

Online Radio & Electronics Course

Reading 29

Ron Bertrand VK2DQ
<http://www.radioelectronicschool.com>

AMPLITUDE MODULATION

Welcome to this reading, this is perhaps the first reading where we really get into some real radio stuff.

We're going to be talking about amplitude modulation. This is the first time we will be discussing how radio is used to transfer information across a long-distance in virtually no time at all. Therefore, it is my intention to cover the concepts completely without the use of unnecessary mathematics. Please do not think that without all of the mathematics we are shortchanging the subject. I have work with many trainees, who have come straight out of university, and when I have asked them about amplitude modulation they reel off a whole set of mathematical equations in an attempt to explain it. However, when I have asked what it *really is* I have often received a blank look for an answer. Mathematics alone does not *always* explain everything in such a way that you really get a *feel* for the subject. I hope that during this reading you will really gain an understanding of what radio transmission and amplitude modulation is really all about. The reading may be verbose and repetitive, but this is done to help you understand the important *real* concepts.

Firstly, let's have a look at what radio transmission is really all about. It's fine to have a microphone to speak into and expect to be heard somewhere else in the world, however what are the real principles involved. This does not apply just to amplitude modulation but to all types of modulation.

Modulation - in radio communications means combining signals, one of which is normally the information we want to transmit (which is usually a low frequency such as audio or voice), with a radio frequency.

Radio communication uses the ability of an electromagnetic wave (or if you like electromagnetic energy) to transfer information from one point to another. It is very easy to create an electromagnetic wave, send this wave to an antenna, and have this wave propagate over a long distance at the speed of light. It's another matter to have the electromagnetic wave contain useful information such as voice, or some other form of intelligence, which after all is the main objective of radio communication.

Remember that electromagnetic radiation takes many forms. Light is electromagnetic radiation. The infrared signal from your TV remote control is electromagnetic radiation. The heat given off by a bar radiator is electromagnetic radiation. The whole secret of radio communication is to place information onto the signal, to convey intelligence or a message from one point to another.

Amplitude modulation (AM) is just one method of doing this, and is really the first method ever used to transfer voice information from one place to another. There are many misconceptions about how this is done, and what AM really is. The first thing that should be understood is that voice frequencies range from about **50 Hz to about 3000 Hz**. Would it not be simple if we could just talk into a microphone, convert the sounds from our voice

to electrical energy, feed to an antenna, and have it transmitted to anywhere we wanted. **The problem is that voice frequencies are not high enough in frequency to be radiated by an antenna.** This will not work. Voice frequencies will not be radiated as voice frequencies are not electromagnetic radiation. The basic principle involves converting voice frequencies, without losing any intelligence, to a higher frequency, radiating them via an antenna, and at the other end converting the electromagnetic radiation back to voice frequencies so that they can be heard.

Again, the basic principle is to move voice frequencies to radio frequencies, radiate them from an antenna, have them propagate at light speed, and at the other end recover those voice frequencies.

The basic principle of amplitude modulation is to take voice frequencies, and mix (or modulate) them with a radio frequency signal so that they are converted to radio frequencies which will radiate or propagate through free space.

It is very easy to convert low frequencies to high frequencies so that they will radiate. All we need to do is take our voice frequencies (say from a microphone), then **mix** them with a radio frequency carrier and convert them to radio frequencies. The term **"carrier"** is very misleading, it is simply a sinusoidal high frequency radio signal which we use to convert voice frequencies to radio frequencies for the purpose of radiation. I have often heard and read that the carrier 'carry's the voice frequencies' - this is a misconception.

Rather than talk about the whole range of voice frequencies from 50 Hz to 3000 Hertz, we will look at the subject of amplitude modulation as if we were trying to transmit a single audio frequency of say, 1000 Hz, to a distant location. Using one audio frequency rather than the entire voice frequency range just simplifies our explanation.

