EE 115 Lecture Note 3
Analog Communication Techniques (Cont.)

Y. Hua

I. Angle Modulation

An angle modulated signal has the following form:

$$u(t) = A_c \cos(2\pi f_c t + \theta(t)) \tag{1}$$

where $\theta(t)$ is a function of the message $m(t)$. The angle modulation keeps a constant envelope of the modulated signal, which no longer requires the linearity for power amplification. (All methods for amplitude modulation require linear amplification.)

For phase modulation (PM),

$$\theta(t) = k_p m(t) \tag{2}$$

where k_p is the *phase modulation constant*.

For frequency modulation (FM),

$$\theta(t) = \theta(0) + 2\pi k_f \int_0^t m(\tau) d\tau \tag{3}$$

where k_f is the *frequency modulation constant*. In this case, the frequency offset is proportional to the message, i.e., $k_f m(t)$.

We see that FM consists of PM driven by an integral of the message. If $m(t)$ is a binary signal equal to 1 or -1 at any time t, the PM signal has a step change of its phase at each transition, and the FM signal has a step change of its frequency at each transition. FM is also known as *continuous phase modulation*.

Since FM has a smoother phase change, FM has a smaller bandwidth than PM and hence is preferred to PM for analog modulation (where the system designer has little control over the variation of the message $m(t)$). But for digital communication, the system designer has more control over the shape of $m(t)$, and hence PM or often called *phase-shift keying (PSK)* is commonly used.

We will focus on FM. Some key notions:

1) The maximum frequency deviation is $\Delta f_{max} = k_f \max_t |m(t)|$.
2) The *FM modulation index* is defined as $\beta = \frac{\Delta f_{max}}{B}$ where B is the bandwidth of the message.

3) If $\beta < 1$, we have a narrowband FM signal.
4) If $\beta > 1$, we have a wideband FM signal

Assume $m(t) = A_m \cos(2\pi f_m t)$. Then $\Delta f_{max} = k_f A_m$ and $\beta = \frac{k_f A_m}{f_m}$. Also note

$$\theta(t) = 2\pi k_f \int_0^t A_m \cos(2\pi f_m \tau) d\tau = \frac{A_m k_f}{f_m} \sin(2\pi f_m t) = \beta \sin(2\pi f_m t). \tag{4}$$

A. FM Modulator

An FM modulator is a *voltage-controlled oscillator* (VCO). A direct approach is to directly produce

$$u(t) = A_c \cos(2\pi f_c + 2\pi k_f \int_0^t m(\tau) d\tau) \tag{5}$$

where $m(t)$ is the input voltage to a VCO. This is a common approach.

If there is a VCO that can produce

$$u_1(t) = A_1 \cos(2\pi f_1 + 2\pi k_1 \int_0^t m(\tau) d\tau) \tag{6}$$

where $f_1 = f_c/M$, we can then use *frequency amplification* to produce

$$u(t) = A_c \cos[M(2\pi f_1 + 2\pi k_1 \int_0^t m(\tau) d\tau)] = A_c \cos(2\pi f_c t + 2\pi k_f \int_0^t m(\tau) d\tau) \tag{7}$$

where $k_f = M k_1$. If $u_1(t)$ is a narrowband FM signal, then $u(t)$ can be a wideband FM signal.

For example, if we have

$$u_1^2(t) = A_1^2 \cos^2(2\pi f_1 + 2\pi k_1 \int_0^t m(\tau) d\tau) \tag{8}$$

then we can extract the following from $u_1^2(t)$ using a passband filter:

$$u_2(t) = A_2 \cos(4\pi f_1 + 4\pi k_1 \int_0^t m(\tau) d\tau). \tag{9}$$

In general, a nonlinear function of $u_1(t)$ is needed to achieve frequency amplification.

B. Limiter-Discriminator Demodulation (FM Demodulator)

Assume that the received signal is

$$y(t) = A(t) \cos(2\pi f_c t + \theta(t)) + w(t) \tag{10}$$

where $A(t)$ can be a function of time t due to channel distortion and $w(t)$ is any other noise. To reduce the distortion and noise, it is common to pass $y(t)$ through a limiter which consists of a sign function and a bandpass filter around f_c. The output of the limiter is

$$u(t) = A \cos(2\pi f_c t + \theta(t)) \tag{11}$$

which is then passed through a discriminator whose output is $\frac{d\theta(t)}{dt}$.

