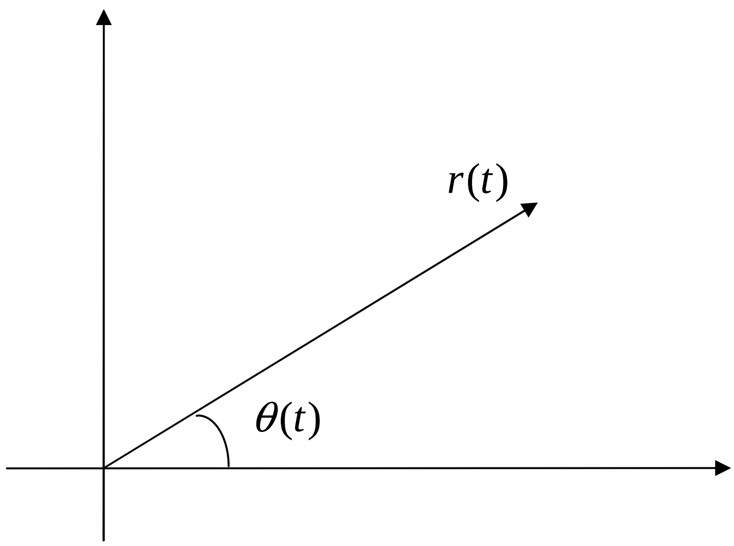
Exponential modulation

Instantaneous frequency

Up to now, we have defined the frequency as the speed of rotation of a phasor (constant frequency phasor) $\phi(t) = A_0 \exp\left[j\left(\omega_0 t + \theta_0\right)\right]$. We are going to generalize this definition to general complex functions of the real variable *t*. Consider such function: $z(t) = r(t)\exp\left(j\theta(t)\right)$. It is obvious that it is a generalization of the constant frequency phasor. It can be represented graphically as a vector with modulus (amplitude) $r(t)$ and argument $\theta(t)$.



The argument $\theta(t)$ of $z(t)$ is called the instantaneous phase. Since this phase varies, the generalized phasor is going to rotate. However, it is not going to rotate at a constant speed. We can thus define an instantaneous speed of rotation for this function. It is the instantaneous frequency:

$$\omega(t) = \frac{d\theta}{dt} \text{ rd/s}$$

We can measure this frequency in Hertz:

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

In this section, we are interested in constant amplitude phasors. Furthermore, we assume that the instantaneous frequency has an average value f_0 with a deviation around it $d(t)$:

$$f(t) = f_0 + d(t)$$

The average value of $d(t)$ is zero. This means that the instantaneous phase can be expressed as:

$$\theta(t) = 2\pi f_0 t + \phi(t) = \omega_0 t + \phi(t)$$

$\phi(t)$ is called the instantaneous phase deviation and we have:

$$d(t) = \frac{1}{2\pi}\frac{d\phi}{dt}$$

To this generalized phasor, we can associate the following signal:

$$x(t) = \mathrm{Re}\left[w(t)\right] = r(t)\cos\left(\omega_0 t + \phi(t)\right)$$

This signal has the general shape of a bandpass signal. In order for $x(t)$ to be bandpass, its quadrature components must be bandlimited to a frequency $W < f_0$. The quadrature components are:

$$a(t) = r(t)\cos\phi(t) \text{ and } b(t) = r(t)\sin\phi(t).$$

This condition is not always satisfied. However, in practical exponential modulation, the carrier f_0 is usually very high (hundreds of MHz). So, we can consider that the obtained modulated signals are bandpass signals.

Frequency and Phase Modulation (FM & PM)

For both phase and frequency modulation, the modulated signal must have constant amplitude. The information is carried in the phase deviation. These modulations are called "*exponential Modulation*" because the signal has always the following shape:

$$x(t) = \mathrm{Re}\left[A_0 \exp\left(j\phi(t)\right)\exp\left(j\omega_0 t\right)\right]$$

Phase modulation is a modulation process that makes the phase deviation $\phi(t)$ proportional to the baseband signal $s(t)$.

$$\phi(t) = k_\phi \tilde{s}(t) = (\Delta\phi)s(t)$$

The constant $\Delta\phi = k_\phi \left|\tilde{s}(t)\right|_{\max}$ is called the maximum phase deviation. In order to avoid phase ambiguity, this constant cannot exceed π.

$$0 \le \Delta\phi \le \pi$$

The expression of a real phase modulated signal is:

$$x(t) = A_0 \cos\left(\omega_0 t + (\Delta\phi)s(t)\right)$$

Frequency modulation, on the other hand, is a modulation system where the frequency deviation is made proportional to the information signal.

$$d(t) = k_f \tilde{s}(t) = (\Delta f)s(t)$$

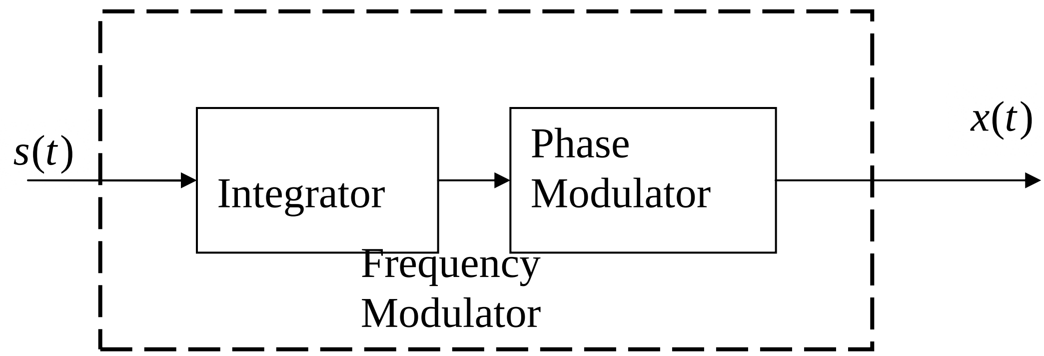
The constant $\Delta f = k_f \left|\tilde{s}(t)\right|_{\max}$ is called the maximum frequency deviation. The instantaneous frequency is:

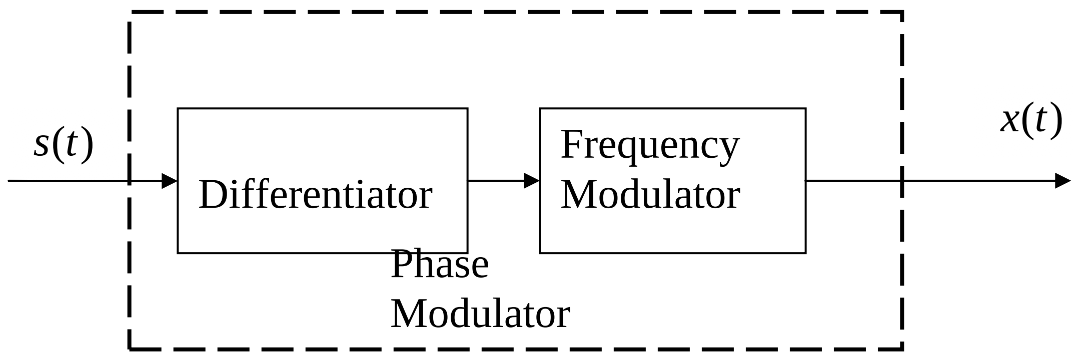
$$f(t) = f_0 + (\Delta f)s(t)$$

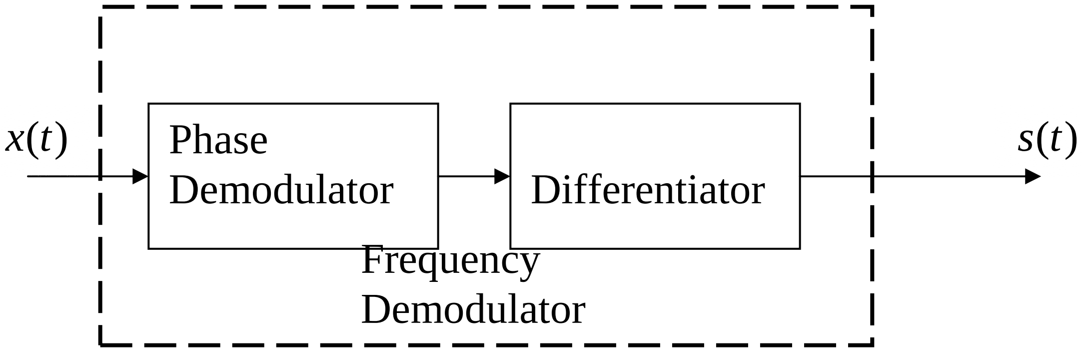
In order to have always a positive frequency, we must have $\Delta f \le f_0$. The instantaneous phase deviation is: $\phi(t) = 2\pi(\Delta f)\int^t s(\lambda)d\lambda$. The lower bound of the integral is not indicated to take into account any initial phase. Sometimes, the lower bound is assumed to be $-\infty$. So, the expression of a frequency modulated signal is:

