

Digital Modulation Techniques

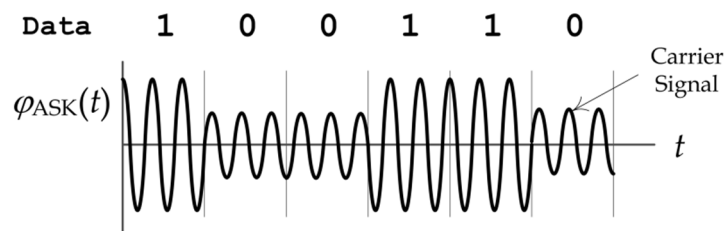
For more information: read Chapters 7 and 10 in your textbook.

There are four major *modulation* techniques used by communication systems nowadays to transport baseband digital data onto a carrier. These modulation techniques are:

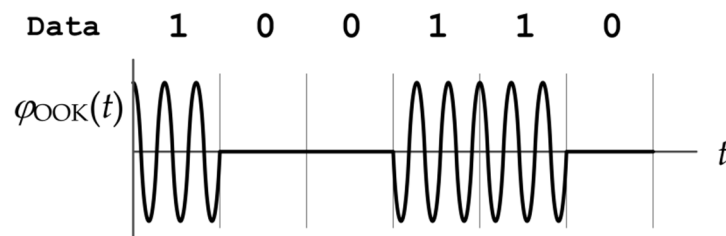
- **Amplitude-Shift Keying (ASK)**
- **Frequency-Shift Keying (FSK)**
- **Phase-Shift Keying (PSK)**
- **Quadrature Amplitude Modulation (QAM)**

Amplitude-Shift Keying (ASK):

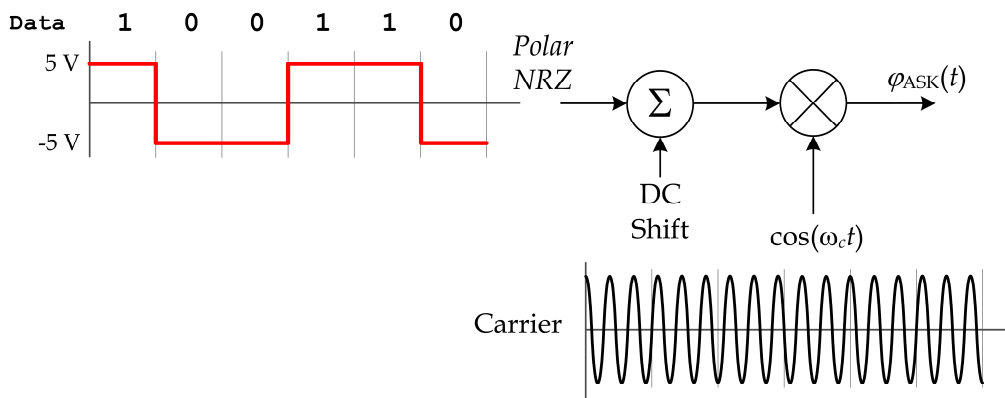
ASK represents digital data as variations in the amplitude of a carrier signal. For example the transmitter could send the carrier $2A \cos \omega_c t$ to represent a logic **1**, while using the carrier $A \cos \omega_c t$ to represent a logic **0**. This is shown in the diagram below. The receiver detects the amplitude of the carrier to recover the original bit stream.



A special case of ASK is when a logic **1** is represented by $A \cos \omega_c t$ (i.e., the presence of a carrier) and a logic **0** is represented by a zero voltage (i.e., the absence of a carrier). This special case is called **On-Off Keying (OOK)** and is shown below.



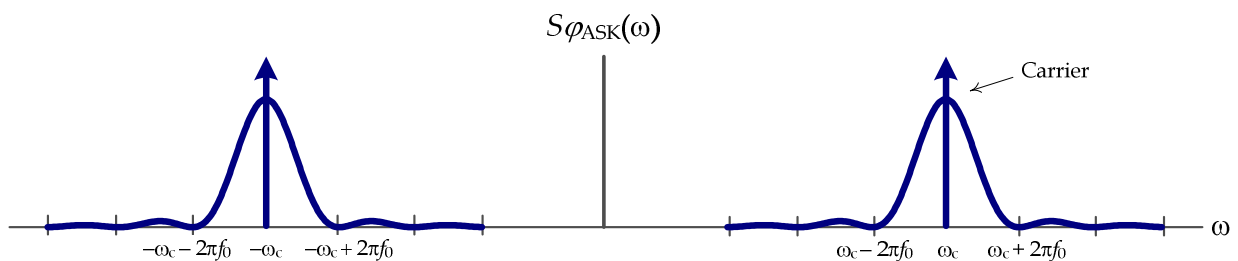
Notice that you can visualize ASK as the process of **Amplitude Modulation (AM)** using a “**Polar NRZ**” digital baseband message signal. In other words, we say that ASK is the result of multiplying a binary Polar NRZ signal $m(t)$ (with appropriate DC shift) times a sinusoidal carrier. This is shown in the diagram below:



The above diagram shows that a general ASK signal is simply an AM signal with a modulation index $m < 1$, while an OOK is an AM signal with a modulation index $m = 1$. Hence, an envelope detector can be used at the receiver to demodulate the ASK signal. Actually, all types of modulators and demodulators we explained earlier for AM are now applicable to ASK provided that the message signal $m(t)$ used is a digital baseband Polar NRZ signal.

In addition, since ASK is a special case of AM modulation, the **bandwidth of ASK is $2B$** centered around the carrier frequency ω_c , where B is the bandwidth of the Polar NRZ signal. Since the bandwidth of Polar NRZ is equal to the data bit rate (f_o) of the bit stream to be sent, the **bandwidth of ASK is $2f_o$ (Hz)**.

The following is a sketch of the PSD for an ASK signal. It consists of two replicas of the PSD for a Polar NRZ signal with additional carrier impulses. You can see that the bandwidth of this ASK signal is approximately $2f_o$ (Hz).

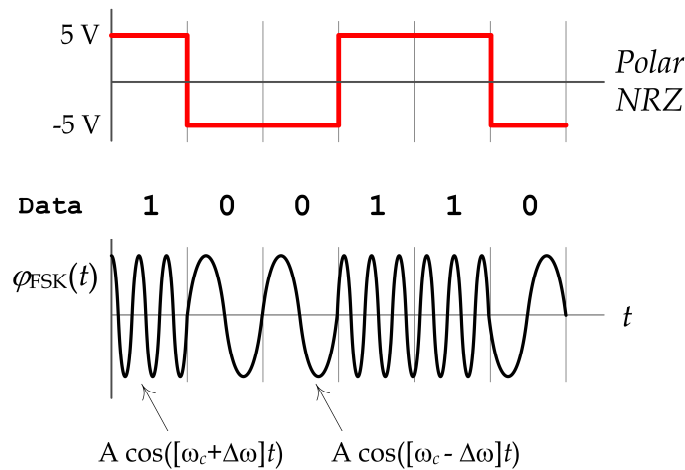


The advantages and disadvantages of ASK are summarized below:

Advantages	Disadvantages
<ul style="list-style-type: none"> ASK is the simplest kind of modulation to generate and detect. 	<ul style="list-style-type: none"> It can be used only when the signal-to-noise ratio (SNR) is very high. Its bandwidth is too big (equals $2f_o$).

