Sirindhorn International Institute of Technology
Thammasat University
School of Information, Computer and Communication Technology

ECS 332: Principles of Communications (Fourier Transform and Communication Systems)

Prapun Suksompong, Ph.D.
prapun@siit.tu.ac.th

August 10, 2012

Communication systems are usually viewed and analyzed in frequency domain. This note reviews some basic properties of Fourier transform and introduce basic communication systems.

Contents

ECS332 2012/1 Part I Dr.Prapun

1 Introduction to communication systems

1.1. Shannon's insight [8]:

> The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.

Definition 1.2. Figure 1 [8] shows a commonly used model for a (single-link or point-to-point) communication system. All information transmission systems involve three major subsystems–a transmitter, the channel, and a receiver.

(a) **Information**[1] source: produce a **message**

- Messages may be categorized as **analog** (continuous) or **digital** (discrete).

(b) **Transmitter**: operate on the message to create a **signal** which can be sent through a channel

(c) **Channel**: the medium over which the signal, carrying the information that composes the message, is sent

- All channels have one thing in common: the signal undergoes **degradation** from transmitter to receiver.

[1]The concept of information is central to communication. But information is a loaded word, implying semantic and philosophical notions that defy precise definition. We avoid these difficulties by dealing instead with the message, defined as the physical manifestation of information as produced by the source. [3, p 2]

- Although this degradation may occur at any point of the communication system block diagram, it is customarily associated with the channel alone.
- This degradation often results from ***noise***[2] and other undesired signals or ***interference***[3] but also may include other ***distortion***[4] effects as well, such as fading signal levels, multiple transmission paths, and filtering.

(d) **Receiver**: transform the signal back into the message intended for delivery

(e) **Destination**: a person or a machine, for whom or which the message is intended

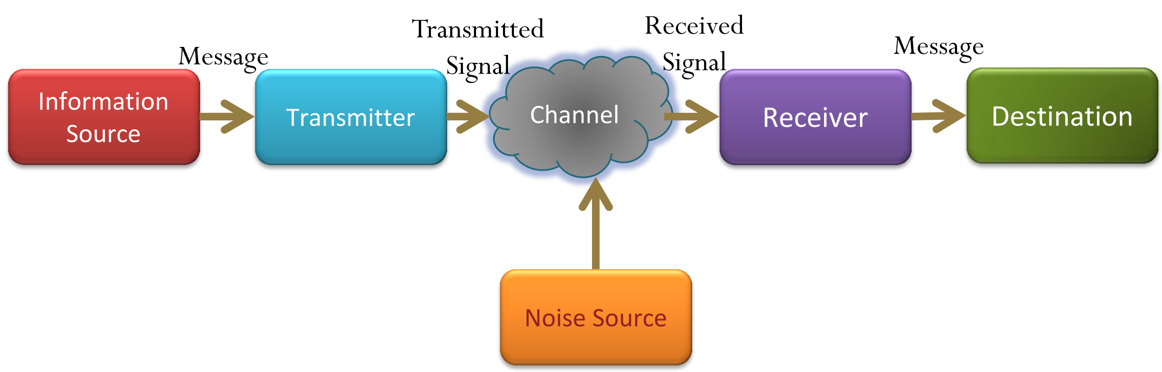


Figure 1: Schematic diagram of a general communication system

[2]Noise refers to random and unpredictable electrical signals produced by natural processes both internal and external to the system. [3, p 4]

[3]Interference is contamination by extraneous signals from human sourcesother transmitters, power lines and machinery, switching circuits, and so on. Interference occurs most often in radio systems whose receiving antennas usually intercept several signals at the same time. [3, p 4]

[4]Distortion is waveform perturbation caused by imperfect response of the system to the desired signal itself. Unlike noise and interference, distortion disappears when the signal is turned off. If the channel has a linear but distorting response, then distortion may be corrected, or at least reduced, with the help of special filters called equalizers. [3, p 4]

2 Frequency-Domain Analysis

Electrical engineers live in the two worlds, so to speak, of time and frequency. Frequency-domain analysis is an extremely valuable tool to the communications engineer, more so perhaps than to other systems analysts. Since the communications engineer is concerned primarily with signal bandwidths and signal locations in the frequency domain, rather than with transient analysis, the essentially steady-state approach of the (complex exponential) **Fourier series** and **transforms** is used rather than the Laplace transform.

2.1 Math background

2.1. ***Euler's formula***: $\boxed{e^{jx} = \cos x + j\sin x.}$

$$\cos(A) = \text{Re}\left\{e^{jA}\right\} = \frac{1}{2}\left(e^{jA} + e^{-jA}\right)$$
$$\sin(A) = \text{Im}\left\{e^{jA}\right\} = \frac{1}{2j}\left(e^{jA} - e^{-jA}\right).$$

2.2. We can use $\cos x = \frac{1}{2}\left(e^{jx} + e^{-jx}\right)$ and $\sin x = \frac{1}{2j}\left(e^{jx} - e^{-jx}\right)$ to derive many trigonometric identities.

Example 2.3. $\cos^2(x) = \frac{1}{2}(\cos(2x) + 1)$

2.4. Similar technique gives

(a) $\cos(-x) = \cos(x)$,

(b) $\cos\left(x - \frac{\pi}{2}\right) = \sin(x)$,

(c) $\sin(x)\cos(x) = \frac{1}{2}\sin(2x)$, and

(d) the **product-to-sum formula**

$$\cos(x)\cos(y) = \frac{1}{2}\left(\cos(x+y) + \cos(x-y)\right). \tag{1}$$

2.2 Continuous-Time Fourier Transform

Definition 2.5. The (direct) **Fourier transform** of a signal $g(t)$ is defined by

$$G(f) = \int_{-\infty}^{+\infty} g(t)e^{-j2\pi ft}dt \tag{2}$$

This provides the frequency-domain description of $g(t)$. Conversion back to the time domain is achieved via the **inverse (Fourier) transform**:

$$g(t) = \int_{-\infty}^{\infty} G(f)\, e^{j2\pi ft}df \tag{3}$$

- We may combine (2) and (3) into one compact formula:

$$\boxed{\int_{-\infty}^{\infty} G(f)\, e^{j2\pi ft}df = g(t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} G(f) = \int_{-\infty}^{\infty} g(t)\, e^{-j2\pi ft}dt.} \tag{4}$$

- We may simply write $G = \mathcal{F}\{g\}$ and $g = \mathcal{F}^{-1}\{G\}$.
- Note that $G(0) = \int g(t)dt$ and $g(0) = \int G(f)df$.

2.6. In some references[5], the (direct) Fourier transform of a signal $g(t)$ is defined by

$$G_2(\omega) = \int_{-\infty}^{+\infty} g(t)e^{-j\omega t}dt \tag{5}$$

[5]MATLAB uses this definition.

In which case, we have

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} G_2(\omega)\, e^{j\omega t} d\omega = g(t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} G_2(\omega) = \int_{-\infty}^{\infty} g(t)\, e^{-j\omega t} dt \tag{6}$$

- In MATLAB, these calculations are carried out via the commands `fourier` and `ifourier`.
- Note that $\hat{G}(0) = \int g(t)dt$ and $g(0) = \frac{1}{2\pi}\int G(\omega)d\omega$.
- The relationship between $G(f)$ in (2) and $G_2(\omega)$ in (5) is given by

$$G(f) = G_2(\omega)\big|_{\omega=2\pi f} \tag{7}$$

$$G_2(\omega) = G(f)\big|_{f=\frac{\omega}{2\pi}} \tag{8}$$

2.7. Q: The relationship between $G(f)$ in (2) and $G_2(\omega)$ in (5) is given by (7) and (8) which do not involve a factor of 2π in the front. Why then does the factor of $\frac{1}{2\pi}$ shows up in (6)?

Example 2.8. Rectangular and Sinc:

$$1\left[|t| \le a\right] \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} \frac{\sin(2\pi f a)}{\pi f} = \frac{2\sin(a\omega)}{\omega} = 2a\ \operatorname{sinc}(a\omega) \tag{9}$$

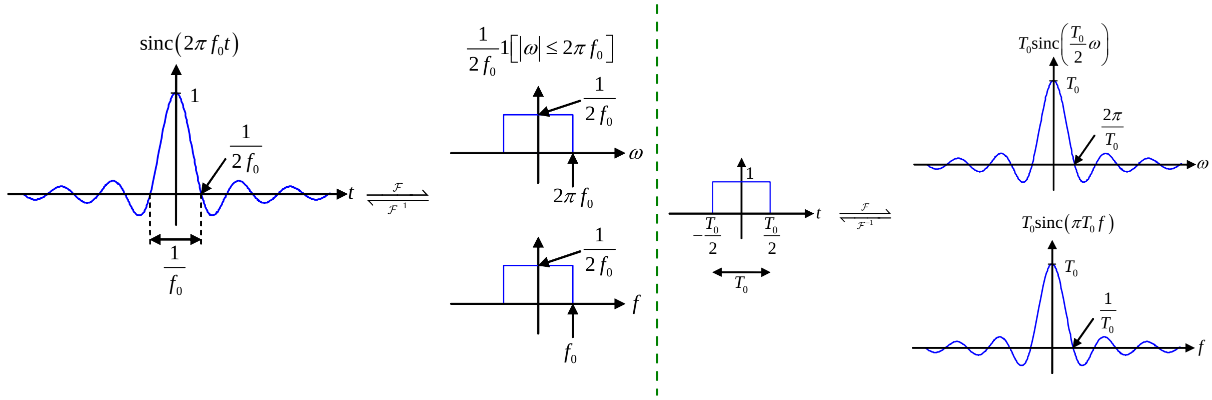


Figure 2: Fourier transform of sinc and rectangular functions

- By setting $a = T_0/2$, we have

$$1\left[|t| \leq \frac{T_0}{2}\right] \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} T_0 \operatorname{sinc}(\pi T_0 f). \tag{10}$$

- In [4, p 78], the function $1[|t| \leq 0.5]$ is defined as the **unit gate** function $\operatorname{rect}(x)$.

Definition 2.9. The function $\operatorname{sinc}(x) \equiv (\sin x)/x$ is plotted in Figure 3.

