Chapter 5

Amplitude Modulation

So far we have developed basic signal and system representation techniques which we will now apply to the analysis of various analog communication systems. In particular, we will study:

- **Amplitude modulation** (AM) and its variants;
- Angle modulation including **Frequency modulation** (FM) and **Phase modulation** (PM).

Let $m(t)$ be a baseband message signal that carries the information that we would like to transmit in an efficient and reliable manner. If we assume that $m(t)$ is band-limited to B-Hz, then its transmission in its baseband format requires a transmission bandwidth of B-Hz. In this chapter we will address the question of transmitting $m(t)$ over a bandpass communication channel. Transmission over a bandpass communications channel requires a shifting the spectrum of $m(t)$ to higher frequencies compatible with the characteristics of the communication channel. At the receiver we reverse this process to recover the message signal from the received waveform.

Before we proceed to fully formulate this problem let us remind ourselves why we need to shift the spectrum of $m(t)$ to a higher frequencies for transmission. In Section 3.9 we stated that the design of practical, economically implementable bandpass systems support transmission bandwidths that are within 1–10% of their center frequencies. Furthermore, we need to devise a mechanism that will allow multiple signals to share the same communication channel; one of the most commonly used methods to achieve this objective is frequency-division multiplexing (FDM). In its most basic form, FDM requires that message signals to be transmitted occupy non-overlapping frequency bands; hence, we need to be able to shift the spectra of these message signals to different frequency bands. In addition, the channel loss characteristics, availability of channel bandwidth, constraints on physical size of the equipment (e.g., antenna size, packaging dimensions, etc.) must also be taken into consideration.

In this chapter we will first formulate a general signal processing framework for the signal processing operation **modulation**. We will then discuss **amplitude modulation** and **amplitude demodulation** techniques and their variants. Chapter **??** will discuss **phase** and **frequency modulation/demodulation** techniques.

Definition 5.1. The process by which some characteristic of a carrier signal signal is varied in accordance with a modulating signal is called **modulation**.

5.1 Modulation

Let:

$$m(t) = \text{modulating signal};$$
$$c(t) = \text{carrier signal};$$
$$\varphi(t) = \text{modulated signal}.$$

The modulating signal $m(t)$ is the message signal which carries the information, e.g. voice, image, data, video, etc., that we want to transmit. Under the assumption that $c(t)$ is a sinusoidal signal (the

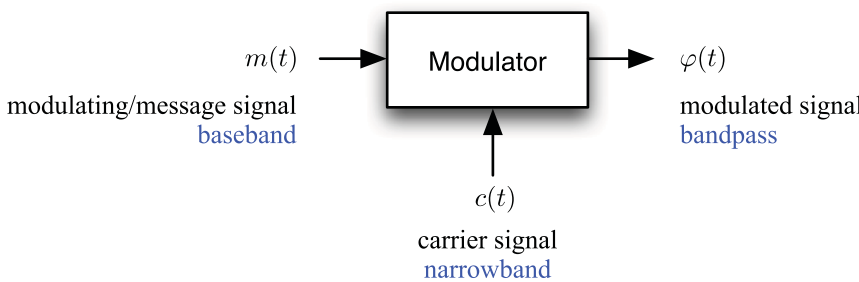


Figure 5.1: The **modulation** process.

most common form of a carrier signal), the modulated signal $\varphi(t)$ can be expressed in a generic form as:

$$\varphi(t) = a(t)\cos\theta(t), \tag{5.1}$$

where $a(t)$ is the (possibly) time-varying amplitude and $\theta(t)$ is the (possibly) time-varying phase of the modulated signal. Observe that $\varphi(t)$ represents a rotating phasor of time-varying amplitude—its instantaneous amplitude is determined by $a(t)$—and generalized angle $\theta(t)$. The instantaneous frequency $f_i(t)$ of the modulated signal $\varphi(t)$ can be calculated as:

$$f_i(t) = \frac{1}{2\pi}\frac{d}{dt}\theta(t). \tag{5.2}$$

An **unmodulated** carrier signal will have:

$$a(t) = A_c, \tag{5.3a}$$
$$\theta(t) = 2\pi f_c t + \theta_0, \tag{5.3b}$$

where A_c is the unmodulated carrier amplitude, θ_0 is the arbitrary initial phase term (which we can assume $\theta_0 = 0$ without loss of generality) and f_c is the carrier frequency. Observe that for an unmodulated carrier the signal amplitude will be constant and independent of $m(t)$ and the generalized angle term will also be independent from the modulating signal $m(t)$.

Different modulation schemes will alter different parts of $\varphi(t)$: amplitude modulation will render $a(t)$ to be a function of $m(t)$ whereas the phase and frequency modulation will render $\theta(t)$ to be a function of $m(t)$. In particular, we can formulate the amplitude, phase and frequency modulation schemes as follows:

Amplitude Modulation (AM): For this modulation scheme we have:

$$a(t) = g\big[m(t)\big]; \tag{5.4a}$$

$$\theta(t) = \text{independent of } m(t); \tag{5.4b}$$

for some function $g[\,.\,]$. Observe that in the AM case the information contained in $m(t)$ is embedded in the time-varying amplitude $\varphi(t)$.

Frequency Modulation (FM): For the FM case we have:

$$a(t) = \text{constant and independent of } m(t); \tag{5.5a}$$

$$f_i(t) = f_c + K_f m(t). \tag{5.5b}$$

$$\begin{aligned} \theta(t) &= 2\pi \int_0^t f_i(\lambda) d\lambda; \\ &= 2\pi f_c t + 2\pi K_f \int_0^t m(\lambda) d\lambda + \theta_0. \end{aligned} \tag{5.5c}$$

where K_f is the frequency sensitivity parameter. For FM signals the amplitude of $\varphi(t)$ is constant whereas the information contained in $m(t)$ is embedded in its instantaneous frequency $f_i(t)$.

Phase Modulation (PM): For the PM case we have

$$a(t) = \text{constant and independent of } m(t); \tag{5.6a}$$

$$\theta(t) = 2\pi f_c t + 2\pi K_p m(t) + \theta_0; \tag{5.6b}$$

$$\begin{aligned} f_i(t) &= \frac{1}{2\pi}\frac{d}{dt}\theta(t); \\ &= f_c + K_p \frac{d}{dt} m(t). \end{aligned} \tag{5.6c}$$

where K_p is the phase sensitivity parameter. For PM signals the amplitude of $\varphi(t)$ is constant and the the information contained in $m(t)$ is embedded in its generalized phase $\theta(t)$.

Observe that we can consider PM as a special case of FM or conversely FM as a special case of PM. We will further study these two important angle modulation techniques in Chapter **??**.

Let us consider the modulated waveform $\varphi(t)$ with spectrum $\Phi(f)$. We will frequently refer to and/or differentiate modulation schemes based on:

- **Carrier Term:** The modulated waveform $\varphi(t)$ may or may not include a separate **carrier term**. If a separate carrier term is present, the spectrum of $\varphi(t)$ shows line spectrum components at $\pm f_c$ where f_c is the carrier frequency.
- **Lower Side-Band:** This terms refers to frequency components in $\Phi(f)$ for $|f| < f_c$. Depending on the modulation scheme the lower side-band may or may not be present.

- **Upper Side-Band:** This terms refers to frequency components in $\Phi(f)$ for $|f| > f_c$. Depending on the modulation scheme the upper side-band may or may not be present.

Figure 5.2 shows these components as part of the spectrum of a typical modulated waveform.

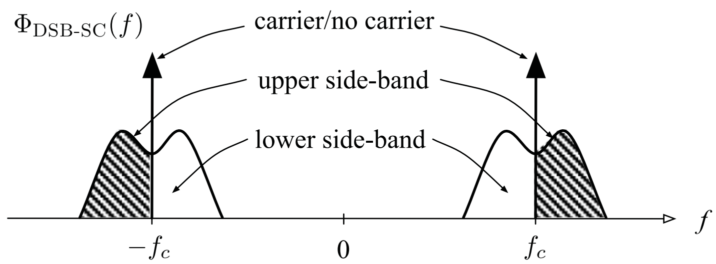


Figure 5.2: Elements of the spectrum of a modulated waveform.

5.2 Double Sideband Amplitude Modulation

Let us consider the simplest and the most intuitive amplitude modulation case. Let $m(t)$ be the baseband message/modulating signal as before, band-limited to B-Hz, and let the carrier be a sinusoid with amplitude A_c and carrier frequency f_c:

$$c(t) = A_c \cos 2\pi f_c t, \tag{5.7}$$

such that

$$\varphi_{\text{DSB-SC}}(t) = A_c m(t) \cos 2\pi f_c t, \tag{5.8}$$

with $f_c \gg B$. It then follows that the spectrum of $m(t)$ consists of amplitude-scaled and frequency-shifted versions of $M(f)$:

$$\Phi_{\text{DSB-SC}}(f) = \frac{A_c}{2}\left[M(f - f_c) + M(f + f_c)\right]. \tag{5.9}$$

Figure 5.3 displays sample waveforms and their respective spectra for the message, the carrier and the modulated signal. We use the term **double sideband suppressed carrier** (DSB-SC) amplitude modulation to refer to this modulation scheme—double sideband as both sidebands are present in $\Phi_{\text{DSB-SC}}(f)$ and suppressed carrier as $\varphi_{\text{DSB-SC}}(t)$ does not have a separate carrier term indicated by the lack of line spectral components at $\pm f_c$ in $\Phi_{\text{DSB-SC}}(f)$. Upon closed inspection of the modulated waveform $\varphi_{\text{DSB-SC}}(t)$ and its respective spectrum $\Phi_{\text{DSB-SC}}(f)$, we further make the following observations:

- If the baseband message signal $m(t)$ has the bandwidth B-Hz, then the DSB-SC amplitude modulated signal $\varphi_{\text{DSB-SC}}(t)$ requires a transmission bandwidth of $2B$-Hz.

- We can use a product modulator to generate $\varphi_{\text{DSB-SC}}(t) = m(t)c(t)$.

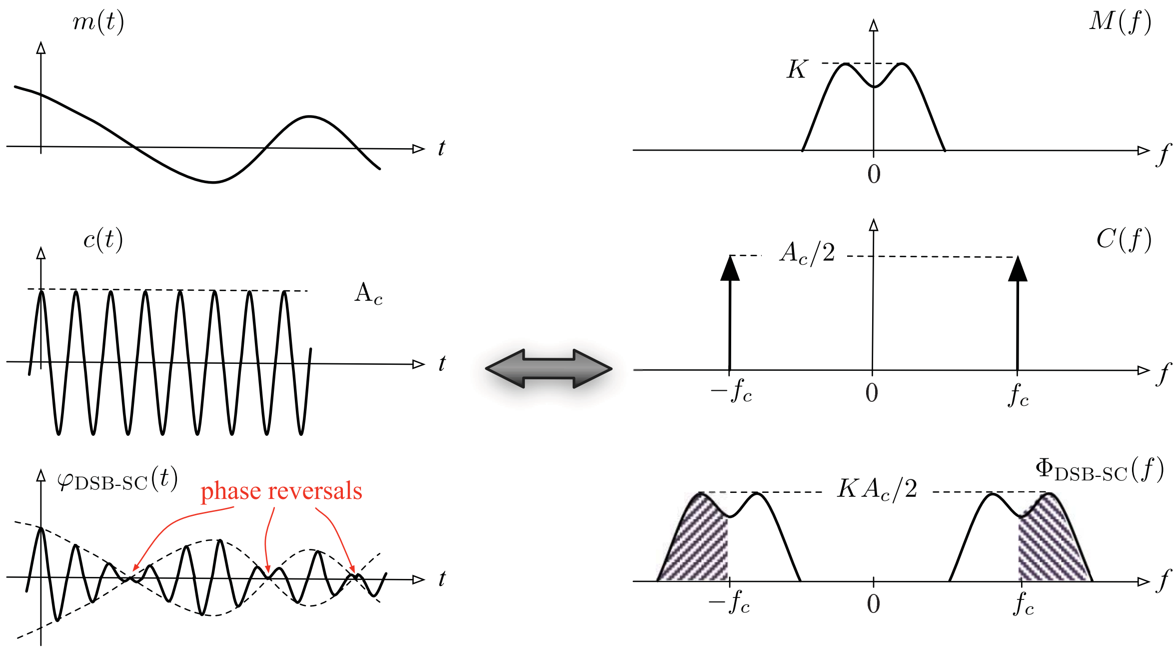


Figure 5.3: DSB-SC amplitude modulation.

