AM & SSB

Introduction

Up to now we have been concerned primarily with power levels – received power, signal-to-noise, signal-to-interference, and so on. Signal power is only a means to an end, the end being to move *information* from one point to another. In this lecture we want to begin examining how we can use modulation to transform an RF carrier so that the resulting radio signal contains information. In a communications course (e.g., EE341 at WSU) you have probably been introduced to AM and FM, so the next two lectures will largely be review. Our main focus will eventually be on digital modulation techniques.

Modulation

An RF carrier is a simple sinusoid that has the form $a\cos[\omega_c t+\phi]$ where a is the amplitude, $\omega_c = 2\pi f_c$ with f_c is the carrier frequency, and ϕ is the carrier phase. For an ideal carrier, a and ϕ are constants. The phase depends on our time reference, i.e., what point we choose to call $t=0$. In theory we can always choose this to make $\phi=0$.

A carrier delivers no information. Once you know it's frequency, amplitude, and phase, you can predict the signal infinitely far into the future. As the name implies the carrier's job is to "carry" information contained in *modulation*. Modulation involves making the amplitude and/or the phase change with time according to some modulating signal. The most general modulation can be expressed as

$$s(t) = a(t)\cos[\omega_c t + \phi(t)] \tag{19.1}$$

where $a(t)$ is *amplitude modulation*, $\phi(t)$ is *phase modulation*. We'll assume that both of these are slowly varying with respect to the carrier frequency.

In principle you can use any combination of amplitude and phase modulation, and in order to get the maximum amount of information through a band-limited channel (e.g., a computer modem operating over a phone line) you may use both. However, as we have seen so far in this course, in a multipath wireless channel amplitude fluctuations can be introduced by small amounts of motion, and the fluctuations become more rapid as the frequency increases. So we would expect that using received signal amplitude to convey information might be problematic, especially at higher frequencies. For this reason, mobile radio channels almost exclusively use phase modulation.

AM

If in (19.1) we keep ϕ constant, and for simplicity we'll assume $\phi=0$, then we have *amplitude modulation* or AM. In this case $s(t)=a(t)\cos(\omega_c t)$. In principle $a(t)$ could be any signal, but it is difficult to tell positive from negative values, as illustrated in Fig. 19.1. In order to do so, the receiver would need to be able to figure out the time/phase reference of the transmitter, that is we would need *phase coherence*. Although this is possible, it is difficult, and in the early days of radio it was not feasible. Most AM systems, therefore, use a form of AM where phase coherence in not required by requiring $a(t)$ to be non-negative.

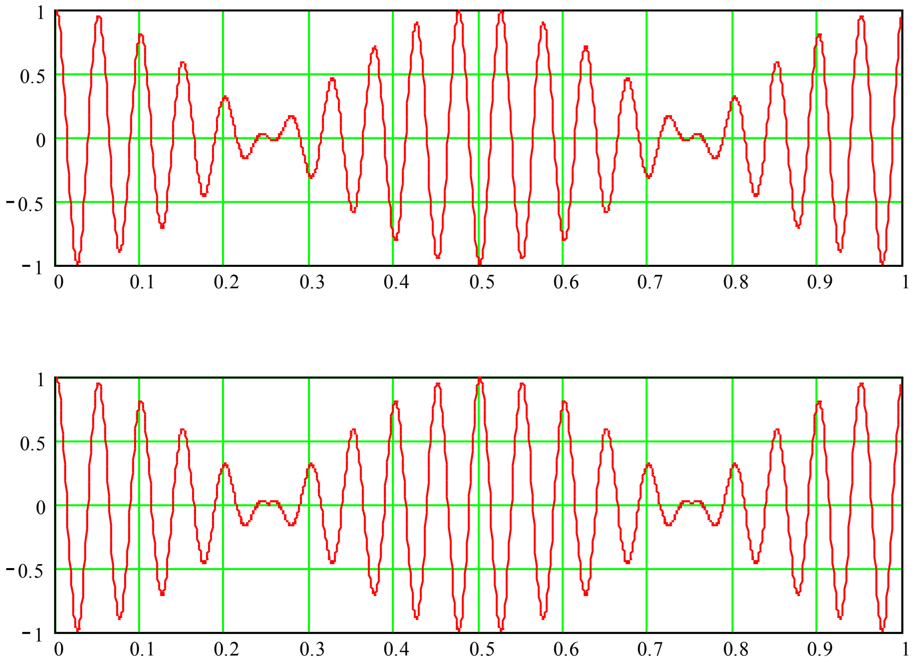

Figure 19.1:AM modulation. At top is $s(t) = (\cos 2\pi t)\cos 2\pi 20t$*, at bottom is* $s(t) = |\cos 2\pi t|\cos 2\pi 20t$*. You can easily make out that the amplitude of the modulation is the same in both cases, but the fact that the two cases have different signs in some intervals is only apparent if you examine the phase differences.*