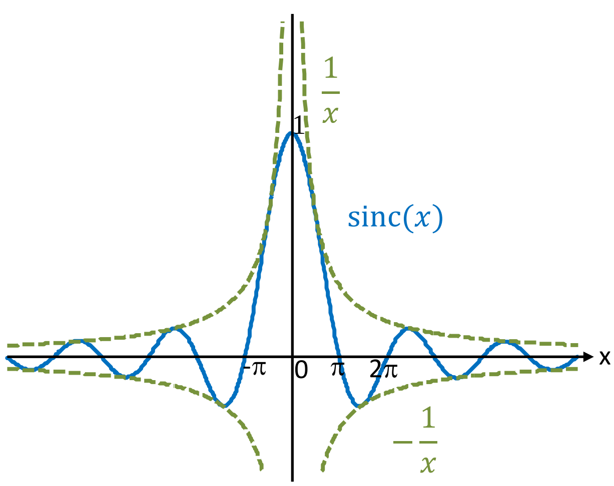


Figure 3: Sinc function

- This function plays an important role in signal processing. It is also known as the filtering or interpolating function.

- Using L'Hôpital's rule, we find $\lim_{x \to 0} \operatorname{sinc}(x) = 1$.
- $\operatorname{sinc}(x)$ is the product of an oscillating signal $\sin(x)$ (of period 2π) and a monotonically decreasing function $1/x$. Therefore, $\operatorname{sinc}(x)$ exhibits sinusoidal oscillations of period 2π, with amplitude decreasing continuously as $1/x$.
- In `MATLAB` and in [10, eq. 2.64], $\operatorname{sinc}(x)$ is defined as $(\sin(\pi x))/\pi x$. In which case, it is an even damped oscillatory function with zero crossings at integer values of its argument.

Definition 2.10. The (Dirac) **delta function** or (unit) impulse function is denoted by $\delta(t)$. It is usually depicted as a vertical arrow at the origin. Note that $\delta(t)$ is not a true function; it is undefined at $t = 0$. We define $\delta(t)$ as a generalized function which satisfies the **sampling property** (or **sifting property**)

$$\int_{-\infty}^{\infty} \phi(t)\delta(t)dt = \phi(0) \tag{11}$$

for any function $\phi(t)$ which is continuous at $t = 0$.

- In this way, the delta "function" has no mathematical or physical meaning unless it appears under the operation of integration.
- Intuitively we may visualize $\delta(t)$ as an infinitely tall, infinitely narrow rectangular pulse of unit area: $\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} 1\left[|t| \le \frac{\varepsilon}{2}\right]$.

2.11. Properties of $\delta(t)$:

- $\delta(t) = 0$ for $t \ne 0$.
  $\delta(t - T) = 0$ for $t \ne T$.
- $\int_A \delta(t)dt = 1_A(0)$.
  (a) $\int \delta(t)dt = 1$.
  (b) $\int_{\{0\}} \delta(t)dt = 1$.
  (c) $\int_{-\infty}^{x} \delta(t)dt = 1_{[0,\infty)}(x)$. Hence, we may think of $\delta(t)$ as the "derivative" of the unit step function $U(t) = 1_{[0,\infty)}(x)$.

- $\int \phi(t)\delta(t-c)dt = \phi(c)$ for ϕ continuous at T. In fact, for any $\varepsilon > 0$,

$$\int_{T-\varepsilon}^{T+\varepsilon} \phi(t)\delta(t-c)dt = \phi(c).$$

- Convolution property:

$$(\delta * \phi)(t) = (\phi * \delta)(t) = \int_{-\infty}^{\infty} \phi(\tau)\delta(t-\tau)d\tau = \phi(t) \tag{12}$$

where we assume that ϕ is continuous at t.

- $\delta(at) = \frac{1}{|a|}\delta(t)$. In particular,

$$\delta(\omega) = \frac{1}{2\pi}\delta(f) \tag{13}$$

and

$$\delta(\omega - \omega_0) = \delta(2\pi f - 2\pi f_0) = \frac{1}{2\pi}\delta(f - f_0), \tag{14}$$

where $\omega = 2\pi f$ and $\omega_0 = 2\pi f_0$.

Example 2.12. $\delta(t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} 1.$

Example 2.13. $e^{j2\pi f_0 t} \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} \delta\left(f - f_0\right).$

Example 2.14. $e^{j\omega_0 t} \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} 2\pi\delta\left(\omega - \omega_0\right).$

Example 2.15. $\cos(2\pi f_0 t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} \frac{1}{2}\left(\delta\left(f-f_0\right)+\delta\left(f+f_0\right)\right)$.

2.16. Conjugate symmetry[6]: If $x(t)$ is **real**-valued, then $X(-f) = (X(f))^*$

Observe that if we know $X(f)$ for all f positive, we also know $X(f)$ for all f negative. Interpretation: Only half of the spectrum contains all of the information. Positive-frequency part of the spectrum contains all the necessary information. The negative-frequency half of the spectrum can be determined by simply complex conjugating the positive-frequency half of the spectrum.

2.17. *Shifting* properties

- ***Time-shift***:

$$g\left(t-t_1\right) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} e^{-j2\pi f t_1} G\left(f\right)$$

    - Note that $\left|e^{-j2\pi f t_1}\right| = 1$. So, the spectrum of $g\left(t-t_1\right)$ looks exactly the same as the spectrum of $g(t)$ (unless you also look at their phases).

- ***Frequency-shift*** (or modulation):

$$e^{j2\pi f_1 t} g\left(t\right) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} G\left(f-f_1\right)$$

[6]Hermitian symmetry in [7, p 17].

2.18. Let $g(t)$, $g_1(t)$, and $g_2(t)$ denote signals with $G(f)$, $G_1(f)$, and $G_2(f)$ denoting their respective Fourier transforms.

(a) ***Superposition theorem*** (linearity):

$$a_1 g_1(t) + a_2 g_2(t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} a_1 G_1(f) + a_2 G_2(f).$$

(b) ***Scale-change*** theorem (scaling property [4, p 88]):

$$g(at) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} \frac{1}{|a|} G\left(\frac{f}{a}\right).$$

- The function $g(at)$ represents the function $g(t)$ *compress*ed in time by a factor a (when $|a| > l$). Similarly, the function $G(f/a)$ represents the function $G(f)$ *expand*ed in frequency by the same factor a.
- The scaling property says that if we "squeeze" a function in t, its Fourier transform "stretches out" in f. It is not possible to arbitrarily concentrate both a function and its Fourier transform.
- Generally speaking, the more concentrated $g(t)$ is, the more spread out its Fourier transform $G(f)$ must be.
- This trade-off can be formalized in the form of an *uncertainty principle*. See also 2.28 and 2.29.
- Intuitively, we understand that compression in time by a factor a means that the signal is varying more rapidly by the same factor. To synthesize such a signal, the frequencies of its sinusoidal components must be increased by the factor a, implying that its frequency spectrum is expanded by the factor a. Similarly, a signal expanded in time varies more slowly; hence, the frequencies of its components are lowered, implying that its frequency spectrum is compressed.

(c) ***Duality theorem*** (Symmetry Property [4, p 86]):

$$G(t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} g(-f).$$

- In words, for any result or relationship between $g(t)$ and $G(f)$, there exists a dual result or relationship, obtained by interchanging the roles of $g(t)$ and $G(f)$ in the original result (along with some minor modifications arising because of a sign change).

  In particular, if the Fourier transform of $g(t)$ is $G(f)$, then the Fourier transform of $G(f)$ with f replaced by t is the original time-domain signal with t replaced by $-f$.
- If we use the ω-definition (5), we get a similar relationship with an extra factor of 2π:

$$G_2(t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} 2\pi g(-\omega).$$

Example 2.19. $x(t) = \cos(2\pi a f_0 t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} \frac{1}{2}\left(\delta(f - af_0) + \delta(f + af_0)\right).$

Example 2.20. From Example 2.8, we know that

$$1\left[|t| \le a\right] \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} 2a \ \operatorname{sinc}\left(2\pi a f\right) \tag{15}$$

By the duality theorem, we have

$$2a \operatorname{sinc}(2\pi a t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} 1[|-f| \le a],$$

which is the same as

$$\operatorname{sinc}(2\pi f_0 t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} \frac{1}{2f_0} 1[|f| \le f_0]. \tag{16}$$

Both transform pairs are illustrated in Figure 2.

Example 2.21. Let's try to derive the time-shift property from the frequency-shift property. We start with an arbitrary function $g(t)$. Next we will define another function $x(t)$ by setting $X(f)$ to be $g(f)$. Note that f here is just a dummy variable; we can also write $X(t) = g(t)$. Applying the duality theorem to the transform pair $x(t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} X(f)$, we get another transform pair $X(t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} x(-f)$. The LHS is $g(t)$; therefore, the RHS must be $G(f)$. This implies $G(f) = x(-f)$. Next, recall the frequency-shift property:

$$e^{j2\pi ct}x\left(t\right) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} X\left(f-c\right).$$

The duality theorem then gives

$$X\left(t-c\right) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} e^{j2\pi c-f}x\left(-f\right).$$

Replacing $X(t)$ by $g(t)$ and $x(-f)$ by $G(f)$, we finally get the time-shift property.

Definition 2.22. The **convolution** of two signals, $x_1(t)$ and $x_2(t)$, is a new function of time, $x(t)$. We write

$$x = x_1 * x_2.$$

It is defined as the integral of the product of the two functions after one is reversed and shifted:

$$x(t) = (x_1 * x_2)(t) \tag{17}$$

$$= \int_{-\infty}^{+\infty} x_1(\mu)x_2(t-\mu)d\mu = \int_{-\infty}^{+\infty} x_1(t-\mu)x_2(\mu)d\mu. \tag{18}$$

- Note that t is a parameter as far as the integration is concerned.
- The integrand is formed from x_1 and x_2 by three operations:
  (a) time reversal to obtain $x_2(-\mu)$,
  (b) time shifting to obtain $x_2(-(\mu - t)) = x_2(t-\mu)$, and
  (c) multiplication of $x_1(\mu)$ and $x_2(t-\mu)$ to form the integrand.
- In some references, (17) is expressed as $x(t) = x_1(t) * x_2(t)$.

