ECS332 2017/1 Part II.4 Dr.Prapun

5 Angle Modulation: FM and PM

5.1. We mentioned in 4.2 that a sinusoidal carrier signal

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

DSB-SC : $\sqrt{2}\, m(t)$

AM : $A + m(t)$

$f(t)$

$\phi(t)$

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 4.61, 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***):

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

- max phase deviation:

shift

$-m_p \le m(t) \le m_p$

$-k_p m_p \le k_p m(t) \le k_p m_p$

$\phi_\Delta = k_p m_p$

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

$$f(t) = f_c + km(t), \qquad (65)$$

$-m_p \leq m(t) \leq m_p$

$-k_f m_p \leq k_f m(t) \leq k_f m_p$

where k is an arbitrary constant.

- The arbitrary constant k is sometimes denoted by k_f to distinguish it from a similar constant in PM.
- $f(t)$ is varied from $f_c - km_p$ to $f_c + km_p$.

  max. freq. deviation shift $f_\Delta = k_f m_p$

- f_c is assumed to be large enough such that $f(t) \geq 0$.

Example 5.5. Figure 31 illustrates the outputs of PM and FM modulators when the message is a unit-step function.

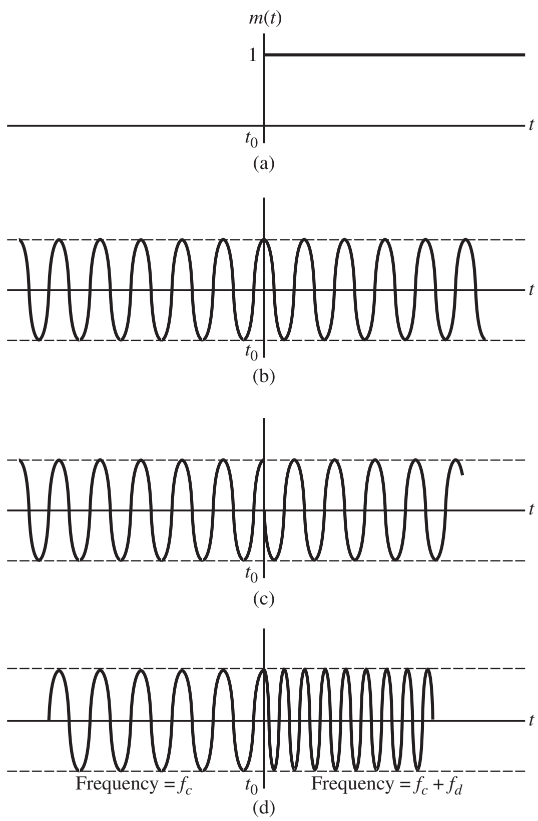


Figure 31: Comparison of PM and FM modulator outputs for a unit-step input. (a) Message signal. (b) Unmodulated carrier. (c) Phase modulator output (d) Frequency modulator output. [15, Fig 4.1 p 158]

---

[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 32a, the frequency deviation of the FM modulator output in Figure 32d 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.

