ELEG 2310B: Principles of Communication Systems **2021-22 First Term**

Handout 8: Aspects of FM

Instructor: Wing-Kin Ma

Suggested Reading: Chapter 4 of Simon Haykin and Michael Moher, *Communication Systems (5th Edition)*, Wily & Sons Ltd; and Chapter 5 of B. P. Lathi and Z. Ding, *Modern Digital and Analog Communication Systems (4th Edition)*, Oxford University Press.

This handout continues with our study on FM.

1 Generation of FM Signals

The modulation process of FM requires a device that can output a sinusoidal wave whose instantaneous frequency varies in accordance with the message signal $m(t)$. In analog circuits, FM signals may be directly generated by a device called the *voltage-controlled oscillator (VCO)*. A classical example of VCO is a *Hartley oscillator*, shown in Figure 1. Simply speaking, the circuit shown in Figure 1 is a highly selective frequency-determining resonant network, whose oscillating frequency is

$$f_i(t) = \frac{1}{2\pi\sqrt{(L_1 + L_2)C(t)}},$$

where L_1 and L_2 are the two inductances of the Hartley oscillator, and $C(t)$ is the capacitance that can vary with time. In particular, the capacitive component for $C(t)$ consists of a fixed capacitor shunted by a voltage-variable capacitor. A voltage-variable capacitor is one whose capacitance depends on the voltage applied across its electrodes; it is also known as a *varactor* or *varicap*. By selecting the inductances and capacitances appropriately, an FM wave may be accurately generated by the Hartley oscillator.

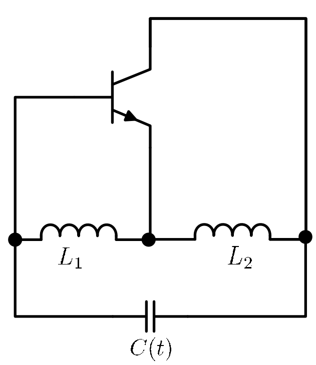

Figure 1: A Hartley oscillator.

Note that there are more sophisticated VCOs, as well as more advanced systems for FM modulation wherein the VCO is just a building block of the whole system. Such designs aim at providing good oscillator stability, and this is more related to the circuits topic. Also, in Armstrong's original method for FM signal generations, VCOs are not used. Interested students are referred to the textbooks for details.

2 Demodulation of FM Signals via Differentiation

There is more than one way to demodulate FM signals. Here, we focus on concepts by considering a simple method based on time differentiation. Let $s'(t) = \frac{ds(t)}{dt}$ denote the time differentiation of the FM signal. We have

$$\begin{aligned} s'(t) &= \frac{d}{dt}\left\{A_c \cos\left[2\pi f_c t + 2\pi k_f \int_{-\infty}^{t} m(\tau)d\tau\right]\right\} \\ &= -A_c[2\pi f_c + 2\pi k_f m(t)] \sin\left[2\pi f_c t + 2\pi k_f \int_{-\infty}^{t} m(\tau)d\tau\right]. \end{aligned} \tag{1}$$

Eq. (1) suggests that the envelope of $s'(t)$ is $A_c[2\pi f_c + 2\pi k_f m(t)]$, if we assume $1 + \frac{k_f}{f_c} m(t) \geq 0$ for all t. In fact, the latter condition is justified since f_c is often large in real-world applications. Hence, by extracting the envelope of $s'(t)$, we may recover the message signal $m(t)$[1].

The mathematical operations shown above can be practically realized by the system in Figure 2. We first apply a differentiator to carry out the differentiation process in (1). The differentiator can be implemented by a filter; in an ideal setting the filter response should achieve $H(f) = j2\pi f$ (why?). Then, an envelope detector, like those used in AM, is applied to extract the envelope of $s'(t)$. The FM demodulation system in Figure 2 is commonly called a *frequency discriminator*, since it discriminates the instantaneous frequency from the FM signal. Note that a frequency discriminator is a noncoherent detector.

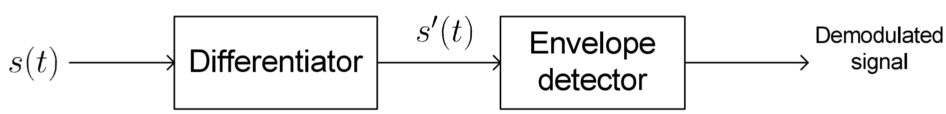

Figure 2: Block diagram of a frequency discriminator.