FREQUENCY CONVERSION

The method used to convert one frequency to another is quite simple. You take the signal that you want to convert (in our case a 1000 Hz audio frequency) and a radio frequency signal, and we are going to use a radio frequency of 1 MHz to illustrate how 1000 Hz can be converted to a radio frequency for transmission. All that is required to do this is to feed the 1000 Hz signal and the 1 MHz signal into a **non-linear device**. The output of that non-linear device will contain our converted frequencies, some of which we want and some of which we don't. **In radio transmission the non-linear device is called a *modulator*, in radio reception the non-linear device is called a *mixer*.**

The apparatus we need is an oscillator for the radio frequency source, a mixer (modulator), and our audio to be frequency converted. The old name for this combination (mixer and oscillator) was called more clearly a **frequency changer**.

Remember, we are using a 1000 Hz audio frequency for the purposes of simplification. In practice, we would be using the whole range of audio frequencies from 50 Hz to 3000 Hertz and the same principles would apply.

Now then, we want to transmit a 1000 Hz audio frequency using radio frequency transmission.

We take the 1000 Hz audio frequency from the microphone, and the radio frequency (badly called the carrier) and inject them into a mixer or a modulator (which are just non-

linear devices) - the output of the non-linear device will contain our wanted frequencies. The output of the non-linear device will contain the original frequency that is 1000 Hz audio, the 1 MHz radio frequency, and also the sum and difference of these two frequencies. The output of the non-linear device will contain a 1 MHz radio frequency signal, a 1.001 MHz signal (ie. sum) and a 0.999 MHz signal (ie. difference). There will be other mixing products produced as well, but these will be eliminated in the following stages by filters. The 1.001 MHz signal is called the upper side frequency. The 0.999 MHz signal is called the lower side frequency. If we were using the whole range of audio frequencies from 50 Hz to 3000 Hertz, then we would have an upper side band (USB) and a lower side band (LSB) of frequencies.

WHAT IS AMPLITUDE MODULATION?

In amplitude modulated (AM) systems, the modulation audio is applied to the radio frequency carrier in such a way that the total power of the transmitted wave is made to vary in amplitude, in accordance or in sympathy with the power of the modulating audio.

It is a popular misconception that in an amplitude modulated system the carrier power (in our case the 1 MHz signal) is made to vary with the application of the modulated audio. Contrary to this, it is in fact the total wave power which varies in amplitude, and not the carrier power. The carrier power in an amplitude modulated system remains constant.

You might ask, "Why modulate the audio frequency at all" (yes, I know I am repeating myself). If audio signals are fed to an antenna, radiation will not occur, because the frequencies are too low to radiate. When an alternating current is made to flow in an antenna, a system of alternating electric and magnetic fields is developed in the vicinity of the antenna. With each 180 degrees of alternation, these fields will return the energy to the antenna by collapsing into it. At high frequencies the fields do not have time to collapse into the antenna and are left stranded in the space about it. This is called a free space field. Successive free space fields exhibit a repulsive force on their predecessors causing the fields to move out from the antenna at the velocity of light. This mechanism of radiation begins to occur at around 15 kHz.

The function of the carrier in amplitude modulation is simply to provide a signal to heterodyne (mix) with the modulated audio, to convert all the audio frequency components to a higher frequency so that the mechanism of radiation will occur. The carrier contains no intelligence. The carrier can be removed before transmission, as is the case with single side band (SSB).

One possibility for the misconception of the modulated audio riding on top of the carrier may have arisen from the pattern that an AM signal produces on the screen of a cathode-ray oscilloscope (the oscilloscope is to be discussed). The pattern produced does appear to support the fallacy. The oscilloscope is limited in its operation by the fact that it can only display the resultant instantaneous voltage amplitude of all the signals present on its deflection plates.

When discussing AM, it is important to realise that the term refers to several modes of transmission wherein the total wave power transmitted is made to vary in accordance with the applied modulating audio. Double sideband (DSB), single side band (SSB), vestigial side band, are common modes used on all amateur and commercial bands. All of these are correctly referred to as amplitude modulation. The reference to double side band as AM and single side band as SSB on some amateur transceivers gives credence to the

misconception that SSB is not an amplitude modulated wave when in fact it is.

HOW IS DOUBLE SIDE BAND PRODUCED?

Regardless of the type of AM system, the principle is the same. The radio frequency carrier is combined with the modulated audio in a non-linear device. The output of the non-linear device will contain the sum and difference frequencies of the original signals (with the exception of balanced mixers used for SSB).

WHAT DETERMINES THE BANDWIDTH OF AN AM (DSB) SIGNAL?

