FM Modulation and Demodulation

EE340: Prelab Reading Material for Experiment 3

AUTUMN 2024

1 Frequency Modulation

In the case of FM, the instantaneous frequency of the carrier wave, $c(t) = cos(2\pi f_c t)$, is varied with the amplitude of message signal x(t). The instantaneous frequency of the modulated signal can be written as $f(t) = f_c + f_\Delta$.x(t),where f_Δ is the maximum frequency deviation away from f_c if x(t) is normalized such that the maximum value $|x(t)| = 1$. Therefore, the phase of the FM signal is

$$\phi(t) = \int_0^t 2\pi f(T)dT = 2\pi f_c t + 2\pi f_\Delta \int_0^t x(T)dT$$

and the modulated FM signal can be written as

$$s(t) = cos(2\pi f_c t + 2\pi f_\Delta \int_t^0 x(T)dT$$

If the message signal is represented by the sinusoid $x(t) = sin(2\pi f_m t)$, the FM signal, ignoring the constant of integration, becomes

$$s(t) = cos(2\pi f_c t + \frac{f_\Delta}{f_m} cos(2\pi f_m t)$$

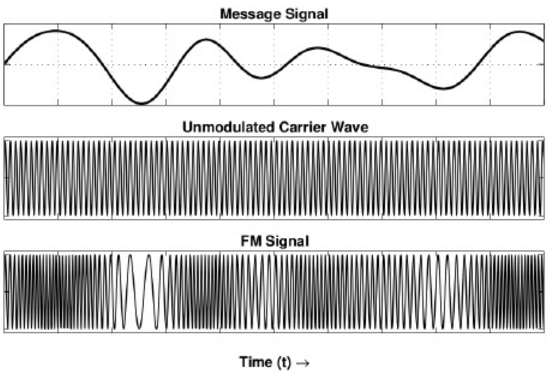

Figure 1: Time domain representation of an FM (Frequency Modulated) signal

The ratio $h = \frac{f_\Delta}{f_m}$ is called the modulation index of the FM signal, where f_m is the highest frequency component in x(t). If h << 1, the modulation is called Narrow-band FM (NBFM), while for $h \leq 1$, it is called Wide-band FM (WBFM). In theory, an FM signal occupies an infinite bandwidth. However, as per the *Carson's Rule* for practical purposes, most of the power in the FM signal spectrum is restricted between the frequencies $(f_c - f_m - f_\Delta)$ and $(f_c + f_m + f_\Delta)$. The bandwidth occupied by the FM signal $is BW \approx 2(f_m + f_\Delta)$. In general, the WBFM signals have much better signal-to-noise ratio compared to the NBFM signals, but occupy much larger bandwidth.

1.1 Demodulation of the FM Signals

In the analog domain, the FM signals can be demodulated using a phase locked loop (PLL), in which a feedback loop tracks the incoming signal frequency by adjusting the control voltage applied to the VCO (voltage controlled oscillator) in the loop, as shown in Fig. 2.

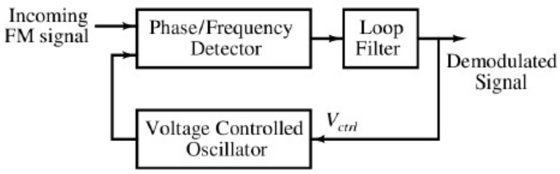

Figure 2: Use of a PLL for demodulating FM signals

The Phase/Frequency Detector (PFD) block compares the phase (represented by the time instants of the zero crossings) of the incoming signal with that of the fed-back signal. The output of the PFD followed by the Loop Filter block is a voltage signal which is proportional to the difference between the two phases.Therefore, this control voltage changes with the instantaneous frequency of the incoming signal (which depends on the message signal amplitude), and hence demodulates the message signal. Some other analog domain techniques for demodulating FM signal are also used, but will not be a part of this discussion. Generally, to demodulate the FM signal in a digital signal processor or in a software, first the phase of the incoming signal is extracted. The frequency information is then obtained by differentiating the phase with respect to time, to obtain the demodulated message signal. You will be using this approach for demodulating FM signals in GNU radio.

