# ECS332 2019/1 Part II.4 Dr.Prapun

## 5 Angle Modulation: FM and PM

**5.1.** We mentioned in 4.2 that a sinusoidal carrier signal

$$A\cos(2\pi f_c t + \phi)$$

has three basic parameters: amplitude, frequency, and phase. Varying these parameters in proportion to the baseband signal results in amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM), respectively.

**5.2.** As in 4.62, we will again assume that the baseband signal $m(t)$ is

(a) band-limited to $B$; that is, $|M(f)| = 0$ for $|f| > B$

and

(b) bounded between $-m_p$ and $m_p$; that is, $|m(t)| \le m_p$.

**Definition 5.3.** ***Phase modulation*** (***PM***):

$$x_{\text{PM}}(t) = A\cos\left(2\pi f_c t + \phi + k_p m(t)\right)$$

- max phase deviation:

**Definition 5.4.** The main characteristic[22] of ***frequency modulation*** (**FM**) is that the carrier frequency $f(t)$ would be varied with time so that

$$f(t) = f_c + k_f m(t), \tag{72}$$

where $k_f$ is an arbitrary constant.

- The subscript "$_f$" in $k_f$ is there to distinguish the constant from a similar constant in PM.
- $f(t)$ is varied from $f_c - k_f m_p$ to $f_c + k_f m_p$.
- $f_c$ is assumed to be large enough such that $f(t) \geq 0$.

**Example 5.5.** Figure 33 illustrates the outputs of PM and FM modulators when the message is a unit-step function.

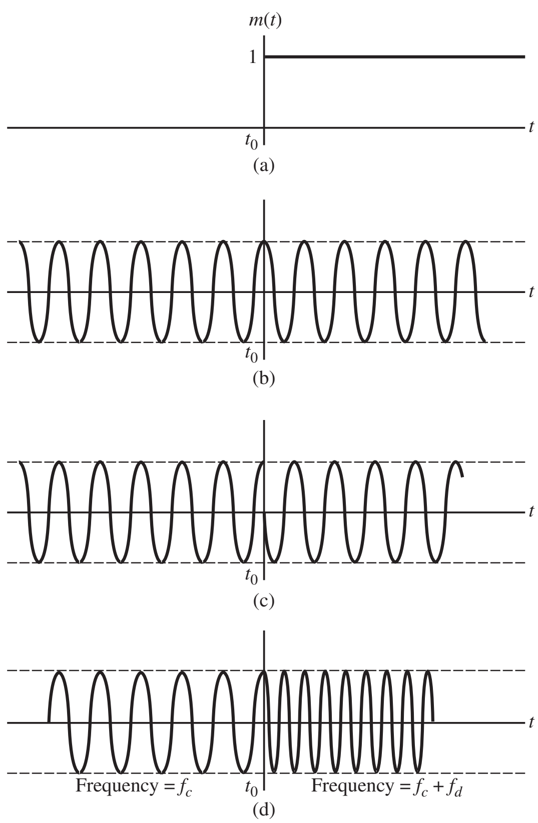


Figure 33: Comparison of PM and FM modulator outputs for a unit-step input. (a) Message signal. (b) Unmodulated carrier. (c) Phase modulator output (d) Frequency modulator output. [15, Fig 4.1 p 158]

[22] Treat this as a practical definition. The more rigorous definition will be provided in 5.15.

- For the PM modulator output,
  - the (instantaneous) frequency is $f_c$ for both $t < t_0$ and $t > t_0$
  - the phase of the unmodulated carrier is advanced by $k_p = \frac{\pi}{2}$ radians for $t > t_0$ giving rise to a signal that is discontinuous at $t = t_0$.
- For the FM modulator output,
  - the frequency is $f_x$ for $t < t_0$, and the frequency is $f_c + f_d$ for $t > t_0$
  - the phase is, however, continuous at $t = t_0$.

**Example 5.6.** With a sinusoidal message signal in Figure 34a, the frequency deviation of the FM modulator output in Figure 34d is proportional to $m(t)$. Thus, the (instantaneous) frequency of the FM modulator output is maximum when $m(t)$ is maximum and minimum when $m(t)$ is minimum.

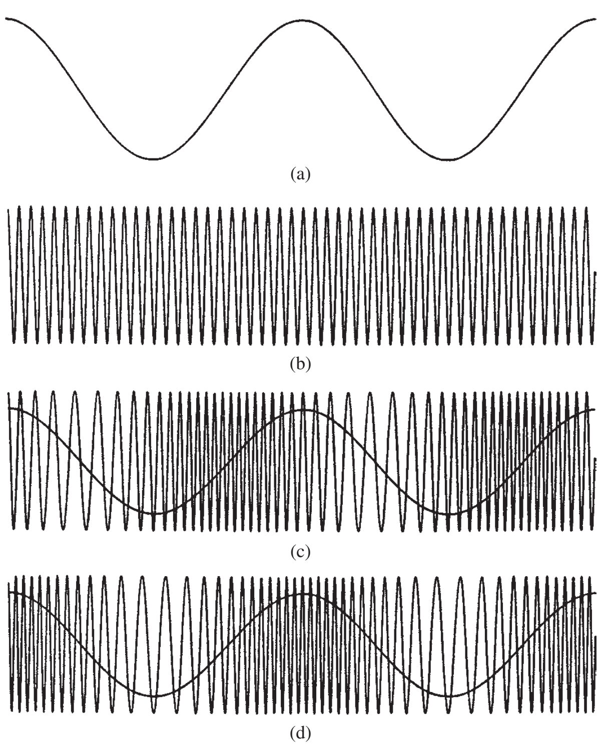


Figure 34: Different modulations of sinusoidal message signal. (a) Message signal. (b) Unmodulated carrier. (c) Output of phase modulator (d) Output of frequency modulator [15, Fig 4.2 p 159 ]

The phase deviation of the PM output is proportional to $m(t)$. However, because the phase is varied continuously, it is not straightforward (yet) to see how Figure 34c is related to $m(t)$. In Figure 38, we will come back to this example and re-analyze the PM output.