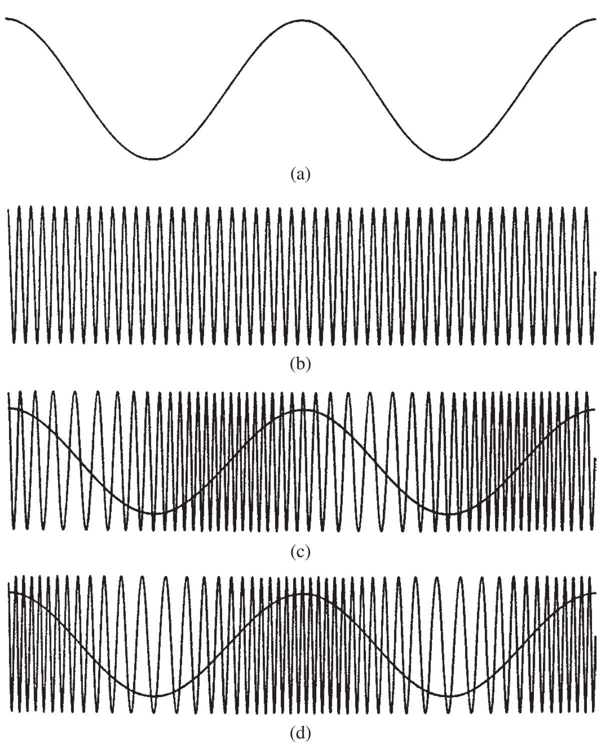


Figure 32: Different modulations of sinusoidal message signal. (a) Message signal. (b) Unmodulated carrier. (c) Output of phase modulator (d) Output of frequency modulator [15, Fig 4.2 p 159]

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 32c is related to $m(t)$. In Figure 36, we will come back to this example and re-analyze the PM output.

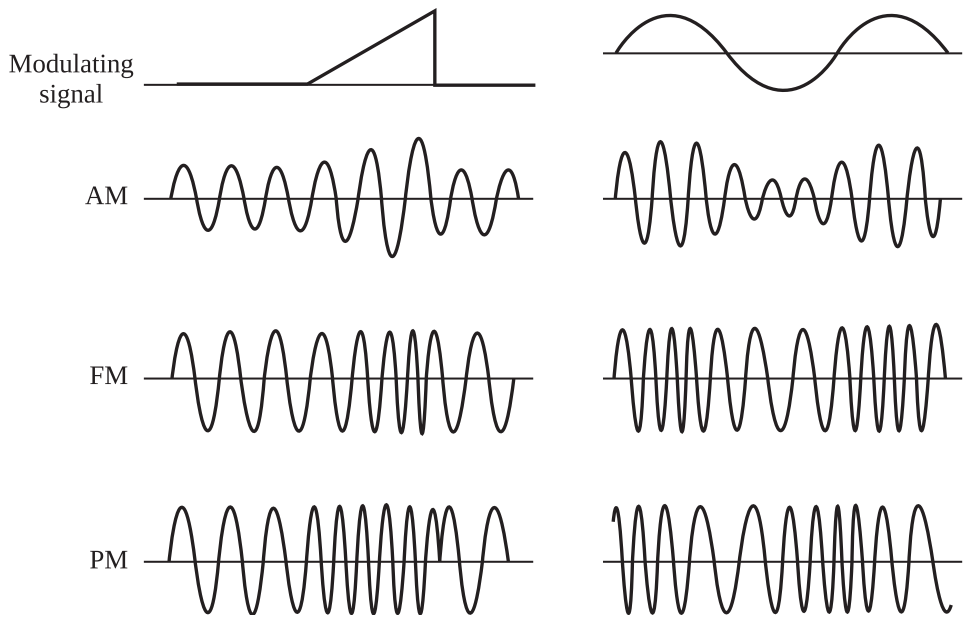


Figure 33: Illustrative AM, FM, and PM waveforms. [3, Fig 5.1-2 p 212]

Example 5.7. Figure 33 illustrates the outputs of AM, FM, and PM modulators when the message is a triangular (ramp) pulse.

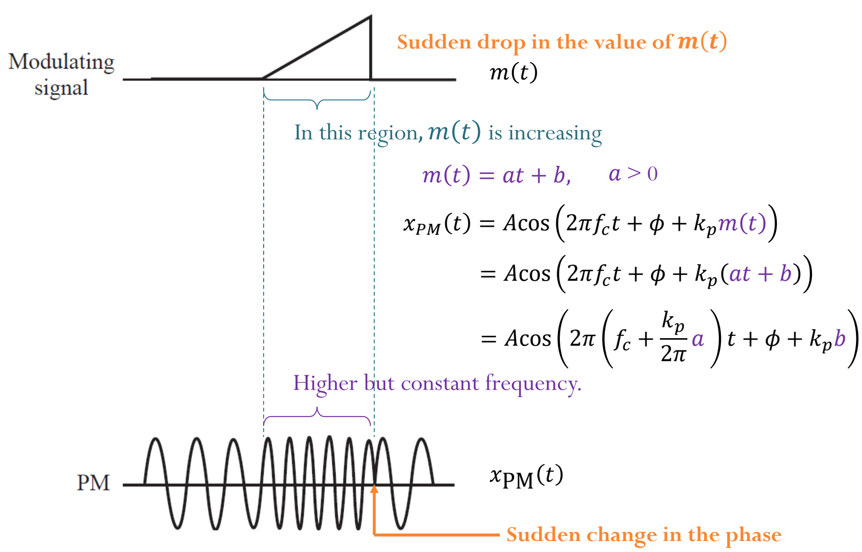


Figure 34: Explaining PM waveform in Figure 33.