Let's write our AM signal as

$$s(t) = A_c[1 + \alpha m(t)]\cos(\omega_c t) \tag{19.2}$$

where $m(t)$ is our modulating signal, for example, an audio signal. We assume $m(t) \geq -1$, which for most signals of interest is the same as $|m(t)| \leq 1$. Here A_c is the carrier amplitude and $0 \leq \alpha \leq 1$ is the *modulation index*. AM signals can be generated using an analog multiplier ("mixer") to form the product of the modulation $[1 + \alpha m(t)]$ and the output of an RF oscillator.

Consider the case where the modulating signal is a sinusoid, $m(t) = \cos \omega_m t$. Then

$$s(t) = A_c[1 + \alpha \cos(\omega_m t)]\cos(\omega_c t) \tag{19.3}$$

An example is shown in Fig. 19.2. Note that the maximum peak-to-peak amplitude of s is $s_{pp,\max} = 2A_c[1 + \alpha]$ while the minimum is $s_{pp,\max} = 2A_c[1 - \alpha]$. Therefore, by looking at an AM signal with sinusoidal modulation, we can calculate the modulation index as

$$\alpha = \frac{s_{pp,\max} - s_{pp,\min}}{s_{pp,\max} + s_{pp,\min}} \tag{19.4}$$

A modulation index of 1 (100% modulation) requires $s_{pp,\min} = 0$; the signal modulates to zero amplitude.

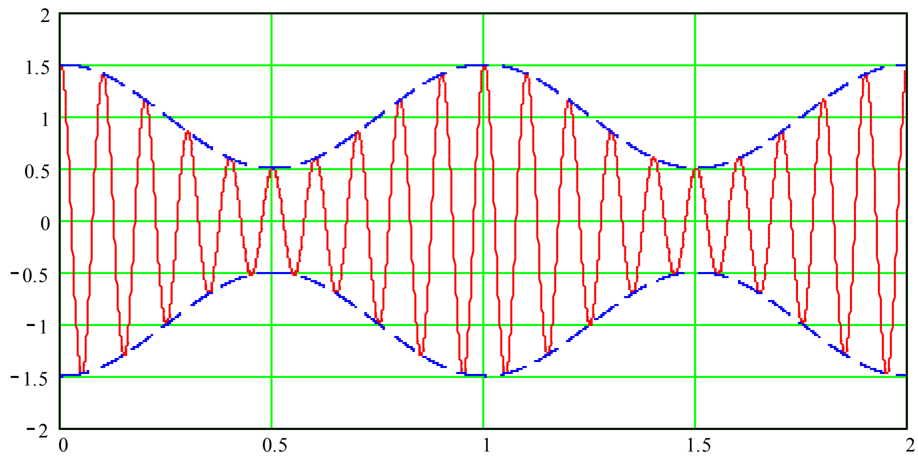

Figure 19.2: AM modulation $s(t) = [1 + 0.5\cos(2\pi 2t)]\cos(2\pi 20t)$. *The expression in brackets is the envelope. The envelope determines the instantaneous amplitude of the carrier.*

Recall from Lecture 9 that for a moving receiver fading can introduce amplitude fluctuations with a maximum frequency of v/λ . For highway speeds of around 100 km/hr and a PCS frequency of 1900 MHz, this is around 175 Hz. That is well into the audio region. If you tried to use AM under those conditions it would be a disaster because you could not distinguish amplitude fluctuations due to the modulating signal from those due to fading. On the other hand, for an AM radio frequency of around 1 MHz, the Doppler shift is 1900 times less, about 1/10 Hz. That is well below the audible range, so AM radio is feasible at low-enough frequencies. AM receivers typically use *automatic gain control* circuits to level out these very slow fluctuations due to fading.

