Intuitive Guide to Principles of Communications
www.complextoreal.com

All About Modulation - Part II

The main Figure of Merit for measuring the quality of digital signals is called the Bit Error Rate (BER) vs. E_b/N_0. Bit error rate is the ratio of number of bits received in error vs. total number of bits sent.

When we talk about bit errors, we need to distinguish between two types of signals.

1. A coded signal - This is a signal that has coding applied to it at the source so that only some of the bits in the stream are important. We can tolerate a higher gross error rate in this channel, as long we can decode the information bits with much better integrity. The BER of the information bits is much less than that of the coded signal.

   $BER_{\text{info bits}} << BER_{\text{coded bits}}$

   If Eb_c/N_0 is defined as the bit energy of this signal, then E_b/N_0 of the information bits is approximately

   $E_b/N_0 = E_{bc}/N_0$ - Code rate (in dBs)

   A coded bit stream may have a bit error rate of 10^{-1} to 10^{-3}, but when coding is applied, the rate experienced by the information bits is quite small, on the order of 10^{-4} to 10^{-11}.

2. An uncoded signal - There is no coding so all bits are equally important and the stated BER is the BER of the information bits, and E_b/N_0 is equal to E_{bc}/N_0.

Since most signals are coded, we need to keep in mind which BER we are talking about. The BER of the received coded signal or the BER of the information bits? Normally even that distinction is not there except in simulation and analysis because signals are soft-decision demodulated and this soft-decision is handed to the coding decoder for the final bit decoding.

For example, we have two QPSK links, of bit rate R_b, one with coding of rate one-half and the other without. Both operate at an E_b/N_0 of 7 dB. Both have a gross BER of 10-3. Bit the one that is coded will ultimately give a BER of 10-10 on the information bits. The only difference is that with link 1 we get full R_b bits of information whereas with the coded link the price of purity is that only half of the bits are information bits.

Bit error statistics, although a discrete phenomena, are generally described by continuous functions. In absence of unusual sources of interference, rf signals are subject primarily to Gaussian random errors. The Gaussian noise environment is coming from the hardware and from the air/medium in which the signal is traveling and these errors are generally random, coming in one's, and two's. A large number of consecutive errors in a Gaussian environment are less likely. A lot of errors at once, called burst errors, are usually characterized as coming from fading, shadowing and Doppler effects.

Errors don't actually occur, but are made by the receiver. Prior to the receiver, the signal is in an analog waveform and in that context bit errors have no meaning. How errors are made in the receiver is a function of modulation. Modulation determines how tightly various symbols are packed in a signal space and how sensitive the receiver must be in making a decision.

Let's begin by looking at the transmission of a polar NRZ signal through a communications channel. The signal has amplitude of ±1 volt, where 1v represents a 1 bit and a -1 v represents a 0 bit.

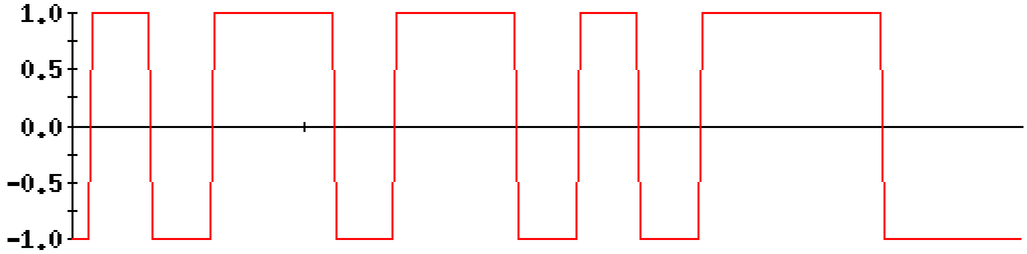


Figure 1 - Transmitted binary signal

The signal picks up noise along the way and at the receiver we see instead of ±1 volts, amplitudes that include noise, sometimes higher than transmitted value, sometimes lower such that

Received signal amplitude = 1 + n, for 1 bit = $s_1(t)$
Received signal amplitude = -1 + n, for 0 bit = $s_0(t)$

where n represents the noise amplitude and is changing from sample to sample. (We are sampling each bit more than once.)

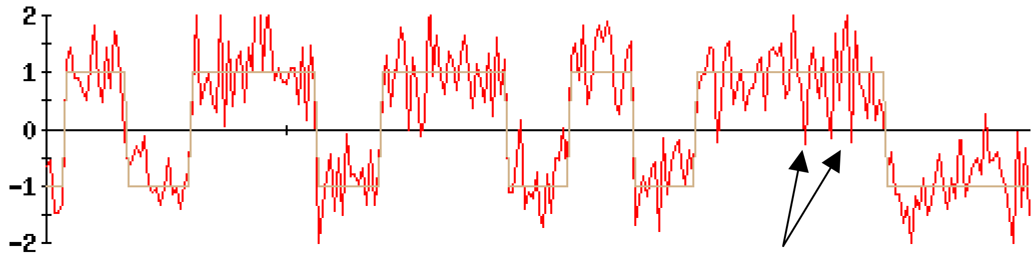


Figure 2 - The binary signal has been acted upon by random noise. The received amplitude now varies at each sample.

Making decision based on the fluctuating values of Figure 2 is like trying to read a tossed coin in a room where with each toss, the light level goes up or down randomly. Sometimes, the light is bright enough and you can clearly read the coin and other times, it's too dark to read correctly. Each toss will be either a head or a tail, and equal number of heads and tails are expected. The fact that an equal number of heads and tails are expected is called the base probability of the process. But how we read the coin (or decode) depends on how bright the light was at the moment of the toss.

The fact that there are just two choices to pick between (H or T) is akin to symbol choices of a modulation. A modulation with fewer choices will lead to smaller errors, this is obvious.

The value of noise, n varies with time and so is different from sample to sample. It is assumed to be normally distributed with a zero mean and variance of quantity σ^2, where σ is the standard deviation of the noise process.

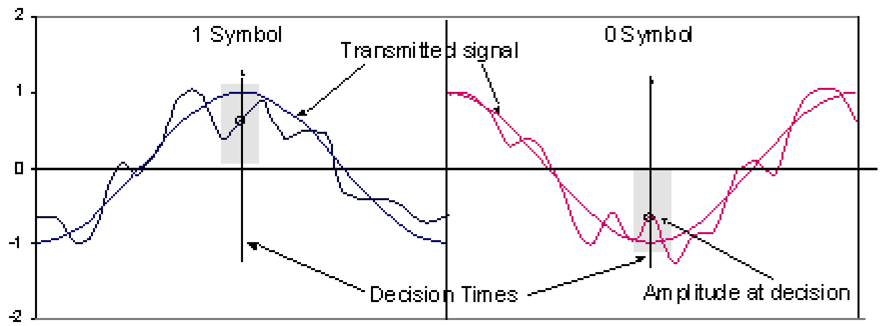


Figure 3 – Effect of noise on the symbol

In Figure 3, we see two symbols, a 0 and 1 symbol with the transmitted shape and result of the addition of random noise at the receiver.