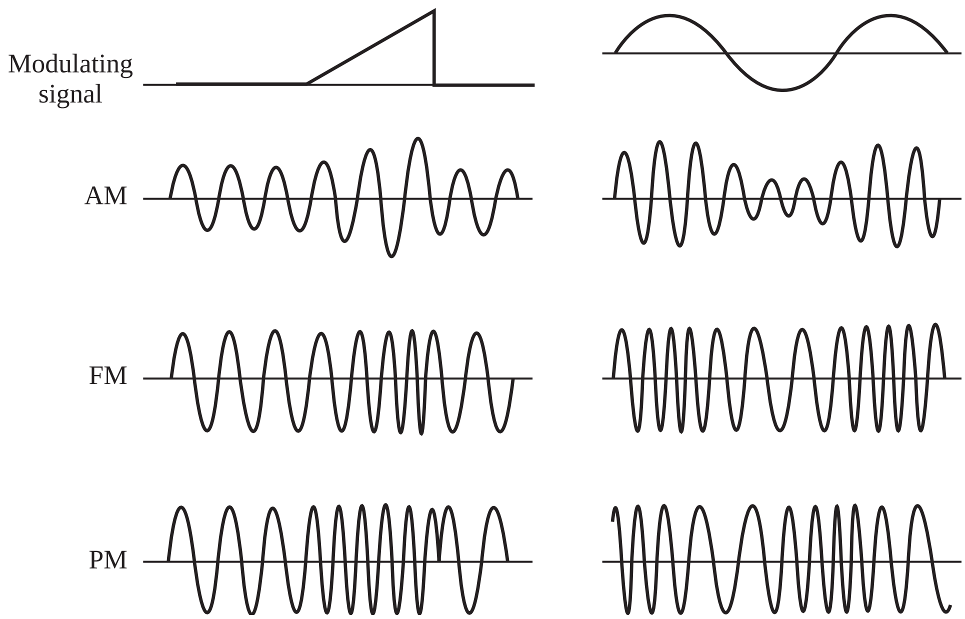


Figure 35: Illustrative AM, FM, and PM waveforms. [3, Fig 5.1-2 p 212]

**Example 5.7.** Figure 35 illustrates the outputs of AM, FM, and PM modulators when the message is a triangular (ramp) pulse.

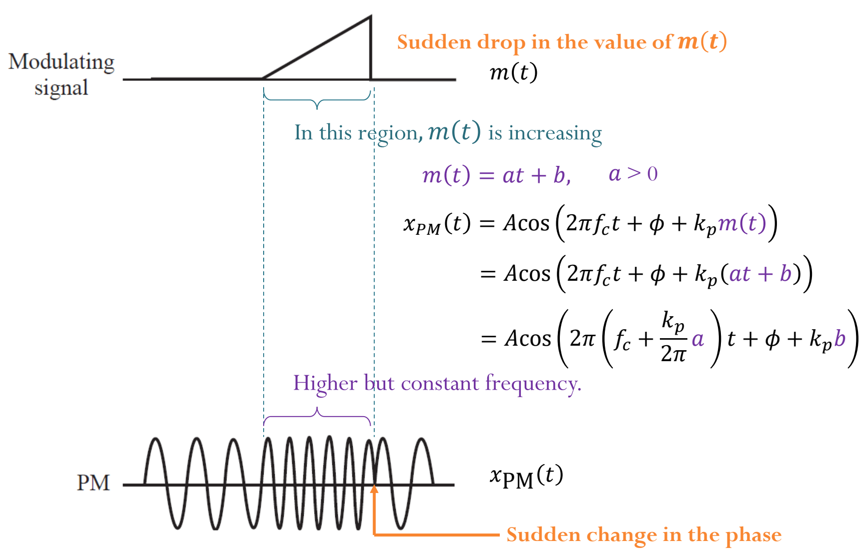


Figure 36: Explaining PM waveform in Figure 35.

To understand more about FM, we will first need to know what it actually means to vary the frequency of a sinusoid.

## 5.1 Instantaneous Frequency

**Definition 5.8.** The ***generalized sinusoidal*** signal is a signal of the form

$$x(t) = A\cos(\theta(t)) \tag{73}$$

where $\theta(t)$ is called the ***generalized angle***.

- The generalized angle for conventional sinusoid is $\theta(t) = 2\pi f_c t + \phi$.
- In [3, p 208], $\theta(t)$ of the form $2\pi f_c t + \phi(t)$ is called the **total instantaneous angle**.

**Definition 5.9.** If $\theta(t)$ in (73) contains the message information $m(t)$, we have a process that may be termed **angle modulation**.

- The amplitude of an angle-modulated wave is constant.
- Another name for this process is **exponential modulation**.
  - The motivation for this name is clear when we write $x(t)$ as $A\mathrm{Re}\left\{e^{j\theta(t)}\right\}$.
  - It also emphasizes the nonlinear relationship between $x(t)$ and $m(t)$.
- Since exponential modulation is a nonlinear process, the modulated wave $x(t)$ does not resemble the message waveform $m(t)$.

