Understanding Frequency Modulation (FM), Frequency Shift Keying (FSK), Sunde's FSK and MSK and some more

The process of modulation consists of mapping the information on to an electromagnetic medium (a carrier). This mapping can be digital or it can be analog. The modulation takes place by varying the three parameters of the sinusoid carrier.

1. Map the info into amplitude changes of the carrier
2. Map the info into changes in the phase of the carrier
3. Map the info into changes in the frequency of the carrier.

The first method is known as **amplitude modulation.** The second and third are both a from of **angle modulation**, with second known as phase and third as frequency modulation.

Let's start with a sinusoid carrier given by its general equation

$$c(t) = A_c \cos(2\pi f_c t + \phi_0)$$

This wave has an amplitude A_c a starting phase of fð$_0$ and the carrier frequency, f_c. The carrier in Figure 1 has amplitude of 1 v, with f_c of 4 Hz and starting phase of 45 degrees.

Generally when we refer to amplitude, we are talking about the maximum amplitude, but amplitude also means any instantaneous amplitude at any time t, and so it is really a variable quantity depending on where you specify it.

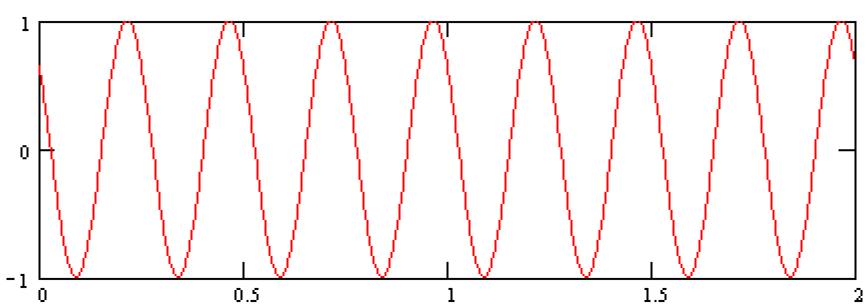
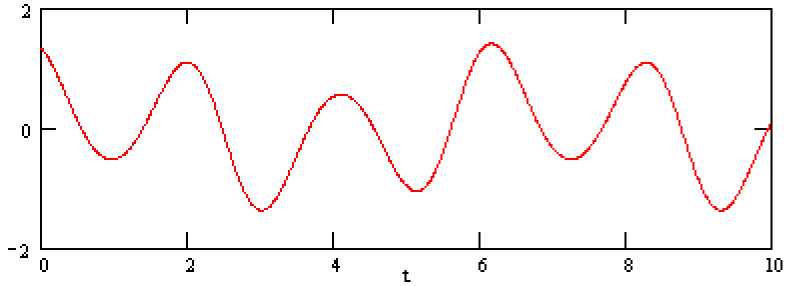
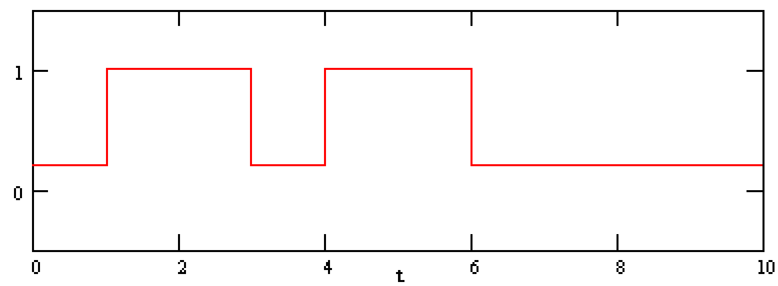


Figure 1 – A sinusoid carrier of frequency 4, starting phase of 45 degrees and amplitude of 1 volt.

The amplitude modulation changes the amplitude (instantaneous and maximum) in response to the information. Take the following two signals; one is analog and the other digital



a. Analog message signal, m(t)



b. Binary message signal, m(t)

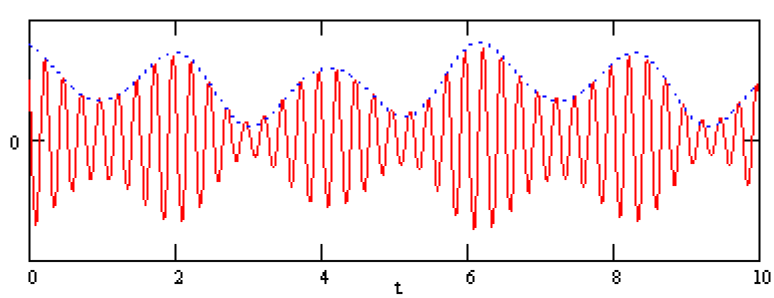
Fig 2 – Two arbitrary message signals, m(t)

About Amplitude Modulation (AM)

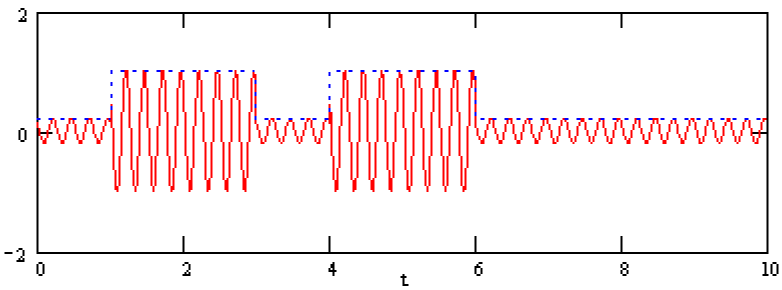
The amplitude modulated wave is created by multiplying the *amplitude* of a sinusoid carrier with the message signal.

$$s(t) = m(t)\,c(t)$$
$$= A_c m(t)\cos(2\pi f_c t + \phi_0)$$

The modulated signal shown in Figure 3, is of carrier frequency f_c but now the amplitude changes in response to the information. We can see the analog information signal as the envelope of the modulated signal. Same is true for the digital signal.



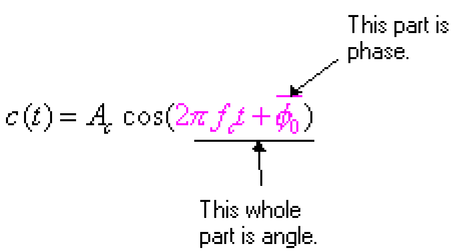
a. Amplitude modulated carrier shown with analog message signal as its envelope



c. Amplitude modulated carrier shown with binary message signal as its envelope

Figure 3 – Amplitude modulated carrier, a. analog, b. digital

Let's look at the argument of the carrier. What is it?



The argument is an angle in radians. The argument of a cosine function is always an angle we know that from our first class in trigonometry. The second term is what is generally called the **phase**. In amplitude modulation only the amplitude of the carrier changes as we can see above for both binary and analog messages. Phase and frequency retain their initial values.

Any modulation method that changes the angle instead of the amplitude is called *angle modulation*. The angle consists of two parts, the phase and the frequency part. The modulation that changes the phase part is called phase modulation (PM) and one that changes the frequency part is called frequency modulation (FM).

How do you define frequency? Frequency is the number of 2 pð revolutions over a certain time period. Mathematically, we can write the expression for average frequency as

$$f_{\Delta t} = \frac{\phi_i(t+\Delta t) - \phi_i(t)}{2\pi\,\Delta t} \qquad 1$$

This equation says; *the average frequency is equal to the difference in the phase at time t + Dðt and time t, divided by 2pð* (or 360 degrees if we are dealing in Hz.)

Example: a signal changes phase from 45 to 2700 degrees over 0.1 second. What is its average frequency?

$$= \frac{2700 - 45}{360 * 0.1} = 73.75 \text{ Hz}$$

This is the average frequency over time period t = 0.1 secs. Perhaps it will be different over 0.2 secs or some other time period or maybe not, we don’t know.

What is the instantaneous frequency of this signal at any particular moment in the 0.1 second period? We don’t really know given this information.

The instantaneous frequency is defined as the limit of the average frequency as Dðt gets smaller and smaller and approaches 0. So we take limit of equation 1 to create an expression for instantaneous frequency, $f_i(t)$,.

The $f_i(t)$ is the limit of $f_{D\eth t}(t)$ as Dðt goes to 0. The phase change over time Dðt is changed to a differential to indicate change from discrete to continuous.

$$f_i(t) = \lim_{\Delta t \to 0} f_{\Delta t}(t)$$
$$= \lim_{\Delta t \to 0} \frac{\phi_i(t + \Delta t) - \phi_i(t)}{2\pi \, \Delta t}$$
$$= \frac{1}{2\pi} \frac{d\phi_i(t)}{dt}$$

This last result is very important in developing understanding of both phase and frequency modulation! The 2pð factor has been moved up front. The remaining is just the differential of the phase.