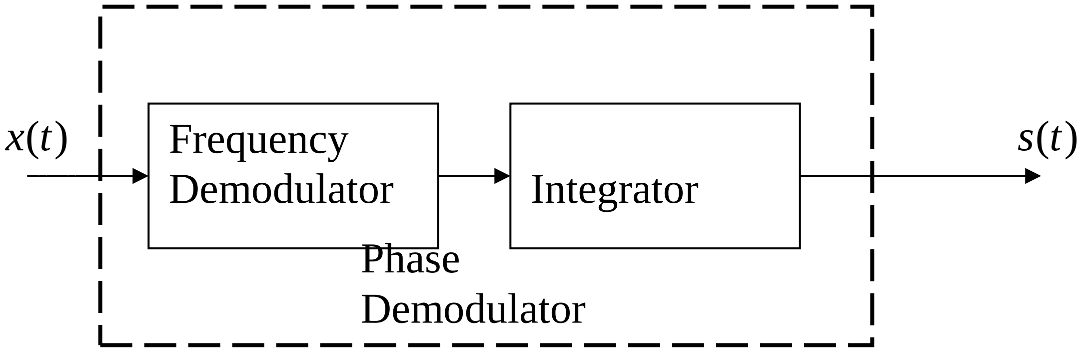
$$x(t) = A_0 \cos\left(\omega_0 t + 2\pi(\Delta f)\int^t s(\lambda)d\lambda\right)$$

If we look at the relations that exist between the phase and the frequency, we remark that the two modulations are related. In fact, we can built a frequency modulator using a phase modulator, a frequency demodulator using also a phase demodulator and vice versa.









The above four figures show how we can build one type of modulator or demodulator using the other.

Exponential modulation is a highly nonlinear modulation. This means that it is very hard to relate the spectrum of the modulated waveform with the one of the baseband as we did with the linear modulations. So, an analysis of the modulated signal in the frequency domain is quite difficult in the general case.

There are two special cases where this analysis is not very complicated: the narrowband phase and frequency modulation where the phase deviation is very small and the sinusoidal modulation where the baseband signal is sinusoidal.

Narrowband phase and frequency modulation

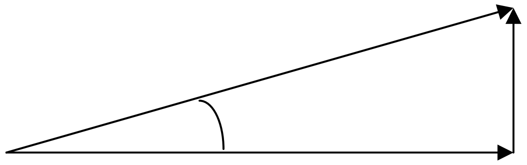
The modulated signal in this case has the general shape of:

$$x(t) = A_0 \cos\left(\omega_0 t + \phi(t)\right) \text{along with } \left|\phi(t)\right|_{\max} << 1$$

Developing the cosine, $x(t)$ becomes:

$$x(t) = A_0 \cos\phi(t)\cos\omega_0 t - A_0 \sin\phi(t)\sin\omega_0 t$$

Using the fact that $\left|\phi(t)\right|_{\max} << 1$, we have: $\cos\phi(t) \approx 1$ and $\sin\phi(t) \approx \phi(t)$, giving: $x(t) = A_0 \cos\omega_0 t - A_0\phi(t)\sin\omega_0 t$.



The above phasor diagram illustrates that the signal x(t) is the projection on the real axis of the sum of two phasors rotating at the same speed ω_0 and making an angle of 90° between them. We see that this phase shift produces the phase modulation. Furthermore, the Fourier transform of the expression of $x(t)$ can be evaluated. So, if $\Phi(f) = \mathcal{F}[\phi(t)]$, then

$$X(f) = \frac{A_0}{2}[\delta(f - f_0) + \delta(f + f_0)] - \frac{A_0}{2j}[\Phi(f - f_0) - \Phi(f + f_0)].$$

If the signal is PM, then $\phi(t) = (\Delta\phi)s(t)$ giving $\Phi(f) = (\Delta\phi)S(f)$. So, if the signal is bandlimited to W, then the PM signal will be limited to a bandwidth $B = 2W$.

If the signal is FM, then $\phi(t) = 2\pi(\Delta f)\int^t s(\lambda)d\lambda$ giving $\Phi(f) = \frac{(\Delta f)}{jf}S(f)$. Here also the bandwidth of the FM signal is $2W$.

Sinusoidal modulation

The other case that has a simple analytic expression is when the modulating signal is sinusoidal. When a signal is sinusoidal, its derivative is also sinusoidal. So, we can use the same analysis for both frequency and phase modulation. The modulated signal in both cases is $x(t) = A_0 \cos\left(\omega_0 t + \phi(t)\right)$. $\phi(t) = (\Delta\phi)s(t)$ for PM and $\phi(t) = 2\pi(\Delta f)\int^t s(\lambda)d\lambda$ for FM.

For FM modulation, we assume that $s(t) = \cos\omega_m t$. This gives:

$\phi(t) = \dfrac{2\pi(\Delta f)}{\omega_m}\sin\omega_m t$. $\beta = \dfrac{\Delta f}{f_m}$ is called the modulation index. So, $x(t) = A_0 \cos\left(\omega_0 t + \beta\sin\omega_m t\right)$.

For PM modulation, the modulating signal is $s(t) = \sin\omega_m t$. The instantaneous phase deviation becomes: $\phi(t) = (\Delta\phi)\sin\omega_m t$. In this case, the modulation index is $\beta = (\Delta\phi)$ and we obtain the same expression. So, for both cases, the expression of the modulated signal is:

$$x(t) = A_0 \cos\left(\omega_0 t + \beta\sin\omega_m t\right) = A_0 \operatorname{Re}\left[\exp\left(j\beta\sin\omega_m t\right)\exp\left(j\omega_0 t\right)\right]$$

In the above expression, the function $\exp\left(j\beta\sin\omega_m t\right)$ is periodic with a period $T_m = \dfrac{2\pi}{\omega_m}$. It can be developed in Fourier series. The development is:

$$\exp\left(j\beta\sin\omega_m t\right) = \sum_{n=-\infty}^{+\infty} J_n(\beta)\exp\left(jn\omega_m t\right)$$

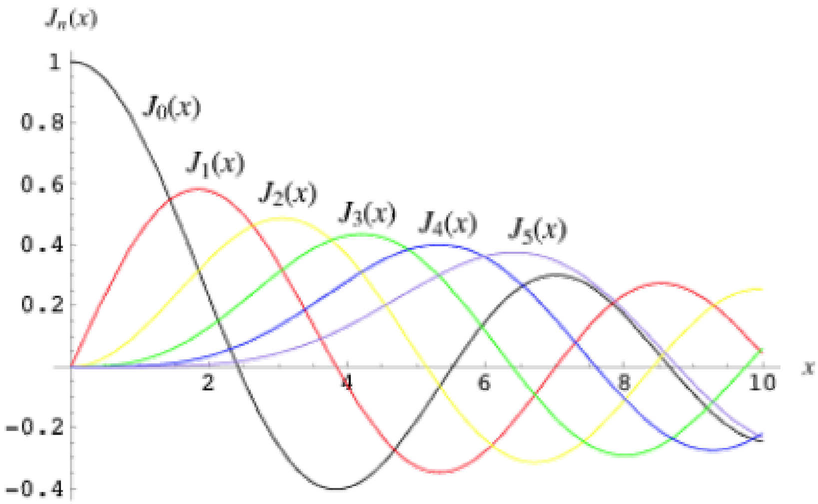
The Fourier coefficients $J_n(\beta)$ are the Bessel functions of the first kind of order *n* and argument β. These functions are tabulated and can be easily computed. They appear as solutions of differential equations. For positive order, we can use the following Mc Lauren series:

$$J_n(\beta) = \left(\frac{\beta}{2}\right)^n \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(n+k)!} \left(\frac{\beta}{2}\right)^k$$

and when *n* is negative, we use the following relation:

$$J_{-n}(\beta) = (-1)^n J_n(\beta)$$

The following figure shows the behavior of the first 6 Bessel functions.



They look like damped sinewaves. From the Mc Lauren series we can deduce their properties for β around 0. For very small values of β, we have the following approximations:

$$J_0(\beta) \approx 1$$

$$J_n(\beta) \approx \frac{1}{n!}\left(\frac{\beta}{2}\right)^n \quad \text{for } n > 0$$

This means that the only functions that we should consider around zero are J_0 and J_1. So, for $\beta < 0.1$, $J_0(\beta) \approx 1$ and $J_1(\beta) \approx \frac{\beta}{2}$.