**5.10.** Suppose we want the frequency $f_c$ of a carrier $A\cos(2\pi f_c t)$ to vary with time as in (72). It is tempting to consider the signal

$$A\cos(2\pi g(t)t), \tag{74}$$

where $g(t)$ is the desired frequency at time $t$.

**Example 5.11.** Consider the generalized sinusoid signal of the form 74 above with $g(t) = t^2$. We want to find its frequency at $t = 2$.

(a) Suppose we guess that its frequency at time $t$ should be $g(t)$. Then, at time $t = 2$, its frequency should be $t^2 = 4$. However, when compared with $\cos(2\pi(4)t)$ in Figure 37a, around $t = 2$, the "frequency" of $\cos(2\pi\left(t^2\right)t)$ is quite different from the 4-Hz cosine approximation. Therefore, 4 Hz is too low to be the frequency of $\cos(2\pi\left(t^2\right)t)$ around $t = 2$.

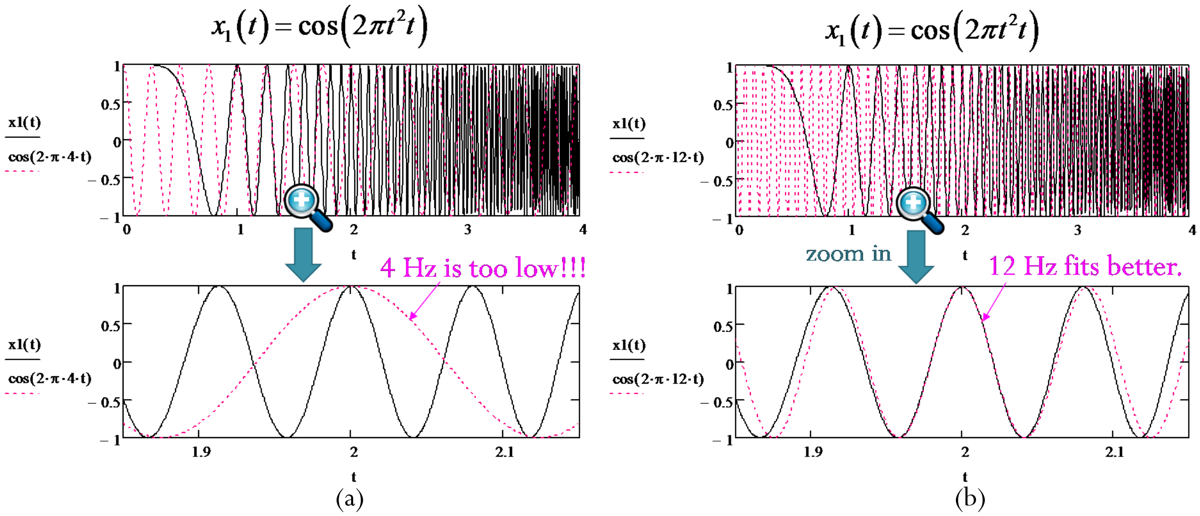


Figure 37: Approximating the frequency of $\cos(2\pi\,(t^2)\,t)$ by (a) $\cos(2\pi(4)t)$ and (b) $\cos(2\pi(12)t)$.

(b) Alternatively, around $t = 2$, Figure 37b shows that $\cos(2\pi(12)t)$ seems to provide a good approximation. So, 12 Hz would be a better answer.

**Definition 5.12.** For generalized sinusoid $A\cos(\theta(t))$, the ***instantaneous frequency***[23] at time $t$ is given by

$$f(t) = \frac{1}{2\pi}\frac{d}{dt}\theta(t). \tag{75}$$

**Example 5.13.** For the signal $\cos(2\pi\left(t^2\right)t)$ in Example 5.11,

$$\theta(t) = 2\pi\left(t^2\right)t$$

and the instantaneous frequency is

$$f(t) = \frac{1}{2\pi}\frac{d}{dt}\theta(t) = \frac{1}{2\pi}\frac{d}{dt}\left(2\pi\left(t^2\right)t\right) = 3t^2.$$

In particular, $f(2) = 3 \times 2^2 = 12$.

**5.14.** The instantaneous frequency formula (75) implies

$$\theta(t) = 2\pi\int_{-\infty}^{t} f(\tau)d\tau = \theta(t_0) + 2\pi\int_{t_0}^{t} f(\tau)d\tau. \tag{76}$$

[23] Although $f(t)$ is measured in hertz, it should not be equated with spectral frequency. Spectral frequency $f$ is the independent variable of the frequency domain, whereas instantaneous frequency $f(t)$ is a time-dependent property of waveforms with exponential modulation.

## 5.2 FM and PM

**Definition 5.15.** ***Frequency modulation*** (***FM***):

$$x_{\mathrm{FM}}(t) = A\cos\left(2\pi f_c t + \phi + 2\pi k_f \int_{-\infty}^{t} m(\tau)d\tau\right). \tag{77}$$

Its instantaneous frequency is

$$f(t) = f_c + k_f m(t).$$

**5.16.** ***Phase modulation*** (***PM***): The phase-modulated signal is defined in Definition 5.3 to be

$$x_{\mathrm{PM}}(t) = A\cos\left(2\pi f_c t + \phi + k_p m(t)\right)$$

When $m(t)$ is differentiable, the instantaneous frequency of $x_{\mathrm{PM}}(t)$ is

(78)

Therefore, the instantaneous frequency of the PM signal varies in proportion to the slope of $m(t)$.