Another way we can state this is by recognizing that radial frequency wð is equal to

$$\omega = 2\pi \, f_i(t)$$

It is also equal to the rate of change of phase,

$$\omega = \frac{d(\phi_i(t))}{dt}$$

so again we get,

$$f_t(t) = \frac{1}{2\pi} \frac{d\phi_i(t)}{dt} \qquad 2$$

Intuitively, it says; the frequency of a signal is equal to its phase change over time. When seen as a phasor, the signal phasor rotates in response to phase change. The faster it spins (phase change), the higher its frequency.

What does it mean, if I say: the phasor rotates for one cycle and then changes directions, goes the opposite way for one cycle and then changes direction again? This is a representation of a phase modulation. Changing directions means the signal has changed its phase by 180 deg.

We can do a simple minded phase modulation this way. Go N spins in clockwise directions in response to a 1 and N spins in counterclockwise directions in response to a 0. Here N represents frequency of the phasor.

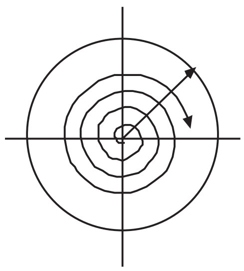


Figure 4 – The carrier as a phasor, the faster it spins, the higher the frequency.

If frequency is the rate of change of phase, then what is phase in terms of frequency? As we know from definition of frequency that it is number of full 2 pð rotations in a time period.

Given, a signal has traveled for 0.3 seconds, at a frequency of 10 Hz with a starting phase of 0, what is its phase now?

$$Phase\ now\ \phi = 2\pi f_i t = 10\,Hz \times 2\pi \times .3 = 20\ radians$$

This is an integration of the total number of radians covered by the signal in 0.3 secs. Now we write this as an integral,

$$\phi_i(t) = 2\pi \int_0^{.3} f_i(t)dt$$

and since this is average frequency, it is constant over this time period, we get

$$Phase\ now\ \phi = 2\pi f_i t = 10\,Hz \times 2\pi \times .3 = 20\ radians$$

We note the phase and frequency are related by

Phase a Integral of frequency

Frequency a Differential of phase

Phase modulation

Let the phase be variable. Going back to the original equation of the carrier, change the phase (the underlined term only) from a constant to a function of time.

$$c(t) = A_c \cos(2\pi f_c t + \underline{\phi(t)}\ )$$

We can phase modulate this carrier by changing the phase in response to the message signal.

$$\phi_i(t) = k_p m(t)$$

Now we can write the equation for the carrier with a changing phase as

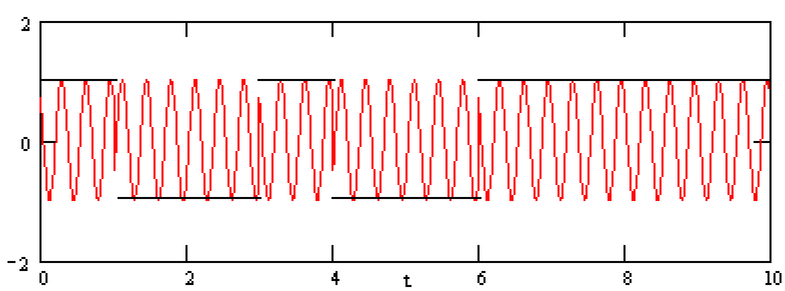
$$s(t) = A_c \cos(2\pi f_c t + \underline{k_p m(t)}) \qquad 4$$

The factor k_p is called the phase sensitivity factor or the modulation index of the message signal. For analog modulation, this expression is called the phase modulation. In phasor representation of an analog PM, the phasor slows down gradually to full stop and then again picking up speed in the opposite direction gradually.

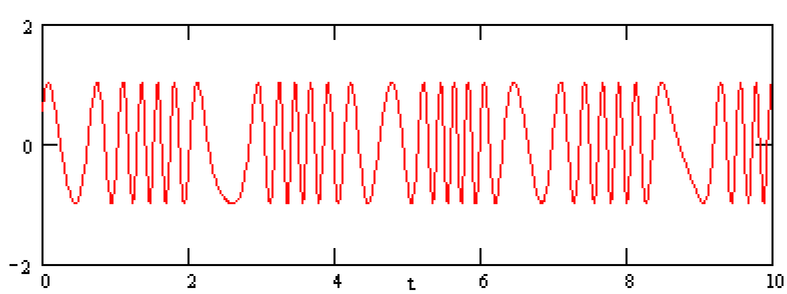
In binary case, the phasor does not slow down gradually but stops abruptly. This is easy to imagine. Take a look at equation 4. Replace the underlined terms by a 0 or a 1, or better yet, replace it by $180^{o\eth}$ if m(t) is 1 and $-180^{o\eth}$ if it is a 0. Now, we have a binary PSK signal. The phase changes from $-180^{o\eth}$ to $+180^{o}$ in response to a message bit.

This is the main difference between analog and digital phase modulation in that in digital, the phase changes are discrete and in analog, they are gradual and not obvious. For binary PSK, phase change is mapped very simply as two discrete values of phase.

$$s(t) = a_c \cos(2\pi f_c t + \phi_j) \qquad \phi_j = 0 \; or \; \pi$$



a. Phase modulation is response to a binary message. The phase changes are abrupt.



b. Phase modulation is response to an analog message. The phase changes are smooth.

Figure 5 – Phase modulated carrier for both binary and analog messages.

Frequency modulation

FM is a variation of angle modulation where instead of phase, we change the frequency of the carrier in response to the message signal.

Vary the frequency by adding a time varying component to the carrier frequency.

$$f_t(t) = f_c + k_f m(t)$$

where f_c is the frequency of the unmodulated carrier, and k_f a scaling factor, and m(t), the message signal. The term k_f m(t) can called a deviation from the carrier frequency.

Example, a carrier with f_c = 100, k_f = 8 and message bit rate = 1. Assume the message signal is polar so we have 1 for a 1 and -1 for 0.

For m(t) = 1, we get

$$f_1 = 100 + 8(1) = 108 \text{ Hz}$$

for m(t) = -1

$$f_2 = 100 - 8(1) = 92 \text{ Hz}$$

The phasor rotates at a frequency of 108 Hz, as long as it has a signal of 1 and at 92 Hz for a signal indicating a 0 bit.

Remember the equation relating phase and frequency

$$f_i(t) = \frac{1}{2\pi}\frac{d\theta_i(t)}{dt} \qquad 5$$

which can also be written as

$$\phi_i(t) = 2\pi \int_{-\infty}^{t} f_i(t)dt \qquad 6$$

Now look at the carrier equation,

$$s(t) = A_c \cos(2\pi f_c t + \phi_i) \qquad 7$$

We define the instantaneous frequency of this signal as the sum of a constant part, which is the carrier frequency, and a changing part as show below.

$$f_t(t) = f_c + k_f m(t) \qquad 8$$

The argument of the carrier is an angle. So we need to convert this frequency term to an angle. Do that by taking the integral of this expression (invoke equation 6)

$$\theta_i(t) = 2\pi \int_{-\infty}^{t} f_c dt + 2\pi k_f \int_{-\infty}^{t} m(t)dt$$

and since f_c is a constant, this is just equal to

$$\theta_i(t) = 2\pi f_c t + 2\pi k_f \int_{-\infty}^{t} m(t)dt \qquad 9$$

Now we plug this as the argument of the carrier into equation 7 and we get

$$s(t) = A_c \cos\underline{(2\pi f_c t + 2\pi k_f \int_{-\infty}^{t} m(t)dt)} \qquad 10$$

This represent the equation of a FM modulated signal. A decidedly unpleasant looking equation! This is about as far as you can get from intuitive. There is no resemblance at all to anything we can imagine.

Analog FM is one of the more complex ideas in communications. It is very difficult to develop an intuitive feeling for it. Bessel functions rear their multi-heads here and things go from bad to worse. Thankfully this is where binary signals come to the rescue. The binary or digital form of Frequency Modulation is very easy to understand .

But before we do that, examine the following relationship between FM and PM.

Let's write out both equations side by side so we can see what's going on.

PM: $$s(t) = A_c \cos(2\pi f_c t + \underline{k_p m(t)}) \qquad 11$$

FM: $$s(t) = A_c \cos\underline{(2\pi f_c t + 2\pi k_f \int_{-\infty}^{t} m(t)dt)} \qquad 12$$

Note that in FM, we integrate the message signal before modulating. Both FM and PM modulated signals are conceptually identical, the only difference being in the first case phase is modulated directly by the message signal and in FM case, the message signal is first integrated and then used in place of the phase.

