Charan Langton, Editor

SIGNAL PROCESSING & SIMULATION NEWSLETTER

Baseband, Passband Signals and Amplitude Modulation

The most salient feature of information signals is that they are generally low frequency. Sometimes this is due to the nature of data itself such as human voice which has frequency components from 300 Hz to app. 20 KHz. Other times, such as data from a digital circuit inside a computer, the low rates are due to hardware limitations.

Due to their low frequency content, the information signals have a spectrum such as that in the figure below. There are a lot of low frequency components and the one-sided spectrum is located near the zero frequency.

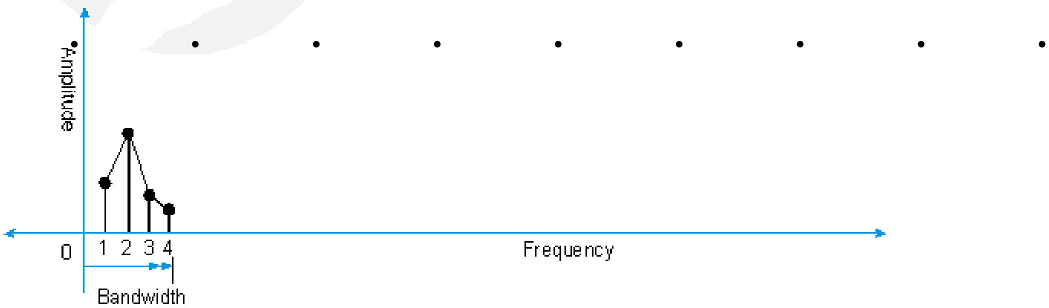


Figure 1 - The spectrum of an information signal is usually limited to low frequencies

The hypothetical signal above has four sinusoids, all of which are fairly close to zero. The frequency range of this signal extends from zero to a maximum frequency of f_m. We say that this signal has a bandwidth of f_m.

In the time domain this 4 component signal may looks as shown in Figure 2.

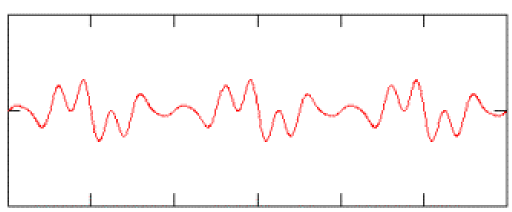


Figure 2 - Time domain low frequency information signal

Now let's modulate this signal, which means we are going to transfer it to a higher (usually much higher) frequency. Just as information signals are characterized by their low frequency, the transmission medium, or carriers are characterized by their high frequency.

The simplest type of modulator for nearly all modulation schemes is called the Product Modulator consisting of a multiplier or a mixer and a band-pass filter. Let's modulate the above signal using the Product Modulator, where m(t) is the low frequency message signal and c(t) is the high frequency carrier signal. The modulator takes these two signals and multiplies them.

$$f(t) = m(t) \times c(t)$$

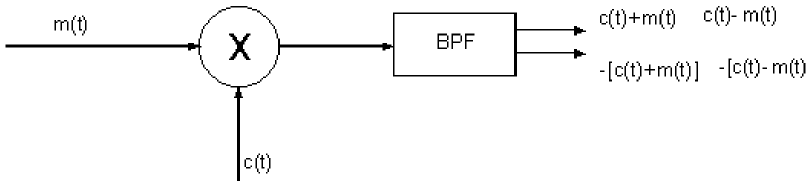


Figure 3 - A Product Modulator

The frequency domain representation of a Product Modulator or a mixer has a curious quality that instead of producing the products of the input frequencies which is what we really want, it produces sums and differences of the frequencies of the two input signals in both the positive and negative frequency domains. Is this a problem? The answer depends on what we want to do with the output. In most case if no non-linearity is present, we can predict exactly where these components will lie and we can filter out what we do not want.

What if the carrier frequency source in a product modulator is not perfectly stable? In this case, each deviation frequency will also produce its own sum and difference frequencies with the baseband signal. These are called spurs and are inherent to the mixer process. In addition phase oscillations of the carrier also affect the output. For this reason simple mixer modulators and demodulators do not work well and further complexity in form of phase lock loops etc. is introduced into the receiver design.

In Figure 4a, we see the two sided spectrum of the message signal. After mixing, modulating or heterodyning (all of these terms refer to the same thing), we get a spectrum such as in Figure 4b. The spectrum is now shifted up to the carrier frequency and we see that it is replicated on both sides of the y-axis.

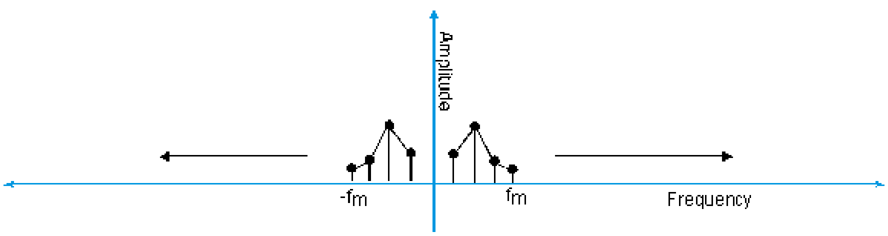


Figure 4a - the Baseband Spectrum

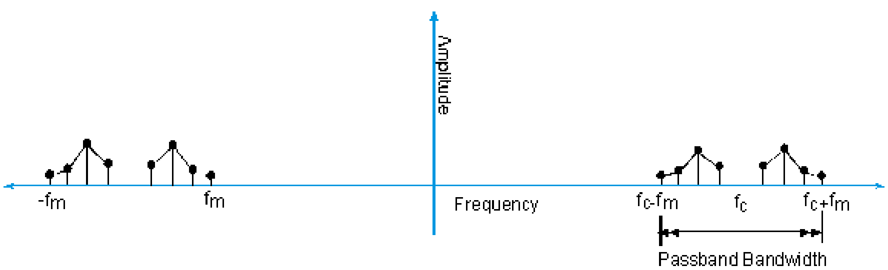


Figure 4b - the Passband Spectrum of the same signal

Another way to describe the process is that multiplication by a sinusoid, shifts one copy of the spectrum to f_c and an another to $-f_c$. Why does this happen? The reason is explained by the Fourier Transform of this signal which is a product of two signals, one of them a sinusoid.

