bandwidth.

When the size of alphabet is M, the maximum possible spectral efficiency is

$2\log_2 M$ bps/Hz for baseband PAM
$\log_2 M$ bps/Hz for I/Q modulated passband systems

It is more fair to compare N-level real alphabet in the baseband system and N^2-level complex alphabet in the passband system. In this case, the spectral efficiency of both systems is

$2\log_2 N$ bps/Hz

These are theoretical values when the excess bandwidth is 0 % and the channel is ideal. In practice, the spectral efficiency is lower. It is reduced by the exess band, guard bands in frequency domain, training symbols, and error control coding, etc.

Power efficiency measures how much signal power must be received to achieve a given SER/BER level. Spectral efficiency and power efficiency are somewhat contradictory requirements. Also in digital transmission systems, with a given transmission power, it is possible to improve the power efficiency by increasing the bandwidth. This is possible until a certain limit, as will be seen shortly.

Power efficiency is, of course, an important consideration in mobile and wireless communication systems with handheld terminals, as well as in satellite communications, and many other applications were the availability of electric power is limited.

Spectral Efficiency *(continued)*

***Example 1**: Voiceband modem*

The bit rate of 56 kbps is commonly used, so the spectral efficiency assuming a bandwidth of 3.2 kHz is

17.5 bps/Hz

***Example 2:** DVB-C, European cable TV system*

The basic configuration uses 64 QAM modulation and 15 % excess bandwidth which would result in theoretical (raw) spectral efficiency of

6/1.15=5.22 bps/Hz

However, the used frequency domain guard intervals between channels reduce the raw spectral efficiency, and the error control coding reduces the overall spectral efficiency to about

4.76 bps/Hz.

***Example 3:** GSM system*

The GSM system has 200 KHz channel spacing, the maximum user data rate is 8x14.4 kbps (in the HSCSD system using GMSK modulation). So (ignoring the signaling overhead), the spectral efficiency is

0.576 bps/Hz

The new enhancements (EDGE using basically 8PSK modulation) are raising this link-level spectral efficiency to about 1 bps/Hz.

The low spectral efficiency in mobile systems is due to the characteristics of the frequency selective fading channels.

Comparison to Channel Capacity

As seen earlier, the capacity of a continuous-time AWGN channel is (*Hartley-Shannon law)*

$$C = W \log_2(1 + \frac{S}{N})$$

where S/N is the received SNR in case of maximum transmission power.

Now we can calculate the maximal spectral efficiency as

$$v = \frac{C}{W} = \log_2\left(1 + \frac{S}{N}\right).$$

So for a given SNR, this is the maximum spectral efficiency that can be achieved.

Also, the minimum received signal power for given spectral effiicency can be calculated as

$$P = \sigma^2(2^v - 1).$$

To improve the spectral efficiency by 1 bps/Hz, the signal power has to be approximatively doubled.

Comparison to Channel Capacity *(continued)*

In an *M*-level PAM/QAM/PSK system, the following relation between the bit rate and SNR can be shown:

$$B = 2W \log_2 M = W \log_2 (1 + \frac{12}{\gamma^2} \frac{S}{N})$$

where γ is a factor that depends on the desired SER (minimum distance $d = \gamma\sigma$).

For example, with binary PSK, $\gamma = 8$ corresponds to the SER of 3×10^{-5}. In this case

$$\frac{\gamma^2}{12} = 5.3 \approx 7 \text{ dB}$$

So to achieve the SER of 3×10^{-5} without error control coding, the SNR should be 7 dB higher than what the channel capacity theorem tells.

For the SER of 10^{-7} dB, this difference is 9.5 dB.

The difference can be reduced by using error control coding.

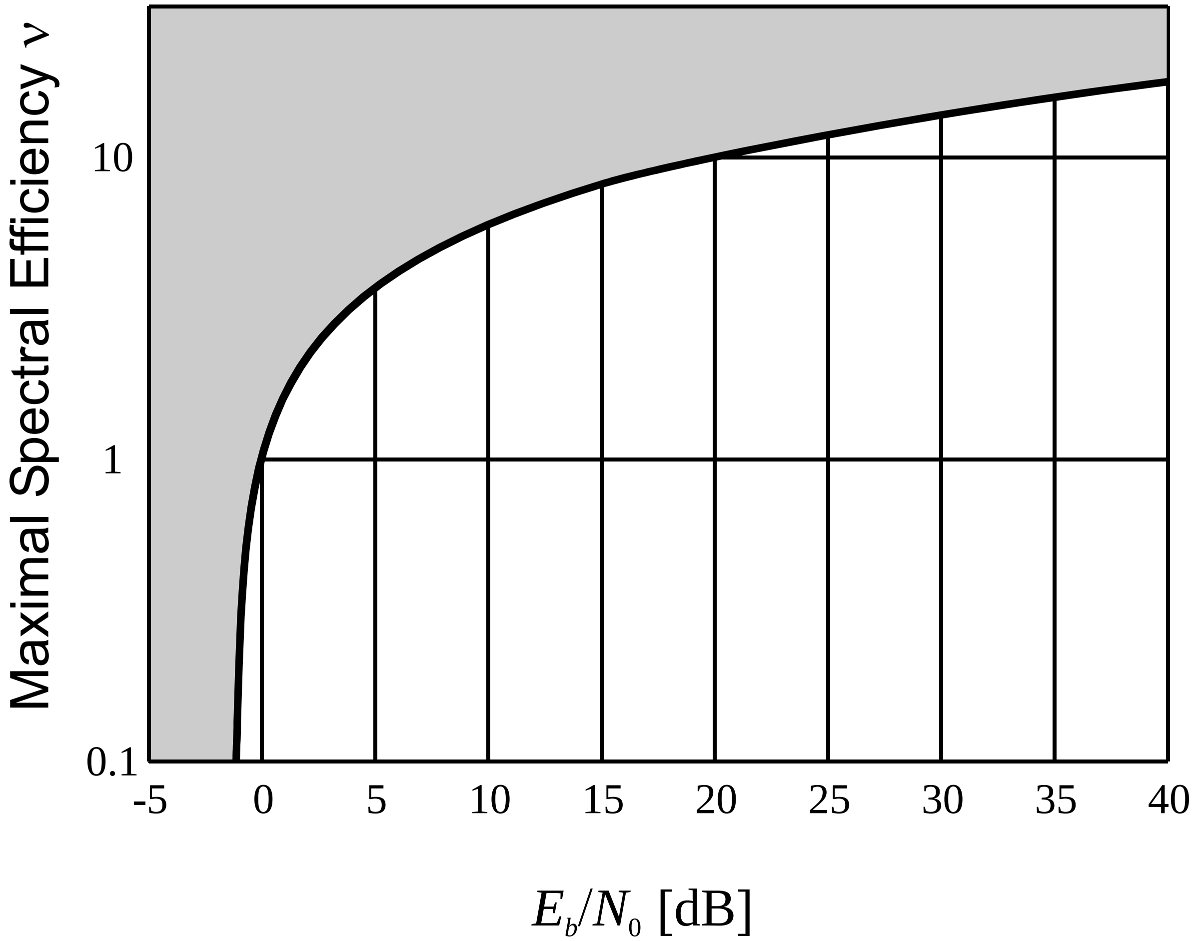
Comparison to Channel Capacity *(continued)*