Using a Phase modulator we can create a FM signal by just integrating the message signal first. Similarly with a FM modulator we can create a PM signal by differentiating the message signal before modulation.

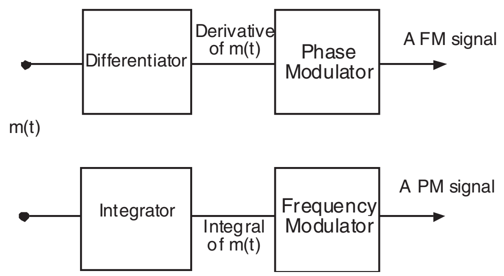


Figure 6 – Phase and Frequency modulators are interchangeable by just changing the form of the message signal. Differentiate the message signal and then feed to a PM modulator to get a FM signal and vice-versa for FM.

Frequency Shift Keying – the digital form of FM

The binary version of FM is called the Frequency Shift Keying or FSK. Here the frequency does not keep changing gradually over symbol time but changes in discrete amounts in response to a message similar to binary PSK where phase change is discrete.

The modulated signal can be written very simply as consisting of two different carriers.

$$s_1(t) = A_c \cos(2\pi f_1 t)$$
$$s_2(t) = A_c \cos(2\pi f_2 t)$$

$s_1(t)$ in response to a 1 and $s_2(t)$ in response to a 0.

Getting sophisticated, this can be written as a deviation from the carrier frequency.

$$s_1(t) = A_c \cos(2\pi (f_c - \Delta f)t)$$
$$s_2(t) = A_c \cos(2\pi (f_c + \Delta f)t)$$

Here Dðf is called the **frequency deviation**. This is excursion of the signal above and below the carrier frequency and indicates the quality of the signal such in stereo FM reception.

Each of these two frequencies f_1 and f_2, are an offset from the carrier frequency, f_c. Let's call the higher of these f_h and lower f_l,. Now we can create a very simple FM modulator as shown the Figure 7. We can use this table to modulate the incoming message signal.

	Amplitude of	
m(t)	f_h	f_l
-1	0	1
+1	1	0

We send f_l in response to a -1 and 0 in response to a +1. When we get a change from a -1 to +1, we switch frequency, so that only one signal, either fh or fl is transmitted.

The combined signal can be written as, keeping mind that two terms are orthogonal in time and never occur at the same time.

$$s_{BFSK}(t) = A_h \cos(2\pi f_h t + \phi_h) + A_l \cos(2\pi f_l t + \phi_l) \qquad 13$$

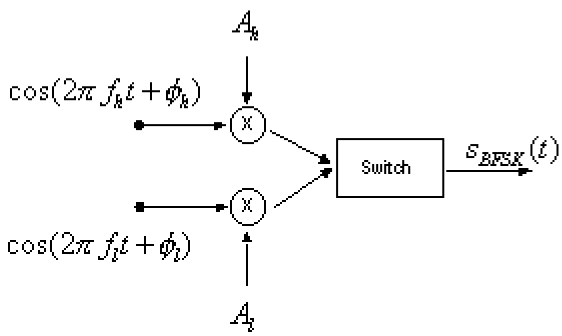


Figure 7 – A FSK modulator

In the modulator shown here, we are changing the amplitude A_l and A_h which are either 0 or 1 in response to the message signal, but never present at the same time.

The above process looks very much like amplitude modulation. The only difference is that the signal polarity goes from 0 to 1(because A_h is either 0 or 1), rather than -1 to +1 as it does in BPSK. This is an important difference and we can try to understand FM better by making this clever substitution (Ref. 1).

$$\begin{aligned} A_h(t) &= \frac{1}{2} + \frac{1}{2} A_h^{'}(t) \\ A_l(t) &= \frac{1}{2} + \frac{1}{2} A_l^{'}(t) \end{aligned} \qquad 14$$

These transformations turn A_h and A_l from 0,1 to -1, +1.
Now we substitute these into equation 13 and get,

$$\begin{aligned} s_{BFSK}(t) = {} & \cos(2\pi f_h t + \phi_h) + \cos(2\pi f_l t + \phi_l) \\ & + A'_h \cos(2\pi f_h t + \phi_h) + A'_l \cos(2\pi f_l t + \phi_l) \end{aligned} \qquad 15$$

The first two terms are just cosines of frequencies f_h and f_l, so their spectrum contribution is an impulse at frequencies f_h and f_l. The second two terms are amplitude modulated carriers. These last two terms give rise to a sin x/x type of spectrum of the square pulse as shown below.

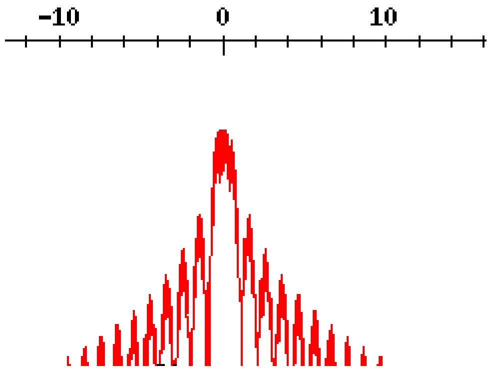
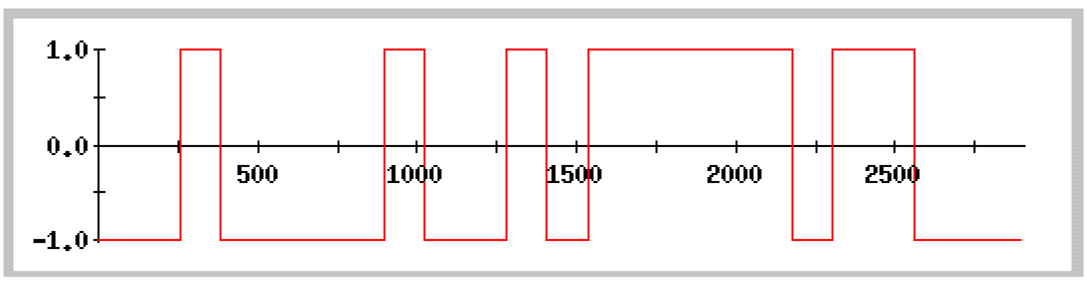


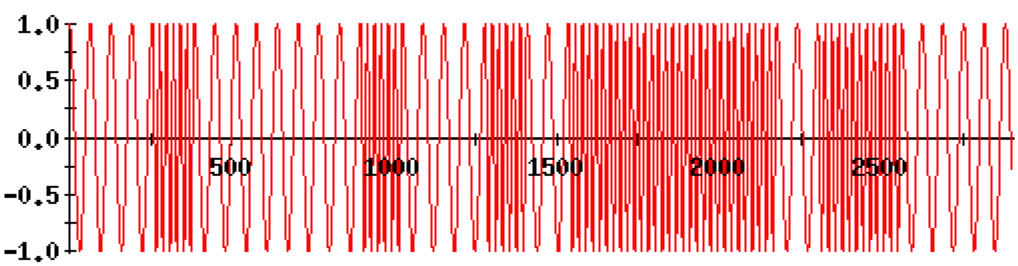
Figure 8 – Sin x/ x spectrum of a square signal pulse

The basic idea behind a FM spectrum is that it is a superimposition of impulses and sin x/x spectrum of the pulses. However, this is true only if the separation or deviation is quite large and the overlap is small. The combinations are not linear and not easily predicted unless of course you use a program like SPW, Matlab or Mathcad as I do.

The following figures show the spectrum of an FSK case, where the carrier frequency is 10, the message signal frequency is 1, Dðf is equal 2.



a. Binary message signal, m(t) = 2



b. Modulated signal, s(t) with fc = 4

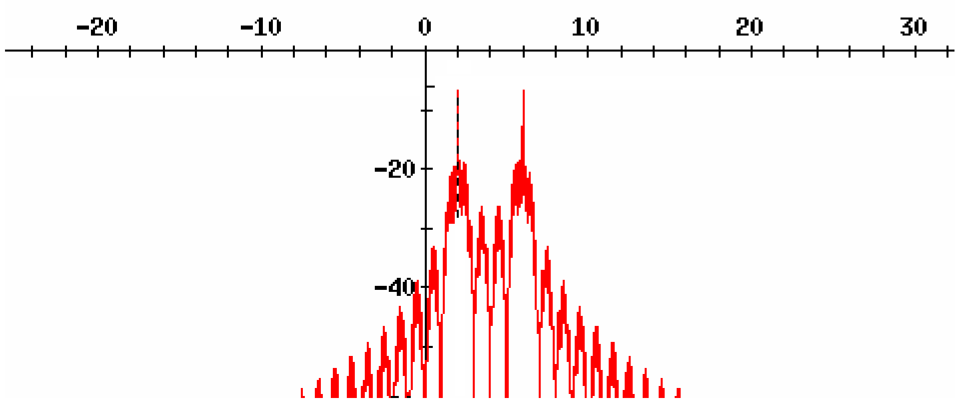


Figure 9- Spectrum of the modulated signal, Dðf is equal 2. Note the two impulses at frequencies 2 and 6 Hz.