$$f(t) = m(t) \times \cos w_c t$$

The Fourier transform of f(t) is just the Fourier Transform of the signal m(t), half of it shifted up and half of it down.

$$F(f) = \frac{1}{2}[\underbrace{M(f - f_c)}_{\text{Half is up shifted.}} + \underbrace{M(f + f_c)}_{\text{This half is down shifted.}}]$$

In Figure 4b the two-sided spectrum of the signal is shifted up to the plus and minus carrier frequency. The negative frequency twin on the other side of y-axis is usually no problem and can be easily filtered out by a real passband filter. And now we just work with half of the spectrum, usually the positive half recognizing that it has one-half the magnitude of the actual signal.

In time domain, we see that this signal has much higher frequency. But its envelope is still the original low frequency signal of Figure 2.

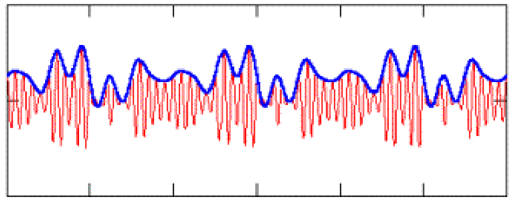


Figure 5 - Output signal of a product modulator, the envelope of which is the information signal (see also Figure 2)

Now we define some new terms.

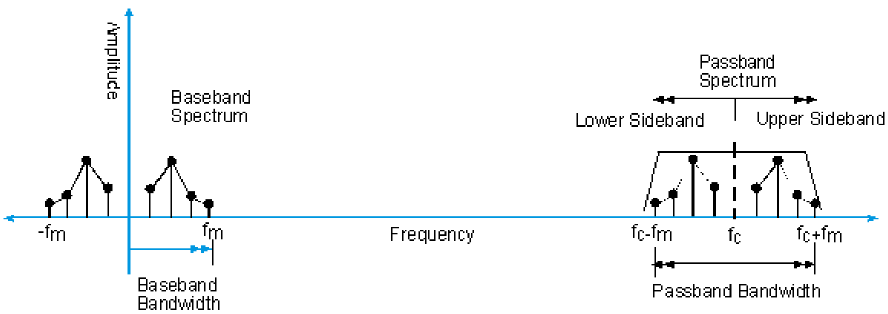


Figure 6 - Baseband becomes Passband by translation to higher frequency
The positive frequency spectrum becomes the upper side-band and the negative frequency spectrum become the lower side band.

Baseband Signal - The information signal is called the baseband signal. The bandwidth is always a positive quantity so the bandwidth of this signal is f_m.

Passband Signal - The multiplication of this signal with a sinusoid carrier signal translates the whole thing up to f_c. This signal is now called the passband signal. This signal extends in range from $(-f_c - f_m)$ to $(f_c + f_m)$. The new signal has doubled in bandwidth. The passband signal bandwidth is double that of the baseband signal.

The fact that the same signal has double the bandwidth in passband is often confusing. We think of bandwidth as something physical so how can it just double? The answer is imbedded in the question itself. In keeping with our concept of bandwidth as something real, we do not allow it to cross from the positive to the negative domain. It exists as a separate quantity on each side of the y-axis and does not cross it. There is no free lunch even in signal processing, so another simplistic way of considering this fact is that the passband signal contains not just the message signal but the carrier as well, so wouldn't you expect it to have a larger bandwidth?

Sidebands

Now note that in Figure 6, the passband spectrum has two parts (on each side of f_c) that are identical.

The upper part of the passband spectrum above the carrier is called the upper sideband and the one below is called the lower sideband.

We notice that since the passband spectrum is symmetrical (not only about the y-axis but also about the carrier frequency) the upper sideband is the mirror image of the lower sideband. Do we need the whole spectrum to recover the baseband signal? Perhaps we can get by with only half.

This intuitive observation is correct. We can recover the original information signal from just the upper band or the lower band. We do not need both halves.

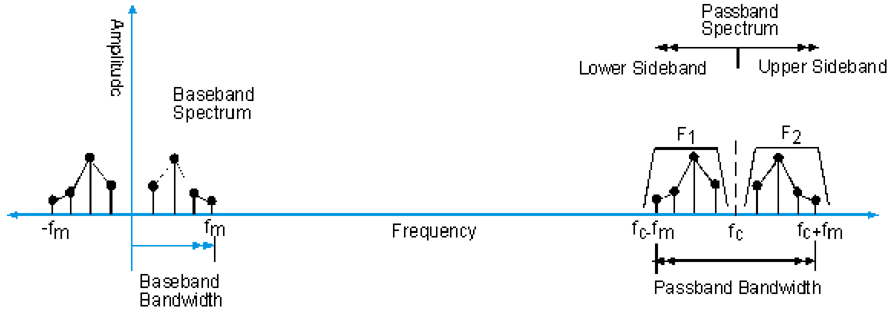


Figure 7 - Filter passband to for the upper and lower side-band as separate signals.

