

Noise:

Background

The performance of any communication system, is ultimately limited by two factors: (i) the transmission bandwidth, and (ii) the noise[3]. Bandwidth is a resource that must be conserved as much as possible, since only a finite electromagnetic spectrum is allocated for any transmission medium. Whatever the physical medium of the channel, the transmitted signal is corrupted in a random manner by a variety of possible mechanisms as it propagates through the system[4]. The term *noise* is used to denote the unwanted waves that disturb the transmission of signals, and over which we have incomplete control. As we will see throughout this course, bandwidth and noise are intimately linked. In this chapter our aim is to develop a model for the noise, with the ultimate goal of using this model in later chapters to assess the performance of various modulation schemes when used in the presence of noise.

A Model of Noise

Sources of noise

In a practical communication system, there are many sources of noise. These source may be external to the system (e.g., atmospheric, b galactic, c and synthetic noised) or internal to the system.[4]

Internal noise arises due to spontaneous fluctuation of current or voltage in electrical circuits, and consists of both *shot noise* and *thermal noise*.

Shot noise arises in electronic devices and is caused by the random arrival of electrons at the output of semiconductor devices. Because the electrons are discrete and are not moving in a continuous steady flow, the current is randomly fluctuating. The important characteristic of shot noise is that it is *Gaussian* distributed with zero mean (i.e, it has the Gaussian probability densityfunction shown in Figure). This follows from the *central limit theorem*, which states that the sum of η independent random variables approaches a Gaussian distribution as $\eta \rightarrow \infty$. In practice, engineers and statisticians usually accept that the theorem holds when $\eta \geq 6$.

Thermal noise is associated with the rapid and random motion of electrons within a conductor due to thermal agitation[4]. It is also referred to as Johnson noise, since it was first studied experimentally in 1928 by Johnson, e who showed that the average power in a conductor due to thermal noise is

$$P_{\text{thermal}} = kTB$$

where k is Boltzman's constant (1.38×10^{-23}), T is the absolute temperature in Kelvin, and B is the bandwidth in hertz. f Again, because the number of electrons in a conductor is very large, and their random motions are statistically independent, the central limit theorem indicates that thermal noise is Gaussian distributed with zero mean.

The noise power from a source (not necessarily a thermal source) can be specified by a number called the *effective noise temperature*:

$$T_e = \text{---}$$

Effective noise temperature can be interpreted as the temperature of a fictitious thermal noise source at the *input*, that would be required to produce the same noise power at the *output*.

Note that if the noise source is not thermal noise, then T_e may have nothing to do with the physical temperature of the device[4].

The important thing to note from this section is that *noise is inevitable*.

The additive noise channel

The simplest model for the effect of noise in a communication system is the additive noise channel, shown in Fig. Using this model the transmitted signal $s(t)$ is corrupted by the addition of

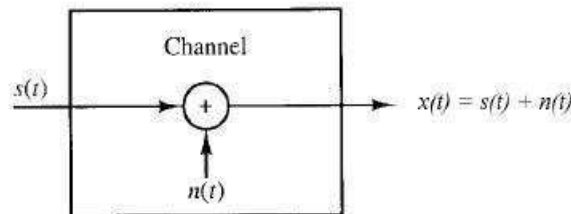


Figure: Additive noise Channel

a random noise signal $n(t)$. If this noise is introduced primarily by electronic components and amplifiers at the receiver, then we have seen that it can be characterized statistically as a Gaussian process. It turns out that the noise introduced in most physical channels is (at least approximately) Gaussian, and thus, this simple model is the predominant one used in communication system analysis and design[3].

A Statistical Description of Noise

As we have already hinted at, noise is completely random in nature. The noise signal $n(t)$ is a time-varying waveform, however, and just like any other signal it must be affected by the system through which it passes. We therefore need a model for the noise that allows us to answer questions such as: How does one quantitatively determine the effect of systems on noise? What happens when noise is picked up at the receiver and passed through a demodulator? And what effect does this have on the original message signal? Here we seek a representation that will enable us to answer such questions in the following chapters.

Background on Probability

Before developing a mathematical model for the noise, we need to define a few terms[3].

Random Variable

Consider a *random experiment*, that is, an experiment whose outcome cannot be predicted precisely.

The collection of all possible separately identifiable outcomes of a random experiment is

called the *sample space*, S . A *random variable* is a rule or relationship (denoted by x) that assigns a real number x_i to the i^{th} sample point in the sample space. In other words, the random variable x can take on values $x_i \in S$. The *probability* of the random variable x taking on the value x_i is denoted $P_x(x_i)$.