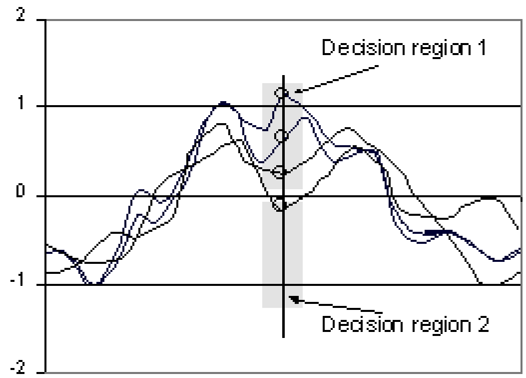


Figure 4 – Symbols arrives in various states of distortion at the receiver. By sampling the amplitude at the decision instant, the receiver makes a decision about what was sent. This decision is made strictly based on the received amplitude.

In Figure 4, we superimpose several received symbols for bit 1. We can see that the received amplitudes at the sampling instants vary. For most samples, the receiver would make a correct decision, because the added noise does not cause signal voltage to change polarity, but there is one instance where noise is large enough that a positive transmitted voltage became a negative voltage in the eyes of the receiver. The receiver will make an error for this bit.

The receiver makes decisions by the following rules;

If observed voltage at the receiver is > **0**, then transmitted bit was 1 bit.
If observed voltage at the receiver is < **0**, then transmitted bit was 0 bit.

We call this process "Threshold Comparison." The selection of 0 volts as a threshold is called the decision region boundary or Threshold, T. A threshold is like a net or a line between the players in a game. The threshold is optimal when it is at the mean of the two values we are trying to distinguish between. However this is true only if the apriori probability of the two outcomes is equal. In a normal communication signal, the number of 0's sent (represented by a -1v signal level) is approximately equal to number of 1's sent, so the optimum threshold is at 0 volts (average of -1v and +1v).

If at the time of decision, the value of noise n exceeds V volts, the transmitted level, then the observed signal will flip from the correct decision region and a wrong decision will be made. We can write this probability of error as

$$P(e|1) = P(n < -V) \qquad 1$$

Read this as: the probability of making a decision error when a 1 is sent is the probability that noise at this moment is less than -V. It is obvious that the noise present must be negative in order to pull the positive voltage value into the negative region. The converse holds for the 0 bit.

$$P(e|0) = P(n > V) \qquad 2$$

By Bayes Rule of probability, which states that the overall rate of making a bad decision is

$$P(e) = P(e|0)P(0) + P(e|1)P(1) \qquad 3$$

This is called the conditional probability. The probability of making a decision error comes from making errors in interpreting both 0 and 1 bit. If equal number of 1s and 0s are sent (0s and 1s are equi-probable) then we can rewrite this equation as

$$P(e) = \frac{1}{2}P(e|0) + \frac{1}{2}P(e|1) \qquad 4$$

To evaluate this expression, we need to calculate error probabilities $P(e|0), P(e|1)$ which maybe assumed to be equal when the threshold is exactly in the middle.

Gaussian distribution and density

We know intuitively that probability of an event is defined to be between 0 and 1. The mathematical definition of a Probability Density Function, f(x), is that it satisfies the following properties.

1. The probability that x, which can be any real quantity, positive or negative, is between two points a and b is

$$P(a \le x \le b) = \int_a^b f(x)\,dx \qquad 5$$

2. This probability is a positive number for all real x.

3. The area under the Probability Density Function is 1.

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1 \qquad 6$$

The variable x can be discrete or continuous. Height of a person is continuous, dice throws are discrete, as well as decoding of a 1 or 0 is considered discrete but the received voltage may be thought of as a continuous variable.

The **Probability Density Function (PDF)**, if you remember from your statistics class, is a histogram. The y axis is the ration of desired events in the universe of events that take on the x value. For a Gaussian process, the probability density function of Figure 5 is given by the expression

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-V)^2/2\sigma^2} \qquad 7$$

where V is the average value and σ^2 is the variance of the noise process.

Two different distributions are shown below, 1. for height of women and 2. height of men in inches. Since the variance is different for them, the height distribution for men spreads out further.

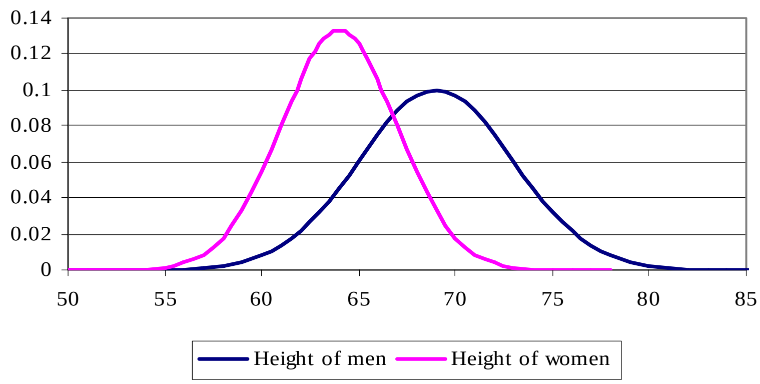


Figure 5 - Gaussian Probability Density Function for variable x, the height of men and women in inches.

If we integrate the PDF, we get its area. This integrated graph is called, the **Cumulative Probability Function**, CDF is a very useful graph, in fact far more than the PDF. This is the graph we use to answer all kinds of useful questions such as about BER of signals.

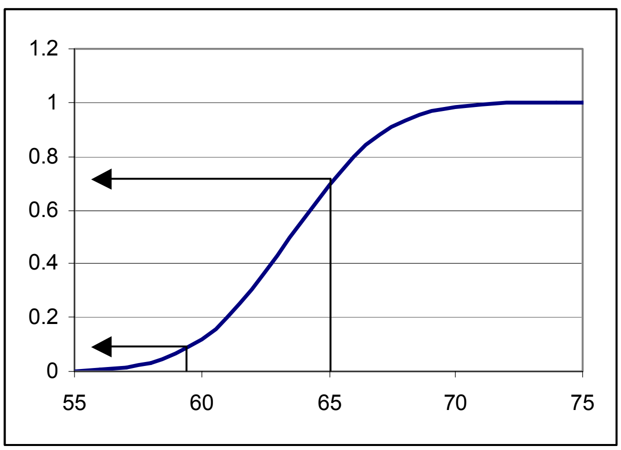


Figure 6 - Cumulative Probability Function - Integral of the probability density function

Question: We ask, what is the probability that a woman is between 4' 11"' and 5'5" tall? By using the CDF in Figure 5 we can answer this question with ease. This is the difference between the two y-axis values, one for 5'5" and the other 4'11".

$$P(4'11" \le x \le 5'5") = F(5'5") - F(4'11") \approx .65 \text{ or } (65\%)$$

The Cumulative probability distribution of Figure 6 is given by

$$F(x) = \int_{-\infty}^{x} f(x)\, dx \qquad 8$$

Integrating the PDF is difficult. The erf function which is approximated by Eq. 10 can be used to develop the CDF. Here μ is the mean value of the function, and σ the standard deviation.

$$F(x)=\frac{1}{2}\left[1+erf\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right]=\frac{1}{2}erfc\left(\frac{\mu-x}{\sqrt{2}\sigma}\right) \qquad 9$$

The Error function is given by (an approximation)

$$erf(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-n^2}\,dn \qquad 10$$

The complementary error function, erfc function is equal to

$$erfc(x)=1-erf(x) \qquad 11$$

The cumulative probability function is always positive since probability is always positive. This is intuitive. So are the erf and the erfc functions shown below. We can however shift them around and normalize them to any mean value. The F(x) function shown below is for mean value = 0.