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

$$x(t) = A\cos\left(\theta(t)\right) \tag{66}$$

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 (66) 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 (65). It is tempting to consider the signal

$$A\cos(2\pi g(t)t), \tag{67}$$

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

Example 5.11. Consider the generalized sinusoid signal of the form 67 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 $\cos\left(2\pi(4)t\right)$ in Figure 35a, 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 $t = 2$.

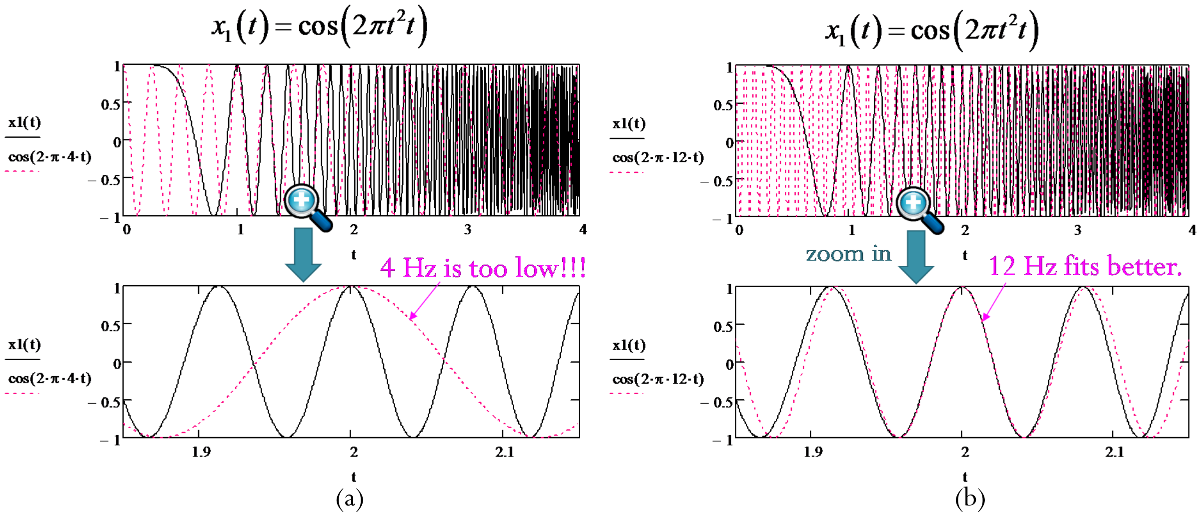


Figure 35: Approximating the frequency of $\cos(2\pi\,(t^2)\,t)$ by (a) $\cos(2\pi(4)t)$ and (b) $\cos(2\pi(12)t)$.

(b) Alternatively, around $t = 2$, Figure 35b 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

$\omega(t)$ instantaneous angular freq.

$$f(t) = \frac{1}{2\pi}\underbrace{\frac{d}{dt}\theta(t)}_{\omega(t)}. \quad [\text{Hz}] \tag{68}$$

[rad/s]

"rate of change of the angle"

Example 5.13. For the signal $\cos(2\pi\,(t^2)\,t)$ 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, $f(2) = 3 \times 2^2 = 12$.

$$f \underset{2\pi\int}{\overset{\frac{1}{2\pi}\frac{d}{dt}}{\rightleftarrows}} \theta$$

5.14. The instantaneous frequency formula (68) implies

$$\theta(t) = 2\pi\int_{-\infty}^{t} f(\tau)d\tau = \theta(t_0) + 2\pi\int_{t_0}^{t} f(\tau)d\tau. \tag{69}$$

[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.