Here we see clearly that there are two impulses representing the carrier frequencies, 2 and 6. If you ignore, these, the remaining are just two sin x/ x spectrums centered at frequencies 2 and 6, added together.

We define a term called the modulation index of a FM signal.

$$m = \frac{\Delta f}{f_m} \qquad 16$$

where Dðf is what is called the frequency deviation, and f_m is the frequency of the message signal.

$$\Delta f = \frac{f_2 - f_1}{2}$$

This factor m, determines the occupied bandwidth and is a measure of the bandwidth of the signal.

The FM broadcasting in the US takes place at carrier frequencies of app. 50 MHz. The message signal which is music, has frequency content up to about 15 kHz. The FCC allows deviation of 75 kHz. For this, the m is

$$m = \frac{\Delta f}{f_m} = \frac{75}{5} = 15$$

m, is independent of the carrier frequency and depends only on the message signal frequency and the allowed deviation frequency. This factor is not constant for a particular signal as is the AM modulation index. In FM broadcasting m can go very high, since the message signal has some pretty low frequencies, such as 50 Hz. For FM it can vary any where from 1500 to 5 on the low end.

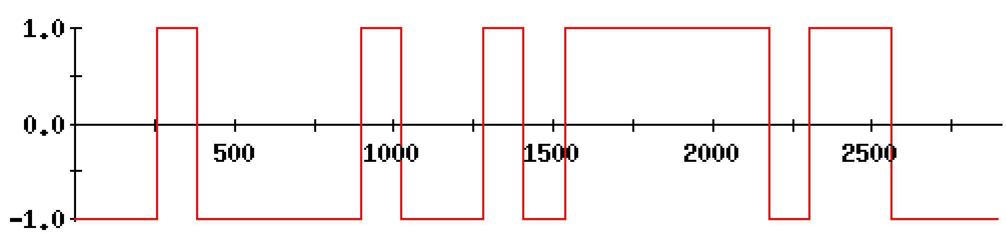
FM signal is classified in two categories,

Signals with $m \ll 1$ are called **Narrowband FM**

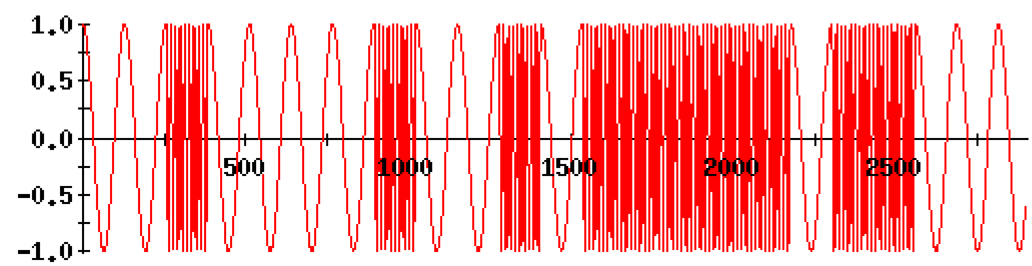
Signals with $m \gg 1$ are called **Wideband FM**

FM radio is an obviously a wideband signal by this definition.

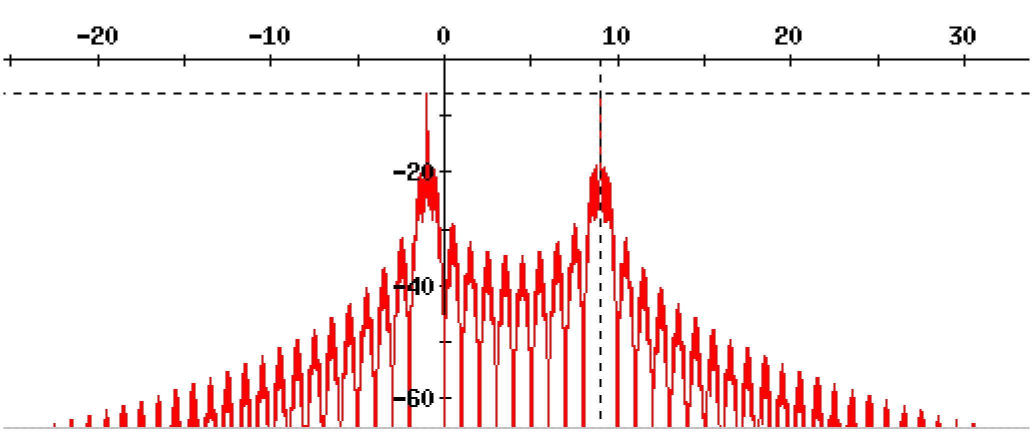
The smaller the deviation, the closer the main lobes lie and in fact overlap and larger the deviation, the further out the signal spreads. The following figures show the effect of increasing modulation index on the signal spectrum and bandwidth.



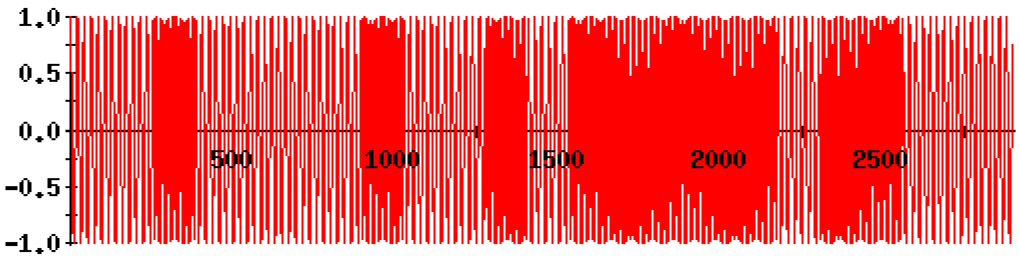
a. Binary message signal, m(t) = 1



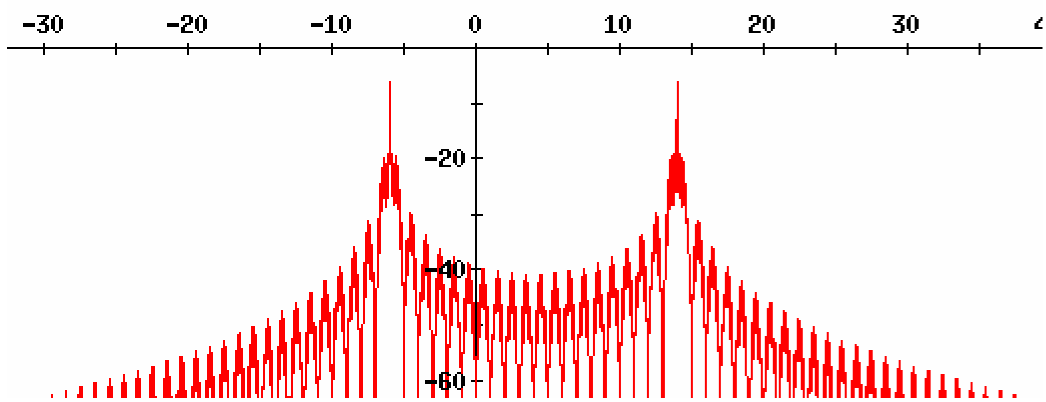
b. Modulated signal f_c = 4, Dðf = 5



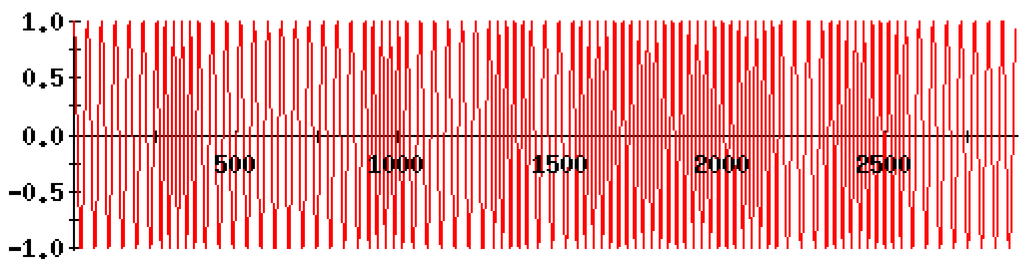
c. Spectrum shows two impulses at f; 5 + 4 = 9 and 4 - 5 = -1