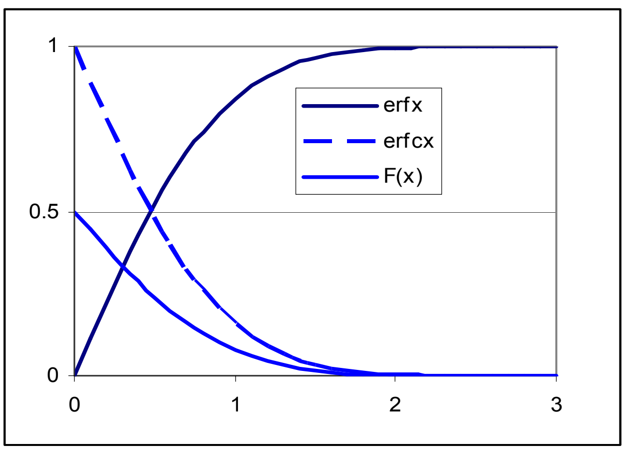


Figure 7 – Plotting erf(x), erfc(x) and the Cumulative probability distribution F(x)

Bit Error Measurement

To apply these ideas to determination of Bit Error rate, we start with the simplest of all modulations, BPSK. It has just two symbols, one for 0 and another for 1 bit. A simple link consists of a binary signal generator, a PSK modulator which multiplies the binary signal with a carrier which then picks up noise. The signals is then down converted to low-pass but the noise remains. The signal is filtered by a filter called "Integrate and Dump" to give us the waveforms shown below.

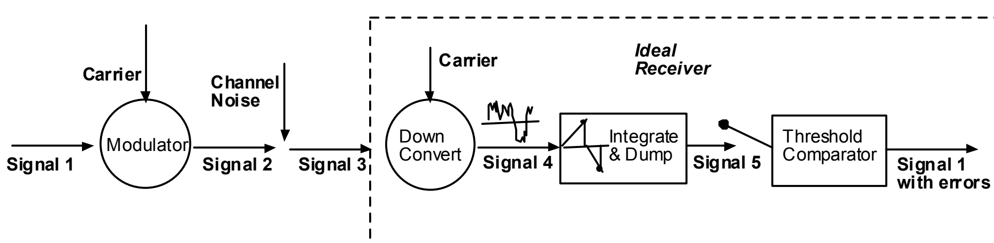
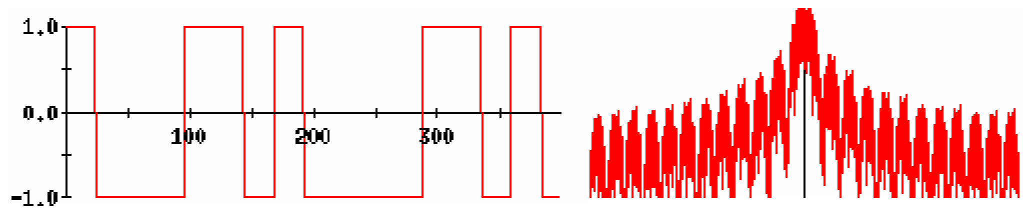
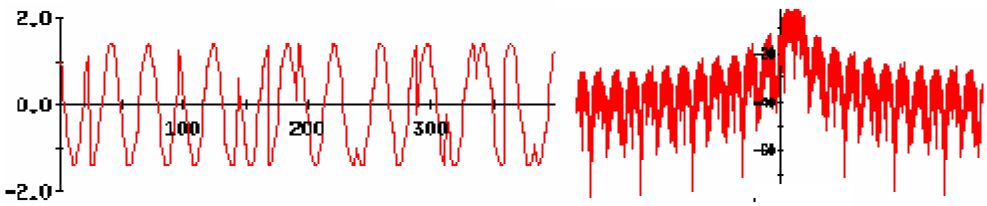


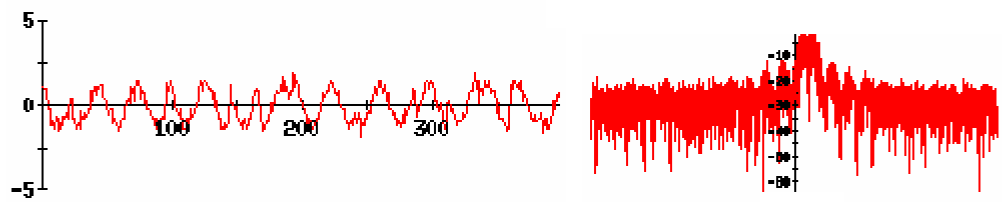
Figure 8 – A BPSK link



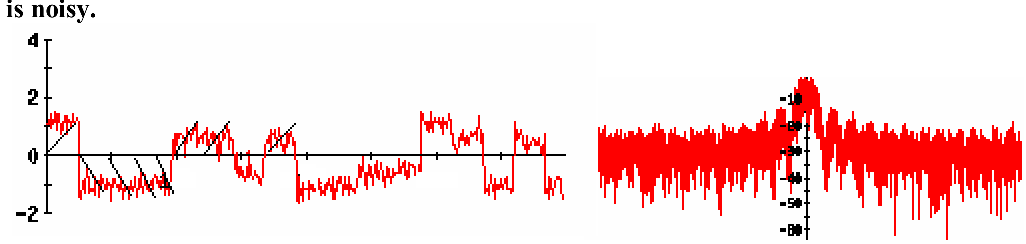
(a) Signal 1 - Baseband binary signal and its spectrum , 1 0 0 0 0 1 1 1 0 1 0 0 0 0



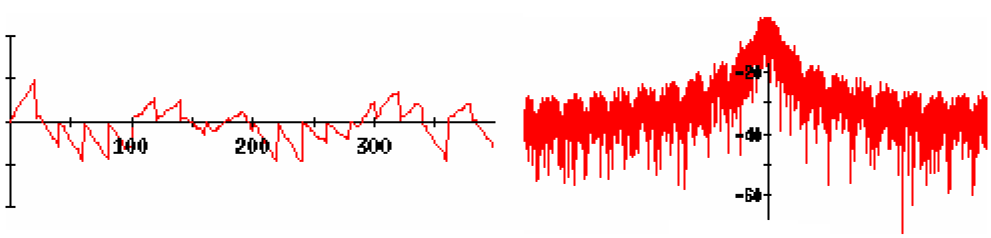
(b) Signal 2 - Modulated signal and its spectrum. Note that spectrum has shifted to carrier frequency.



(c) Signal 3 - Modulated signal with noise. The signal shape is the same as Signal 2 and the spectrum is noisy.



(d) Signal 4 - Converted signal at the front end of receiver, same as Signal 1 but has noise.



(e) Signal 5 - Output of Integrate and Dump filter. Bit decision is made at then edge of the symbol period based on accumulated value.
Figure 9 – Signal transformation through the link

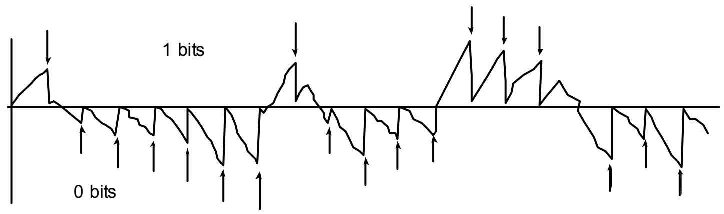


Figure 10 - Threshold Comparator output = 1 0 0 0 0 0 0 1 0 0 0 0 1 1 1 0 0 0 , Underlined bits received in error.