Example 2.23. We can get a triangle from convolution of two rectangular waves. In particular,

$$1[|t| \leq a] * 1[|t| \leq a] = (2a - |t|) \times 1[|t| \leq 2a].$$

2.24. Convolution theorem:

(a) Convolution-in-time rule:

$$x_1 * x_2 \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} X_1 \times X_2. \tag{19}$$

(b) Convolution-in-frequency rule:

$$x_1 \times x_2 \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} X_1 * X_2. \tag{20}$$

Example 2.25. We can use the convolution theorem to "prove" the frequency-sift property in 2.17.

2.26. From the convolution theorem, we have

- $g^2 \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} G * G$
- if g is band-limited to B, then g^2 is band-limited to $2B$

2.27. Parseval's theorem (Rayleigh's energy theorem, Plancherel formula) for Fourier transform:

$$\int_{-\infty}^{+\infty} |g(t)|^2 dt = \int_{-\infty}^{+\infty} |G(f)|^2 df. \tag{21}$$

The LHS of (21) is called the (total) **energy** of $g(t)$. On the RHS, $|G(f)|^2$ is called the energy spectral density of $g(t)$. By integrating the energy spectral density over all frequency, we obtain the signal 's total energy. The energy contained in the frequency band B can be found from the integral $\int_B |G(f)|^2 df$.

More generally, Fourier transform preserves the inner product [2, Theorem 2.12]:

$$\langle g_1, g_2 \rangle = \int_{-\infty}^{\infty} g_1(t) g_2^*(t) dt = \int_{-\infty}^{\infty} G_1(f) G_2^*(f) df = \langle G_1, G_2 \rangle.$$

2.28. (Heisenberg) **Uncertainty Principle** [2, 9]: Suppose g is a function which satisfies the normalizing condition $\|g\|_2^2 = \int |g(t)|^2 dt = 1$ which automatically implies that $\|G\|_2^2 = \int |G(f)|^2 df = 1$. Then

$$\left(\int t^2 |g(t)|^2 dt \right) \left(\int f^2 |G(f)|^2 df \right) \geq \frac{1}{16\pi^2}, \tag{22}$$

and equality holds if and only if $g(t) = Ae^{-Bt^2}$ where $B > 0$ and $|A|^2 = \sqrt{2B/\pi}$.

- In fact, we have

  $$\left(\int t^2 |g(t - t_0)|^2 dt \right) \left(\int f^2 |G(f - f_0)|^2 df \right) \geq \frac{1}{16\pi^2},$$

  for every t_0, f_0.

- The proof relies on Cauchy-Schwarz inequality.

- For any function h, define its dispersion Δ_h as $\frac{\int t^2 |h(t)|^2 dt}{\int |h(t)|^2 dt}$. Then, we can apply (22) to the function $g(t) = h(t)/\|h\|_2$ and get

  $$\Delta_h \times \Delta_H \geq \frac{1}{16\pi^2}.$$

2.29. A signal cannot be simultaneously time-limited and band-limited.

Proof. Suppose $g(t)$ is simultaneously (1) time-limited to T_0 and (2) band-limited to B. Pick any positive number T_s and positive integer K such that $f_s = \frac{1}{T_s} > 2B$ and $K > \frac{T_0}{T_s}$. The sampled signal $g_{T_s}(t)$ is given by

$$g_{T_s}(t) = \sum_k g[k]\delta(t - kT_s) = \sum_{k=-K}^{K} g[k]\delta(t - kT_s)$$

where $g[k] = g(kT_s)$. Now, because we sample the signal faster than the Nyquist rate, we can reconstruct the signal g by producing $g_{T_s} * h_r$ where the LPF h_r is given by

$$H_r(\omega) = T_s 1[\omega < 2\pi f_c]$$

with the restriction that $B < f_c < \frac{1}{T_s} - B$. In frequency domain, we have

$$G(\omega) = \sum_{k=-K}^{K} g[k]e^{-jk\omega T_s} H_r(\omega).$$

Consider ω inside the interval $I = (2\pi B, 2\pi f_c)$. Then,

$$0 \overset{\omega>2\pi B}{=} G(\omega) \overset{\omega<2\pi f_c}{=} T_s \sum_{k=-K}^{K} g(kT_s) e^{-jk\omega T_s} \overset{z=e^{j\omega T_s}}{=} T_s \sum_{k=-K}^{K} g(kT_s) z^{-k} \tag{23}$$

Because $z \neq 0$, we can divide (23) by z^{-K} and then the last term becomes a polynomial of the form

$$a_{2K} z^{2K} + a_{2K-1} z^{2K-1} + \cdots + a_1 z + a_0.$$

By fundamental theorem of algebra, this polynomial has only finitely many roots– that is there are only finitely many values of $z = e^{j\omega T_s}$ which satisfies (23). Because there are uncountably many values of ω in the interval I and hence uncountably many values of $z = e^{j\omega T_s}$ which satisfy (23), we have a contradiction. □

2.30. The observation in 2.29 raises concerns about the signal and filter models used in the study of communication systems. Since a signal cannot be both bandlimited and timelimited, we should either abandon bandlimited signals (and ideal filters) or else accept signal models that exist for all time. On the one hand, we recognize that any real signal is timelimited, having starting and ending times. On the other hand, the concepts of bandlimited spectra and ideal filters are too useful and appealing to be dismissed entirely.

The resolution of our dilemma is really not so difficult, requiring but a small compromise. Although a strictly timelimited signal is not strictly bandlimited, its spectrum may be negligibly small above some upper frequency limit B. Likewise, a strictly bandlimited signal may be negligibly small outside a certain time interval $t_1 \leq t \leq t_2$. Therefore, we will often assume that signals are essentially both bandlimited and timelimited for most practical purposes.

ECS332 2012/1 Part II.1 Dr.Prapun

3 Modulation and Frequency Shifting

Definition 3.1. The term **baseband** is used to designate the band of frequencies of the signal delivered by the source.

Example 3.2. In telephony, the baseband is the audio band (band of voice signals) of 0 to 3.5 kHz.

Example 3.3. For digital data (sequence of two voltage levels representing 0 and 1) at a rate of R bits per second, the baseband is 0 to R Hz.

Definition 3.4. **Modulation** is a process that causes a shift in the range of frequencies in a signal.

- The modulation process commonly translates an information-bearing signal to a new spectral location depending upon the intended frequency for transmission.

Definition 3.5. In **baseband communication**, baseband signals are transmitted without modulation, that is, without any shift in the range of frequencies of the signal.

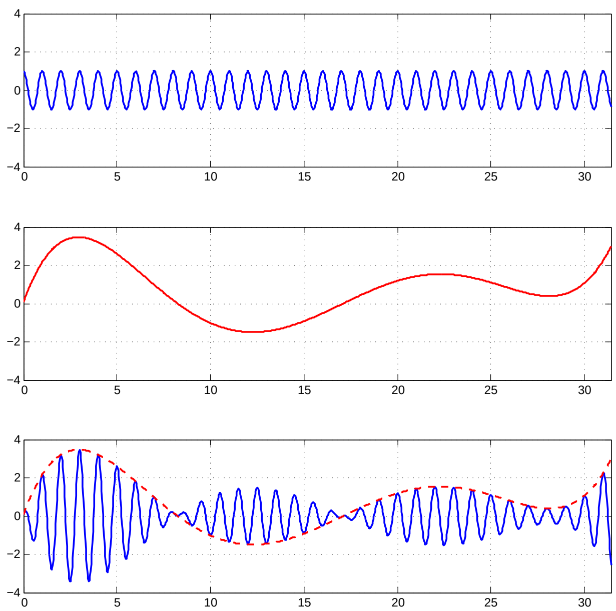
3.6. Recall the frequency-shift property:

$$e^{j2\pi f_c t} g(t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} G(f - f_c).$$

This property states that multiplication of a signal by a factor $e^{j2\pi f_c t}$ shifts the spectrum of that signal by $\Delta f = f_c$.

3.7. Frequency-shifting (frequency translation) in practice is achieved by multiplying $g(t)$ by a sinusoid:

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



Similarly,

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

Definition 3.8. $\cos(2\pi f_c t + \phi)$ is called the (sinusoidal) **carrier signal** and f_c is called the **carrier frequency**. In general, it can also has amplitude A and hence the general expression of the carrier signal is $A\cos(2\pi f_c t + \phi)$.

3.9. Examples of situations where modulation (spectrum shifting) is useful:

(a) **Channel passband matching**: Recall that, for a linear, time-invariant (LTI) system, the input-output relationship is given by

$$y(t) = h(t) * x(t)$$

where $x(t)$ is the input, $y(t)$ is the output, and $h(t)$ is the **impulse response** of the system. In which case,

$$Y(f) = H(f)X(f)$$

where $H(f)$ is called the **transfer function** or **frequency response** of the system. $|H(f)|$ and $\angle H(f)$ are called the **amplitude response** and **phase response**, respectively. Their plots as functions of f show at a glance how the system modifies the amplitudes and phases of various sinusoidal inputs.

(b) **Reasonable antenna size**: For effective radiation of power over a radio link, the antenna size must be on the order of the wavelength of the signal to be radiated.

- Audio signal frequencies are so low (wavelengths are so large) that impracticably large antennas will be required for radiation. Here, shifting the spectrum to a higher frequency (a smaller wavelength) by modulation solves the problem.

(c) **Frequency-Division Multiplexing** (FDM) and Frequency-Division Multiple Access (FDMA):

- If several signals, each occupying the same frequency band, are transmitted simultaneously over the same transmission medium, they will all interfere; it will be difficult to separate or retrieve them at a receiver.
- For example, if all radio stations decide to broadcast audio signals simultaneously, the receiver will not be able to separate them.
- One solution is to use modulation whereby each radio station is assigned a distinct carrier frequency. Each station transmits a modulated signal, thus shifting the signal spectrum to its allocated band, which is not occupied by any other station. A radio receiver can pick up any station by tuning to the band of the desired station.

