

Handout 4: Amplitude Modulation

Instructor: Wing-Kin Ma

Suggested Reading: Chapter 3 of Simon Haykin and Michael Moher, *Communication Systems (5th Edition)*, Wiley & Sons Ltd; or Chapter 4 of B. P. Lathi and Z. Ding, *Modern Digital and Analog Communication Systems (4th Edition)*, Oxford University Press.

Let $m(t)$ denote a signal that contains information to be transmitted. The information can take analog form or digital form. In traditional analog radio broadcast, $m(t)$ would be an audio signal. In digital communication systems, $m(t)$ may be a sequence of pulses that carries binary data. The information-bearing signal $m(t)$ will be called a *message signal* for convenience. Also, we will assume that $m(t)$ is a baseband signal with bandwidth W Hz.

In baseband transmission systems such as wireline telephone systems, we may directly transmit $m(t)$. However, there are many scenarios where the physical medium or channel is bandpass. In particular, in wireless communications, a channel operates over a certain frequency range (or frequency band); the operating frequencies are from 30 kHz and upward. In such scenarios, it would be appropriate to consider *carrier modulation*. Let

$$c(t) = A_c \cos(2\pi f_c t)$$

denote a *sinusoidal carrier wave*, where A_c is the carrier amplitude and f_c is the carrier frequency. We wish to use the carrier wave to carry the message signal, so that the message signal can appropriately be transmitted over a bandpass channel.

There are various carrier modulation techniques. Among them, amplitude modulation (AM) is considered the oldest. This handout considers AM.

1 Amplitude Modulation Principle

The amplitude-modulated wave may be described by the following formula

$$s(t) = [1 + k_a m(t)] c(t) \tag{1}$$

$$= A_c [1 + k_a m(t)] \cos(2\pi f_c t), \tag{2}$$

where k_a is a constant and is called the *amplitude sensitivity*. Figures 1.(a)-(b) gives an illustration of the AM process. In the illustration of the AM wave in Figure 1(b), the constant k_a is adjusted such that $1 + k_a m(t) \geq 0$ for all t . It is observed that the *envelope* of the AM wave takes the same shape as the message signal (more precisely, the waveform $1 + k_a m(t)$). In fact, the idea of AM is to *use the envelope of the modulated wave $s(t)$ to carry the message signal*.

There is a requirement for AM to operate properly. Specifically, we must have

$$|k_a m(t)| \leq 1, \quad \text{for all } t. \tag{3}$$

The condition in (3) implies that $1 + k_a m(t) \geq 0$ for all t . Figure 1.(c) shows a situation where $1 + k_a m(t) < 0$ for some t . We see that the envelope of the AM wave now becomes a distorted

version of the message signal. This phenomenon is sometimes known as *overmodulation*, which can happen when k_a is set too large.