The recovery of the message signal $m(t)$, i.e., the demodulation of $\varphi_{\text{DSB-SC}}(t)$, can be achieved in a manner similar to generating $\varphi_{\text{DSB-SC}}(t)$ from $m(t)$. We observe that $\varphi_{\text{DSB-SC}}(t)$ has been generated from $m(t)$ by frequency shifting $M(f)$ by modulating the carrier $c(t)$. Therefore, we can demodulate $\varphi_{\text{DSB-SC}}(t)$ by first generating the product:

$$\varphi_{\text{DSB-SC}}(t)\cos\omega_c t = A_c\, m(t)\cos^2\omega_c t, \tag{5.10a}$$

$$= \frac{A_c}{2}m(t) + \frac{A_c}{2}m(t)\cos 2\omega_c t, \tag{5.10b}$$

where we used the trigonometric identity $\cos^2 x = (1+\cos 2x)/2$. The first term on the right-hand side of Equation (5.10b) is the magnitude-scaled baseband message signal whereas the second term has a narrowband spectrum centered at $\pm 2f_c$. Under the assumption that $f_c \gg B$ we can recover $m(t)$ by lowpass filtering $\varphi_{\text{DSB-SC}}(t)\cos\omega_c t$. Figure 5.5 depicts the block of the modulator used

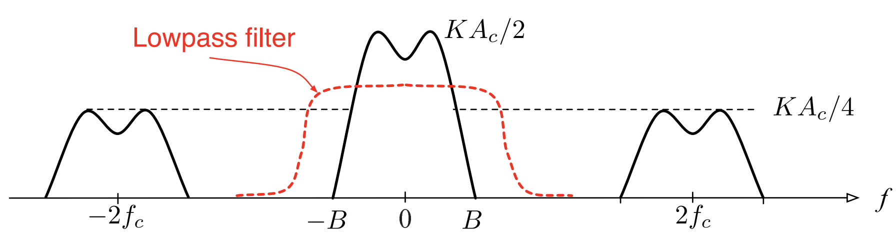


Figure 5.4: Spectrum of $\varphi_{\text{DSB-SC}}(t)\cos\omega_c t$ and the recovery of $m(t)$.

to generate a DSB-SC amplitude modulated signal and the corresponding demodulator used to recover the message signal. A closer look at the time-domain waveforms generated at various points

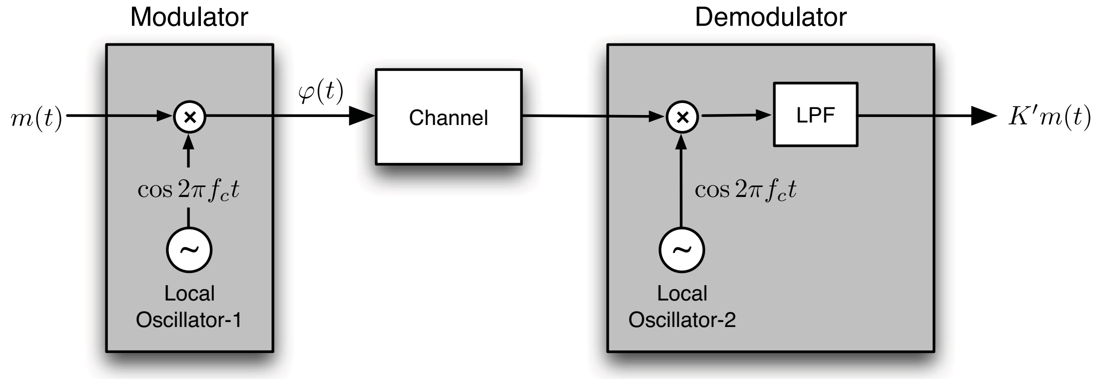


Figure 5.5: Generation and demodulation of DSB-SC amplitude modulated signals.

within the demodulator allow us to have abetter understanding of the operational characteristics of the demodulator.

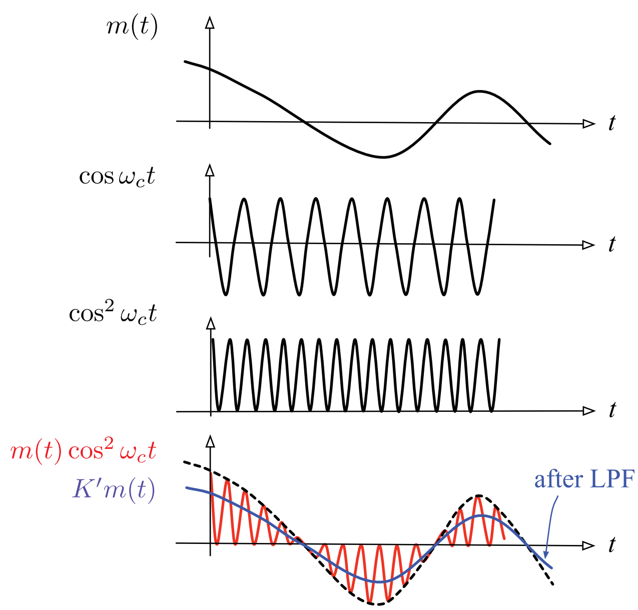


Figure 5.6: Time-domain waveforms at the demodulator.

5.2.1 Coherent Detection

Proper demodulation of DSB-SC amplitude modulated signals as presented in Equation (5.10) requires that the local oscillators at the modulator and the demodulator are synchronized, i.e., both oscillators should generate sinusoids with identical frequencies that are phase coherent: if the output of the local oscillator at the modulator is $\cos(2\pi f_1 t + \phi_1)$ and the output of the local oscillator at the demodulator is $\cos(2\pi f_2 t + \phi_2)$, then $f_1 = f_2 = f_c$ and $\phi_1 = \phi_2$.

To illustrate the significance of synchronization, let $\varphi_{\text{DSB-SC}}(t) = A_c m(t) \cos 2\pi f_c t$ and let $l_d(t)$ be the sinusoid generated by the local oscillator in the demodulator:

$$l_d(t) = \cos\big(2\pi(f_c + \Delta f)t + \phi_0\big). \tag{5.11}$$

Δf represents deviation from the carrier frequency and ϕ_0 represents phase difference between the modulator and the demodulator. The demodulator operates by first multiplying the received waveform $\varphi_{\text{DSB-SC}}(t)$ with the output of local oscillator. This operation generates the signal:

$$\varphi_{\text{DSB-SC}}(t) l_d(t) = A_c\, m(t) \cos\big(2\pi f_c t\big) \cos\big(2\pi(f_c + \Delta f)t + \phi_0\big), \tag{5.12a}$$

$$= \frac{A_c}{2} m(t) \cos\big(2\pi \Delta f t + \phi_0\big) + \frac{A_c}{2} m(t) \cos\big(2\pi(2f_c + \Delta f)t + \phi_0\big). \tag{5.12b}$$

We assume that the mismatch between the local oscillators in the modulator and the demodulator is small, i.e., $\Delta f \ll f_c$. Therefore, the spectrum of the second term in Equation (5.12b) will be centered on $\pm 2f_c$ and will be filtered out by the lowpass filter in the demodulator such that the lowpass filter output equals:

$$y(t) = \frac{A_c}{2} m(t) \cos\big(2\pi \Delta f t + \phi_0\big). \tag{5.13}$$

Thus, $y(t)$ oscillates at the slow rate Δf which is due to the frequency mismatch between the two local oscillators. We can safely assume $\Delta f = 0$ by arguing that we have full knowledge of the carrier frequency of the signal that we want to demodulate and that modern oscillators can be tuned very precisely to any desired frequency. Under the $\Delta f = 0$ assumption we can express the output of the demodulator becomes:

$$y(t) = \frac{A_c}{2} m(t) \cos \phi_0. \tag{5.14}$$

If we are lucky and the local oscillators are perfectly synchronized that is $\phi_0 = 0$, then $y(t) = K' m(t)$—the constant K' takes into account not only the term $Ac/2$ in Equation (5.14), but also any further magnitude changes the signal may experience during transmission and/or processing. The worst case scenario occurs when $\phi_0 = \pi/2$ radians such that $\cos \pi/2 = 0$ and therefore $y(t) = 0$. Other values for ϕ_0 will result in varying degrees of attenuation or even phase reversal at the demodulator output. Hence, we conclude that successful demodulation of DSB-SC amplitude modulated signals requires accurate synchronization of the two local oscillators. We will use the term **synchronous detection** or **coherent detection** to refer to signal demodulation where the local oscillator in the demodulator is synchronized with the local oscillator used at the modulator.

5.3 Generation of AM Signals

In this section we we will introduce and briefly discuss some of the most prominent methods for generating AM signals.

5.3.1 Non-linear Modulator

Consider the non-linear device described by the input-output relation:

$$[\text{output}] = a[\text{input}] + b[\text{input}]^2, \tag{5.15}$$

where $a, b \in \Re$. Figure 5.7 shows how we can use this non-linear device together with an appropriately designed bandpass filter to generate a DSC-SC signal. To show that the output of this system

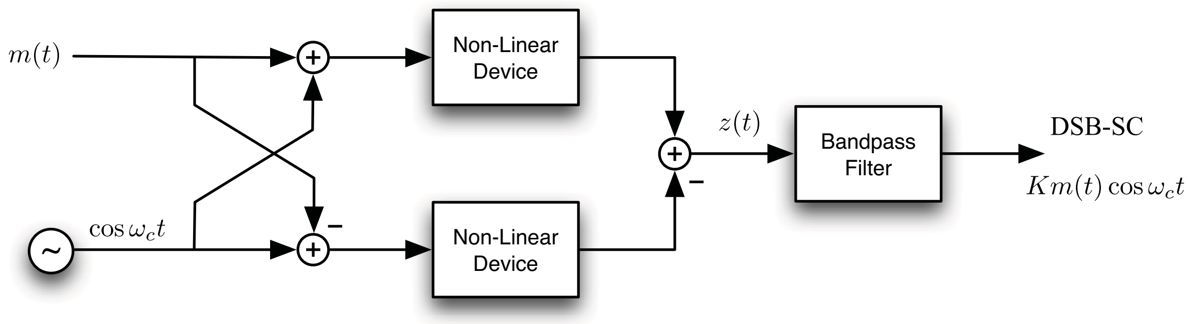


Figure 5.7: AM signal generation using non-linear modulator.

indeed represents a DSC-SC signal we first evaluate $z(t)$ as the signal generated by combining the outputs of the non-linear devices:

$$\begin{aligned} z(t) &= a\,[m(t) + \cos\omega_c t] + b\,[m(t) + \cos\omega_c t]^2 \\ &\qquad - a\,[-m(t) + \cos\omega_c t] - b\,[-m(t) + \cos\omega_c t]^2; && (5.16a) \\ &= a\,m(t) + a\,\cos\omega_c t + b\,m^2(t) + 2b\,m(t)\cos\omega_c t + b\,\cos^2\omega_c t \\ &\qquad + a\,m(t) - a\,\cos\omega_c t - b\,m^2(t) + 2b\,m(t)\cos\omega_c t - b\,\cos^2\omega_c t; && (5.16b) \\ &= 2a\,m(t) + 4b\,m(t)\cos\omega_c t. && (5.16c) \end{aligned}$$

The first term on the right-hand side of Equation (5.16c) is a baseband signal that will be filtered out by the bandpass filter, whereas the second term represents the DSB-SC signal that we want to retain. Therefore, the bandpass filter must have its passband centered at f_c and have a bandwidth of $2B$-Hz where B is the bandwidth of the baseband message signal $m(t)$.

5.3.2 Switching Modulator

The multiplication operation required for modulation can be replaced by a much simpler operation, namely **switching**. Observe that the DSB-SC amplitude modulated signal $\varphi_{\text{DSB-SC}}(t)$ can be obtained by multiplying $m(t)$ with any periodic signal with fundamental frequency f_c. This was precisely the same approach we used as part of our discussion of the natural sampling process in Section 4.5. Let $\phi(t)$ be such a periodic waveform; as $\phi(t)$ is a periodic waveform it can be expanded in a complex Fourier series with coefficients $\{C_n\}$:

$$\phi(t) = \sum_{n=-\infty}^{\infty} C_n e^{jn\omega_c t}$$

such that

$$\mathcal{F}\big[\, m(t)\phi(t) \,\big] = M(f) * \mathcal{F}\big[\, \phi(t) \,\big]; \tag{5.17a}$$
$$= M(f) * \sum_n C_n \delta(f - nf_c); \tag{5.17b}$$
$$= \sum_n C_n M(f - nf_c). \tag{5.17c}$$

Thus the spectrum of the product $m(t)\phi(t)$ is the spectrum of the modulating waveform $M(f)$ shifted to $\pm f_c, \pm 2f_c, \ldots$. If we pass this product through a bandpass filter with bandwidth $2B$-Hz and tuned to f_c then the output of the bandpass filter will be the desired DSB-SC amplitude modulated waveform. Figure 5.9 shows the time-domain signals $m(t)$, $w(t)$ and $\varphi_{\text{DSB-SC}}(t)$ that

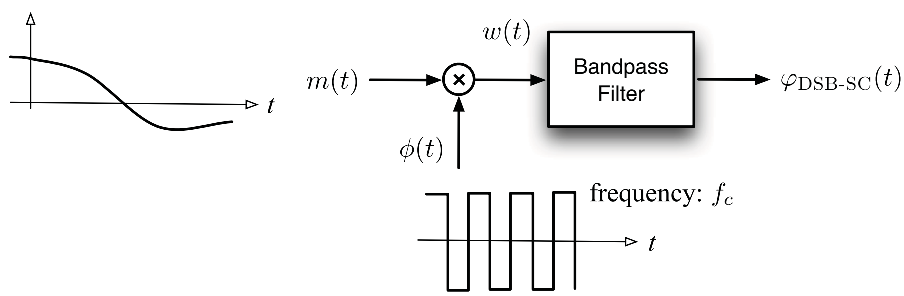


Figure 5.8: DSB-SC AM signal generation using a switching modulator.

we encounter in the switching modulator and their respective spectra.