Spectrum of an AM Signal

Consider the sinusoidally modulated signal (19.3). Using the trig identity $\cos a \cos b = \frac{1}{2}\cos(a+b) + \frac{1}{2}\cos(a-b)$, we can write this as

$$s(t) = A_c\left[\cos\omega_c t + \frac{\alpha}{2}\cos(\omega_c + \omega_m)t + \frac{\alpha}{2}\cos(\omega_c - \omega_m)t\right] \tag{19.5}$$

This consists of three frequency components. First, we have the carrier at frequency f_c and amplitude A_c. This component is unaffected by the modulation frequency or the modulation index. Then we have two components, each with amplitude of $\alpha/2$ relative to the carrier. One is at a frequency f_m above the carrier frequency while the other is f_m below the carrier. The higher frequency component is called the *upper sideband* while the lower frequency component is called the *lower sideband.* Figures 19.3 and 19.4 show AM modulation measurements in the time and frequency domains.

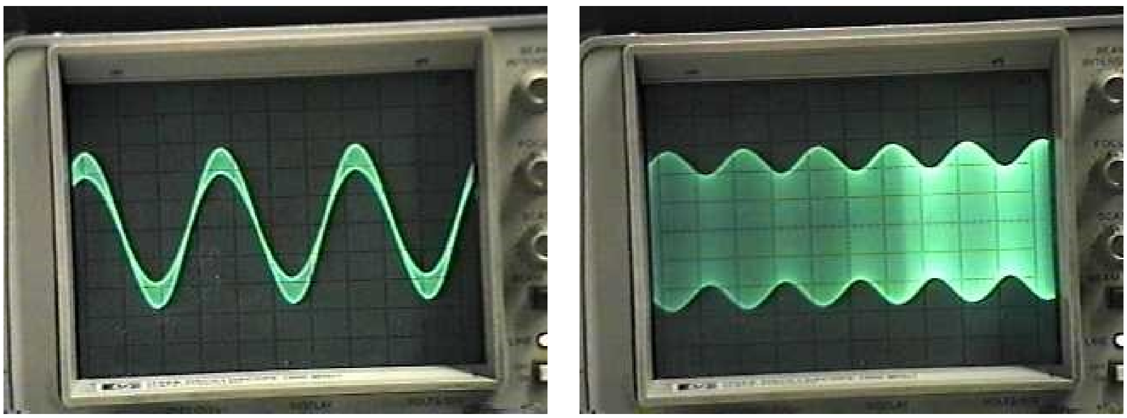

Figure 19.3: AM modulation. A 1-kHz audio tone is driving a 27-MHz AM radio (CB radio). At left we see successive carrier cycles superimposed over a relatively short time scale. At right we see the signal at a much long time scale where several cycles of the audio modulation can be observed.

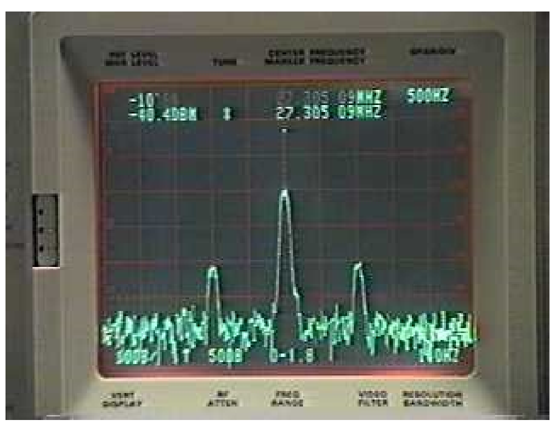

Figure 19.4: Spectrum analyzer display corresponding to the time-domain signal shown in Fig. 19.3. The resolution is 500 Hz per horizontal division and 10 dB per vertical division. The central peak is the carrier. On either side are the upper and lower sidebands.

In Fig. 19.5 the modulation index is increased to nearly 100%. We see that the carrier component of the spectrum in unchanged, but the amplitudes of the sidebands increase.