Frequency-shift keying (FSK):

In FSK the instantaneous frequency of the carrier signal is shifted between two possible frequency values termed the *mark* frequency (representing a logic 1) and the *space* frequency (representing a logic 0). This is shown in the diagram below.



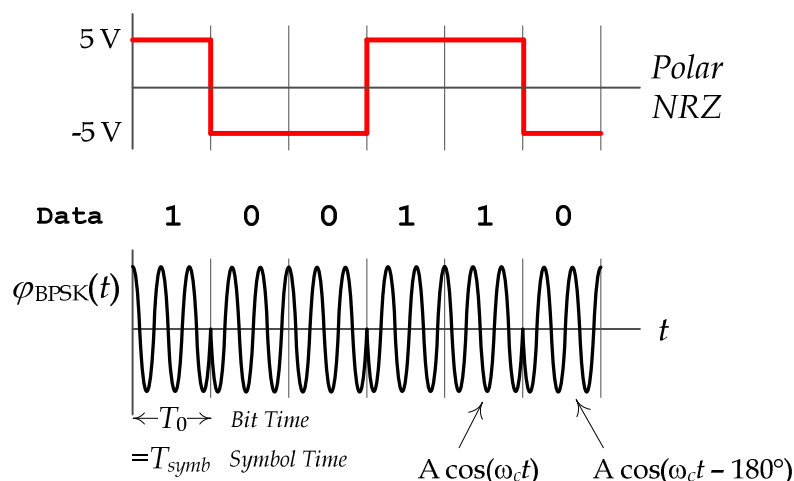
Notice that FSK can be thought of as **Frequency Modulation (FM)** using a “**Polar NRZ**” digital baseband signal as the message, and hence FSK can be seen as a subset of FM modulation.

Since FSK is a special case of FM modulation, the **bandwidth of FSK** is given by Carson’s rule which says that $B_{FM} \approx 2\Delta f + 2B$, where B is the bandwidth of the Polar NRZ signal (equal to f_o (the bit rate)). Hence, the **bandwidth of FSK is $2\Delta f + 2f_o$ (Hz)**. In addition, all modulator and demodulator circuits for FM are still applicable for FSK.

FSK has several advantages over ASK due to the fact that the carrier has a constant amplitude. These are the same advantages present in FM which include: immunity to non-linearities, immunity to rapid fading, immunity to adjacent channel interference, and the ability to exchange SNR for bandwidth. FSK was used in *early slow dial-up modems*.

Phase-shift keying (PSK):

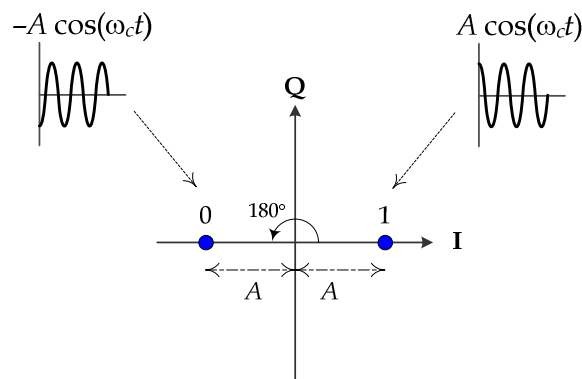
In PSK, the data is conveyed by changing the phase of the carrier wave. One possible representation (called Binary Phase-Shift Keying or **BPSK**) is to send logic **1** as a cosine signal with zero phase shift and a logic **0** as a cosine signal but with a 180° phase shift. We say in this case that the BPSK signal can assume one of two possible *symbols*: 0° and 180° . This case is shown in the following Figure.



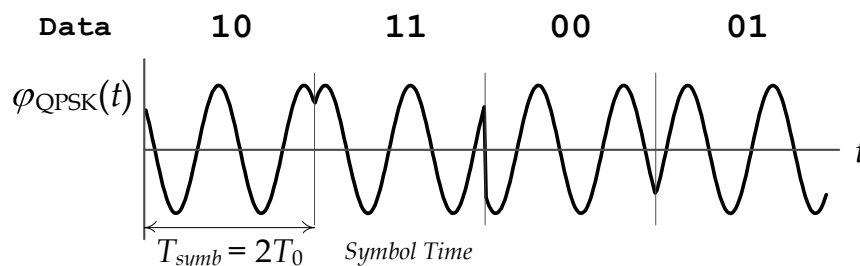
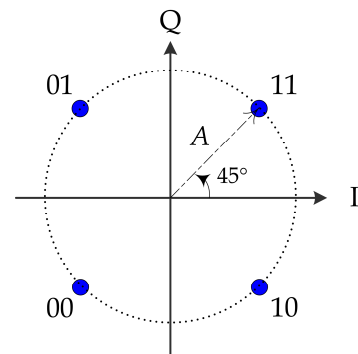
BPSK can be thought of as a special case of **Phase Modulation (PM)** using a “**Polar NRZ**” digital baseband message¹. In the case of BPSK, we select the peak phase deviation to be $\Delta\theta = \pi/2$ (i.e., $2\Delta\theta = \theta_{max} - \theta_{min} = \pi$). This value maximizes immunity to phase noise.

Since BPSK is a special case of PM, the **bandwidth of PSK is $2B + 2\Delta f$** , where B is the bandwidth for the polar NRZ signal and $\Delta f = 0$ since the sinusoidal carrier signal does not change its frequency. Hence, the **bandwidth of BPSK is $2f_o$ (Hz)**.

A convenient way to represent PSK modulation is using a **constellation diagram**. A constellation diagram consists of a group of points representing the different symbols the carrier in a PSK modulated signal can assume. For example, for **BPSK**, in which each bit is represented by one symbol (i.e., either $A \cos(\omega_c t)$ or $A \cos(\omega_c t - 180^\circ)$), the constellation diagram consists of two points (see Figure below). These two points have the same amplitude A , but they are 180° apart. This means that a logic 1 corresponds to $A \cos(\omega_c t)$, while a logic 0 corresponds to $A \cos(\omega_c t - 180^\circ)$.



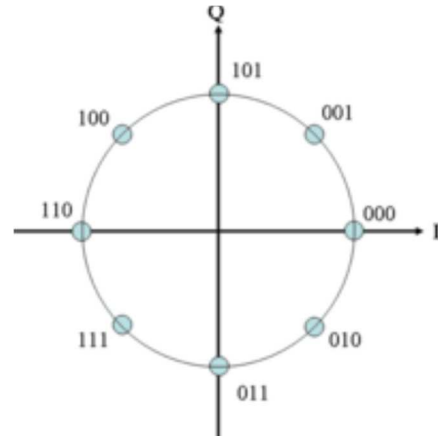
Another common example of PSK is **Quadrature (or Quaternary) Phase-Shift Keying (QPSK)**. QPSK uses four possible phases for the carrier ($45^\circ, 135^\circ, 225^\circ, 315^\circ$) but with the same carrier amplitude, as shown in the constellation diagram to the right. With four phases, QPSK can encode two bits per one symbol (see Figure below).



