

RADIO COMMUNICATIONS: AM AND FM

Prepared by Dr. Bhaskar Ramamurthi

November 1998

- If you wish to broadcast two or more signals which occupy the same frequency band, you have to translate them to non-overlapping frequency bands. That is, if you wish to transmit two signals both of which occupy the band 0–10kHz, one of them (at least) has to be shifted to a different band. In practice, both will be shifted to non-overlapping bands beyond 500kHz at least, since radio transmission at lower frequencies is impractical. Frequency translation in electrical circuits is achieved by **mixers**.
- At the receiver, one has to select the band of interest from a composite signal consisting of many non-overlapping bands. This is achieved by an electrical circuit known as **filter**. Radio transmitters and receivers consist primarily of mixers, filters, and amplifiers that amplify signals).
- Let us first look at the principle behind a mixer. We have a **baseband** signal from 0 to W Hz. For ease of understanding we will assume the baseband signal is a sinusoid with frequency f_b : $\cos 2\pi f_b t$. Let us multiply this signal by a **carrier** sinusoid: $A_c \cos 2\pi f_c t$. From elementary trigonometry, we know that the result is $(A_c/2) \cos 2\pi(f_c + f_b)t + (A_c/2) \cos 2\pi(f_c - f_b)t$. Here we have implicitly assumed that f_c is larger than f_b , which will always be the case. We can see that the baseband signal has been sifted in frequency. However, two copies have been created: one at a frequency higher than f_c , and the other at a frequency lower. This will happen to all frequency components in the baseband signal. So, we will get a so-called **upper sideband** from f_c to $f_c + W$ Hz, and a **lower sideband** from $f_c - W$ to f_c Hz. The key principle in mixing is therefore multiplication.
- The word mixing has come about because electrical engineers are particularly sensitised to the fact that new frequencies can be created only when signals mix, as opposed to taking weighted linear combinations, which cannot create new frequency components that do not already exist. Linear operations can only modify the strengths of the components already present. The filter we have referred to above, for example, can be obtained by linear processing. The word mix connotes that in the sense of allowing non-linear operations, the “purity” of the signal is not being maintained.
- Now let us see how the mixer can be used to shift the high-frequency signal back to baseband. Let us once again consider the signal above, consisting of two sinusoids at $f_c + f_b$ and $f_c - f_b$, for ease of understanding. Let us mix this with $\cos 2\pi f_c t$. We will now get three components: at f_b , $2f_c + f_b$ and $2f_c - f_b$. Work out the precise expressions to verify this. Actually, we will get the baseband signal back as well as frequency-shifted signals around $2f_c$. This unwanted band, called the **image**, has to be eliminated by filtering.
- In the above, if we use $\sin 2\pi f_c t$, note that we get only the unwanted image band. Work out the trigonometry to see that the baseband component vanishes. If there is an arbitrary phase difference θ , we will get the baseband attenuated by $\cos \theta$. When θ is $\pi/2$, we get no baseband, as we have just seen. This is the main problem with the scheme, that we require what we call a **coherent local oscillator** whose phase difference with respect to the incoming carrier is nearly zero. We will shortly see how this is got around in commercial broadcasting.

- A receiver based on the principle shown in Fig. 1 is called a **super-heterodyne receiver**. The word heterodyne means “different frequencies”. The prefix super is not to indicate its superb nature, but to distinguish it from an earlier version which employed multiple frequencies, but was inferior. The simultaneous tuning of the coarse filter and the oscillator is achieved by a **ganged capacitor** (two capacitors whose value change proportionately by one control). It is this ganged capacitor that one varies as one turns the knob on a radio. A string and needle attached to it move along on a dial as you turn the knob to tell you what station you will receive for each position of the capacitor. The needle has nothing to do with the operation of the radio.
- Due to the cumbersome requirement of coherent reception, some modification is needed to the transmitted signal that will enable simpler operation at the receiver. This is particularly important in broadcasting, where there is one transmitter but millions of receivers. This is achieved by what is called **Amplitude Modulation** or **AM**. The word **modulation** refers to the variation of a particular parameter of the carrier, in this case the amplitude, in proportion to the message signal.
- In AM, baseband signal $m(t)$ is not directly multiplied with the carrier to produce the required frequency translation. Instead, the signal $(1+k_a m(t))$ is mixed with (or, modulates) the carrier. The constant k_a is chosen such that the sum above never becomes negative. The AM signal, therefore, is $A_c(1+k_a m(t)) \cos 2\pi f_c t$. A sample AM waveform is shown in Fig 2.