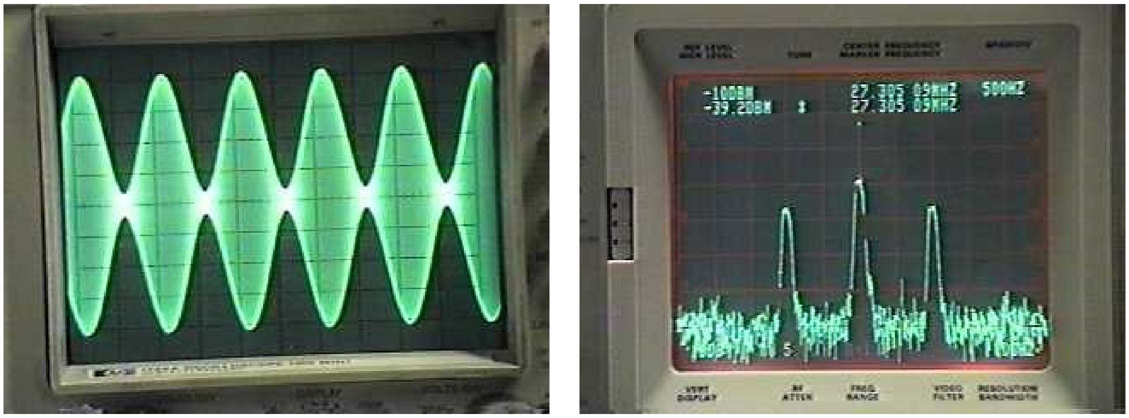

Figure 19.5: The signals of Figs. 19.3 and 19.4 when the modulation index is increased (by turning up the audio volume).

Even with 100% modulation, the amplitude of a sideband is only ½ that of the carrier. Yet the carrier contains no information about the modulation. Moreover, the sidebands are redundant, that is, they both contain the same information about the amplitude, frequency, and phase of the modulation. In principle we could get rid of the carrier and one of the sidebands without loosing any information about $m(t)$. That is, in Fig. 19.4 we could cut out the central peak and one of the other peaks. This is, in fact, what is done in *single sideband* modulation (SSB). We'll talk about SSB later. So, AM is not *power efficient*. If one unit of power is in a sideband, then another unit is wasted in the other sideband, and four units are wasted in the carrier. So only 1/6 of the total power (at most) is used to convey real information. And AM is not *bandwidth efficient*. With reference to Fig. 19.4, ½ of the total bandwidth is wasted on redundant information.

The reason for keeping the carrier and both sidebands is that AM signals are very easy to *demodulate*. With reference to Figs. 19.3 and 19.5, all we need is a circuit that can track the envelope of our signal. This can be done with a diode, a capacitor and a resistor. This simplicity, and the corresponding low cost, was a major factor in the commercial success of radio in the early part of the 20th century. In these times of integrated circuits, however, AM in not a very attractive modulation scheme.

Consider a more general modulating signal represented as a superposition of sinusoids: $m(t) = \sum_k a_k \cos(\omega_k t + \theta_k)$. In this case we would apply (19.5) for each of the sinusoidal components to get the AM signal:

$$s(t) = A_c \left[\cos \omega_c t + \frac{\alpha}{2} \sum_k a_k \cos[(\omega_c + \omega_k)t + \theta_k] + \frac{\alpha}{2} \sum_k a_k \cos[(\omega_c - \omega_k)t - \theta_k] \right] \qquad (19.6)$$

The first term is again the carrier. The first sum is the upper sideband while the second is the lower sideband. Again, they are redundant as each of them conveys everything you need to know about the modulation independently of the other sideband. If the modulation contains frequencies from DC to f_m, the bandwidth of the AM signal is $B = 2f_m$. So, the bandwidth of the AM signal is twice that of the modulation.

Noise Representation

We've noted that for a receiver with bandwidth B and noise temperature T_N the noise power is $P_N = kT_N B$, with k being Boltzmann's constant. Let's call

$$N_0 = kT_N \tag{19.7}$$

the *noise spectral density*. This has units of Watts per Hz, or Joules. Then $P_N = N_0 B$. For example, for $T_N = 290K$ ("room temperature"), which represents a noise figure of 3 dB, the noise spectral density is $4 \cdot 10^{-21}$ Watts per Hz.