Definition 3.10. Communication that uses modulation to shift the frequency spectrum of a signal is known as **carrier communication**. [4, p 151]

3.11. 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. Collectively, these techniques are called **continuous-wave modulation** in [10, p 111].

We will use $m(t)$ to denote the baseband signal. We will assume that $m(t)$ is band-limited to B; that is, $|M(f)| = 0$ for $|f| > B$. Note that we usually call it the **message** or the **modulating signal**.

Definition 3.12. The process of recovering the signal from the modulated signal (retranslating the spectrum to its original position) is referred to as **demodulation**, or **detection**.

4 Amplitude modulation: DSB-SC

Definition 4.1. Amplitude modulation is characterized by the fact that the amplitude A of the carrier $A\cos(2\pi f_c t + \phi)$ is varied in proportion to the baseband (message) signal $m(t)$.

- Because the amplitude is time-varying, we may write the modulated carrier as
$$A(t)\cos(2\pi f_c t + \phi)$$
- Because the amplitude is linearly related to the message signal, this technique is also called **linear modulation**.

4.1 Double-sideband suppressed carrier (DSB-SC) modulation

4.2. Basic idea:

$$\text{LPF}\left\{\underbrace{\left(m\left(t\right)\times\sqrt{2}\cos\left(2\pi f_c t\right)\right)}_{x(t)}\times\left(\sqrt{2}\cos\left(2\pi f_c t\right)\right)\right\} = m\left(t\right). \qquad (24)$$

$$\begin{aligned}
x\left(t\right) &= m\left(t\right)\times\sqrt{2}\cos\left(2\pi f_c t\right) = \sqrt{2}m\left(t\right)\cos\left(2\pi f_c t\right)\\
X\left(f\right) &= \sqrt{2}\left(\frac{1}{2}\left(M\left(f-f_c\right)+M\left(f+f_c\right)\right)\right)\\
&= \frac{1}{\sqrt{2}}\left(M\left(f-f_c\right)+M\left(f+f_c\right)\right)
\end{aligned}$$

Similarly,

$$\begin{aligned}
v\left(t\right) &= y\left(t\right)\times\sqrt{2}\cos\left(2\pi f_c t\right) = \sqrt{2}x\left(t\right)\cos\left(2\pi f_c t\right)\\
V\left(f\right) &= \frac{1}{\sqrt{2}}\left(X\left(f-f_c\right)+X\left(f+f_c\right)\right)
\end{aligned}$$

Alternatively, we can use the trig. identity from Example 2.3:

$$\begin{aligned} v(t) &= \sqrt{2}x(t)\cos(2\pi f_c t) = \sqrt{2}\left(\sqrt{2}m(t)\cos(2\pi f_c t)\right)\cos(2\pi f_c t) \\ &= 2m(t)\cos^2(2\pi f_c t) = m(t)\left(\cos(2(2\pi f_c t)) + 1\right) \\ &= m(t) + m(t)\cos(2\pi(2f_c)t) \end{aligned}$$

4.3. In the process of modulation, observe that we need $f_c > B$ in order to avoid overlap of the spectra.

4.4. Observe that the modulated signal spectrum centered at f_c, is composed of two parts: a portion that lies above f_c, known as the **upper sideband** (USB), and a portion that lies below f_c, known as the **lower sideband** (LSB). Similarly, the spectrum centered at $-f_c$ has upper and lower sidebands. Hence, this is a modulation scheme with **double sidebands**.

4.5. Observe that (24) requires that we can generate $\cos(\omega_c t)$ both at the transmitter and receiver. This can be difficult in practice. Suppose that the frequency at the receiver is off, say by Δf, and that the phase is off by θ. The effect of these frequency and phase offsets can be quantified by calculating

$$\text{LPF}\left\{\left(m(t)\sqrt{2}\cos\omega_c t\right)\sqrt{2}\cos((\omega_c + \Delta\omega)t + \theta)\right\},$$

which gives

$$m(t)\cos((\Delta\omega)t + \theta).$$

Of course, we want $\Delta\omega = 0$ and $\theta = 0$; that is the receiver must generate a carrier in phase and frequency synchronism with the incoming carrier. These demodulators are called ***synchronous*** or ***coherent*** (also ***homodyne***) demodulator [4, p 161].

4.6. Effect of **time delay**: Suppose the propagation time is τ, then we have

$$\begin{aligned} y(t) = x(t-\tau) &= \sqrt{2}m(t-\tau)\cos(2\pi f_c(t-\tau)) \\ &= \sqrt{2}m(t-\tau)\cos(2\pi f_c t - 2\pi f_c \tau) \\ &= \sqrt{2}m(t-\tau)\cos(2\pi f_c t - \phi_\tau). \end{aligned}$$

Consequently,

$$\begin{aligned} v(t) &= y(t) \times \sqrt{2}\cos(2\pi f_c t) \\ &= \sqrt{2}m(t-\tau)\cos(2\pi f_c t - \phi_\tau) \times \sqrt{2}\cos(2\pi f_c t) \\ &= m(t-\tau)\, 2\cos(2\pi f_c t - \phi_\tau)\cos(2\pi f_c t). \end{aligned}$$

Applying the product-to-sum formula, we then have

$$v(t) = m(t-\tau)\left(\cos(2\pi(2f_c)t - \phi_\tau) + \cos(\phi_\tau)\right).$$

4.2 Fourier Series

Let the (real or complex) signal $r(t)$ be a *periodic* signal with period T_0. Suppose the following ***Dirichlet*** conditions are satisfied

(a) $r(t)$ is absolutely integrable over its period; i.e., $\int_0^{T_0} |r(t)| dt < \infty$.

(b) The number of maxima and minima of $r(t)$ in each period is finite.

(c) The number of discontinuities of $r(t)$ in each period is finite.

Then $r(t)$ can be expanded in terms of the complex exponential signals $\left(e^{jn\omega_0 t}\right)_{n=-\infty}^{\infty}$ as

$$\tilde{r}(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t} = c_0 + \sum_{k=1}^{\infty}\left(c_k e^{jk\omega_0 t} + c_{-k} e^{-jk\omega_0 t}\right) \tag{25}$$

where

$$\omega_0 = 2\pi f_0 = \frac{2\pi}{T_0},$$

$$c_k = \frac{1}{T_0}\int_{\alpha}^{\alpha+T_0} r(t)\, e^{-jk\omega_0 t} dt, \tag{26}$$

for some *arbitrary* α. In which case,

$$\tilde{r}(t) = \begin{cases} r(t), & \text{if } r(t) \text{ is continuous at } t \\ \frac{r(t^+)+r(t^-)}{2}, & \text{if } r(t) \text{ is not continuous at } t \end{cases}$$

We give some remarks here.

- The parameter α in the limits of the integration (26) is arbitrary. It can be chosen to simplify computation of the integral. Some references simply write $c_k = \frac{1}{T_0} \int\limits_{T_0} r(t) e^{-jk\omega_0 t} dt$ to emphasize that we only need to integrate over one period of the signal; the starting point is not important.
- The coefficients $c_k = \frac{1}{T_0} \int\limits_{T_0} r(t) e^{-jk\omega_0 t} dt$ are called the (k^{th}) ***Fourier (series) coefficients*** of (the signal) $r(t)$. These are, in general, complex numbers.
- $c_0 = \frac{1}{T_0} \int\limits_{T_0} r(t)\, dt$ = average or DC value of $r(t)$
- The quantity $f_0 = \frac{1}{T_0}$ is called the ***fundamental frequency*** of the signal $r(t)$. The nth multiple of the fundamental frequency (for positive n's) is called the nth ***harmonic***.
- $c_k e^{jk\omega_0 t} + c_{-k} e^{-jk\omega_0 t}$ = the k^{th} ***harmonic component*** of $r(t)$.
  $k = 1 \Rightarrow$ ***fundamental component*** of $r(t)$.

4.7. Consider a restricted version $r_{T_0}(t)$ of $r(t)$ where we only consider $r(t)$ for one specific period. Suppose $r_{T_0}(t) \underset{\mathcal{F}^{-1}}{\overset{\mathcal{F}}{\rightleftharpoons}} R_{T_0}(f)$. Then,

$$c_k = \frac{1}{T_0} R_{T_0}(kf_0).$$

So, the Fourier coefficients are simply scaled samples of the Fourier transform.

4.8. Parseval's Identity: $P_r = \frac{1}{T_0} \int\limits_{T_0} |r(t)|^2 dt = \sum\limits_{k=-\infty}^{\infty} |c_k|^2$

4.3 Fourier series expansion for real valued function

4.9. Suppose $r(t)$ in the previous section is real-valued; that is $r^* = r$. Then, we have $c_{-k} = c_k^*$ and we provide here three alternative ways to represent the Fourier series expansion:

$$\tilde{r}(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t} = c_0 + \sum_{k=1}^{\infty} \left(c_k e^{jk\omega_0 t} + c_{-k} e^{-jk\omega_0 t} \right) \tag{27}$$

$$= c_0 + \sum_{k=1}^{\infty} \left(a_k \cos(k\omega_0 t) \right) + \sum_{k=1}^{\infty} \left(b_k \sin(k\omega_0 t) \right) \tag{28}$$

$$= c_0 + 2\sum_{k=1}^{\infty} |c_k| \cos(k\omega_0 t + \angle c_k) \tag{29}$$

where the corresponding coefficients are obtained from

$$c_k = \frac{1}{T_0} \int_{\alpha}^{\alpha+T_0} r(t) e^{-jk\omega_0 t} dt = \frac{1}{2}(a_k - jb_k) \tag{30}$$

$$a_k = 2\mathrm{Re}\{c_k\} = \frac{2}{T_0} \int_{T_0} r(t) \cos(k\omega_0 t)\, dt \tag{31}$$

$$b_k = -2\mathrm{Im}\{c_k\} = \frac{2}{T_0} \int_{T_0} r(t) \sin(k\omega_0 t)\, dt \tag{32}$$

$$2|c_k| = \sqrt{a_k^2 + b_k^2} \tag{33}$$

$$\angle c_k = -\arctan\left(\frac{b_k}{a_k}\right) \tag{34}$$

$$c_0 = \frac{a_0}{2} \tag{35}$$

The Parseval's identity can then be expressed as

$$P_r = \frac{1}{T_0} \int_{T_0} |r(t)|^2 dt = \sum_{k=-\infty}^{\infty} |c_k|^2 = c_0^2 + 2\sum_{k=1}^{\infty} |c_k|^2$$

4.10. To go from (27) to (28) and (29), note that when we replace c_{-k} by c_k^*, we have

$$\begin{aligned} c_k e^{jk\omega_0 t} + c_{-k} e^{-jk\omega_0 t} &= c_k e^{jk\omega_0 t} + c_k^* e^{-jk\omega_0 t} \\ &= c_k e^{jk\omega_0 t} + \left(c_k e^{jk\omega_0 t}\right)^* \\ &= 2\operatorname{Re}\left\{c_k e^{jk\omega_0 t}\right\}. \end{aligned}$$

- Expression (29) then follows directly from the phasor concept:

$$\operatorname{Re}\left\{c_k e^{jk\omega_0 t}\right\} = |c_k| \cos\left(k\omega_0 t + \angle c_k\right).$$

- To get (28), substitute c_k by $\operatorname{Re}\{c_k\} + j\operatorname{Im}\{c_k\}$ and $e^{jk\omega_0 t}$ by $\cos(k\omega_0 t) + j\sin(k\omega_0 t)$.