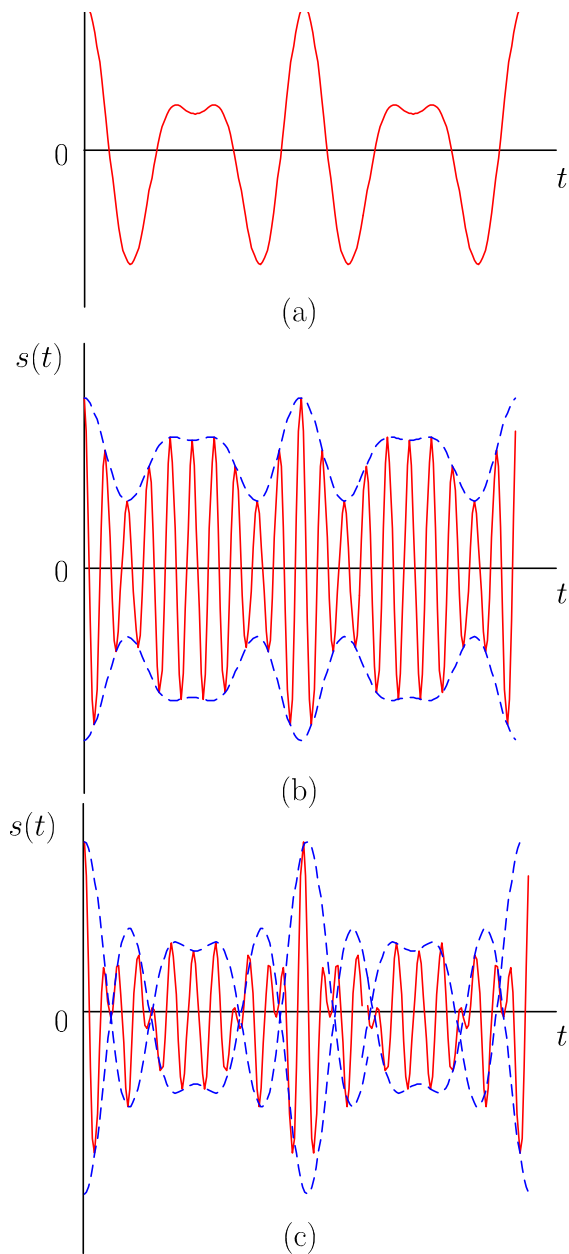


Figure 1: An illustration of the AM process. (a) The message signal $m(t)$. (b) The AM wave $s(t)$ when $|k_a m(t)| \leq 1$ holds for all t . (c) The AM wave $s(t)$ when we have $|k_a m(t)| > 1$ for some t .

2 AM Spectrum

We are interested in the spectrum of the AM signal. From (2), it can be easily shown that the Fourier transform of $s(t)$ is

$$\begin{aligned} S(f) &= F[A_c \cos(2\pi f_c t)] + F[A_c k_a m(t) \cos(2\pi f_c t)] \\ &= \frac{A_c}{2} [\delta(f - f_c) + \delta(f + f_c)] + \frac{A_c k_a}{2} [M(f - f_c) + M(f + f_c)] \end{aligned} \quad (4)$$

where $\delta(f)$ is the Dirac delta function, and $M(f)$ is the Fourier transform of $m(t)$. Figure 2 illustrates the AM spectrum. In the figure, we assume that the carrier frequency f_c is much greater than the message bandwidth W , which is true in practically relevant scenarios. We observe that the message signal $m(t)$ is frequency-shifted from baseband to a frequency interval $[f_c - W, f_c + W]$ (if we just look at positive frequencies). Also, there is a pure carrier term located at frequency f_c , that is, $\frac{A_c}{2}\delta(f - f_c)$ (over positive frequencies). From Figure 2 we conclude that the transmission bandwidth of AM is

$$B_T = 2W \text{ Hz.}$$

Notice that the AM bandwidth is twice of the message bandwidth.

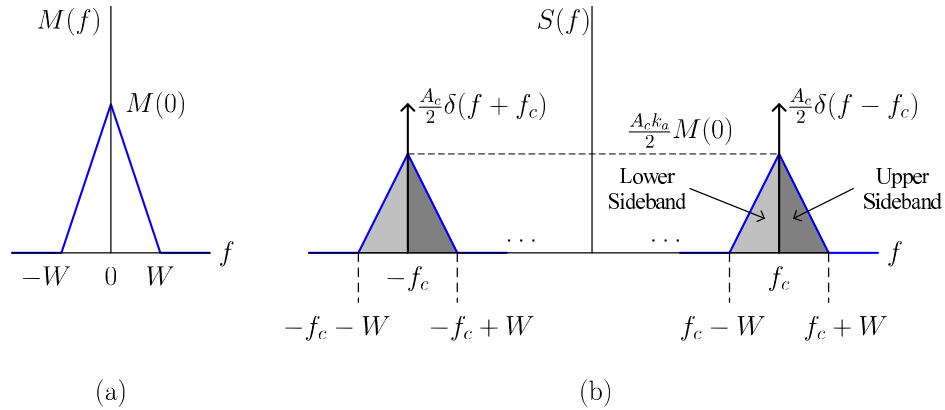


Figure 2: (a) Spectrum of the message signal. (b) Spectrum of the corresponding AM signal.

We should also note that over positive frequencies, the portion of the AM spectrum lying above the carrier frequency f_c is called the *upper sideband*; and the portion of the AM spectrum lying below the carrier frequency f_c is called the *lower sideband*.