The Integrate and Dump filter works by adding up the individual voltage values at each sample during the bit. In that it is an integrator. At the end of the bit period it resets itself to start anew for the next bit. The sum of all the values give a triangle-like shape and bit decision is made based on the last value.

This basic link assumes that this system has infinite bandwidth and has no other distortions such as inter-symbol interference, non-linearities, etc. And most importantly the noise is purely Gaussian, or AWGN, Additive White Gaussian Noise, and this is a very common assumption in communications.

Below we draw two events; one of transmitting a signal of 1 volt peak amplitude and the other a signal of - The figures shows three cases. In each case the nominal voltages are +1 and -1 volts but the variance (the amount of noise) is different. In the first case it is 1, the second case it is .5 and third case .1 v. The pictures show the distribution of received voltages for each of the two signals.

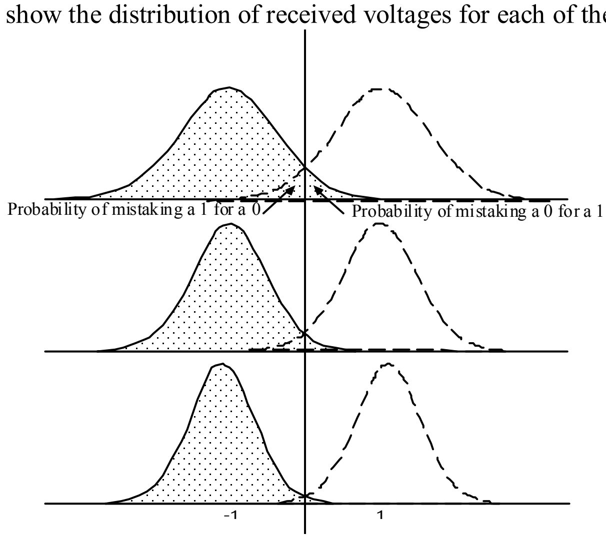


Figure 11 - The received voltage of two signals of 1 and -1 volts are distributed around the mean value with variance equal to the noise variance of the channel.

Even though all three cases have the same mean values, they overlap by a different amount. The amount of overlap is a function of how much noise is present. The larger the noise variance, the larger the overlapping and intuitively that means a larger probability of error.

Computing the Bit Error Probability

How do we figure out the probability of being in this overlapped area?
Let's take a binary signal, where a 0 is sent by signal of amplitude V_0 and a 1 with amplitude of V_1.

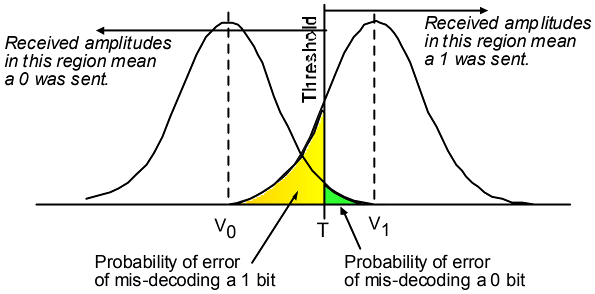


Figure 12 – Two signals with an arbitrary threshold

s_1 is the signal received.(It is function of time t, but I am leaving out the index to make it easier to read.) It includes noise, n (also function of time, t.) so that it is really equal to the original amplitude V_1 plus noise.

$$s_1(t) = V_1 + n(t) \qquad 12$$

and the noise is given by

$$n = s_1 - V_1$$

Similarly for a 0 bit, the received signal is a sum of the original amplitude of V_0 plus noise.

$$s_0 = V_0 + n$$
$$n = s_0 - V_0 \qquad 13$$

We set the decision boundary at an arbitrary point between the two signals, at amplitude T.

A decision error is made when noise at the sampling instant exceeds one-half the difference between the two discriminating amplitudes. It is also obvious we want this difference to be as large as possible to minimize errors. The BER, we will see shortly is a function of this "distance" between the signals and we want this distance large. Signals like countries want as few neighbors and as far away as possible! This noise level is given by

$$|n| \geq \frac{V_1 - V_0}{2} \qquad 14$$

If V_1 = 1 v and V_0 = -1 v, then the noise level that will cause error is greater than 1. Two different errors can happen, one for mistaking a 1 for a 0 and vice versa. The probability of each is the area under the curve starting at the threshold, T and going to infinity in the opposite direction as we can see in Figure 10.

Let's write the expression for this error. We recognize that the received voltages are randomly distributed with mean equal to the transmit voltage and variance equal to noise power. We use the probability density function (Eq. 7) of a Gaussian variable as our starting point with mean of the process equal to V_0 and V_1 and s_0 an s_1 being the received random variables. The variance of the process is σ.

$$P(e|1) = \int_{-\infty}^{T} \frac{1}{\sigma\sqrt{2\pi}} e^{-(s_1 - V_1)^2 / 2\sigma^2} ds_1 \qquad 15$$

This is equal to the shaded area shown on the left in Figure 12. Similarly, the probability of error when a 0 is sent is, shown in right shaded in Figure 12.

$$P(e|0) = \int_T^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(s_0 - V_0)^2 / 2\sigma^2} ds_0 \qquad 16$$

The total probability of both events, using Eq.7 is

$$P(e) = \frac{1}{2}\int_{-\infty}^{T} \frac{1}{\sigma\sqrt{2\pi}} e^{-(s_1 - V_1)^2 / 2\sigma^2} ds_1 + \frac{1}{2}\int_{-\infty}^{T} \frac{1}{\sigma\sqrt{2\pi}} e^{-(s_0 - V_0)^2 / 2\sigma^2} ds_0$$

17

Without making it any more complicated, let me say that when we take the differential of the above equations and set it to zero, we obtain the following value for the optimum threshold, T.

$$Threshold = \frac{V_1 + V_0}{2} \qquad 18$$

Makes sense! This is the average value, assuming that both 0s and 1s are equiprobable as they usually are. Now substitute the value of the threshold T, into the Eq. 17, and using Eq. 9 as a guide, we can write Eq. 17 in at least shorter if not totally friendly form as

$$P(e) = \frac{1}{2} erfc\left(\frac{V_1 - V_0}{2\sqrt{2}\sigma}\right) \qquad 19$$

This can also be expressed in the alternate Q function format, which is equal to the erfc function by

$$erfc(x) = 2Q(\sqrt{2}x) \qquad 20$$

Q function is often used in books but since erf and erfc tabulations are more readily available in Excel and Matlab, Q function is used less often. Here is the above BER equation in Q format.