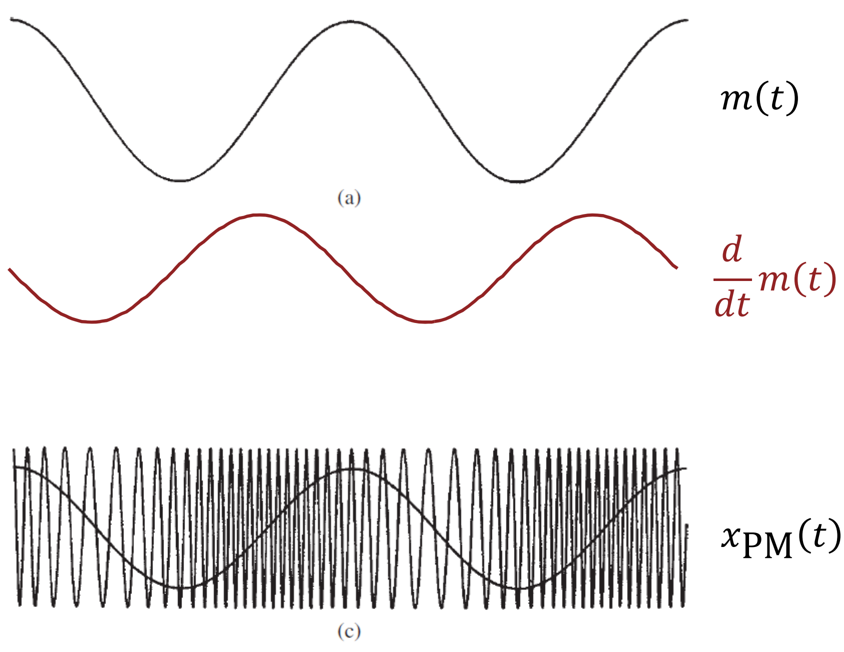


Figure 38: A revisit of the PM signal in Figure 34.

In particular, the instantaneous frequency of the PM signal is maximum when the slope of $m(t)$ is maximum and minimum when the slope of $m(t)$ is minimum.

**Example 5.17.** Sketch FM and PM waves for the modulating signal $m(t)$ shown in Figure 39a.

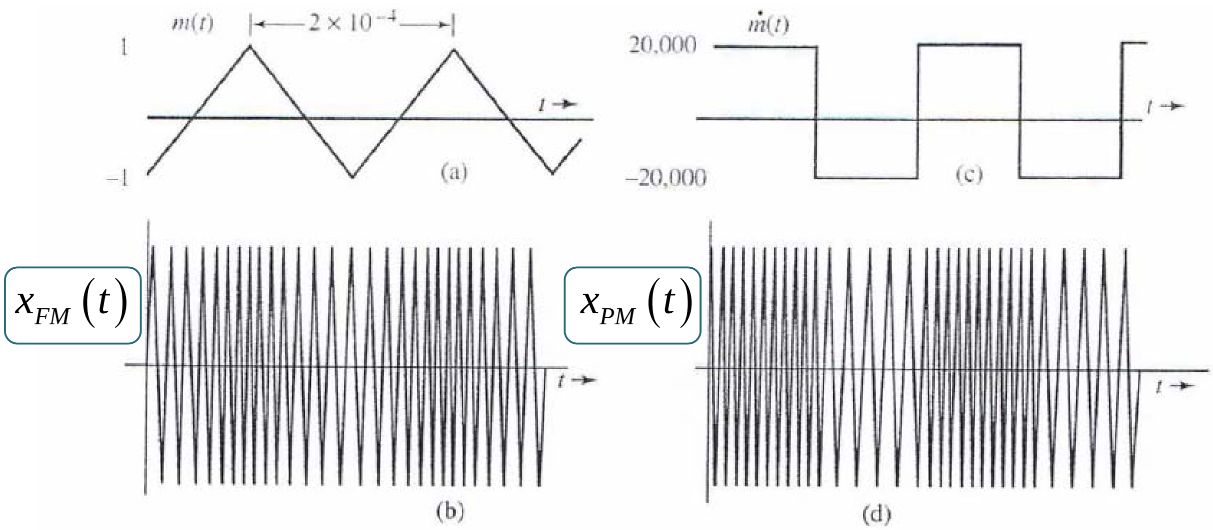


Figure 39: FM and PM waveforms generated from the same message.

**5.18.** The "indirect" method of sketching $x_{\text{PM}}(t)$ (using $\dot{m}(t)$ to frequency-modulate a carrier) works as long as $m(t)$ is a continuous signal. If $m(t)$ is discontinuous, this indirect method fails at points of discontinuities. In such a case, a direct approach should be used to specify the sudden phase changes. This is illustrated in Example 5.20.

**5.19.** Summary: To sketch $x_{\text{PM}}(t)$ from $m(t)$,

(a) in the region where $m(t)$ is differentiable, vary the the instantaneous frequency of $x_{\text{PM}}(t)$ in proportion to the slope of $m(t)$

(b) at the location where $m(t)$ is discontinuous (has a jump), calculate the amount of phase shift from the jump amount:

$$\Delta\theta = \theta(t_0^+) - \theta(t_0^-) = k_p \left(m(t_0^+) - m(t_0^-)\right) = k_p \Delta m.$$