Bandwidth: means the signal width in hertz, kilohertz, and in some cases megahertz.

The bandwidth of an AM signal is equal to twice the highest modulating audio frequency.

As the highest voice frequency that needs to be transmitted (for acceptable intelligibility) is 3 kHz, and the bandwidth of a double side band signal is twice the highest modulating frequency, a double side band amateur transmission must therefore have a bandwidth of: $2 \times 3 \text{ kHz} = 6 \text{ kHz}$.

Many modern transmitters have audio filters that slice off some of the voice frequencies above 2.8 kHz making the bandwidth even narrower. **The bandwidth is not dependent upon the power of the modulating audio** (provided the transmitter is not over modulated). In commercial AM broadcast transmitters, the frequency of the modulating audio is permitted to be as high as 4.5 kHz, and since a double side system is used, commercial stations therefore have a bandwidth of 9 kHz.

HIGH AND LOW LEVEL MODULATION

When the modulating audio is applied to the carrier in a low power stage (that is, before significant amplification), this is said to be low level AM.

On the other hand, **if the carrier is amplified to a high power level** before the modulating audio is applied, then the system of modulation is called high level AM.

With low level modulation the double side band (DSB) signal is produced at low power levels. The advantage here is that the audio power required is low, no high power audio frequency amplifier is required. If it is necessary to raise the power of the DSB signal (and it usually is), **then all subsequent amplification must be done with linear amplifiers**. In a high power AM system, the use of linear power amplifiers is usually avoided because of their lower efficiency. Low level modulation is usually restricted to transmission systems that operate at low power levels such as small hand-held transceivers and telemetry devices.

Telemetry: distance metering or measurements at a distance, via radio.

Note: once a signal has been modulated ie. it contains intelligence, all amplification must be linear otherwise distortion will occur.

It is interesting to note that **SSB is in fact a low level AM modulation system**, since it is far more efficient (overall) to remove the carrier when it is in the milliwatt region, than to amplify the modulated signal and then have to remove a high power carrier which by then

could be a hundred watts or more.

When very high transmitter output power is required, it is usual to employ high level modulation. In these systems (AM Broadcasters) the carrier is developed by an oscillator and amplified to the final transmit power, then the modulating audio is applied. Since the carrier has no intelligence before modulation, high efficiency non-linear stages (Class C 60-80% efficiency) can be used for amplification. In a broadcast transmitter, modulating audio is frequently applied to the Power Amplifiers plate (anode) circuit (high level plate modulation). If the final carrier power is 10 kW then the modulator must supply 5 kW of audio power to the modulated wave for one hundred percent modulation.

The relationship between carrier power and side band power in AM system is:

$$P_t = P_c + 0.5m^2P_c$$

where:

P_t = the total power transmitted

P_c = carrier power

m = modulation index

Examination of the equation shows that since the **carrier power (P_c) remains constant**, then the side band power is given by the equation as:

$$\text{Power Sidebands} = 0.5m^2P_c.$$

The modulation index (m) is simply the percentage of modulation divided by one hundred. As the percentage of modulation goes from 0% to 100%, the modulation index goes from 0 to 1.

You do not need to know the above the equation for either NAOCP or AOCP exam.

A WORKED EXAMPLE

A 100 watt double side band (AM) transmitter is modulated firstly to 50% and then to 100%. What is the total power transmitted in each case?

From the equation given, at 50% modulation:

$$P_t = 100 + 0.5 \times (0.5)^2 \times 100 = 112.5 \text{ watts,}$$

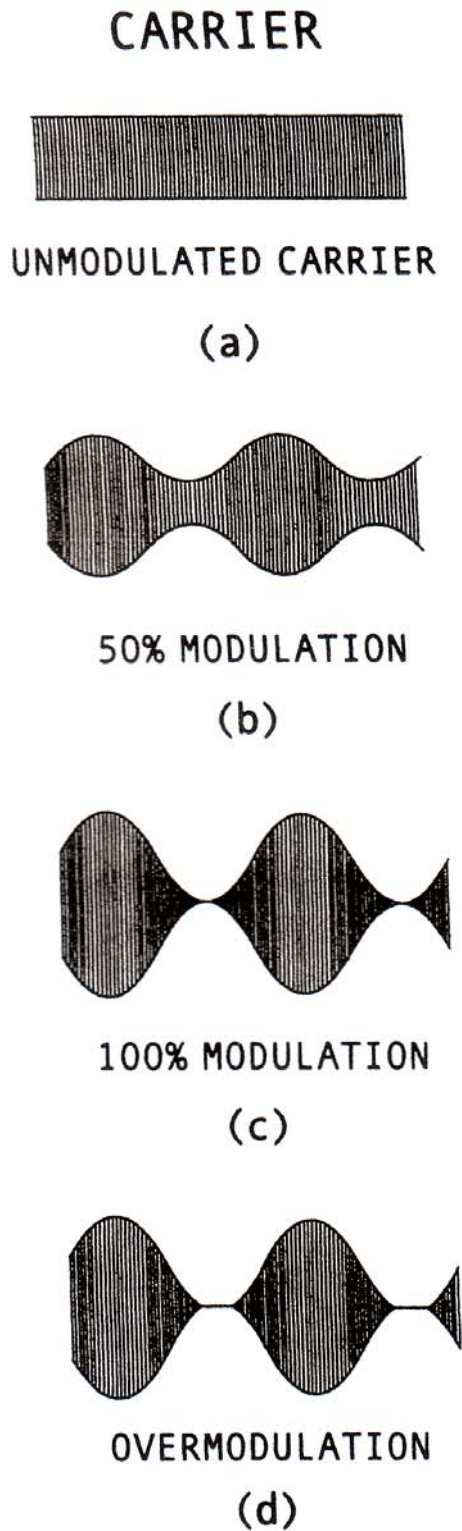
of which one hundred watts is the carrier power and **only 12.5 watts** is the total side band power. That is, 6.5 watts in each side band. At 100% modulation:

$$P_t = 100 + 0.5 \times (1)^2 \times 100 = 150 \text{ Watts ie. } 25 \text{ watts in each sideband.}$$

The only reason I've shown these calculations is to highlight the **inefficiency of a double side band AM system**. At best (m=1) the power transmitted in the sidebands is only **one-third** of the total power being transmitted (eg. in example above for 100% modulation with m=1, 50 watts total sideband is one third of total power of 150 watts). The carrier does not convey any intelligence to the receiver. The carrier can be removed and all the information in the sidebands will remain.

AMPLITUDE MODULATION ON AN OSCILLOSCOPE

Figure 1



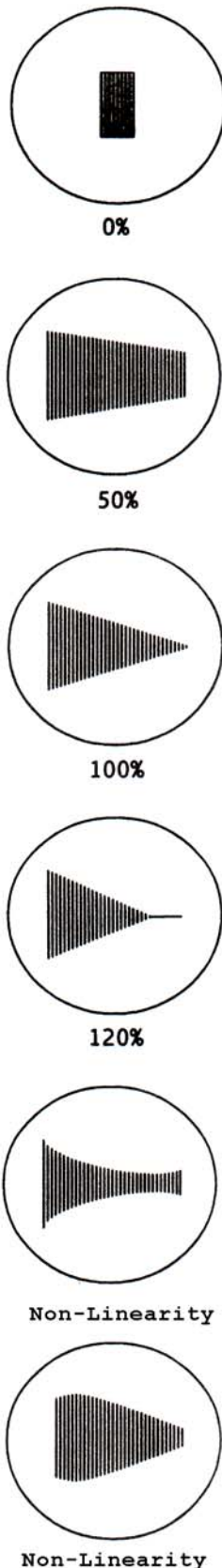
You do not need to know how these waveforms on an oscilloscope are produced, however you may be asked in the exam to identify one of them. As you can see from the way the waveform of an amplitude modulated signal is displayed on an oscilloscope, it does appear at a glance that the power in the carrier varies.

An oscilloscope displays resultant voltage and not power. Carrier power remains constant in an AM system.

TRAPEZOID PATTERNS

The trapezoid patterns (figure 2) are another method of showing the percentage of modulation on the oscilloscope. Again, you do not need to know how this is done with the oscilloscope, however, such patterns have been shown in examinations, and you may be presented with one pattern in the exam and asked a multiple choice question about what the display means.