First-order (straight-line) approximation/linearization

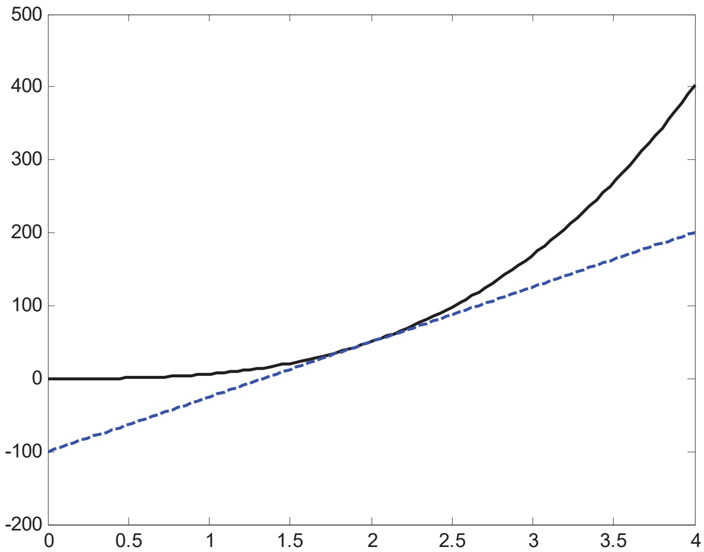
During a small time interval, if $x(t) \approx A\cos(\alpha t+\beta)$, then the freq. is $\frac{\alpha}{2\pi}$.

- How does the formula $f(t) = \frac{1}{2\pi}\frac{d}{dt}\phi(t)$ work?
- Technique from Calculus: first-order (tangent-line) approximation/linearization

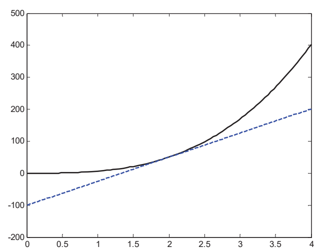
Here, we have $x(t) = A\cos(\underbrace{\theta(t)})$.

During the small time interval around $t=t_0$, calculus says we can approx.

$\theta(t) \approx \alpha t+\beta$



First-order (straight-line) approximation/linearization



- How does the formula $f(t) = \frac{1}{2\pi}\frac{d}{dt}\theta(t)$ work?
- Technique from Calculus: first-order (tangent-line) approximation/linearization
  - When we consider a function $\theta(t)$ near a particular time, say, $t = t_0$, the value of the function is approximately

$$\theta(t) \approx \underbrace{\theta'(t_0)}_{\text{slope}}(t-t_0)+\theta(t_0) = \underbrace{\theta'(t_0)}_{\text{slope}}t + \underbrace{\theta(t_0)-t_0\theta'(t_0)}_{\text{constant}}$$

  - Therefore, near $t = t_0$,

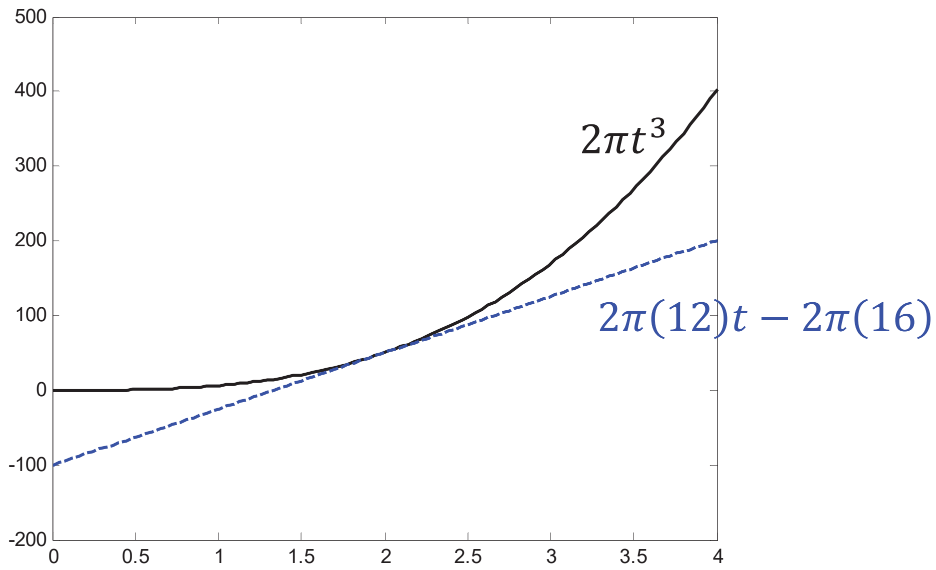
$$\cos(\theta(t)) \approx \cos(\theta'(t_0)t+\theta(t_0)-t_0\theta'(t_0))$$