So can we just transmit only half of the signal? Can we figure out some way of transmitting an another signal in the rejected half? Then we can transmit two signals for the price of one!

This realization leads to the single and double side-band modulation techniques. In double side-band, we use the whole spectrum just as we show above. Both halves are used. In single sideband modulation, we filter out the lower or the upper band to separate out these signals as if they were two independent signals. Each half is enough to recover the signal.

Filter 1 and Filter 2 in Figure 7 do just that and show how we could transmit two signals in the place of one. Use F1 before transmitting, and you get only the lower side band, and use F2 and you get only the upper side band. We get two channels in place of one. Where ever bandwidth limitations exist, SSB is used. Most notable application is in telephony. Telephony signals have ideal characteristics for the use of SSB. There is very little signal content below 300 Hz so the SSB signal does not suffer much distortion. Also telephone signals are bandwidth-limited, and SSB maximizes bandwidth usage. HAM radio and HF communications is one area where the Single Side Band (SSB) modulation is used to this advantage.

Amplitude Modulation

We have already discussed much of the building blocks of Amplitude Modulation as SSB is a form of Amplitude Modulation. The simplest form of Amplitude Modulation is the Double Sideband Modulation.

Double Side Band Modulation

Let's take the information signal m(t). The output of the mixer gives us

$$c(t) = m(t)\ A_c \cos(2\pi fct)$$

now we add to this signal the carrier (the second term).

$$c(t) = m(t)\ A\cos(2\pi fct) + A\cos(2\pi fc\ t)$$

Now instead of transmitting just the signal times the carrier, we add the carrier to the to the product. The block diagram of this, called the AM product modulator, would look like this.

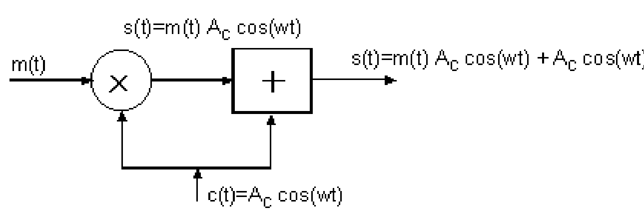


Figure 8 - A basic AM modulator, its output contains the modulated signal and the carrier

What is the Fourier Transform of this signal?

$$C(f) = \frac{1}{2}[\underbrace{A_c M(f - f_c) + A_c \delta(f - f_c)}_{\text{Spectrum on the positive x-axis}} + \underbrace{A_c M(f + f_c) + A_c \delta(f + f_c)}_{\text{Spectrum on the negative x-axis}}]$$

(We are using properties of the Fourier Transform here; the first term comes from the fact that the FT of a signal, multiplied by a cosine is just the same

spectrum shifted, and the second term is just a delta function times the amplitude of the original carrier. Fourier Analysis is the absolute fundamental of all signal processing and I suggest reading tutorials 6 and 7 so you are clear on the main concepts. You are welcome to email me your questions.)

Here is the spectrum of this signal.

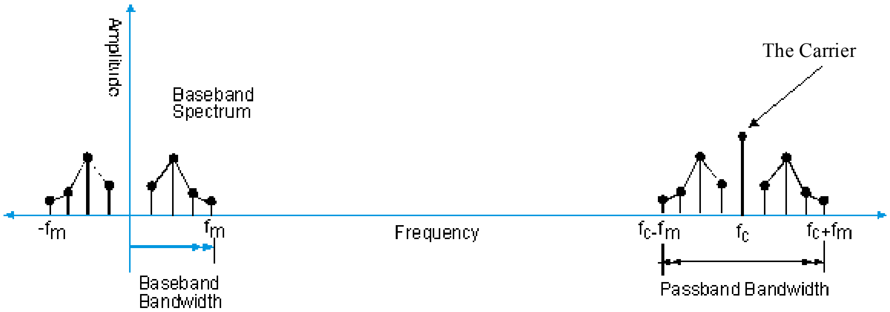


Figure 9 - Double Side Band Modulation Spectrum of received Signal
(Note only the positive side of the spectrum is shown.)

Now you see the carrier signal pop up in the middle of the spectrum. We can put a filter around this signal and recover the carrier at the receiver. This is then fed to the demodulation circuitry later.

This modulation is called Double Side Band (DSB) modulation. It is the most basic form of the AM modulation. From here on, we can do a variety of things such as suppress the carrier, use one band or the other etc.. All of these are variations of the Double Side Band (DSB) Amplitude Modulation.

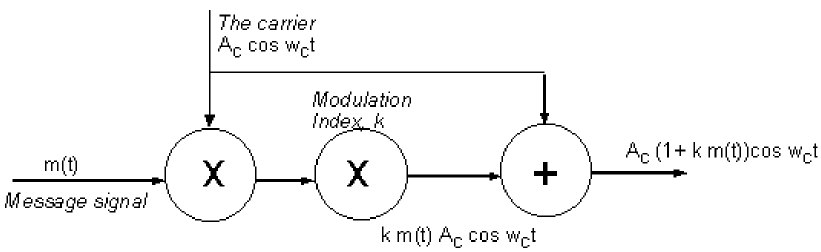
We can rearrange terms to write the amplitude modulation equation as

$$c(t) = A_c(1 + km(t))\cos w_c t$$

By varying the amplitude of the carrier vs. the amplitude of the information signal, we can create different looking waveforms. As long as certain parameters are not exceeded, the envelope of this signal would look like the information signal and using an Envelope Detector (demodulation) we can recover this signal.