**Example 5.20.** Sketch FM and PM waves for the modulating signal $m(t)$ shown in Figure 40a.

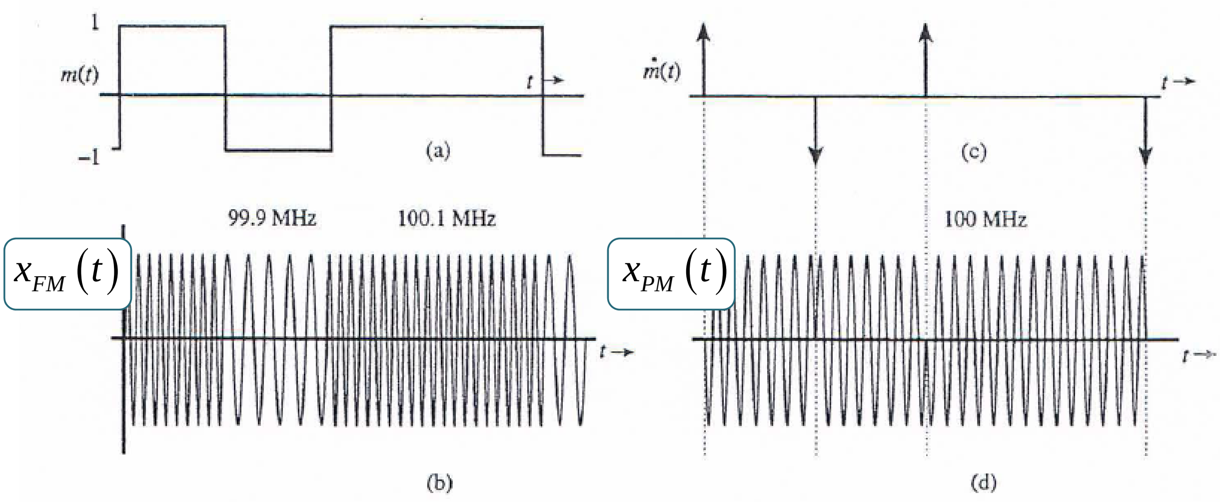


Figure 40: FM and PM waveforms generated from the same message.

**5.21.** ***Generalized angle modulation*** (or ***exponential modulation***):

$$x(t) = A\cos\left(2\pi f_c t + \phi + (m * h)(t)\right)$$

where $h$ is causal.

(a) ***Frequency modulation*** (***FM***): $h(t) = 2\pi k_f 1[t \geq 0]$

(b) ***Phase modulation*** (***PM***): $h(t) = k_p \delta(t)$.

**5.22.** Relationship between FM and PM:

- Equation (77) implies that one can produce frequency-modulated signal from a phase modulator.
- Equation (78) implies that one can produce phase-modulated signal from a frequency modulator.
- The two observations above are summarized in Figure 41.

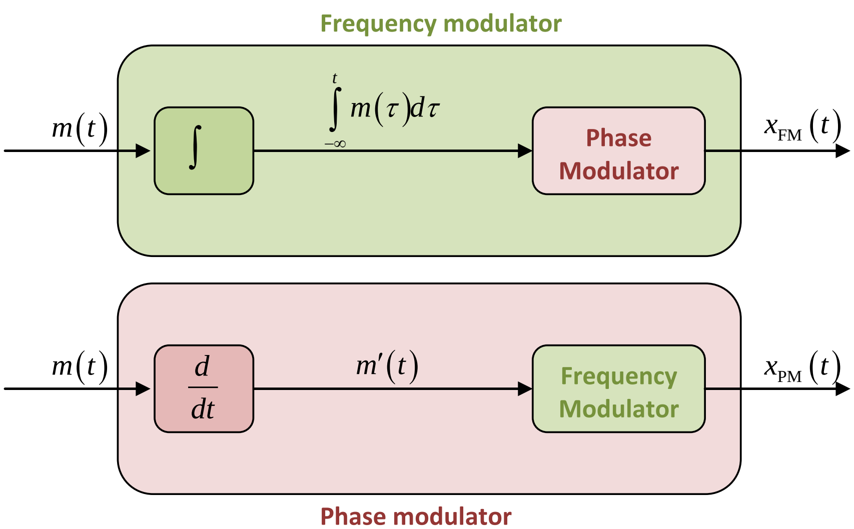


Figure 41: With the help of integrating and differentiating networks, a phase modulator can produce frequency modulation and vice versa [5, Fig 5.2 p 255].

- By looking at an angle-modulated signal $x(t)$, there is no way of telling whether it is FM or PM.
    - Compare Figure 34c and 34d in Example 5.6.
    - In fact, it is meaning less to ask an angle-modulated wave whether it is FM or PM. It is analogous to asking a married man with children whether he is a father or a son. [6, p 255]

**5.23.** So far, we have spoken rather loosely of amplitude and phase modulation. If we modulate two real signals $a(t)$ and $\phi(t)$ onto a cosine to produce the real signal $x(t) = a(t)\cos(\omega_c t + \phi(t))$, then this language seems unambiguous: we would say the respective signals amplitude- and phase-modulate the cosine. But is it really unambiguous?

The following example suggests that the question deserves thought.

**Example 5.24.** [9, p 15] Let's look at a "purely amplitude-modulated" signal

$$x_1(t) = a(t)\cos(\omega_c t).$$

Assuming that $a(t)$ is bounded such that $0 \leq a(t) \leq A$, there is a well-defined function

$$\theta(t) = \cos^{-1}\left(\frac{1}{A}x_1(t)\right) - \omega_c t.$$

Observe that the signal

$$x_2(t) = A\cos\left(\omega_c t + \theta(t)\right)$$

is exactly the same as $x_1(t)$ but $x_2(t)$ looks like a "purely phase-modulated" signal.

**5.25.** Example 5.24 shows that, for a given real signal $x(t)$, the factorization $x(t) = a(t)\cos(\omega_c t + \phi(t))$ is not unique. In fact, there is an infinite number of ways for $x(t)$ to be factored into "amplitude" and "phase".