  - Now, we can directly compare the terms with $\cos(2\pi f_0 t + \phi)$.

First-order (straight-line) approximation/linearization

- For example, for t near $t = 2$,

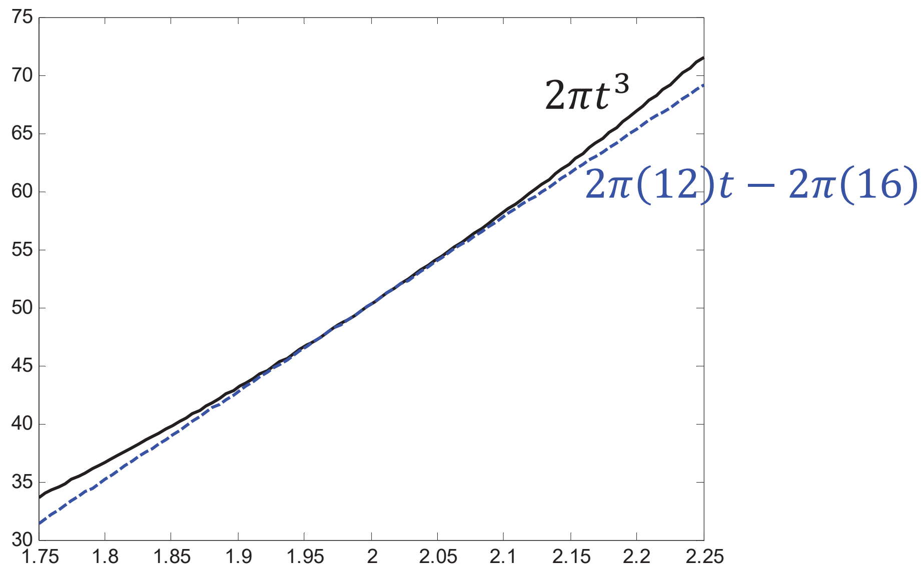
$$2\pi t^3 \approx 2\pi\left(3t^2\right)\Big|_{t=2}(t-2) + 2\pi t^3\Big|_{t=2} = 2\pi(12)t - 2\pi(16)$$



First-order (straight-line) approximation/linearization

- For example, for t near $t = 2$,

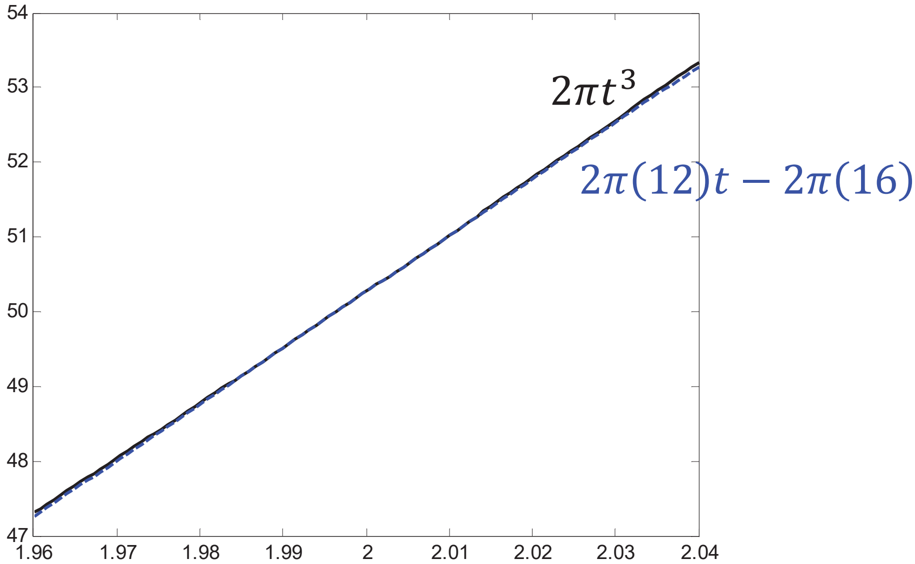
$$2\pi t^3 \approx 2\pi\left(3t^2\right)\Big|_{t=2}(t-2) + 2\pi t^3\Big|_{t=2} = 2\pi(12)t - 2\pi(16)$$