Figure 2 – Trapezoid Patterns



I could explain to you how this is done, however, I think you have enough to remember without adding superfluous non-exam essential information. The trapezoid patterns have one advantage over the previous patterns - they show if the transmitter is linear. Look at the last trapezoid pattern and you will see that sides of the trapezoid are bowed inward (concave); this indicates the transmitter is operating in a non-linear way. The sides of the trapezoid could also be bowed outward (convex) this would also indicate non-linearity. You may be wondering what all the fuss about linearity is. Non-linearity causes interference, so it is highly undesirable. Whenever signals are fed to a non-linear device in an amplitude modulated system, some of the signals mix due to the nonlinearity and can create unwanted frequencies that can produce interference to other services.

A NOTE ON EFFICIENCY

In an amplitude modulated system no intelligence is transmitted by the carrier, and the wanted intelligence is transmitted in both the upper and the lower sidebands. It is no wonder then that amplitude modulation is not a popular method of voice communication, though it is still used by AM commercial broadcasters (though many of these are converting to frequency modulation).

On a CB radio there is often a switch to change between AM (double sideband), upper sideband and lower sideband. Many CB radio operators have the impression that these are three different channels, when in fact it is only one channel, with three different methods of transmission - all of them being amplitude modulation.

SINGLE SIDEBAND SUPPRESSED CARRIER (SSB)

SSB is a form of amplitude modulation (AM) where the carrier is suppressed (typically 40 to 60 dB below the carrier), and one of the sidebands is removed.

For an interesting experiment on air particularly with amateur and CB stations: Ask the other nearby station to key his or her transmitter say on upper side band and ask them not to say anything (don't modulate). Then you tune down about 1.5 kHz and you will hear the suppressed but not eliminated carrier. You will hear a tone – this is the carrier.

Since all of the intelligence in a double side band signal is present in each of the sidebands, and none of the intelligence is in the carrier, it is possible to remove one sideband and the carrier before transmission. The advantages of this are:

- Conservation of spectrum space (an SSB signal occupies only half the band of a DSB signal);
- The useful power output of the transmitter is greater since the carrier is not amplified;
- The SSB receiver is quieter due to the narrower bandwidth (receiver noise is a function of bandwidth - this will be discussed later).

The major disadvantage of SSB is that the carrier must be reinserted in the receiver, to down convert the sideband back to the original modulating audio. Another important fact that many textbooks on SSB fail to emphasise is, that the carrier must be reinserted with the same frequency and phase it would have **if it were still present**. This is why clarity control on a CB radio is so fiddly, if the carrier is not inserted exactly correct you get the Donald Duck type voice. The amateur operator will use the VFO or RIT (receiver incremental tune) to do the same thing.

A single sideband signal on entering a receiver will go through many frequency conversions before being converted to the original audio. If the carrier was left with the sideband, it (the carrier), would experience the same conversions, so the carrier must be reinserted at a frequency and a phase relative to the sidebands current position in the spectrum for demodulation of the original audio (demodulation is the opposite of modulation).

Because of the necessity for the carrier to be reinserted, the SSB receiver is more complex than the DSB (AM) receiver, and requires continuous attention from the operator to keep the reinserted carrier exactly on the correct frequency and phase. If the receiver weren't so finicky, the public broadcasting service would have adopted the SSB system long ago.

TWO-TONE TEST SIGNAL

It is unlikely that you will ever perform a two-tone test. However, you do need to know what it is for examination purposes.

A two-tone test signal provides a standard whereby measurements made on an SSB transmitter can be duplicated. If an SSB transmitter is voice modulated and the output fed to the vertical plates of an oscilloscope, **the pattern produced is so complex and**

dependent on the voice characteristics of the individual operator that it defies interpretation. A two-tone test signal consists of two sinusoidal modulating audio frequencies of equal amplitude which are *not* harmonically related (one is not a multiple of the other). The two tone test signal produces the identifiable and repeatable patterns that I've shown in this reading.

The reason the two tones of the two tone test signal must not be harmonically related is that the linearity of any system can be determined by injecting the two tones in, and looking for distortion products of these frequencies at the output.

