Chan and Mitran
Amplitude Modulation

Introduction

Welcome one, welcome all. Today's topic is amplitude modulation. The initials - AM - may ring a bell here. It is this exact technology or concept that AM radio uses. First, we need some definitions.

Baseband - this is the frequencies the original signal occupies
- for example, the baseband of a human voice would be approximately [300 Hz, 3500 Hz]

Carrier signal - usually a sinusoid
- we'll change how the sinusoid behaves so it can "carry" our message

Modulation - the process by which the carrier signal is changed with respect to the original signal

$$f(t) \longrightarrow \boxed{\text{Modulator}} \longrightarrow \Phi(t)$$

$$\uparrow$$

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

$c(t)$, the carrier signal, with amplitude A_c and frequency ω_c, is sent into a modulator with $f(t)$. $c(t)$ gets modulated in some way by $f(t)$, producing $\Phi(t)$, our amplitude-modulated, or AM signal.

We'll assume the following of $f(t)$:

1) Min freq of $f(t)$ = − Max freq of $f(t)$.

This is because real-valued signals have mirrored frequencies: if a frequency exists in the positive region, a negative one equal in magnitude will also exist.

$$\cos(\theta) = \frac{e^{j\theta} + e^{-j\theta}}{2}$$

$$= \underbrace{\frac{e^{j(\theta)}}{2}}_{\text{positive part}} + \underbrace{\frac{e^{j(-\theta)}}{2}}_{\text{equal but opposite negative part}}$$

2) $$\overline{f(t)} = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(t)\, dt$$

$$= 0$$

The average value of $f(t)$, denoted in shorthand by $\overline{f(t)}$, is 0. It's actually not excruciatingly difficult to take into account amplitude shifts, but we'll use a simpler case to illustrate the concept.

Double Sideband − Large Carrier (DSB − LC)

This is the first modulation strategy we'll learn, simply because of how simple it is.

$$\Phi(t) \triangleq A_c \left(1 + k_a f(t)\right) \cos(\omega_c t)$$

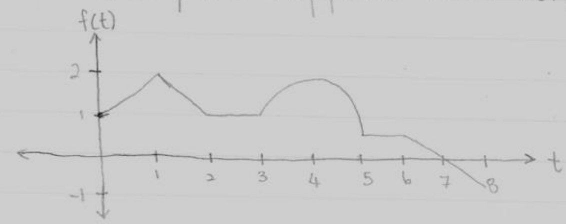
K_a is a constant denoting "amplitude sensitivity".

$$= [A_c + A_c K_a f(t)] \cos(\omega_c t)$$

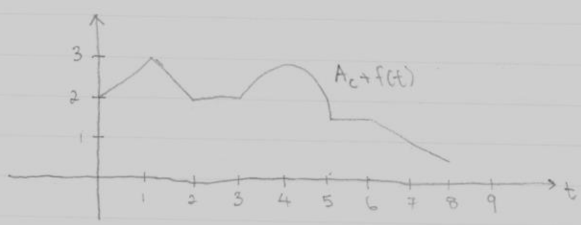
But honestly it's easier to pretend that the original signal f(t) is just the already transformed $A_c K_a f(t)$. That is, we'll just say

$$f(t) = A_c K_a f(t)$$
$$\therefore \Phi(t) = (A_c + f(t)) \cos(\omega_c t)$$

So what exactly has happened here? Let's graph it out.

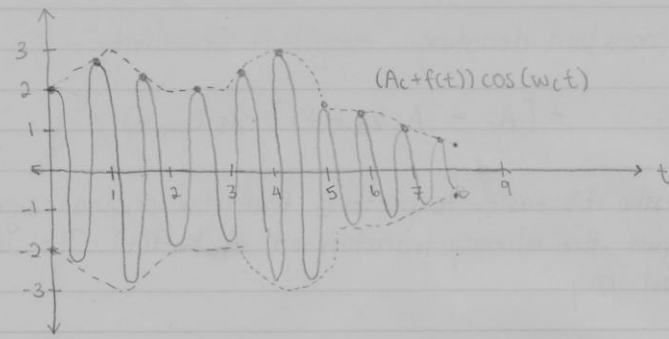


Let's pretend this weird signal is f(t). Let's pretend $A_c = 1$, so $A_c + f(t)$ is just f(t) shifted up by 1.