## 5.3 Bandwidth of FM Signals

**5.26.** FM: The "Holy Grail" Technique for BW Saving?

In the 1920s, the idea of frequency modulation (FM) was naively proposed very early as a method to conserve the radio spectrum. The argument was presented as follows:

- If $m(t)$ is bounded between $-m_p$ and $m_p$, then the maximum and minimum values of the (instantaneous) carrier frequency would be $f_c + k_f m_p$ and $f_c - k_f m_p$, respectively. (Think of this as a delta function shifting to various location between $f_c + k_f m_p$ and $f_c - k_f m_p$ in the frequency domain.)

- Hence, the spectral components would remain within this band with a bandwidth $2k_f m_p$ centered at $f_c$.
- Conclusion: By using an arbitrarily small $k_f$, we could make the information bandwidth arbitrarily small (much smaller than the bandwidth of $m(t)$.

In 1922, Carson argued that this is an ill-considered plan. We will illustrate his reasoning later. In fact, experimental results shows that

As a result of his observation, FM temporarily fell out of favor.

**5.27.** Armstrong (1936) reawakened interest in FM when he realized it had a much different property that was desirable. When the $k_f$ is large, the inverse mapping from the modulated waveform $x_{\mathrm{FM}}(t)$ back to the signal $m(t)$ is much less sensitive to additive noise in the received signal than is the case for amplitude modulation. FM then came to be preferred to AM because of its higher fidelity. [1, p 5-6]

Finding the "bandwidth" of FM Signals turns out to be a difficult task. Here we present a few approximation techniques.

**5.28.** First, from 5.21, we see that both FM and PM can be viewed as

$$x(t) = A\cos\left(2\pi f_c t + \theta_0 + \phi(t)\right) \tag{79}$$

where $\phi(t) = (m * h)(t)$ if $h(t)$ is selected properly.

The Fourier transform of $\phi(t)$ is $\Phi(f) = M(f)H(f)$. So, if $M(f)$ is band-limited to $B$, we know that $\Phi(f)$ is also band-limited to $B$ as well.

Now, let us rewrite (79) as

$$x(t) = A\,\mathrm{Re}\left\{e^{j(2\pi f_c t + \theta_0 + \phi(t))}\right\} = A\,\mathrm{Re}\left\{e^{j(2\pi f_c t + \theta_0)}e^{j\phi(t)}\right\} \tag{80}$$

Recall that Taylor series expansion of $e^z$ around $z = 0$ is

$$e^z = \sum_{k=0}^{\infty}\frac{z^k}{k!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots.$$

Plugging in $z = j\phi(t)$ gives

$$e^{j\phi(t)} = 1 + j\phi(t) + \frac{(j\phi(t))^2}{2!} + \frac{(j\phi(t))^3}{3!} + \cdots = 1 + j\phi(t) - \frac{\phi^2(t)}{2!} + (-j)\,\frac{\phi^3(t)}{3!} + \cdots \tag{81}$$

Applying the Euler's formula

$$e^{j(2\pi f_c t + \theta_0)} = \cos\left(2\pi f_c t + \theta_0\right) + j\sin\left(2\pi f_c t + \theta_0\right)$$

and (81) to (80) gives

$$x\left(t\right) = A\left(\cos\left(2\pi f_c t + \theta_0\right) - \phi(t)\sin\left(2\pi f_c t + \theta_0\right) - \frac{\phi^2(t)}{2!}\cos\left(2\pi f_c t + \theta_0\right) + \frac{\phi^3(t)}{3!}\sin\left(2\pi f_c t + \theta_0\right) + \cdots\right).$$

Recall that if $\phi(t)$ is band-limited to $B$, then $\phi^n(t)$ is band-limited to $nB$. With such series, there is no bound for the value of $n$ and therefore, we conclude that the absolute bandwidth would be infinite.

**5.29. Narrowband Angle Modulation**: When $\phi(t)$ is small, we may approximate $e^z$ by $z + 1$. Therefore,

$$e^{j\phi(t)} \approx 1 + j\phi(t). \tag{82}$$

Applying the Euler's formula

$$e^{j(2\pi f_c t + \theta_0)} = \cos\left(2\pi f_c t + \theta_0\right) + j\sin\left(2\pi f_c t + \theta_0\right)$$

and (82) to (80) gives

$$x(t) = A\,\mathrm{Re}\left\{e^{j(2\pi f_c t+\theta_0)}e^{j\phi(t)}\right\}$$

$$\begin{aligned}&\approx A\,\mathrm{Re}\left\{\left(\cos\left(2\pi f_c t+\theta_0\right)+j\sin\left(2\pi f_c t+\theta_0\right)\right)\left(1+j\phi(t)\right)\right\}\\ &= A\cos\left(2\pi f_c t+\theta_0\right)-A\phi(t)\sin\left(2\pi f_c t+\theta_0\right)\end{aligned}$$

- The "approximated" expression of $x(t)$ is similar to AM.
  - The first term yields a carrier component. The second term generates a pair of sidebands. Thus, if $\phi(t)$ has a bandwidth $B$, the bandwidth of $x(t)$ is $2B$.
- The important difference between AM and angle modulation is that the sidebands are produced by multiplication of the message-bearing signal, $\phi(t)$, with a carrier that is in phase quadrature with the carrier component, whereas for AM they are not.
- The FM signal whose $\left|2\pi k_f \int\limits_{-\infty}^{t} m(\tau)d\tau\right| \ll 1$ is called **narrowband FM (NBFM)**. The PM signal whose $|k_p m(t)| \ll 1$ is called **narrowband PM (NBPM)**. Note that these conditions are satisfied when $k_f \ll 1$ or $k_p \ll 1$, respectively. [6, p 260]
- For larger values of $|\phi(t)|$ the terms $\phi^2(t)$, $\phi^3(t)$, ...in (81) cannot be ignored and will increase the bandwidth of $x(t)$.
- Recall, from (32) that

$$g(t)\cos(2\pi f_c t+\phi) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} \frac{1}{2}\left(e^{j\phi}G(f-f_c)+e^{-j\phi}G(f+f_c)\right).$$