Cumulative and Probability Density Functions

The *cumulative density function* (cdf) of a random variable x is

$$F_x(x) = P_x(X \leq x)$$

The *probability density function* (pdf) of a random variable x is

$$P_x(x) = -F_x'(x)$$

Note that we use upper case P to denote probability, and lower case p to denote a pdf. We also have that

$$P(x_1 < x \leq x_2) = \int_{x_1}^{x_2} p(x) dx$$

One specific pdf that we will be particularly interested in is the *Gaussian* pdf, defined as

$$P_x(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

where m is the *mean* of the distribution and σ^2 is the *variance*. This pdf is shown in Figure.

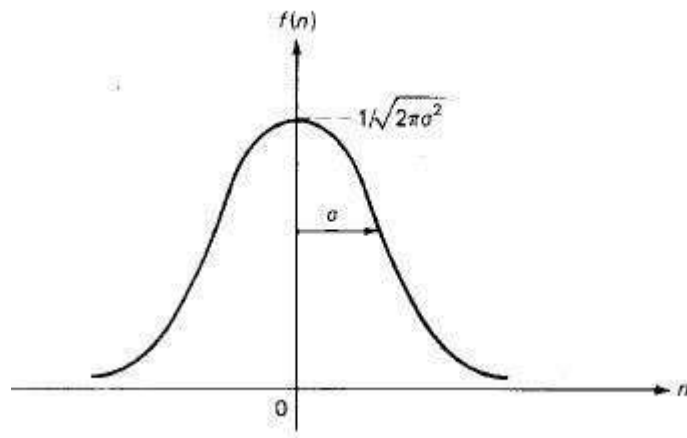


Figure:Gaussian Probilty Density Function

Statistical Averages

One is often interested in calculating averages of a random variable. Denote the *expected value* (or mean value) of a random variable x as $E\{x\}$. If the random variable x has a pdf $p_x(x)$, then the expected value is defined as

where $E\{\cdot\}$ denotes the expectation operator.

$$E\{x\} = \int_{-\infty}^{\infty} x p_x(x) dx \quad (1)$$

It is also often necessary to find the mean value of a function of a random variable, for example, the mean square amplitude of a random signal. Suppose we want to find $E\{y\}$ where y is a random variable given by

$$y = g(x);$$

where x is a random variable with a known pdf, and $g(\cdot)$ is an arbitrary function. Then,

$$E\{y\} = E\{g(x)\} = \int_{-\infty}^{\infty} g(x) p_x(x) dx \quad (2)$$

The *variance* of a random variable x is defined as

$$\begin{aligned} \sigma_x^2 &= E\{(x - E\{x\})^2\} \\ &= E\{x^2\} - E^2\{x\} \end{aligned}$$

Note that for a zero-mean random variable, the variance is equal to the mean square. **Random Processes**

A random time-varying function is called a *random process*[4]. Let $n(t)$ be such a process. A sample n of $n(t)$ taken at any time t is a random variable with some probability density function. Here we will only consider *stationary* processes, where a stationary process is defined as one for which a sample taken at time t is a random variable whose pdf is independent of t .

Recall that one can view a random variable as an outcome of a random experiment, where the outcome of each trial is a number. Similarly, a random process can be viewed as a random experiment where the outcome of each trial is a waveform that is a function of time. The collection of all possible waveforms is the *ensemble* of the random process.

Two averages exist for random processes. Consider the ensemble of noise processes shown in Figure. Each waveform in this figure represents a different outcome of a random experiment

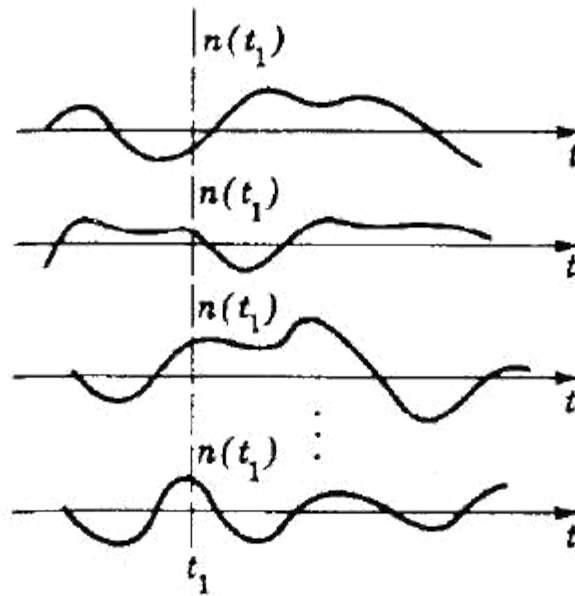


Figure. Ensemble Averages

Time Average

For a specific waveform, one could find the *time average*, defined as

$$\langle n(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T n^2(t) dt$$

where $\langle \cdot \rangle$ denotes the time average operator. Note that average power (1.2) is just the time average of the magnitude-square of the waveform[3].