The discriminator can be implemented via a differentiator, an envelope detector and a DC blocker.

Note

$$v(t) = \frac{du(t)}{dt} = -A\sin(2\pi f_c t + \theta(t))[2\pi f_c + 2\pi k_f m(t)] \tag{12}$$

whose envelope is $2\pi f_c + 2\pi k_f m(t)$ (assuming $f_c > k_f \max_t |m(t)|$).

An ideal differentiator has the frequency response $H(f) = j2\pi f$ for all f. But for the bandpass signal $u(t)$, we can achieve the same goal by using a filter (or slope detector) whose frequency response is

$$|H(f)| \approx H_0 + \alpha(f - f_c) \tag{13}$$

for $|f - f_c| \leq W/2$ where W is the bandwidth of $u(t)$.

C. Spectrum of FM Signal

Recall the FM signal $u_p(t) = A_c\cos(2\pi f_c + \theta(t))$. Since $\theta(t) = \theta(0) + 2\pi k_f \int_0^t m(\tau)d\tau$, $\theta(t)$ has the same bandwidth B as $m(t)$. Since $\theta(0)$ does not affect the bandwidth of $u_p(t)$, we will set $\theta(0) = 0$.

The complex envelope of $u_p(t)$ (with respect to $\cos(2\pi f_c t)$) is now denoted by $u(t)$, which is

$$u(t) = A_c e^{j\theta(t)} = A_c\cos\theta(t) + jA_c\sin\theta(t) \tag{14}$$

If k_f is so small that $\theta(t) \ll 1$, we have

$$u(t) \approx A_c + jA_c\theta(t) \tag{15}$$

which has the bandwidth B, and hence

$$u_p(t) \approx A_c\cos(2\pi f_c) - A_c\theta(t)\sin(2\pi f_c t) \tag{16}$$

which has the bandwidth equal to $2B$.

In general, the bandwidth of $u_p(t)$ is

$$B_{FM} \approx 2B + 2\Delta f_{max} = 2B(1 + \beta) \tag{17}$$

which is known as *Carson's rule*.

1) Single-tone case: Let $m(t) = \cos(2\pi f_m t)$ and then $\theta(t) = 2\pi k_f \int_0^t m(\tau)d\tau = \frac{k_f}{f_m}\sin(2\pi f_m t) = \beta \sin(2\pi f_m t)$. The complex envelope of the FM signal is

$$u(t) = e^{j\theta(t)} = e^{j\beta \sin(2\pi f_m t)}. \tag{18}$$

Since $u(t)$ has the period $1/f_m$, it follows that

$$u(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_m t} \tag{19}$$

where

$$c_n = f_m \int_{-\frac{1}{2f_m}}^{\frac{1}{2f_m}} u(t) e^{-j2\pi n f_m t} dt = f_m \int_{-\frac{1}{2f_m}}^{\frac{1}{2f_m}} e^{j\beta \sin(2\pi f_m t)} e^{-j2\pi n f_m t} dt. \tag{20}$$

Let $x = 2\pi f_m t$. Then

$$c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j(\beta \sin x - nx)} dx = \frac{1}{\pi}\int_0^{\pi} \cos(\beta \sin x - nx) dx = J_n(\beta) \tag{21}$$

which is the *Bessel function* of the first kind of order n.

Therefore, the spectrum of $u(t)$ is

$$U(f) = \sum_{n=-\infty}^{\infty} J_n(\beta)\delta(f - nf_m). \tag{22}$$

For illustration of $J_n(\beta)$, see page 115. Some properties of $J_n(\beta)$ are:

1) For integer n, $J_n(\beta) = (-1)^n J_{-n}(\beta) = (-1)^n J_n(-\beta)$. Hence, $|J_n(\beta)| = |J_{-n}(\beta)| = |J_n(-\beta)|$.
2) For a fixed β, $J_n(\beta)$ approaches zero rapidly as $|n|$ becomes large. In particular, $J_n(\beta)$ is small if $|n| > \beta + 1$. More specifically, if we let

$$u_K(t) = \sum_{n=-K}^{K} J_n(\beta) e^{j2\pi f_m n t} \tag{23}$$

then

$$\frac{\overline{u_K^2(t)}}{\overline{u^2(t)}} = \frac{J_0^2(\beta) + 2\sum_{n=1}^{K} J_n^2(\beta)}{J_0^2(\beta) + 2\sum_{n=1}^{\infty} J_n^2(\beta)} \tag{24}$$

which is close to one (e.g., 99%) if $K = \beta + 1$.

