Chan and Mitran
Angle Modulation

Introduction

Remember that some carrier wave $c(t)$ is a simple sinusoid.

$$c(t) = A_c \cos(\omega_c t + \theta)$$

Like all sinusoids, it's got three parts. An amplitude, a frequency, and a phase. To communicate any information, there has to be some change in some medium that is detectable.

Maybe you're waving a flag, tapping morse code, or flashing a bunch of LEDs. In all of these forms of communication, there is change.

So far, we've only changed one thing: the amplitude. So wouldn't it make sense to try changing frequency or phase, too?

The $(\omega_c t + \theta)$ part is, as a whole, the angle of the sinusoid. So varying ω_c or θ is called angle modulation.

A Throwback to High School Physics

Before we get into the nitty-gritty of angle modulation, it might be helpful to review some physics concepts.

If some object has a time-varying position, $x(t)$, how do we determine its instantaneous velocity?

We look at how far the object has travelled in time Δt, and we divide that distance by Δt.

$$v(t) = \frac{x(t+\Delta t) - x(t)}{\Delta t}$$

Then, we shrink Δt to as small as we can.

$$v(t) = \lim_{\Delta t \to 0} \frac{x(t+\Delta t) - x(t)}{\Delta t}$$

This is, by definition, the derivative with respect to time.

$$v(t) = \frac{d}{dt} x(t)$$

So in reverse, if we wanted to figure out where we are at some time past $t=0$ given our velocity, we take the integral.

$$x(t_0) = x(0) + \int_0^{t_0} v(\tau)\, d\tau$$

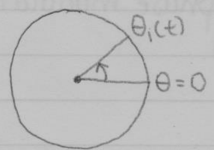
Angle Modulation

By definition, the modulated signal $\phi(t)$ is

$$\phi(t) = A_c \cos(\theta_i(t))$$

where $\theta_i(t)$ represents the instantaneous phase, which is based on our message $f(t)$.

$\theta_i(t)$ is analagous to $x(t)$. $\theta_i(t)$ represents some point on the unit circle in radians. It is a "position" in radians. We can determine the instantaneous angular velocity, aka frequency, in the exact same way we find velocity.



$$\omega_i(t) = \lim_{\Delta t \to 0} \frac{\theta_i(t+\Delta t) - \theta_i(t)}{\Delta t}$$

$$= \frac{d}{dt}\,\theta_i(t)$$

Conversely, we can find the phase at any time t_0 by integrating the frequency.

$$\theta_i(t_0) = \theta_i(0) + \int_0^{t_0} \omega_i(\tau)\,d\tau$$

We can modify one or the other with respect to time, though maybe that's not technically correct to say it that way because they are dependent on each other through the derivative/integral relationship.

What's more correct to say is that we intentionally control one of these variables at a time. Which one we want to control leads to different "types" of angle modulation: phase modulation and frequency modulation.

In each case, the variable in question is changed LINEARLY with $f(t)$, the message.

Phase vs. Frequency Modulation

As the names suggest, $\theta_i(t)$ is varied in phase modulation, and $\omega_i(t)$ is varied in frequency modulation.

For PM:

time, initial phase

$$\theta_i(t) = \omega_c t + \theta_0 + k_p f(t)$$

carrier frequency, "phase sensitivity"

$$\therefore \ \phi_{PM}(t) = A_c \cos [\omega_c t + \theta_0 + \underbrace{k_p f(t)}_{\text{changing phase with respect to time}}]$$

For FM:

carrier frequency, "frequency sensitivity"

$$\omega_i(t) = \omega_c + k_f f(t)$$

However, we need $\theta_i(t)$ to write $\phi_{FM}(t)$, as that's how angle modulation is defined. To do this, we integrate $\omega_i(t)$.

$$\theta_i(t) = \theta_0 + \int_0^t \omega_i(t)\, dt$$

$$= \theta_0 + \int_0^t \omega_c + k_f f(t)\, dt$$

$$= \theta_0 + \omega_c t + \int_0^t k_f f(t)\, dt$$

$$\therefore \ \phi_{FM}(t) = A \cos [\omega_c t + \theta_0 + \underbrace{k_f \int_0^t f(t)\, dt}_{\text{changing phase with respect to time}}]$$

So both frequency and phase modulation end up with some time-dependent phase change. Are they really that different?