What happens when we multiply by $\cos(\omega_c t)$? The min and max of cos are -1 and 1. As such, we can never go above $+(A_c + f(t))$ or below $-(A_c + f(t))$, and we'll have the standard sinusoidal action between the extremes.

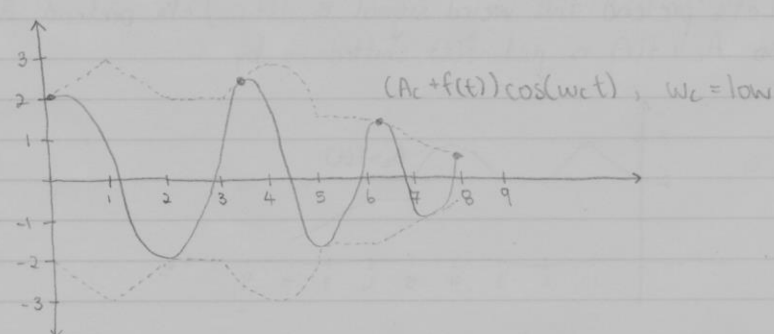
We'll say ω_c = very high. I'll explain this later



This is the signal we'd actually send out. So how do we get back our original signal? All we'd need to do is interpolate the peaks of $\phi(t)$.

Which is just a fancy way of saying "let's trace the top of the wave". The proper terminology is similar. This is called "tracing the envelope of $\phi(t)$".

Notice what happens if w_c is not "very high".

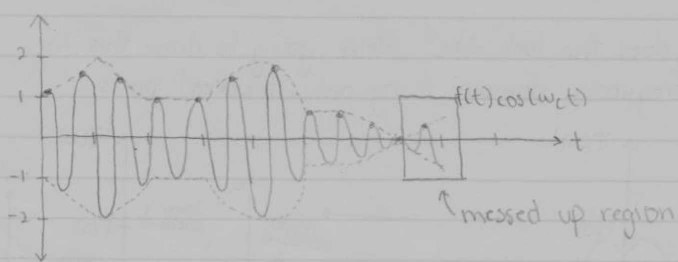


We get just as bad of an approximation of $\phi(t)$ as my drawing of this sinusoid. As such, w_c MUST be much greater than the highest frequency in the Fourier expansion of $f(t)$.

We don't really quantify this, so it's not exact how much larger ω_c should be.

So double sideband - large carrier shifts the wave upward, and then uses its extremes to create a sinusoid. But how much do we have to shift by? What happens if we don't shift enough?

In this extremely convoluted example, we'll say $A_c = 0$, so $\Phi(t) = f(t)\cos(\omega_c t)$.



For the most part, it seems okay. But in the last part, the envelope tracing will see a positive amplitude, whereas in the original, it goes negative. This mistake is known as "envelope distortion", due to "overmodulation".

We can conclude from this that

1) If $A_c + f(t) > 0$, the envelope is approximately the same as $f(t)$, as long as ω_c is high enough

2) If $A_c + f(t) < 0$ at any point, we get envelope distortion from overmodulation.

Like good math students, let's take a look at this in the frequency domain.

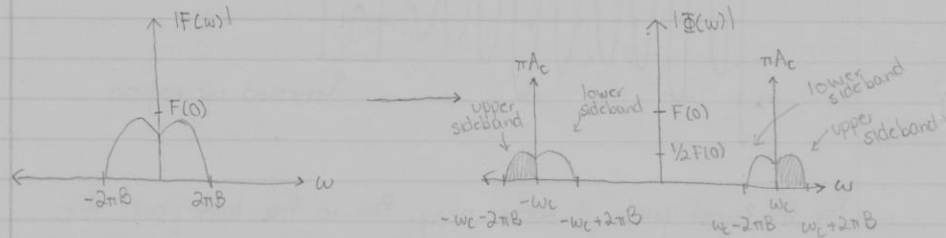
$$\Phi(t) = \mathcal{F}\{\phi(t)\}$$
$$= \mathcal{F}\{[A_c + f(t)]\cos(\omega_c t)\}$$
$$= \mathcal{F}\{A_c \cos(\omega_c t)\} + \mathcal{F}\{f(t)\cos(\omega_c t)\}$$