In terms of E_b/N_0 ratio, the maximum spectral efficiency, ν, can be written as:

$$\nu = \frac{C}{W} = \log_2\left(1+\frac{S}{N}\right) = \log_2\left(1+\frac{E_b}{N_0}\frac{C}{W}\right) = \log_2\left(1+\frac{E_b}{N_0}\nu\right).$$

Here we have replaced R/W by C/W in the SNR expression, i.e., taking the bit rate equal to the channel capacity, which is appropriate when we are after the theoretical limit for maximum spectral efficiency.

The above relation can be plotted as:



Practical systems operate always in the white area; it is impossible to realize any system that would operate in the grey area.

The Shannon Limit

From the expression on the previous slide it follows that

$$\frac{E_b}{N_0} = \frac{2^{\nu} - 1}{\nu}$$

and

$$\lim_{\nu \to 0} \frac{E_b}{N_0} = \lim_{\nu \to 0} \frac{2^{\nu} - 1}{\nu} = \ln(2) \approx 0.693 \equiv -1.6 \text{ dB}$$

This indicates (as can be seen also from the figure of the previous page) that, whatever modulation and error control coding methods might be utilized, reliable communication with an E_b/N_0 ratio below –1.6 dB is impossible. This is the so-called Shannon limit.

FSK–TYPE MODULATION METHODS

Here we consider a class of nonlinear digital modulation methods that are alternatives to the earlier described linear digital modulation techniques. These include

FSK	Frequency Shift Keying
CPSK	Continuous-Phase FSK
MSK	Minimum Shift Keying
GMSK	Gaussian Minimum Shift Keying

These methods differ very much from the linear digital modulation methods. For example, pulse shaping principle is not utilized, at least not in the same way as in linear methods.

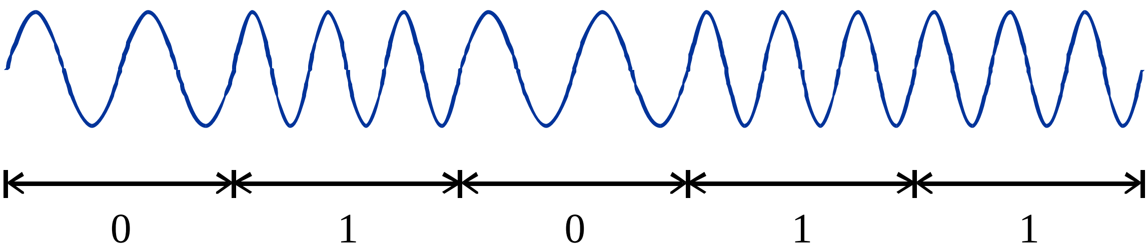
FSK has some similarties with analog FM, but also a lot of differences, and not much of the analog FM results can be utilized in the digital FSK case.

Source: Lee&Messerschmitt, Chapter 6.

FSK: Principle

In case of an *M* symbol alphabet, the idea is to use *M* different frequencies for each of the symbols.

The simplest case is binary FSK, where 0 and 1 correspond to different frequencies:



FSK properties

+ Non-coherent detection possible

⇒ Carrier synchronization not necessary

  ⇒ Easy to implement, simple hardware.

  o However, coherent detection gives better performance, also in case of FSK.

⇒ Robust also in fast fading cases

+ FSK and many (but not all) of its variants are strictly constant envelope modulations.

⇒ Immunity to non-linearities because envelope carries no information and information is at zero-crossings.

  ⇒ Non-linear power amplifiers can be used resulting in low power consumption in transmitter.

  o In simple schemes (e.g., when there are no channel equalizers involved), non-linear (e.g., hard-limiting) amplifiers can be used also on the receiver side, but only after channel selection filtering.

- FSK Issues
  - o 3dB higher SNR needed in coherent detection of binary FSK than in binary PSK.
  - o Multilevel FSK has good performance with low SNRs, but they waste bandwidth.
  - o Basic FSK has low spectral efficiency.
  - o More advanced techniques closer to linear modulations in terms of spectral efficiency.
  - o It is difficult to analyse FSK due to modulation non-linearities. For example, analytic expressions for power spectrum are difficult to find.