The following table gives the value of the first Bessel functions.

				$J_n(x)$					
x / **n**	**0.5**	**1**	**2**	**3**	**4**	**6**	**8**	**10**	**12**
0	0.9385	0.7652	0.2239	- 0.2601	-0.3971	0.1506	0.1717	-0.2459	0.0477
1	0.2423	0.4401	0.5767	0.3391	-0.0660	-0.2767	0.2346	0.0435	-0.2234
2	0.0306	0.1149	0.3528	0.4861	0.3641	- 0.2429	-0.1130	0.2546	-0.0849
3	0.0026	0.0196	0.1289	0.3091	0.4302	0.1148	-0.2911	0.0584	0.1951
4	0.0002	0.0025	0.0340	0.1320	0.2811	0.3576	-0.1054	-0.2196	0.1825
5	—	0.0002	0.0070	0.0430	0.1321	0.3621	0.1858	-0.2341	-0.0735
6		—	0.0012	0.0114	0.0491	0.2458	0.3376	-0.0145	-0.2437
7			0.0002	0.0025	0.0152	0.1296	0.3206	0.2167	-0.1703
8			—	0.0005	0.0040	0.0565	0.2235	0.3179	0.0451
9				0.0001	0.0009	0.0212	0.1263	0.2919	0.2304
10				—	0.0002	0.0070	0.0608	0.2075	0.3005
11					—	0.0020	0.0256	0.1231	0.2704
12						0.0005	0.0096	0.0634	0.1953
13						0.0001	0.0033	0.0290	0.1201
14						—	0.0010	0.0120	0.0650

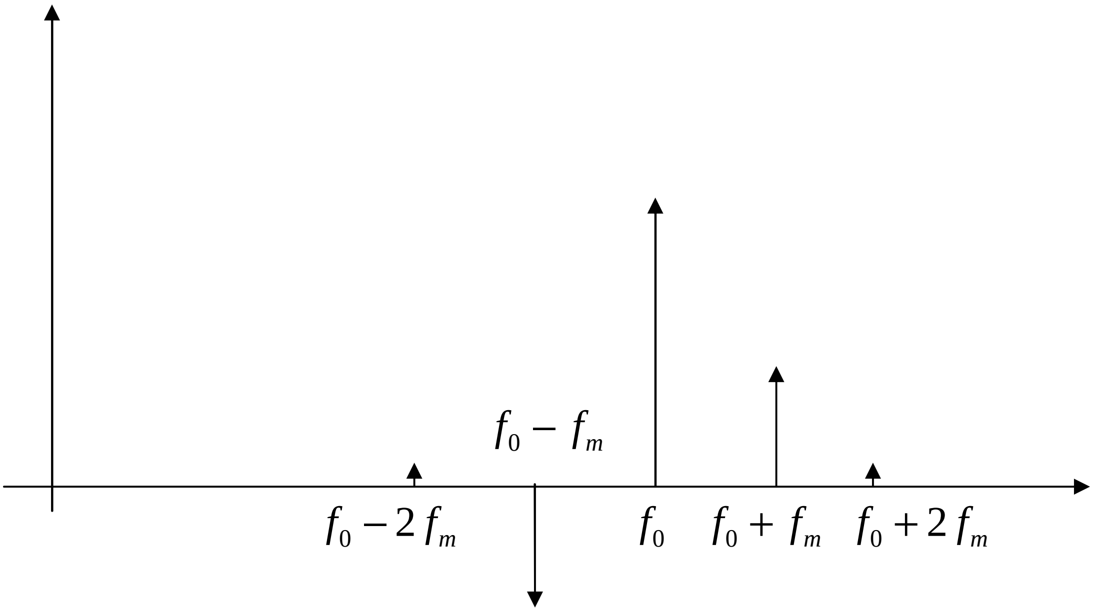
The exponentially modulated signal is:

$$x(t) = A_0 \sum_{n=-\infty}^{+\infty} J_n(\beta)\cos\left[(\omega_0 + n\omega_m)t\right]$$

We see that the signal contains a large number of components around the carrier frequency. For example, let us assume $\beta = 1$. Then only J_0, J_1 and J_2 are significantly different from zero. We can then write:

$$x(t) \approx A_0 \sum_{n=-2}^{2} J_n(1)\cos[(\omega_0 + n\omega_m)t]$$

The modulated signal is the sum of five sinewaves. The spectrum is displayed below. The values of the different amplitudes are read from the above table.



Single sided spectrum of the signal

The above spectrum is approximately bandlimited. For the exponentially modulated signal, we can use as transmission bandwidth the band of frequency that contains most of the power of the signal. The total power of the transmitted waveform can be computed as follows. From the series development of $x(t)$, we obtain:

$$P = <x^2(t)> = < A_0^2 \left(\sum_{n=-\infty}^{+\infty} J_n(\beta) \cos\left[(\omega_0 + n\omega_m) t \right] \right)^2 >$$

If the different sinewave are independent, the total power is the sum of the individual powers. We obtain:

$$P = \frac{A_0^2}{2} \sum_{n=-\infty}^{+\infty} J_n^2(\beta) = \frac{A_0^2}{2}$$

We have used the following property of the Bessel functions:

$$\sum_{n=-\infty}^{+\infty} J_n^2(\beta) = 1$$

If we keep *n* components on each side of the carrier, we obtain:

$$P(n,\beta) = \frac{A_0^2}{2} \sum_{k=-n}^{n} J_k^2(\beta)$$

The ratio of this power to the total power is:

$$\frac{P(n,\beta)}{P} = \sum_{k=-n}^{n} J_k^2(\beta)$$

This ratio is very close to 1 (≈ 0.95) for $n = \lfloor \beta + 1 \rfloor$. So, the band of frequency components containing the sidebands from $f_0 - nf_m$ to $f_0 + nf_m$ can be used as an approximation for the bandwidth of the exponentially modulated waveform for sinusoidal modulating waveform. So the transmission bandwidth is approximately:

$$B = 2nf_m = 2(\beta+1)\, f_m$$

For FM signal, $\beta = \frac{\Delta f}{f_m}$, we obtain:

$$B = 2(\Delta f) + 2\, f_m$$

The above rule is called the "*Carson's Rule*". It has been found empirically that this rule can be applied even if the modulating signal is not sinusoidal. In this case, we can define the "*Deviation Ratio*";

$\delta = \frac{\Delta f}{W}$ where *W* is the bandwidth of the baseband modulating signal.

Carson's rule generalizes to $B = 2(\delta+1)\, f_m = 2(\Delta f) + 2W$. For example, broadcast FM uses a value of Δf = 75 kHz for transmitting a baseband signal having a bandwidth W = 15 kHz. This gives a transmission bandwidth B = 2×75 + 2×15 = 180 kHz. The normalized bandwidth is 200 kHz. Carson's rule underestimates it, but the error is small.

A general remark about FM and PM is that the bandwidth of the resulting signal is in general much larger than $2W$. This is due to the fact that these modulations are highly non-linear. Another remark about FM is that it is very resistant to perturbations induced by noise and interference. So, we can say that FM protects the information of the signal at the expense of a bandwidth increase.

When β is very small, we can use $J_0(\beta) \approx 1$ and $J_1(\beta) \approx \frac{\beta}{2}$ to express the modulated signal:

$$x(t) = A_0 \cos \omega_0 t + A_0 \frac{\beta}{2} \cos\left[(\omega_0 + \omega_m)t\right] - A_0 \frac{\beta}{2} \cos\left[(\omega_0 - \omega_m)t)\right]$$

giving $x(t) = A_0 \left[\cos \omega_0 t - \beta \sin \omega_m t \sin \omega_0 t\right] = A_0 \left[\cos \omega_0 t - \beta s(t) \sin \omega_0 t\right]$ which is the narrow band approximation.

Filtering the FM signal

Being a non linear modulation, the usual method of filtering the complex envelop of the FM signal by the equivalent lowpass filter does not work for general filter shapes. In some specific cases, this technique can be used. In order to use it, this FM signal must be bandpass. In this case, the complex envelop is easily extracted.