Remember our property from the introduction - that multiplying something by cos splits the spectrum in two halves, and shifts them to the right and left:

$$= \pi A_c\, \delta(\omega - \omega_c) + \pi A_c\, \delta(\omega + \omega_c)^{*}$$
$$+ \tfrac{1}{2}F(\omega - \omega_c) + \tfrac{1}{2}F(\omega + \omega_c)$$

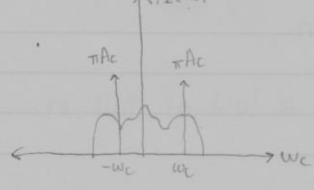
*should this be $A_c/2$?

What does this look like? We're going to draw this in the frequency domain using only the real parts.



We've just used some arbitrary shape as our original baseband spectrum. We've halved the amplitude, shifted the centers by $\pm \omega_c$, and added delta functions of amplitude πA_c there.

The new bandwidth exists at $[\omega_c - 2\pi B, \omega_c + 2\pi B]$, and is $4\pi B$ in length. Each individual sideband does not overlap with the other as long as we've moved them far enough away: $2\pi B$ or greater.



If they're too close together, the overlapping messes with the spectrum, and we aren't able to recover the baseband anymore.

Modulation Index

How can we tell if we modulate enough? To know that, we'll have to define the modulation index. We'll assume the max height of a signal $f(t)$ equals its min height.

DSB-LC is given by:

$$\phi(t) = [A_c + f(t)]\cos(\omega_c t)$$
$$= A_c\,[1 + f(t)/A_c]\cos(\omega_c t)$$

Since

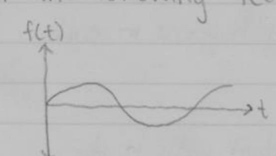
$$f(t)_{min} \le f(t) \le f(t)_{max}$$
$$\frac{f(t)_{min}}{A_c} \le \frac{f(t)}{A_c} \le \frac{f(t)_{max}}{A_c}$$

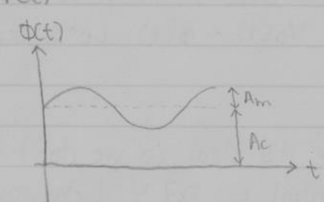
We'll call $|f(t)_{min}| = |f(t)_{max}| = A_m$ for brevity

$$-\frac{A_m}{A_c} \le \frac{f(t)}{A_c} \le \frac{A_m}{A_c}$$
$$1 - \frac{A_m}{A_c} \le 1 + \frac{f(t)}{A_c} \le 1 + \frac{A_m}{A_c}$$

We can then say $\mu = \frac{A_m}{A_c}$ is the modulation index itself.

For the following $f(t)$ and $\phi(t)$:





μ relates the amplitude of the signal with how much the carrier signal shifts it upward

Remember that overmodulation occurs when we don't shift $f(t)$ high enough. That is, when

$$\text{amount of shift} < \text{largest amplitude of } f(t)$$
$$A_c < A_m$$

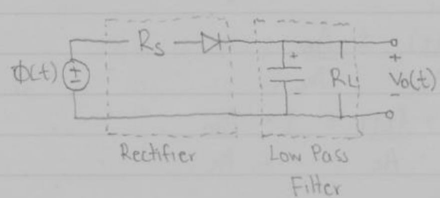
Therefore, we can conclude as long as $A_c \geq A_m$, or

$$\mu = \frac{A_m}{A_c} \leq 1$$

We are undermodulated as long as $\mu \leq 1$.

<u>Demodulating Double Sideband - Large Carrier</u>

THIS IS NOT NEEDED FOR EXAMINATIONS. We're going to show how simple it is to create an envelope detector, and why this is the first modulation concept we learn.

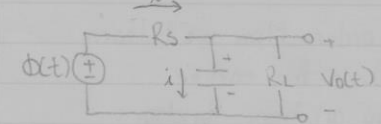


$R_L >> R_s$

This is it. Four parts. $V_o(t)$ is the final, demodulated signal. $\phi(t)$ is our input, the modulated signal. The diode is off (open circuit) when $V_o(t) > \phi(t)$. It is on (short circuit) when $V_o(t) < \phi(t)$. Let's explore what happens in each case.