So, an approximate bandwidth of $u(t)$ is $f_m(\beta + 1)$, and hence an approximate bandwidth of $u_p(t)$ is $2f_m(\beta + 1)$. This is consistent with Carson's rule.
3) For fixed n, $J_n(\beta) = 0$ for specific values of β. This property can be used for spectral shaping. For example, if we want $u_p(t)$ to have a spectral null at $f_c \pm 10 f_m$, then we should choose such β that $J_{10}(\beta) = 0$.

II. Superheterodyne Receiver

For more bandwidth and smaller antenna, it is necessary to exploit the radio spectrum at higher frequency. Many many decades ago, bandpass filtering and power amplification at high frequency (such as several MHz) was difficult. This motivated the idea of superheterodyne (or superhet) receiver.

A superhet receiver uses a low-quality (broadband) frontend RF filter/amplifier to pick up a combination $y_{RF}(t)$ of several signals at multiple carrier frequencies (corresponding to multiple "stations"). The composite signal $y_{RF}(t)$ is then translated in frequency to an intermediate frequency (IF) region. At IF, the signal $y_{IF}(t)$ is further bandpass-filtered and amplified (with desired narrower band), which can be done with high quality due to $f_{IF} < f_{RF}$.

Note that a quality factor (Q-factor) of an analog bandpass filter is the center-frequency to bandwidth ratio. A good Q factor is around 100. (The higher Q, the more costly.) Assuming $Q = 100$, the bandwidth of a filter at the carrier frequency 10MHz would be no less than $100kHz$, which includes many voice channels each of $10kHz$ (in passband).

Assume that the RF frontend picks up the following signal

$$y_{RF}(t) = A(t)\cos(2\pi f_{RF}t + \theta(t)) \tag{25}$$

To reduce its center frequency, we need a LO to generate $\cos(2\pi f_{LO}t)$, which is used to mix with $y_{RF}(t)$ to produce:

$$z(t) = y_{RF}(t)2\cos(2\pi f_{LO}t) = A(t)\cos(2\pi(f_{LO} - f_{RF})t - \theta(t)) + A\cos(2\pi(f_{LO} + f_{RF})t + \theta(t)) \tag{26}$$

where the first term has the center frequency at $\pm|f_{LO} - f_{RF}|$ and the second term has the center frequency at $\pm(f_{LO} + f_{RF})$. The gap between these two center frequencies is large (either $2f_{RF}$ if $f_{LO} > f_{RF}$ or $2f_{LO}$ if $f_{RF} > f_{LO}$) and hence it is easy to filter out the second term.

We call $f_{IF} \doteq |f_{LO} - f_{RF}|$ the intermediate frequency (IF). The first term in $z(t)$ is the IF signal, i.e.,

$$y_{IF}(t) = A(t)\cos(2\pi(f_{LO} - f_{RF})t - \theta(t)). \tag{27}$$

To show the relationship between the spectrum $Y_{IF}(f)$ of $y_{IF}(t)$ and the spectrum $Y_{RF}(f)$ of $y_{RF}(t)$, let us assume $f_{LO} = f_{RF} + f_{IF}$. In this case, $Y_{IF}(f)$ consists of the right side of $Y_{RF}(f)$ shifted to the left by f_{LO} (mostly residing in the negative frequency) and the left side of $Y_{RF}(f)$ shifted to the right by f_{LO} (mostly residing in the positive frequency).