2 Discrete Time Filters for Implementation in GNU Radio

In a computer or a digital signal processor, signals are represented as discrete-time samples. Consider an LTI system for which the nth input and output samples are represented by x[n] and y[n], respectively. Therefore, for a causal system, one can write

$$a_0 y[n] = (a_1 y[n-1] + a_2 y[n-2] + ...) + (b_0 x[n] + b_1 x[n-1] + b_2 x[n-2] + ...)$$

A discrete-time signal can be represented in the z-domain, which is equivalent to frequency-domain representation of the discrete-time signal. In the "z-domain", the delay by one sample period is represented by z^{-1}, which means $z^{-1} = e^{-j\omega T}$, where T is the sampling period. The delay by one sample period in the time-domain implies multiplication of the signal by z^{-1} in the z-domain. Therefore, the z-transform of x[n-n$_0$] becomes X(z)z^{-n0} if X(z) is the z-transform of x[n]. Further, if X(z) and Y (z) are z-transforms of x[n] and y[n] respectively, (1) can be rewritten in the z-domain as

$$a_0 Y(z) = Y(z)(a_1 z^{-1} + a_2 z^{-2} + ...) + X(z)(b_0 + b_1 z^{-1} + b_2 z^{-2} + ...)$$

Therefore, for this system, the transfer function in z-domain can be expressed as

$$H(z) = \frac{Y(z)}{X(z)} = \frac{(b_0 + b_1 z^{-1} + b_2 z^{-2} + ...}{(a_0 - a_1 z^{-1} - a_2 z^{-2} - ...)}$$

The above expression represents an IIR (infinite impulse response) filter, because y[n] depends on the previous output samples $y[n-1], y[n-2], y[n-3], ...$, in addition to the input samples, and therefore may last for an infinite duration for an impulse input. However, if a_1, a_2, a_3, ... are all zero (and assuming that the number of non-zero coefficients $b_0, b_1, b_2, ...$ is finite), the response to an impulse inputs lasts for a finite duration only, and the corresponding filter is known as an FIR (finite impulse response) filter.

The IIR block in GNU radio software implements the above transfer function, where $b_0, b_1, b_2, ...$

represent feed-forward taps and $a_0, a_1, a_2, ...$ represent feedback taps. The FIR block can be used for implementing an FIR filter. Alternatively, the FIR filter can also be implemented using the IIR block and setting the filter coefficients $a_1, a_2, a_3, ... = 0$.

Important Notes:

- For implementing (1) and (2) using the IIR block, use the default setup, i.e., Old style of taps = **TRUE** in the IIR block. If Old style of taps = **FALSE** is used, the following equation is implemented:
$$H(z) = \frac{Y(z)}{X(z)} = \frac{(b_0 + b_1 z^{-1} + b_2 z^{-2} + ...}{(a_0 + a_1 z^{-1} + a_2 z^{-2} + ...)}$$
which means the signs of $a_1, a_2, a_3, ...$ have to be inverted. In this lab course, we will always be assuming that the default setup, i.e.,Old style of taps = TRUE is used.

- The value of a_0 is always taken as 1 by the software, even if you specify some other value. Therefore, make sure that you have used $a_0 = 1$ in your calculations.

FM Modulator and Demodulator Implementation: The FM modulator and demodulator implementations in GNU Radio will require use of an integrator and a differentiator, respectively. In the software, the discrete time equations y[n] = y[n-1]+T *x[n] and y[n] = (x[n]- x[n-1])/T can be used for implementing an integrator using the IIR block and a differentiator using the FIR block, respectively, where T is the sampling period. More details about the implementation of the modulator and demodulator blocks will be made available in the lab-sheets.

3 Noise in FM Transmission

The channel adds white Gaussian noise to the signal, which basically implies that the power spectral density (PSD) of the added noise is flat (i.e. independent of frequency) in the frequency band of interest, as shown in Fig. 3.

Figure 3: Block diagram showing message transmission using FM through an AWGN (Additive White Gaussian Noise) channel.