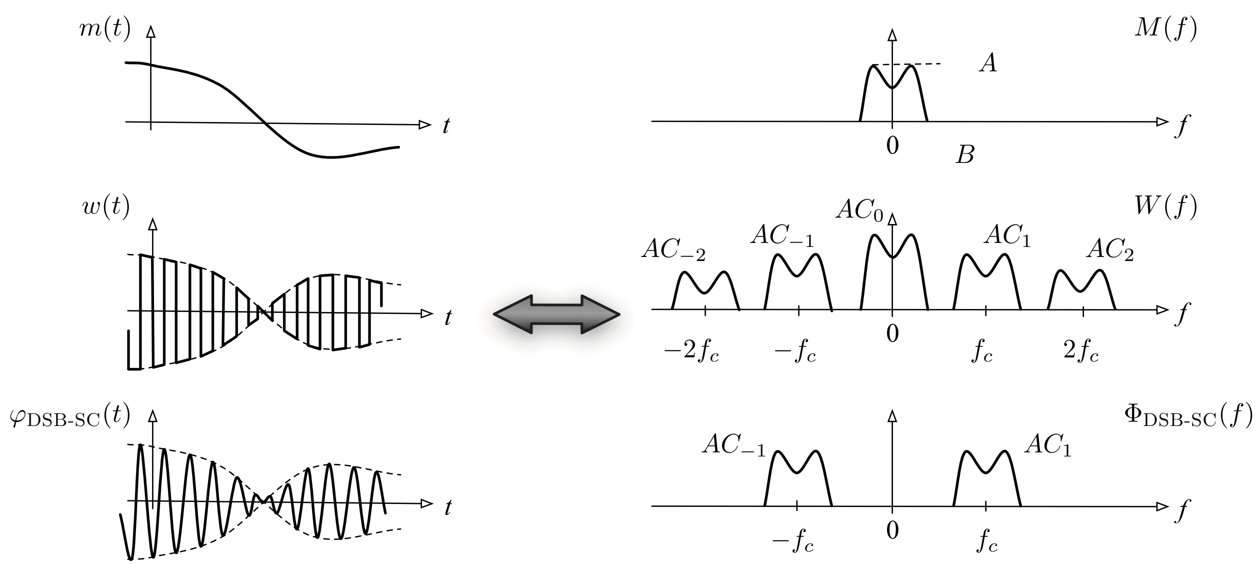


Figure 5.9: Time-domain waveforms and their spectra encountered in the switching modulator.

There are other AM generation techniques. Please refer to the course reference text for further information and discussion. We will use the terms **mixing/frequency conversion/heterodyning** interchangeably to refer to the process of multiplication followed by bandpass filtering as demonstrated in the two AM generation techniques discussed above.

5.4 Amplitude Modulation (AM)

Modulated waveform with suppressed carrier terms require fairly complex circuitry at the receiver to acquire and maintain phase synchronization which make the receivers expensive to manufacture. In applications where we have one or few transmitters and a much, much larger number of receivers (e.g. AM/FM radio broadcasting) it makes economic sense that the receivers are as simple as possible.

To facilitate simple demodulation we consider the idea of transmitting a separate carrier term in the same frequency band as a DSB-SC amplitude modulated signal. This approach will adversely affect the d.c. response of the modulating signal $m(t)$; however, due to the lack of any significant d.c. content of typical message signals this will still be an acceptable solution. Let $\varphi_{\text{AM}}(t)$ represents the corresponding modulated waveform:

$$\varphi_{\text{AM}}(t) = m(t)\cos\omega_c t + A_c \cos\omega_c t, \tag{5.18a}$$

$$= \left[\, A_c + m(t) \,\right] \cos\omega_c t, \tag{5.18b}$$

such that the spectrum of the resulting amplitude modulated waveform equals:

$$\Phi_{\text{AM}} = \frac{1}{2}\big[M(f - f_c) + M(f + f_c)\big] + \frac{A_c}{2}\big[\delta(f - f_c) + \delta(f + f_c)\big]. \tag{5.19}$$

Figure 5.10 shows that $\Phi_{\text{AM}}(t)$ includes a separate carrier term in addition to both the lower and the upper side-bands. Therefore, this signal processing technique can be described as **double side-band, large carrier** (DSB-LC) amplitude modulation. However, due to the prevalence of this technique, as in radio broadcasting, we use the term **amplitude modulation** (AM) to describe the technique described in Equation (5.18). Figure 5.11 compares the time-domain waveforms and

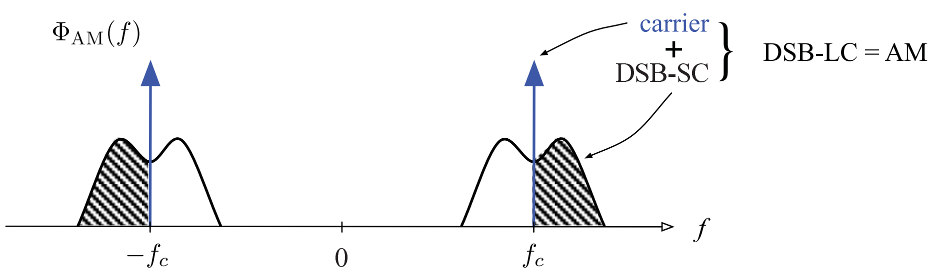


Figure 5.10: Spectrum of an DSB-LC/AM signal.

their respective spectra for DSB-SC and AM signals.

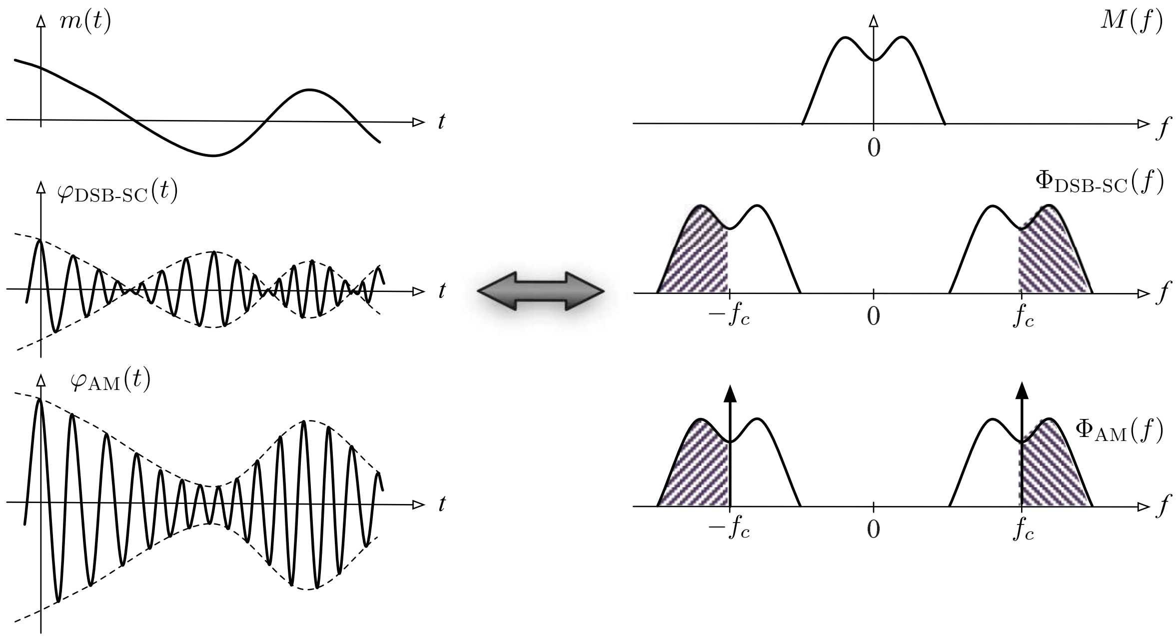


Figure 5.11: Time-domain waveform and spectrum comparison for DSB-SC and AM signals.

The *amplitude* and *envelope* functions are important components of an AM signal that allow the calculation of key AM signal parameters.

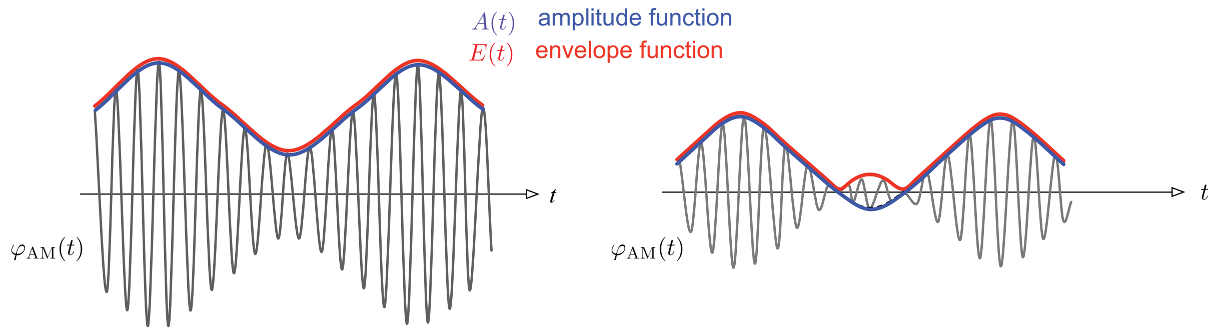
Definition 5.2. The **amplitude** function $A(t)$ of the AM signal $\left[\, A_c + m(t) \,\right] \cos \omega_c t$ is:

$$A(t) = \left[A_c + m(t)\right]. \tag{5.20}$$

Definition 5.3. The **envelope** function $E(t)$ of the AM signal $\left[\, A_c + m(t) \,\right] \cos \omega_c t$ is:

$$E(t) = \left|A_c + m(t)\right|. \tag{5.21}$$

The use of an **envelope detector** as an AM demodulator requires that the amplitude function $A(t)$ remains positive, i.e., $A(t) = [A_c + m(t)] \geq 0$, at all times. If this happens to be the case, then the envelope function equals to the amplitude function: $E(t) = \left|A_c + m(t)\right| = [A_c + m(t)] = A(t)$. Under these conditions an envelope detector together with a DC blocker can easily extract the modulating signal $m(t)$ from the envelope function $E(t)$. Conversely, if $A(t) = [A_c + m(t)] < 0$ for some t, then $E(t) =$ envelope will no longer be a scaled and shifted version of the modulating signal $m(t)$ and $[A_c + m(t)]$ will experience phase reversals at zero crossings. As a result, we will have to use a synchronous detector similar to the demodulator as in the case with DSB-SC amplitude modulated signals, eliminating all the inherent advantages of simple demodulation of AM signals. Figure 5.12 depicts the amplitude and envelope functions of a single-tone modulated AM signal for different cases.



Case 1: $A(t) \geq 0, \ \forall t$
$E(t) = A(t)$

Case 2: $A(t) \geq 0, \ \text{for some } t \text{ and}$
$A(t) < 0, \ \text{for some } t.$
$E(t) = \big|A(t)\big|$

Figure 5.12: Amplitude and envelope functions of a modulated waveform.

Let:

$$m_p = \max_t |m(t)|, \tag{5.22}$$

be the maximum absolute value of the message/modulating signal $m(t)$ which will in turn allow us to define another key parameter associated with AM signals.

Definition 5.4. The **modulation index** of an AM signal represented by μ, is defined as:

$$\mu = \frac{m_p}{A_c} = \frac{\max_t |m(t)|}{A_c}, \tag{5.23}$$

where m_p is the maximum absolute value of $m(t)$ and A_c is the carrier amplitude.

Definition (5.4) implies that the condition "$[A_c + m(t)] \geq 0$ *at all times*" is equivalent to

$$0 \leq \mu \leq 1. \tag{5.24}$$

If the positive and negative swings of the modulating signal $m(t)$ are equal in magnitude, i.e., if $|\max_t m(t)| = |\min_t m(t)|$, then the definition of modulation index given in Equation (5.23) is sufficient. However, if $m(t)$ is not symmetric, we need to extend the definition of the modulation index. As a matter of fact, AM broadcasting standards impose separate conditions on the modulation index based on positive and negative swings of the amplitude function $A(t)$. Let

$$A_{\text{max}} = \max_t A(t) = \max_t \big[A_c + m(t)\big], \tag{5.25a}$$

$$A_{\text{min}} = \min_t A(t) = \min_t \big[A_c + m(t)\big], \tag{5.25b}$$

Definition 5.5.

$$\textbf{Modulation Index:} \qquad \mu = \frac{A_{\text{max}} - A_{\text{min}}}{2A_c}. \tag{5.26a}$$

$$\textbf{Positive Modulation Index:} \qquad \mu_+ = \frac{A_{\text{max}} - A_c}{A_c}. \tag{5.26b}$$

$$\textbf{Negative Modulation Index:} \qquad \mu_- = \frac{A_c - A_{\text{min}}}{A_c}. \tag{5.26c}$$

Observe that if the positive and negative swings of $m(t)$ are equal in magnitude, i.e., if $|\max_t m(t)| = |\min_t m(t)|$, then $\mu = \mu_+ = \mu_-$. The following examples demonstrate how μ, μ_+ and μ_- are related when working with a variety of message signals.

Example 5.1. Consider the single-tone signal message signal $m(t) = A_m \cos\omega_m t$ which we use to modulate a carrier and generate the AM signal:

$$\varphi_{\text{AM}}(t) = \left[\, A_c + A_m \cos\omega_m t \,\right] \cos\omega_c t. \tag{5.27}$$

Assume $\omega_m \ll \omega_c$. We want to evaluate μ, μ_+ and μ_- using the definitions provided above, and then sketch the amplitude function $A(t)$ and the envelope function $E(t)$ for: (i) $A_m/A_c < 1$, (ii) $A_m/A_c = 1$ and (iii) $A_m/A_c > 1$.