In above equation, quantity A^c represents the power of the modulated signal. Both the carrier and the message signal are assumed to have normalized amplitude. The quantity $|km(t)|$ is called the modulation index of the signal. The index effects how the received signal looks. Modulation index larger that 100% distorts the signal so an envelope detector can not be used to demodulate it any longer.

The following figure shows how we might create this signal.



The following two figures show the effect of the modulation index on the received AM signal.

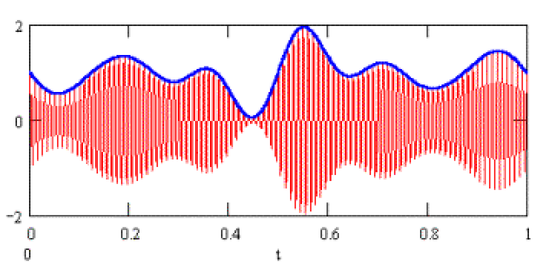


Figure 10a - DSB Modulated signal with Modulation Index = 100%
Note that the envelope of this signal is the same as the baseband signal.

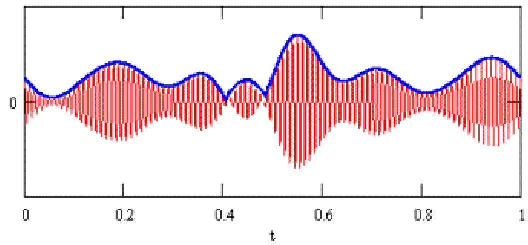


Figure 10b - DSB Modulated signal with Modulation Index = 120%
Note that the envelope of this signal is not the same as the baseband signal.

As long as the modulation index is less than 100%, the envelope of the signal can be used to remove the information signal. For index greater than 100% as shown in figure above, the envelope detector will no longer be able to correctly detect the signal. We see that the envelope in the lower figure is no longer a copy of the original signal in Figure 2.

Standard DSB Modulation is used in AM Radio broadcasting. It offers the advantage of using a simple receiver based on a Envelope Detector.

Double Side Band - Suppressed Carrier

We just added the carrier, but now we realize that it actually takes a lot of power to include the carrier and perhaps it makes no sense to do that after all. But we want to somehow include the carrier information but without actually doing so. And we want to use the envelope detector as the receiver. How can we do that?

We rely on the symmetry of the signal spectrum now. Consider a modulation scheme called the Double Side Band - Suppressed carrier, or DSB-SC modulation, everything is same as DSB except that no carrier in included. DSB-SC signals are created by a modulator called the Balanced Modulator. The following figure shows the basic block diagram of a Balanced Modulator.

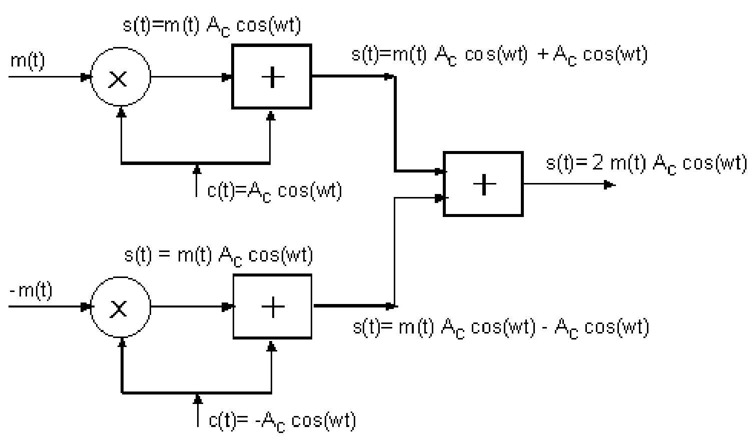


Figure 11 - A balanced Modulator results in suppression of the carrier

This balanced modulator is basically two product modulators added together. The input to one is a negative information signal and a negative carrier. The product of this modulator when added to its positive counterpart results in canceling the carrier as we can see in the output. (The math above is quite straightforward and worth checking for that wonderful feeling that comes when you really understand something.)

Envelope detector can not be used with DSB-SC carrier because the envelope of the DSB-SC signal is not the same as the baseband signal. A more sophisticated modulator is needed with this signal.

The DSB-SC modulation is identical to BPSK, which we will discuss later.

Generating Single Side Band (SSB) signals

In essence the SSB transmission that we discussed before is a bandwidth conserving technique. The most notable point of SSB is that the SSB passband signal and the baseband signal occupy the same bandwidth, so cutting spectrum needs in half.

How do we create a SSB signal? There are two main ways that SSB signals can be generated.

1. Filtering the unwanted side-band
2. Phasing Method

The simplest solution would be to just take the DSB-SC signal and filter the unwanted band before transmission so that the unwanted side is not sent at all as shown in the figure below. By keeping only the part shown, we have gotten rid of all the other images, all of the negative components and the upper side-band.

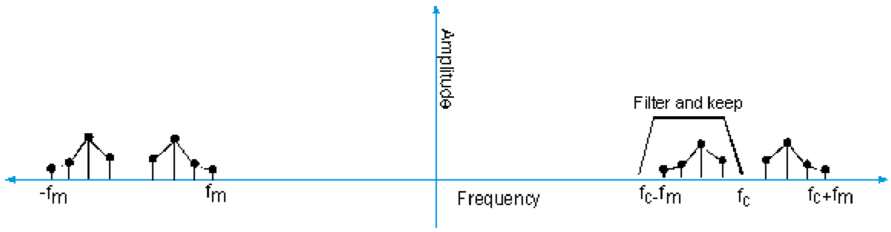


Figure 12 - A passband filter after DSB-SC modulation results in getting rid all but one band.