Assume it's ideal so we don't have to deal with nasty exponential or 0.7 V behaviour.

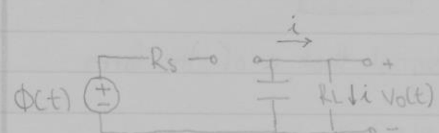
1) $V_o(t) < \Phi(t)$: diode ON.



Since R_L is very large, current will travel through the capacitor, charging it up. This increases the voltage across it, which is exactly increasing $V_o(t)$.

Once $V_o(t)$ increases enough, it'll enter the second state.

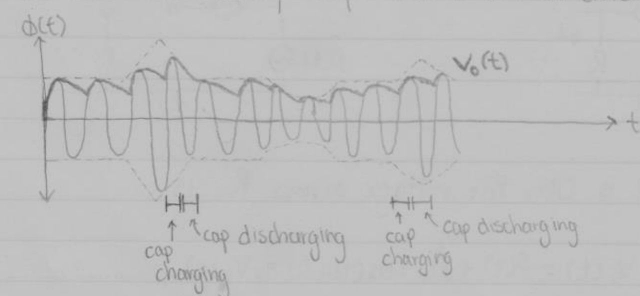
2) $V_o(t) > \Phi(t)$: diode OFF



The capacitor begins to discharge, sending current upward through R_L, as it has no other path to go. As such, $V_o(t)$ falls.

Once $V_o(t)$ decreases enough, it'll enter the first state.

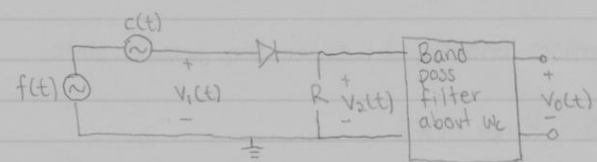
How does this actually help? How is this an envelope detector?



The capacitor's continuous charge-discharge cycles allow it to somewhat follow the peaks of $\Phi(t)$. All we need to do is to pinpoint what time constant $\tau = R_s C$, we'd need to best track $\Phi(t)$.

Generating Double Sideband Large Carrier

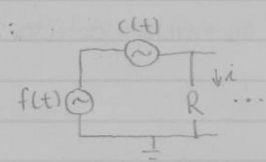
We kind of did this in the opposite order. Now we'll learn how to construct $\Phi(t)$ out of $f(t)$. In the earlier analog circuitry days, multiplying two arbitrary functions is difficult. So here's what we're going to do instead.

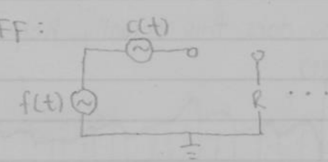


$V_0(t)$ is going to be $\Phi(t)$: a signal that is of the form

$$\Phi(t) = [X A_c + Y f(t)] \cos(\omega_c t)$$

where X and Y are some arbitrary constants. Let's see why this is true. Again, we've got an ideal diode.





When the diode is ON, the voltage across R, is

$$V_1(t) = f(t) + A_c \cos(\omega_c t) = V_2(t)$$

When it's OFF, R no longer receives current, as it goes directly into ground. As such,

$$V_2(t) = 0$$
$$V_1(t) = f(t) + A_c \cos(\omega_c t).$$

So now, the question is: when is the diode on or off?
We'll assume $|f(t)| << A_c$, so the voltage $v_1(t)$ will largely be dominated by $c(t)$.

Remember that a diode is only ON when the voltage across it is >0. So we can say that the diode is ON when $c(t) > 0$ and off otherwise.

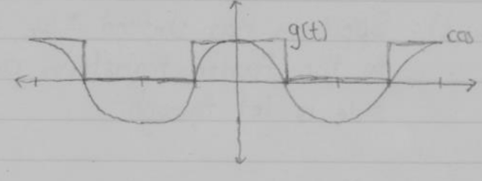
Let's define the diode's behaviour as $g(t)$.

$$g(t) = \begin{cases} \text{ON} & c(t) > 0 \\ \text{OFF} & c(t) \leq 0. \end{cases}$$

Since $v_2(t)$'s voltage is dictated by the behaviour of the diode, we can say that

$$\begin{aligned} V_2(t) &= \begin{cases} v_1(t) & g(t) = \text{ON} \\ 0 & g(t) = \text{OFF} \end{cases} \\ &= \begin{cases} v_1(t) & c(t) > 0 \\ 0 & c(t) \leq 0 \end{cases} \\ &= [f(t) + A_c \cos(\omega_c t)]\, g(t) \end{aligned}$$

What IS $g(t)$? $g(t)$ is a half-wave rectifier. It turns any negative amplitudes to zero. $g(t)$ is essentially a periodic rectangular wave.