As we begin to examine the effect of noise on our ability to extract transmitted information, we will need to be even more specific. Receivers are designed so that they ideally pass only the spectrum of the desired signal and block all other frequencies, so, B will typically be the bandwidth of the transmitted signal. The noise that makes it through with the signal is therefore some random process occupying a spectrum of width B centered on the carrier frequency f_c. We can represent the noise in the same way we represent a modulated signal, i.e., (19.1)

$$n(t) = r_n(t)\cos[\omega_c t + \phi_n(t)] \tag{19.8}$$

except that $r_n(t)$ and $\phi_n(t)$ are random processes instead of deterministic functions.

Since $n(t)$ represents a voltage, and power is voltage squared over resistance, the average power into a resistance R would be $P_N = \langle n(t)^2 \rangle / R$. We will typically be interested in *ratios* of signal to noise power, so the factor of $1/R$ will be cancelled by a similar term for the signal power. Therefore, R doesn't really matter, and for simplicity we will take R to be 1 Ω. With an R of 1 Ω we have $P_N = \langle n(t)^2 \rangle$. Keep in mind that *power* is the physically relevant quantity. We can always impedance match a signal to any load thereby transforming the voltage, but the power will be conserved. But, we will always transform the signal and noise voltages by the same amount, so their ratio will be unchanged.

In many cases a more convenient representation of the noise is

$$n(t) = n_c(t)\cos\omega_c t - n_s(t)\sin\omega_c t \tag{19.9}$$

where

$$\begin{aligned} n_c(t) &= r_n(t)\cos\phi_n(t) \\ n_s(t) &= r_n(t)\sin\phi_n(t) \end{aligned} \tag{19.10}$$

We assume that $n_c(t)$ and $n_s(t)$ are independent, zero-mean Gaussian random processes, each with variance $\sigma_n^{\ 2}$ and having bandwidth B. The noise power is

$$
\begin{aligned}
P_N &= \langle n(t)^2 \rangle \\
&= \langle [n_c(t)\cos\omega_c t - n_s(t)\sin\omega_c t]^2 \rangle \\
&= \langle n_c(t)^2 \rangle \langle \cos^2 \omega_c t \rangle + \langle n_s(t)^2 \rangle \langle \sin^2 \omega_c t \rangle - 2\langle n_c(t) n_s(t) \rangle \langle \cos\omega_c t \sin\omega_c t \rangle \\
&= \sigma_n{}^2 \frac{1}{2} + \sigma_n{}^2 \frac{1}{2} - 2(0)(0) \\
&= \sigma_n{}^2
\end{aligned}
\tag{19.11}
$$

With

$$
\begin{aligned}
r_n(t) &= \sqrt{n_c(t)^2 + n_s(t)^2} \\
\tan\phi_n(t) &= \frac{n_s(t)}{n_c(t)}
\end{aligned}
\tag{19.12}
$$

you can show that $r_n(t)$ has a Rayleigh distribution with $\langle r_n(t)^2 \rangle = 2\sigma_n{}^2$, and $\phi_n(t)$ is uniformly distributed over 0 to 2π.

Effect of Noise in an AM System

Let the AM signal in a receiver be $s(t) = A_c[1+\alpha m(t)]\cos(\omega_c t)$. Let's represent the noise by (19.9). The carrier power (into 1 Ω) is $A_c{}^2/2$, so the RF *S/N* is $(A_c/\sigma_n)^2/2$. The total signal in the receiver is

$$
\begin{aligned}
x(t) &= A_c[1+\alpha m(t)]\cos(\omega_c t) + [n_c(t)\cos(\omega_c t) - n_s(t)\sin(\omega_c t)] \\
&= \{A_c[1+\alpha m(t)] + n_c(t)\}\cos(\omega_c t) - n_s(t)\sin(\omega_c t)
\end{aligned}
\tag{19.13}
$$

Since $a\cos(\omega t) + b\sin(\omega t) = r\cos(\omega t + \theta)$ where $r = \sqrt{a^2+b^2}$, we have that the envelope of $x(t)$ is

$$
r(t) = \sqrt{\left(A_c[1+\alpha m(t)] + n_c(t)\right)^2 + \left(n_s(t)\right)^2}
\tag{19.14}
$$