$$P(e) = Q\left(\frac{V_1 - V_0}{2\sigma}\right) \qquad 21$$

We can further manipulate this equation by noting that we can write this equation as

$$P(e) = \frac{1}{2} erfc\left(\frac{A}{2\sqrt{2}\sigma}\right) \qquad 22$$

where

$$A = V_1 - V_0$$

By Parsevals' Theorem, we determine the power of this difference signal as

$$
\begin{aligned}
A^2 &= \int_{-\infty}^{\infty} (V_1 - V_0)^2 dt \\
&= \left[\int_{-\infty}^{\infty} (V_1)^2 dt + \int_{-\infty}^{\infty} (V_0)^2 dt - 2\int_{-\infty}^{\infty} (V_1 V_0)\ dt \right] \\
&= E_1 + E_0 - 2\sqrt{E_1 E_0}\sigma_{10}
\end{aligned}
\qquad 23
$$

We note that bit energy is defined as the product of bit time with the square of the instantaneous amplitude squared.

$$E_0 = \int_{-\infty}^{\infty} (V_0)^2 dt \qquad 24$$

To deal with the last term in Eq, 23, we define correlation factor ρ, as

$$\rho_{10} = \frac{1}{\sqrt{E_0 E_1}} \int_{-\infty}^{\infty} V_1 V_0\ dt \qquad 25$$

Now rewrite the Eq. 23 as

$$P(e) = \frac{1}{2} erfc\left(\sqrt{\frac{E_1 + E_0 - 2\sqrt{E_1 E_0}\rho_{10}}{2N_0}} \right)$$

$$where \qquad 26$$

$$N_0 = \frac{\sigma^2}{2}$$

The average bit energy of a signal is the average energy of all its individual bit energies. In this binary case, we have

$$E_{b(avg)} = \frac{E_1 + E_0}{2} \qquad 27$$

Now set $\rho_{10} = -1$ because this will give us the conditions for the lowest BER.

$$P(e) = \frac{1}{2} erfc\left(\sqrt{\frac{E_b(1-R\ )}{2N_0}} \right) \qquad 28$$

where

$$R = \frac{2\sqrt{E_1 E_0}}{E_1 + E_0}\rho_{10} \qquad 29$$

The value of R is lowest when $E_1 = E_0$ and $\rho_{10} = -1$, all these conditions are true for BPSK.

So now we get the easy to use (Ha!) BER equation for a binary BPSK signal with unshaped pulses (unlimited bandwidth), no distortions and Gaussian noise environment using an "optimum receiver". We know that BPSK satisfies the important condition that its $\rho = -1$, since cos wt and sin wt, the two pulses used are exactly anti-podal.

$$P(e) \quad = \quad \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{N_0}}\right) \qquad 30$$

Since erfc is a monotonically decreasing function, we conclude

1. The bit error rate experienced by the link is proportional to the "distance" or dissimilarity between the symbol.

2. The most dissimilar signals are BPSK because we use a sine and a cosine to represent the symbols. Hence BPSK BER forms the limit of lowest BER for a signal.

3. All other modulations have a ρ which is larger than -1 and so their BERs must also be larger than BPSK.

4. The bit error rate is proportional to the variance of the noise. Large variance means larger BER.

Bit error of a QPSK signal

A QPSK signal can be seen as two independent BPSK signal, on I and Q channels. If P_{eI} is the probability of error of the I channel (computed by Eq. 31) and P_{eQ} is the probability of errors of the Q channels, then we can write the probability of QPSK symbol error (that is either the I or Q or both channels make an incorrect decision, causing the QPSK symbol to be mis-decoded.)

$$P(e) = P_{eI} + P_{eQ} - \underline{P_{eI} P_{eQ}} \qquad 31$$

Assuming that these probabilities are small, and both channels have same error rate, we can ignore the last term.

$$P(e)_{QPSK} \approx 2P_{eI} = 2P_{eQ}$$

This says that the QPSK symbol error rate is twice that of BPSK. Now write the error probability of the QPSK.

$$If \quad P(e)_{sym-BPSK} = \frac{1}{2} erfc\sqrt{\frac{E_s}{N_0}}$$

$$then \quad P(e)_{sym-QPSK} = erfc\sqrt{\frac{E_s}{N_0}} \qquad 32$$

Note that although I have been kind of loose with my terms, mentioning bits and symbols in the same breath, I need now to explicitly state that the error rate we computed for BPSK case was the error rate of its symbol and not bit. But since for BPSK, a symbol is equal to just 1 bit, it causes no errors. That's no longer true for QPSK which has 2 bits per symbol. An error in a symbol can mean 1 or 2 errors. So the above calculation is for a QPSK symbol. To compute the bit error rate, need to consider Gray coding.

Bit and Symbol error rates for QPSK

Example: Imagine two independent BPSK streams. We transmit 1000 bits and see two errors in one and three in the other. The bit errors rate individually is .002 and .003 and is .0025 on the average.

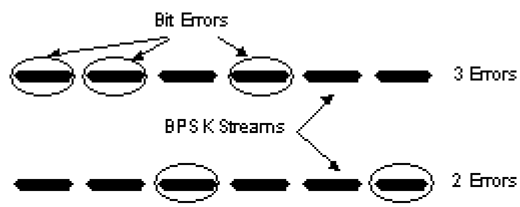


Figure 13 – QPSK can be represented by two independent BPSK streams. Its BER is the sum of bit errors of both streams vs. total bits.

Now imagine the same thing but as a QPSK signal, which is what it is when you have two BPSK streams together.

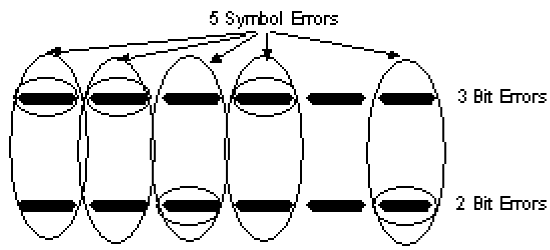


Figure 14 – Each bit error represents a symbol mistake in QPSK.

Each time a bit is decoded incorrectly, the whole symbol is wrong. Now we have 5 symbol errors.

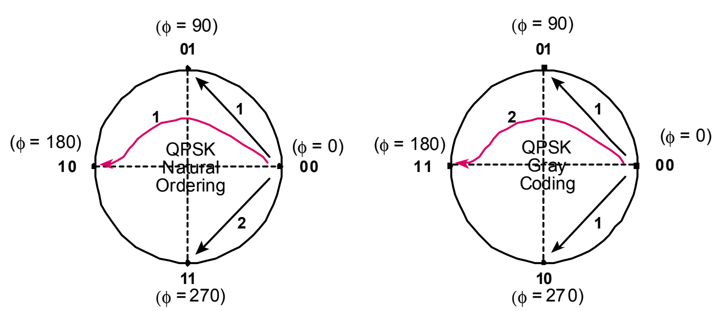


Figure 15 a – On the average if a symbol is mistaken, there will be 1.33 bit errors. If we consider only the adjacent symbol errors, (ignore the red link) then on the average, there will be 1.5 errors per symbol for a QPSK that is numbered according to natural numbering and one bit per symbol for Gray coded QPSK.

If Symbol 1 is mis-decoded, then we can have on the average 1+1+2/3 =1.33 bits in error per miscoded symbol If we ignore the possibility that symbol 1 will be mistaken for symbol 3 – they are 180 degrees apart in phase – the average number of bits decoded incorrectly is 1.5.

But we can reduce this BER by utilizing Gray coding, which will reduce average miscoded bits down to one per symbol.