Using the single-tone $m(t)$ we first evaluate the signal parameters:

$$m_p = \max_t |m(t)| = A_m, \qquad A_{\text{max}} = A_c + A_m, \qquad A_{\text{min}} = A_c - A_m,$$

such that from Equation (5.23) we calculate:

$$\mu = \frac{m_p}{A_c} = \frac{A_m}{A_c}, \tag{5.28}$$

Equivalently, from Equations (5.26a–5.26c) we have:

$$\mu = \frac{A_{\text{max}} - A_{\text{min}}}{2A_c} = \frac{[A_c + A_m] - [A_c - A_m]}{2A_c} = \frac{A_m}{A_c}, \tag{5.29a}$$

$$\mu_+ = \frac{A_{\text{max}} - A_c}{A_c} = \frac{[A_c + A_m] - A_c}{A_c} = \frac{A_m}{A_c}, \tag{5.29b}$$

$$\mu_- = \frac{A_c - A_{\text{min}}}{A_c} = \frac{A_c - [A_c - A_m]}{A_c} = \frac{A_m}{A_c}. \tag{5.29c}$$

Observe that in this example $m(t)$ is symmetric with respect to the horizontal axis so that the modulation indices evaluated using Equation (5.23) and Equation (5.26a) are identical, and $\mu = \mu_+ = \mu_-$. Let us now sketch the envelope function for the three cases considered:

- **Case (i):** $A_m/A_c < 1$. In this case $\mu = A_m/A_c < 1$ and $E(t) = A(t) = [A_c + A_m \cos\omega_m t]$ resulting in $E(t) \sim m(t)$. This is precisely how we want the envelope to be.

- **Case (ii):** $A_m/A_c = 1$. In this "borderline" case $\mu = A_m/A_c = 1$ and $E(t) = A(t) = [A_c + A_m \cos\omega_m t]$ still follows $m(t)$.

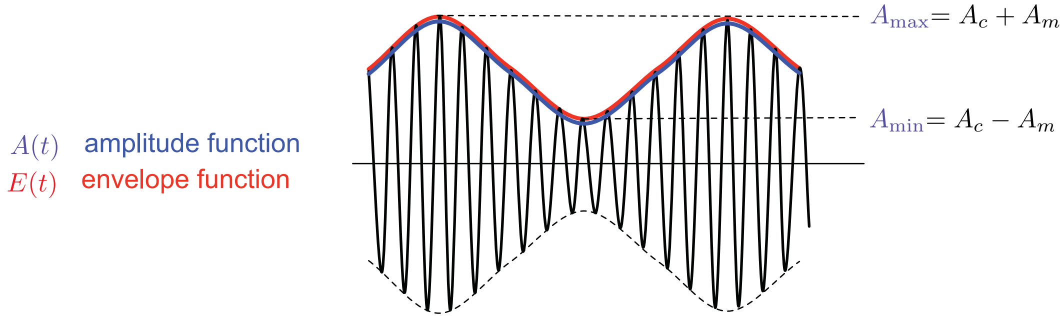


Figure 5.13: Modulated waveform for $\mu = A_m/A_c < 1$.

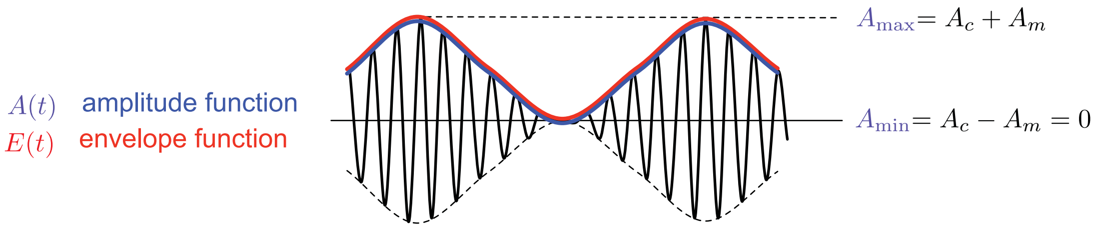


Figure 5.14: Modulated waveform for $\mu = A_m/A_c = 1$.

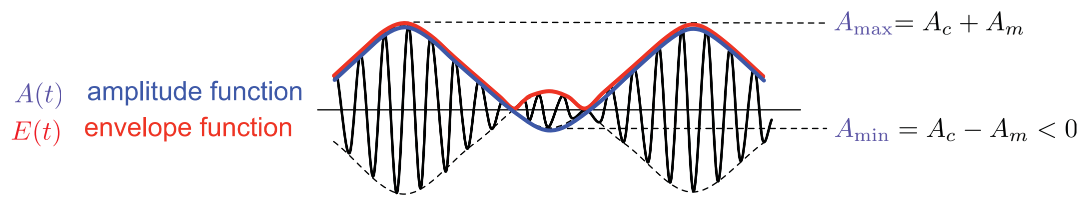


Figure 5.15: Modulated waveform for $\mu = A_m/A_c > 1$.

- **Case (iii):** $A_m/A_c > 1$. In this "over-modulation" case $\mu = A_m/A_c > 1$ and $E(t) = |A_c + A_m \cos\omega_m t| \neq A(t)$ such that $E(t) \not\propto m(t)$ resulting in **envelope distortion**.

Let us also determine the spectrum of the AM signal in this single-tone modulation example. Using the formulation of $\varphi_{\mathrm{AM}}(t)$ given in Equation (5.27) we can express $\Phi_{\mathrm{AM}}(f)$ using the frequency shifting/modulation property:

$$\Phi_{\mathrm{AM}}(f) = \frac{1}{2}\left[X(f - f_c) + X(f + f_c)\right], \tag{5.30}$$

where $x(t) = A_c + A_m \cos\omega_m t$, with the corresponding Fourier transform $X(f)$:

$$X(f) = A_c\,\delta(f) + \frac{A_m}{2}\left[\delta(f - f_m) + \delta(f + f_m)\right]. \tag{5.31}$$

Substituting Equation (5.34) in Equation (5.30) we obtain:

$$\begin{aligned}\Phi_{\text{AM}}(f) = \frac{A_c}{2}\left[\,\delta(f-f_c)+\delta(f+f_c)\,\right] \\ +\frac{A_m}{4}\left[\,\delta\big(f-(f_c+f_m)\big)+\delta\big(f-(f_c-f_m)\big)\,\right] \\ +\frac{A_m}{4}\left[\,\delta\big(f+(f_c-f_m)\big)+\delta\big(f+(f_c+f_m)\big)\,\right].\end{aligned} \tag{5.32}$$

Figure 5.16 shows the spectrum of the AM signal. Observe that $\Phi_{\text{AM}}(f)$ includes all three components that are part of an AM signal: the lower side-band (frequency components at $\pm(f_c - f_m)$), the upper side-band (frequency components at $\pm(f_c + f_m)$) and the carrier (frequency components at $\pm f_c$). As an alternate method, we can determine the spectrum of the AM signal by first

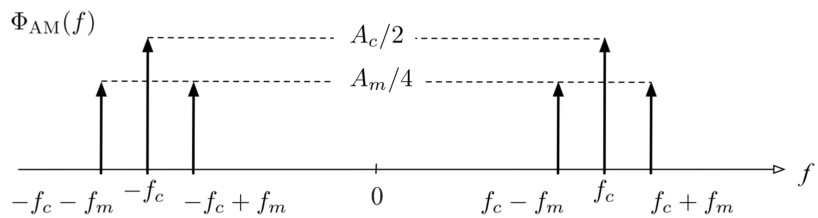


Figure 5.16: $\Phi_{\text{AM}}(f)$ resulting from single-tone amplitude modulation.

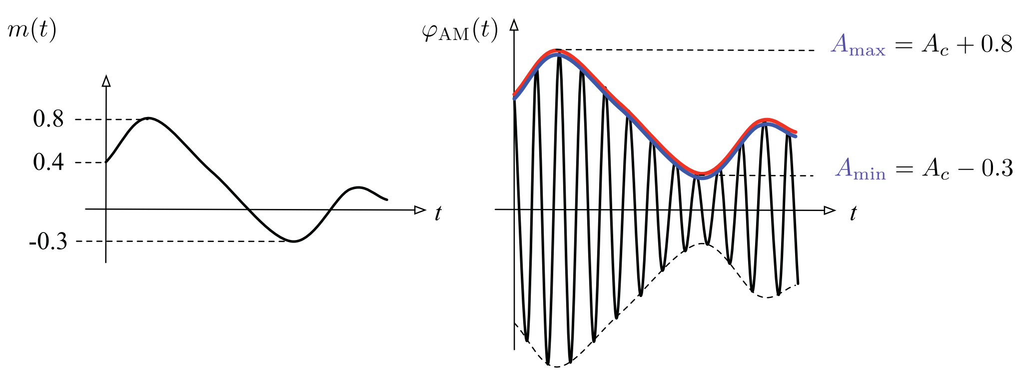
re-formulating $\varphi_{\text{AM}}(t)$ using trigonometric identities:

$$\varphi_{\text{AM}}(t) = A_c \cos\omega_c t + A_m \cos\omega_m t \cos\omega_c t, \tag{5.33a}$$

$$= A_c \cos\omega_c t + \frac{A_m}{2}\cos(\omega_c+\omega_m)t + \frac{A_m}{2}\cos(\omega_c-\omega_m)t \tag{5.33b}$$

It is then a straight-forward process to transform the time-domain expression in Equation (5.33b) using the Fourier transform tables.

Example 5.2. Consider the modulating signal $m(t)$ and the corresponding AM signal $\varphi_{\text{AM}}(t) = [A_c + m(t)]\cos\omega_c t$:



We want to determine the modulation indices using the various definitions provided. Using Equation (5.25) we first determine

$$A_{\text{max}} = A_c + 0.8, \qquad \text{and} \qquad A_{\text{min}} = A_c - 0.3,$$

which in turn allow us to evaluate the modulation indices as defined Equations (5.26a–5.26c):

$$\begin{aligned} \mu &= \frac{A_{\text{max}} - A_{\text{min}}}{2A_c} = \frac{0.55}{A_c}, \\ \mu_+ &= \frac{A_{\text{max}} - A_c}{A_c} = \frac{0.8}{A_c}, \\ \mu_- &= \frac{A_c - A_{\text{min}}}{A_c} = \frac{0.3}{A_c}. \end{aligned}$$

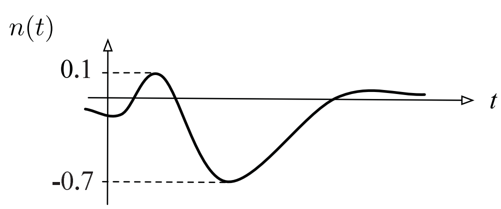
On the other hand if we were to use the modulation index as defined in Equation (5.23), we obtain:

$$\mu = \frac{m_p}{A_c} = \frac{0.8}{A_c},$$

where $m_p = \max_t |m(t)| = 0.8$.

Observe that $m(t)$ is not symmetric in the sense that its maximum equals 0.8 whereas its minimum is -0.3. Therefore, the values of the modulation index μ calculated using Equation (5.23) and Equation (5.26a) differ. The modulation index from Equation (5.23) uses $\mu = \max\{\mu_+, \mu_-\}$ whereas the modulation index from Equation (5.26a) uses $\mu = \text{average}\{\mu_+, \mu_-\}$.

To demonstrate how the shape of the modulating waveform affects the values of the modulation indices, let us now consider $n(t)$ as the new modulating signal used to generate the AM signal $\varphi_{\text{AM}}(t) = [A_c + n(t)] \cos \omega_c t$. From $n(t)$ we first determine $A_{\text{max}} = A_c + 0.1$, $A_{\text{min}} = A_c - 0.7$



and $n_p = \max_t |n(t)| = 0.7$. From Equations (5.26a–5.26c) we have:

$$\begin{aligned} \mu &= \frac{A_{\text{max}} - A_{\text{min}}}{2A_c} = \frac{0.4}{A_c}, \\ \mu_+ &= \frac{A_{\text{max}} - A_c}{A_c} = \frac{0.1}{A_c}, \\ \mu_- &= \frac{A_c - A_{\text{min}}}{A_c} = \frac{0.7}{A_c}, \end{aligned}$$

and from Equation (5.23):

$$\mu = \frac{n_p}{A_c} = \frac{0.7}{A_c}.$$

As in the previous case $n(t)$ is not symmetric with its maximum being equal to 0.1 and its minimum at -0.7. As before, values of the modulation index μ calculated using different equations differ. In particular, the modulation index from Equation (5.23) yields the result $\mu = 0.7/A_c = \max\{\mu_+, \mu_-\}$ whereas Equation (5.26a) results in uses $\mu = 0.4/A_c = \text{average}\{\mu_+, \mu_-\}$.

