Back to basics – modulation by Andrew Barron ZL3DW

All hams know a bit about modulation, or they did when they sat the exam anyway. So we will treat this as a refresher on different modulation types and why we need to understand how they affect our signal. Our radio transmitters all generate RF energy on frequencies within the allocated amateur radio bands, whether that is around 146MHz to talk on a 2m repeater, 3.6Mhz for an 80m net, or 14.010MHz for a cw contact. Sending the RF wave on its own is not enough because it carries no 'intelligence', for example keying up your FM handheld on the local repeater and not talking, just sends the message "I am being a nuisance" and very little else. We need to add some information to the signal such as an audio voice signal or digital data. The process of adding an information signal to the RF energy is called modulation. The RF energy or 'wave' is called a 'carrier' wave because it carries the information signal that you want to send.

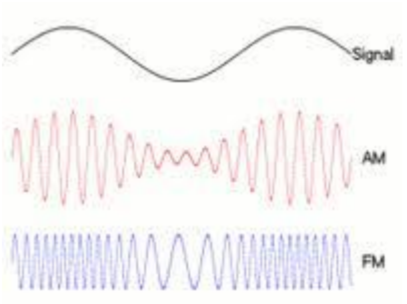
There are only three ways to modulate an RF wave so that it can carry information. You can change the amplitude (level) of the RF signal, you can change the frequency, and you can change the phase. All modulation types use changes of one or more of these factors to carry the information intelligence. I know about 50% of you are saying "what about cw then". But you can consider cw to be an, amplitude modulated digital mode. When you send cw, the RF carrier amplitude is changed from full output level to zero output level and that is enough to convey the digital code, in this case Morse code. Spark transmitters transmitted wide band noise as RF energy, but to convey information it was turned on and off using the Morse key and so 'spark' is still effectively an amplitude modulated signal.

It does not matter whether we are talking about analogue modulation where the information, for example your voice, modulates the amplitude, frequency, or phase of the RF carrier signal, or digital modulation where the information is coded into data bits and then that data is used to modulate the amplitude, frequency and/or phase of the RF carrier, those three methods are all there are. I know you are thinking that in SSB no carrier wave is transmitted, but the RF is still generated by the oscillator in the rig. What happens is that only the modulated RF energy is transmitted and that brings us to the issue of how does the modulation affect the RF signal being transmitted and why should we care about that.

A pure RF sine wave signal has a very narrow bandwidth (theoretically zero), phase noise, minor amplitude variations and jitter cause the bandwidth to be a little wider because they cause noise modulation of the pure wave. The wanted modulation increases the bandwidth further. Bandwidth is the range of frequencies that are required to transmit your information. Generally the more information you want to send, the wider the signal. For example hi fidelity music needs a wider bandwidth than 300Hz-3kHz communications quality voice audio, higher data speeds require more bandwidth as well. An example of this is looking at PSK31, PSK63 and PSK125 signals on the waterfall of your digi-mode software. You find that the faster modes are wider. In most cases if you over modulate the RF wave you will cause distortion which usually makes the bandwidth of your signal much wider, potentially causing interference to other users, or even out of band to other services. In (analogue) voice modes this is called splatter. Your signal would sound distorted and might be heard across a wide area of the band. In digital modes the distortion caused by over modulation can cause the signal to be wider or fuzzy. Severe distortion could cause the phase transitions to be masked causing poor or no decoding of your signal even when the receive level is strong. Most digital modes use phase transitions or frequency shifts to convey the data information. High capacity digital systems usually use QAM (quadrature amplitude modulation) which is a combination of amplitude and phase changes.

Why is bandwidth important? The main reasons are that you must not transmit outside of the allocated bands and you must try to not cause interference to other users of the band. If you are transmitting an USB signal on a frequency of 14.347MHz and the audio you are using for modulation has been limited to a maximum of 3kHz then all is fine, but what about if you change those mic equaliser settings in your fancy new transceiver to a 4kHz filter? Yes it will sound more hi-fi but you will be transmitting out of band. Further down the band, you will interfere with other stations 3kHz above you and unless the ham at the other end of the QSO has a wide receiver filter they won't hear the high notes in your signal anyway. Another downside of wide signals is that your RF power is distributed over a wider spectrum so your signal will have a lower signal to noise ratio at the receiver.