Orthonormal Modulations

An important general class of modulation methods are orthogonal modulations, and with suitable scaling they are orthonormal. This means that the basic symbol waveforms, $g_i(t)$, $i=1,\ldots,M$, are orthonormal, i.e.

$$\int_{-\infty}^{\infty} g_i(t) g_j(t) dt = \delta_{i-j} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}.$$

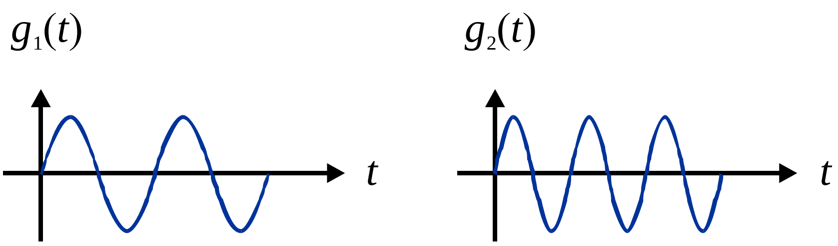
In general, the modulated signal can be written as

$$S(t) = \sum_{m=-\infty}^{\infty} A g_m (t - mT).$$

In the FSK case we can define

$$g_i(t) = \begin{cases} \frac{1}{\sqrt{2T}} \cos \omega_i t & \text{for } 0 \leq t \leq T \\ 0 & \text{otherwise.} \end{cases}$$

Example: Binary FSK

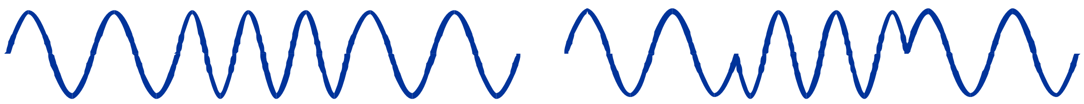


With suitable choice of the frequencies FSK is orthonormal. The most common cases of FSK are orthonormal, but there are also some interesting cases where this condition is not satisfied.

It should be noted that there are many other orthonormal modulation methods than the FSK family.

Continuous-Phase FSK

The phase of FSK may be continuous or discontinuous:



Continuous phase is better in terms of bandwidth because discontinuities cause high frequency components to the spectrum. Continuous-phase is also better when then transmission link has non-linearities, e.g., in the transmitter power amplifier.

In FSK, continuous phase is easily obtained when every pulse is composed of full carrier cycles, i.e., when the symbol interval T and the FSK frequencies f_i satisfy

$$f_i T = M_i, \qquad i = 1, \cdots, M.$$

where M_i are integers.

It is easy to prove that with this choice of the frequencies (and suitable scaling) FSK is orthonormal.

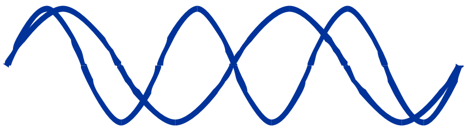
Continuous-Phase FSK *(continued)*

It is sensible to minimize the frequency separation to obtain minimum bandwidth while maintaining continuous phase and orthogonality of pulses. Using higher separation doesn't improve the performance in any way, it would only waste bandwidth.

In case of binary FSK we have

$$fT_i = M_1 \qquad\qquad f_1 T = M_2$$

Minimum frequency separation $|f_1 - f_2|$ is achieved when $|M_1 - M_2| = 1$.



More generally, we can choose

$$f_i - f_{i-1} = 1/T \qquad i = 1, \cdots, M.$$

to achieve orthogonality.

It should be noted that the orthogonality does not depend in any way on the phase of the sinusoids of the basic symbol waveforms and instead of the earlier expression, we can use

$$\frac{1}{\sqrt{2T}} \cos(\omega_i t + \phi_i)$$

with arbitrary ϕ_i.

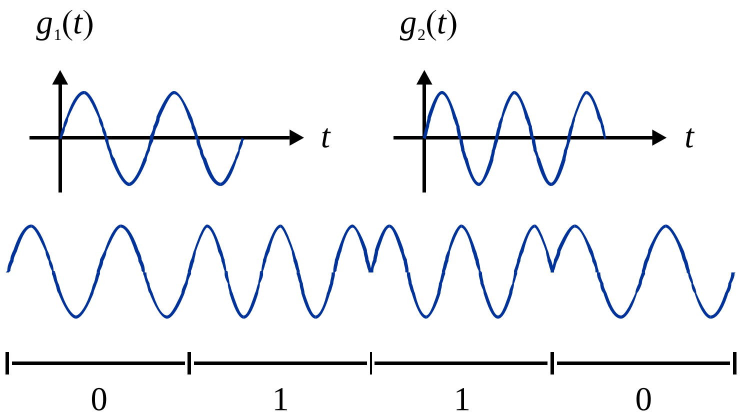
MSK, Minimum Shift Keying

Here the frequency differences are half of the previous:

$$f_i - f_{i-1} = 1/2T \qquad i = 1, \cdots, M.$$

This makes the bandwidth more narrow. It can be proven that the error probability does not depend on frequency difference.

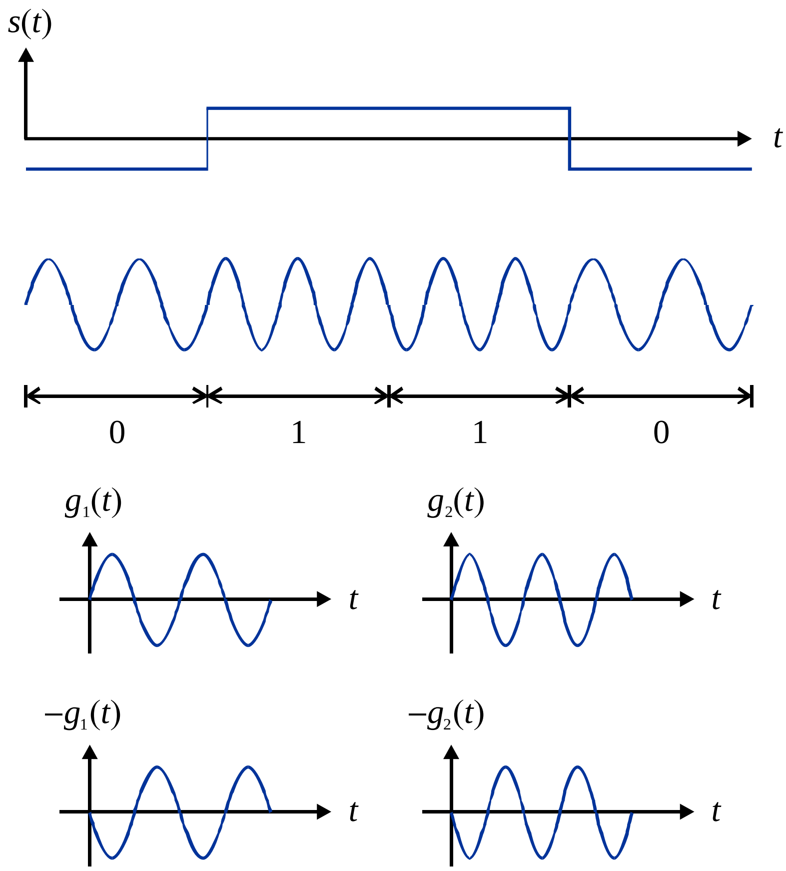
Example: Binary FSK with $|f_2 - f_1| = 1/2T$:



CPFSK principle gives better result and we obtain the following MSK modulation waveforms:

Phase continuity is obtained by changing the sign of $g_i(t)$ appropriately.

It can be proven that orthogonality condition is valid in this case, too.



Offset QPSK Interpretation of MSK

Binary MSK signal can be written in the baseband I/Q format as

$$s(t) = \sum_{k=-\infty}^{\infty} (-1)^{s_k} \Pi(t - kT - \tfrac{T}{2}, T)\, e^{j(-1)^{b_k} \pi t/(2T)}$$

Here the sequence s_k is chosen to satisfy the phase continuity, b_k is the transmitted bit sequence, and $\Pi(t,T)$ is a square pulse of width T (=bit interval) and centered at 0. For convenience of notation, the kth bit is here defined during the interval $kT \le t \le (k+1)T$, and in the baseband model it appears as a complex exponential with frequency $1/4T$ or $-1/4T$ depending on the bit value.

The above expression can be rewritten as:

$$s(t) = \sum_{k=-\infty}^{\infty} (-1)^{s_k} \Pi(t - kT - \tfrac{T}{2}, T)\left(\cos(\tfrac{\pi}{2T} t) + j(-1)^{b_k} \sin(\tfrac{\pi}{2T} t)\right)$$

$$= \sum_{k=-\infty}^{\infty} (-1)^{s_k} \Pi(t - kT - \tfrac{T}{2}, T)\left(\cos(\tfrac{\pi}{2T} t) + j(-1)^{b_k} \cos(\tfrac{\pi}{2T} (t - T)\right)$$

This signal is built up using a basic pulse shape in the form of quarter cycle of cosine wave. To satisfy the phase continuity according to the CPFSK principle, these quarter cycle pulses must be preceeded/followed by other quarter cycle pulses in such a way that half-cycle cosine pulses are obtained. Then we can write the signal as

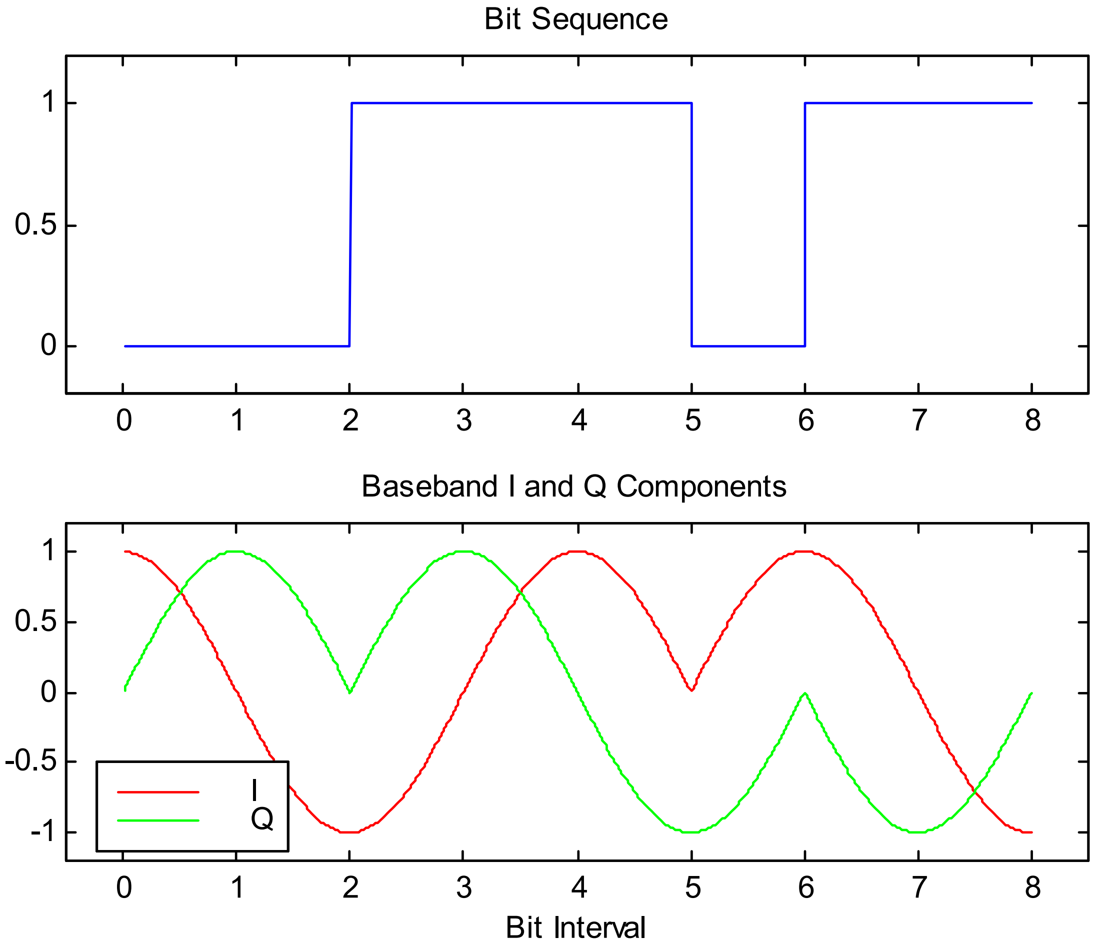
$$s(t) = \sum_{k=-\infty}^{\infty} (-1)^{\tilde{b}_{2k}} \Pi(t - 2kT, 2T) \cos(\tfrac{\pi}{2T} t) +$$

$$j(-1)^{\tilde{b}_{2k+1}} \Pi(t - (2k+1)T, 2T) \cos(\tfrac{\pi}{2T} (t - T))$$

where $\tilde{b}_k$ is a modified bit sequence. This can be interpreted as an offset QPSK signal with half-cycle of cosine wave as the pulse shape.