5.4.1 Sideband and Carrier Power

We developed the DSB-LC/AM technique by adding a separate carrier term to be transmitted within the same frequency band as the corresponding DSB-SC signal. Our main objective has been the development of a new amplitude modulation scheme that would allow demodulation using simple, inexpensive signal processing operations. Indeed, the presence of the carrier term within the AM signal eliminates the need for coherent detection; instead we will use an envelope detector which we will introduce and analyze in Section 5.4.3. In order to have a better appreciation of the effects of adding the carrier term we will now turn our attention to the calculation of the power of the AM signal and in particular to its constituent parts.

Let us consider the AM signal $\varphi_{\text{AM}}(t)$ generated by the baseband message/modulating signal $m(t)$ bandlimited to B_m-Hz. We express $\varphi_{\text{AM}}(t)$ by separating the carrier and sideband terms:

$$\varphi_{\text{AM}}(t) = \underbrace{A_c \cos \omega_c t}_{\text{carrier}} + \underbrace{m(t) \cos \omega_c t}_{\text{sidebands}}. \tag{5.34}$$

Let $\overline{g(t)}$ represent the time average of the waveform $g(t)$ and let P_x be the power delivered by the signal $x(t)$ across the load R:

$$P_x = \frac{\overline{x^2(t)}}{R}. \tag{5.35}$$

We now want to compute the power of the AM signal P_φ and express it in terms of the carrier power P_c and the sideband power P_s. From the definition of the AM signal given in Equation (5.34) and under the normalization assumption $R = 1\text{-}\Omega$ we write:

$$P_\varphi = \overline{A_c^2 \cos^2 \omega_c t + 2A_c m(t) \cos^2 \omega_c t + m^2(t) \cos^2 \omega_c t} \tag{5.36}$$

We simply the above expression for P_φ we first recognize that $m(t)$ changes slowly with respect to the carrier term $\cos \omega_c t$. This observation follows from the assumption $f_c \gg B_m$. Furthermore, we assume that $\overline{m(t)} = 0$ since the lack of d.c. content in $m(t)$ is a pre-condition for generating AM signals. It then follows:

- $\overline{A_c^2 \cos^2 \omega_c t} = A_c^2/2$,
- $2A_c\, \overline{m(t) \cos^2 \omega_c t} = 2A_c\, \overline{m(t)}\; \overline{\cos^2 \omega_c t} = 0$,
- $\overline{m^2(t) \cos^2 \omega_c t} = \overline{m^2(t)}\; \overline{\cos^2 \omega_c t} = \overline{m^2(t)}/2$,

where we used the assumption $f_c \gg B_m$ to separate the time averaging operations involving $m(t)$ and the carrier term. We further define:

$$P_c = \frac{A_c^2}{2} \qquad \text{and} \qquad P_s = \frac{\overline{m^2(t)}}{2}, \tag{5.37}$$

such that the total power in the AM signal can be expressed as:

$$P_\varphi = P_c + P_s. \tag{5.38}$$

In an AM signal, the carrier term does not carry any information. Hence, any power used to transmit the carrier is *wasted*. It is the sidebands that are a function of the message signal $m(t)$ and therefore our objective will be the maximization of the sideband power in order to enhance the efficiency of transmitting the message signal $m(t)$ in amplitude modulated format. We define η as the power efficiency:

$$\eta = \frac{\text{useful power}}{\text{total power}} = \frac{P_s}{P_\varphi} = \frac{P_s}{P_c + P_s}. \tag{5.39}$$

It is instructive to further investigate the power efficiency of an AM signal in the case of single-tone modulation. As before, let $m(t) = A_m \cos\omega_m t$ with $f_c \gg f_m$ such that the modulation index $\mu = A_m/A_c$ (see Example 5.1). Rewriting the message signal amplitude as $A_m = \mu A_c$ we calculate the sideband power:

$$P_s = \frac{\overline{m^2(t)}}{2} = \frac{\mu^2 A_c^2}{4}. \tag{5.40}$$

The resulting power efficiency equals:

$$\eta = \frac{P_s}{P_c + P_s} = \frac{\mu^2 A_c^2/4}{A_c^2/2 + \mu^2 A_c^2/4} = \frac{\mu^2}{2 + \mu^2}. \tag{5.41}$$

A closer inspection of the power efficiency as a function of modulation index reveals a monotonically increasing function of μ. This result is expected as the sideband power, i.e., the useful power within an AM signal, as expressed in Equation (5.40), increases with increasing modulation index. However, for AM signals the modulation index is constrained by the condition $0 \leq \mu \leq 1$ which ensures that the envelope of the AM signal remains undistorted. Therefore, for single-tone modulation the maximum power efficiency is achieved at $\mu = 1$:

$$\eta_{\max} = \left.\frac{\mu^2}{2 + \mu^2}\right|_{\mu=1} = \frac{1}{3}. \tag{5.42}$$

We conclude that in an AM signal no more than 33% of the total signal power is used to transmit the message itself; the remaining power is used to transmit the carrier. This result stands in stark contrast with the power efficiency that can be achieved with a DSB-SC amplitude modulated signal. Furthermore, the 33% power efficiency is achieved only with single-tone modulation, i.e., a test signal; real-life modulating signals such as speech or music, are dynamic in nature and the signal amplitude changes considerably over the duration of the program, resulting in reduced μ and therefore in lower power efficiency. The low power efficiency of AM signals is its main drawback; it is a price we pay in order to simplify the demodulation process.

5.4.2 AM Broadcasting Standards

The following list provides some of the key AM radio broadcasting standards. These standards are formulated, published and enforced by regulatory agencies in each country/region; they include:

CRTC, the Canadian Radio-Television and Telecommunications Commission in Canada, **FCC**, the Federal Communications Commission in the USA and **EBU**, the European Broadcasting Union in the European Union and affiliated countries.

- **Assigned carrier frequency:** 540–1600 kHz in 10 kHz increments.
- **Channel bandwidth:** 10 kHz.
- **Carrier frequency stability:** ± 20 Hz.
- **Percentage Modulation:** maintain 85–95% modulation; maximum modulation values allowed: $\mu = 100\%$, $\mu_+ = 125\%$ and $\mu_- = 100\%$.
- **Audio frequency response:** 100 Hz–5 kHz, ± 2 dB with 0 dB at 1 kHz.
- **Harmonic distortion:** for $\mu < 85\%$ harmonic distortion should remain below 5%, for $\mu \in [85\%, 95\%]$ harmonic distortion should remain below 7.5%.
- **Noise and hum:** 45% below 100% in the 20–30 Hz range.
- **Maximum licensed power:** 50 kW.

Broadcasters use limiting and compression on the program material to reduce the dynamic range of the signal (this is true both for AM and FM broadcasting) so that the transmitter can operate at close to the maximum allowed modulation index value most of the time and thus achieve higher power efficiency.

5.4.3 Generation of AM Signals

We can apply the DSB-SC generation techniques (see Section 5.3) to $[A_c + m(t)]$ instead of $m(t)$. Howeevr, since we do not need to suppress the carrier as in the case of DSB-SC signals, we no longer require balanced modulators (balanced with respect to the carrier). This simplifies the modulator structure considerably. Please refer to the course reference text for further information.

5.4.4 Demodulation of AM Signals

While it is possible to demodulate an AM signal using a coherent detector with a locally generated carrier as with DSB-SC signals, such an approach would defy the purpose of transmitting a large carrier as part of the AM signal. Therefore, we will not pursue this approach any further.

Rectifier Detector: Please see the course reference text for a description of the rectifier detector and for a detailed analysis of how it can be used to demodulate AM signals.

Envelope Detector: The envelope detector is a simple but effective device well-suited to the demodulation of narrowband AM signals with $\mu \leq 1$. Ideally, the output of the envelope detector follows $E(t)$, the envelope of $\varphi_{\mathrm{AM}}(t)$. To understand how the envelope detector functions, let us consider the following cases:

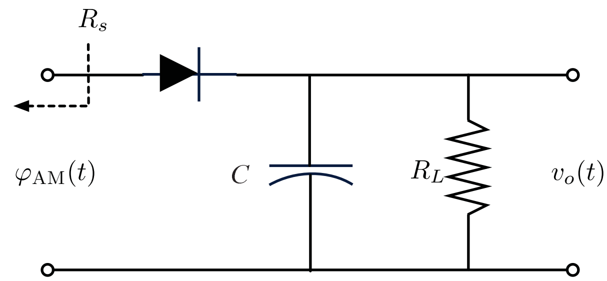


Figure 5.17: The envelope detector (R_L : load resistance, R_s : source resistance).

Charging capacitor: During the positive cycles, i.e., when $\varphi_{\text{AM}}(t) > 0$ and when the input voltage exceeds the voltage across the capacitor C, the diode is forward biased and conducts, which results in the capacitor C charging quickly to the maximum value of $\varphi_{\text{AM}}(t)$.

Discharging capacitor: During the negative cycles, i.e., when $\varphi_{\text{AM}}(t) < 0$ and when the the capacitor voltage exceeds the input voltage, the the diode is reverse biased and does not conduct; the capacitor C discharges over R_L until the charge cycle.

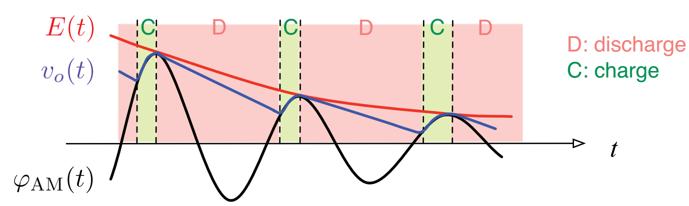


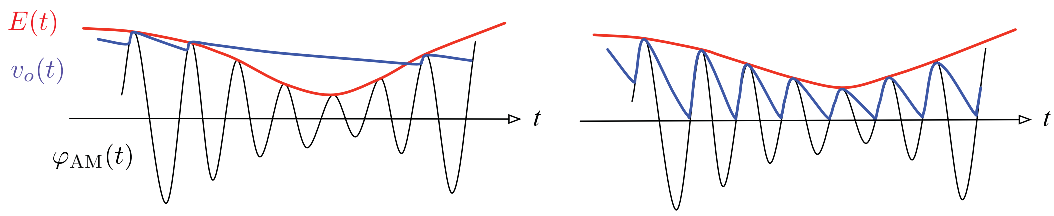
Figure 5.18: Charging and discharging of the capacitor in the envelope detector.

Under the assumption that the diode is an ideal device, we calculate the charging time constant of the envelope detector as $\tau_C = (R_s + R_L)\,C$ and the discharging time constant as $\tau_D = R_L C$. For proper operation of the envelope detector the charging time constant τ_C must be small compared to the carrier period $1/f_c$, i.e., $\tau_C \ll 1/f_C$ so that during the charging cycles the capacitor can quickly charge to the peak voltage of the input signal.

On the other hand the discharging constant τ_D must be long enough compared to the carrier period $1/f_c$ but not so long such that the capacitor will not discharge at the maximum rate of change of the modulating signal $m(t)$. A good compromise can be expressed as:

$$\frac{1}{f_c} \ll \tau_D \ll \frac{1}{B_m}, \tag{5.43}$$

where B_m is the bandwidth of the modulating signal $m(t)$. Figure 5.19 demonstrates how the output of the envelope detector may look like when τ_D is too large or too small.



(a) τ_D is too large **(b)** τ_D is too small

Figure 5.19: Envelope detector output: **(a)** τ_D too large, **(a)** τ_D too small.

In Figure 5.19a the envelope detector cannot follow the envelope $E(t)$ as τ_D is too large such that the capacitor in the envelope detector cannot discharge fast enough relative to the rate of change of the modulating signal $m(t)$. Figure 5.19b shows that if τ_D is too small then during the discharge cycles the detector output will quickly decay such that the output of the envelope detector will exhibit significant ripple which may prove to be difficult to remove even with additional processing.

Figure 5.18 shows that the output of the envelope detector $v_o(t)$ has ripples and also exhibits a d.c. component which need to be eliminated for the full recovery of the modulating/message signal $m(t)$. Therefore, we process $v_o(t)$ first with a lowpass filter to eliminate ripples followed by a d.c. blocking unit. Figure 5.20 presents the complete block diagram of the AM demodulator based on envelope detection.

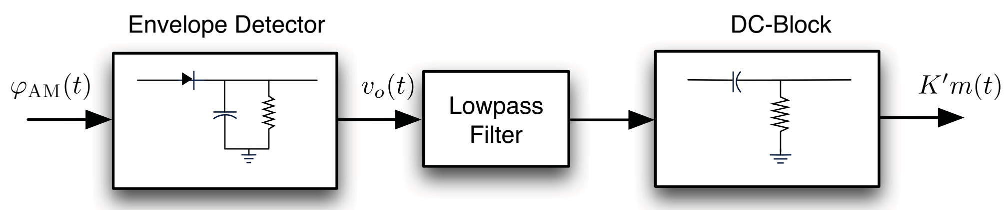


Figure 5.20: AM signal demodulation using envelope detection.

5.5 Quadrature Amplitude Modulation (QAM)

The transmission bandwidth of a DSB amplitude modulated signal, i.e., DSB-SC and AM, is twice the baseband bandwidth of the modulating/message signal. **Quadrature Amplitude Modulation** (QAM) is a technique that allows the transmission of two independent signals without interfering with each other while occupying the same frequency band, eliminating one of the major disadvantages of DSB amplitude modulation. We also use the term *Quadrature Carrier Multiplexing* to describe QAM: *multiplexing* as both message signals that are part of a QAM signal will be transmitted over the same communication channel, and *quadrature carrier* as a QAM transmitter uses two carriers that are in quadrature with respect to each other, that is two carriers with $\pi/2$ phase difference, e.g. $\cos\omega_c t$ and $\sin\omega_c t$.