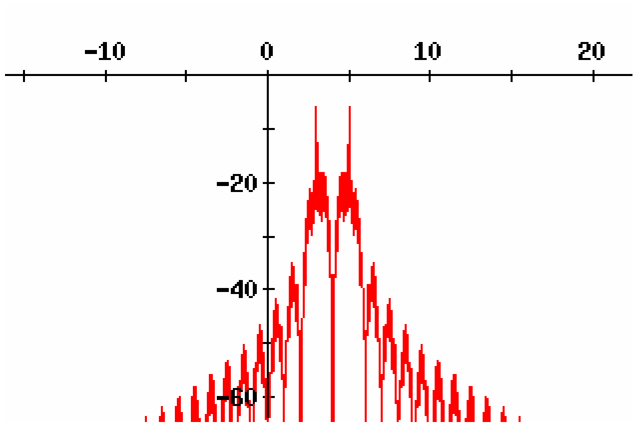
d. Modulated signal, f_c = 4, Dðf = 10



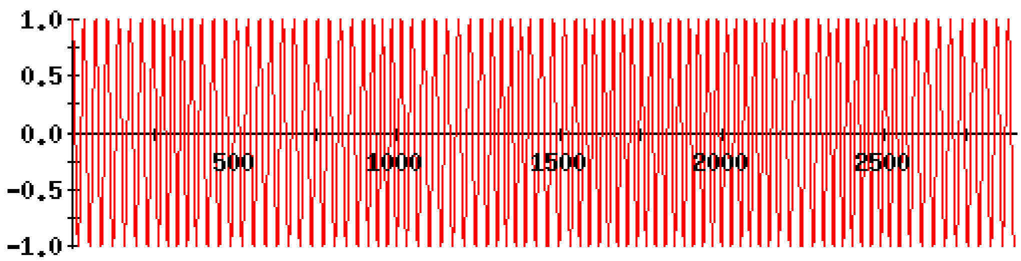
e. Spectrum shows two impulses at f; 10 + 4 = 14 and 4-10 = -6
Note that the spectrum does look like a superposition of impulses and sinc functions.



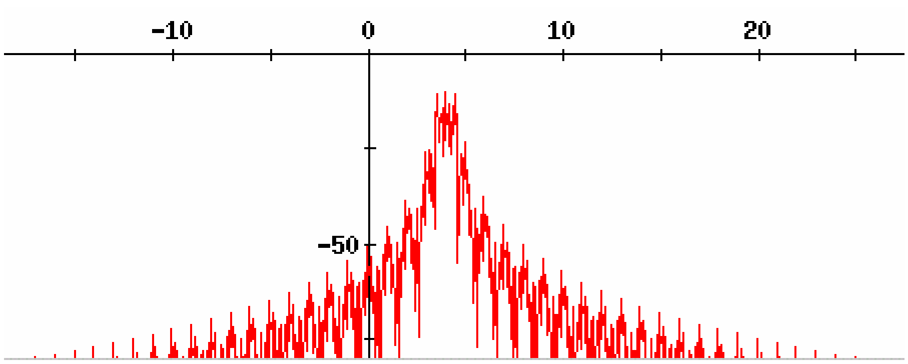
f. Modulated signal, f_c = 4, Dðf = 1



g. Spectrum shows two impulses at f; 1 + 4 = 5 and 4 - 1 = 3



h. Modulated signal, f_c = 4, Dðf = .4



i. Spectrum shows two impulses at f; 4 + .1 = 4.1 and 4 - .11 = 3.9

Figure 10 – Relationship of FM modulation index and its spectrum. As index gets large, the signal bandwidth increases. In these figures, we can define the bandwidth as the space between the main lobes or impulses.

In fact as k_f gets very small, the general rules of identifying the impulses do not apply and we have to know about Bessel functions in order to compute the spectrum.

There are two very special cases of FM that bear discussing.

Sundes' FSK

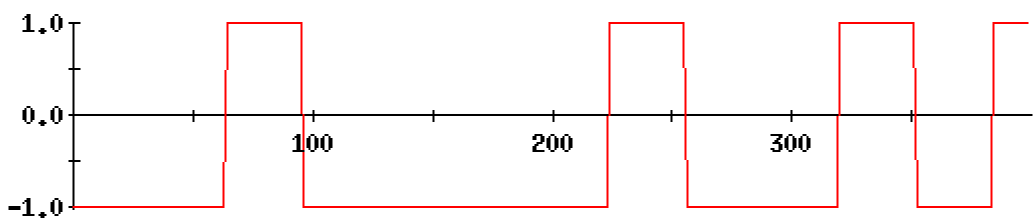
Sunde's FSK is a special case of the general form we have described above. Here the frequency spacing of the two carriers, f_h and f_l is exactly equal to the symbol rate.

$$D\delta f = f_m$$

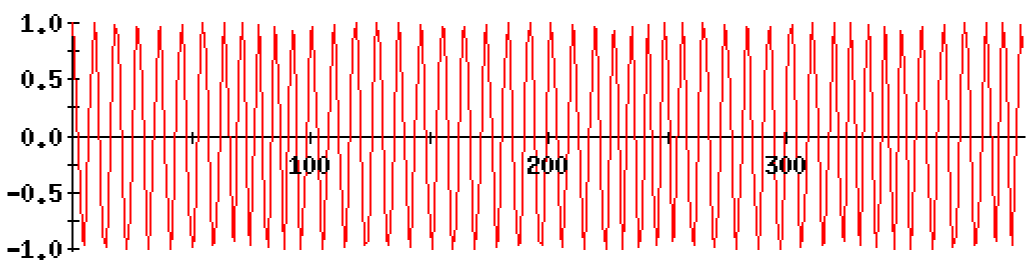
or

$$m = 1$$

Two tones appear at frequencies which are exactly the symbol rate and these help in the demodulation of this signal without external timing information. When you can extract timing information from a signal, the detection task is called coherent and generally has better BER performance.



a. Message signal with bit rate = 1.0



b. Modulated signal, f_c = 8, deviation = 1.0 k_f = 1.0

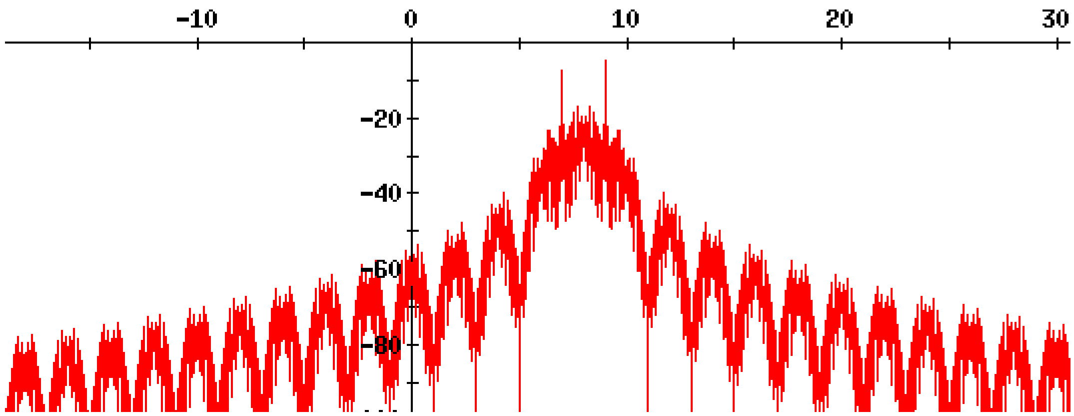


Figure 11 – Sunde's FSK spectrum with tones at = 9 and 7 Hz.

The spectrum of Sunde's FSK is given by

$$S(f) = \frac{1}{4}\left[\underline{\delta(f - .5\frac{1}{T_s})} + \underline{\delta(f + .5\frac{1}{T_s})} + \frac{4T_s}{\pi^2}\left[\frac{\cos(\pi f T_s}{4f^2T_s^2 - 1}\right]^2\right] \qquad \underline{17}$$

The first two underlined terms are the two impulses (batman's ears in above spectrum) and the last term is the spectrum of the message signal, which is not quite sin x/x as I had said earlier.