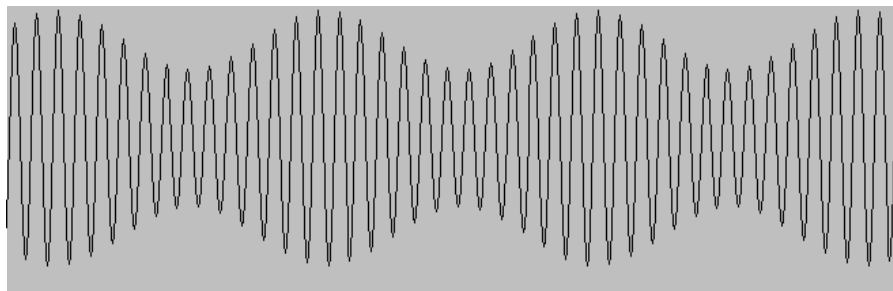


Fig. 2 : AM waveform

- The main advantage of AM is that detection of $m(t)$ can be done simply by what is called an **envelope detector** circuit. This circuit is basically related to the peak detector, which you might have seen earlier. A coherent local oscillator is not needed. A super-heterodyne receiver is still required, both to select the station and to bring the frequency down to the IF, a convenient frequency for implementing the detector.
- The signal-to-noise ratio (SNR) at the output of an AM receiver is directly related to the signal level at the input. The only way to improve the SNR is to increase the input power, since the noise power generated in the receiver is fixed.
- **Frequency modulation**, or **FM**, overcomes this limitation. In FM, the frequency of the carrier is modulated proportional to the message signal $m(t)$. What does one mean by the statement frequency is modulated? Here, we interpret frequency as the time-derivative of the argument of the sinusoid. That is, the FM signal is $A_c \cos(2\pi f_c t + 2\pi k_f x(t))$, where $x(t)$ is the integral of $m(t)$. If you differentiate the argument, and divide by 2π to convert radian-frequency to Hz, we get $f_c + k_f m(t)$. This is referred to as **instantaneous frequency**, and this lends meaning to the term frequency modulation. It can also be called phase modulation though we usually reserve that term to the case where $x(t)$ is replaced by $m(t)$.

- First, however, let us consider the other important operation of filtering. You are already familiar with inductors and capacitors, whose behavior is frequency-dependent. Consider, for example, a current signal flowing into a parallel combination of resistor, inductor, and capacitor. Let the resistor represent the circuit into which the filtered current is flowing. Let the signal be a sinusoid whose frequency can be changed. At low frequencies, the inductor has a very low impedance. The current will take the path of least resistance and flow through the inductor. Very little will flow through the resistor. At very high frequencies, the capacitor behaves similarly, and once again very little current flows through the resistor. At a certain frequency, the parallel combination of inductor and capacitor presents infinite impedance.

Try it out for yourself, if you know how: compute the equivalent impedance of the combination. You will see that the frequency at which this happens is $1/(2\pi\sqrt{LC})$. This is the **resonant frequency** of the L-C circuit, at which the inductor and capacitor keep exchanging energy back and forth and do not draw any current from elsewhere to maintain the voltage across them. At this frequency, any current driven into the circuit will flow through the resistor. At frequencies around this value, some of the current will flow in to the inductor and capacitor, and the current through the resistor will drop. We can see, therefore, that this parallel combination acts as a filter for frequencies around the resonant frequency of the L-C combination. The precise response of the filter to frequencies around this resonant frequency (also called **centre frequency** in this context) is determined by the relative values of R, L and C. The frequency band which the filter allows through with significant strength (some acceptable definition of what is significant has to be used here) is called the **passband** of the filter.