$x(t) = A_0 \cos\left(\omega_0 t + \phi(t)\right)$ giving a complex envelop $m_x(t) = A_0 \exp\left(j\phi(t)\right)$. One case where this technique can be used is the case of a filter with an amplitude response of the type:

$H(f) = M(f)\exp\left(j\theta(f)\right)$ where $M(f) = M_0 + k(f - f_0)$ for $f > 0$ and $\theta(f) = -2\pi\tau_g(f - f_0) + \theta_0$. The equivalent lowpass filter is:

$H_{lp}(f) = [M_0 + kf]\exp\left(-j2\pi\tau_g f\right)\exp(j\theta_0)$. In the frequency domain, the complex envelop of the output is:

$M_y(f) = M_x(f)H_{lp}(f) = [M_0 + kf]\exp\left(-j2\pi\tau_g f\right)\exp(j\theta_0)M_x(f)$

So:

$$M_y(f) = M_0\exp\left(j\theta\right)M_x(f)\exp\left(-j2\pi\tau_g f\right) + kf\exp\left(j\theta\right)M_x(f)\exp\left(-j2\pi\tau_g f\right)$$

Given that $\mathscr{F}\left[\frac{dx(t)}{dt}\right] = 2\pi jfX(f)$, the multiplication by f in the frequency domain becomes a differentiation in the time domain. So,

$$m_y(t) = \left[M_0 m_x(t-\tau_g) + \frac{k}{2\pi j}\frac{dm_x(t-\tau_g)}{dt}\right]\exp\left(j\theta\right)$$

From the expression of the input complex envelop, we obtain:

$\frac{dm_x(t-\tau_g)}{dt} = jA_0\frac{d\phi(t-\tau_g)}{dt}\exp\left(j\phi(t)\right)$ and finally:

$$m_y(t) = A_0\left[M_0 + \frac{k}{2\pi}\frac{d\phi(t-\tau_g)}{dt}\right]\exp\left(j\phi(t-\tau_g)\right)\exp\left(j\theta\right)$$

The output signal is then:

$$y(t) = A_0\left[M_0 + \frac{k}{2\pi}\frac{d\phi(t-\tau_g)}{dt}\right]\cos\left(\omega_0 t + \phi(t-\tau_g) + \theta\right)$$

If the signal is FM, $\frac{d\phi(t)}{dt} = 2\pi(\Delta f)s(t)$. In this particular case:

$$y(t) = A_0\left[M_0 + k(\Delta f)s(t-\tau_g)\right]\cos\left(\omega_0 t + 2\pi(\Delta f)\int^{t-\tau_g} s(\lambda)d\lambda + \theta\right)$$

We remark that the output signal has two different modulations: FM and AM. The information signal is contained in the envelop of the output signal. So, an envelop detector will demodulate the signal.

If the filter has a different transfer function, we can use the concept of "*quasi-static*" approximation. If the carrier frequency is much higher than the baseband modulating frequency, then we can safely assume that the frequency is constant over a quite long time. The FM signal will behave almost like a constant frequency sinewave. We know that if the input of a filter with transfer function $H(f)$ is a pure sinewave $A_0\cos\left(2\pi f_0 t\right)$, the output is also a sinewave at the same frequency: $y(t) = A_0\left|H(f_0)\right|\cos\left[2\pi f_0 t + \text{Arg}\left[H(f_0)\right]\right]$. In the quasi- static approximation, we replace f_0 by the instantaneous frequency $f(t)$.

So, given $f(t) = f_0 + (\Delta f)s(t)$, we obtain:

$$y(t) = A_0\left|H(f_0 + (\Delta f)s(t)\right|\cos\left[2\pi f_0 t + 2\pi(\Delta f)\int^t s(\lambda)d\lambda + \text{Arg}\left[H\left(f_0 + (\Delta f)s(t)\right)\right]\right]$$

Example:

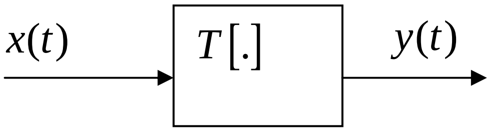
Using $H(f) = M(f)\exp\left(j\theta(f)\right)$ with $M(f) = M_0 + k(f - f_0)$ for $f > 0$ along with $\theta(f) = -2\pi\tau_g(f - f_0) + \theta_0$, we obtain:

$\left|H(f(t))\right| = M_0 + k(\Delta f)s(t)$ and $\text{Arg}\left[H(f(t))\right] = -2\pi\tau_g(\Delta f)s(t) + \theta_0$.

We see that we obtain the same amplitude variation as in the previous case (except for a time delay).

FM through nonlinear system

Consider the following memoriless nonlinear system.



If the input is $x(t) = A_0 \cos\theta(t)$, the output will be $y(t) = T\left[A_0 \cos\theta(t)\right]$. The output signal $T\left[A_0 \cos\theta(t)\right]$ is not periodic as a function of the variable *t*, it is periodic, with a period 2π if we consider it function of θ. We can thus develop the output *y* in Fourier series.

$$y(t) = \frac{a_0}{2} + \sum_{n=1}^{+\infty} a_n \cos n\theta(t)$$

The Fourier series is a cosine series because the input is $A_0 \cos\theta$. It is even. So, $T\left[A_0 \cos\theta\right]$ is also even. The coefficients are given by:

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} T\left[\cos\theta\right]\cos n\theta d\theta$$

Replacing the argument θ by its value leads to:

$$y(t) = \frac{a_0}{2} + \sum_{n=1}^{+\infty} a_n \cos\left[n\omega_0 t + n\phi(t)\right]$$

The above signal is a superposition of an infinite number of exponentially modulated waveforms. For the FM case, we can write:

$$y(t) = \frac{a_0}{2} + a_1 \cos\left(\omega_0 t + 2\pi(\Delta f)\int^t s(\lambda)d\lambda\right) + a_2 \cos\left(2\omega_0 t + 2\pi(2\Delta f)\int^t s(\lambda)d\lambda\right) + \cdots$$

We can remark that the output is a sum of FM signals at carrier frequencies nf_0 each one with a maximum deviation $n\Delta f$. If the different spectra do not intersect, we can select one of them using a bandpass filter tuned at nf_0 and obtain at the output:

$$z(t) = a_n \cos\left(n\omega_0 t + 2\pi(n\Delta f)\int^t s(\lambda)d\lambda\right)$$

FM Generation

Direct Method

The FM signal can be generated directly using a Voltage Controlled Oscillator (VCO) like the one used in the lab generators (GW-Instek GFG8255A). The output signal is a sinewave with an instantaneous frequency given by $f = f_v + k_m v_{in}$. The frequency f_v is called the free running frequency and the constant k_m is called the VCO gain (It is measured in Hz/Volts). In general, these VCOs can use a variable reactance in a parallel RLC circuit used to tune an oscillator. We can use "*varactors*" for example. The output frequency of this type of oscillator is the resonant frequency of the RLC circuit.

$f = \dfrac{1}{\sqrt{LC}}$ where $C = C_0 - cv_{in}$. The input voltage is proportional to $s(t)$. $v_{in} = |V_{in}|_{\max} s(t)$, so: $f(t) = \dfrac{1}{\sqrt{LC}} = \dfrac{1}{\sqrt{LC_0}}\left[1 - \dfrac{c}{C_0}v_{in}(t)\right]^{-\frac{1}{2}}$.

If $\left|\dfrac{c}{C_0}v_{in}(t)\right|_{\max} << 1$, we can use the approximation $[1-\varepsilon]^{-\frac{1}{2}} \approx 1 + \dfrac{1}{2}\varepsilon$.

The instantaneous frequency is given by: $f \approx f_v\left[1 + \dfrac{c}{2C_0}v_{in}\right]$. The free running frequency is $f_v = \frac{1}{\sqrt{LC_0}}$ and the VCO gain is $k_m = \frac{cf_v}{2C_0}$. The maximum frequency deviation is $\Delta f = |V_{in}|_{\max} k_m$. There exist a large number of VCO circuits. The most common ones (the ones that are found in integrated circuits and in signal generators) generate triangular waves using integrators or capacitors charged by controlled current sources. A good reference is "K. K. Clarke & D. T. Hess, *Communication Circuits: Analysis and Design*, 2nd ed. Krieger Pub Co, 1994" which is the textbook for the communication circuit course.

(The part between the two ♦ is for reading only)

<u>Frequency mixing</u>

In this part, we introduce an important technique used in receiver and transmitter design: Frequency mixing. The mixer is a device capable of changing a carrier frequency for any type of modulation. It is based on the frequency translation theorem of Fourier theory. We start first with real signal mixing.