Therefore, when

$$x(t) \approx A\cos\left(2\pi f_c t+\theta_0\right)-A\phi(t)\cos\left(2\pi f_c t+\theta_0-90^\circ\right),$$

we have

$$X(f) \approx \frac{A}{2}\left(e^{j\theta_0}\delta(f-f_c)+e^{-j\theta_0}\delta(f+f_c)-e^{j(\theta_0-90^\circ)}\Phi(f-f_c)-e^{-j(\theta_0-90^\circ)}\Phi(f+f_c)\right)$$
$$= \frac{A}{2}\left(e^{j\theta_0}\delta(f-f_c)+e^{-j\theta_0}\delta(f+f_c)+je^{j\theta_0}\Phi(f-f_c)-je^{-j\theta_0}\Phi(f+f_c)\right).$$

**5.30. Wideband FM (WBFM)**: For potentially wideband $m(t)$, here, we present a technique to roughly estimate the bandwidth of $x_{\mathrm{FM}}(t)$.

To do this, we consider $m(t)$ that is a piecewise constant function (also known as step function or staircase function); this implies that the instantaneous frequency $f(t) = f_c + k_f m(t)$ of $x_{\mathrm{FM}}(t)$ is also piecewise constant as shown in Figure 42.

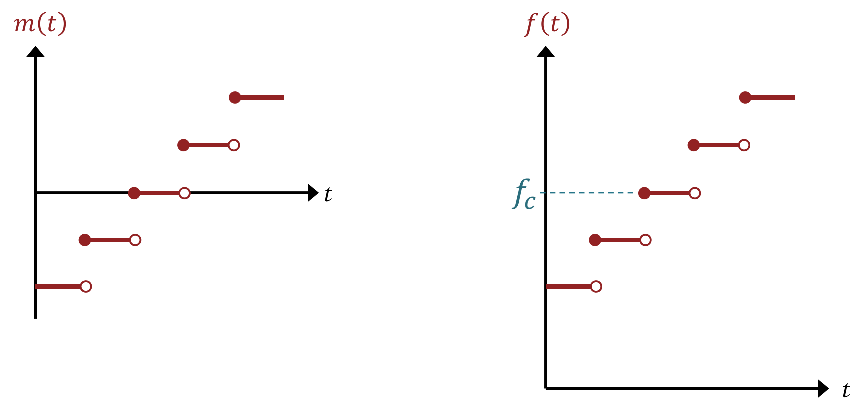


Figure 42: FM for discrete-valued (digital) message

For example, we can consider the transmitted signal $x_{\mathrm{FM}}(t)$ constructed from five different tones. Its instantaneous frequency is increased from $f_1$ to $f_5$.

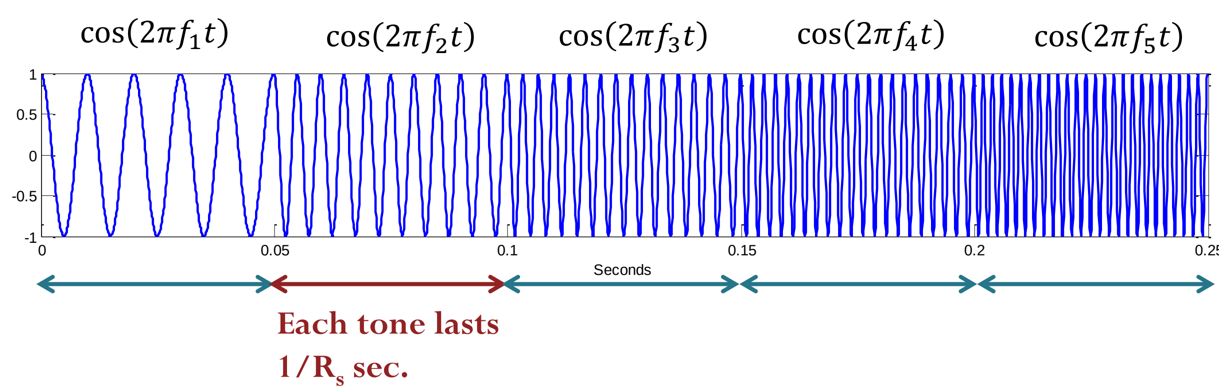


Figure 43: $x_{\mathrm{FM}}(t)$ for discrete-valued (digital) message in Figure 42.

Assume that each tone lasts $T_s = \frac{1}{R_s}$ [s] where $R_s$ is called the "(symbol) rate" of the data transmission. The value of $R_s$ indicates how fast the values of $m(t)$ is changed. Increasing the value of $R_s$ reduces the time to complete the transmission.