- In the context of radio receiver design, we need a filter whose passband can be shifted to select the station we want to pick up. This is difficult to do while maintaining the bandwidth of the passband. If we change L or C to change the centre frequency, the bandwidth will also change and this is not acceptable.
- A Smart way i.e. the Superhet Receiver has been found around this. The idea is to use a mixer to shift the desired station to a pre-determined frequency called the **Intermediate Frequency (IF)**. It is as if the IF is the carrier frequency now, but of the desired station. Of course, the nearby stations will also shift to frequencies on either side of the IF. Now, a fixed filter with the IF as centre frequency is used to **select** the desired station. The block diagram is shown in Fig 1. Note that only the local oscillator frequency changes depending on the station we want to select. The IF filter is unchanged. (It is easier to tune an oscillator by changing a capacitor, because now there is no band-width to worry about.) It turns out that we need a coarse filter at the input before the mixer. This filter will ensure that the **image station** is not picked up along with the desired station. The discussion of this is being purposely left out here, to leave something for you to chew on. Can you figure out what the carrier frequency of the image station is? The centre frequency of this coarse filter has to change, of course, but we do not mind its band-width changing also, because its task is only to reject the image station, and not the stations neighboring the desired station. If the IF is not too low, this can be easily achieved even if the passband width changes.

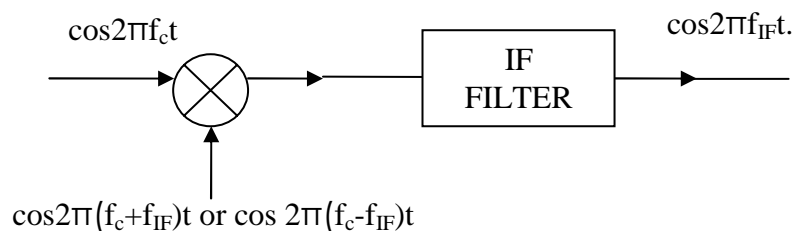


Fig 1: Super-heterodyne receiver principle

- FM signals can also be **de-modulated** (or detected) by a circuit that does not need a coherent local oscillator. We will not go into this topic here. What is more important is the specific advantage accruing from FM with regard to the SNR.
- The noise added to the FM signal by the receiver will also be occupying the same band as the FM signal itself. Such **bandpass noise** will corrupt the instantaneous value of the FM signal, and thus corrupt the zero-crossings. It is the corruption of the zero-crossings that will affect the demodulated signal in FM. Any FM detector, irrespective of the implementation, must in some manner measure the instantaneous frequency. This will mean measurement of the intervals between zero-crossings, at least indirectly. The corruption by the noise of the amplitude, which is what is relevant in the case of AM, does not matter here. It is the corruption of the zero-crossings that matters in FM. Now, the detected signal in noiseless FM will be proportional to $f_c + k_f m(t)$. If the constant is removed, we get a signal proportional to $k_f m(t)$, which is what we want. Now, if k_f is increased, the signal power in the detected baseband message signal can be increased. Due to the corruption of the zero crossings by the noise, we get additive noise in the detected baseband signal. The power in this noise signal does not change as k_f is changed. It is a function only of the noise level in the receiver. Thus the SNR in FM can be changed by changing k_f . This does not increase the power level of the FM signal because the power of the FM signal is $A_c^2/2$ irrespective of the modulation. (This is because the amplitude of the FM wave is always A_c). Thus, in FM the SNR of the demodulated signal can be increased by increasing k_f without affecting the transmit power.
- There is no free lunch in this world. Where does this hit? The answer is that as k_f increases, the bandwidth of the FM signal increases. This is a non-trivial result that you will have to take on faith. Therefore, FM can be said to permit a trade-off between bandwidth and SNR. To achieve a certain SNR at the receiver output, we can set the transmit power to a convenient level and increase the bandwidth allotted to each station accordingly. The transmit power in FM does matter. The larger the amplitude of the received FM signal, the less the variation in zero-crossing instants caused by a given noise level.
- When FM was introduced in broadcasting, the message signal bandwidth was also increased from the value used in AM. This permits more high frequency content in the music signal transmitted, making the music sound truer and richer. This, and the better SNR achieved in FM, are both responsible for the immense popularity of FM. We must remember, however, that an FM station uses about 200 kHz bandwidth compared to 10 kHz for an AM station.