Consider a general bandpass signal $x_{rf}(t) = r(t)\cos\left[\omega_{rf} t + \phi(t)\right]$. If we multiply this signal by a sinewave $x_{lo}(t) = B\cos\omega_{lo} t$, the result is:

$$z(t) = x_{rf}(t) \times x_{lo}(t) = Br(t)\cos\left[\omega_{rf} t + \phi(t)\right]\cos\omega_{lo} t$$

Using trigonometric identities, we see that this signal is the sum of two bandpass signals:

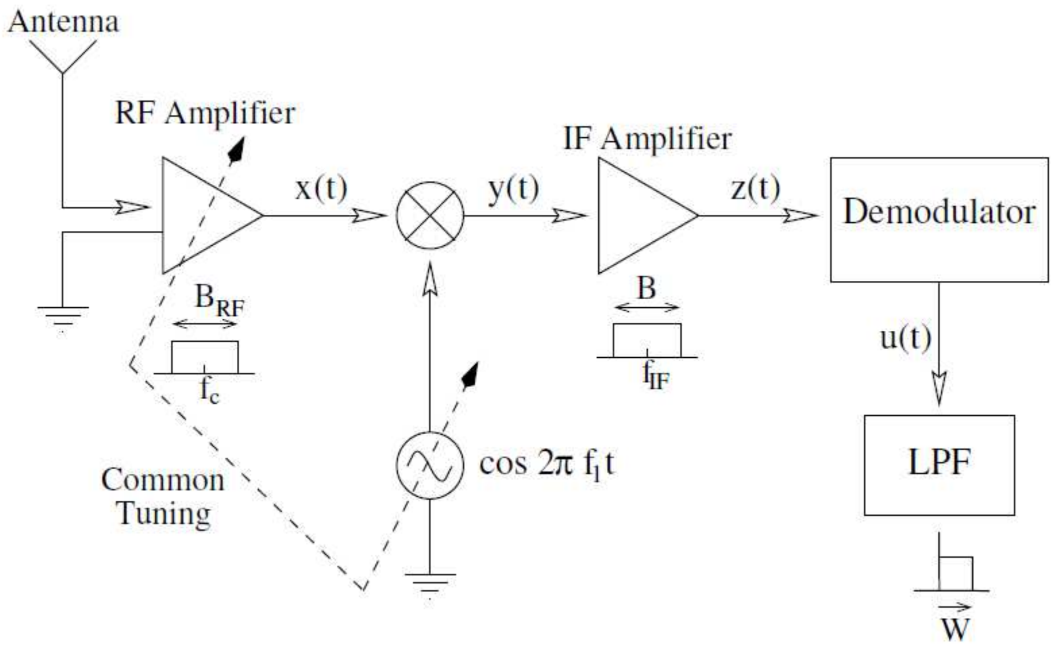
$$z(t) = \frac{B}{2} r(t)\cos\left[\left(\omega_{rf} + \omega_{lo}\right)t + \phi(t)\right] + \frac{B}{2} r(t)\cos\left[\left(\omega_{rf} - \omega_{lo}\right)t + \phi(t)\right].$$

Using a bandpass filter tuned at either the sum or the difference frequency, we obtain a bandpass signal having the same complex envelop (i.e. the same information) but a different carrier frequency. The new frequency is usually called "*intermediate frequency*" f_{if}.

If $f_{if} = f_{rf} + f_{lo}$, we say that we are doing "*up mixing*". On the other hand, if $f_{if} = f_{rf} - f_{lo}$ or $f_{if} = f_{lo} - f_{rf}$, we say that we are performing down mixing.

♦The "*mixer*" is an important electronic subsystem in any communication receiver or transmitter. It is the basic building block of the "*superheterodyne*" receiver. This concept of receiver was introduced in order to solve the very complex problem of amplifying and selecting one radio station among a large number of stations transmitting at different frequencies.

The first solution that comes to mind is to use a "*tunable*" bandpass filter. However, the construction of a very selective tunable bandpass filter is very complex. Furthermore, due to component aging, such system is prone to random changes and mistuning after a certain time. It is much easier to build a fixed frequency very selective filter. So, instead of translating the center frequency of a tunable filter before the different signals, it is much easier to translate the frequency of the signals before the center frequency of a fixed bandpass filter. This is the concept of the super heterodyne receiver. The superheterodyne receiver is composed of a tunable local oscillator ganged with a wide band tunable RF amplifier, a mixer and a fixed frequency IF amplifier. It is built using the block diagram shown below.



Let us assume we are using down mixing. The rf amplifier pre-selects a band of frequencies containing a small number of stations around the station at frequency f_c. The bandwidth B_{RF} is large compared to the bandwidth B required by the modulation used (FM, AM, any linear one) but smaller than $2f_{IF}$, the intermediate frequency. Using down mixing, we must have:

$f_{IF} = f_c - f_1$ giving $f_c = f_1 + f_{IF}$ or $f_{IF} = f_1 - f_c$ giving $f_c = f_1 - f_{IF}$.

From the above two relations, we see that if the rf filter does not exist, then we can receive two different stations if we simply use the local oscillator for tuning. These two stations are separated by $2f_{IF}$. These two frequencies are called "*image frequencies*". The job of the tunable rf amplifier is to eliminate one of them so that it will not interfere with the station that we want to receive. Intermediate frequency for broadcast receivers has been standardized to the values of 455 kHz for AM and 10.7 MHz for FM.

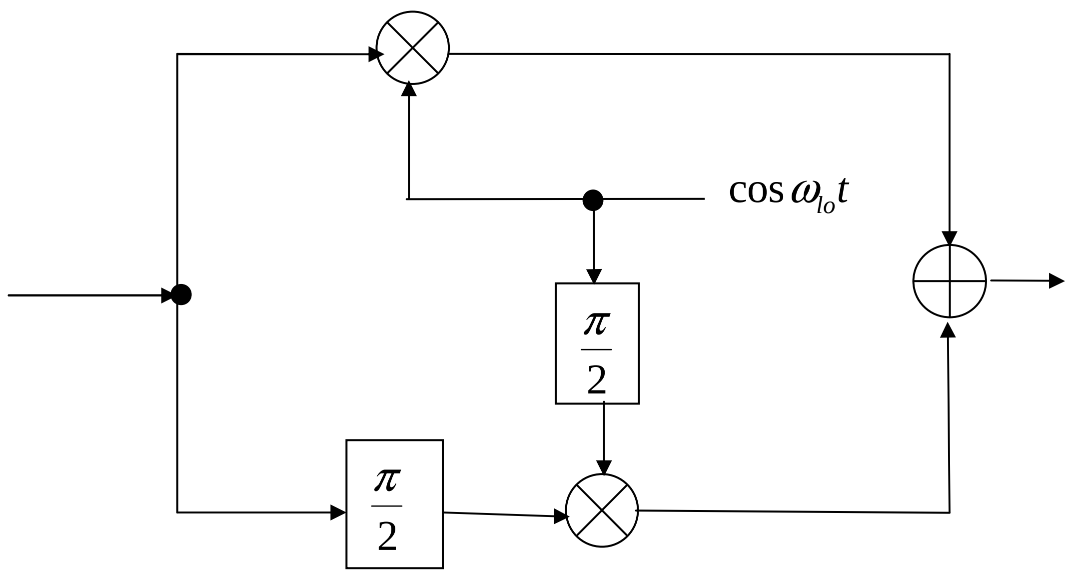
Another technique is used to avoid this problem of image frequencies. It is based on frequency translation using complex phasors. Consider the high frequency signal: $x_{rf}(t) = r(t)\cos\left[\omega_{rf}t + \phi(t)\right]$. The associated analytic signal is: $x_{rf+}(t) = x_{rf}(t) + j\hat{x}_{rf}(t) = r(t)\exp\left(j\phi(t)\right)\exp\left(j\omega_{rf}t\right)$. If we multiply this signal by the phasor: $\exp\left(-j\omega_{lo}t\right)$, we obtain the signal: $z(t) = r(t)\exp\left(j\phi(t)\right)\exp\left(j\left(\omega_{rf} - \omega_{lo}\right)t\right)$. The real part is:

$x_{if}(t) = \text{Re}\left[z(t)\right] = r(t)\cos\left[\left(\omega_{rf} - \omega_{lo}\right)t + \phi(t)\right]$. This is the correct translated signal. So, the process of performing the above operation is:

$x_{if}(t) = \text{Re}\left[\left(x_{rf}(t) + j\hat{x}(t)\right)\left(\cos\omega_{lo}t - j\sin\omega_{lo}t\right)\right]$ giving:

$$x_{if}(t) = x_{rf}(t)\cos\omega_{lo}t + \hat{x}_{rf}(t)\sin\omega_{lo}t$$

This leads to the following block diagram:



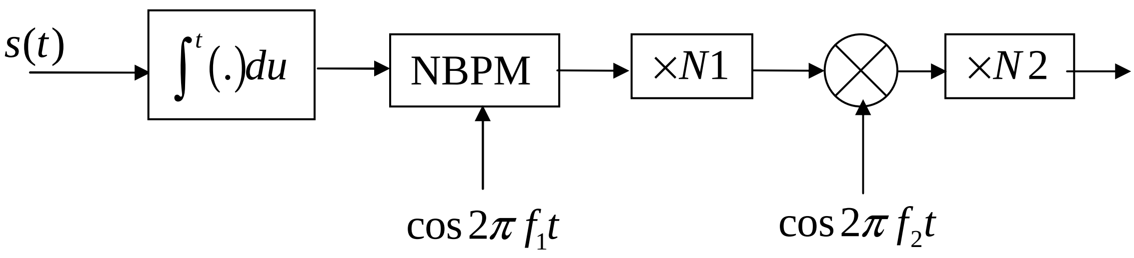
Imageless Mixer

The above circuit can be used without any image rejection filter before.