Example 4.11. Train of impulses:

$$\delta^{(T_0)}(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT_0) = \frac{1}{T_0} \sum_{k=-\infty}^{\infty} e^{jk\omega_0 t} = \frac{1}{T_0} + \frac{2}{T_0} \sum_{k=1}^{\infty} \cos k\omega_0 t \quad (36)$$

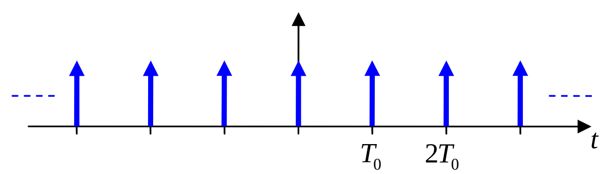


Figure 4: Train of impulses

Example 4.12. Square pulse periodic signal:

$$1[\cos\omega_0 t \ge 0] = \frac{1}{2} + \frac{2}{\pi}\left(\cos\omega_0 t - \frac{1}{3}\cos 3\omega_0 t + \frac{1}{5}\cos 5\omega_0 t - \frac{1}{7}\cos 7\omega_0 t + \dots\right) \quad (37)$$

We note here that multiplication by this signal is a switching function.

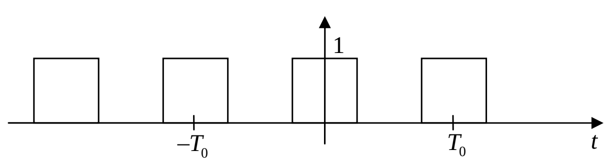


Figure 5: Square pulse periodic signal

Example 4.13. Bipolar square pulse periodic signal:

$$\operatorname{sgn}(\cos\omega_0 t) = \frac{4}{\pi}\left(\cos\omega_0 t - \frac{1}{3}\cos 3\omega_0 t + \frac{1}{5}\cos 5\omega_0 t - \frac{1}{7}\cos 7\omega_0 t + \dots\right)$$

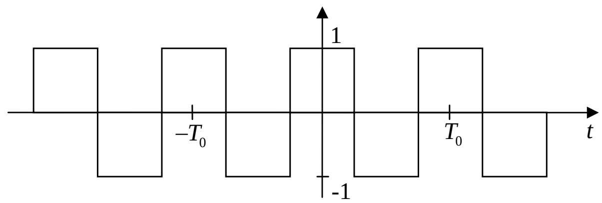


Figure 6: Bipolar square pulse periodic signal

4.4 Producing the modulated signal

To produce the modulated signal $m(t)\cos(2\pi f_c t)$, we may use the following methods which generate the modulated signal along with other signals which can be eliminated by a bandpass filter restricting frequency contents to around ω_c.

4.14. Multiplier Modulators: Here modulation is achieved directly by multiplying $m(t)$ by $\cos(2\pi f_c t)$ using an analog multiplier whose output is proportional to the product of two input signals.

- Such a multiplier may be obtained from a variable-gain amplifier in which the gain parameter (such as the the β of a transistor) is controlled by one of the signals, say, $m(t)$. When the signal $cos(2\pi f_c t)$ is applied at the input of this amplifier, the output is then proportional to $m(t)\cos(2\pi f_c t)$.
- Another way to multiply two signals is through logarithmic amplifiers. Here, the basic components are a logarithmic and an antilogarithmic amplifier with outputs proportional to the log and antilog of their inputs, respectively. Using two logarithmic amplifiers, we generate and add the logarithms of the two signals to be multiplied. The sum is then applied to an antilogarithmic amplifier to obtain the desired product.
- Difficult to maintain linearity in this kind of amplifier.

- Expensive.

4.15. Square Modulator: When it is easier to build a squarer than a multiplier, use

$$\begin{aligned}(m(t) + c\cos(\omega_c t))^2 &= m^2(t) + 2c\,m(t)\cos(\omega_c t) + c^2\cos^2(\omega_c t)\\ &= m^2(t) + +2c\,m(t)\cos(\omega_c t) + \frac{c^2}{2} + \frac{c^2}{2}\cos(2\omega_c t).\end{aligned}$$

- Alternative, can use $\left(m(t) + c\cos\left(\frac{\omega_c}{2}t\right)\right)^3$.

4.16. Multiply $m(t)$ by "any" periodic and even signal $r(t)$ whose period is $T_c = \frac{2\pi}{\omega_c}$. Because $r(t)$ is an even function, we know that

$$r(t) = c_0 + \sum_{k=1}^{\infty} a_k \cos(k\omega_c t).$$

Therefore,

$$m(t)r(t) = c_0 m(t) + \sum_{k=1}^{\infty} a_k m(t)\cos(k\omega_c t).$$

See also [4, p 157]. In general, for this scheme to work, we need

- $a_1 \neq 0$; that is T_c is the "least" period of r;
- $\omega_c > 4\pi B$; that is $f_c > 2B$ (to prevent overlapping).

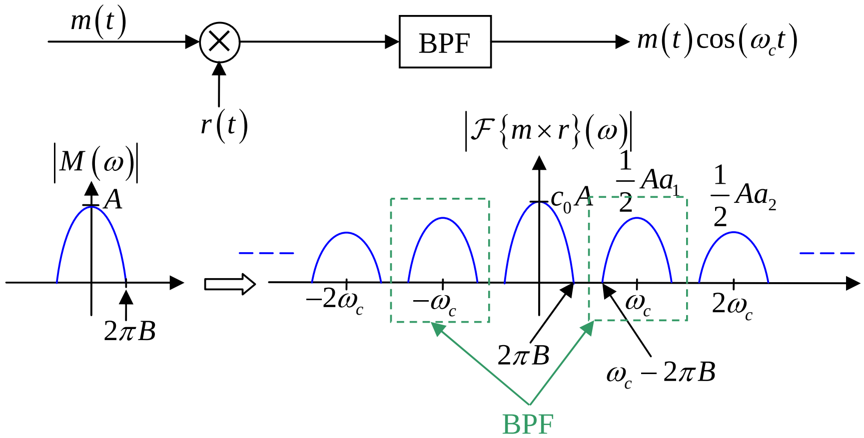


Figure 7: Modulation of $m(t)$ via even and periodic $r(t)$

Note that if $r(t)$ is not even, then by (29), the outputted modulated signal is of the form $a_1 m(t)\cos(\omega_c t + \phi_1)$.

4.17. ***Switching modulator***: Set $r(t)$ to be the square pulse train given by (37):

$$\begin{aligned} r(t) &= 1\left[\cos\omega_0 t \geq 0\right] \\ &= \frac{1}{2} + \frac{2}{\pi}\left(\cos\omega_0 t - \frac{1}{3}\cos 3\omega_0 t + \frac{1}{5}\cos 5\omega_0 t - \frac{1}{7}\cos 7\omega_0 t + \dots\right). \end{aligned}$$

Multiplying this $r(t)$ to the signal $m(t)$ is equivalent to switching $m(t)$ on and off periodically.

It is equivalent to periodically turning the switch on (letting $m(t)$ pass through) for half a period $T_c = \frac{1}{f_c}$.

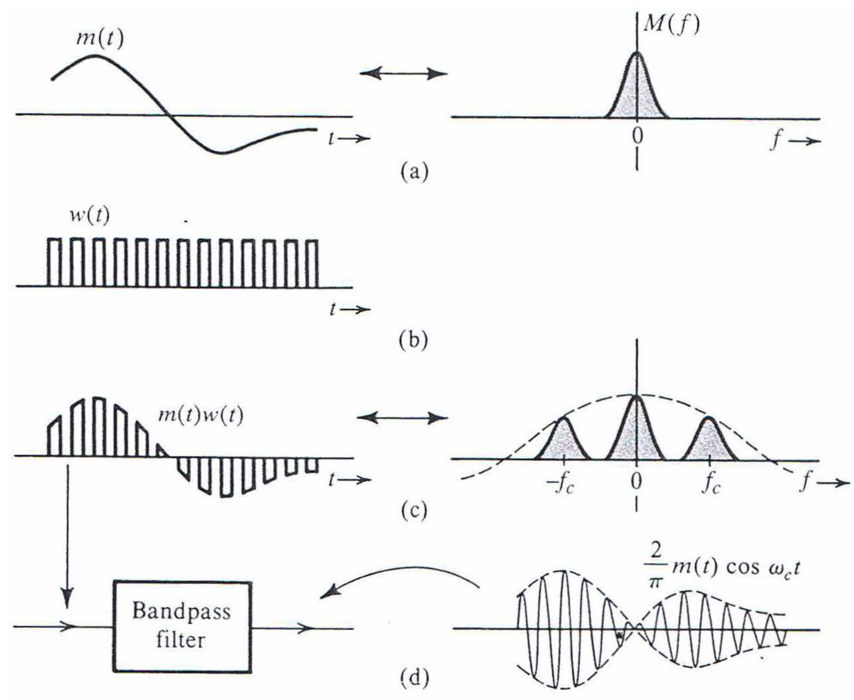


Figure 8: Switching modulator for DSB-SC [4, Figure 4.4].