As the bandwidth of the RF signal becomes wider due to modulation the amount of RF spectrum used increases, usually on frequencies both above and below the carrier frequency. These bands of RF energy on each side of the centre carrier frequency are known as 'side bands'. As explained at the beginning, the carrier on its own contains no useful information. All of the useful information is contained in the sidebands. So some modulation schemes minimise or eliminate the RF power sent on the carrier frequency in order to maximise the RF power containing the useful information. In single sideband transmission we only transmit one sideband and no carrier, so all of the RF energy transmitted contains useful information (speech or data). This is why the recovered audio from an SSB signal is much stronger than from an AM signal, where much of the power radiated is carrier signal.   This increase in received audio signal is the reason that most hams use SSB rather than AM on the HF bands.



Modulating signal

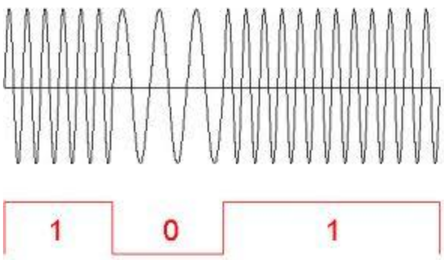
AM modulated RF signal

FM modulated signal

With an analogue FM (frequency modulation) transmission the transmitter radiates a fixed level of RF energy, the amplitude of the audio signal changes the frequency of the transmission. In other words the louder you talk the more the frequency deviates from the carrier frequency. The maximum amount that the RF signal can be deviated is always limited to avoid using too much bandwidth and again hi-fi signals with large dynamic range need more bandwidth. For example a narrow band FM signal from your hand held radio will be limited to +-3kHz deviation while an FM broadcast station will be limited to +-75kHz deviation. The frequency of your audio signal affects the rate that the signal changes. Because changing the instantaneous frequency of the signal also changes the phase of the wave, there is essentially no difference between analogue Frequency Modulation and analogue Phase Modulation. Many FM transmitters actually use phase modulators, check the specification page on your hand-held radio manual. There are differences in the frequency response of the received signal when frequency rather than phase modulators are used and these are dealt with using pre-emphasis at the transmitter with de-emphasis at the receiver.

FM transmission is more immune to noise than amplitude modulated signals like SSB. That is because the amplitude of the FM signal is not very important (as long as it is not zero!). Noise spikes don't affect the deviation much so you don't hear them in the received audio. The main disadvantage is that FM signals take more bandwidth to carry the same quality of audio. This would be a problem on the narrow HF bands, but is less important on the higher bands because there is more room. That is the reason that FM is only used on the 10m ham band and above.

In the digital world FM (frequency modulation) is not the same as PM (phase modulation). The RTTY mode is a good example of a frequency keyed digital mode, two tones are transmitted at the same level. Your transmitter only sends one tone at a time, so the tones don't mix. The way the tones are alternated sends the digital information, in this case the Baudot code. This means that an RTTY signal is narrow bandwidth and because the tone levels are constant an RTTY signal can be amplified with a non linear amplifier. Modes like Domino, Ros and Olivia are also frequency keyed. In these cases a variety of tones are used, but only one tone is transmitted at a time. The sequence of the tones carries the digital information. In other words a transition from tone (a) to tone (b) may mean "0110" and a transition from tone (a) to tone (c) may mean "1011".