♦

Indirect FM generation

This technique of FM generation is the one that is commonly used in FM transmitters. This is due to the fact that the carrier frequency and the maximum frequency deviation can be set with high precision. It is based on a Narrow Band Frequency modulator cascaded with nonlinear amplifiers that are used as frequency multipliers. Mixers are also used to translate the carrier because frequency multiplication leads usual to impractically high carrier frequencies. A general block diagram of such system is:



The frequency multipliers are implemented using a nonlinear amplifier (Class C) followed by a narrow bandpass filter tuned at the proper harmonic.

If we consider the above block diagram, the carrier frequency is given by $f_0 = N2\left(N1\times f_1 - f_2\right)$ or $f_0 = N2\left(f_2 - N1\times f_1\right)$. If the frequency deviation at the output of the NBFM is Δf_1, the final deviation is $\Delta f = N1\times N2\times\Delta f_1$. In general the frequency multiplication is achieved by a cascade of frequency doublers and triplers. It is impossible to achieve an efficient amplifier if the multiplication factor is larger than three.

FM Demodulation

FM demodulation by differentiation (FM to AM conversion):

If we compute the derivative of an FM signal, we obtain:

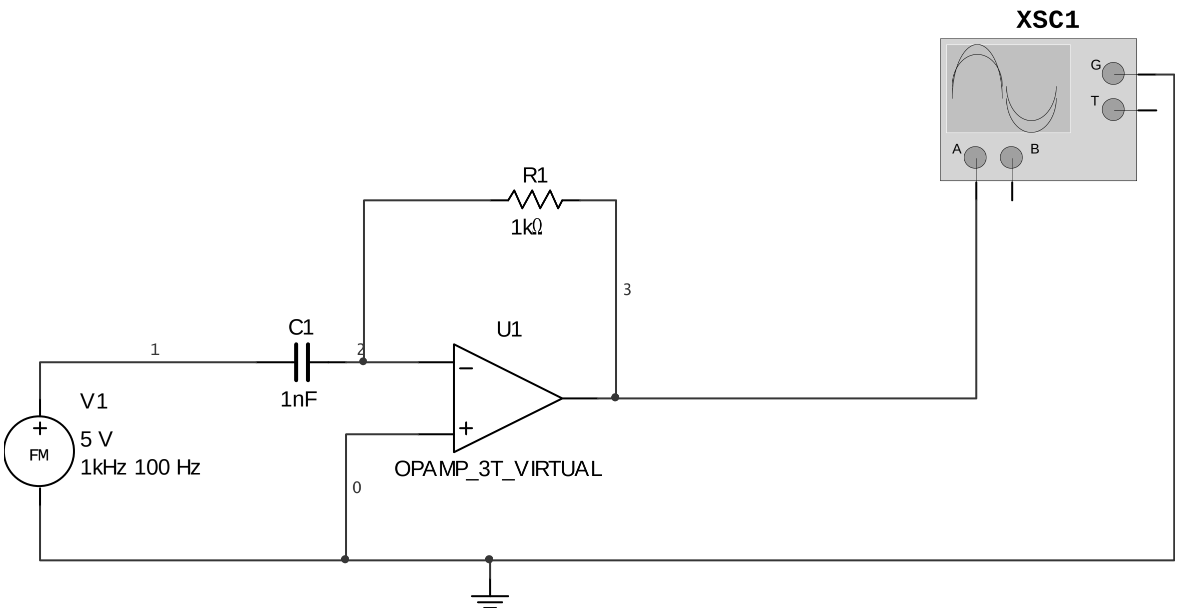
$$\frac{d}{dt}\left[A_0\cos\left(\omega_0 t+\phi(t)\right)\right]=-A_0\left(\omega_0+\frac{d\phi}{dt}\right)\sin\left(\omega_0 t+\phi(t)\right)$$

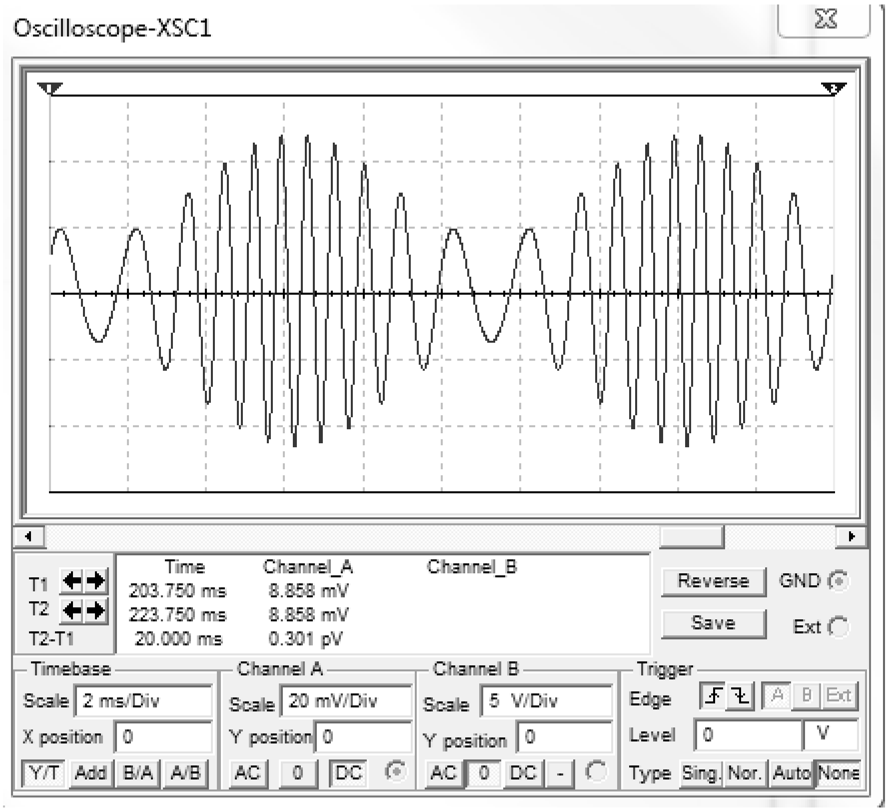
Since the signal is an FM one, $\frac{d\phi}{dt}=2\pi(\Delta f)s(t)$, we get:

$$\frac{d}{dt}\left[A_0\cos\left(\omega_0 t+\phi(t)\right)\right]=-A_0\left(\omega_0+2\pi(\Delta f)s(t)\right)\sin\left(\omega_0 t+2\pi(\Delta f)\int^{t}s(\lambda)d\lambda\right)$$

We can see that the output of a differentiation circuit will produce a modulation of the amplitude. This modulation can be detected using any AM demodulator.

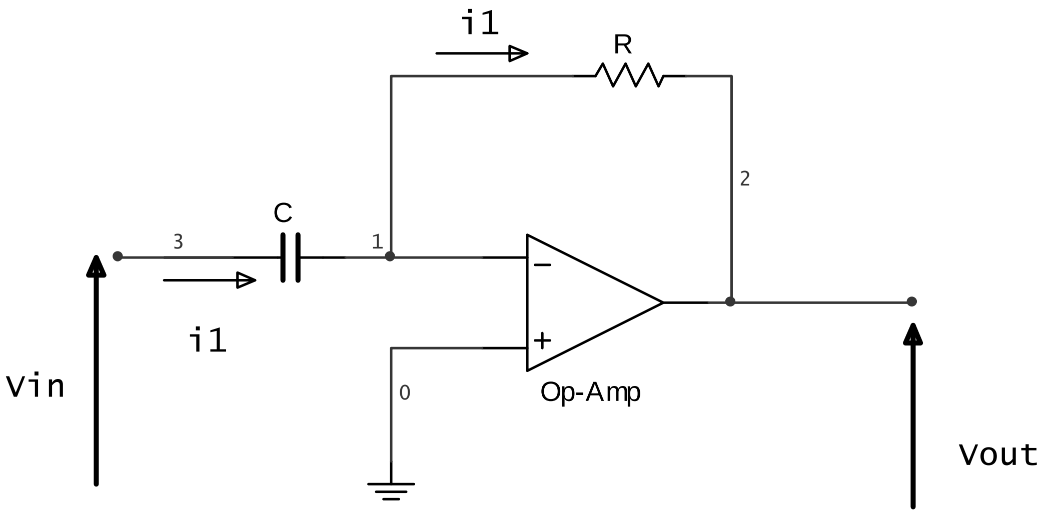
The following circuit is a differentiator built using an OP-Amp simulated using Multisim.