If we organized the symbols such that adjacent symbols are only 1 bit apart then, decoding a symbol incorrectly will lead to making only a 1 bit error. If adjacent symbols are organized in a binary fashion, this is not the case and 50% of the time, a miscoded symbol will mean two bit errors. Gray coding helps us to organize our signaling such that adjacent symbols are different by exactly one bit and hence the Bit Error rate for QPSK is one-half the symbol error rate.

Since there are two bits per symbol, the BER of a QPSK when it is Gray coded can be written as

$$P(e)_{QPSK} = \frac{1}{2} erfc\sqrt{\frac{E_b}{N_0}} \qquad 33$$

This is exactly the same as BPSK.

Gray coding and 8-PSK modulation

In 8-PSK modulation we transmit 8 symbols. Each one of these symbols represents 3 bits. How do we decide which symbol should represents which bit pattern. The first choice is the binary order, called natural order here in Column 2. We can alternately also number the symbols by a different ordering, Gray coding as in Column 4.

Table I – Natural vs. Gray coding of symbols

Symbol Number	**Natural Ordering**	**Bit Diff. to Next Neighbor**	**Gray Coding**	**Bit Diff. to Next Neighbor**
1	000	3	000	1
2	001	1	001	1
3	010	2	011	1
4	011	1	010	1
5	100	3	110	1
6	101	1	111	1
7	110	2	101	1
8	111	1	100	1

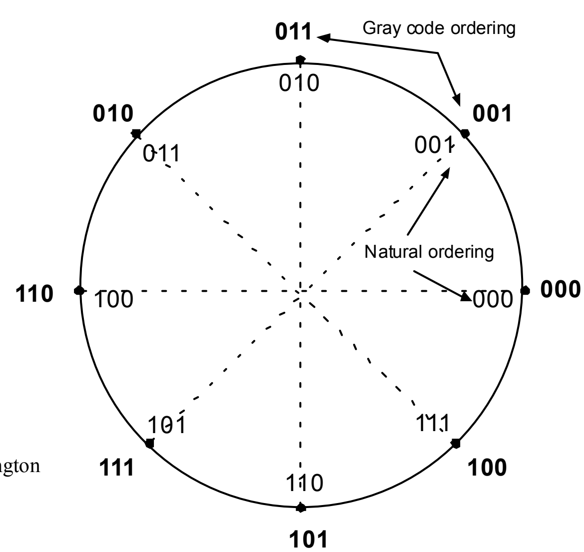


Figure 16 - Mapping bits to symbols, natural and Gray Coding

Natural Ordering

Notice that symbol 1 differs from symbol 2 by only 45 degrees of phase. So is it likely that when an error is made, one symbol will be mistaken for one that looks similar to it. And that means we may confuse symbol 2 for either 3 or symbol 1. It is more unlikely that we will mistake symbol 2 for symbol 6 as these two differ by 180 degrees. When a symbol is misidentified then the bit pattern that it stands for is also mis-decoded.

- Symbol 4 is mistaken for Symbol 5; the bit pattern of symbol 4, 011 is decoded as 100. Three bit errors are made in the final information signal.
- Symbol 4 is mistaken for Symbol 3; the bit pattern of symbol 4, 011 is decoded as 010. One bit error is made in the final information signal.
- The average number of bits that are decoded incorrectly for this symbol is 2 bits.

Gray Coding

Now examine a different type of mapping where adjacent bit mappings differ by just one bit. Now when a mistake is made in identifying a symbol, then the maximum number of bits that will be wrong is one and not two as was the case for natural ordering.

This way of mapping the bit groups to the symbols is called Gray coding. Gray coding is far more natural then binary system. In nature such as in DNA and chemical structures we find that Mother Nature prefers Gray coding and uses it liberally where errors can be devastating. Gray coding is used for nearly all digital modulations starting with QPSK

Relationship of Bit Error Rate with Symbol Error Rate

If a symbol consists of N bits, and if it is decoded incorrectly then the net effect is that anywhere from 1 to N bits could be decoded incorrectly. Let's take a look at QPSK which has 4 symbols. Since only one is correct, the probability of a bit error comes from

Symbol 2 - 1 bit error out of 2
Symbol 3 - 2 bit error out of 2
Symbol 4 - 1 bit error out of 2

So now we can write the bit error as

$$P_b = \frac{\frac{1}{2}(1\ bit\ error) + \frac{1}{2}(1\ bit\ error) + \frac{2}{2}(2\ bit\ error)}{3}$$

There are three symbols that have errors, and each adds to the total bit errors as shown above.

$$P_b = \frac{2}{3} P_s \qquad 34$$

And if the signal is gray coded this improves by factor of N which is the number of bits per symbol.

$$P_b = \frac{M/2}{M-1} P_s \qquad 35$$

The following table gives a list of BER relationships. One of the outstanding things is that BER is a function of only one variable, Eb/N0. However, don't forget that these are theoretical equations under ideal conditions, that is no ISI, no distortions, no filtering effects, no amplitude non-linearities, no fading. In fact when we incorporate all these effects, the required E_b/N_0 to achieve a certain BER goes up considerably. What these numbers do provide is a baseline from which we can judge improvements and degradations.

Inter-symbol distance concept

Constellation diagrams are very important in developing intuitive understanding of BER. The constellation lets us see the signals as population of objects from which the receiver must pick out the correct one. How densely packed are the symbols? This is obviously the key to how good the bit error rate of a signal will be. We want a lot of signals but we also want them to be "different" from each other so mistakes are not frequent.

Note that the bit energy of a modulated signal is equal to

$$E_b = P_c T_b \qquad 36$$

Where P_c is carrier power and T_b the bit time of the information signal.
The carrier power is defined as

$$P_c = \frac{A^2}{2}$$

The symbol energy is defined as

$$E_s = P_c T_s \qquad 37$$

The bit time and symbol time of BPSK is the same, so its bit and symbol energies are the same. For QPSK, which has two independent I and Q channels, we have

$$E_s = P_c T_s = \frac{A^2 T_s}{2} \qquad 38$$

This energy is divided equally between the I and Q channels, so we have

$$E_s = P_c T_s = \frac{A^2 T_s}{2} \rightarrow \begin{array}{l} \dfrac{A^2 T_s}{4} \rightarrow E_b = \dfrac{A^2 T_b}{2} \\ \dfrac{A^2 T_s}{4} \rightarrow E_b = \dfrac{A^2 T_b}{2} \end{array} \qquad 39$$

Bit times for QPSK are related by

$T_b = 2T_s$ from which we get that

$$E_{b(BPSK)} = \frac{A^2 T_b}{2}$$

$$E_{b(QPSK)} = \frac{A^2 T_b}{2} \qquad 40$$

Now lets' examine the constellation of a BPSK signal. Each point is located at an amplitude equal to $\sqrt{E_b}$ from the origin.

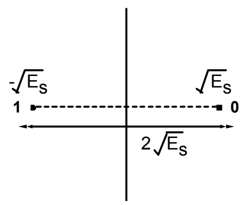


Figure 17 – The distance between the two BPSK symbols is the largest that is possible and that means that these two symbols are least like each other.

We define distance as the degree of dissimilarity between signals. The distance of BPSK symbols to the origin is $\sqrt{E_b}$ and the distance between the two points is

$$d = 2\sqrt{P_c T_b} = 2\sqrt{E_b} \qquad 41$$

This distance although expressed as a scalar quantity is the property of the signal space. It has no physical meaning but is very very helpful in understanding how different modulations result in different bit error rates.