Minimum Shift Keying - MSK

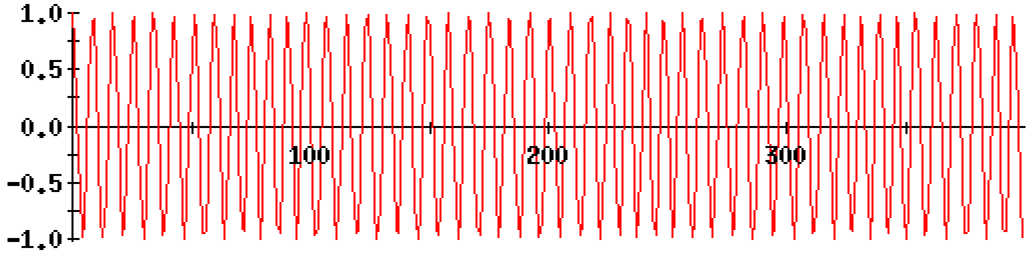
Minimum Shift Keying, another special from of FSK (also called Continuous phase FSK or CPSFK) and a very important one. It is used widely in cellular systems.. MSK is also a special from of PSK owing to the equivalence of phase and frequency modulation which we will discuss in more detail again. For now, let it suffice to say that MSK is a special case where

Dðf = .5 f_m

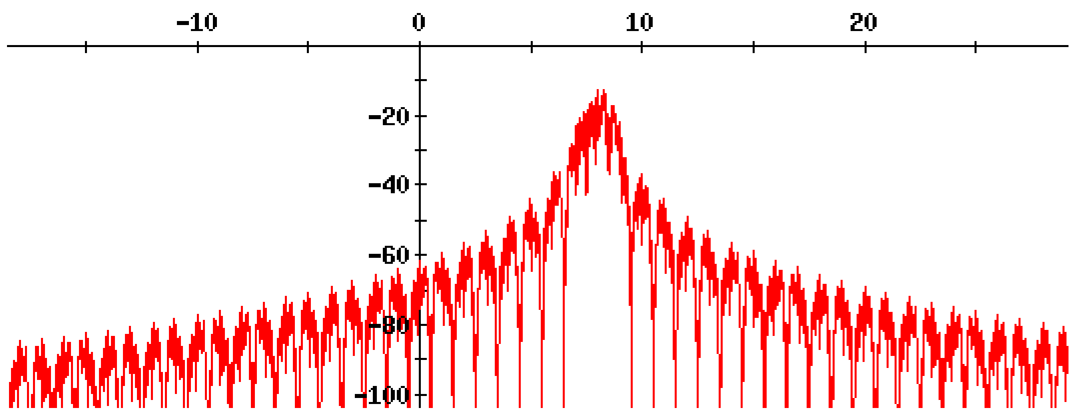
or

m = 0.5

The MSK spectrum has no discrete components unlike all the other FSKs including Sunde's above. The spectrum is just a bit wider than QPSK but its side lobes fall off much more quickly than QPSK In wireless systems, it often offers better properties than QPSK due to its constant envelope characteristics, as we can see in the comparison above.



a. MSK modulated signal, m(t) = 2.0, Dðf = 1



MSK spectrums does look a lot like a QPSK spectrum but is not the same.

Figure 12 – MSK signal and spectrum. Note that it has no "Batman ears" as in all other FSKs.

The spectrum of MSK is given by

$$S(f) = \left[\frac{16T_s}{\pi^2} \left[\frac{\cos(2\pi f T_s}{16 f^2 T_s^2 - 1} \right]^2 \right] \qquad 18$$

(We will discuss MSK again from PSK point of view in the next tutorial.)

M-ary FSK

A M-ary FSK is just an extension of BFSK. Instead of two carriers, we have M. These carriers can be orthogonal or not, but the orthogonal case would obviously give better BER.

M-ary FSK requires considerably larger bandwidth than M-PSK but as M increases, the BER goes down unlike M-PSK. Infact if number of frequencies are increased, M-FSK becomes an OFDM like modulation.

Analog FM

Response to a sinusoid – not as simple as we would like

I have side-skirted the issue of analog FM's complicated spectrum. Binary or digital FSK allows us to delve into FM and is easy to understand but a majority of FM transmission such for radio is analog. So we will touch on that aspect to give you the full understanding.

Now we will look at what happens to a message signal that is a single sinusoid of frequency f_m when it is FM modulated.

$$s(t) = A_c \cos\underline{(2\pi f_c t + 2\pi k_f \int_{-\infty}^{t} m(t)dt)} \qquad 19$$

Let's take a message signal as shown below.

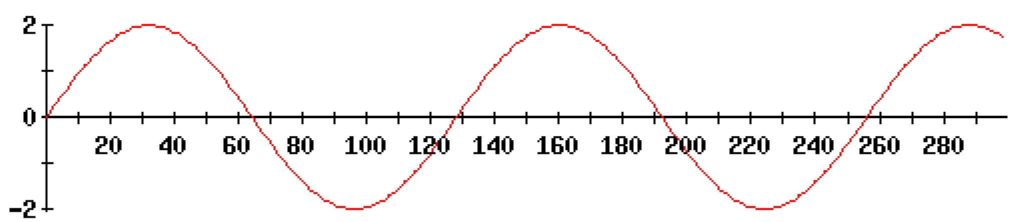


Figure 13 – Just a lowly sinusoid that is about to be mangled to high-fidelity heights by a FM modulator.

$$s(t) = A_c \cos(2\pi f_c t + 2\pi k_f \sin f_m t)$$

Now set $A_c = 1$ and switch to radial frequency for convenience (less typing.)

$$s(t) = \cos(\omega_c t + k_f \sin \omega_m t)$$

By the well-known trigonometric relationship, this equation is becomes

$$\begin{aligned} s(t) &= \cos(\omega_c t + k_f \sin \omega_m t) \\ &= \cos \omega_c t \ \underline{\cos(k_f \sin \omega_m t)} - \sin \omega_c t \, \underline{\sin(k_f \sin \omega_m t)} \end{aligned} \qquad 20$$

Now the terms $\underline{\cos(k_f \sin \omega_m t)}$ can be expanded. How and why, we leave it to the mathematicians, and see that we again get a pretty hairy looking result.

$$\cos(k_f \sin \omega_m t) = \underline{J_0(k_f)} + 2\underline{J_2(k_f)}\cos(2\omega_m t) + 2\underline{J_4(k_f)}\cos(4 \qquad 21$$

The $\underline{\sin(k_f \sin\omega_m t)}$ is similarly written as

$$\sin(k_f \sin\omega_m t) = 2\underline{J_1(k_f)}\sin(\omega_m t) + 2\underline{J_3(k_f)}\sin(3\omega_m t) + 2\underline{J_5(}k$$

22

(Note $k_f = m$)

The underlined functions are Bessell functions and are function of the modulation index m or k_f. They are seen many places where harmonic signals analysis and despite my bad-mouthing, they are quite harmless and benign. There are just so many of them. They ones here are of first kind and of order n. The terms of the cosine expansion since it is an even function, contain only even harmonics of w_m and sin expansion has only odd harmonics and this is obvious if you will examine the above equations even casually.

Figure 14 shows the general shape of 4 Bessel functions as the order is increased. The x-axis is modulation index k_f and the y-axis is the value of the Bessel function value. For example, for $k_f = 2$, the values of the various Bessel functions are

J0 = .25
J1 = .65
J2 = .25
J3 = .18

The FM carriers take their amplitude values from these functions as we will see below.

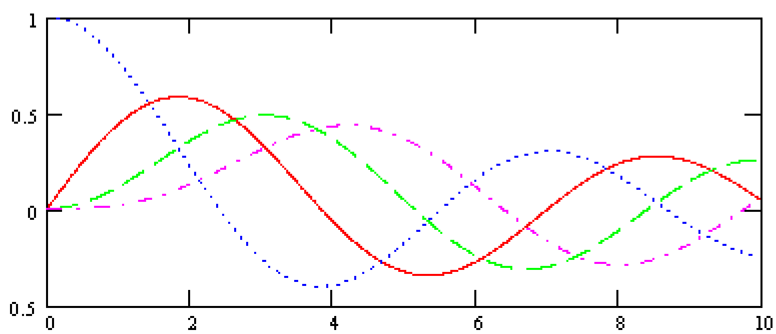


Figure 14 – Bessel function of the first kind of order n, shown for order = 0, 1, 2 and 4
X-axis is read as the modulation index k_f, and y-axis is the value of the amplitude of the associated harmonic.
Except for the 0 – order function, all others start at 0.0 and damp down with cycle.

Below we see a plot of what the cos and sin expansions (equations 22, 21) containing the various Bessel functions look like for just four terms.