Problem with this method is that it is hard to build practical filters with steep enough cut-offs at high frequencies. Such a filter ends up distorting the desired signal as well as including some of the unwanted side-band anyway.

The second method involves the use of Hilbert Transform and the Analytic signal we talked about in the last Tutorial. As a way of review, the figure below shows the baseband spectrum of our signal. The second part shows the Hilbert Transform of the same signal. (Recall that the Hilbert Transform rotates the positive frequency components.)

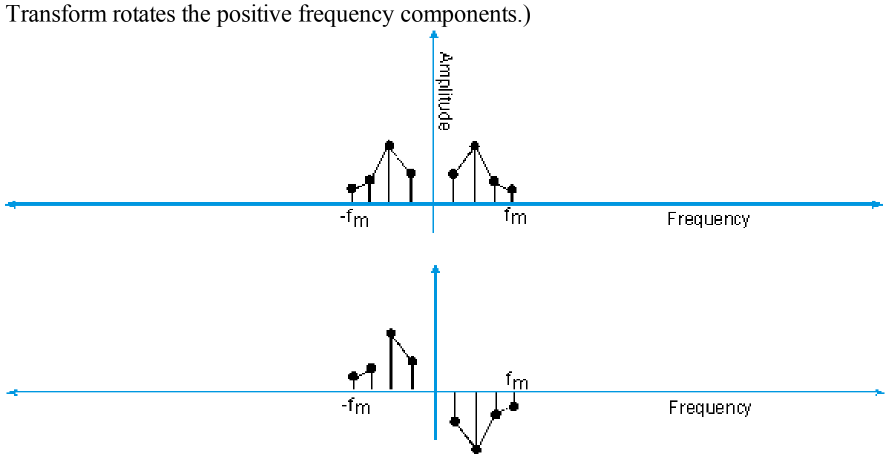


Figure 13 - a. Baseband spectrum (symmetric about the y-axis) b. Hilbert transform of the same signal (antisymmetric about the y-axis)

Now let's take this signal and modulate it up, we get

$$m(t)\cos(\omega_c t)$$

Now let's take the Hilbert transform of this signal and modulate it by a sine wave, so we get

$$\hat{m}(t)\sin(\omega_c t)$$

Now we create a carrier which is the sum of these two parts.

$$c(t) = m(t)\cos(\omega_c t) \pm \hat{m}(t)\sin(\omega_c t)$$

The signal Its Hilbert Transform

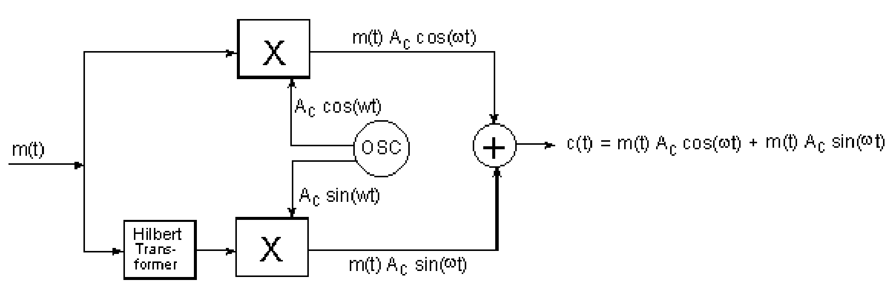


Figure 14 - SSB Modulator using the Phasing method

The SSB signal created in this way is essentially two signals in quadrature. The combination gives us the equation for the SSB signal. By changing the sign of the analytic signal, we can create either the upper sideband or the lower.

$$c(t) = m(t)\cos(\omega_c t) \pm \hat{m}(t)\sin(\omega_c t)$$

Now let's take the Fourier Transform of each part. The Fourier Transform of the first part is

$$\tfrac{1}{2}\left(M(\omega-\omega_c)+M(\omega+\omega_c)\right)$$

The Fourier Transform of the second part is

$$\tfrac{1}{2j}\left(M(\omega-\omega_c)(-j)+M(\omega+\omega_c)(j)\right)$$

(the presence of j is due to the Hilbert transform, see Tutorial 7)

Figure below shows the two spectrums and we see at once that adding these two representations give us a nice clean signal with only one side band, upper or lower as we desire.

Thanks to Dr. Hilbert and his analytic signal there is nothing to filter, just a clean single band.

Another interesting fact is that the sum of the two side bands give us the DSB-SC waveform.