¹ Notice that BPSK can also be thought of as a special case of DSB-SC in which the Polar NRZ signal DSB-SC modulates a sinusoidal carrier. This is because multiplying a carrier by positive and negative values switches its phase by 180° .

You can imagine QPSK as a special case of Phase Modulation (PM) in which the baseband message signal $m(t)$ is a digital M -ary signal (with $M = 4$). In this case, the bandwidth of the M -ary baseband signal is $B = \text{Baud Rate} = f_o / 2$, which means that the bandwidth of the QPSK signal is $2B + 2\Delta f = f_o$ instead of $2f_o$ for BPSK. Hence, QPSK can be used to double the data rate compared to a BPSK system while maintaining the same bandwidth of the modulated signal.

Notice that any number of phases may be used to construct a PSK constellation. Usually, **8-PSK** is the highest order PSK constellation deployed in practice (see the figure to the right). In this case, each carrier symbol represents three bits.



With more than eight phases, the error-rate becomes too high and there are better, though more complex, modulation schemes available (such as QAM).

Notice that in PSK, the constellation points are usually positioned with uniform angular spacing **around a circle**. This gives **maximum phase-separation** between adjacent points and thus the best immunity to noise. Points are positioned on a circle so that all the different phases can be transmitted with the **same carrier amplitude**.

The axes in a constellation diagram are called the **in-phase (I)** and **quadrature (Q)** axes, respectively, due to their 90° separation. The nice thing about a constellation diagram is that it lends itself to straightforward and simple implementation of PSK modulation in hardware. This is because the PSK modulated signal can be generated by individually DSB-SC modulating both a sine wave and a cosine wave and then adding the resulting modulated carriers to each other. In such case, the constellation diagram is extremely helpful since the amplitude of each point along the in-phase axis is the one used to modulate the cosine wave and the amplitude along the quadrature axis is the one used to modulate the sine wave. This procedure will be much more obvious when we discuss QAM modulation in the next section.

It is worth mentioning that BPSK and QPSK can be regarded special cases of the more general QAM modulation, where the amplitude of the modulating signal is constant (see next section).

Example: Find the bandwidth of an 8-PSK modulated signal if the data bit rate is 100 kbit/s.

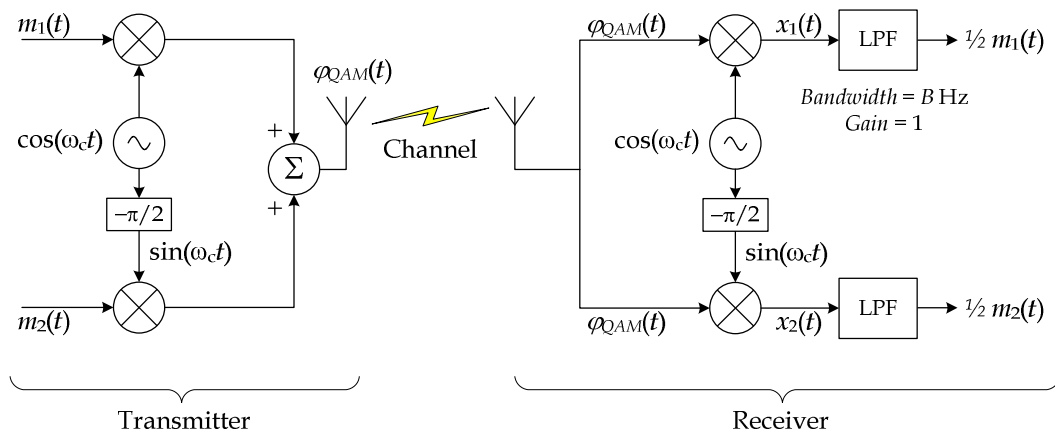
Solution: For 8-PSK, Bandwidth = $2B = 2 \times \text{Baud Rate} = 2 \times \frac{100\text{kbps}}{\log_2(8)} = 2 \times \frac{100\text{kbps}}{3 \text{ bits/symb}} = 66.67 \text{ kHz}$.

Quadrature Amplitude Modulation (QAM):

QAM is a modulation scheme which conveys data by modulating the **amplitude of two carrier waves**. These two waves (a cosine and a sine) are **out of phase** with each other by 90° and are thus called quadrature carriers – hence the name of the scheme.

Both **analog and digital QAM** are possible. Analog QAM was used in **NTSC and PAL television systems**, where the I- and Q-signals carry the components of chrominance (color) information.

Let us start by remembering analog QAM, which allowed us to transmit two message signals using two orthogonal carriers of the same frequency. The following Figure shows this scheme. Notice that both modulated signals will **occupy the same frequency band** around ω_c .

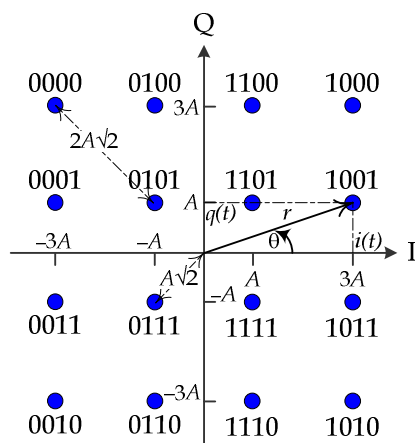


The two baseband signals **can be separated** at the receiver by **synchronous detection** using two **local carriers in phase quadrature**. This can be shown by considering the multiplier output $x_1(t)$ of the top branch (see Figure above):