There are improved or more sophisticated versions of the frequency discriminator, e.g., the balanced frequency discriminator. They follow a similar principle as that described above. For details, please see the Haykin-Moher textbook or other communication textbooks.

3 Phase-Locked Loop

Simply speaking, a phase-locked loop (PLL) is a device that can track the instantaneous frequency of an incoming angle-modulated signal. Thus, it serves the purpose of FM demodulation. It is worthwhile to note that PLLs also have several other applications, such as synchronization, frequency division/multiplication, and frequency modulation.

[1]As in AM, we may assume that the the presence of the DC component in the envelope is not a problem.

Figure 3 shows a system diagram of the PLL. It consists of a multiplier, a loop filter and a VCO. The objective is to have the phase angle of the VCO output locked to that of the input FM signal—and this is done so via feedback control.

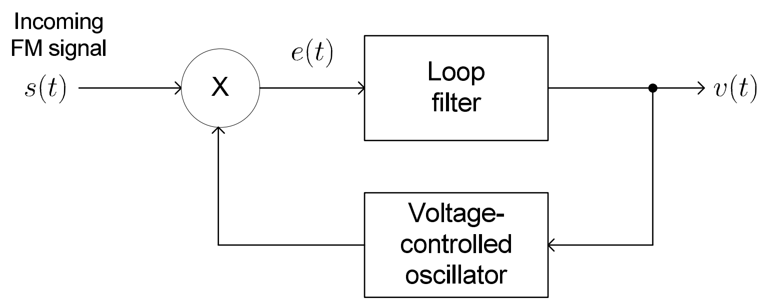

Figure 3: Block diagram of a phase-locked loop.

3.1 Problem Setup

Let us describe the problem setup. The input signal of the PLL is an FM signal

$$s(t) = A_c \sin[2\pi f_c t + \phi_1(t)], \tag{2}$$

where

$$\phi_1(t) = 2\pi k_f \int_{-\infty}^{t} m(\tau) d\tau. \tag{3}$$

The VCO generates an angle-modulated wave whose instantaneous frequency depends on the VCO input $v(t)$. Specifically, the VCO output is modeled as

$$r(t) = A_c \cos[2\pi f_c t + \phi_2(t)], \tag{4}$$

where

$$\phi_2(t) = 2\pi k_v \int_{-\infty}^{t} v(\tau) d\tau, \tag{5}$$

with k_v denoting the frequency sensitivity of the VCO. At the multiplier, we have an output

$$e(t) = s(t) \cdot r(t). \tag{6}$$

The output $e(t)$ passes through a loop filter, and the consequent filtered signal is fed to the VCO input $v(t)$. The loop filter is a linear time-invariant system and thus its input-output relation can be described by the convolution formula

$$v(t) = \int_{-\infty}^{\infty} h(t - \tau) e(\tau) d\tau, \tag{7}$$

where $h(t)$ is the impulse response of the loop filter. Note that the loop filter requires design.

3.2 A Simplified Analysis

To understand why and how the PLL works, we resort to analysis. In this regard, it is worthwhile to mention that analysis of the PLL can be a sophisticated topic. Here we give a rough but simplified analysis, with an emphasis on insights. We divide the analysis into three steps.

Step 1) We aim at writing out the relationship between $\phi_1(t)$ and $\phi_2(t)$. Consider the multiplier output

$$\begin{aligned} e(t) &= s(t) \cdot r(t) \\ &= A_c A_v \sin[2\pi f_c t + \phi_1(t)] \cos[2\pi f_c t + \phi_2(t)] \\ &= \frac{A_c A_v}{2} \left\{ \sin[\phi_1(t) - \phi_2(t)] + \sin[4\pi f_c t + \phi_1(t) + \phi_2(t)] \right\}. \end{aligned} \tag{8}$$

The purpose of the multiplier is to help us extract the phase difference $\phi_1(t) - \phi_2(t)$ from the signals. Thus, we may be interested in the first term of (8) only. With this objective this mind, we make the following assumption: *the loop filter is a lowpass filter such that the second term in* (8) *is removed from the loop filter output.* Consequently, the loop filter output $v(t)$ in (7) can be simplified to

$$v(t) = \frac{A_c A_v}{2} \int_{-\infty}^{\infty} h(t-\tau) \sin[\phi_1(\tau) - \phi_2(\tau)] d\tau. \tag{9}$$