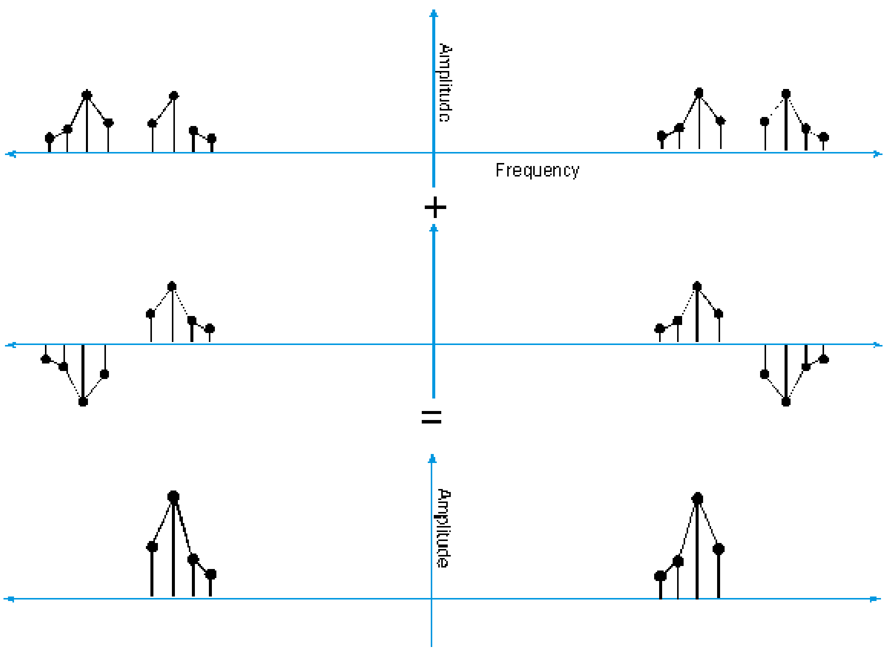


Figure 15 - a. Spectrum of part one, b. spectrum of part two, d. the sum of these two gives us the lower side band, the difference would give the upper side band.

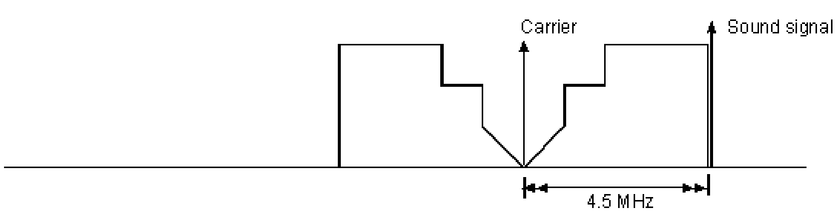
AM Modulation and Video broadcasting

Vestigial Sideband Modulation

A variation of DSB is used for broadcast TV. Under the FCC requirements, the standard video signal occupies a bandwidth of 4.5 MHz. The sound signal is separate and is transmitted at the upper edge of this signal. When carrier is shifted to bandpass, this one sided bandwidth becomes 9 MHz. This is nearly ten times as large as the total bandwidth occupied by all the channels of the AM radio. Use of SSB modulation would cut this in half but SSB is not used for video signals because of the complexity of the SSB receivers. TV manufacturers particularly American companies were instrumental in setting these standards like to keep the cost of the TV's as low as possible so SSB receivers are not used.

A modulation technique used for commercial video broadcasting which lies some where in the middle of SSB and DSB is called theVestigial sideband Modulation (VSB).

In figure below we show a hypothetical bandpass video signal. The sound signal which is sent separately is at the upper edge of the spectrum.



a.Video signal DSB Spectrum

Figure 16a - Video Signal bandwidth

In Figure b, we show a peculiar kind of filtering of this video signal that takes place after modulation with a carrier but before transmission.

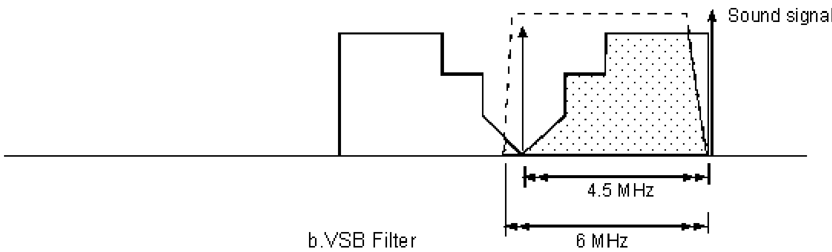


Figure 16b - Vestigial Filter

This filter takes in a small part of the upper edge of the lower sideband, starting from -1.25 MHz. The signal is attenuated in this range from -1.25 MHz to -.75 MHz. From here on to 4 MHz, the signal is transmitted full strength. At 4 MHz it is once again attenuated down to 4.5 MHz so as not to interfere with the sound carrier which is demodulated separately. The shaded portion is what is transmitted.

The term vestigial is used since a tiny trace part of the lower sideband is also included in the transmission. The net result is that instead of transmitting a 9 MHz signal, we transmit only 6 MHz, the standard video signal today.

Unlike voice signals which have no components near the zero frequency, Video signals are very sensitive to their low frequency content. Distortion in these components degrades the picture. So extra care has to be taken to make sure that all the low frequency components (which are located in the center) are transmitted without distortion. VSB modulation transmits these low frequencies at the twice level. The motivation for filtering the signal in this way also comes from the desire to use a diode demodulator which requires an explicit carrier. But to recover the carrier we need to go a little to the other side of the carrier frequency and take in an attenuated part of the signal because of the limitations of practical filters. The development of this filter was a function of a compromise between bandwidth and the TV receiver complexity.

The new HDTV standard is also based on VSB.

About Amplitude Demodulation

Product Demodulator

All AM signals discussed here, DSB, DSB-SC, SSB and VSB can be demodulated using a product demodulator. In principle it is the reverse of the modulation process. We take the incoming signal, which now also includes noise and we multiply it by a known carrier. The product obtained is then low pass filtered and what remains then is the information signal.

The main problem with the product demodulator is that the carrier phase is not known. We do not know if the starting phase was 30° or 45° or 90° or some other number. For some signals this is not such a big problem. An audio signal can be demodulated incoherently which means that the phase of the carrier at the receiving end is not synchronized with the transmitter. In radio AM broadcasting we can get away with ignoring the phase because our ears are not very sensitive to phase deviations of the signals. We can hear and understand the signal just fine. In such cases, an incoherent product demodulation makes sense and would be the cheapest solution.