4.18. ***Switching Demodulator***:

$$\text{LPF}\{m(t)\cos(\omega_c t) \times 1[\cos(\omega_c t) \geq 0]\} = \frac{1}{\pi}m(t) \tag{38}$$

[4, p 162]. Note that this technique still requires the switching to be in sync with the incoming cosine as in the basic DSB-SC.

ECS332 2012/1 Part II.2 Dr.Prapun

5 Quadrature Amplitude Modulation (QAM)

Definition 5.1. One of the possible definition for the ***bandwidth*** (***BW***) of a signal is the difference between the highest frequency and the lowest frequency in the positive-f part of the signal spectrum.

Example 5.2.

5.3. Rough Approximation: If $g_1(t)$ and $g_2(t)$ have bandwidths B_1 and B_2 Hz, respectively, the bandwidth of $g_1(t)g_2(t)$ is $B_1 + B_2$ Hz.

This result follows from the application of the width property[7] of convolution[8] to the convolution-in-frequency property.

Consequently, if the bandwidth of $g(t)$ is B Hz, then the bandwidth of $g^2(t)$ is $2B$ Hz, and the bandwidth of $g^n(t)$ is nB Hz. We mentioned this property in 2.26.

5.4. BW Inefficiency in DSB-SC: Recall that for real-valued baseband signal $m(t)$, the conjugate symmetry property from 2.16 says that

$$M(-f) = (M(f))^*.$$

[7]This property states that the width of $x * y$ is the sum of the widths of x and y.

[8]The width property of convolution does not hold in some pathological cases. See [4, p 98].

The DSB spectrum has two sidebands: the upper sideband (USB) and the lower sideband (LSB), both containing complete information about the baseband signal $m(t)$. As a result, DSB signals occupy twice the bandwidth required for the baseband. To improve the spectral efficiency of amplitude modulation, there exist two basic schemes to either utilize or remove the spectral redundancy:

(a) Single-sideband (SSB) modulation, which removes either the LSB or the USB so that for one message signal $m(t)$, there is only a bandwidth of B Hz.

(b) Quadrature amplitude modulation (QAM), which utilizes spectral redundancy by sending two messages over the same bandwidth of $2B$ Hz.

We will only discussed QAM here. SSB discussion can be found in [3, Sec 4.4], [10, Section 3.1.3] and [4, Section 4.5].

Definition 5.5. In ***quadrature amplitude modulation*** (***QAM***) or ***quadrature multiplexing***, two baseband real-valued signals $m_1(t)$ and $m_2(t)$ are transmitted simultaneously via the corresponding QAM signal:

$$x_{\text{QAM}}(t) = m_1(t)\sqrt{2}\cos(\omega_c t) + m_2(t)\sqrt{2}\sin(\omega_c t).$$

- QAM operates by transmitting two DSB signals via carriers of the same frequency but in phase quadrature.
- QAM can be exactly generated without requiring sharp cutoff bandpass filters.

- Both modulated signals simultaneously occupy the same frequency band.
- The upper channel is also known as the ***in-phase*** (***I***) channel and the lower channel is the ***quadrature*** (***Q***) channel.

5.6. ***Demodulation***: The two baseband signals can be separated at the receiver by synchronous detection:

$$\mathrm{LPF}\left\{x_{\mathrm{QAM}}(t)\sqrt{2}\cos(\omega_c t)\right\} = m_1(t)$$
$$\mathrm{LPF}\left\{x_{\mathrm{QAM}}(t)\sqrt{2}\sin(\omega_c t)\right\} = m_2(t)$$

- $m_1(t)$ and $m_2(t)$ can be separately demodulated.

5.7. Sinusoidal form:

$$x_{\mathrm{QAM}}(t) = \sqrt{2}E(t)\cos(2\pi f_c t + \theta(t)),$$

where

$$E(t) = \sqrt{m_1^2(t) + m_2^2(t)}$$
$$\theta(t) = -\tan^{-1}\left(\frac{m_2(t)}{m_1(t)}\right)$$

5.8. Complex form:

$$x_{\mathrm{QAM}}(t) = \sqrt{2}\mathrm{Re}\left\{(m(t))\,e^{j2\pi f_c t}\right\}$$

where $m(t) = m_1(t) - jm_2(t)$.

- If we use $-\sin(\omega_c t)$ instead of $\sin(\omega_c t)$,

$$x_{\text{QAM}}(t) = m_1(t)\sqrt{2}\cos(\omega_c t) - m_2(t)\sqrt{2}\sin(\omega_c t)$$

and

$$m(t) = m_1(t) + jm_2(t).$$

- We refer to $m(t)$ as the ***complex envelope*** (or ***complex baseband signal***) and the signals $m_1(t)$ and $m_2(t)$ are known as the ***in-phase*** and ***quadrature(-phase)*** components of $x_{\text{QAM}}(t)$.
- The term "quadrature component" refers to the fact that it is in phase quadrature ($\pi/2$ out of phase) with respect to the in-phase component.
- Key equation:

$$\text{LPF}\left\{\underbrace{\left(\operatorname{Re}\left\{m(t)\times\sqrt{2}e^{j2\pi f_c t}\right\}\right)}_{x(t)}\times\left(\sqrt{2}e^{-j2\pi f_c t}\right)\right\} = m(t).$$

5.9. Three equivalent ways of saying exactly the same thing:

(a) the complex-valued envelope $m(t)$ complex-modulates the complex carrier $e^{j2\pi f_c t}$,

(b) the real-valued amplitude $E(t)$ and phase $\theta(t)$ real-modulate the amplitude and phase of the real carrier $\cos(\omega_c t)$,

(c) the in-phase signal $m_1(t)$ and quadrature signal $m_2(t)$ real-modulate the real in-phase carrier $\cos(\omega_c t)$ and the real quadrature carrier $\sin(\omega_c t)$.

5.10. References: [3, p 164–166], [10, Sect. 2.9.4], [4, Sect. 4.4], and [7, Sect. 1.4.1]

5.11. Question: In engineering and applied science, measured signals are real. Why should real measurable effects be represented by complex signals?

Answer: One complex signal (or channel) can carry information about two real signals (or two real channels), and the algebra and geometry of analyzing these two real signals as if they were one complex signal brings economies and insights that would not otherwise emerge.

6 Amplitude modulation: AM

6.1. The analysis of DSB-SC in the earlier sections illustrates that the spectrum of a DSB signal does not contain a **discrete** spectral component at the carrier frequency unless $m(t)$ has a DC component. This is why we referred to it as a ***suppressed carrier*** system.

6.2. DSB-SC amplitude modulation is easy to understand and to analyze in both time and frequency domains. However, analytical simplicity is not always accompanied by an equivalent simplicity in practical implementation.

Problem: The (coherent) demodulation of DSB-SC signal requires the receiver to possess a carrier signal that is synchronized with the incoming carrier. This requirement is not easy to achieve in practice because the modulated signal may have traveled hundreds of miles and could even suffer from some unknown frequency shift.

6.3. If a carrier component is transmitted along with the DSB signal, demodulation can be simplified.

(a) The received carrier component can be extracted using a narrowband bandpass filter and can be used as the demodulation carrier. (There is no need to generate a carrier at the receiver.)

(b) If the carrier amplitude is sufficiently large, the need for generating a demodulation carrier can be completely avoided.

- This will be the focus of this section.

Definition 6.4. For AM, the transmitted signal is typically defined as

$$x_{\mathrm{AM}}(t) = (A + m(t))\cos(2\pi f_c t) = \underbrace{A\cos(2\pi f_c t)}_{\text{carrier}} + \underbrace{m(t)\cos(2\pi f_c t)}_{\text{sidebands}}$$

6.5. Trade-off:

(a) *Disadvantage*:

- Higher power and hence higher cost required at the transmitter
- The carrier component is wasted power as far as information transfer is concerned.
- This fact can completely preclude the use of AM in power-limited applications.

(b) *Advantage*:

- Coherent reference is not needed for demodulation.
- Demodulator becomes simple and inexpensive.
- For broadcast system such as commercial radio (with a huge number of receivers for each transmitter,
  - any cost saving at the receiver is multiplied by the number of receiver units.
  - it is more economical to have one expensive high-power transmitter and simpler, less expensive receivers.

(c) Conclusion: Broadcasting systems tend to favor the trade-off by migrating cost from the (many) receivers to the (fewer) transmitters.

6.6. Spectrum of x_{AM}:

- Basically the same as that of DSB-SC except for the two additional impulses at $\pm f_c$.

Definition 6.7. Consider a signal $A(t)\cos(2\pi f_c t)$. If $A(t)$ varies slowly in comparison with the sinusoidal carrier $\cos(2\pi f_c t)$, then the ***envelope*** $E(t)$ of $A(t)\cos(2\pi f_c t)$ is $|A(t)|$.

6.8. *Envelope of AM signal*: See Figure 9. For AM signal, $A(t) = A + m(t)$.

(a) If $\forall t, A(t) > 0$, then $E(t) = A(t) = A + m(t)$

- The envelope has the same shape as $m(t)$.
- We can detect the desired signal $m(t)$ by detecting the envelope (envelope detection).

(b) If $\exists t, A(t) < 0$, then $E(t) \neq A(t)$.

- The envelope shape differs from the shape of $m(t)$ because the negative part of $A + m(t)$ is rectified.