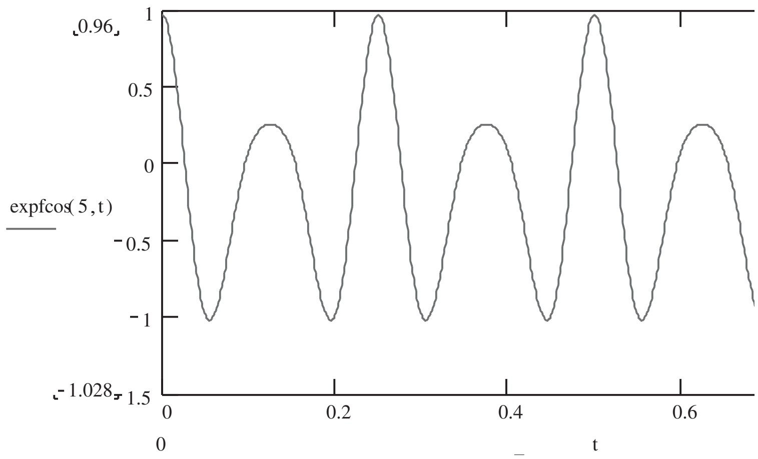


Figure 15 – Plot of $\cos(k_f \sin \omega_m t) = J_0(k_f) + 2J_2(k_f)\cos(2\omega_m t) + 2J_4(k_f)\cos(4$

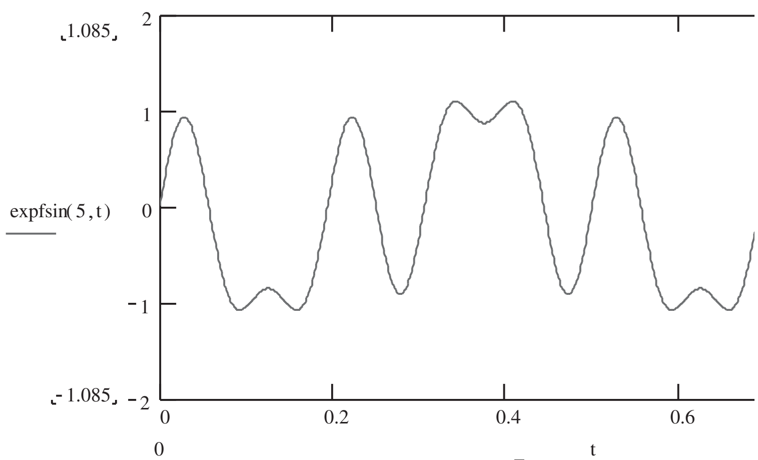


Figure 16 – Plot of $\sin(k_f \sin \omega_m t) = 2J_1(k_f)\sin(\omega_m t) + 2J_3(k_f)\sin(3\omega_m t) + 2J_5(k$

Now we put it all together (plug equations 21, 22 into equation 20)
And by trigonometric magic get, an expression for the FM modulated signal.

$$
\begin{aligned}
s(t) = {} & J_0(k_f)\cos(\omega_c t) \\
& -J_1(k_f)\left[\cos(\omega_c - \omega_m)t - \cos(\omega_c + \omega_m)t\right] \\
& +J_2(k_f)\left[\cos(\omega_c - 2\omega_m)t - \cos(\omega_c + 2\omega_m)t\right] \\
& -J_3(k_f)\left[\cos(\omega_c - 3\omega_m)t - \cos(\omega_c + 3\omega_m)t\right] \\
& +\ldots
\end{aligned}
$$

Remember this is a response to just one single solitary sinusoid. This signal contains the carrier with amplitude set by Bessel function of 0 order (the first underlined term), and sideband on each side of the carrier at harmonically related separations of $\omega_m, 2\omega_m, 3\omega_m, 4\omega_m, \ldots$. This is very different from AM, in that we know that in AM, a single sinusoid would give rise to just two sidebands (FFT of the sinusoid) on each side of the carrier. So many sidebands, in fact an infinite number of them, for just one sinusoid!

Each of these components has a Bessel function as its amplitude value. For certain values of k_f, we can see that J0 function can be 0. In such case there would no carrier at all (MSK) and all power is in sidebands. For the case of $k_f = 0$, which also means that there is no modulation, all power is in the carrier and all sideband Bessel function values are zero.

FM is called a constant envelope modulation. As we know the power of a signal is function of its amplitude only. We note that the total power of the signal is constant and not a function of the frequency. For FM, the power gets distributed amongst the sidebands, the total power always remains equal to the square of the amplitude and is constant..

Here we plot the above signal for $k_f = 2$ and $k_f = 1$.

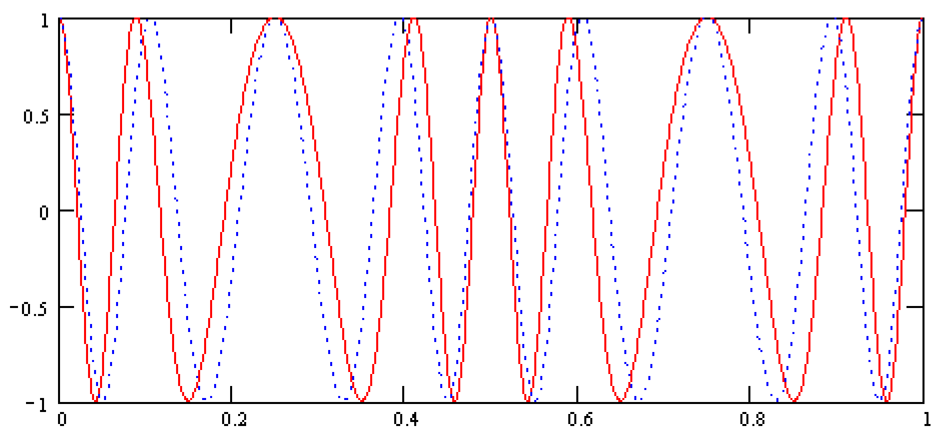
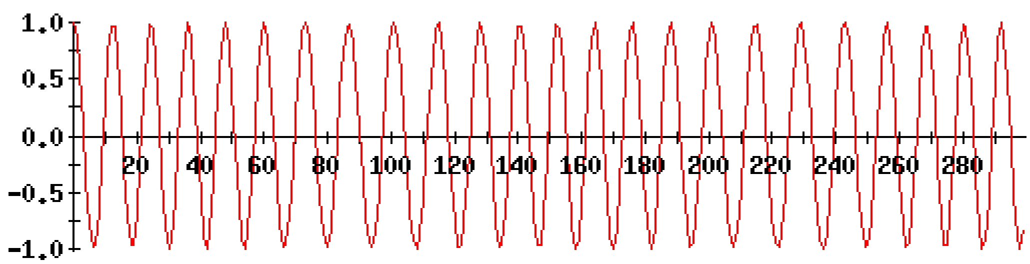


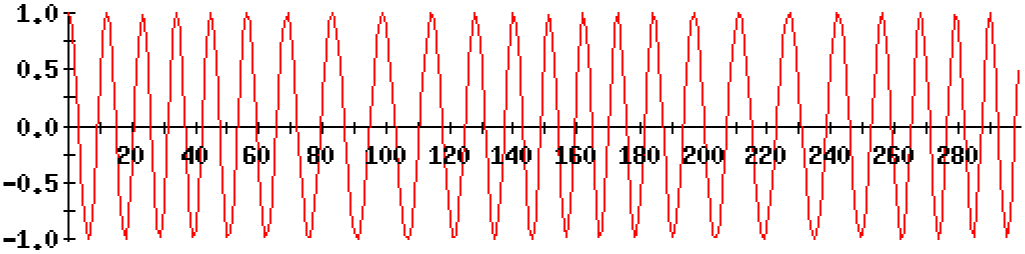
Figure 17 – Modulated FM signal for $k_f = 2$ and $k_f = 1$, the red signal is moving faster indicated higher frequency components .

Examples of FM spectrums to a sinusoid.

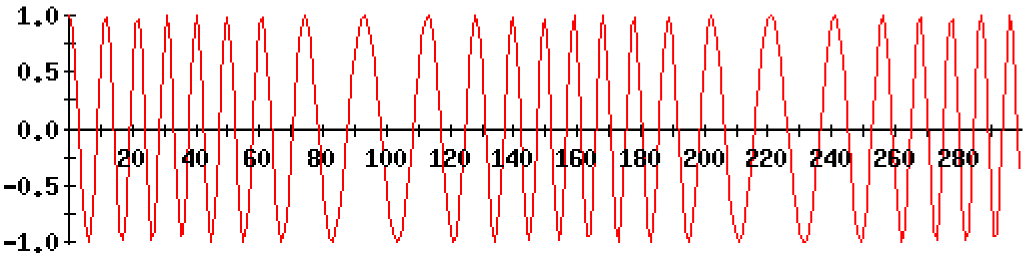
In the following signals, the only thing we are changing is the modulation index k_f. This increases the frequency separation (or deviation and increases the bandwidth.)