Now if we are sending data, this is indeed a big problem and we need to exactly recover the phase of the transmitted carrier. Even video signals are not forgiving of phase errors. Phase information for nearly all signals except, telephone and radio signals is considered very important.

There are two methods of making sure that we know the phase of the incoming signal; 1. The Costas loop and 2. The phase locked loop. Both are variations of a technique to find and lock on to the phase (we will discuss these in another tutorial in detail.) This variation of the product demodulation where we make a special effort to determine the phase of the transmitted carrier is called coherent demodulation.

Square-Law Demodulator

Non-linearity usually has a bad name in communications. We don't like it because it distorts the signal and produces unwanted products. But here is a way non-linearity is actually used to advantage in demodulation of some AM signals.

Let's take a non-linear device with the following behavior.

$$y = kx^2$$

Now let's take an amplitude modulated signal

$$g(t) = A_c[1 + m(t)]\cos\omega_c t$$

Putting this through the above non-lienearity, after some manipulations and clever trigonometric substitutions, we get

$$y(t) = g^2(t) = \underbrace{kA_c^2}_{\text{DC term}} + \underbrace{[kA_c m(t) + \frac{k}{2} A_c m^2(t)]}_{\text{Information signal}} + \underbrace{\cos 2\omega_c t[kA_c + \frac{k}{2} A_c^2 m^2(t)]}_{\text{Signal at } 2w_c}$$

Now throw away the DC term, filter out the terms at two times the frequency and what we have left is

$$y(t) = [kA_c m(t) + \frac{k}{2} A_c m^2(t)]$$

The term $m^2(t)$ is not a big problem if the modulation index is small. This term disappears and for audio broadcasting this term makes no discernible difference.

One by-product of this method is that if no carrier is included, we can still recover the carrier. This technique can also be used to recover the carrier. Take a signal

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

squaring it gives

$$= \frac{1}{2}[m^2(t) + m^2(t)\cos(2\omega_c t)]$$

The second term is the carrier at twice its frequency which we recover by filtering at this frequency.

Figure 17 - Non-linearity used to recover the carrier

Envelope Detector

The envelope of a signal is its maximum value over a set sampling period. A diode circuit used most often to detect the envelope of AM signals is the simplest and the universal method of demodulating AM signals. The prerequisite for the use of this demodulation method is the presence of a strong carrier and high SNR. Excessive amount of noise causes severe envelope fluctuations and makes this method less effective. We all know of the AM radio's vulnerability to noise and other atmospheric perturbations.

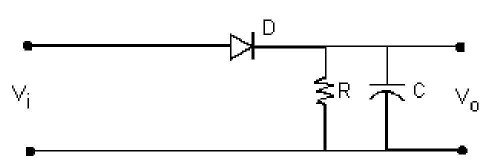


Figure 18 - RC-Diode Circuit used for Envelope Demodulation

The envelope detector is basically a Diode-RC circuit as shown above. The signal is applied to the terminals of the circuit. The Diode conducts as the voltage(amplitude) increases and the capacitor charges up. Now as the voltage begins to go down, resistor discharges and the capacitor lets go of its charge. The cycle continues and each charge of the capacitor indicates the maximum value over that period. In fact the capacitor discharges slightly between cycles as shown in the figure below but this can be compensated for easily.

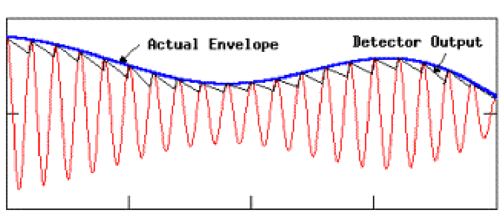


Figure 19 - Envelope Detection up-close

Summary

Baseband Signal - The baseband signal is usually the message signal. It has a bandwidth of B. See Figure 20.

Passband Signal - The passband signal is one that has been multiplied by a carrier. It is centered at the carrier frequency and has a bandwidth of 2B.

Double Sideband - When both sidebands and the carrier is transmitted, this is called the AM or DSB modulation. DSB signals which are passband signals have a bandwidth of 2B.

Double Sideband - Suppressed Carrier - When we remove the carrier to conserve power, the DSB signal is called the DSB-SC signal. It has a bandwidth of 2B.

Single Sideband - When either by filtering or phasing only one band is transmitted the signal is called SSB. It has a bandwidth of B.

Vestigial Sideband - VSB is used for video broadcasting. VSB is a compromise between SSB and DSB and has a bandwidth of .666B.

AM demodulators - There are three main types of AM demodulators or receivers. Envelope Detector is the simplest and senses the maximum amplitude of the in coming signal which happens to be the message signal. The Product Demodulator is next in complexity and is used for nearly all AM signals. Costas or Phase locked loops are used when phase is important. Squaring Demodulator is often used to recover the carrier as well as for demodulation of DSB-SC signal.

Figure 20 - follows

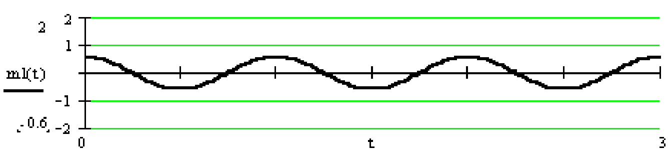
Charan Langton, Nov 4, 1998
Previous Tutorials are kept at the Advanced Systems Web site under CAP.