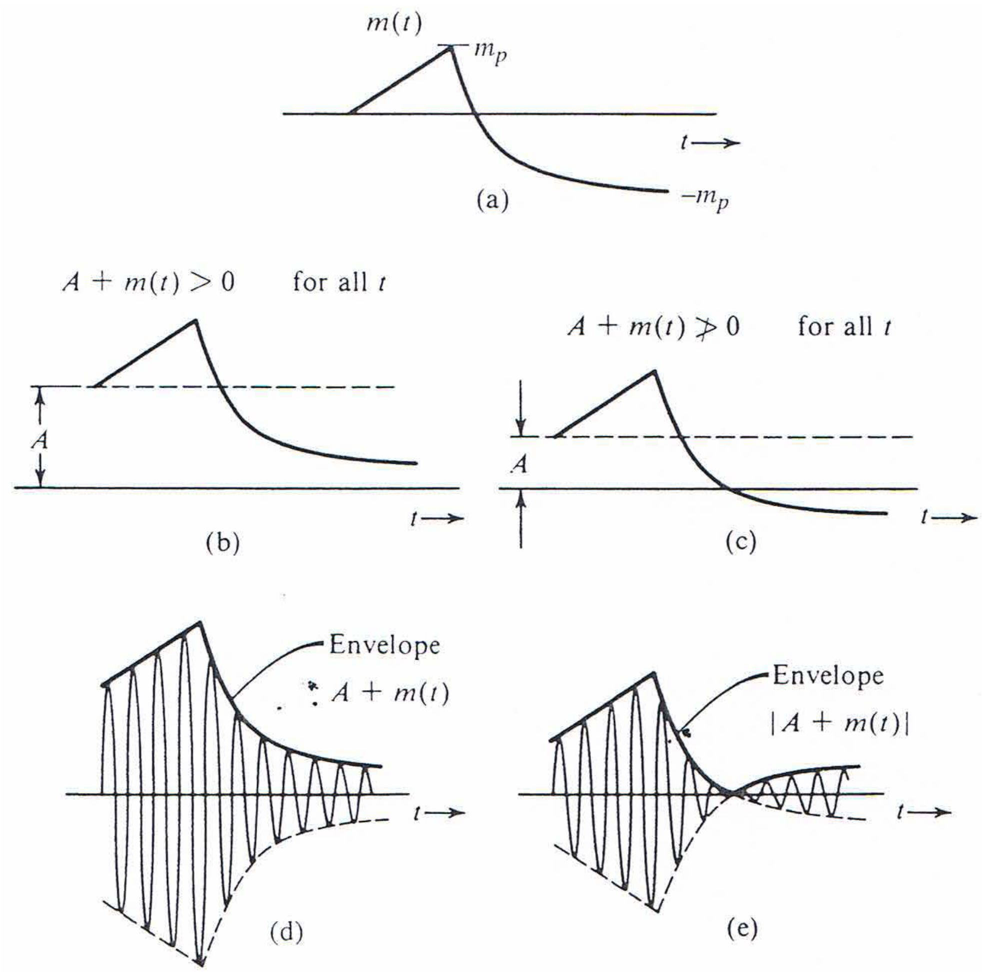


Figure 9: AM signal and its envelope [5, Fig 4.8]

6.9. Summary of AM Concept:

- The carrier term $A\cos(2\pi f_c t)$ is added.
- The size of A affects the time domain envelope of the modulated signal.
- A should be large enough to ensure that $A+m(t)$ is always nonnegative.
  - If $\forall t, m(t) \geq 0$, then there is no need to add any carrier. The DSB-SC signal can be detected by envelope detection.

6.10. ***Demodulation*** of AM Signals via rectifier detector: The receiver will first recover $A + m(t)$ and then remove A. Note that, conceptually, the received signal is the same as DSB-SC signal except that the $m(t)$ in the DSB-SC signal is replaced by $A(t) = A + m(t)$. We will also assume that A is large enough so that $A(t) \geq 0$.

Recall the key equation of ***switching demodulator*** (38):

$$\text{LPF}\{A(t)\cos(2\pi f_c t) \times 1[\cos(2\pi f_c t) \geq 0]\} = \frac{1}{\pi}A(t) \tag{39}$$

We noted before that this technique requires the switching to be in sync with the incoming cosine.

When $\forall t, A(t) \geq 0$, we can replace the switching demodulator by the ***rectifier demodulator/detector***. In which case, we suppress the negative part of $m(t)\cos(\omega_c t)$ using a diode (half-wave rectifier). This is mathematically equivalent to switching demodulator in (38) and (39).

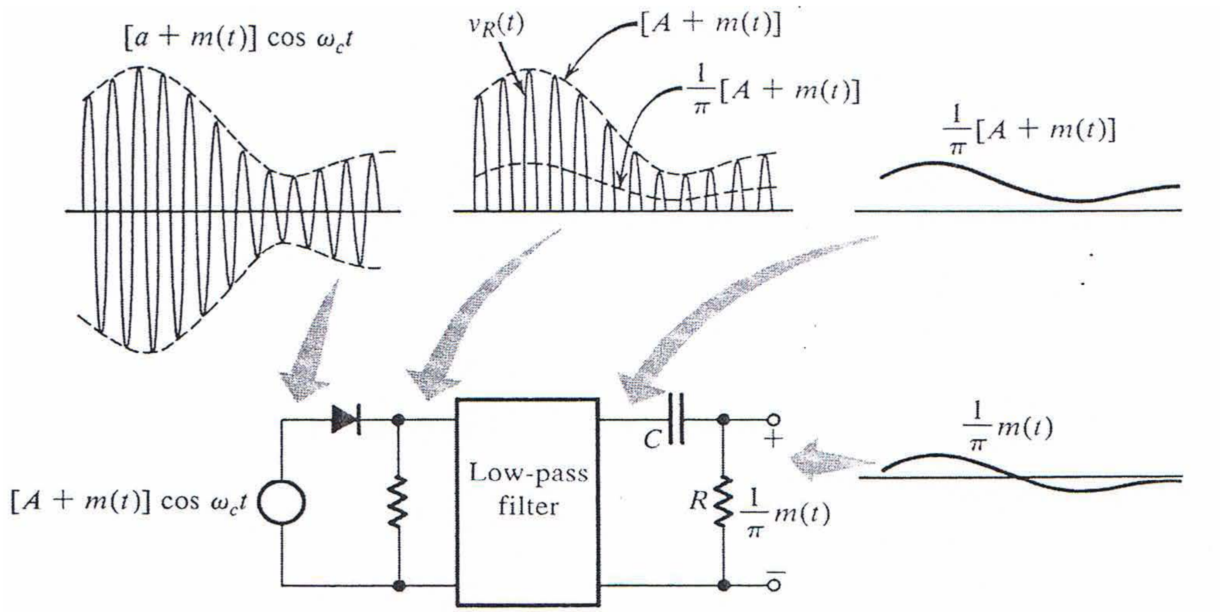


Figure 10: Rectifier detector for AM [5, Fig. 4.10].

- It is in effect synchronous detection performed without using a local carrier [4, p 167].
- This method needs $A(t) \geq 0$ so that the sign of $A(t)\cos(\omega_c t)$ will be the same as the sign of $\cos(\omega_c t)$.
- The dc term $\frac{A}{\pi}$ may be blocked by a capacitor to give the desired output $m(t)/\pi$.

6.11. Demodulation of AM signal via ***envelope detector***:

- Design criterion of RC:

$$2\pi B \ll \frac{1}{RC} \ll 2\pi f_c.$$

- The envelope detector output is $A + m(t)$ with a ripple of frequency f_c.
- The dc term can be blocked out by a capacitor or a simple RC high-pass filter.
- The ripple may be reduced further by another (low-pass) RC filter.

6.12. References: [3, p 198–199], [5, Section 4.3] and [10, Section 3.1.2].

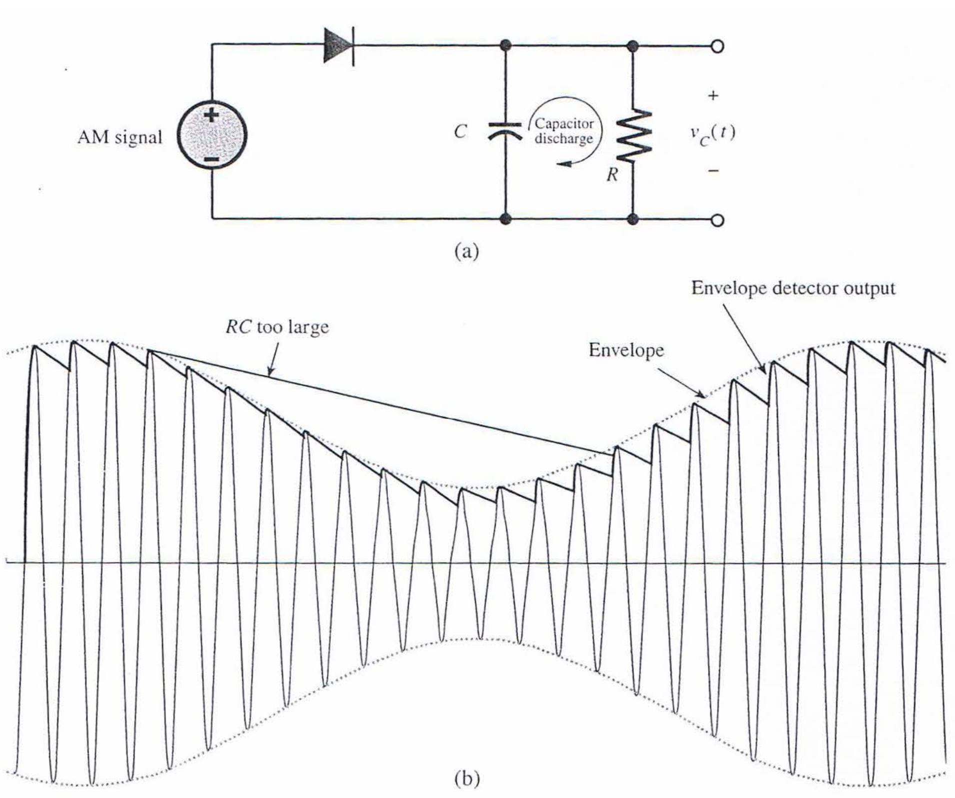


Figure 11: Envelope detector for AM [5, Fig. 4.11].