The desired message of bandwidth B is around $\pm f_{RF}$ of $Y_{RF}(f)$, or equivalently around $\pm f_{IF}$ of $Y_{IF}(f)$. To avoid any interference inside $Y_{IF}(f)$, we need the single-sided bandwidth of $A(t)e^{j\theta(t)}$ (the complex envelope of $y_{RF}(t)$ with respect to $\cos(2\pi f_{RF}t)$) or equivalently $A(t)e^{-j\theta(t)}$ (the complex envelope of $y_{IF}(t)$ with respect to $\cos(2\pi f_{IF}t)$) to be less than $2f_{IF}-B$.

It is rather cheap to have a RF frontend amplifier/filter with the center frequency f_{RF} and the bandwidth no larger than $4f_{IF}-2B$ (as long as the latter is not much smaller than the former). This RF filter is also called *image-reject filter*.

In practice (see page 119), the image-reject filter allows us to tune f_{RF} and f_{LO} at the same time (in synch) such that $|f_{LO}-f_{RF}|$ is always a fixed value equal to f_{IF}. In this way, the circuit needed to process $z(t)$ can be all based on the same center frequency f_{IF}, which makes the IF circuit design much easier.

For example, $f_{IF} = 455$kHz (a typical value).

Given f_{IF}, there are two choices for f_{LO}: $f_{LO} = f_{RF} + f_{IF}$ and $f_{LO} = f_{RF} - f_{IF}$.

For the same range $f_{min} \leq f_{RF} \leq f_{max}$, the first choice of f_{LO} has a smaller maximum-to-minimum ratio, which makes the circuit easier to design. If $f_{IF} = 455$kHz, $f_{min} = 540$kHz and $f_{max} = 1600$kHz (for AM radio), then the first choice corresponds to $995kHz \leq f_{LO} \leq 2055kHz$ while the second choice corresponds to $85kHz \leq f_{LO} \leq 1145kHz$.

Following the mixer which yields $z(t)$, we need a high quality bandpass filter/amplifier at IF to extract out a desired narrow-passband signal $u_p(t)$ from $z(t)$ (or equivalently from $y_{IF}(t)$). If the desired message $m(t)$ was transmitted via the conventional AM at $f_c = f_{RF}$, then

$$u_p(t) = A(a_{mod}m_n(t) + 1)\cos(2\pi f_{IF}t + \phi). \tag{28}$$

Following the bandpass filter, an envelope detector can be used to extract out $m(t)$.

There is a fundamental tradeoff in superhet receiver. To relax the quality of image-reject filter (for a fixed f_{RF}), we should increase f_{IF}. But to relax the quality of the IF filter (for a fixed bandwidth), we should decrease f_{IF}. In practice, we may need multiple stages for frequency translation and image rejection. We can move the message from f_{RF} to $f_{IF,1}$, $f_{IF,2}$, ..., $f_{IF,N}$ sequentially while we reject enough images at each stage of the frequency translation. Here, $f_{RF} > f_{IF,1} > f_{IF,2} > ... > f_{IF,N}$.

A. Direct Conversion

If we have a *reliable LO signal* at f_{RF}, then we can perform

$$z(t) = y_{RF}(t) 2\cos(2\pi f_{RF} t) = A(t)\cos(\theta(t)) + A\cos(4\pi f_{RF} t + \theta(t)) \tag{29}$$

where the second term is easy to be filtered out. It can be seen in the frequency domain that the spectrum of the first term $z_1(t) = A(t)\cos(\theta(t))$ consists of the desired message and other components that do not overlap with the desired message. In this case, no image-reject filter is needed before the frequency translation.

To extract out the desired message $m(t)$ of bandwidth B from $z_1(t)$, we can first pass $z_1(t)$ through an analog lowpass filter with a bandwidth $W > B$, and then sample the output $z_2(t)$ of the analog lowpass filter at a sampling rate $f_s > 2W$. We can then apply a digital filter to extract out the message $m(t)$ in its (equivalent) discrete form.

This direct conversion method has been used in modern digital communication transceivers (since 1990) where f_{RF} ranges (for example) in the licensed bands from 900MHz to 2GHz, and in the unlicensed bands around 2.4GHz and 5GHz.

Digital signal processing (on small chips) was not available until 1970's. Analog lowpass filtering would be too bulky. That is a main reason why the superhet receiver was widely used (instead of the direct conversion).