We recall from our earlier discussion of coherent detection of DSB-SC signals that if the phase difference between the local oscillators in the modulator and demodulators equals $\pi/2$, then the detector output becomes 0 (see the discussion at the end of Section 5.2.1 on page 69). It is precisely this idea that we exploit in designing the QAM system. Let $m_1(t)$ and $m_2(t)$ be two independent baseband message signals with B_m-Hz bandwidth. The QAM signal at carrier frequency f_c with $f_c \gg B_m$ is generated by first modulating $\cos\omega_c t$ with $m_1(t)$, $\sin\omega_c t$ with $m_2(t)$ and then combining the two DSB-SC signals:

$$\varphi_{\text{QAM}}(t) = m_1(t)\cos\omega_c t + m_2(t)\sin\omega_c t. \tag{5.44}$$

As $m_1(t)$ and $m_2(t)$ are bandlimited to B_m-Hz and both carriers operate at f_c, $\varphi_{\text{QAM}}(t)$ occupies the frequency band $|f \pm f_c| \leq B_m$ resulting in a transmission bandwidth of $2B_m$-Hz. Figure 5.21 depicts the generation and demodulation of a QAM signal. The QAM-receiver consists of

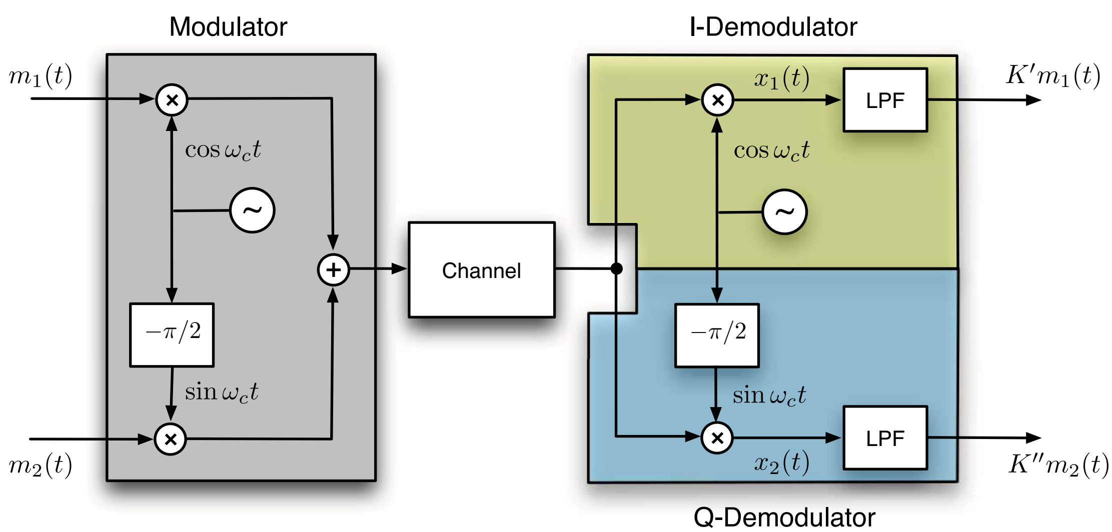


Figure 5.21: QAM modulator and demodulator.

two coherent detectors, the I-Demodulator (I: in-phase) and the Q-Demodulator (Q: quadrature), operating in parallel. We start our analysis of the demodulator by first considering the signal $x_1(t)$ at the output of the multiplier in the I-Demodulator:

$$x_1(t) = \varphi_{\text{QAM}}(t)\cos\omega_c t, \tag{5.45a}$$
$$= m_1(t)\cos^2\omega_c t + m_2(t)\cos\omega_c t\sin\omega_c t, \tag{5.45b}$$
$$= \frac{1}{2}m_1(t) + \frac{1}{2}m_1(t)\cos 2\omega_c t + \frac{1}{2}m_2(t)\sin 2\omega_c t. \tag{5.45c}$$

Observe that the last two terms in Equation (5.45c) are narrowband signals centered at $2f_c$ and therefore will be eliminated by the lowpass filter with cutoff frequency set at B_m-Hz. Therefore, the output of the lowpass filter in the I-Demodulator can be expressed as:

$$h_{lpf}(t) * x_1(t) = \frac{1}{2} m_1(t), \tag{5.46}$$

where $h_{lpf}(t)$ is the impulse response of the lowpass filter. Equivalently, we can represent the output of the lowpass filter as $K' m_1(t)$ to allow any non-unity gain the signal may experience during the demodulation process. The analysis of the Q-Demodulator follows the same pattern; the signal $x_2(t)$ at the output of the multiplier equals:

$$\begin{aligned} x_2(t) &= \varphi_{\text{QAM}}(t) \sin \omega_c t, && (5.47a) \\ &= m_1(t) \cos \omega_c t \sin \omega_c t + m_2(t) \sin^2 \omega_c t, && (5.47b) \\ &= \frac{1}{2} m_1(t) \sin 2\omega_c t + \frac{1}{2} m_2(t) - \frac{1}{2} m_2(t) \cos 2\omega_c t. && (5.47c) \end{aligned}$$

Similarly, the output of the lowpass filter in the Q-Demodulator (which has identical characteristics to the lowpass filter used in the I-Demodulator) becomes:

$$h_{lpf}(t) * x_2(t) = \frac{1}{2} m_2(t), \tag{5.48}$$

or equivalently $K'' m_2(t)$. Demodulation of QAM signals requires phase and frequency synchronization that can be achieved by using synchronous/coherent detectors with carrier acquisition.

There are numerous applications of QAM in signal processing and communications. For example, the colour information (the chrominance signal) in broadcast TV is transmitted in quadrature to the luminance signal. This decision was dictated by the frequency and channel allocations for TV broadcasting which were established when all TV signals were *monochrome*. Therefore, the engineers had to use a modulation scheme that allowed the transmission of the chrominance signal within the allocated frequency band. Today, the digital version of QAM, where the message signals $m_1(t)$ and $m_2(t)$ are digital, forms the foundation of many digital data communication systems.

5.6 Single Sideband Modulation (SSB)

Standard AM and DSB-SC amplitude modulation techniques are wasteful of bandwidth because they both require transmission bandwidths of $2B_m$-Hz where B_m is the bandwidth of the baseband modulating signal $m(t)$. Doubling of the transmission bandwidth is the result of transmitting both sidebands, upper and lower sidebands, as part of the modulated waveform.

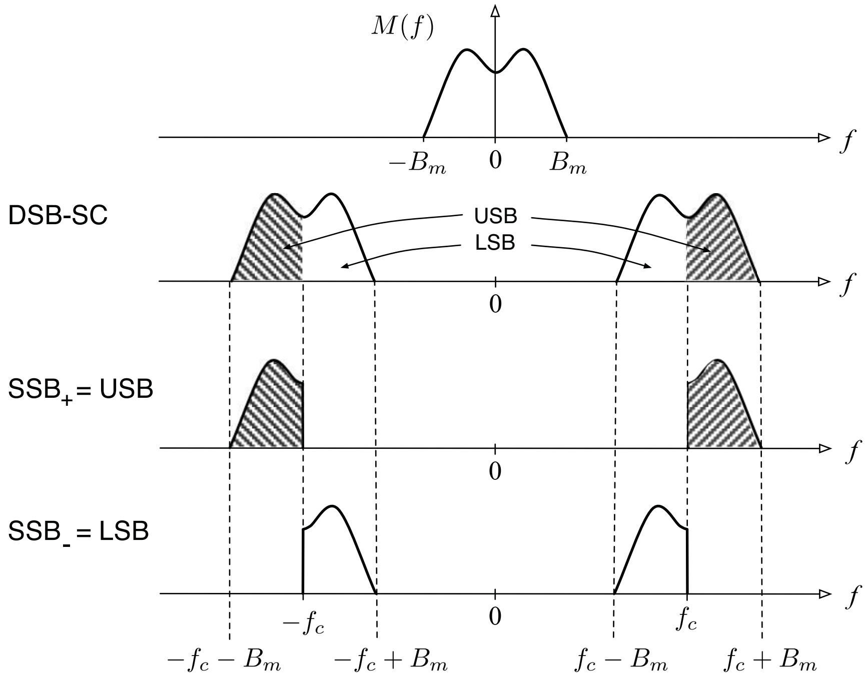


Figure 5.22: Spectra of DSB and SSB modulated waveforms.

The upper sideband (USB) and the lower sideband (LSB) signals in a modulated waveform are uniquely related to each other as they are symmetric with respect $f = f_c$. Therefore, we need to transmit *only one* sideband while retaining all the information in $m(t)$. We will use the notation SSB_+ and SSB_- to refer to single sideband modulated waveforms which retain only the USB and the LSB from the double sideband modulated signal, respectively.

5.6.1 Representation of Single Sideband Signals

Let $m(t)$ be a real-valued baseband message signal bandlimited to B_m-Hz. We will use the notation $m_+(t)$ and $m_-(t)$ to refer to the single sideband signals generated from $m(t)$. Let $M_+(f)$ and $M_-(f)$ be their respective spectra.

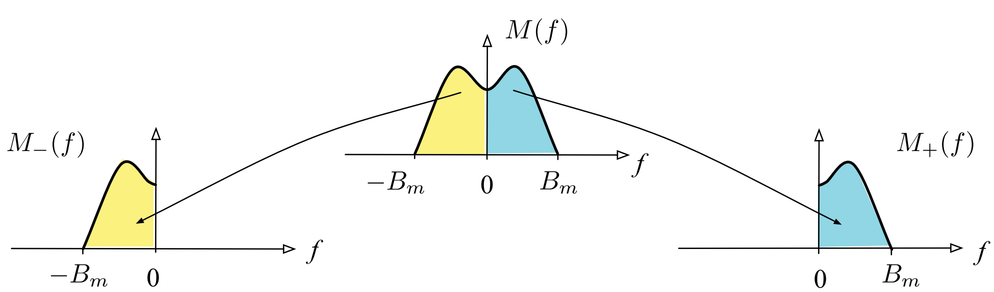


Figure 5.23: Single sideband spectra $M_+(f)$ and $M_-(f)$ generated from $M(f)$.

We observe that $m_+(t)$ and $m_-(t)$ cannot be real valued signals as their magnitude spectra no longer exhibit even symmetry, i.e., $|M_+(f)| \neq |M_+(-f)|$ and $|M_-(f)| \neq |M_-(-f)|$. In order to have a time-domain formulation of single sideband signals we first observe from Figure 5.23:

$$M(f) = M_+(f) + M_-(f), \tag{5.49}$$

or equivalently

$$m(t) = m_+(t) + m_-(t). \tag{5.50}$$

From the definitions of $M_+(f)$ and $M_-(f)$ and the conjugate symmetry property[1] of the Fourier transform we can write:

$$M_+(f) = M_-^*(-f). \tag{5.51}$$

The symmetry property of the single sideband Fourier transform corresponds to the following relation for the corresponding time signals:

$$m_+(t) = m_-^*(t). \tag{5.52}$$

In view of our earlier observation that $m_+(t)$ and $m_-(t)$ cannot be real valued signals and the symmetry properties stated in Equations (5.51) and (5.52), we assume:

$$m_+(t) = \frac{1}{2}\big[\, m(t) + jm_h(t) \,\big], \tag{5.53a}$$

$$m_-(t) = \frac{1}{2}\big[\, m(t) - jm_h(t) \,\big], \tag{5.53b}$$

for some signal $m_h(t)$. Clearly, the time-domain waveforms defined in Equation (5.53) satisfies the symmetry conditions stated in Equations (5.51) and (5.52).

To formulate the relation between the spectra of the single sideband signals obtained from $M(f)$ we recognize that $M_+(f)$ can be expressed as:

$$M_+(f) = M(f)\, u(f), \tag{5.54a}$$

$$= M(f)\left[\frac{1}{2}\big(1 + \text{sgn}(f)\big)\right], \tag{5.54b}$$

$$= \frac{1}{2}M(f) + \frac{1}{2}M(f)\text{sgn}(f), \tag{5.54c}$$

where $u(x)$ is the unit step function and sgn(x) is the sign function:

$$\text{sgn}(x) = \begin{cases} 1, & \text{if } x \geq 0; \\ -1, & \text{if } x < 0. \end{cases} \tag{5.55}$$

From Equation (5.53a) we can write:

$$M_+(f) = \frac{1}{2}M(f) + \frac{j}{2}M_h(f). \tag{5.56}$$

[1] The conjugate symmetry property states that $\mathcal{F}[x^*(t)] = X^*(-f)$ where $X(f) = \mathcal{F}[x(t)]$.