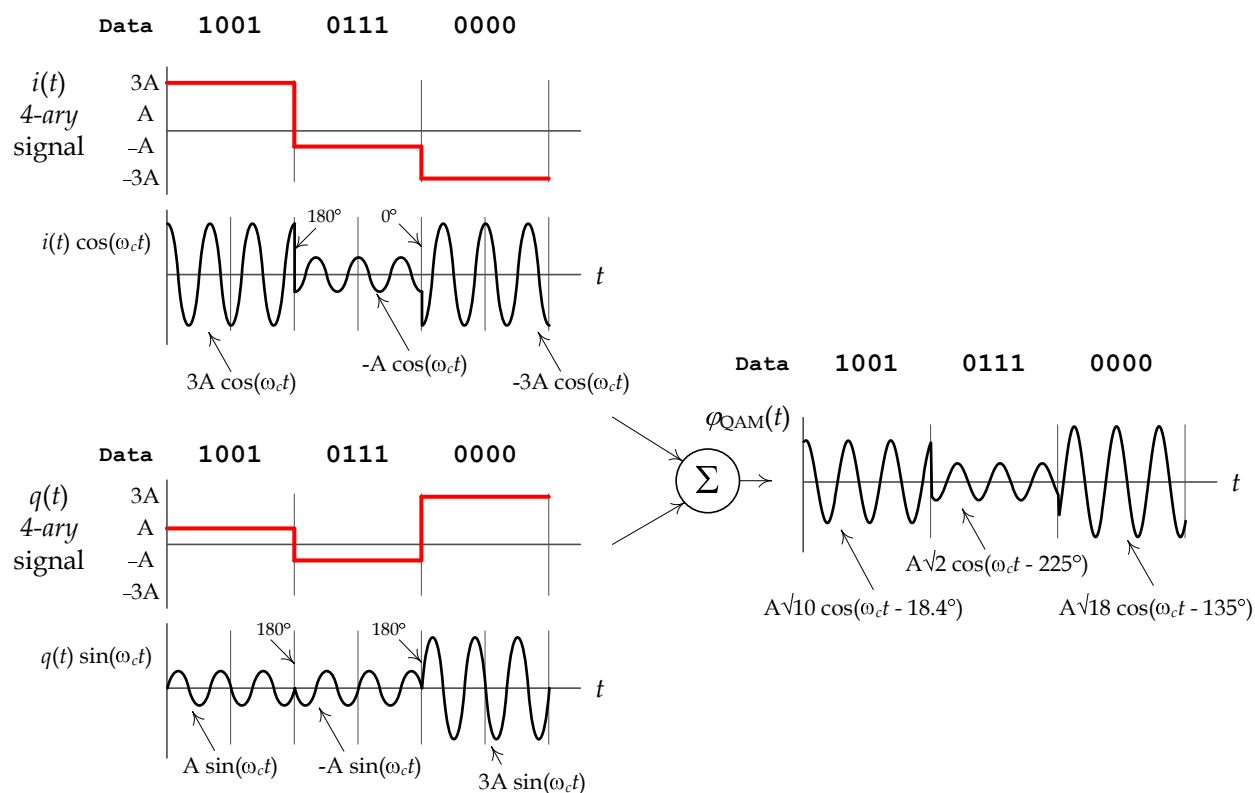
$$\begin{aligned}
 x_1(t) &= \phi_{QAM}(t) \times \cos(\omega_c t) = [m_1(t) \cos(\omega_c t) + m_2(t) \sin(\omega_c t)] \times \cos(\omega_c t) \\
 &= \frac{1}{2} m_1(t) + \frac{1}{2} m_1(t) \cos(2\omega_c t) + \frac{1}{2} m_2(t) \sin(2\omega_c t)
 \end{aligned}$$

The last two terms are suppressed by the **lowpass filter (LPF)**, yielding the desired output $m_1(t)/2$. Thus, in QAM two signals can be transmitted simultaneously over a bandwidth of $2B$, and still get separated at the receiver.

Digital QAM, on the other hand, is constructed using two M -ary baseband signals (called $i(t)$ and $q(t)$) modulating the two quadrature carriers. For example, in **16-QAM** both $i(t)$ and $q(t)$ are 4-ary digital baseband signals, which means each one of them can assume one of four possibilities. This results in $4 \times 4 = 16$ possible carrier symbols as shown in the constellation diagram below. Hence, 16-QAM uses 16 symbols, with each symbol representing a specific four-bit pattern.



For example, to send the bit sequence 100101110000 using 16-QAM, the bit stream is split into 4-bit groups, with each 4-bit pattern affecting $i(t)$ and $q(t)$ as shown in the figure below.



Notice that the baud rate (i.e., the symbol rate) of the resulting 16-QAM signal is *one fourth* that of the data bit rate. This is why the bandwidth of 16-QAM is $2 \times \text{Baud Rate} = 2 f_o / 4 = f_o / 2$. You can see that this is correct because the bandwidth of each one of the 4-ary signals is $B = f_o / 4$ (one symbol per four bits). Performing DSB-SC modulation for each one of these signals (i.e., QAM) results in a total bandwidth of $2B = 2 (f_o / 4) = f_o / 2$.

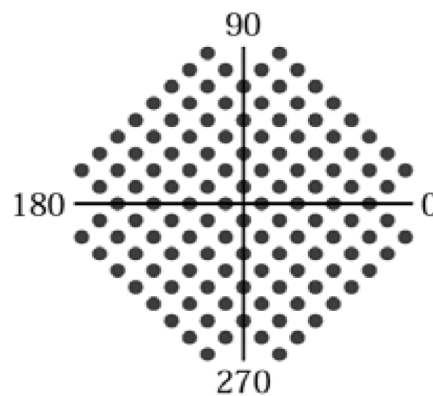
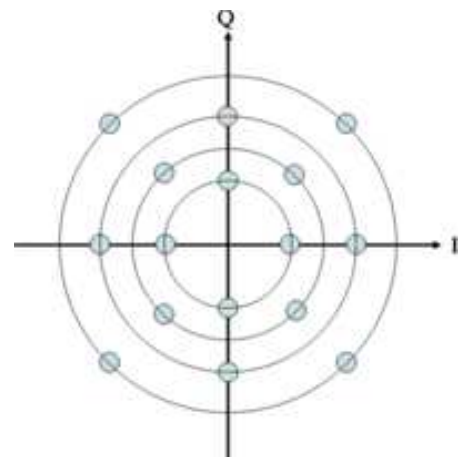
Example: Find the bandwidth of an 16-QAM modulated signal if the data bit rate is 8 Mbit/s.

Solution: For 16-QAM, Bandwidth = $2 \times \text{Baud Rate} = 2 \times \frac{8 \text{ Mbps}}{\log_2(16)} = 2 \times \frac{8 \text{ Mbps}}{4 \text{ bits/symb}} = 4 \text{ MHz}$.

In QAM, the constellation points are usually arranged in a square grid with equal vertical and horizontal spacing called **rectangular QAM** (see the above constellation diagram). The number of points in the grid is usually a power of 2 (2, 4, 8...). The most common forms of QAM are **16-QAM**, **64-QAM**, **128-QAM** and **256-QAM**. By moving to higher-order constellations, it is possible to transmit more bits per symbol, which reduces bandwidth. However, if the mean energy of the constellation is to remain the same, the points must be closer together and are thus more susceptible to noise; this results in a higher bit error rate (BER) and, hence, higher-order QAM can deliver more data *less* reliably than lower-order QAM unless, of course, the SNR is increased.

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 are easier to generate and demodulated using simple hardware. Non-square constellations achieve marginally better performance but are harder to modulate and demodulate.

For example, the diagram of **circular 16-QAM** constellation is shown to the right. The constellation diagram shown below is the one used in the **V.32bis dial-up modem**. This modem provides **14.4 kbit/s** using only **2400 baud rate**. Can you calculate the number of constellation points from these numbers?²



Note: It is worth mentioning that in practical systems, M -ary signals are shaped using a raised-cosine pulse before modulating the two quadrature carriers. In such case, the bandwidth of QAM (or PSK) becomes $2 \times \text{Baud} \times (1 + \beta)/2$ instead of just $2 \times \text{Baud}$.

² Remember that baud rate = symbol rate = bit rate / bits-per-symbol. This means that the number of data bits-per-symbol is $k = 14400 / 2400 = 6$ bits. However, if you count the number of symbols in the above constellation diagram, you will notice 128 symbols corresponding to $k = \log_2(128) = 7$ bits. This seeming contradiction is a result of using one of the seven bits as a parity bit not as a data bit. Using parity bits combined with high-order modulation is called *Trellis Coded Modulation* (TCM).