Next, we apply another assumption: *the PLL is in near-phase lock, such that $\phi_1(t) - \phi_2(t)$ is small and the term $\sin[\phi_1(t) - \phi_2(t)]$ may be accurately approximated by*

$$\sin[\phi_1(t) - \phi_2(t)] \approx \phi_1(t) - \phi_2(t). \tag{10}$$

By the above approximation, we may further simplify (9) to

$$v(t) = \frac{A_c A_v}{2} \int_{-\infty}^{\infty} h(t-\tau)[\phi_1(\tau) - \phi_2(\tau)] d\tau. \tag{11}$$

Moreover, in the VCO relationship of $v(t)$ and $\phi_2(t)$ in (5), we can differentiate (5) to obtain

$$\frac{d\phi_2(t)}{dt} = 2\pi k_v v(t). \tag{12}$$

By substituting (12) into (11), we obtain

$$\frac{1}{2\pi k_v} \frac{d\phi_2(t)}{dt} = \frac{A_c A_v}{2} \int_{-\infty}^{\infty} h(t-\tau)[\phi_1(\tau) - \phi_2(\tau)] d\tau. \tag{13}$$

In particular, Eq. (13) describes the relation of $\phi_1(t)$ and $\phi_2(t)$ under the aforementioned assumptions and approximation.

Step 2) We wish to understand when $\phi_1(t)$ equals $\phi_2(t)$ under the relation in (13). To do so, we take the Fourier transform of (13) to obtain

$$\frac{1}{2\pi k_v} \times [j2\pi f \cdot \Phi_2(f)] = \frac{A_c A_v}{2} H(f)[\Phi_1(f) - \Phi_2(f)]. \tag{14}$$

Eq. (14) can be reorganized as

$$\Phi_2(f) = \frac{(\pi k_v A_c A_v) \cdot H(f)}{(\pi k_v A_c A_v) \cdot H(f) + j2\pi f} \Phi_1(f) \tag{15}$$

To simplify our analysis, we choose $H(f)$ in the following way: Let W denote the bandwidth of $\phi_1(t)$. Particularly, we assume $\Phi_1(f) = 0$ for all $|f| > W$. The frequency response $H(f)$ is chosen to be such that

$$H(f) = k_\ell, \text{ for all } |f| \leq W, \tag{16}$$

where k_ℓ is a constant and describes the loop filter gain. In words, the loop filter is almost like a straight amplifier for the frequency band of interest, although we should also note that the loop filter has to be lowpass so as to be able to eliminate the second (and high-frequency) term of $e(t)$ in (8). By the above choice of the loop filter, (15) can be reduced to

$$\Phi_2(f) = \frac{K_0}{K_0 + j2\pi f} \Phi_1(f), \tag{17}$$

where $K_0 = \pi k_v A_c A_v k_\ell$. Interestingly, in (17), the term $\frac{K_0}{K_0 + j2\pi f}$ is equivalent to the frequency response of a lowpass filter whose bandwidth increases with K_0. By choosing a sufficiently large K_0 compared to W, which can be made possible by increasing the loop filter gain k_ℓ, we may further reduce (17) to

$$\Phi_2(f) \simeq \Phi_1(f). \tag{18}$$

This means that

$$\phi_2(t) \simeq \phi_1(t), \tag{19}$$

that is, the result we desire to achieve.

Step 3) It is shown in Step 2) that if the loop filter gain k_ℓ is sufficiently large, then we have $\phi_2(t) \simeq \phi_1(t)$. As the last step, we are interested in recovering the message signal from the PLL. It turns out that the loop filter output $v(t)$ already outputs the message signal. To see this, consider (12) under the condition $\phi_2(t) \simeq \phi_1(t)$:

$$\begin{aligned} v(t) &= \frac{1}{2\pi k_v} \frac{d\phi_2(t)}{dt} \\ &\simeq \frac{1}{2\pi k_v} \frac{d\phi_1(t)}{dt} \\ &= \frac{k_f}{k_v} m(t), \end{aligned} \tag{20}$$

where (20) follows from (3).

3.3 Discussions

The PLL considered above, which uses a simple loop filter [See Step 2) in the analysis in the last subsection], is commonly called a *first-order PLL*. The first-order PLL is possibly the simplest form of PLLs, but is seldom used in practice. Better designs for PLLs would consider a more complex loop filter response $H(f)$ to improve performance under practical constraints.

4 Advantages of FM

We describe two advantages for using FM.

4.1 Trading the Transmission Bandwidth for Improved Performance in Noise

In the study of amplitude modulation, one of our interests has been in the transmission bandwidth requirements of the various AM-based schemes. We should first mention that FM has no advantage in terms of bandwidth savings. Carson's rule states that the FM transmission bandwidth approximately equals $B_T = 2W(D+1)$, where W is the message bandwidth and $D = \frac{\Delta f}{W} > 0$ is the deviation ratio. Since $2W(D+1) \geq 2W$, the transmission bandwidth of FM is no better than that of AM.