Recall that the Fourier transform of a cosine contains simply (two shifted and scaled) delta functions at the (plus and minus) frequency of the cosine. However, recall also that when we consider the cosine pulse, which is time-limited, its Fourier transform contains (two) sinc functions. In particular, the cosine pulse

$$p(t) = \begin{cases} \cos(2\pi f_0 t), & t_1 \le t < t_2, \\ 0, & \text{otherwise}, \end{cases}$$

can be viewed as the pure cosine function $\cos(2\pi f_0 t)$ multiplied by a rectangular pulse $r(t) = 1[t_1 \le t < t_2]$. By (31), we know that multiplication by $\cos(2\pi f_0 t)$ will shift the spectrum $R(f)$ of the rectangular pulse to $\pm f_c$ and scaled its values by a factor of $\frac{1}{2}$: $P(f) = \frac{1}{2}R(f - f_0) + \frac{1}{2}R(f + f_0)$

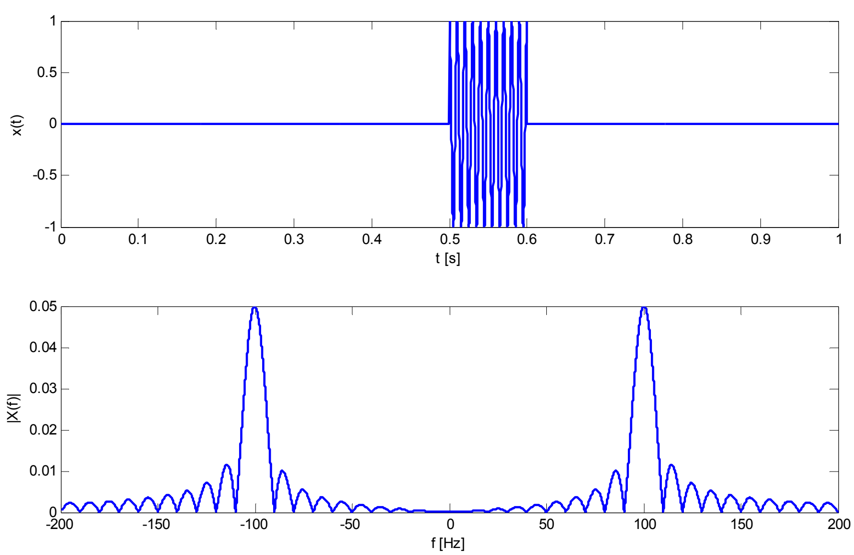


Figure 44: Cosine pulse and its spectrum which contains two sinc functions at $\pm$ freqeuncy of the cosine (which is 100 Hz in the figure). When the pulse only lasts for a short time period, the sinc pulses in the frequency domain are wide.

where the Fourier transform[24] $R(f)$ of the rectangular pulse is given by

$$R(f) = (t_2 - t_1)\, e^{-j\pi f(t_1+t_2)} \operatorname{sinc}(\pi f (t_2 - t_1)).$$

[24] To get this, first consider the rectangular pulse of width $t_2 - t_1$ centered at t = 0. From (15), the corresponding Fourier transform is $2\left(\frac{t_2-t_1}{2}\right)\operatorname{sinc}\left(2\pi\left(\frac{t_2-t_1}{2}\right)f\right)$. Finally, by time-shifting the rectangular pulse in the time domain by $\frac{t_2+t_1}{2}$, we simply multiply the Fourier transform by $e^{-2\pi f\left(\frac{t_2-t_1}{2}\right)}$ in the frequency domain.

See Figure 44 for an example.

When $m(t)$ is piecewise constant, $x_{\mathrm{FM}}(t)$ is a sum of cosine pulses. Therefore, its spectrum $X(f)$ will be the sum of the sinc functions centered at the frequencies of the pulses as shown in Figure 45.

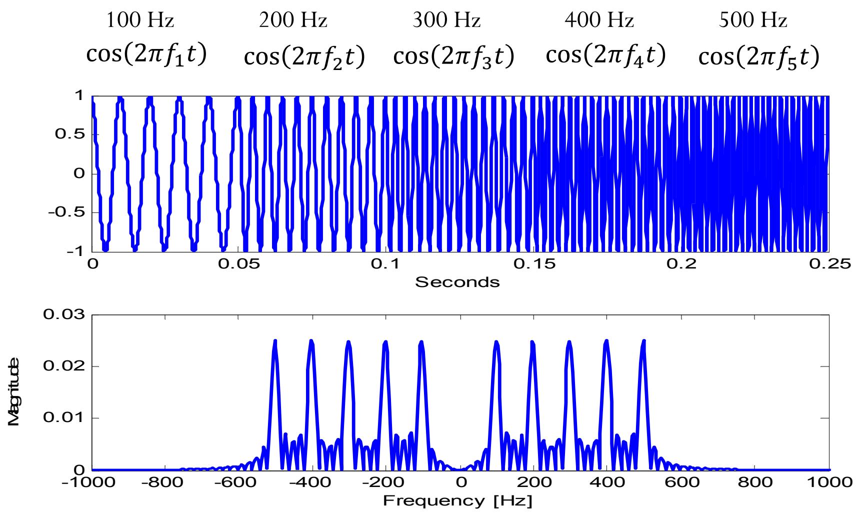


Figure 45: A digital version of FM: $x_{\mathrm{FM}}(t)$ and the corresponding $X_{\mathrm{FM}}(f)$.

- $X(f)$ extends to $\pm\infty$. It is not band-limited.
- One may approximate its bandwidth by assuming that "most" of the energy in the sinc function is contained in its main lobe which is at $\pm\frac{1}{T_s} = \pm R_s$ from its peak. Therefore, the bandwidth of $x_{\mathrm{FM}}(t)$ becomes

$$\mathrm{BW}_{\mathrm{FM}} \approx R_s + (f_{\max} - f_{\min}) + R_s = (f_{\max} - f_{\min}) + 2R_s$$