Offset QPSK Interpretation of MSK *(continued)*

Example:



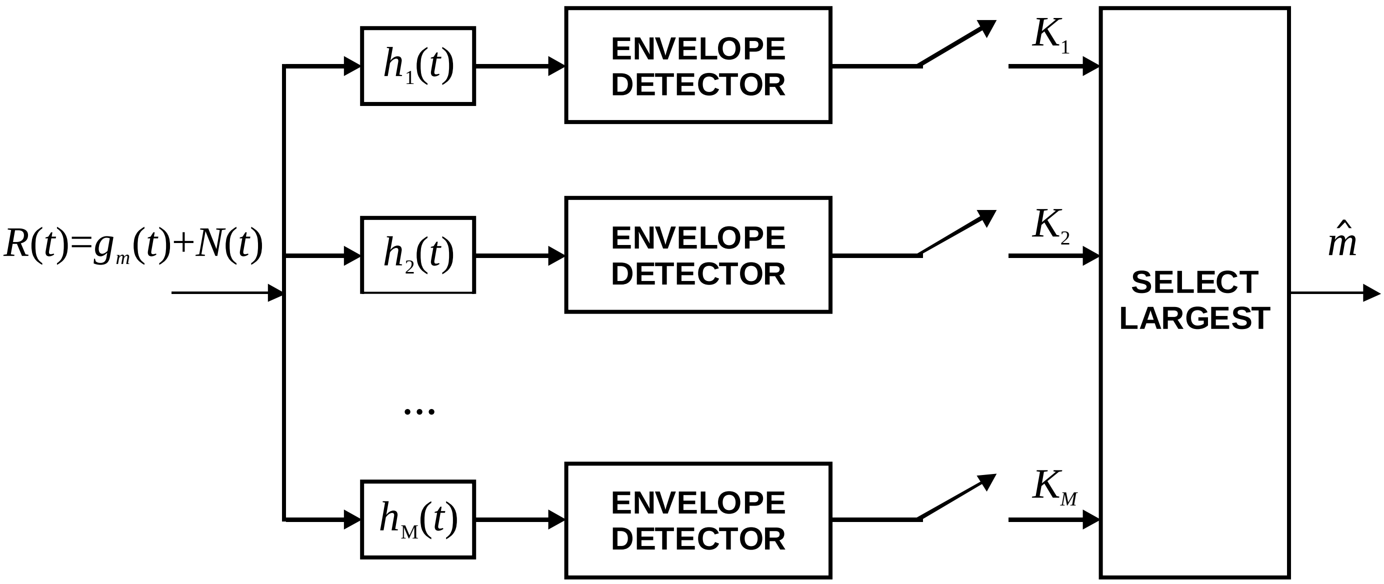
Notice that for 0-bits, the I component is 90 degrees ‘ahead of’ the Q component and for 1-bits, it is 90 degrees ‘behind’ the Q component. Phase continuity is achived by proper sign inversions.

The O-QPSK inperpretation can be seen quite clearly in this example.

We can note that in this very special case, the FSK family and the linear digital modulation methods coincide, i.e., we have a modulation method that has both interpretations.

Detection of FSK

Non-coherent detection can be implemented in a simple ad-hoc way as follows:

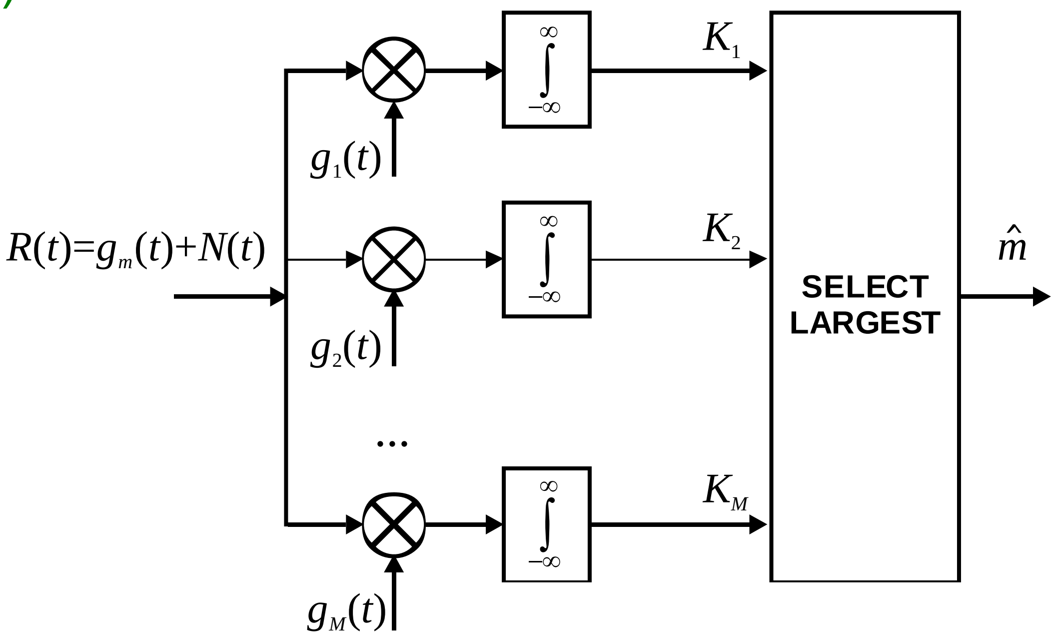


The structure includes basically a bank of bandpass filters, each of them tuned to one of the frequencies used in the modulation. Then, envelope detector and maximum selection is used to find the frequency with highest energy for each symbol.