Performance of digital modulation techniques in presence of Noise:

Remember that we measured the performance for analog modulation techniques in terms of signal quality, which was related to **output signal-to-noise ratio** (SNR_{out}). For digital modulation techniques, the performance is measured in terms of **output bit error rate (BER)**, which represents the number of erroneous bits that the receiver expects per second. For example, a $BER = 10^{-4}$ means that we expect *on average* 1 bit error out of every 10,000 transmitted bits. We say the system exhibits good performance if the $BER \leq 10^{-6}$.

Remember that we are using the Additive White *Gaussian* Noise (AWGN) mathematical model to describe the noise on a communication channel. Hence, the noise $n(t)$ is considered as a Gaussian random process with zero average and a variance σ^2 . The variance of the noise σ^2 is its **average power**.

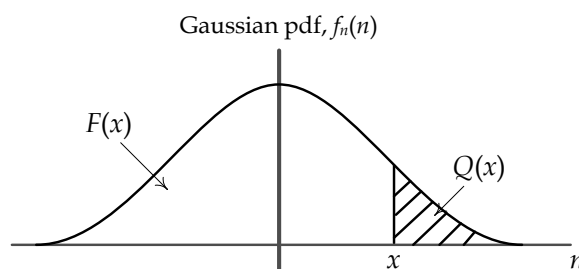
Recall that for a standard Gaussian random variable X with zero-mean and unity-variance, the *probability density function* (pdf) is:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

For the purpose of our performance analysis, we will define the *Quantile function* $Q(x)$ as the complement of the *cumulative distribution function* $F(x)$ of the standard Gaussian random variable, i.e.,

$$Q(x) = 1 - F(x) = 1 - \int_{-\infty}^x f(\alpha) d\alpha = \int_x^{\infty} f(\alpha) d\alpha = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{\alpha^2}{2}} d\alpha$$

The diagram below gives a visual representation for $Q(x)$ which represents the shaded area under the standard Gaussian density curve:

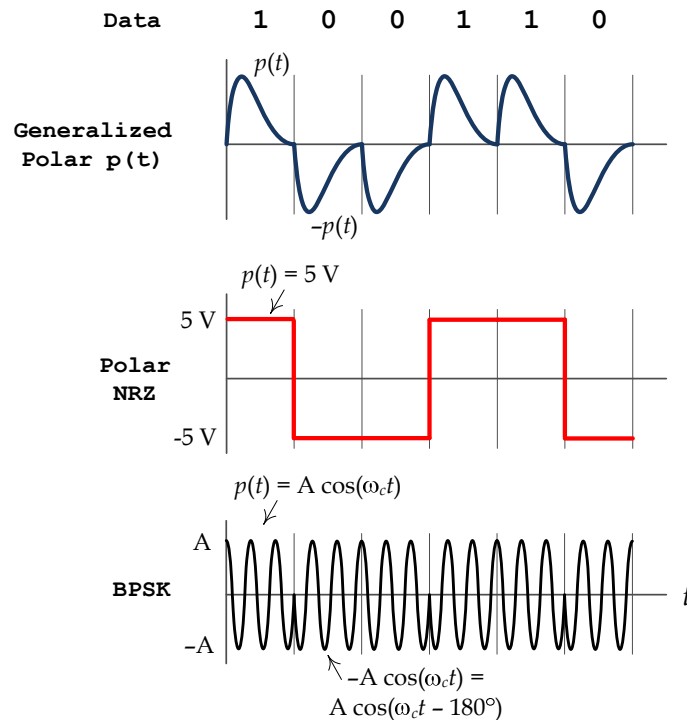


Usually we use a table (similar to the one shown below) to lookup $Q(x)$ values for specific x arguments since the above integral has no closed form solution.

x	$Q(x)$	x	$Q(x)$	x	$Q(x)$	x	$Q(x)$
2.00	2.28E-02	3.00	1.35E-03	4.00	3.17E-05	5.00	2.87E-07
2.05	2.02E-02	3.05	1.14E-03	4.05	2.56E-05	5.05	2.21E-07
2.10	1.79E-02	3.10	9.68E-04	4.10	2.07E-05	5.10	1.70E-07
2.15	1.58E-02	3.15	8.16E-04	4.15	1.66E-05	5.15	1.30E-07
2.20	1.39E-02	3.20	6.87E-04	4.20	1.33E-05	5.20	9.96E-08
2.25	1.22E-02	3.25	5.77E-04	4.25	1.07E-05	5.25	7.60E-08
2.30	1.07E-02	3.30	4.83E-04	4.30	8.54E-06	5.30	5.79E-08
2.35	9.39E-03	3.35	4.04E-04	4.35	6.81E-06	5.35	4.40E-08
2.40	8.20E-03	3.40	3.37E-04	4.40	5.41E-06	5.40	3.33E-08
2.45	7.14E-03	3.45	2.80E-04	4.45	4.29E-06	5.45	2.52E-08
2.50	6.21E-03	3.50	2.33E-04	4.50	3.40E-06	5.50	1.90E-08
2.55	5.39E-03	3.55	1.93E-04	4.55	2.68E-06	5.55	1.43E-08
2.60	4.66E-03	3.60	1.59E-04	4.60	2.11E-06	5.60	1.07E-08
2.65	4.02E-03	3.65	1.31E-04	4.65	1.66E-06	5.65	8.02E-09
2.70	3.47E-03	3.70	1.08E-04	4.70	1.30E-06	5.70	5.99E-09
2.75	2.98E-03	3.75	8.84E-05	4.75	1.02E-06	5.75	4.46E-09
2.80	2.56E-03	3.80	7.23E-05	4.80	7.93E-07	5.80	3.32E-09
2.85	2.19E-03	3.85	5.91E-05	4.85	6.17E-07	5.85	2.46E-09
2.90	1.87E-03	3.90	4.81E-05	4.90	4.79E-07	5.90	1.82E-09
2.95	1.59E-03	3.95	3.91E-05	4.95	3.71E-07	5.95	1.34E-09

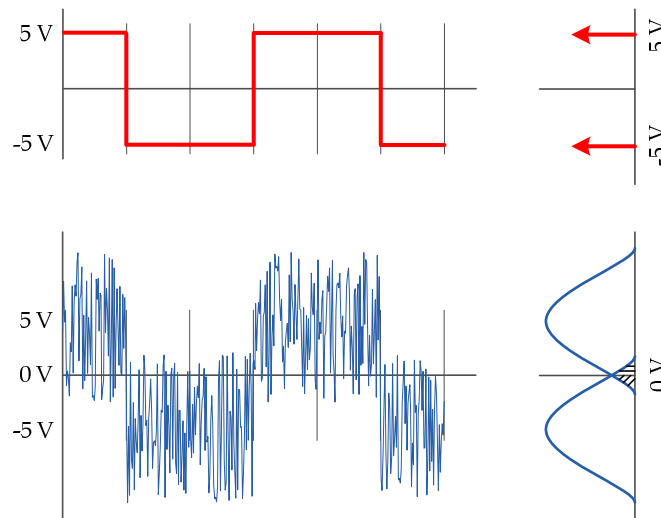
This is not part of the exam material ...