It has an amplitude of $1/2$ and a DC shift upward of $1/2$.

And because it's periodic, it's got a Fourier expansion:

$$g(t) = \frac{1}{2} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{2n-1} \cos[\omega_c t (2n-1)]$$

How we derived this is inconsequential. What is of importance, though, is what happens when we send it through the bandpass filter centered at ω_c. Let's expand.

$$\begin{aligned} g(t) = & \ \tfrac{1}{2} f(t) \\ & + \tfrac{1}{2} A_c \cos(\omega_c t) \\ & + \tfrac{2}{\pi} f(t) \cos(\omega_c t) \\ & + \tfrac{2}{\pi} A_c \cos(\omega_c t) \cos(\omega_c t) \\ & - \tfrac{2}{3\pi} f(t) \cos(3\omega_c t) \\ & - \tfrac{2}{3\pi} A_c \cos(\omega_c t) \cos(3\omega_c t) \\ & + \tfrac{2}{5\pi} f(t) \cos(5\omega_c t) \\ & + \tfrac{2}{5\pi} A_c \cos(\omega_c t) \cos(5\omega_c t) \\ & - \ldots \end{aligned}$$

Now, we can look at each portion of $g(t)$ and see if it passes through the filter.

$\frac{1}{2} f(t)$: Frequency of 0.* Not close to ω_c so filtered out.

$\frac{1}{2} A_c \cos(\omega_c t)$: Right on ω_c, so it gets let through.

$\frac{2}{\pi} f(t) \cos(\omega_c t)$: Spectrum has shifted $\pm \omega_c$ in accordance to the Fourier transform property. $+\omega_c$ side is let through.

*not necessarily, but we assume the spectrum is centered at 0.

$2/\pi\, A_c \cos(w_c t)\cos(w_c t)$: Can be rewritten as $2/\pi\, A_c\,[1/2 + 1/2\cos(2w_c t)]$.
A_c/π is DC, so it gets filtered out.
$A_c/\pi \cos(2w_c t)$ is $2 \times w_c$, which is pretty far as well. Filtered out.

$-2/3\pi\, f(t)\cos(3w_c t)$: $3w_c$ is really far. Filtered out.

$-2/3\pi\, A_c \cos(w_c t)\cos(3w_c t)$: Rewritten as $-2/3\pi\, A_c\,[1/2\cos(-2w_c t) + 1/2\cos(4w_c t)]$. Both frequencies are far from w_c, so they get filtered.

If we continue, things just get further and further, so they all get filtered out. So what are we left with?

$$V_o(t) = 1/2\, A_c \cos(w_c t) + 2/\pi\, f(t)\cos(w_c t)$$
$$= [1/2\, A_c + 2/\pi\, f(t)]\cos(w_c t)$$
is of the form
$$= [X A_c + Y f(t)]\cos(w_c t)$$
$$= \phi(t)$$

$V_o(t)$ is $\phi(t)$, the double sideband large-carrier modulated signal. In order to ensure we keep the original frequencies of f, the bandpass filter's bandwidth must simply be 2 times the highest frequency of f.

Power Efficiency

Now we know how to modulate and demodulate DSB-LC signals. But is this method an efficient use of our resources?

Remember that

$$P = \lim_{T\to\infty} 1/T \int_{-T/2}^{T/2} |\phi^2(t)|\, dt$$

and that we use $\overline{f(t)} = \lim_{T\to\infty} \int_{-T/2}^{T/2} f(t)\, dt$ as shorthand.