To obtain optimum performance, and make the system to work also in the cases where the frequency differences are minimized, we have to use optimum receiver principles to be discussed later on in the course.

Correlation Receiver for FSK

The basic coherent optimum receiver structure is the correlation receiver (or the equivalent matched filter receiver):



This structure is correlating the received signal (considering an isolated symbol)

$$r(t) = g_m(t) + N(t)$$

with all the used symbol waveforms as

$$K_i = \int_0^T r(t) g_i(t) dt \quad \text{for } i = 1, \ldots, M$$

and choosing the largest among these.

It is clear that for orthonormal modulation and in ideal ISI and noise free conditions, this principle works perfectly.

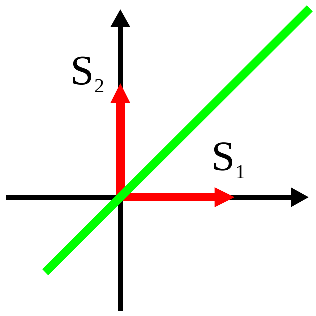
More generally, the decision is based on *M*-dimensional vectors composed of the correlator outputs.

Detecting Binary FSK

In the orthonormal binary FSK case, the signal vectors and noisy observation vectors are

$$\mathbf{S_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \mathbf{S_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \mathbf{R_1} = \begin{bmatrix} 1+n_1 \\ n_2 \end{bmatrix} \quad \mathbf{R_2} = \begin{bmatrix} n_1 \\ 1+n_2 \end{bmatrix}$$

and the correlation receiver produces the following signal geometry (this will be better understood after the detection theoretical part later):



The alternative symbol values appear as orthonormal signal vectors, and with the usual assumptions, it can be shown that the channel noise appears as additive Gaussian noise component. The decision boundary is defined in the same way as earlier. Then the symbol error probability becomes:

$$\Pr[symbol\ error] = Q\left(\frac{d}{2\sigma}\right) = Q\left(\frac{1}{\sigma\sqrt{2}}\right)$$

assuming that the ‘signal vector length’ is unity. Recalling that binary PSK (antipodal signalling) would give

$$\Pr[symbol\ error] = Q\left(\frac{1}{2\sigma}\right)$$

we see that coherent detection of binary FSK has 3 dB worse performance than binary PSK.

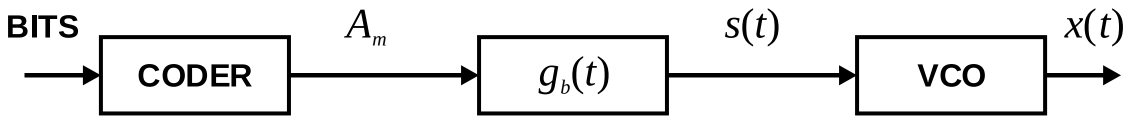
Continuous Phase Modulations

CPM can be implemented by FM-modulating the carrier by using a real baseband signal with proper pulse shape:

$$x(t) = K\cos\left[\omega_c t + \omega_d \int_{-\infty}^{t} s(\tau)\,d\tau\right]$$

$$s(t) = \sum_{m=-\infty}^{+\infty} A_m g_b(t - mT), \qquad |s(t)| \le 1.$$

This can be implemented, for example, by using voltage controlled oscillator (VCO):



Here, the continuous phase property is guaranteed automatically, for any frequency separation (i.e., this structure can be used also for non-orthonormal cases).

A special case, where $g_b(t)$ is rectangular pulse is called CPFSK.

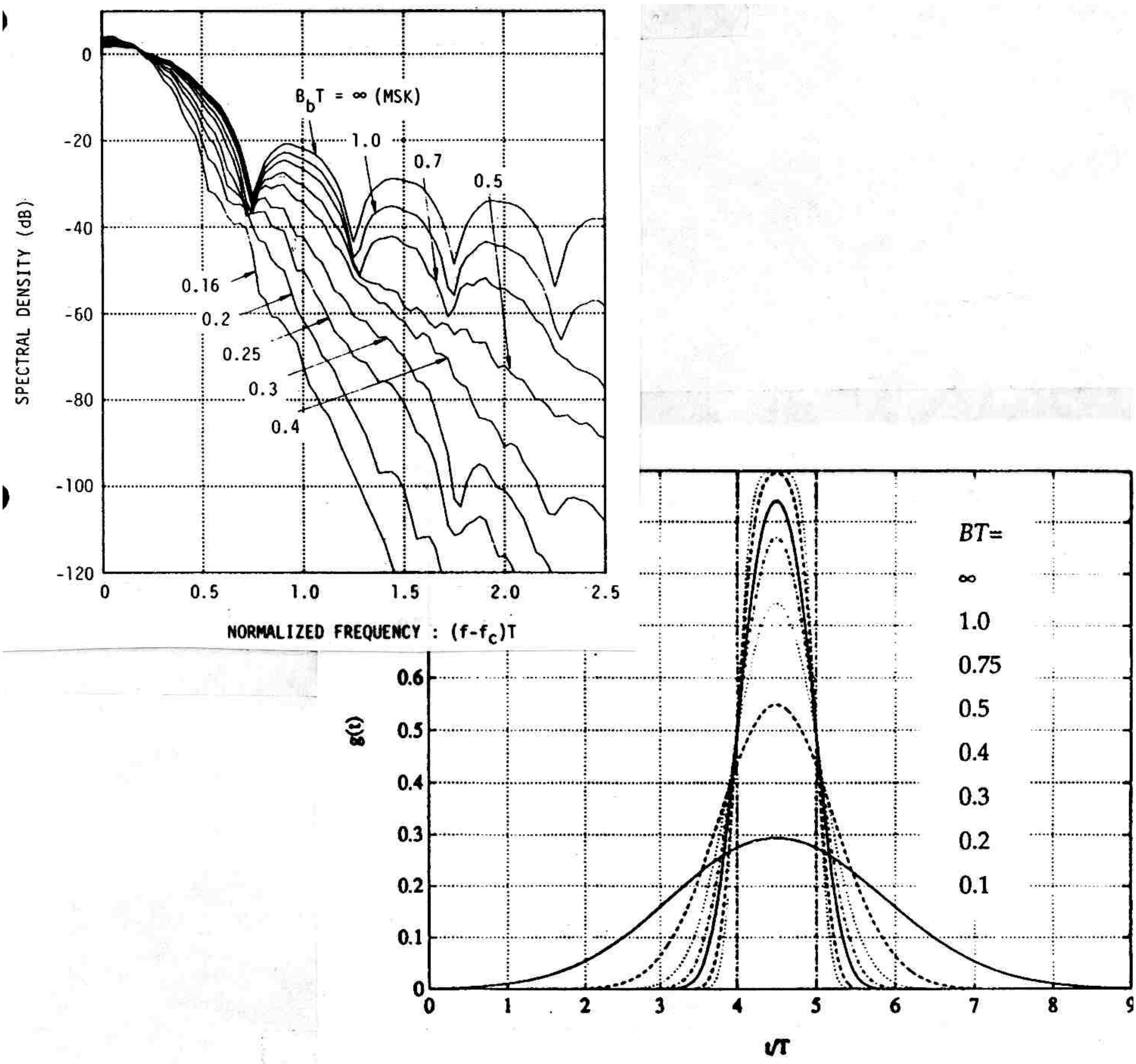
The non-linearity of CPFSK is clearly seen from the previous equation. In general, it is difficult to compute FSK or CPFSK spectrum. General observation is that the bandwidth of FSK is greater than in linear modulation methods with the same symbol rate.