Suppose tones of 1500 Hz and 2300 Hz were fed to an SSB transmitter in the upper sideband mode. The upper sideband output should then consist of two upper side frequencies: carrier + 1500 Hz and carrier + 2300 Hz. If any other upper side frequencies are present, this would indicate nonlinearity in what should be a linear system. If the tones were harmonically related, it would not be possible to distinguish easily between harmonics of the two tones. Unfortunately for the radio amateur, such linearity measurements are best done with an expensive device called a spectrum analyser.

So in reality, two-tone measurements are rarely done by many amateurs. However you need to know that a two tone (audio oscillator) is used to test an amateur SSB transmitter for non-linearity, and produces the oscilloscope patterns shown earlier.

SINGLE TONE TEST

If an SSB transmitter were modulated with a single audio tone, what would you expect to appear on the oscilloscope? (a typical exam question).

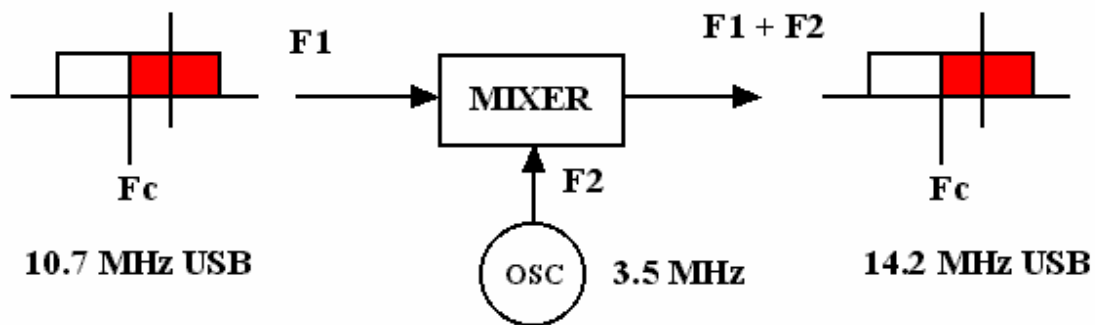
Answer: an unmodulated carrier wave - see earlier diagram.

In fact, if a SSB transmitter were modulated with a single audio tone it would be very hard to tell the difference between it and a CW (morse code) transmitter.

If a single audio tone is fed to the microphone of an SSB transmitter, then the sideband output would consist of a single upper side frequency - the transmission would be indistinguishable from that of a telegraphy transmitter with the key down. In fact, if the audio tone could be keyed on and off, telegraphy signals could be sent. With the addition of a second tone, and alternation between the two tones, code transmissions of teletype (RTTY) or facsimile (FAX) type are possible.

This method of modulating further SSB signals is called "audio frequency shift keying" or AFSK. This technique is extended further with slow scan television where a number of discrete audio tones are used sequentially to represent different shades of grey (these special modes will be covered more fully later).

In figure 3 below a 10.7 MHz upper sideband signal is mixed with a variable frequency oscillator (VFO) on 3.5 MHz. The output is the sum of F1 and F2. What sideband will the SSB signal now be on?



MIXING DOES NOT INVERT SIDEBANDS

Figure 3.

When a single sideband signal is mixed with an oscillator to move it in frequency, the sideband that is the upper or lower sideband, remains an upper or lower sideband.

This question has been asked in an AOCPE exam, and seems to confuse many students. Adding 3.5 MHz to each frequency component in a sideband signal will have the effect of shifting each frequency component up the spectrum by 3.5 MHz. The relative distance between all frequency components (all of the individual side frequencies) is unaffected by the shift, and the sideband remains as it was, whether it was upper or lower.

Here are some other interesting aspects of the question. What does it mean to have an SSB signal on 10.7 MHz? Where is the 10.7 MHz measured? This may seem simple, but remember a sideband is not a single frequency but a small band of frequencies.

In the example given above, the 10.7 MHz SSB signal may be passing through an intermediate frequency amplifier (IF). The IF amplifier is tuned to 10.7 MHz with a passband of 1.5 kHz on either side of the frequency to allow the sideband through. Obviously any calculations to do with mixing must consider the centre of the sideband to be 10.7 MHz.