By comparing the expression in Equation (5.59) with Equation (5.54c) we can obtain the frequency domain expression:

$$M_h(f) = -jM(f)\text{sgn}(f) = H_h(f)M(f), \tag{5.57}$$

where $H_h(f)$ is the **Hilbert transformer**:

$$H_h(f) = -j\text{sgn}(f) = \begin{cases} -j, & \text{if } f \geq 0; \\ +j, & \text{if } f < 0. \end{cases} \tag{5.58}$$

Observe that the Hilbert transformer functions as a wideband $\pi/2$-radians phase shifter. Thus,

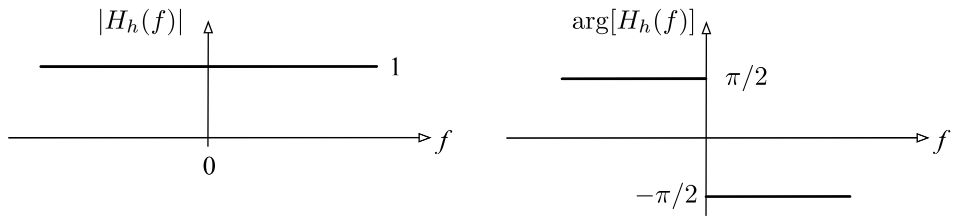


Figure 5.24: Magnitude and phase response functions of the Hilbert transformer.

$m_h(t)$ as the output of the Hilbert transformer when $m(t)$ is the input, represents a $\pi/2$ phase-shifted version of the modulating signal, i.e., every frequency component in $m(t)$ experiences the same $\pi/2$ phase shift as $m(t)$ is processed by the Hilbert transformer.

Using the results for expressing the single sideband signals both in time and frequency domains, we can now formulate the single sideband modulated signals with ease. Let $\varphi_{\text{SSB}_+}(t)$ be the upper single sideband modulated signal with the corresponding Fourier transform $\Phi_{\text{SSB}_+}(f)$; similarly we will use the notation $\varphi_{\text{SSB}_-}(t)$ to represent the lower single sideband modulated signal with $\Phi_{\text{SSB}_-}(f)$ its Fourier transform. Comparing the shapes of the single sideband spectra in Figure 5.23 with those of the modulated waveforms depicted in Figure 5.22 we write:

$$\Phi_{\text{SSB}_+}(f) = M_+(f - f_c) + M_-(f + f_c), \tag{5.59a}$$

$$\Phi_{\text{SSB}_-}(f) = M_-(f - f_c) + M_+(f + f_c). \tag{5.59b}$$

The corresponding time domain waveforms can be determined by taking the inverse Fourier transform of the expressions given in Equation (5.59). In particular, Equation (5.59b) together with Equation (5.53) allows us to express $\varphi_{\text{SSB}_+}(t)$ as:

$$\varphi_{\text{SSB}_+}(t) = m_+(t)e^{j\omega_c t} + m_-(t)e^{-j\omega_c t}, \tag{5.60a}$$

$$= \frac{1}{2}\big[\, m(t) + jm_h(t)\,\big]e^{j\omega_c t} + \frac{1}{2}\big[\, m(t) - jm_h(t)\,\big]e^{-j\omega_c t}, \tag{5.60b}$$

$$= m(t)\cos\omega_c t - m_h(t)\sin\omega_c t. \tag{5.60c}$$

Similarly, for the lower sideband modulated signal we obtain the expression:

$$\varphi_{\text{SSB}_-}(t) = m(t)\cos\omega_c t + m_h(t)\sin\omega_c t. \tag{5.60d}$$

5.6.2 Generation of Single Sideband Signals

Selective Filtering Method

The **selective filtering method** is the most commonly used technique for generating SSB signals.

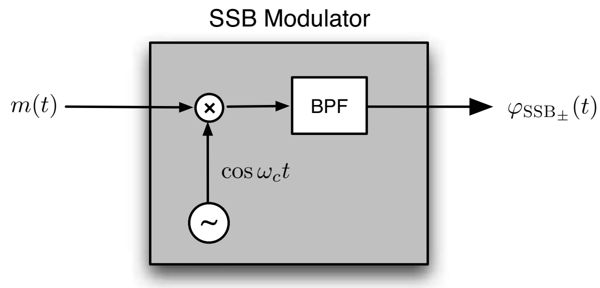


Figure 5.25: SSB signal generation using the selective filtering method.

This method first generates the DSB-SC amplitude modulated signal $m(t)\cos\omega_c t$, and then filters out one of its sidebands using a bandpass filter.

Successful implementation of the selective filtering method requires that $B_m \ll f_c$ and the modulating signal $m(t)$ has little or no low-frequency content, i.e., $|M(f)| = 0$ for $|f| < f_1$ where f_1 is the low-frequency edge for $m(t)$. Modulating signals used in many practical applications easily conform to these conditions, e.g. voice-grade speech signals occupy the frequency band $[300, 3400]$ Hz and therefore has a 300 Hz "frequency hole".

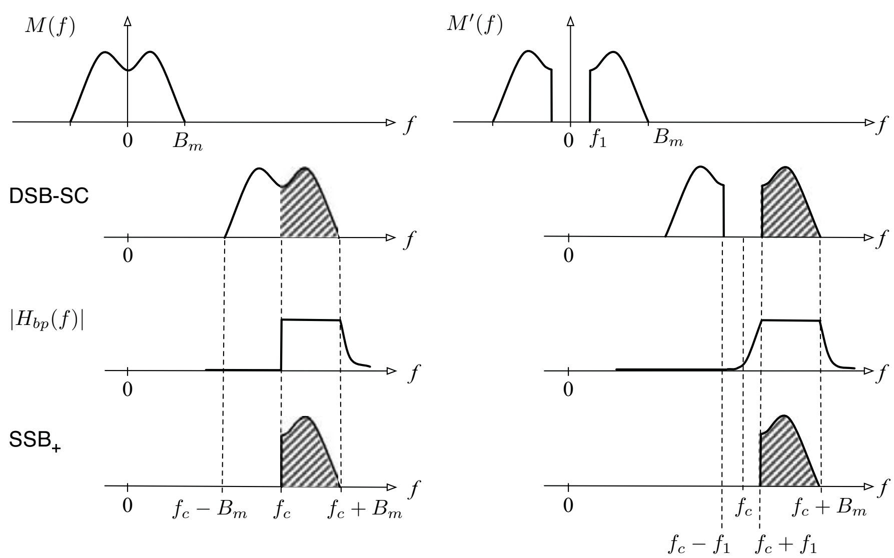


Figure 5.26: Generating a SSB_+ signal from a DSB-SC signal using the selective filtering method based on **(a)** $m(t)$ with low-frequency content and **(b)** $m'(t)$ with a low-frequency hole.

But why does a modulating signal have to have a "frequency hole" at low frequencies for the selective filtering method to work? Figure 5.26 demonstrates SSB_+ signal generation from a DSB-SC signal using the selective filtering method for two different modulating signals. **Case (a):** The modulating signal $m(t)$ with Fourier transform $M(f)$ has no "frequency hole" at low frequencies. The bandpass filter $H_{bp}(f)$ required to generate the SSB$_+$ signal from the DSB-SC signal $m(t)\cos\omega_c t$ is not realizable because the filter's first stopband $[0, f_c]$ Hz is adjacent to its passband $[f_c, f_c + B_m]$ with no provision for a transition band. **Case (b):** The second modulating signal $m'(t)$ with Fourier transform $M'(f)$ has the "frequency hole" $[0, f_1]$ Hz. As a result the bandpass filter $H_{bp}(f)$ that can generate the SSB$_+$ from the DSB-SC signal $m'(t)\cos\omega_c t$ can operate with the transition band $[f_c - f_1, f_c + f_1]$ and is therefore realizable.

In cases when the carrier frequency f_c of the modulated signal is high, we use a multi-stage approach for the generation of SSB amplitude modulated signals. In the first stage, we generate a SSB signal from $m(t)$ using a SSB modulator as shown in Figure 5.25; the resulting SSB signal occupies a frequency band below the desired carrier frequency f_c. The second stage uses the SSB signal generated by the first stage as an input and shifts the SSB signal further up in the frequency band. This process is repeated until the output of the final stage results in a SSB signal at the desired carrier frequency f_c. A major advantage of the multi-stage approach is the less stringent requirements on the bandpass filters used in the SSB modulators.

Phase-Shift Method

The time domain expressions we developed in Section 5.6.1 provide an alternate method for the generation of SSB signals. Under the assumption that we have access to a Hilbert transformer we can generate SSB signals using the **phase-shift method**: in this method we first generate the quadrature modulated waveforms $m(t)\cos\omega_c t$ and $m_h(t)\sin\omega_c t$ using the message signal $m(t)$ and its Hilbert transformed version $m_h(t)$, we then compute the difference of the quadrature modulated waveforms for SSB$_+$ or their sum for SSB$_-$ as per Equation (5.60):

$$\varphi_{\text{SSB}_+}(t) = m(t)\cos\omega_c t - m_h(t)\sin\omega_c t,$$
$$\varphi_{\text{SSB}_-}(t) = m(t)\cos\omega_c t + m_h(t)\sin\omega_c t.$$

Figure 5.27 shown the block diagram representation of the phase-shift method.

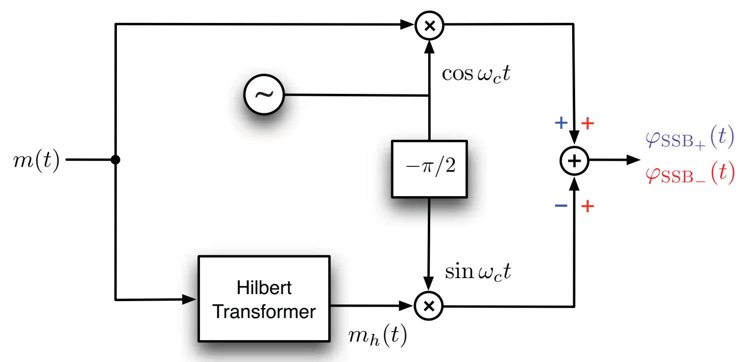


Figure 5.27: Generating SSB signals using the phase-shift method.

5.6.3 Demodulation of Single Sideband Signals

The demodulation of SSB signals can be achieved using a coherent detector with an identical structure as a coherent detector used to demodulate DSB-SC modulated signals. Observe that

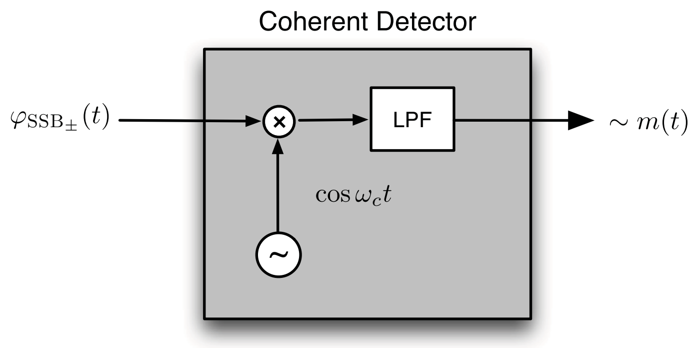


Figure 5.28: Demodulation of SSB signals using a coherent detector.

$$\varphi_{\mathrm{SSB}_\pm}(t)\cos\omega_c t = m(t)\cos^2\omega_c t \mp m_h(t)\sin\omega_c t\cos\omega_c t, \tag{5.61a}$$

$$= \frac{1}{2}m(t) + \frac{1}{2}m(t)\cos 2\omega_c t \mp \frac{1}{2}m_h(t)\sin 2\omega_c t. \tag{5.61b}$$

The last two terms in the right-hand side of Equation (5.61b) are centered about $\pm 2f_c$ and will be filtered out by the lowpass filter with passband $[0, B_m]$ Hz which has been designed to extract the baseband modulating modulating signal $m(t)$. Hence, any coherent demodulation technique developed for DSB-SC modulated signals can be used with SSB modulated signals as well.

5.7 Vestigial Sideband Modulation (VSB)

Single sideband modulation is a technique suitable for the transmission of signals with a "frequency hole" in the low frequency content of the modulating signal. On the other hand if the modulating signal has significant low-frequency energy then single sideband modulation is no longer a viable technique. For example, video signals have significant low frequency content and therefore do not allow the use of the bandwidth preserving/minimizing SSB modulation techniques. The **vestigial sideband** (VSB) modulation technique represents a compromise between the single sideband and double sideband modulation systems: a VSB modulated signal includes one *almost* complete sideband of the modulating waveform together with just a trace or *vestige* of the other sideband.

5.7.1 Generation of VSB Signals

The selective filtering method discussed in the context of generating SSB modulated signals also forms the basis for generating VSB modulated waveforms.

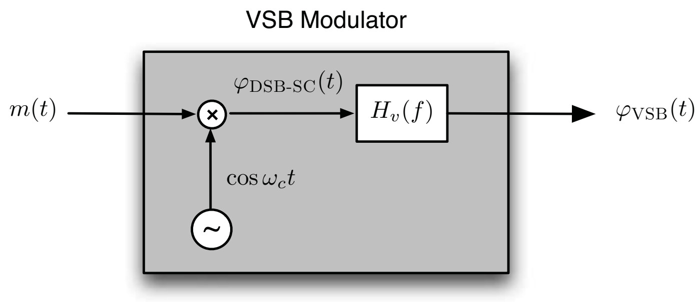


Figure 5.29: VSB signal generation using filtering.

In the selective filtering method, we first generate a DSB-SC modulated waveform $m(t)\cos\omega_c t$ followed by bandpass filtering where the filter is designed to keep only one sideband of the modulated waveform. VSB modulated signals are generated using the same approach: sideband shaping of a DSB-SC modulated waveforms using an appropriately designed bandpass filter. Figure 5.30 shows the spectra of the signals generated at the VSB-Modulator based on the filtering method.