In fact, could we create a frequency modulated signal using a phase modulator, and/or create a phase modulated signal using a frequency modulator?

We can: check it out.

$$f(t) \longrightarrow \boxed{\text{Integrator}} \longrightarrow \underset{\substack{\uparrow \\ k_I}}{\otimes} \longrightarrow \underset{(k_p)}{\boxed{\text{Phase Modulator}}} \longrightarrow \Phi(t)$$

Here, we're creating an FM signal using a phase modulator.

① $f(t) \rightarrow \text{integrator} = \int_0^t f(t)\,dt$

② $\rightarrow \underset{\substack{\uparrow \\ k_I}}{\otimes} = k_I \int_0^t f(t)\,dt$, where k_I is some arbitrary constant

③ $\rightarrow \text{PM w/ } k_p = A_c \cos\left[\omega_c t + \left(\theta_0 + k_p\left[k_I \int_0^t f(t)\,dt\right]\right)\right]$

$$\Phi(t) = A_c \cos\left[\omega_c t + \theta_0 + k_p k_I \int_0^t f(t)\,dt\right]$$

This is FM, where the frequency sensitivity $k_f = k_p k_I$.
Let's do the opposite.

$$f(t) \longrightarrow \boxed{\text{Differentiator}} \longrightarrow \underset{\substack{\uparrow \\ k_D}}{\otimes} \longrightarrow \underset{(k_f)}{\boxed{\text{Freq Modulator}}} \longrightarrow \Phi(t)$$

① $f(t) \rightarrow \text{differentiator} = \frac{d}{dt} f(t)$

② → $\otimes$ $= k_D \frac{d}{dt} f(t)$, where k_D is some arbitrary constant
$\uparrow$
k_D

③ → FM w/ k_f $= A_c \cos\left[\omega_c t + \theta_0 + k_f \int_0^t \left(k_D \, {}^d\!/\!_{dt} f(t)\right) dt\right]$

$$= A_c \cos\left[\omega_c t + \theta_0 + k_f k_d f(t)\right]$$

This is PM, where the phase sensitivity $k_p = k_f k_d$. So we can trivially create both modulated signals using the others' modulator.

FM with input $f(t)$ = PM with input $k_I \int_0^t f(t)\,dt$

FM with input $k_D \frac{d}{dt} f(t)$ = PM with input $f(t)$

Since these are so similar, it makes sense to focus on just one - we'll be doing frequency modulation, because it's more widely used in the real world.

So, before we head into the next section on types of FM, we'll make a few assumptions about the kinds of messages we want to modulate.

1) The maximum height of $f(t)$ is the same magnitude of the minimum.

$$\max f(t) = -\min f(t)$$

2) We also assume this is true of their integrals.

$$\max \int_0^t f(\tau)\,d\tau = -\min \int_0^t f(\tau)\,d\tau$$

Since both $\theta_i(t)$ and $\omega_i(t)$ are related linearly to $f(t)$,

3) The change in frequency is limited by the maximum and minimum of $f(t)$.

$$\omega_i(t) = \omega_c + k_f f(t) \rightarrow \Delta\omega = k_f \left[\max f(t)\right] = k_f \left[-\min f(t)\right]$$

This is the peak frequency deviation.

4) The change in phase is therefore limited by the maximum and minimum of the integral of $f(t)$, as phase is the integral of frequency.

$$\theta_i(t) = \omega_c t + k_p \int_0^t f(\tau)\,d\tau \rightarrow \beta = k_f \left[\max \int_0^t f(\tau)\,d\tau\right] = -k_f \left[-\min \int_0^t f(\tau)\,d\tau\right]$$

This is the peak phase deviation, aka the modulation index.

If: $\beta \ll 1$, this is called narrowband FM.
$\beta > 1$, this is called wideband FM.

This leads us to our next chapter: narrowband and wideband FM.