In commercial and amateur radio it is customary to specify the transmitter frequency for an SSB signal as that of the suppressed carrier. An amateur station transmitting on 28.5 MHz USB means that the suppressed carrier is on 28.5 MHz, on upper sideband the actual centre of the emission is 1.5 kHz higher than the carrier, and on lower sideband 1.5 kHz lower.

CONVERTING AN SSB SIGNAL BACK TO THE ORIGINAL MODULATING AUDIO

After all, the whole idea of radio transmission is to take an audio frequency from the microphone, convert it to a radio frequency, transmit or radiate it, receive at a distant location, and recover the original audio.

With a single sideband signal the missing carrier is reinserted into the sideband at the receiver in the exact position it would be had it not been removed. The carrier re-insertion is done in the *product detector*. The difference frequency between the carrier and each individual side frequency component is the demodulated audio.

To put this in other words, and demonstrate the whole aspect of radio communications regardless of the mode used, the problem is that audio frequencies cannot be radiated from an antenna. So, regardless of the mode of transmission, the audio frequency from the microphone must be converted in frequency to a radio frequency, so that it can be transmitted and received at some distant location. The purpose of the receiver then, is to convert that radio frequency back down to the original audio frequency that entered the microphone at the transmitter.

SCHEMATIC DIAGRAM OF A MIXER

The combination of a mixer and a local oscillator is often called a frequency converter. A mixer takes two input signals. Usually a signal such as an AM, SSB or FM plus another signal from an oscillator are fed to the mixer. The mixer is just a *non-linear* device. The non-linearity will cause the mixing to take place. If you feed two signals into a *linear* device you would get the same two signals out and that's all.

We can even just use a diode as a non-linear device. However, better mixers consist of active devices biased to operate on the non-linear portion of their operating curve.

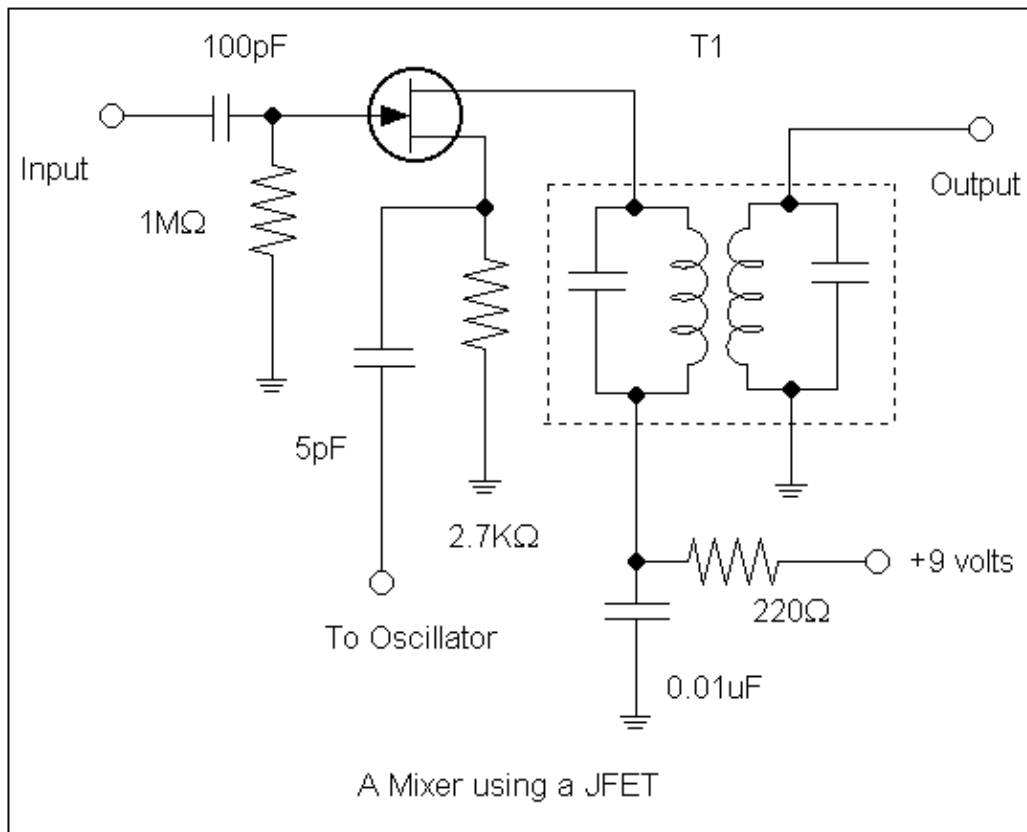


Figure 4.