III. The Phased-Locked Loop (PLL)

The PLL has many important applications including FM demodulation. The concept of PLL consists of a phase detector, a loop filter and a voltage-controlled oscillator (VCO). The output $\theta_0(t)$ of VCO is controlled by its input voltage $x(t)$ while $x(t)$ is the output of the loop filter driven by $\theta_i(t) - \theta_0(t)$ with $\theta_i(t)$ being the input to the PLL. The loop filter is such that $\theta_i(t) - \theta_0(t)$ is as small as possible for almost all the time.

In practice, the phase detector is replaced by a mixer which performs the following:

$$A_c\cos(2\pi f_c t + \theta_i(t)) \times [-A_v \sin(2\pi f_c t + \theta_0(t))] = \frac{1}{2} A_c A_v \sin(\theta_i(t) - \theta_0(t)) - \frac{1}{2} A_c A_v \sin(4\pi f_c t + \theta_i(t) + \theta_0(t)) \tag{30}$$

where we have used $2\sin\alpha\cos\beta = \sin(\alpha - \beta) + \sin(\alpha + \beta)$. And the second term is easy to be filtered out as long as the loop filter is a lowpass.

Given $\theta_0(t) = K_v \int_0^t x(\tau)d\tau$ and $\theta_i(t) = \theta_i(0) + 2\pi k_f \int_0^t m(\tau)d\tau$, then $\theta_0(t) = \theta_i(t)$ implies $\frac{d\theta_0(t)}{dt} = \frac{d\theta_i(t)}{dt}$ or equivalently

$$K_v x(t) = 2\pi k_f m(t). \tag{31}$$

A. Analysis of PLL

Assume $\theta_i - \theta_0 \ll 1$. Then $\sin(\theta_i - \theta_0) \approx \theta_i - \theta_0$. Then, in the s-domain, we know

$$(\Theta_i(s) - \Theta_0(s))KG(s)/s = \Theta_0(s) \tag{32}$$

where $G(s)$ is the transfer function of the loop filter, $K = \frac{1}{2}A_c A_v K_v$ is the overall loop gain, and $1/s$ is due to the VCO. It then follows that

$$H(s) \doteq \frac{\Theta_0(s)}{\Theta_i(s)} = \frac{KG(s)}{s + KG(s)} \tag{33}$$

and

$$H_e(s) \doteq \frac{\Theta_e(s)}{\Theta_i(s)} \doteq \frac{\Theta_i(s) - \Theta_0(s)}{\Theta_i(s)} = 1 - H(s) = \frac{s}{s + KG(s)}. \tag{34}$$

1) First-Order PLL: Let $G(s) = 1$ (within the bandwidth of interest). Then

$$H(s) = \frac{K}{s + K} \tag{35}$$

which corresponds to the impulse response $h(t) = Ke^{-Kt}u(t)$.

If $\theta_i(t) = \Delta\theta u(t)$ (a step phase input), then

$$\Theta_0(s) = H(s)\Theta_i(s) = \frac{K\Delta\theta}{s(s + K)} = \frac{\Delta\theta}{s} - \frac{\Delta\theta}{s + K} \tag{36}$$

and its inverse Laplace transform is

$$\theta_0(t) = \Delta\theta(1 - e^{-Kt})u(t) \tag{37}$$

which converges to $\Delta\theta$ as $t \to \infty$. It follows that $\theta_e(t) = \theta_i(t) - \theta_0(t) = \Delta\theta e^{-Kt}u(t) \to 0$ as $t \to \infty$. Here we see that the larger is K, the faster is the convergence.