Let's rewrite the BER equation of a BPSK signal by setting

$$E_b = \frac{d^2}{2}$$

Now substitute d for E_b,

$$P(e) = \frac{1}{2} erfc \sqrt{\frac{E_b}{N_0}}$$

$$= \frac{1}{2} erfc \sqrt{\frac{d^2}{4N_0}} \qquad 42$$

This equation tells us that the BER is directly related to intersymbol distance d. Let's see how this applies to other PSK modulations.

BER and QPSK inter-symbol distance

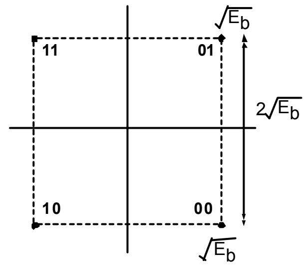


Figure 18 – QPSK inter-symbol distance is same as BPSK

The QPSK inter-symbol distance is equal to

$$2\sqrt{E_b} \qquad 43$$

when we substitute this into Eq. 42, we get

$$\begin{aligned} P(e) &= \frac{1}{2}erfc\sqrt{\frac{d^2}{4N_0}} \\ &= \frac{1}{2}erfc\sqrt{\frac{4E_b}{4N_0}} \\ &= \frac{1}{2}erfc\sqrt{\frac{E_b}{N_0}} \end{aligned} \qquad 44$$

This is same as BPSK, again.

BER and Inter-symbol distance for 6-PSK

Now examine a M-PSK signal space where M = 8. The signal space for a M-PSK is always a circle, and each bit sequence of N bits is represented by one of these M symbols.

$M = 2^N$ and M = 8, N = 3, for 8-PSK, where three bits can be represented by one symbol. Each signal is this circular space is located at a distance $\sqrt{P_s T_s}$ from the origin. Each symbol is 360/M degrees apart. So 8-PSK signals are 360/8 = 45 degrees apart. For small M, the distance d between symbols is given by

$$d^2 = 4NE_b \sin^2\left(\frac{\pi}{M}\right) \qquad 45$$

for 8-PSK, M = 8 and N=3, we get

$$d^2 = 12E_b \qquad 46$$

Plug this into equation 42, we get BER of 8-PSK,

$$\begin{aligned} P(e) &= \frac{1}{2} erfc\sqrt{\frac{E_b}{N_0}} \\ &= \frac{1}{2} erfc\sqrt{\frac{d^2}{4N_0}} \\ &= N\frac{1}{2} erfc\sqrt{\frac{12E_b}{4N_0}} \\ &= \frac{3}{2} erfc\sqrt{\frac{3E_b}{N_0}} \end{aligned} \qquad 47$$

Of course, we can extrapolate to this any M-PSK by using consistent values of M and N in equation.

BER and Inter-symbol distance of 16-QAM

Now let's do M-QAM, specifically 16-QAM. We will compute its BER using the concept of the distance. Now we have 16 symbols to transmit, with each symbol standing for 4 bits.

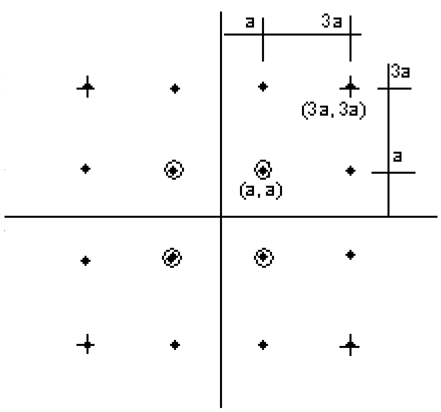


Figure 19 – 16-QAM constellation. Its inter-symbol distance is smaller than 16-PSK.

In PSK, all signals were at the same distance from its neighbor. There was only one inter-symbol distance to be computed. Here we have three different types of symbols, those that are in the middle and have neighbors on all sides, and those that are on the sides so have neighbors on only one side and those in the corners.

We assume that the inner most symbol is at (a, a) volts. First we will compute the signal energy. Energy of a symbol is equal to the sum the energy of all the individual symbols. Another way to look at symbol energy is to see it as the average energy of all different types of symbols. In case of PSK, all symbols have the same energy (since all the amplitudes are the same), PAM and others where the amplitudes vary would have different energies for the symbols. So we write the average energy as

$$E_{av} = \frac{1}{M}\sum_{m=1}^{M} E_m \qquad 48$$

The symbol energy is just the sum of all the individual energies of the different looking symbols.

$$E_b = \frac{E_{av}}{N} = \frac{E_{av}}{\log_2 M} \qquad 49$$

Using a as the distance, we can express the average energy of the symbol as

$$E_s = \frac{1}{4}\left[(a^2 + a^2) + (9a^2 + a^2) + (a^2 + 9a^2) + (9a^2 + 9a^2)\right]$$
$$= 10a^2$$

Now we have

$$a = \sqrt{.1E_s}$$
$$d = 2\sqrt{.1E_s} \qquad 50$$

And since in this modulation, each symbol represents 4 bits, we have

$$E_s = 4E_b$$
$$d = 2\sqrt{.4E_b} \qquad 51$$

Compare this to the inter-symbol distance of QPSK which is

$$d = 2\sqrt{E_b}$$

And 16-PSK, which is

$$d = 2\sqrt{16E_b \sin^2 \frac{\pi}{16}} = 2\sqrt{.15E_b} \qquad 52$$

The distance we just computed for 16-QAM is less than QPSK but more than 16-PSK. And what this says is that 16-QAM has a BER that is in between those two other schemes.

The BER of a 16-QAM is computed by plugging in this value of inter-symbol distance d in Eq 42. And we get,

$$P(e) = \frac{1}{2} erfc\sqrt{\frac{E_b}{N_0}}$$
$$= \frac{1}{2} erfc\sqrt{\frac{d^2}{4N_0}}$$
$$= 4\frac{1}{2} erfc\sqrt{\frac{.4E_b}{N_0}} \qquad 53$$
$$= 2erfc\sqrt{\frac{.4E_b}{N_0}}$$

BER and inter-symbol distance of a FSK signal

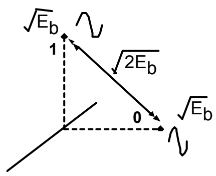


Figure 20 – FSK with two signals of two different frequencies

In a FSK system, we have symbol signals given by

$$u_1(t) = \sqrt{\frac{2}{T_b}} \cos 2\pi m\, f_c t$$

$$u_2(t) = \sqrt{\frac{2}{T_b}} \cos 2\pi n\, f_c t \qquad 54$$

These are the FSK basis functions and are harmonics. The symbol signals are now given by (after scaling with E_b)

$$s_1(t) = \sqrt{E_b}\; u_1(t)$$

$$s_2(t) = \sqrt{E_b}\; u_2(t) \qquad 55$$

The signals are orthogonal when n and m are integers. Now imagine just a two dimensional system. The signals lie on the two axes and the distance between these two is easily computed as

$$d = \sqrt{2E_b} \qquad 56$$

This distance is larger than 4-PSK but smaller than 16QAM. Now the BER can be written as