The schematic in figure 4 shows a JFET mixer. One signal is fed to the gate of the JFET while the other is injected into the source leg. Mixing will occur because the JFET will be operating as a non-linear device. The bias used is 'gate-leak' formed by the 100pF capacitor and 1MΩ resistor.

At the output of the mixer is an IF transformer with acts as a bandpass filter passing the mixing products we want and rejecting the others.

The 220-ohm resistor and the 0.01uF capacitor act as a lowpass filter to prevent RF from entering the power supply.

Another popular method of mixing is to use a dual gate MOSFET. This is shown in the schematic of figure 5 below.

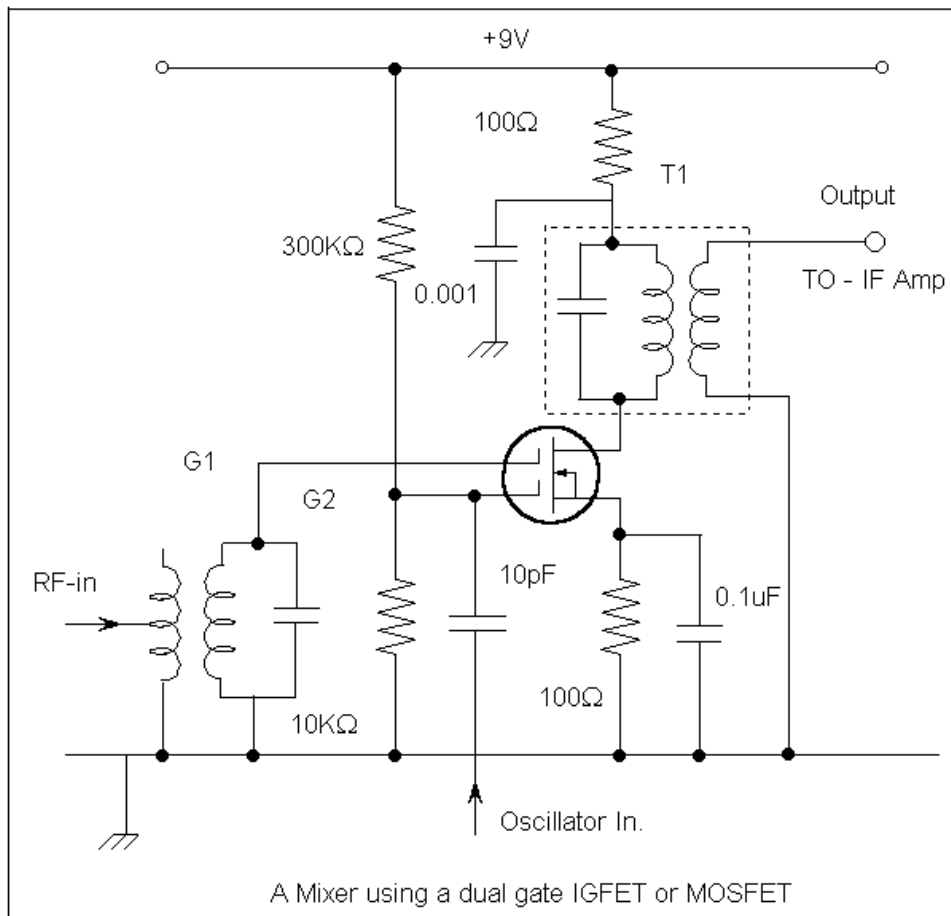


Figure 5.

The signal that is to be moved in frequency is fed to gate 1 and the signal from an oscillator is fed to gate 2. This circuit uses voltage divider bias on gate 2.

This circuit is quiet easy to identify as a mixer. You could be asked to identify a mixer circuit in the exam.

End of Reading 29
Last revision: April 2002
Copyright © 1999-2004 Ron Bertrand
E-mail: manager@radioelectronicschool.com
<http://www.radioelectronicschool.com>
Free for non-commercial use with permission