First-order (straight-line) approximation/linearization

- For example, for t near $t = 2$,

$$2\pi t^3 \approx 2\pi\left(3t^2\right)\Big|_{t=2}(t-2) + 2\pi t^3\Big|_{t=2} = 2\pi(12)t - 2\pi(16)$$



Same idea

- Suppose we want to find $\sqrt{15.9}$.
- Let $g(x) = \sqrt{x}$.
  - Note that $\frac{d}{dx} g(x) = \frac{d}{dx}\sqrt{x} = \frac{1}{2\sqrt{x}}$.
- Approximation: $g(x) \approx g'(x_0)(x - x_0) + g(x_0)$
- 15.9 is near 16.
- $\sqrt{15.9} = g(15.9)$
  - $\approx g'(16)(15.9 - 16) + g(16)$
  - $= \frac{1}{2\sqrt{16}}(-0.1) + \sqrt{16} = -\frac{0.1}{8} + 4 = 3.9875$
- MATLAB:

```
>> sqrt(15.9)
ans =
    3.987480407475377
```

5.2 FM and PM

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

$\theta(t)$

$$x_{\mathrm{FM}}(t) = A\cos\left(2\pi f_c t + \phi + 2\pi k_f \int_{-\infty}^{t} m(\tau)d\tau\right). \tag{70}$$

$\frac{1}{2\pi}\frac{d}{dt}\theta(t)$

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

$\theta(t)$

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

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

$$f(t) = \frac{1}{2\pi}\frac{d}{dt}\theta(t) = f_c + \frac{k_p}{2\pi}\frac{d}{dt}m(t) \tag{71}$$

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

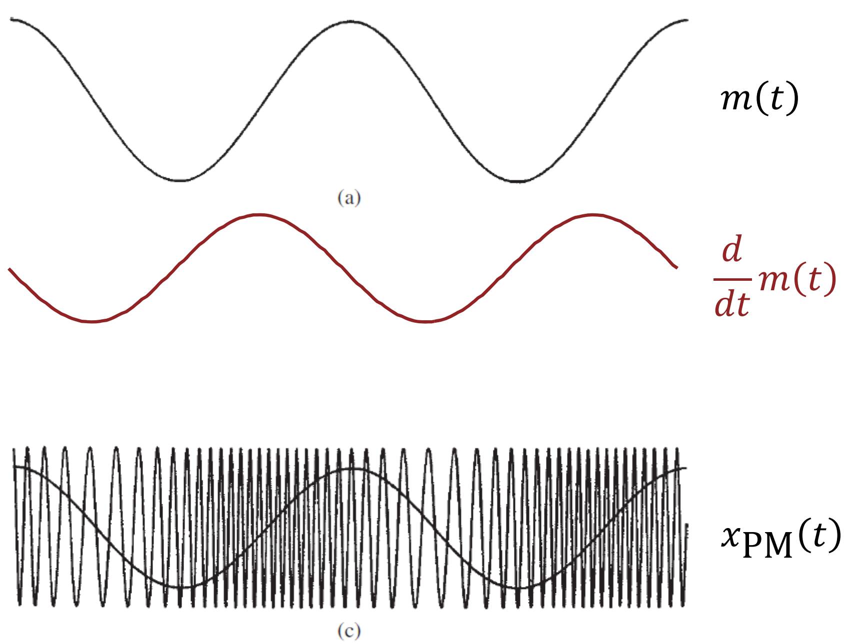


Figure 36: A revisit of the PM signal in Figure 32.