The analysis presented here is quite general and applies for both digital *baseband* and digital *modulated* signals. We will start with the special case of sending a pulse $p(t)$ when transmitting a **1** and sending the negative of the same pulse, i.e., $-p(t)$ when transmitting a **0**. Notice that this analysis is equally applicable to Polar NRZ signals (where $p(t) = \text{rect}(t)$) and to BPSK signals (where $p(t) = A \cos(\omega_c t)$). See the following figure.



We start by assuming that our receiver will sample the received pulse $p(t)$ in the middle of the pulse (also called symbol) period to read its amplitude. For example, if the transmitted signal is Polar NRZ, as shown above, the detected samples are supposed to be $A = +5\text{V}$ or $-A = -5\text{V}$, which decides whether the received bit is a logic **1** or a logic **0**.

However, because of the channel noise the detected samples will be $\pm A + n$, where n is a zero-mean normally (Gaussian) distributed random noise, as shown below. In *threshold detection*, and because of the symmetry of the situation, the preferred detection threshold is zero; that is, the received pulse is detected as a **1** or **0** depending on whether the sample value is positive or negative.



Notice that because of added noise, occasionally a transmitted **0** will be sampled as a positive voltage, causing the received **0** to be read as **1**. This means that an error can occur with a probability equal to the area of the Gaussian function corresponding to a transmitted **0** but falling above the zero-voltage threshold (see above figure). Hence, for Gaussian noise with zero-mean and variance σ^2 the probability of making an error in reading a **0** at the receiver is given by³:

$$\Pr[\text{error}|\mathbf{0}] = \Pr[n > A] = Q\left(\frac{A}{\sigma}\right)$$

And similarly,

$$\Pr[\text{error}|\mathbf{1}] = \Pr[n < -A] = Q\left(\frac{A}{\sigma}\right)$$

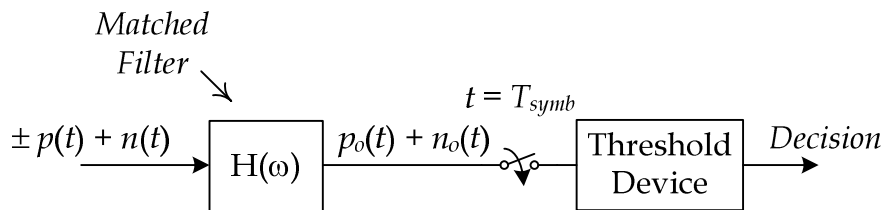
³ Notice that for a Gaussian random variable that has zero mean and variance σ^2 , the probability that the random variable exceeds a value x is $\int_x^\infty f(\alpha) d\alpha = \frac{1}{\sigma\sqrt{2\pi}} \int_x^\infty e^{-\frac{\alpha^2}{2\sigma^2}} d\alpha = \frac{1}{\sqrt{2\pi}} \int\left(\frac{x}{\sigma}\right) e^{-\frac{\alpha^2}{2}} d\alpha = Q\left(\frac{x}{\sigma}\right)$.

Hence, using conditional probabilities, we have:

$$\begin{aligned} \Pr[\text{error}] &= \sum_{\text{all symbols}} Q\left(\frac{\text{peak sample value}}{\text{rms of noise}}\right) \Pr[\text{symbol}] \\ &= \Pr[\text{error}|\mathbf{1}] \Pr[\mathbf{1}] + \Pr[\text{error}|\mathbf{0}] \Pr[\mathbf{0}] = Q\left(\frac{A}{\sigma}\right) \times \frac{1}{2} + Q\left(\frac{A}{\sigma}\right) \times \frac{1}{2} \\ &= Q\left(\frac{A}{\sigma}\right) \end{aligned}$$

It is important to notice, however, that for AWGN noise the above threshold detection is not optimal. For example, a better method is to sample the pulse $p(t)$ three times; if at least two out of three samples are positive, the received bit is considered $\mathbf{1}$. It turns out that the optimal scenario is to sample the pulse infinitely many times, which is *equivalent* to calculating the **energy** of the received symbol instead of just using one sample value to make the final decision.

To achieve this optimal behavior, receivers use a **matched filter** before sampling the output of the filter at the end of each symbol time (see the Figure below). The matched filter is simply a convolution of the received symbol with itself, resulting in the energy of the incoming symbol. The matched filter is represented by $H(\omega)$ in the Figure below.



To derive the mathematical equations at the receiver, we notice that instead of sampling the received symbol $p(t) + n(t)$, the receiver now samples the *output* of the matched filter, denoted by $p_o(t) + n_o(t)$. Of course, we know that in the frequency domain:

$$P_o(\omega) = P(\omega)H(\omega)$$

and,

$$p_o(t) = \mathcal{F}^{-1}\{P_o(\omega)\} = \mathcal{F}^{-1}\{P(\omega)H(\omega)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega)H(\omega)e^{j\omega t} d\omega$$

where $P(\omega)$ is the Fourier transform of the incoming symbol $p(t)$. Since the output of the matched filter is sampled at $t = T_{\text{symb}}$, the peak sample value is now:

$$A_o = p_o(T_{\text{symb}}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega)H(\omega)e^{j\omega T_{\text{symb}}} d\omega$$

But it is also important to realize that the AWGN noise signal $n(t)$ will also pass through the matched filter which means we get an output noise signal $n_o(t)$ with a power spectral density of:

$$S_{n_o}(\omega) = S_n(\omega)|H(\omega)|^2$$

Hence, the rms of output noise is:

$$\begin{aligned}\sigma_o &= \sqrt{\sigma_{n_o}^2} = \sqrt{n_o^2(t)} = \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} S_{n_o}(\omega) d\omega} = \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} S_n(\omega)|H(\omega)|^2 d\omega} \\ &= \sqrt{\frac{N_o}{4\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega}\end{aligned}$$

We explained earlier that the probability of error is given by the following equation:

$$\Pr[\text{error}] = \sum_{\text{all symbols}} Q\left(\frac{\text{peak sample value}}{\text{rms of noise}}\right) \Pr[\text{symbol}]$$

And since the quantile function $Q(x)$ decreases monotonically with x , our objective is to design the matched filter $H(\omega)$ such that the ratio A_o/σ_o is maximum. It can be shown that (see Chapter 10 in your textbook for the analysis), that the matched filter that maximizes the above ratio is:

$$H(\omega) = P(-\omega)e^{-j\omega T_{\text{symp}}}$$

If we use such a matched filter at the receiver (i.e., we use *optimal* threshold detection), then irrespective of the symbol type or shape, we will have:

$$\text{peak sample value} = A_o = \frac{1}{2\pi} \int_{-\infty}^{\infty} |P(\omega)|^2 d\omega = E_{s_i} = \text{energy in the } i^{\text{th}} \text{ symbol}$$

$$\text{rms of noise} = \sigma_o = \sqrt{\frac{N_o}{4\pi} \int_{-\infty}^{\infty} |P(\omega)|^2 d\omega} = \sqrt{\frac{N_o E_{s_i}}{2}}$$