This envelope is what an AM receiver will recover. When the *S/N* is high, $A_c >> \sigma_n$ and the $n_s(t)$ term can be neglected. Therefore, we can write

$$
r(t) \approx A_c[1+\alpha m(t)] + n_c(t)
\tag{19.15}
$$

Removing the DC term, our recovered signal is

$$
m_r(t) = m(t) + \frac{1}{A_c\alpha} n_c(t)
\tag{19.16}
$$

This is the desired signal $m(t)$ plus the "in-phase" component of the noise scaled by $1/A_c\alpha$. Demodulated (typically audio) signal power (into 1 Ω) is $\langle m(t)^2\rangle$ while the demodulated noise power is $\sigma_n^{\;2}/(A_c\alpha)^2$. For a sinusoidal modulation (19.3), $\langle m(t)^2\rangle = 1/2$, and the audio *S/N* is

$$\begin{aligned} S/N_{audio} &= \alpha^2 \frac{A_c^{\;2}}{2\sigma_n^{\;2}} \\ &= \alpha^2 \frac{A_c^{\;2}}{4N_0 f_m} \\ &= \alpha^2 \left(S/N_{RF}\right) \end{aligned} \tag{19.17}$$

(recalling that the RF bandwidth is twice the modulating frequency). Clearly we want the modulation index to be as large as possible to get the best possible audio. With α limited to maximum of 1, the best we can do is to match the RF *S/N*.

In cases where the RF *S/N* is very low, $\sigma_n >> A_c$, and we can write

$$\begin{aligned} r(t) &= \sqrt{n_c^{\;2}(t)+n_s^{\;2}(t)}\sqrt{1+\frac{2A_c n_c(t)\left[1+\alpha m(t)\right]+A_c^{\;2}\left[1+\alpha m(t)\right]^2}{\left[n_c^{\;2}(t)+n_s^{\;2}(t)\right]}} \\ &\approx \sqrt{n_c^{\;2}(t)+n_s^{\;2}(t)} + \frac{A_c n_c(t)\left[1+\alpha m(t)\right]}{\sqrt{n_c^{\;2}(t)+n_s^{\;2}(t)}} \end{aligned} \tag{19.18}$$

Here we've expanded the terms under the radical in (19.14), factored out $\sqrt{n_c^{\;2}(t)+n_s^{\;2}(t)}$, neglected the term proportional to $A_c^{\;2}$, and then used $\sqrt{1+x} \approx 1+x/2$. The first and strongest term here, is pure noise. The only term containing the signal is

$$\frac{A_c \alpha n_c(t) m(t)}{\sqrt{n_c^{\;2}(t)+n_s^{\;2}(t)}} \tag{19.19}$$

But, here the signal is *multiplied* by noise. That is, this term doesn't even look like the signal; it's equivalent to the signal being AM modulated by noise. The point is, as the *S/N* decreases we eventually get to a point where the signal is effectively wiped out. This is known as the *threshold effect.*

SSB

As we alluded to above, AM is not power efficient and it is not bandwidth efficient. However, merely filtering out the carrier term and one of the sidebands can rectify this. So if the modulation is $m(t) = \sum_k a_k \cos(\omega_k t + \theta_k)$, then

$$\begin{aligned} s_{USB}(t) &= A_c \sum_k a_k \cos\left[\left(\omega_c + \omega_k\right)t + \theta_k\right] \\ s_{LSB}(t) &= A_c \sum_k a_k \cos\left[\left(\omega_c - \omega_k\right)t - \theta_k\right] \end{aligned} \tag{19.20}$$

are the resulting upper and lower *single sideband* signals. Either of these can be used to communicate. The result is *Single Sideband* (SSB) modulation.

SSB is power efficient. No power is wasted in an informationless carrier or a redundant sideband. And SSB is bandwidth efficient. The bandwidth of the RF signal is the same as the bandwidth of the audio signal. However, SSB is more difficult than AM to generate and to demodulate. It also still suffers from the fact that amplitude fluctuations due small-scale fading can mimic modulation. However, as we've seen above, at lower frequencies, or without motion of course, this is not a problem. In many cases SSB is a very attractive technique for analog modulation. A form of SSB is used to encode video in analog television broadcasts.

References

1. Ziemer, R. E. and W. H. Tranter, Principles of Communcations, Houghton Mifflin, 1985, ISBN 0-395-35724-1.
2. Proakis, J. G. and M. Salehi, Communication Systems Engineering, 2nd ed., Prentice Hall, 2002, ISBN 0-13-061793-8.