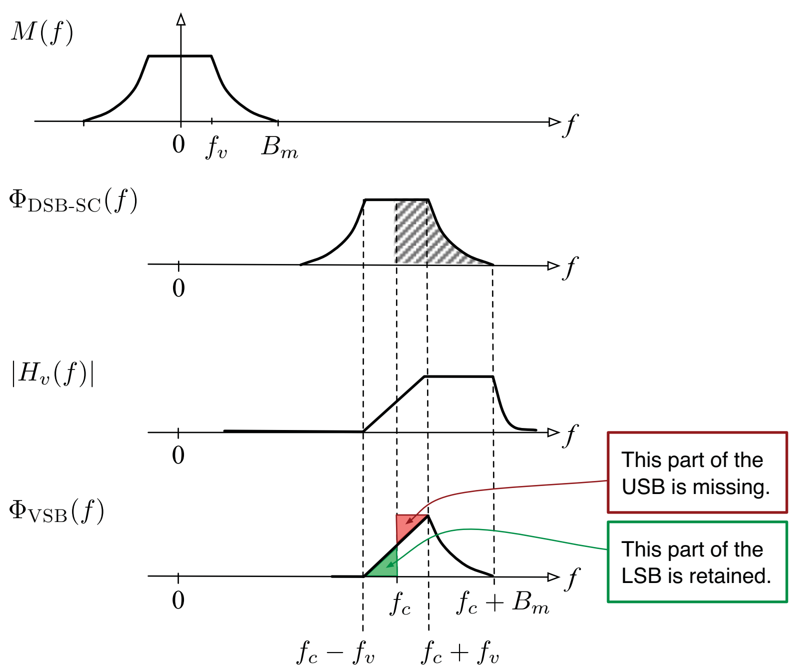


Figure 5.30: Spectra of the signals encountered during VSB signal generation based on selective filtering.

The VSB filter $H_v(f)$ has a bandpass characteristics with passband $[f_c + f_v,\ f_c + f_{B_m}]$ where f_v is the frequency parameter that determines the frequency extent of the vestige of the sideband[2] that will be part of the VSB modulated signal. In particular, $H_v(f)$ forces part of the LSB over the frequency band $[f_c - f_v,\ f_c]$ to be included in the VSB signal whereas part of the USB over the

[2] A VSB modulated waveform may include a vestige of the LSB or USB. The discussion provided in this section uses a VSB modulated waveform that includes part of the LSB and most of the USB of the DSB modulated waveform.

frequency band $[f_c,\ f_c + f_v]$ will be symmetrically shaped. Once we formulate the demodulation of VSB modulated signals, we will develop constraints that the bandpass filter $H_v(f)$ must satisfy such the demodulator can properly recover the modulating waveform $m(t)$ even though the VSB signal $\varphi_{\text{VSB}}(t)$ does not include a complete sideband.

5.7.2 Demodulation of VSB Signals

The demodulation of VSB modulated signals can be achieved using a coherent detector with a structure identical to those used to demodulate both the DSB-SC and the SSB modulated waveforms as depicted in Figures 5.5 and 5.28, respectively. The lowpass filter in the coherent detector will have a passband $[0, B_m]$ Hz where B_m is the bandwidth of the modulating signal $m(t)$.

Observe that the first step of coherent detection is to modulate the incoming signal $\varphi_{\text{VSB}}(t)$ with the output of the local oscillator $2\cos\omega_c t$. This operation will generate the signal $s(t) = 2\varphi_{\text{VSB}}(t)\cos\omega_c t$ with the spectrum

$$S(f) = \Phi_{\text{VSB}}(f - f_c) + \Phi_{\text{VSB}}(f + f_c). \tag{5.62}$$

Figure 5.31 shows the the spectra of the signals generated at the coherent demodulator (this set of spectra assumes that the baseband modulating waveform $m(t)$ and the resulting VSB modulated signal $\varphi_{\text{VSB}}(t)$ are characterized by their spectra shown in Figure 5.30). Clearly applying the signal

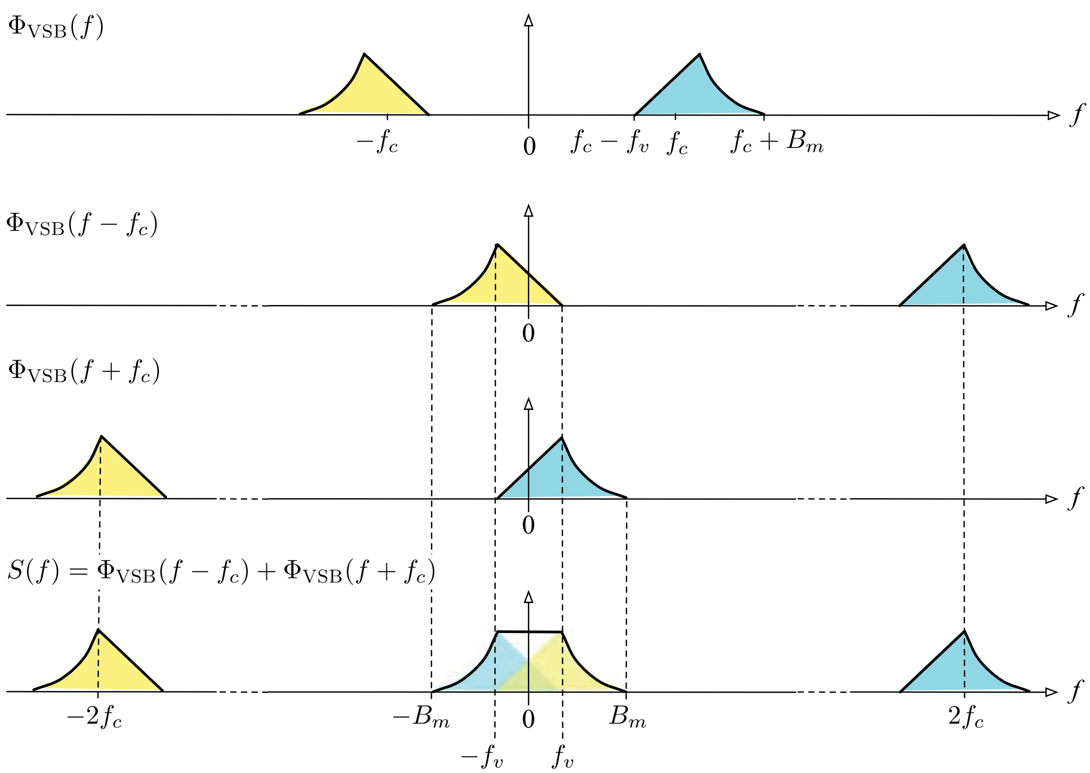


Figure 5.31: Demodulation of VSB modulated waveforms using a coherent detector.

$s(t)$ to the lowpass filter with passband $[0, B_m]$ Hz that is part of the coherent demodulator will result in the recovery of $m(t)$ from $s(t)$.

5.7.3 Choosing $H_v(f)$

From the discussion presented in Section 5.7.1 together with theVSB modulator depicted in Figure 5.29 we can express the spectrum of the VSB modulated signal $\varphi_{\text{VSB}}(t)$ as:

$$\Phi_{\text{VSB}}(f) = K\left[M(f - f_c) + M(f + f_c)\right] H_v(f), \tag{5.63}$$

where the constant K incorporates all scaling parameters encountered throughout the signal processing in the modulator. Let $y(t)$ be the output of the coherent demodulator and let $h_{lpf}(t)$ be the impulse response of the lowpass filter. The spectrum of the demodulated signal $y(t) = h_{lpf}(t) * s(t)$ is given by the expression:

$$\begin{aligned} Y(f) &= S(f) H_{lpf}(f), && (5.64a)\\ &= \left[\Phi_{\text{VSB}}(f - f_c) + \Phi_{\text{VSB}}(f + f_c)\right] H_{lpf}(f), && (5.64b)\\ &= K\Big[\big[M(f - 2f_c) + M(f)\big] H_v(f - f_c) \\ &\qquad + \big[M(f) + M(f + 2f_c)\big] H_v(f + f_c)\Big] H_{lpf}(f), && (5.64c)\\ &= K'\big[H_v(f - f_c) + H_v(f + f_c)\big] M(f). && (5.64d) \end{aligned}$$

Equation (5.64d) follows from the lowpass filter $H_{lpf}(f)$ filtering out the narrowband signal spectra, $M(f - 2f_c)H_v(f - f_c) + M(f + 2f_c)H_v(f + f_c)$ centered at $\pm 2f_c$. Thus, the bandpass filter $H_v(f)$ must satisfy the constraint:

$$H_v(f - f_c) + H_v(f + f_c) = \text{constant}, \qquad -B_m \leq f \leq B_m. \tag{5.65}$$

The constraint given in Equation (5.65) implies that $H_v(f)$ should be symmetric for $|f - f_c| \leq f_v$.

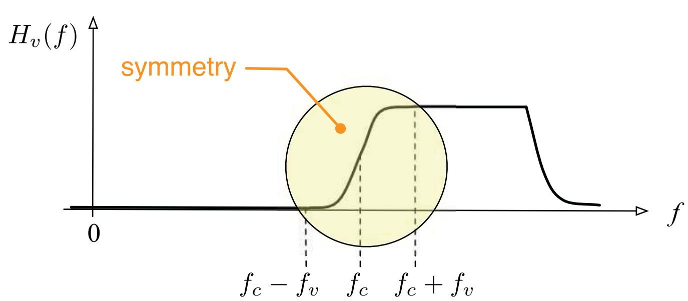


Figure 5.32: Symmetry condition for $H_v(f)$.

5.7.4 Further Comments on VSB Modulation

In this section we present a few observations and an example to complete our discussion of VSB modulation technique.

- Earlier we introduced VSB modulation as a compromise between DSB and SSB modulation. One can further show that

$$\lim_{f_v \to 0} \varphi_{\text{VSB}}(t) = \varphi_{\text{SSB}}(t) \qquad \text{and} \qquad \lim_{f_v \to B_m} \varphi_{\text{VSB}}(t) = \varphi_{\text{DSB}}(t),$$

  where f_v is the frequency parameter that determines the frequency extent of the vestige of the sideband that will be part of the VSB modulated signal.

- Let B_m be the transmission bandwidth of the modulating baseband signal $m(t)$ and let B_T be the transmission bandwidth of a signal. Thus, we have $B_T\big[\,\varphi_{\text{SSB}}(t)\,\big] = B_m$ and $B_T\big[\,\varphi_{\text{DSB}}(t)\,\big] = 2B_m$ such that

$$B_T\big[\,\varphi_{\text{SSB}}(t)\,\big] < B_T\big[\,\varphi_{\text{VSB}}(t)\,\big] < B_T\big[\,\varphi_{\text{DSB}}(t)\,\big].$$

  In practice, a VSB modulated waveform will resull in approximately 25% increase in transmission bandwidth over the baseband or SSB modulated transmission bandwidth of B_m:

$$B_T\big[\,\varphi_{\text{VSB}}(t)\,\big] \approx 1.25 B_T\big[\,\varphi_{\text{SSB}}(t)\,\big] = 1.25 B_m.$$

- The VSB modulation technique we discussed in this section does not include a separate carrier and therefore must be demodulated using a coherent detector. It is possible, however, to create a variant of VSB modulation by adding a carrier that will be transmitted in the same frequency band as the VSB signal; let VSB+C refer to this modulation scheme. Similarly, one can also generate a version of SSB modulation which includes a carrier; let SSB+C be this version of SSB modulation. The presence of the carrier terms both as part of VSB and the SSB modulated signals allows simplified demodulation using an envelope detector. However, both SSB+C and VSB+C are notoriously power inefficient, such that the power efficiency of the modulation schemes which include a carrier are ordered as follows:

$$\eta_{\text{SSB+C}} < \eta_{\text{VSB+C}} < \eta_{\text{AM}} \leq \frac{1}{3}.$$

Example 5.3. The best known application of VSB+C modulation is in commercial TV broadcasting. As analog video signals have large bandwidth and significant low-frequency content, VSB modulation is an obvious choice. Figure 5.33 shows the spectrum of a typical broadcast TV signal which is a composite signal consisting of a VSB+C video component frequency multiplexed with a FM-modulated audio signal. The composite TV signal also includes an audio carrier that will allow simple demodulation of the audio signal.

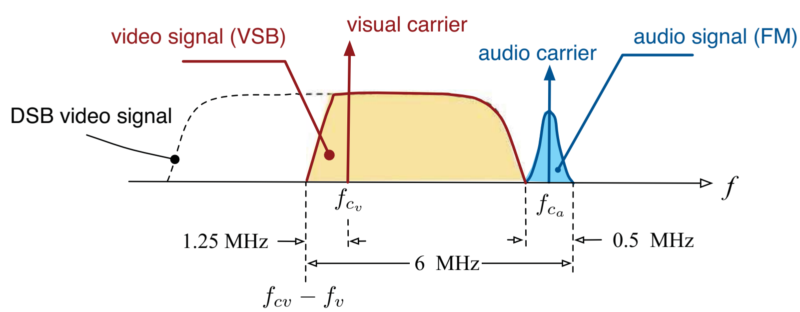


Figure 5.33: Spectrum of a typical composite TV signal (f_{c_v} : visual carrier, f_{c_a} : audio carrier such that $f_{c_a} = f_{c_v} + 4.5$ MHz).