If $f_i(t) = \Delta f u(t)$ (a step input frequency shift), then $\theta_i(t) = 2\pi \int_0^t f_i(\tau)d\tau$ and hence $\Theta_i(s) = \frac{2\pi\Delta f}{s^2}$ and

$$\Theta_0(s) = \Theta_i(s)H(s) = \frac{2\pi\Delta f}{s^2}\frac{K}{s + K} \tag{38}$$

and the phase error in the s-domain is

$$\Theta_e(s) = \Theta_i(s) - \Theta_0(s) = \frac{2\pi\Delta f}{s^2}\frac{s}{s + K}. \tag{39}$$

We see that (via the final value theorem)

$$\theta_e(\infty) = \lim_{s \to 0} s\Theta_e(s) = \frac{2\pi\Delta f}{K} \neq 0 \tag{40}$$

But the output frequency shift (in the s-domain) is

$$F_0(s) = \frac{1}{2\pi} s\Theta_0(s) = \frac{\Delta f}{s}\frac{K}{s+K}. \tag{41}$$

and hence

$$f_0(\infty) = \lim_{s \to 0} sF_0(s) = \Delta f. \tag{42}$$

We see that with the first-order PLL, there is a steady-state phase error although the steady-state frequency error is zero.

2) Second-Order PLL: Now assume that $G(s) = 1 + a/s$ ("proportional plus integral"). It follows that

$$H(s) = \frac{KG(s)}{s + KG(s)} = \frac{K(s+a)}{s^2 + Ks + Ka} \tag{43}$$

$$H_e(s) = 1 - H(s) = \frac{s}{s + KG(s)} = \frac{s^2}{s^2 + Ks + Ka} \tag{44}$$

The poles of the transfer functions are $s = (-K \pm \sqrt{K^2 - 4Ka})/2$, which are stable poles as $K > 0$.

If there is a step frequency shift in the input, i.e., $F_i(s) = \frac{\Delta f}{s}$, then $\Theta_i = 2\pi \frac{1}{s} F_i(s) = \frac{2\pi\Delta f}{s^2}$ and

$$\Theta_e(s) = H_e(s)\Theta_i(s) = \frac{2\pi\Delta f}{s^2 + Ks + Ka} \tag{45}$$

which implies

$$\theta_e(\infty) = \lim_{s \to 0} s\Theta_e(s) = 0. \tag{46}$$

So, with the second-order PLL, both the steady-state phase error and the steady-state frequency error are zero (in response to step frequency shift).

If we choose a 3rd order PLL (i.e., by choosing $G(s) = 1 + a/s + b/s^2$), the steady-state phase error can be similarly shown to be zero in response to a step derivative of the frequency shift/offset $f_i(t) = ctu(t)$ (a linear frequency ramp).

3) Nonlinear Model: Now we consider the exact (nonlinear) model of the phase detector and $G(s) = 1$. In this case, we have

$$\theta_0(t) = K \int_0^t \sin(\theta_i(\tau) - \theta_0(\tau)) d\tau \tag{47}$$

Let $\theta_e(t) = \theta_i(t) - \theta_0(t)$. It follows that

$$\theta_i(t) - \theta_e(t) = K \int_0^t \sin(\theta_e(\tau)) d\tau \tag{48}$$

Taking the derivative, we have

$$\frac{d\theta_i(t)}{dt} - \frac{d\theta_e(t)}{dt} = K \sin(\theta_e(t)). \tag{49}$$

Assume a step frequency offset $\frac{d\theta_i(t)}{dt} = 2\pi\Delta f$. Then

$$\frac{d\theta_e(t)}{dt} = 2\pi\Delta f - K \sin(\theta_e(t)) \tag{50}$$

We can see that

1) If $2\pi\Delta f > K$, then $\frac{d\theta_e(t)}{dt} > 0$ for all t and hence there is no convergence.
2) If $2\pi\Delta f < K$, $\frac{d\theta_e(t)}{dt}$ may be either positive or negative depending on the value of $\theta_e(t)$. The *stationary/equilibrium* points of θ_e are such that $\frac{d\theta_e(t)}{dt} = 2\pi\Delta f - K\sin(\theta_e) = 0$. There are two such points $\theta_{e,0}$ and $\theta_{e,1}$ within each interval of 2π along the axis of θ_e.
3) One of the two stationary points is stable while the other is unstable. The stable point is where $\frac{d}{d\theta_e(t)}[\frac{d\theta_e(t)}{dt}] < 0$ while the unstable point is where $\frac{d}{d\theta_e(t)}[\frac{d\theta_e(t)}{dt}] > 0$. (See the *phase-plane plot* of $\frac{d\theta_e(t)}{dt}$ vs $\theta_e(t)$.)
4) The smallest positive stationary point of θ_e is the stable stationary point.