↑ the average

$$\therefore P = \overline{\phi^2(t)}$$

$$= \overline{[A_c + f(t)]^2 \cos^2(\omega_c t)}$$

$$= \overline{A_c^2 \cos^2(\omega_c t)} + \overline{2A_c f(t) \cos^2(\omega_c t)}$$

$$+ \overline{f(t)^2 \cos^2(\omega_c t)}$$

Let's split these up so it's easier to handle.

$$\overline{A_c^2 \cos^2(\omega_c t)} = \underbrace{\overline{A_c^2}}_{\text{constant}} (\underbrace{1/2}_{\text{constant}} + 1/2 \underbrace{\overline{\cos(2\omega_c t)}}_{0})$$

$$= \frac{A_c^2}{2}$$

$$\overline{2A_c f(t) \cos^2(\omega_c t)} = \overline{2A_c}\; \overline{f(t)}\; \overline{[1/2 + 1/2 \cos(2\omega_c t)]}$$

$$= A_c \underbrace{\overline{f(t)}}_{0} + A_c \underbrace{\overline{f(t)}}_{0} \underbrace{\overline{\cos(2\omega_c t)}}_{0}$$

$$= 0$$

We assume the average of the original signal, $f(t)$, is 0. This is because we want this signal to use as little power as possible to send to look at our best-case power usage.

Since at the end we look at power of $f(t)$ vs. power of $c(t)$, giving $f(t)$ an average $\neq 0$ would skew the scenario.

$$\overline{f^2(t)\cos^2(\omega_c t)} = \overline{\tfrac{1}{2} f^2(t)} + \overline{\tfrac{1}{2} f^2(t)\cos(2\omega_c t)}$$

Here, we'll make another assumption, a little more realistic. We want ω_c to be much higher than $f(t)$'s max frequency so it tracks $f(t)$ well. So $f(t)$ is very slow in comparison - slow enough to be constant for one entire cycle of $c(t)$.

As such, the average of functions like $f^2(t)\cos(2\omega_c t)$ can be assumed to simply be the product of the two.

$$= \tfrac{1}{2}\,\overline{f^2(t)} \quad \tfrac{1}{2}\,\underset{\uparrow \neq 0}{\overline{f^2(t)}}\;\underset{\uparrow 0}{\overline{\cos(2\omega_c t)}}$$

$$= \tfrac{1}{2}\,\overline{f^2(t)}$$

Notice that squaring a function can make its average non-zero. $\cos(t)$'s average is 0 but $\cos^2(t)$'s average is $\tfrac{1}{2}$.

So overall, we're left with

$$P = \underset{\substack{\uparrow \\ \text{power} \\ \text{due to carrier}}}{\tfrac{1}{2} A_c^2} + \underset{\substack{\nwarrow \\ \text{power due} \\ \text{to original}}}{\tfrac{1}{2}\,\overline{f(t)^2}}$$

We can then define modulation efficiency as:

$$\xi \triangleq \frac{\text{sideband power}}{\text{total power}} = \frac{\overline{f^2(t)}}{\overline{f^2(t)} + A_c^2}$$

Carrier power is considered to be wasted because it's the mechanism by which we deliver the message itself. Ideally, we'd expend only the energy needed to convey the message, not deliver it.

To make an analogy - suppose I want to write a letter to someone. I have to spend energy writing it. What I would rather not do, though, is get the mailman to deliver it. Having the mailman deliver is carrier power. It would be ideal if I could use the energy I expended to write the letter to also transport it to where you are instead

Anyway, we're getting off topic. So what is the best case modulation efficiency? We'll use $f(t) = A_m \cos(\omega_m t)$, a signal with zero power.

$$\overline{f^2(t)} = \lim_{T\to\infty} A_m^2/T \int_{-T/2}^{T/2} \cos^2(\omega_m t)\, dt$$

$$= \frac{A_m^2}{2}$$

$$\therefore \xi = \frac{A_m^2/2}{A_m^2/2 + A_c^2}$$

$$= \frac{\frac{A_m^2}{2A_c^2}}{\frac{A_m^2}{2A_c^2} + 1}, \text{ where } \frac{A_m}{A_c} = \mu$$

$$= \frac{\mu^2/2}{\mu^2/2 + 1}$$

Since μ must be ≤ 1 to not be overmodulated:

$$= \frac{1/2}{1/2 + 1}$$

$$= \frac{1}{3}$$

At best, DSB-LC has a power-efficiency of 1/3. Pretty terrible.