Other pulse shapes than rectangular can be much better in terms of bandwidth, as will be seen shortly.

GMSK, Gaussian Minimum Shift Keying

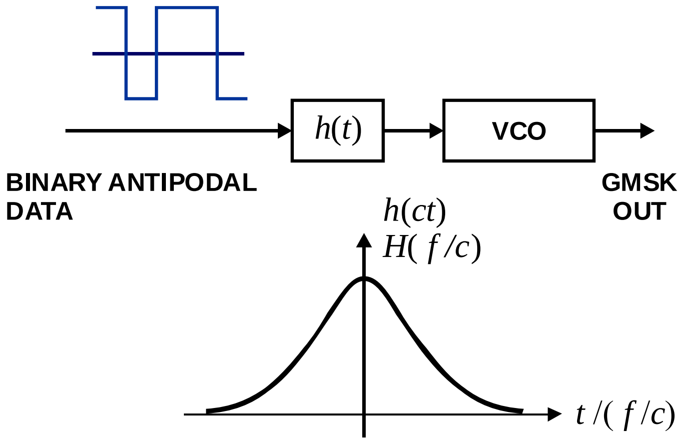
This is a variation of MSK modulation method, where a better pulse shape is used instead of square pulse. The actual pulse shape is determined by a *Gaussian lowpass filter.*

The parameter $B_b T$ is the normalized 3dB bandwidth of the Gaussian LPF. This parameter determines the pulse width in the time domain and the bandwidth of the modulated signal in frequency domain, as follows:

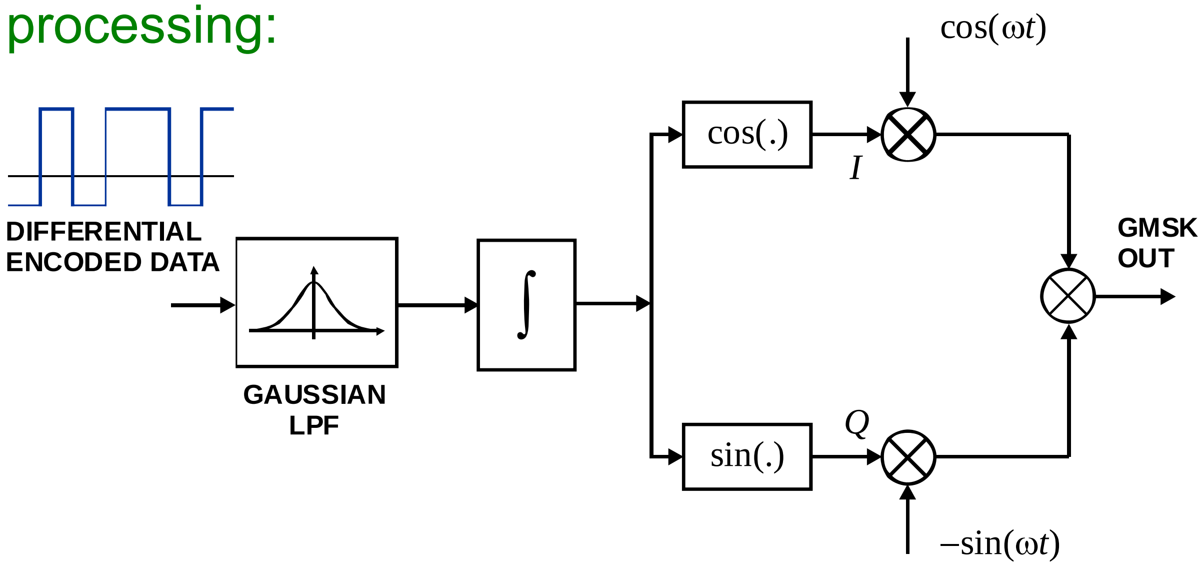


GMSK (cont.)

So, this is the basic structure for GMSK modulator:



And here is an alternative model based on nonlinear I/Q signal processing:



This is a nonlinear modulation method, and the Gaussian filter characteristics have no clear analytic relationship with the modulated waveform or frequency spectrum.