Frequency Keyed modulated signal

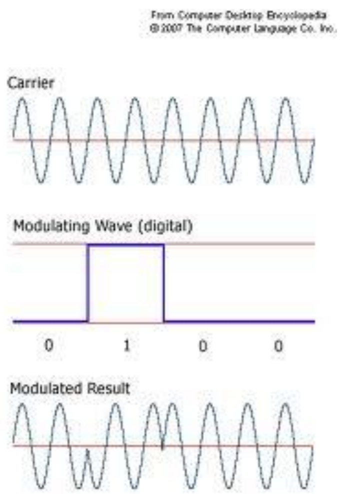
Modulating signal

PSK31 is the simplest phase modulated mode. PSK stands for 'phase shift keying'. In PSK a single tone is transmitted. 180 degree phase changes are used to send the digital information. A more complicated phase modulation method is 4PSK. In that mode 90 degree phase changes each send two bits of the digital signal. Each of the four possible phase states is called a symbol. Because PSK modes only use one frequency they are narrow bandwidth and can be amplified with a non linear amplifier.

Un-modulated RF signal

Modulating signal

Phase Keyed modulated signal

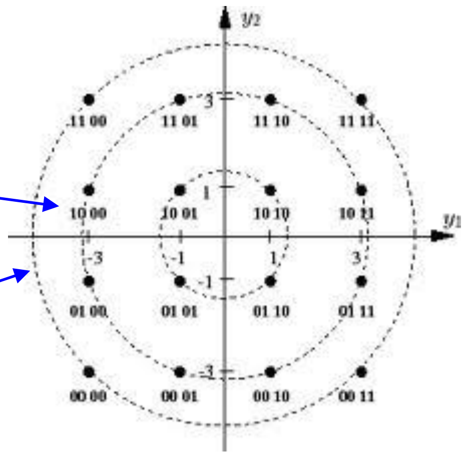


Quadrature modulation is usually a combination of analogue and phase modulation. In 4QAM a two voltage level signal (like a standard binary stream of 1s and 0s) is combined with a second a two voltage level signal which is adjusted to be 90 degrees out of phase with the first stream. This creates an output with four possible phase states 90 degrees apart, each phase state or 'symbol' carries 2 bits of digital information.   If you are still with me, you will be thinking "*huh that's just like 4PSK*" and you would be right! There is no functional difference between a 4PSK signal (with 90 degree shifts) and a 4QAM signal.   Modern digital radio systems use quadrature modulation to combine two, 4 level data signals, with each voltage level equivalent to 2 bits of binary data to create a 16QAM signal with 16 phase states (symbols). Each symbol represents four bits of digital data, two from each of the input data streams. Because you have the voltage levels as well as the phase states, 16QAM mod schemes and above are a combination of amplitude modulation and phase modulation. High capacity radio systems carry on with this idea and use, 64QAM, 256QAM or even 512QAM modulation.   The higher the complexity of the modulation the wider the bandwidth that is required. Because 16QAM and above includes amplitude modulation you must use a linear amplifier. Non linearity would cause the received data streams to have the wrong voltage levels which could cause the data to be decoded incorrectly. Also the amplifiers must have good phase linearity because a phase error on the signal can also cause the received data to be decoded incorrectly.

This is the phase diagram showing the 16 phase states generated by a 16QAM modulator.

Note that each phase state symbol carries 4 bits of data.

The three rings illustrate the three amplitude levels. So each symbol has a phase angle with respect to the carrier wave and also one of the three possible amplitude levels.



FINALLY A WORD OR TWO ON AMPLITUDE LINEARITY. Amplitude linearity often referred to as just 'linearity' is a measurement of how accurately the output of a device follows changes in the level of the input signal. With amplitude modulated signals, particularly digital AM signals, it is important that the output signal of stages like your final amplifier and any additional amplifiers is an accurate, but larger copy of the input signal. Severe amplitude distortions could cause your signal to be distorted and splatter. Many hams use a Linear Amplifier on the HF bands where perhaps the most commonly used modulation method is SSB. Many digital modes and cw don't actually need the amp to be particularly linear and some linear amps have an alternative bias setting for cw. On the VHF bands when you are using FM modulation or digi-modes like JT65, class C amplifiers may be used. Of course for SSB you still need a linear amplifier.

Drawings were gathered from various Internet sources.