Here, we denote the noise added by the channel in the FM signals frequency band as $n_p(t)$, which can be written as (recall Experiment 4 reading material)

$$n_p(t) = n_I(t)cos(2\pi f_c t) - n_Q(t)sin(2\pi f_c t), \tag{3}$$

where $n_I(t$ and $n_Q(t)$ are both independent of each other and also have flat PSDs. Ignoring scaling factors, the signal received at the FM demodulator input can be written as

$$s_r(t) = s_p(t) + n_p(t) = cos(2\pi f_c t + \phi_m(t) + n_I(t)cos(2\pi f_c t) - n_Q(t)sin(2\pi f_c t) \tag{4}$$

It can be shown that for narrow-band FM, i.e. when $|\phi_m << 1$, and when the noise level is much smaller than the signal level, i.e. $|n_I|, |n_Q| << |\phi_m|$, the carrier phase is$\phi_r(t)\phi_m(t) + n_Q(t)$. The demodulated message signal, thus, is

$$x_r(t) = \frac{d}{dt}\phi_r(t) = \frac{d}{dt}\phi_m(t) + \frac{d}{dt}n_Q(t) = x(t) + \frac{d}{dt}n_Q(t) \tag{5}$$

Differentiation of a signal with respect to time results in peaking of high frequency components (differentiation in time domain implies multiplication by $j\omega$ in frequency domain). Since the demodulation process involves differentiation of $\phi_r(t)$, the high frequency components of noise part in it experience a large gain.

4 Use of pre-emphasis and de-emphasis to avoid noise peaking

The problem of noise peaking at high-frequencies in the demodulated signals can be avoided by applying a de-emphasis filter, which is basically a low pass filter with its gain decreasing with frequency. However, simply applying de-emaphasis to the demodulated signal will de-emphasize the high frequency contents of the message signal as well, which will effectively distort the signal. To avoid this problem, the message signal at the transmitter can first be pre-emphasized (i.e. passed through a high-pass filter) before frequency modulation, as shown in Fig.4.

Figure 4: Block diagram showing message transmission using FM through an AWGN channel. Preemphasis and de-emphasis filters are used for avoiding noise peaking and getting undistorted signal as the final output.

The resultant signal can then be frequency modulated and transmitted via the wireless channel. The received FM signal is demodulated and passed through a low pass filter to get back the desired message signal. If the product of the pre-emphasis and de-emphasis transfer functions is a constant, message signal distortion as well as high-frequency noise peaking can be avoided. Typically, the pre-emphasis filter is chosen as a first-order high-pass filter with corner frequency f_{pre}, and the de-emphasis filter as a first-order low-pass filter with corner frequency f_{de}, such that $f_{pre} = f_{de}$. The time constantτ of values 50 μs or 75 μs are commonly used for these filters, for which, the corresponding corner frequency values are given by $f_{3dB} = {}_{3dB}/(2\pi) = 1/(2\pi\tau)$. The pre-emphasis and de-emphasis filters can be implemented using IIR filter blocks in GNU radio (please refer to prelab reading material of Experiment 2).

5 FM demodulation using RTL-SDR dongle

The demodulation of an FM signal (before de-emphasis) can be carried out in the following way. The incoming message signal $x_r[n]$ from RTL-SDR dongle can be multiplied with the conjugate of its one-sample-delayed copy, i.e. $x_r[n1]$. Therefore, if the phase of $x_r[n]$ is $\phi_r[n]$, then the phase of the resultant signal becomes $\phi_r[n]\phi_r[n1]$, which is basically proportional to the desired demodulated signal (before de-emphasis). The proportionality constant 1/T (where T is the sampling period) can be multiplied to this resultant phase to get the correct differential value in the absolute sense. **Important note:** In the discrete-time implementation in GNU radio, the phase value (which is obtained using the 'Complex to Arg' block) has an ambiguity of $2n\pi$, where n is an integer. Therefore, you need to be careful in your implementation to ensure that the $2n\pi$ jumps in the demodulated signal are avoided. When the noise level is high, these phase jumps are sometimes unavoidable, and as a result, shape of the noise spectrum in the demodulated spectrum changes. Try to observe this effect by increasing noise in your simulations when you are carrying out the experiment.