Notice that intersymbol interference is introduced when the pulse shape is other than square. However, this is controlled ISI which can be removed, e.g., by using the Viterbi algorithm for detection.

GMSK is the basic modulation method utilized in the GSM system, with the parameter value $B_bT=0.3$. The following page shows how it is defined in the GSM specs.

GMSK (cont.)

10.6 Modulation technique

Gaussian Minimum Shift Keying (GMSK) shall be used to modulate the high bit rate data for transmission, and Frequency Shift Keying (FSK) shall be used to modulate the low bit rate data, as defined in the following subclauses.

10.6.1 Gaussian Minimum Shift Keying (GMSK)

GMSK is specified as follows.

For the purposes of defining the modulation scheme, the data to be modulated is considered to be a sequence of bits d_i where $d_i \in \{0,1\}$.

The modulating data value input to the modulator is:

$$\alpha_i = 1 - 2d_i \qquad (\alpha_i \in \{-1,+1\})$$

The modulating data values, as represented by Dirac pulses, excite a linear filter with impulse response defined by:

$$g(t) = h(t) * rect(t/T)$$

where the function $rect$ is defined by:

$$rect(t/T) = 1/T \qquad \text{for } |t| < T/2$$

$$rect(t/T) = 0 \qquad \text{otherwise}$$

and * denotes convolution. $h(t)$ is defined by:

$$h(t) = \exp(-t^2/(2\sigma^2T^2))/(\sqrt{(2\pi)}\sigma T)$$

where $\sigma = \sqrt{\ln(2)}/2\pi BT$ and $BT = 0.3$

where B is the 3 dB bandwidth of the filter with impulse response $h(t)$ and T is the duration of one input data bit, equal to the reciprocal of the bit rate defined in subclause 10.6.1 and approximately equal to 42,5 ns.

The phase of the modulated signal is:

$$z(t') = \sum_i \alpha_i \pi h \int_{-\infty}^{t'-iT} g(u)\,du$$

where the modulating index h is 1/2 (maximum phase change in radians is $\pi/2$ per data interval).

The time reference $t' = 0$ is the start of the active part of the burst.

The modulated RF carrier, except during the start and end of the burst may therefore be expressed as:

$$x(t') = \sqrt{(2E_c/T)}\cos(2\pi f_c t' + z(t') + z_0)$$

where E_c is the energy per modulating bit, f_c is the centre frequency and z_0 is a random phase and is constant during one burst.

10.6.2 Frequency Shift Keying (FSK)

FSK is specified as follows.

Each low bit rate bit shall be mapped to a transmitted frequency as shown in the table below.

Table 50: Nominal frequencies for FSK modulation

Performance Comparison of FSK-Based Methods

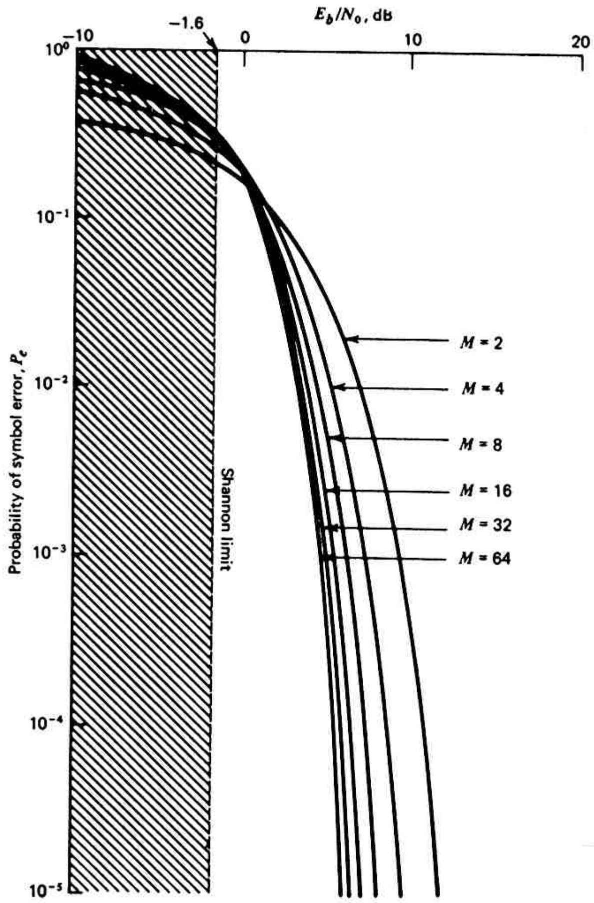


Figure 7.36 Error-rate curves for M-ary FSK.

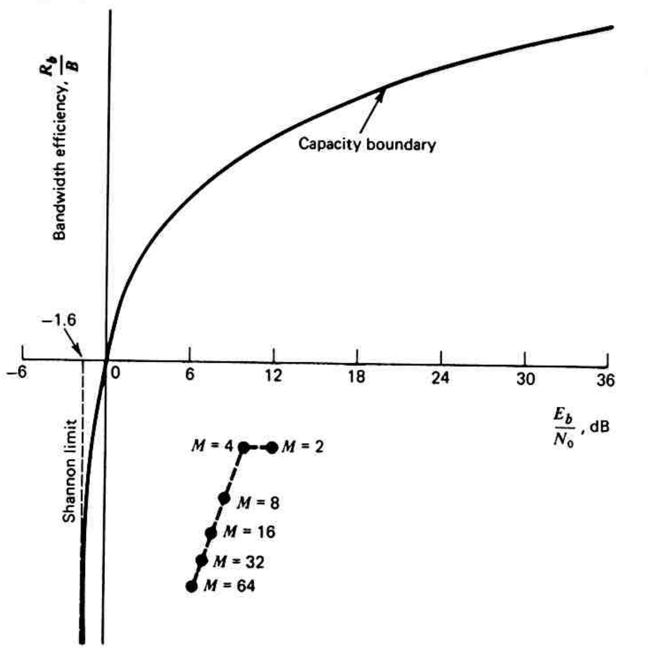


Figure 7.35 Comparison of M-ary FSK with the ideal system.