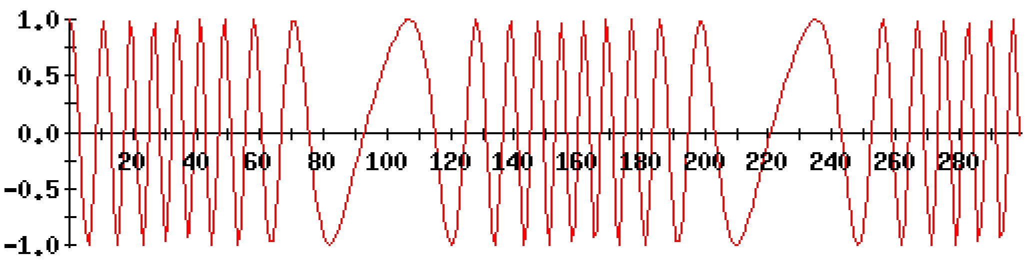
$k_f = 0.5$



$k_f = 1$

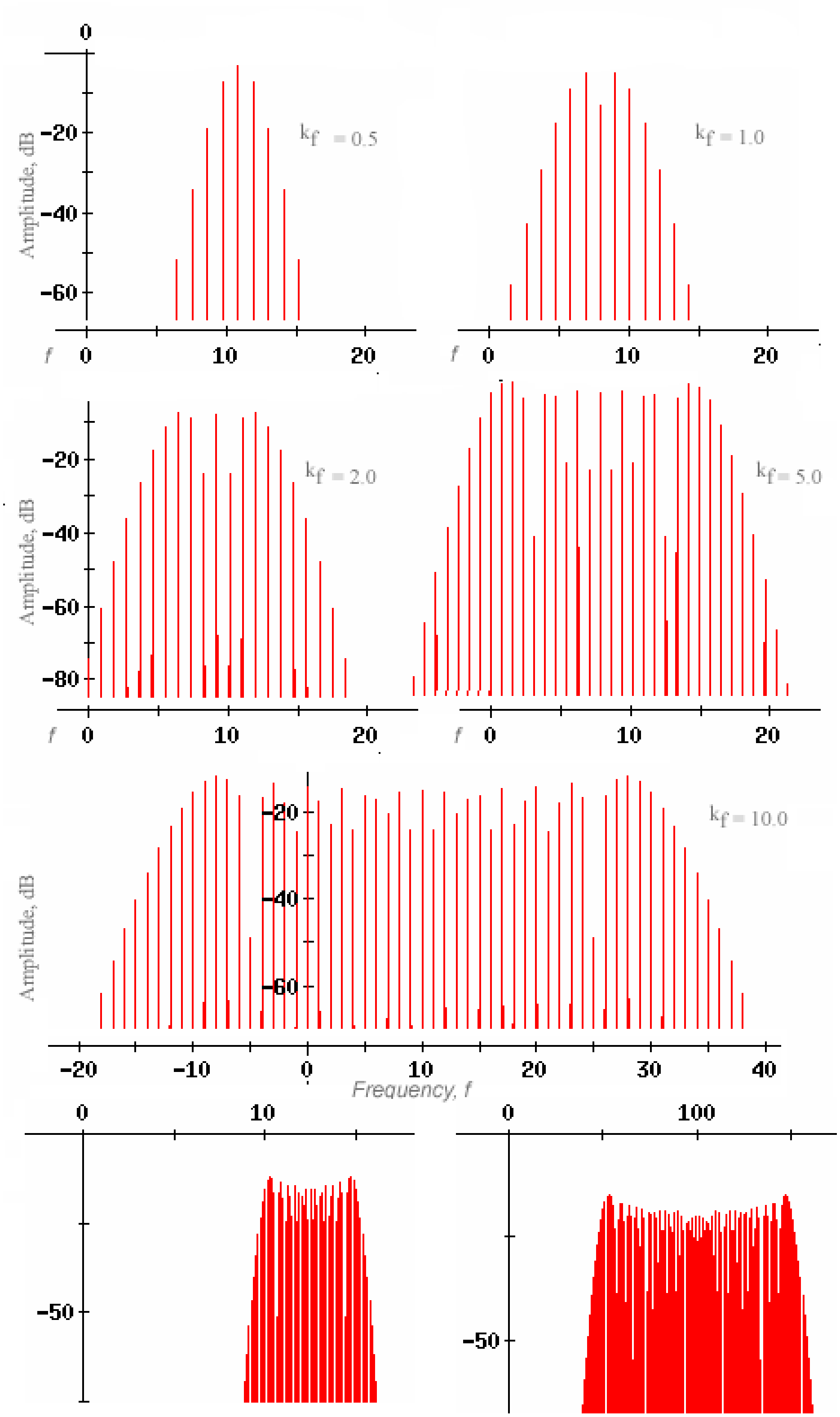


$k_f = 2$



$k_f = 4$

Figure 18 – Time domain FM signals for various k_f values.



Amplitude, dB
kf = 0.5
kf = 1.0
kf = 2.0
kf = 5.0
kf = 10.0
Frequency, f

Figure 19 – Spectrum of FM signals for various k_f values.

Bandwidth of a FM signal

The bandwidth of an FM signal is kind of slippery thing. Carson's rule states that is bandwidth of a FM signal is equal to

$$Bandwidth_{est.} = 2(\Delta f + f_m)$$

Here

$$f_m = \frac{2}{T_s}$$

or is equal to the baseband symbol rate, Rs.

If the frequencies f1 and f2 chosen satisfy the following equation, the cross-correlation between the carriers is zero, then this is an orthogonal set.

$$\rho = \int_0^{T_0} \cos(2\pi f_1 t)\cos(2\pi f_2 t)dt = 0$$

The output of I is maximum just when output of Q channel is zero. The decision at the receiver is a simple matter of determining if the voltage is present.

The BER of a BFSK system is given by

$$P_e = \frac{1}{2}\left[1 - erf\ \frac{1}{\sqrt{2}}\left(\frac{E_s}{N_0}\right)^2\right]$$

If however, if the cross-correlation is not zero,

$$\rho = \int_0^{T_0} \cos(2\pi f_1 t)\cos(2\pi f_2 t)dt \neq 0$$

then energy is present on both channels at the same time and in presence of noise, symbol decision may be flawed. Expanding above equation, we get,

$$\rho = \frac{\sin(\pi(f_2 - f_1)T_s)\ \cos(\pi(f_2 - f_1)T_s)}{\pi(f_2 - f_1)T_s}$$

Now set

$$(f_2 - f_1)T_s = \frac{4\Delta f}{f_m} = 4k_f$$

$$\rho = \frac{\sin(4\pi k_f)\ \cos(4\pi k_f)}{4\pi k_f}$$

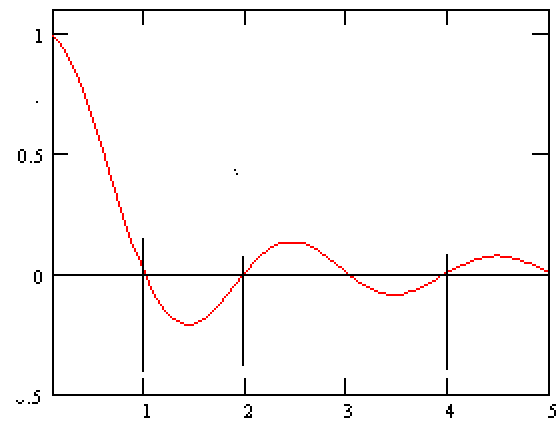


Figure 20 – Cross-correlation between I and Q FSK signals. The zero crossing points are the preferable points of operation.

Plotting this equation against the 2 times modulation index k_f, we get Figure 20. The y-axis which is correlation between I and Q channels is 0 at specific values of k_f. These values are 1, 2, 3 and 4 and so on.

The point with the lowest correlation (negative value is considered the most efficient place to operate an FSK. At first zero crossing, we note that the symbols are located only half a symbol apart so, coherent detection is not possible at this point. The next zero crossing, where kf = 1, is the Sunde's FSK. Coherent detection is possible at this point because symbols are at least ((f2-f1)Ts = 1) one symbol apart. The 4th zero crossing is MSK.

Bandwidth of Sunde's FSK is given by

$$B = (f_2 - f_1) + \frac{2}{T_s} = \frac{3}{T_s} \quad Hz$$

And for MSK,

$$B = (f_2 - f_1) + \frac{4}{T_s} = \frac{5}{T_s} \quad Hz$$

Sunde's FSK gives us the most spectrally efficient form of FSK that can be detected incoherently, whereas MSK gives us a spectrum that is a lot like QPSK but rolls-off much faster.

Questions, corrections?
Please contact me,

Charan Langton
mntcastle@earthlink.net
www.complextoreal.com