Sirindhorn International Institute of Technology
Thammasat University
School of Information, Computer and Communication Technology

ECS332 2012/1 Part II.3 Dr.Prapun

7 Angle Modulation: FM and PM

7.1. Recall 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.

7.2. As usual, we will again assume that the baseband signal $m(t)$ is band-limited to B; that is, $|M(f)| = 0$ for $|f| > B$.

In this section, we will also assume that

$$|m(t)| \le m_p.$$

In other words, $m(t)$ is bounded between $-m_p$ and m_p.

Definition 7.3. The main characteristic[9] of ***frequency modulation*** is that the carrier frequency $f(t)$ would be varied with time so that

$$f(t) = f_c + km(t), \tag{40}$$

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.

7.4. FM: A Magical Technique?

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

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

- Hence, the spectral components would remain within this band with a bandwidth $2km_p$ centered at f_c.

- Conclusion: By using an arbitrarily small k, 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.

[9] Treat this as a practical definition. The more rigorous definition will be provided in 7.11.

7.5. 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 $x_{FM}(t)$ 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]

7.1 Instantaneous Frequency

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

Definition 7.6. The ***generalized sinusoidal*** signal is a signal of the form

$$x(t) = A\cos(\theta(t)) \tag{41}$$

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

- The generalized angle for conventional sinusoid is $\omega_c t + \theta_0$.

7.7. Suppose we want the frequency f_c of a carrier $A\cos(2\pi f_c t)$ to vary with time as in (40). It is tempting to consider the signal

$$A\cos(2\pi f(t)t),$$

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

Example 7.8. See Slides. Consider the generalized sinusoid with $f(t) = t^2$.

Definition 7.9. For generalized sinusoid $A\cos(\theta(t))$, the ***instantaneous frequency***[10] at time t is given by

$$f(t) = \frac{1}{2\pi}\frac{d}{dt}\theta(t). \tag{42}$$

7.10. Equation (42) implies

$$\theta(t) = 2\pi\int_{-\infty}^{t} f(\tau)d\tau = \theta(t_0) + 2\pi\int_{t_0}^{t} f(\tau)d\tau. \tag{43}$$

[10] 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.

Definition 7.11. ***Frequency modulation*** (***FM***):

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

The instantaneous frequency is given by

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

Definition 7.12. ***Phase modulation*** (***PM***):

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

The instantaneous frequency is given by

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

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

where h is causal.

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

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

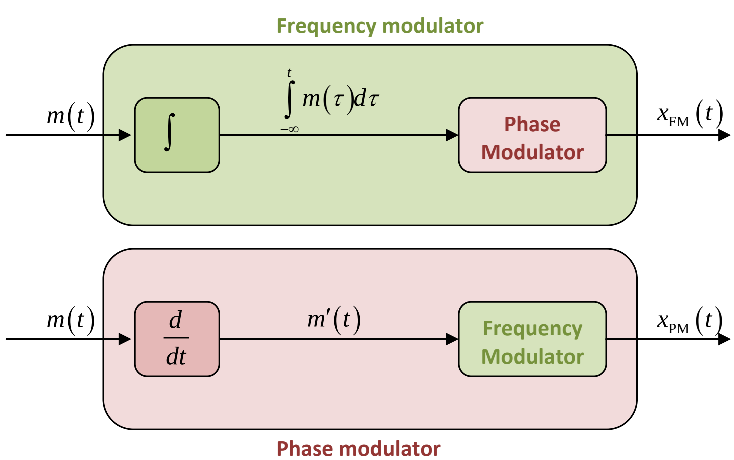


Figure 12: With the help of integrating and differentiating networks, a phase modulator can produce frequency modulation and vice versa [4, Fig 5.2].

A Trig Identities

All of the trigonometric functions of an angle θ can be constructed geometrically in terms of a unit circle centered at origin as shown in Figure 13.

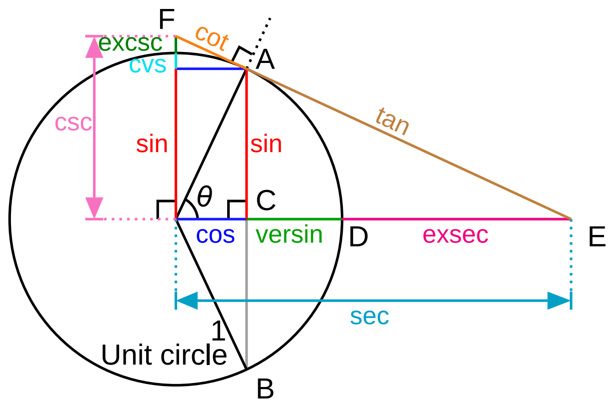


Figure 13: Trigonometric functions on a unit circle.

A.1. Cosine function

(a) Is an even function: $\cos(-x) = \cos(x)$.

(b) $\cos\left(x - \frac{\pi}{2}\right) = \sin(x)$.

(c) Sum formula:

$$\cos(x \pm y) = \cos x \cos y \mp \sin x \sin y. \tag{44}$$

(d) Product-to-Sum Formula:

$$\cos(x)\cos(y) = \frac{1}{2}\left(\cos(x+y) + \cos(x-y)\right).$$

(e) $$\cos^n x = \begin{cases} \frac{1}{2^{n-1}} \sum_{k=0}^{\frac{n-1}{2}} \binom{n}{k} \cos\left((n-2k)x\right), & \text{odd } n \geq 1 \\ \frac{1}{2^n} \left(\sum_{k=0}^{\frac{n}{2}-1} 2\binom{n}{k} \cos\left((n-2k)x\right) + \binom{n}{\frac{n}{2}} \right), & \text{even } n \geq 2 \end{cases}$$

(f) Any two real numbers a, b can be expressed in terms of cosine and sine with the same amplitude and phase:

$$(a, b) = (A\cos(\phi), A\sin(\phi)), \tag{45}$$

where $A = \sqrt{a^2+b^2}$ and $\phi = \tan^{-1}\frac{b}{a}$. This is simply the polar-coordinates from of the point (a, b) on Cartesian coordinates.

A.2. Properties of e^{ix}

(a) ***Euler's formula***: $\boxed{e^{ix} = \cos x + i\sin x.}$ Hence,

$$\cos(A) = \text{Re}\left(e^{jA}\right) = \frac{1}{2}\left(e^{jA} + e^{-jA}\right)$$
$$\sin(A) = \text{Im}\left(e^{jA}\right) = \text{Re}\left(-je^{jA}\right) = \text{Re}\left(+\frac{1}{j}e^{jA}\right) = \frac{1}{2j}\left(e^{jA} - e^{-jA}\right).$$

- We can use $\cos x = \frac{1}{2}\left(e^{ix} + e^{-ix}\right)$ and $\sin x = \frac{1}{2i}\left(e^{ix} - e^{-ix}\right)$ to derive many trigonometric identities.

In fact, we can combine linear combination of cosine and sine of the same argument into a single cosine by

$$A\cos\omega_0 t + B\sin\omega_0 t = \sqrt{A^2+B^2}\cos\left(\omega_0 t - \tan^{-1}\frac{B}{A}\right).$$

To see this, note that

$$\begin{aligned} A\cos\omega_0 t + B\sin\omega_0 t &= \text{Re}\left(Ae^{j\omega_0 t}\right) + \text{Re}\left(-jBe^{j\omega_0 t}\right) = \text{Re}\left((A - jB)\, e^{j\omega_0 t}\right) \\ &= \text{Re}\left(\sqrt{A^2+B^2}e^{-j\tan^{-1}\frac{B}{A}}e^{j\omega_0 t}\right). \end{aligned}$$

Another way to see this is to reexpress the two real numbers A, B using (45) and then use (44).

(b) e^{jx} is periodic with period 2π.

(c) Any complex number $z = x+jy$ can be expressed as $z = \sqrt{x^2+y^2}e^{j\tan^{-1}\left(\frac{y}{x}\right)} = |z|e^{j\phi}$.

- $z^t = |z|^t e^{j\phi t}$.

(d) More relations involving sin and cos.

- $e^{jAt} + e^{jBt} = 2e^{j\frac{A+B}{2}t}\cos\left(\frac{A-B}{2}\right)$.
- $e^{jAt} - e^{jBt} = 2je^{j\frac{A+B}{2}t}\sin\left(\frac{A-B}{2}\right)$
- $\frac{e^{jAt}-e^{jBt}}{e^{jCt}-e^{jDt}} = e^{j\frac{(A+B)-(C+D)}{2}t}\frac{\sin\left(\frac{A-B}{2}\right)}{\sin\left(\frac{C-D}{2}\right)}$.

References

[1] Richard E. Blahut. *Modem Theory: An Introduction to Telecommunications*. Cambridge University Press, 1 edition, December 2009. 7.5

[2] Albert Boggess and Francis J. Narcowich. *First Course in Wavelets with Fourier Analysis*. Prentice Hall, 1 edition, 2001. 2.27, 2.28

[3] A. Bruce Carlson and Paul B. Crilly. *Communication Systems: An Introduction to Signals and Noise in Electrical Communication*. McGraw-Hill, 5th international edition edition, 2010. 1, 2, 3, 4, 5.4, 5.10, 6.12

[4] B. P. Lathi. *Modern Digital and Analog Communication Systems*. Oxford University Press, 1998. 2.2, 2, 3, 3.10, 4.5, 4.16, 8, 4.18, 8, 5.4, 5.10, 6.10, 12

[5] B. P. Lathi and Zhi Ding. *Modern Digital and Analog Communication Systems*. Oxford University Press, 2009. 9, 10, 6.12, 11

[6] C. Britton Rorabaugh. *Communications Formulas and Algorithms: For System Analysis and Design*. Mcgraw-Hill, 1990.

[7] Peter J. Schreier and Louis L. Scharf. *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press, 2010. 6, 5.10

[8] Claude E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3):379–423, July 1948. Continued 27(4):623-656, October 1948. 1.1, 1.2

[9] Elias M. Stein and Rami Shakarchi. *Fourier Analysis: An Introduction*. Princeton University Press, March 2003. 2.28

[10] Rodger E. Ziemer and William H. Tranter. *Principles of Communications*. John Wiley & Sons Ltd, 2010. 2.9, 3.11, 5.4, 5.10, 6.12