Hence,

$$\begin{aligned}\text{BER} = \Pr[\text{error}] &= \sum_{\text{all symbols}} Q\left(\frac{\text{peak sample value of } i^{\text{th}} \text{ symbol}}{\text{rms of noise}}\right) \Pr[i^{\text{th}} \text{ symbol}] \\ &= \sum_{\text{all symbols}} Q\left(\frac{E_{s_i}}{\sqrt{\frac{N_o E_{s_i}}{2}}}\right) \Pr[i^{\text{th}} \text{ symbol}] = \sum_{\text{all symbols}} Q\left(\sqrt{\frac{2E_{s_i}}{N_o}}\right) \Pr[i^{\text{th}} \text{ symbol}] \\ &= Q\left(\sqrt{\frac{2 \sum E_{s_i} \Pr[i^{\text{th}} \text{ symbol}]}{N_o}}\right)\end{aligned}$$

If we use E_s to represent the *average* energy-per-transmitted-symbol (over all possible symbols in the modulated signal), we can write the above equation as follows:

$$BER = Q\left(\sqrt{\frac{2E_s}{N_0}}\right) = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$

The above analysis applies for other types of digital modulated signals, not just polar signals, but with slight variations. A summary of the BER equations for the different modulation techniques we discussed in this document is given below.

For the rest of this document, we will use the following notation:

- M = Number of possible symbols that the modulated signal can assume.
- k = the number of bits sent per transmitted symbol = $\log_2(M)$.
- E_s = Average energy-per-transmitted-symbol in the modulated signal (Joule).
- E_b = Average energy-per-transmitted-bit in the modulated signal (Joule) = E_s/k .
- $S_n(\omega) = \frac{N_0}{2}$ = Double-sided noise power spectral density (in W/Hz = Joule).
- T_o = Bit duration.
- $T_{symp} =$ Symbol duration = $k T_o$
- **BER** = Probability of bit-error = bit error rate.

Modulation with AWGN	Error Probability
ASK	$BER = Q\left(\sqrt{\frac{E_b}{N_0}}\right)$
FSK	$BER = Q\left(\sqrt{\frac{E_b}{N_0}}\right)$
BPSK	$BER = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$
QPSK	$BER = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$
PSK (order M)	$BER \cong \frac{2}{k} Q\left(\sqrt{2k \frac{E_b}{N_0} \times \sin\left(\frac{\pi}{M}\right)}\right)$
QAM (order M) (Rectangular QAM)	$P_{bc} = \frac{4}{k} \left(1 - \frac{1}{\sqrt{M}}\right) Q\left(\sqrt{\frac{3k E_b}{M-1 N_0}}\right)$ $BER = 1 - (1 - P_{bc})^2$

Example:

Find the BER for BPSK if we use an *optimal* detector (a matched filter). Assume the amplitude of the carrier is $A = 0.5$ V, data rate is 2 bps, and $N_0 = 2 \times 10^{-2}$ W/Hz.

Solution:

In BPSK there is one symbol per bit (i.e., a total of two symbols that the modulated signal can assume). The two symbols can be written as:

$$s_1 = A \cos(\omega_c t) \qquad s_2 = -A \cos(\omega_c t) = A \cos(\omega_c t - \pi)$$

The energy-per-symbol here is the same as the energy-per-bit and is equal for both possible symbols. Hence, its average is:

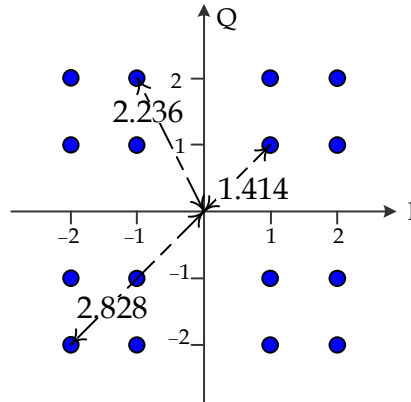
$$E_b = E_s = \left(\frac{A^2}{2} T_{\text{sympb}}\right) \Pr[\mathbf{1}] + \left(\frac{A^2}{2} T_{\text{sympb}}\right) \Pr[\mathbf{0}] = \frac{A^2}{2} T_{\text{sympb}} = \frac{A^2}{2} T_o = \frac{A^2}{2} \frac{1}{f_0}$$

Hence,

$$\begin{aligned} \text{BER} &= Q\left(\sqrt{\frac{2E_b}{N_0}}\right) = Q\left(\sqrt{\frac{A^2}{N_0 f_0}}\right) = Q\left(\sqrt{\frac{0.5^2}{2 \times 10^{-2} \times 2}}\right) = Q(\sqrt{6.25}) = Q(2.5) \\ &= 6.21 \times 10^{-3} \end{aligned}$$

Example:

Find the BER for the 16-QAM constellation shown below if we use an *optimal* detector (a matched filter). Assume the data rate is 4 bps, and $N_0 = 5 \times 10^{-2}$ W/Hz.



Solution:

In this system there are 16 possible symbols, which we assume to be equally probable, i.e., each occurs with a probability of 1/16. Hence, the energy-per-symbol is:

$$E_s = \left(\frac{1.414^2}{2} T_{\text{sympb}}\right) \left(\frac{4}{16}\right) + \left(\frac{2.236^2}{2} T_{\text{sympb}}\right) \left(\frac{8}{16}\right) + \left(\frac{2.828^2}{2} T_{\text{sympb}}\right) \left(\frac{4}{16}\right)$$

$$E_s = [0.25 + 1.25 + 1] T_{\text{sympb}} = 2.5(T_{\text{sympb}})$$

$$E_b = \frac{E_s}{k} = 2.5 \left(\frac{T_{\text{sympb}}}{k}\right) = 2.5(T_0) = \frac{2.5}{f_0} = \frac{2.5}{4} = 0.625 \text{ J}$$

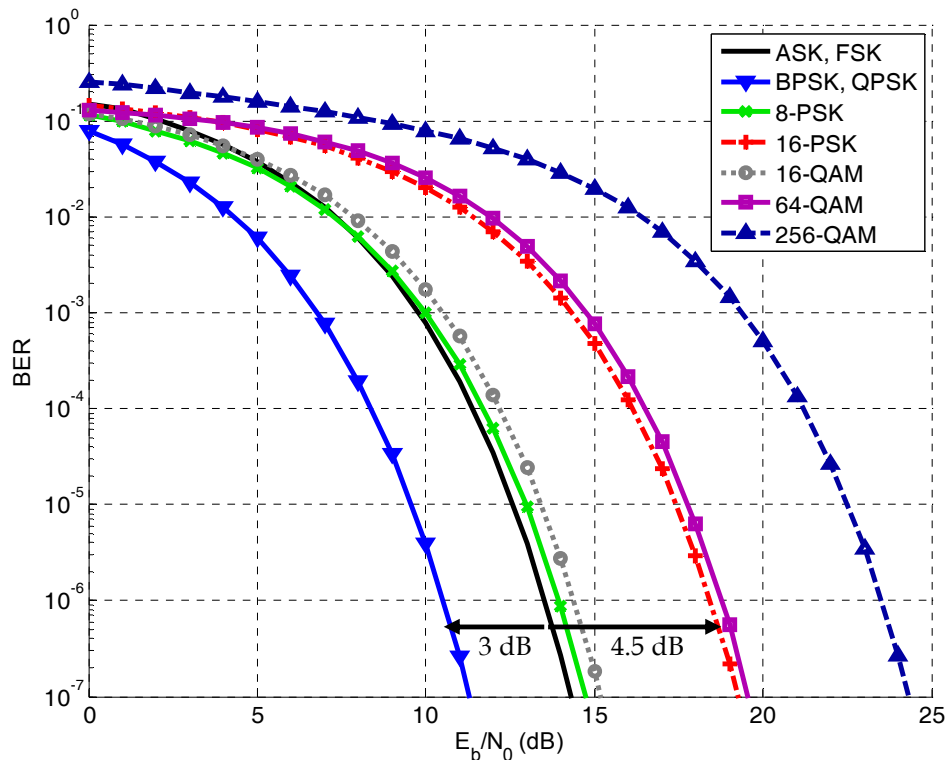
$$P_{bc} = \frac{4}{k} \left(1 - \frac{1}{\sqrt{M}}\right) Q \left(\sqrt{\frac{3k}{M-1} \frac{E_b}{N_0}} \right) = \frac{4}{4} \left(1 - \frac{1}{\sqrt{16}}\right) Q \left(\sqrt{\frac{3 \times 4}{16-1} \times \frac{0.625}{0.05}} \right) = \frac{3}{4} Q(\sqrt{10})$$