$$P_e = \frac{1}{2} erfc \sqrt{\frac{d^2}{N_0}}$$

$$P_e = \frac{1}{2} erfc \sqrt{\frac{E_b}{2N_0}} \qquad 57$$

Relative Inter-symbol distances

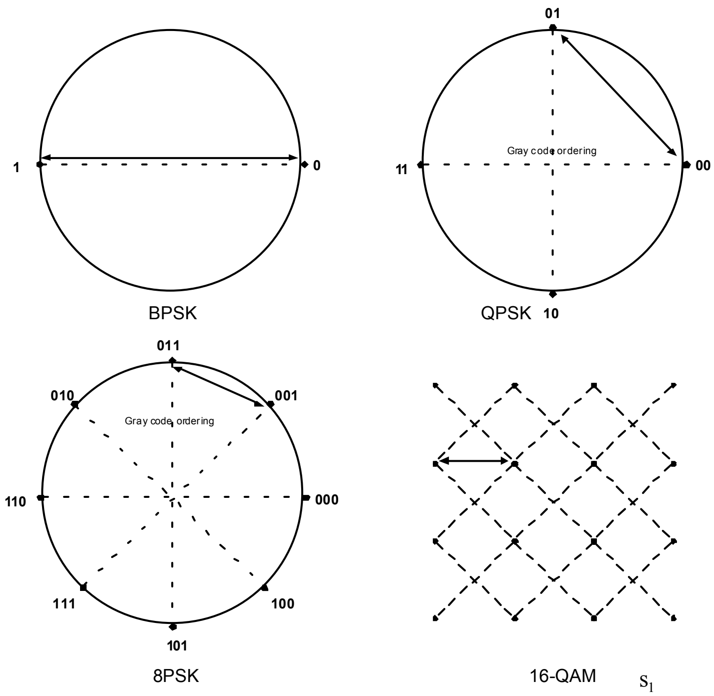


Figure 20 - Inter-symbol distance decreases as number of symbols in a signal space goes up.

The above figure shows that for the same power signal, one with more symbols will have smaller inter-symbol distance. This leads to larger bit error rates. This problem is excerbated by other distortions making the power required to operate these modulations non-linearly related to the linear case. Other techniques can be applied to control the distortions but there is no way to overcome the minimum required E_b/N_0 for these modulations. Only with coding can we reduce the requirements down but then we do pay the price of a smaller throughput.

"No pain, no gain."

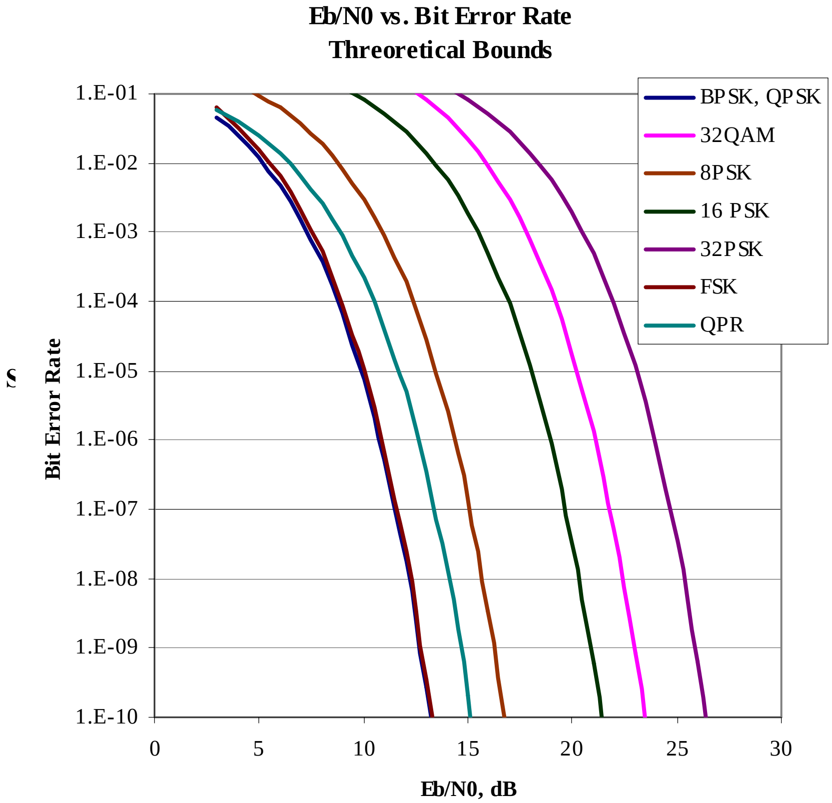


Figure 21 – BER curves for various modulations

Table II - BER Equations

Modulation	$M = 2^N$, N	Pe	Bandwidth
BPSK	2, 1	$\frac{1}{2} erfc \left[\frac{E_b}{N_0}\right]^{1/2}$	R_b
QPSK	4, 2	$\frac{1}{2} erfc \left[\frac{E_b}{N_0}\right]^{1/2}$	$R_b/2$
SQPSK	4, 2	$\frac{1}{2} erfc \left[\frac{E_b}{N_0}\right]^{1/2}$	$R_b/2$
MSK	4, 2	$\frac{1}{2} 1\ erfc \left[\frac{E_b}{N_0}\right]^{1/2}$	$R_b/2$
M-PSK	M, N	$erfc \left[\frac{NE_b}{N_0} \sin^2 \pi / M\right]^{1/2}$	R_b/N
16-QAM	16, 4	$2\ erfc \left[0.4 \frac{E_b}{N_0}\right]^{1/2}$	$R_b/4$

M-QAM	M, N	$2\left[1-\frac{1}{N}\right] erfc\left[\left(\frac{3}{N^2-1}\right)\frac{E_b}{N_0}\right]^{1/2}$	R_b /N
QPR	L levels	$\frac{1}{2}\, erfc\left[\frac{\pi^2}{16}\frac{E_b}{N_0}\right]^{1/2}$	R_b /4
LQPR	L levels	$2\left[1-\frac{1}{L^2}\right] erfc\left[\frac{\pi}{4}(\log_2 L)^{1/2}\left(\frac{6}{L^2-1}\right)^{1/2}\frac{E_b}{N_0}\right]^{1/2}$	R_b /L
M-FSK	M, N	$\frac{M-1}{2}\, erfc\left[\frac{N E_b}{2N_0}\right]^{1/2}$	M R_b /N

http://www.vmsk.org/BERMeas.pdf

http://www.ctr.kcl.ac.uk/lectures/Reza/Lec3354/slides6.pdf

Charan Langton
Errors, corrections, comments, please email me: mntcastle@earthlink.net

Symbols	Bits	AM	FM	PM
1	00	$\underline{A1}\sin(\omega t+\phi)$	$A\sin(\underline{\omega 1}\, t+\phi)$	$A\sin(\omega t+\underline{\phi 1})$
2	01	$A2\sin(\omega t+\phi)$	$A\sin(\omega 2\, t+\phi)$	$A\sin(\omega t+\phi 2)$
3	10	$A3\sin(\omega t+\phi)$	$A\sin(\omega 3\, t+\phi)$	$A\sin(\omega t+\phi 3)$
4	11	$A4\sin(\omega t+\phi)$	$A\sin(\omega 4\, t+\phi)$	$A\sin(\omega t+\phi 4)$