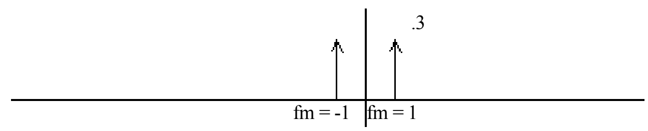
Thanks much to Eric Arakaki and Dave Watson for their invaluable comments and edits.

Figure 20 - AM Waveforms

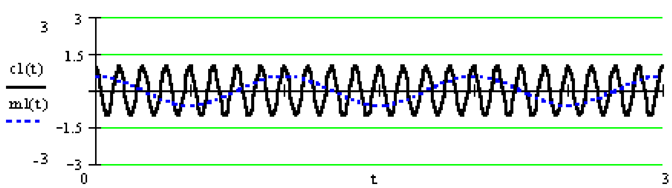
1. **Message Signal** $ml(t) = .6\cos(wt)$

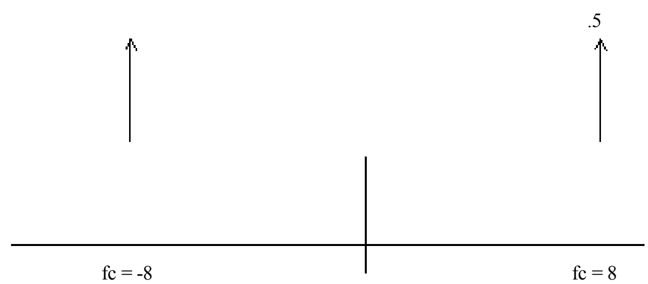
Signal Spectrum



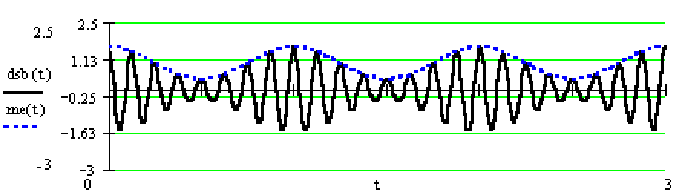


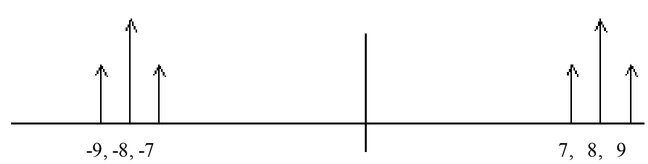
2.
Carrier Signal $c1(t) = \cos(8wt)$





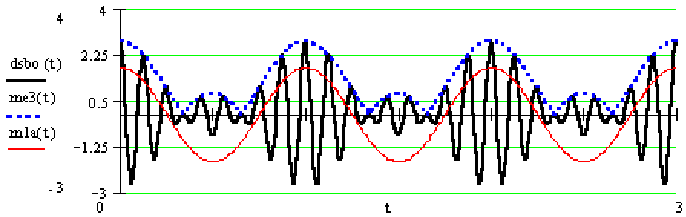
3.
DSB Waveform $dsb(t) = (1 + k\,ml(t))c1(t)$ **k = 60%**

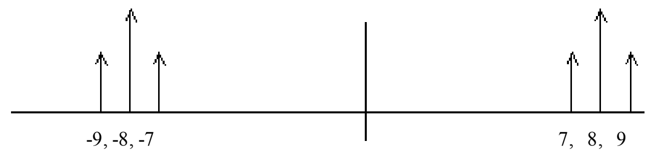




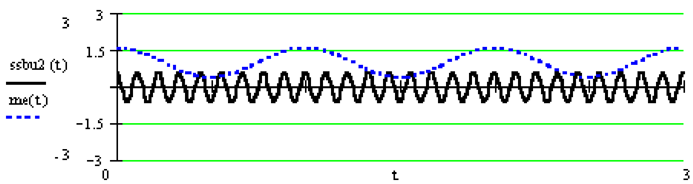
4. **DSB waveform Overmodulated** **k = 150%**

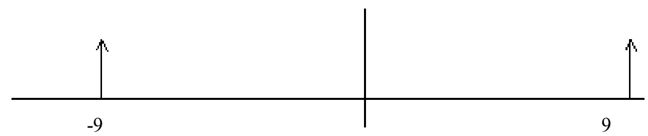
Note the envelope of the signal is no longer same as the baseband signal trace, hence there is no way to demodulate it from the envelope of this signal.



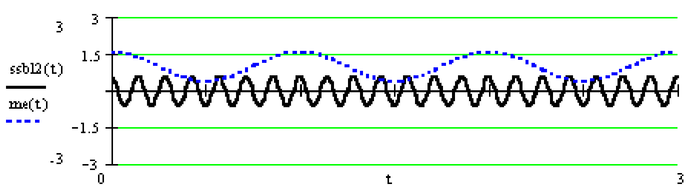


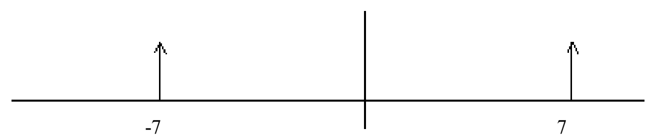
5. SSB - Upper sideband $ssbu(t) = .6\cos(9wt)$





6. SSB - Lower Sideband $ssbl(t) = .6\cos(7wt)$





7. **DSB-Suppressed Carrier** $dsbsc(t) = 2c1(t)m1(t) = ssbl(t) + ssu(t)$