While FM does not help save transmission bandwidth, FM is known to be able to trade the transmission bandwidth for improved performance against noise. Specifically, given a fixed message bandwidth W, the demodulated signals at the receiver exhibit better suppression of noise as the transmission bandwidth B_T increases. This has been found to be so by experiments (for a long time), and there is mathematical analysis that gives us the clue on why the bandwidth-performance tradeoffs are possible in FM. The details are beyond the scope of this course. At this point, we should mention that practical FM systems are often wideband FM, taking a transmission bandwidth several or many times greater than the message bandwidth. For example, in FM broadcast radios, the transmission bandwidth is roughly about 200 kHz and the message bandwidth is 53 kHz. Also, in Advanced Mobile Phone Service (AMPS), a 1G mobile system in North America, voice transmission is done by FM. The transmission bandwidth is 30 kHz for each voice message signal, and the message bandwidth is 3 kHz.

4.2 Resistance to Amplitude Nonlinearities

FM is a constant envelope transmission scheme. This brings about an advantage, namely, that FM signals are immune to a class of amplitude nonlinearity effects introduced by the transmitter. At the RF stage of the transmitter side, there are cases where the power amplifier exhibits nonlinear effects w.r.t. the signal amplitudes. Such nonlinear effects arise from practical limitations of RF amplifier circuitry and/or considerations of power efficiency in specific applications. Simply speaking, "cheap" transmitters tend to exhibit stronger nonlinear effects. Also, amplifier nonlinearities are not too uncommon in existing systems such as satellite communications. To put into context, consider a special (and artificially created) example where the input-output relation of the nonlinear effects is described by

$$v_o(t) = a_1 v_i(t) + a_2 v_i^2(t) + a_3 v_i^3(t), \tag{21}$$

where $v_i(t)$ and $v_o(t)$ are the input and output, respectively; a_1, a_2 and a_3 are some coefficients that characterize the nonlinearities in (21). In particular, $v_i(t)$ is the modulated signal we desire to transmit, and $v_o(t)$ is the signal actually transmitted by the transmitter. Now, consider the FM signal

$$v_i(t) = A_c \cos[2\pi f_c t + \phi(t)], \tag{22}$$

where $\phi(t) = 2\pi k_f \int_{-\infty}^{t} m(\tau) d\tau$. It can be shown that

$$\begin{aligned} v_o(t) = &\frac{1}{2} a_2 A_c^2 + \left(a_1 A_c + \frac{3}{4} a_3 A_c^3 \right) \cos[2\pi f_c t + \phi(t)] \\ &+ \frac{1}{2} a_2 A_c^2 \cos[4\pi f_c t + 2\phi(t)] + \frac{1}{4} a_3 A_c^3 \cos[6\pi f_c t + 3\phi(t)]. \end{aligned} \tag{23}$$

While the transmitted signal $v_o(t)$ is not identical to the original FM signal, the receiver can easily retrieve the desired FM signal component from (23), i.e., $\left(a_1 A_c + \frac{3}{4} a_3 A_c^3\right) \cos[2\pi f_c t + \phi(t)]$. It can be shown that by applying on $v_o(t)$ a bandpass filter with passband $[f_c - B_T/2, f_c + B_T/2]$ (which will be needed even though we have no nonlinear effects), and by assuming a sufficiently large f_c, the receiver may keep the term $\left(a_1 A_c + \frac{3}{4} a_3 A_c^3\right) \cos[2\pi f_c t + \phi(t)]$ and eliminate other terms in (23). As a result, the receiver may safely demodulate the FM signal without being affected by the amplitude nonlinearities. It is important to note that this is *not* the case with any AM-based scheme: Some simple derivations (which you may try) would show that the nonlinear relation in (21) would distort the envelope of an AM modulated wave (as well as DSB-SC, SSB and so on), and the subsequently, the corresponding demodulated signal.[2]

The discussion above by no means implies that FM is robust against any kind of nonlinearities. FM is sensitive to *phase nonlinearities*, as well as *phase noise*. For example, phase noise may be introduced by the oscillator, owing to device non-ideality and some other reasons. Better designs, in terms of employing better devices and/or more sophisticated systems, would be required to suppress such effects.

[2]It should be mentioned though in some recent systems for non-constant envelope modulation, an idea called *pre-distortion* has been used to neutralize or reduce the amplitude nonlinearity effects.