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 37a.

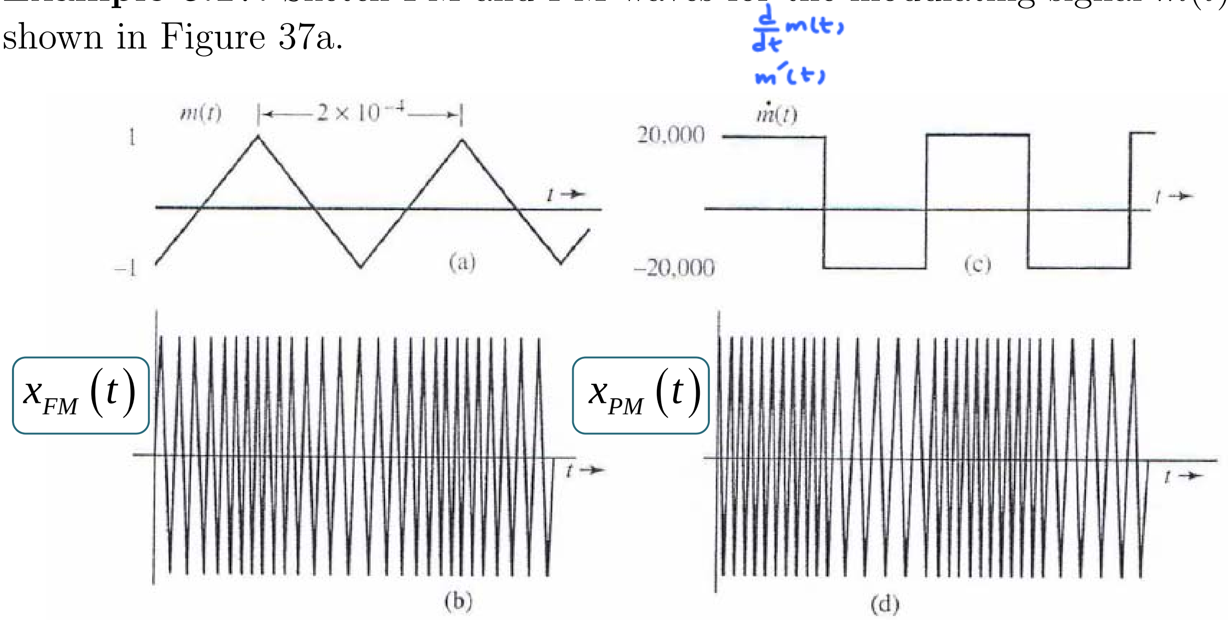


Figure 37: FM and PM waveforms generated from the same message.

5.18. The "indirect" method of sketching $x_{\mathrm{PM}}(t)$ (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 $x_{\mathrm{PM}}(t)$ from $m(t)$,

(a) in the region where $m(t)$ is differentiable, vary the the instantaneous frequency of $x_{\mathrm{PM}}(t)$ 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:

jump size

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

$k_p = \frac{\pi}{2}$ $\Delta\theta = k_p \Delta m = -2k_p = -\pi$

$\Delta m = -2$

$\frac{d}{dt} m(t) = 0$ almost everywhere.