3 AM Modulation and Demodulation

The modulation process of AM, or equivalently, generation of AM waves, can be achieved in several ways; usually it is done by circuits. Here, we consider a modulator approach that will be considered again in some other carrier modulation schemes to be introduced in later lectures. Figure 3 uses a simplified block diagram to illustrate the modulator of interest. The process is nothing more than that of implementing the multiplication process in (1). The local oscillator is a device that generates

the sinusoidal carrier wave. The product modulator is a device that performs the multiplication in (1). Both the local oscillator and the product modulator are built by electronic devices and circuits¹. This modulator is sometimes known as a multiplier modulator, or a mixer.

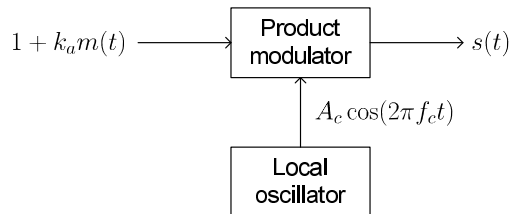


Figure 3: A modulator system diagram for AM.

We are also concerned with *demodulation*, which is the process of recovering the message signal from the modulated signal. As in modulation, there are several ways to achieve demodulation. One common way is to use a simple device called the *envelope detector*, shown in Figure 4. Readers are referred to the textbooks for more insight about how the envelope detector works intuitively.

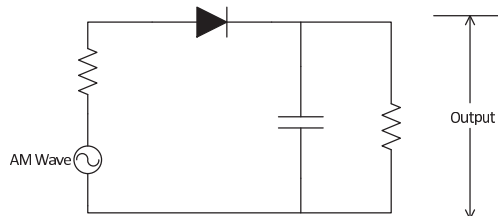


Figure 4: The envelope detector.

3.1 The Rectifier Detector for AM Demodulation: A Case Study

We consider an alternative AM demodulator approach whose block diagram is illustrated in Figure 5. The idea is intuitively similar to the simple envelope detector in Figure 4. The AM wave $s(t)$ first goes through a half-wave rectifier, which clips the signal to zero when the signal value falls below zero; the rectifier can be built by a diode and a resistor circuit. A half-wave rectified signal from an AM wave is illustrated in Figure 6.(c) (it is generated by MATLAB). The signal output at the half-wave rectifier may be expressed as

$$v(t) = [1 + k_a m(t)]w(t) \quad (5)$$

¹It should however be noted that recent technology allows direct generation of a carrier-modulated wave from a digital-to-analog converter [1]. Consequently, the oscillator and product modulator exist in the software or digital signal processing domain!

where $w(t)$ is a periodic signal of period $T_0 = \frac{1}{f_c}$ and its expression under time interval $[-T_0/2, T_0/2]$ is

$$w(t) = \begin{cases} 0, & -\frac{T_0}{2} \leq t \leq -\frac{T_0}{4} \\ A_c \cos(2\pi f_c t), & -\frac{T_0}{4} < t < \frac{T_0}{4}, \\ 0, & \frac{T_0}{4} \leq t \leq \frac{T_0}{2} \end{cases}$$

Then, a lowpass filter is applied to $v(t)$ in an attempt to recover the message signal $m(t)$. The lowpass filter can intuitively be viewed as a smoothing process to remove the fast varying signal components, and to retain the slowly varying signal components which is the message signal itself. Figure 6.(d) gives an illustration on the lowpass filter output.

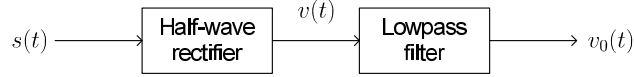


Figure 5: The rectifier detection system diagram.