The above picture displays the output signal. We clearly see the two modulations: AM and FM. The signal produced by the FM source is a sinewave modulated FM signal with β = 5, carrier frequency = 1 kHz and modulating frequency = 100 Hz.

♦ Differentiator using an op-amp



The Operational Amplifier is an amplifier that possesses a very high gain, differential input and very high input impedance. If the voltage at the inverting input is v_-, the input at the non-inverting input is v_+ and the gain of the op-amp is *G*, its output is given by $G(v_+ - v_-)$. *G* is assumed very large. This means that a very small difference will produce a measurable output. We can safely assume that this difference is zero. At that time, the inverting input is practically at ground in the above schematic. This implies that the capacitor is in parallel with Vin. So, $i1 = C\frac{dVin}{dt}$. The input impedance of the op-amp is very large. The same current will flow though the resistance R. So, $Vout = -Ri1$. After the elimination of i1, we obtain:

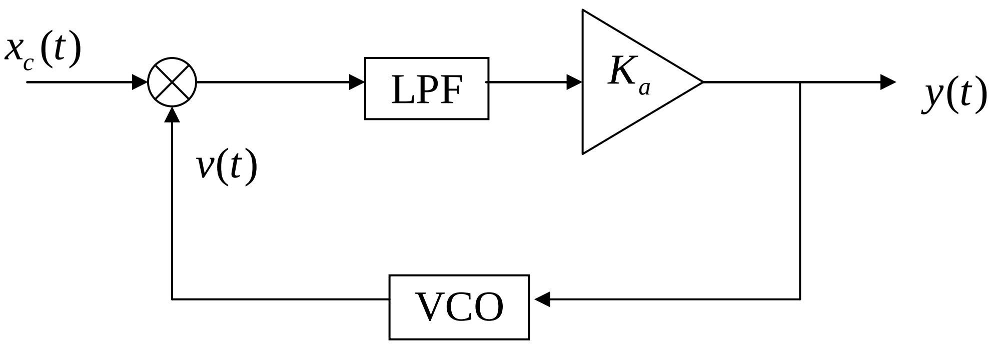
$$Vout = -RC\frac{dVin}{dt}$$

♦

There exist a large number of other FM demodulators. The interested student should consult the previous reference (Clarke & Hess).

The Phase Locked Loop

The phase locked loop (PLL) is a feedback system composed principally of a voltage controlled oscillator (VCO), a phase detector (PD) and a lowpass filter (LPF). The phase detector is usually modeled as a multiplier.



The input signal is $x_c(t) = A_c \cos\theta_c(t)$ and the output of the VCO is $v(t) = A_v \cos\theta_v(t)$. The output of the multiplier is $z(t) = \frac{A_c A_v}{2}\left[\cos\left(\theta_c(t) - \theta_v(t)\right) + \cos\left(\theta_c(t) + \theta_v(t)\right)\right]$. The lowpass filter eliminates the sum term and filters the difference term. So, we can consider that the output of the lowpass filter (after amplification) is

$$y(t) = K_a h(t) * \left\{\frac{A_c A_v}{2}\left[\cos\left(\theta_c(t) - \theta_v(t)\right)\right]\right\}.$$

Let us introduce a variable ε such that: $\theta_v(t) = \theta_c(t) - \varepsilon(t) + \frac{\pi}{2}$. The output of the lowpass filter becomes:

$$y(t) = K_a h(t) * \left\{\frac{A_c A_v}{2}\left[\sin\varepsilon(t)\right]\right\}$$

We have a one to one relationship between $y(t)$ and $\varepsilon(t)$ if $-\frac{\pi}{2} \leq \varepsilon(t) \leq \frac{\pi}{2}$. The relation becomes linear if $\varepsilon << 1$. To simplify the analysis of the PLL and eliminate the effect of the amplitudes, let us make $A_c = 2$ and $A_v = 1$. The input signal has a carrier frequency f_0. This makes $\theta_c(t) = \omega_0 t + \phi(t)$. The VCO free running frequency f_v is shifted from f_0 by an offset f_Δ: $f_v = f_0 - f_\Delta$. The instantaneous phase of the VCO is: $\theta_v(t) = 2\pi f_v t + \phi_v(t) + \frac{\pi}{2}$. The $\frac{\pi}{2}$ constant is added in order to introduce the variable $\varepsilon(t)$ in the following expressions. The instantaneous phase deviation $\phi_v(t)$ is produced by the output signal

$y(t)$: $\dot{\phi}_v(t) = 2\pi K_v y(t)$[4]. The constant K_v is the VCO gain expressed in Hz/Volts. Using the definition of $\varepsilon(t)$, we can write:

$$\varepsilon(t) = \theta_c(t) - \theta_v(t) + \frac{\pi}{2}$$

Replacing each phase, we obtain:

$$\varepsilon(t) = 2\pi f_\Delta t + \phi(t) - \phi_v(t)$$

After differentiation:

$$\dot{\varepsilon}(t) = 2\pi f_\Delta + \dot{\phi}(t) - 2\pi K_v y(t)$$

However, $y(t)$ is the output of the lowpass filter amplified by K_a. So:

$$y(t) = K_a \int_{-\infty}^{+\infty} h(\tau) \sin\left(\varepsilon(t-\tau)\right) d\tau$$

The PLL is thus governed by the following integro-differential equation:

$$\dot{\varepsilon}(t) + 2\pi K \int_{-\infty}^{+\infty} h(\tau) \sin\left(\varepsilon(t-\tau)\right) d\tau = \dot{\phi}(t)$$

K = $K_a K_v$ is called the "*Loop Gain*". The above equation is a nonlinear equation that is quite complex to solve. In our course, we are going to consider two different cases: The first order loop (with no filtering) to analyze in a simple manner the "locking mechanism" and the linear approximation when $\sin\varepsilon \approx \varepsilon$.

Frequency Acquisition

Consider a filter with a transfer function $H(f) = 1$ over the frequency band of interest. The impulse response will be a Dirac impulse: $h(t) = \delta(t)$ and we will have:

[4] A dot above a function means that the function is differentiated.

$$\int_{-\infty}^{+\infty} h(\tau)\sin\left(\varepsilon(t-\tau)\right)d\tau = \sin\left(\varepsilon(t)\right)$$

We obtain the following first order differential equation:

$$\dot{\varepsilon} + 2\pi K \sin\varepsilon = \dot{\phi}(t) + 2\pi f_\Delta$$

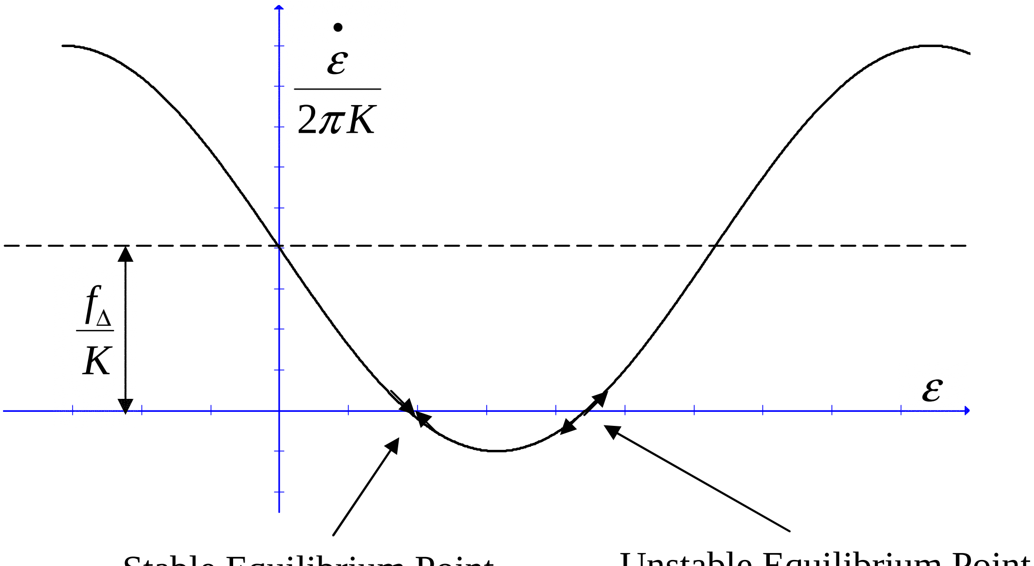
Let us assume that we apply the signal $x_c(t) = 2\cos\left(\omega_0 t + \phi_0\right)$ at the time $t = 0$. In this case, $\phi(t) = \phi_0$ and its derivative is zero. We also assume that the free running frequency of the VCO is different from the carrier we want to lock on ($f_\Delta \neq 0$). The differential equation becomes $\dot{\varepsilon} + 2\pi K \sin\varepsilon = 2\pi f_\Delta$ for $t \geq 0$. This equation can be written as:

$$\frac{\dot{\varepsilon}}{2\pi K} = -\sin\varepsilon + \frac{f_\Delta}{K}$$

This equation relates the derivative of ε with ε. The set of points $\left(\dot{\varepsilon}, \varepsilon\right)$ is called the phase plane. In our case, it is better to use $\left(\frac{\dot{\varepsilon}}{2\pi K}, \varepsilon\right)$. This describes a single curve with *t* as a parameter. As *t* varies, the point is going to move along the curve shown below. The points on the curve where $\dot{\varepsilon} = 0$ are called "*equilibrium points*". We distinguish two different types of equilibrium points: A point is called stable if after a small perturbation around the point, the trajectory in the phase plane will go back to the equilibrium point. The point is called unstable if after a small perturbation, the trajectory will go away from the equilibrium point.

The following mechanical analogy will show the difference between the points.

If a trajectory starts at an equilibrium point, it will remain there.



Phase Plan Plot

The above curve shows the locus of the points $\left(\frac{\dot{\varepsilon}}{2\pi K}, \varepsilon\right)$ as t goes from zero to infinity. The trajectory starts at an initial point ε_0 and it will move. It is easy to see that for any starting point, the trajectory will move toward a stable equilibrium point. However, we can have equilibrium points if and only if the curve intersects the ε axis. This is possible if $\left|\frac{f_\Delta}{K}\right| \le 1$. So, if this condition is satisfied, then $\lim_{t\to\infty} \dot{\varepsilon}(t) = 0$ and the steady state value of ε will be:

$$\varepsilon_{st} = \sin^{-1}\frac{f_\Delta}{K}$$

The steady state value of *y*(*t*) will be:

$$y_{st} = K_a \sin\varepsilon_{st} = \frac{f_\Delta}{K_v}$$

and since $\theta_v(t) = \theta_c(t) - \varepsilon(t) + \frac{\pi}{2}$, the steady state output of the VCO will be:

$$v_{st}(t) = \cos\left(\omega_0 t + \phi_0 - \varepsilon_{st} + \frac{\pi}{2}\right)$$

If the loop gain is very large, then ε_{st} will be very small. The VCO output will have exactly the same frequency as the input signal with a phase shift that is practically $\frac{\pi}{2}$.

If $\left|\frac{f_\Delta}{K}\right| > 1$, it is impossible to have a lock. The trajectory will not intersect the ε axis and there is no equilibrium point and no steady state solution. The VCO frequency will keep changing.

In order to have a lock, the VCO free running frequency must be in the following range: $[f_0 - K, f_0 + K]$. This range is called the "*lock range*". The value found is valid for a first order loop only. It will be different for another filter.

FM Demodulation

In order to analyze the PLL when the input is frequency modulated, we assume that the VCO free running frequency is the same as the input carrier frequency. This means that $f_\Delta = 0$. We also assume that ε remains small all the time. We can replace sinε by ε in our analysis. With these hypotheses, we have:

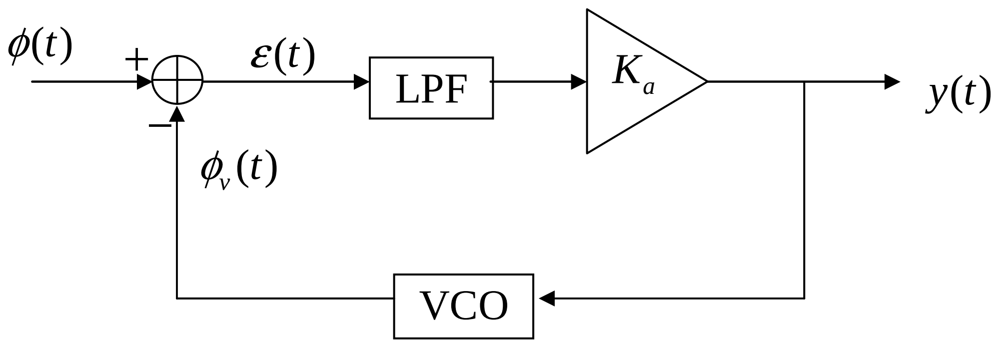
$$x_c(t) = 2\cos\left(\omega_0 t + \phi(t)\right)$$

$$v(t) = \cos\left(\omega_0 t + \phi_v(t) + \frac{\pi}{2}\right)$$

and $y(t) = K_a \int_{-\infty}^{+\infty} h(\tau)\sin\left(\varepsilon(t-\tau)\right)d\tau \simeq K_a \int_{-\infty}^{+\infty} h(\tau)\varepsilon(t-\tau)d\tau$

$$\varepsilon(t) = \phi(t) - \phi_v(t)$$

Since y is linearly related to ε and ϕ_v is also linearly related to y, it is more interesting to use the phase deviations as primary variables instead of using x_c and v. the different relationships are better described by the following block diagram.



The above block diagram is completely linear. The VCO transfer function is given by the following relation:

$$\phi_v(t) = 2\pi K_v \int^t y(u)du$$

This gives the following relation in the frequency domain:

$$\Phi_v(f) = \frac{K_v}{jf} Y(f)$$

The closed loop transfer function (in the frequency domain) is:

$$Y(f) = \frac{1}{K_v} \frac{jfH(f)}{H(f) + j\dfrac{f}{K}} \Phi(f)$$

where $H(f) = \mathscr{F}\left[h(t)\right]$.

If the input is an FM wave, $\phi(t) = 2\pi\Delta f \int^t s(u)du$, then

$$\Phi(f) = \frac{\Delta f}{jf} S(f)$$

The transfer function becomes:

$$Y(f) = \frac{\Delta f}{K_v} \frac{H(f)}{H(f) + j\dfrac{f}{K}} S(f)$$

We see that the output signal is the baseband information signal filtered. If the loop gain is very high, the output signal will be proportional to $s(t)$.

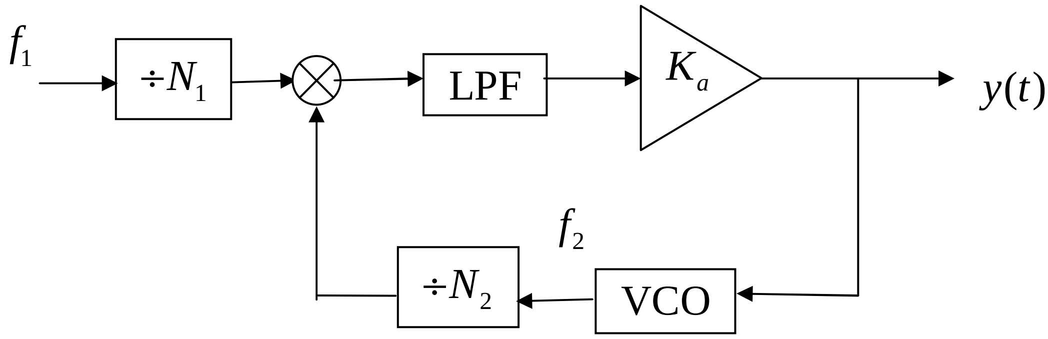
$$y(t) \approx \frac{\Delta f}{K_v} s(t)$$

So, if we can make sure that the error signal is small at all times, the PLL can be used with advantage as a frequency demodulator.

Another important application of the PLL is the implementation of frequency synthesizers.

Frequency Synthesis

When a PLL is locked, the frequencies of the signals arriving at the two inputs of the phase detector (multiplier) are equal.



The blocks labeled $\div N_k$ are frequency dividers (usually implemented by presettable logic counters). At the inputs of the phase detector (assuming that the PLL is locked), we can write:

$$\frac{f_1}{N_1} = \frac{f_2}{N_2}$$

So, the VCO will produce: $f_2 = \frac{N_2}{N_1} f_1$. This means that we can produce a signal with a frequency that can be set using digital hardware and that can be very stable if the reference oscillator producing f_1 is very stable. This technique of frequency generation is commonly used in modern receivers. The PLL is usually built in a microcontroller.