↳ $f(t) \equiv f_c$ almost everywhere

Example 5.20. Sketch FM and PM waves for the modulating signal $m(t)$ shown in Figure 38a. $\Delta\theta = k_p \Delta m = 2k_p = \pi$

$\Delta m = +2$

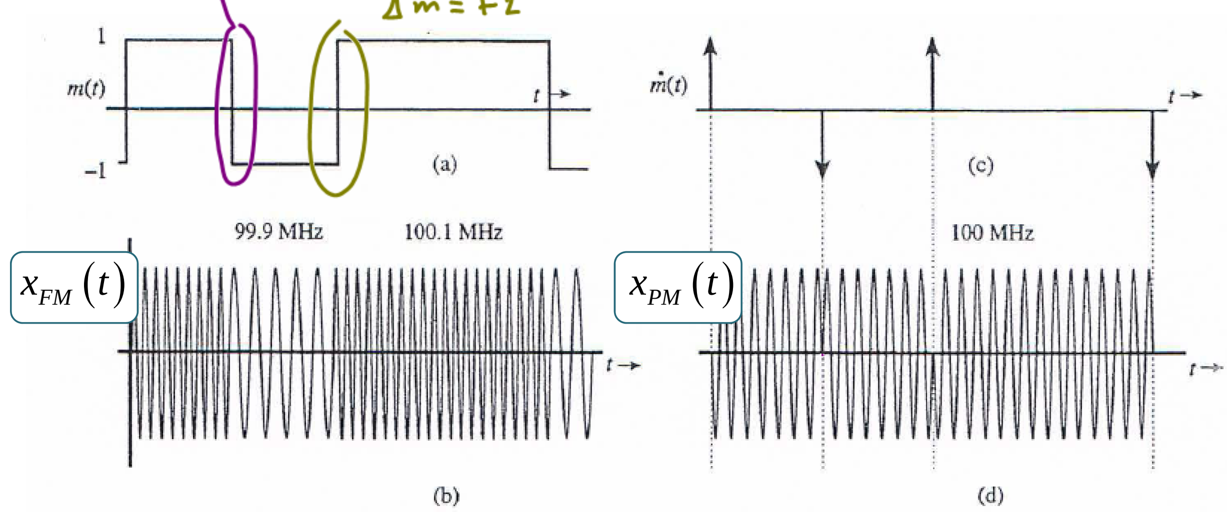


Figure 38: FM and PM waveforms generated from the same message.

$\phi(t) \xrightarrow{\mathcal{F}} M(f)H(f)$

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

$$x(t) = A\cos\left(2\pi f_c t + \theta_0 + (m * h)(t)\right)$$

convolution

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 (70) implies that one can produce frequency-modulated signal from a phase modulator.
- Equation (71) implies that one can produce phase-modulated signal from a frequency modulator.
- The two observations above are summarized in Figure 39.

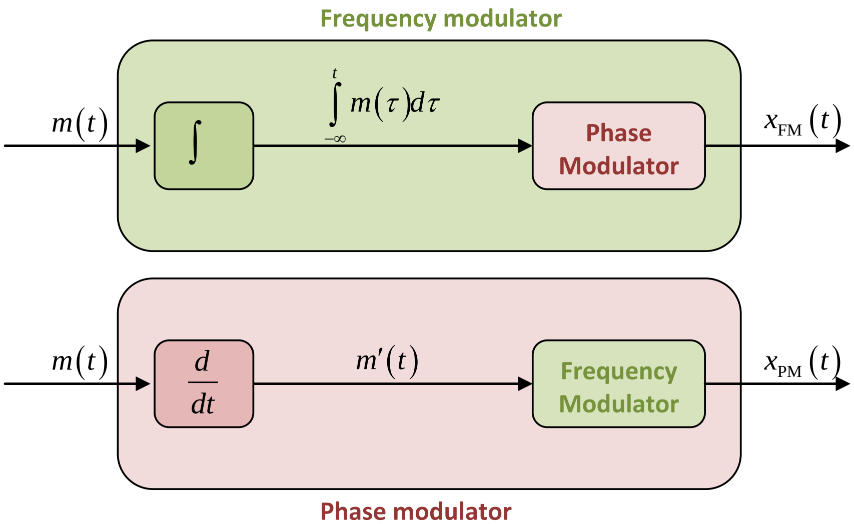


Figure 39: With the help of integrating and differentiating networks, a phase modulator can produce frequency modulation and vice versa [5, Fig 5.2 p 255].

- By looking at an angle-modulated signal $x(t)$, there is no way of telling whether it is FM or PM.
  - Compare Figure 32c and 32d 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 $0 \le a(t) \le A$, 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".