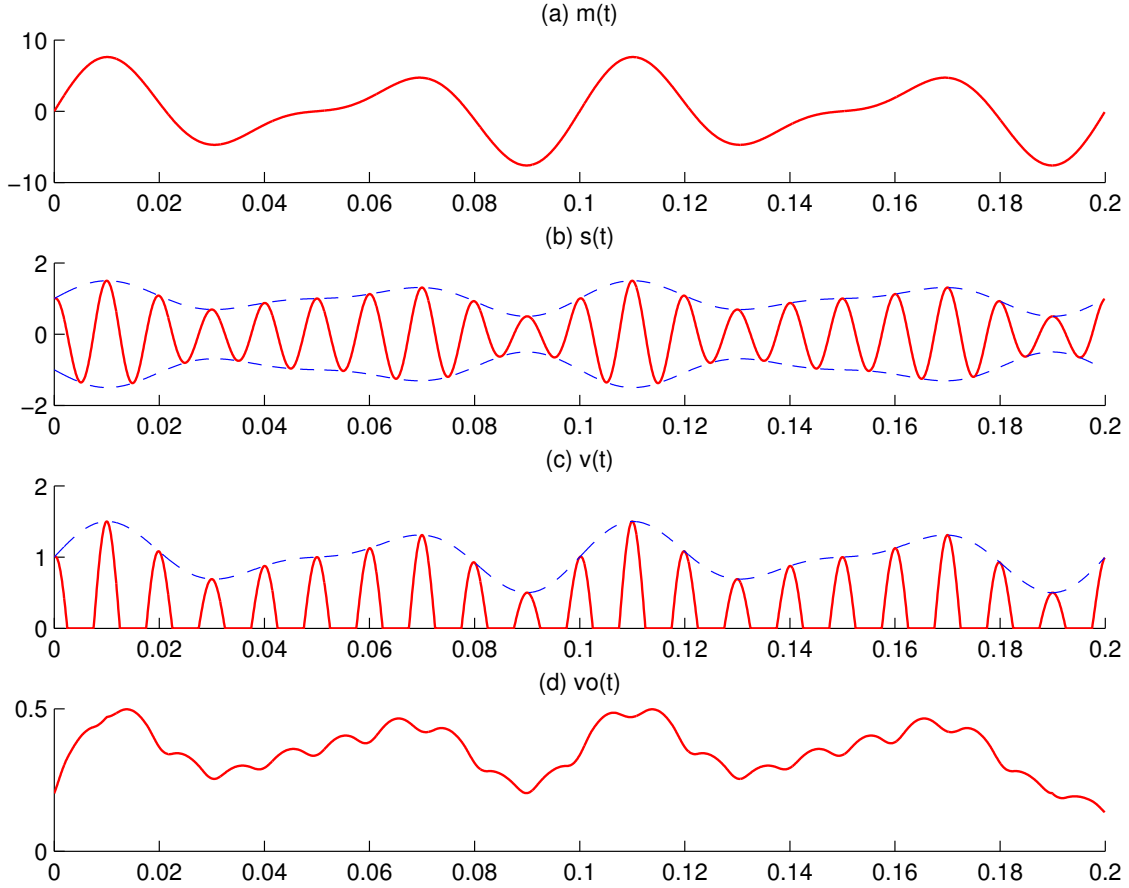


Figure 6: Illustration of the signals in rectifier detection.

Here is an analysis challenge—Can we prove that the rectifier detection method above guarantees perfect recovery of $m(t)$? If yes, how should the lowpass filter be designed? This is a relatively advanced problem, and we will leave it in this handout. We may go through it in class if time allows.

4 Drawbacks

Amplitude modulation and demodulation systems are easy to implement; they can be built by relatively simple circuits. However, AM has two drawbacks. First, to ensure that the envelope of the AM wave faithfully preserves the shape of the message signal, we must spend power on the pure carrier component, which bears no information. In other words, power is wasted on the pure carrier component. Second, it is redundant to transmit both the upper sideband and lower sideband—there is a direct relationship between the upper and lower sidebands (specifically, $M(-f) = M^*(f)$, which is obtained from the conjugate function property and the fact that $m(t)$ is real-valued). The consequence is the use of twice of the message bandwidth to transmit a message signal—a waste of transmission bandwidth.

References

- [1] G. Engel, D. E. Fague, and A. Toledano, “RF digital-to-analog converters enable direct synthesis of communications signals,” *IEEE Communications Magazine*, vol. 50, no. 10, pp. 108–116, 2012.