$$= \frac{3}{4} Q(3.162) = \frac{3}{4} \times 8 \times 10^{-4} = 6 \times 10^{-4}$$

$$BER = 1 - (1 - P_{bc})^2 = 1 - (1 - 6 \times 10^{-4})^2 = 1.2 \times 10^{-3}$$

Comparison of Digital Modulation Schemes

We show below the **BER curves** for the different **digital modulation schemes**:



Comparing BPSK and QPSK with ASK and FSK, we notice that BPSK and QPSK provide smaller bit error rate for the same E_b/N_0 . In other words, for the same bit error rate, we need less signal-to-noise ratio (E_b/N_0) to send BPSK and QPSK. This means that BPSK and QPSK have better immunity to noise than ASK and FSK.

Notice also that the performance of BPSK is the same as that for QPSK, while the performance of 8-PSK and 16-PSK are worse (i.e., they require more signal-to-noise ratio to achieve the same bit error rate). This is an expected result because 8-PSK and 16-PSK have more constellation diagram points (which are now closer and closer to each other).

Also notice how 16-QAM has a superior performance compared to 16-PSK, which is to be expected because the constellation points are further apart in 16-QAM compared to 16-PSK.

The following table shows the bandwidth requirements and the necessary signal-to-noise ratio (E_b/N_0) to achieve near error free transmission (this is $BER \approx 10^{-6}$). Notice that for higher order modulation techniques, we require less bandwidth but we need more signal-to-noise ratio (E_b/N_0) to maintain small bit error rate (i.e., to maintain good performance).

Modulation	Bandwidth	Error free E_b/N_0 (i.e., $BER < 10^{-6}$)
ASK	$2f_o$	13.5 dB
FSK	$2\Delta f + 2B = 2f_o(\beta + 1)$	13.5 dB
BPSK	$2 \times \text{Baud} = 2f_o$	10.5 dB
QPSK	$2 \times \text{Baud} = f_o$	10.5 dB
8-PSK	$2 \times \text{Baud} = 2f_o/3$	14 dB
16-PSK	$2 \times \text{Baud} = f_o/2$	18 dB
16-QAM	$2 \times \text{Baud} = f_o/2$	14.5 dB
64-QAM	$2 \times \text{Baud} = f_o/3$	18.5 dB
256-QAM	$2 \times \text{Baud} = f_o/4$	23.4 dB

Remember this for digital modulation:

The available **bandwidth** of the channel decides the **baud rate** (symbols per second) you can send.

The available **signal-to-noise ratio** (E_b/N_0) decides the level of modulation you can use while still maintaining a small bit error rate (good quality). In other words, it decides the **number of bits you can send per symbol**.

Hence, the two factors together (bandwidth and SNR) decide the total bit rate you can achieve over any single channel. Now compare this to Shannon's Limit!!

Applications of digital modulation techniques:

The following are some current-day communication systems that use digital modulation:

IEEE 802.11 (Wi-Fi): A very important Wireless Local Area Networking technology. Since Wi-Fi has many variants, it uses different modulation techniques such as: BPSK, QPSK, 16-QAM, 64-QAM and CCK (Complementary Code Keying) (CCK is an extension of QPSK).

IEEE 802.16 (Wi-MAX): A very important Wireless Metropolitan Area Network, and currently competes with ADSL for Internet delivery. Wi-MAX switches dynamically between different modulation schemes such as: BPSK, QPSK, 16-QAM, and 64-QAM. It uses these modulation schemes in combination with OFDM (Orthogonal Frequency division multiplexing) (OFDM is an extension of FDM).

DVB (Digital Video Broadcasting): This is the European standard for digital television broadcasting. There are many variants within the standard: DVB-S (for satellite broadcasting) uses QPSK or 8-PSK; DVB-C (for cable) uses 16-QAM, 32-QAM, 64-QAM, 128-QAM or 256-QAM; and DVB-T (for terrestrial television broadcasting) uses 16-QAM or 64-QAM.

DAB (Digital Audio Broadcasting): Future European standard for digital radio broadcasting, which should replace AM and FM radio broadcasting. DAB use DQPSK (Differential QPSK) (DQPSK is a variation of QPSK).

ADSL: Currently one of the main choices for connecting to the Internet. Uses adaptive QAM in a scheme called DMT (Discrete Multi-Tone modulation).

Shannon's Limit:

You see by now that the two main resources in communication systems are the **channel bandwidth** and the **transmitted power** (or SNR). In a given communication channel, one resource may be more valuable than the other, and the communication scheme should be designed accordingly.

The limitation imposed on communication by the channel bandwidth and the SNR is dramatically highlighted by **Shannon's noisy channel theorem**, which applies for channels contaminated with Additive White Gaussian Noise (**AWGN**) noise. Shannon's equation states that:

$$C = B \log_2 (1 + \text{SNR}) \quad \text{bits/second}$$

where C is the rate of information transmission in bits per second, B is the channel bandwidth (in Hz), and SNR is the signal-to-noise ratio (unitless) on the channel.

This rate C (known as the **channel capacity**) is the **maximum** number of binary bits that can be transmitted per second with a probability of error arbitrarily close to zero. In other words, it is impossible to transmit at a rate higher than C without incurring errors. Shannon's equation clearly brings out the limitation on the rate of communication **imposed by B and SNR**. This is why bandwidth and SNR are the main quantities that we study in communication systems. Notice that if noise was zero (i.e., $\text{SNR} = \infty$), or the bandwidth was infinite ($B = \infty$), then we can transmit infinity information and communication would cease to be a problem.

It should be remembered that Shannon's result represents the upper limit on the rate of communication over a channel and can be achieved only with a systems of **great complexity and a time delay in reception approaching infinity**. Practical systems operate at rates **below the Shannon rate**.