Ensemble Average

Alternatively, one could pick a specific time and average across all sample functions of the process at that time. This would give the *ensemble average*

$$E\{n(t)\} = \int_{-\infty}^{\infty} n(t) p(n(t)) dn(t)$$

Comments

Note that the mean value $E\{n(t)\}$ locates the center of gravity of the area under the pdf. Random processes for which the time-average and the ensemble-average are identical are called *ergodic* processes. All the random processes encountered in this course will be assumed to be ergodic. The importance of this is that time averages can be easily measured, whereas ensemble averages cannot.

We can now interpret some engineering measures of the noise in terms of statistical quantities:

$$\text{DC component: } E\{n(t)\} = \langle n(t) \rangle$$

$$\text{Average power: } E\{n^2(t)\} = \langle n^2(t) \rangle$$

Notice that for a zero-mean process, the variance is equivalent to the average power, i.e., $\sigma^2 = E\{n^2(t)\}$. This could be measured in the lab using a power metre[3].

Autocorrelation and Power Spectral Density

To understand bandwidth issues relating to random signals, we must now find a reasonable spectral representation of $n(t)$. In particular, we are interested in a frequency representation that reflects an ensemble average of all possible random processes[3].

Autocorrelation

The frequency content of a process depends on how rapidly the amplitude changes as a function of time. This can be measured by correlating the amplitudes at times t_1 and t_2 . Define the *autocorrelation* of a real random process as

$$R_x(t_1, t_2) = E\{x(t_1)x(t_2)\}$$

For a stationary process, the autocorrelation depends only on the time difference, so

$$R_x(\tau) = E\{x(t)x(t + \tau)\}$$

Recall that the average power of a waveform is the mean square. Hence,

$$P = E\{x^2(t)\} = R_x(0)$$

Power Spectral Density

Power spectral density (PSD) is a function that measures the distribution of power of a random signal with frequency. To illustrate the idea, consider a power meter tuned to a frequency f_0 that measures the power in a very narrow bandwidth around f_0 ; the output of this metre would give a good approximation to the PSD at the frequency f_0 . PSD is only defined for stationary signals.

Theorem (Wiener-Khinchine Theorem)

The power spectral density of a random process is defined as the Fourier transform of the autocorrelation[3]:

$$S_x(f) = \int_{-\infty}^{\infty} R_x(\tau) e^{-j2\pi f\tau} d\tau$$

Since the autocorrelation is thus given by the inverse Fourier transform of the PSD, it follows from the equation that the average power of a random process can be found by integrating the PSD over all frequencies:

$$P = R_x(0) = \int_{-\infty}^{\infty} S_x(f) df$$

One particular example of a PSD that plays an extremely important role in communications and signal processing is one in which the PSD is constant over all frequencies, i.e.,

$$S(f) = \frac{P}{2B}$$

Noise having such a PSD is referred to as *white noise*, and is used in the same sense as white light which contains equal amounts of all frequencies within the visible band of electromagnetic radiation. Note that the factor $1/2$ is included to indicate that half the power is associated with

positive frequency and half with negative.

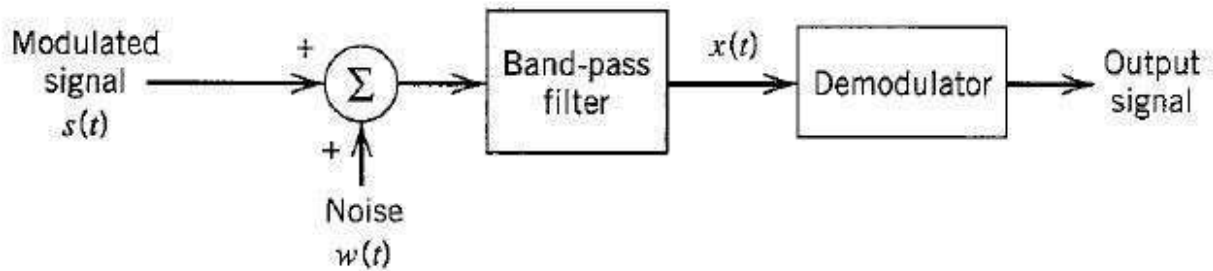


Figure. Receiver Model

Representation of Bandlimited Noise

Development

Any communication system that uses carrier modulation will typically have a bandpass filter at the front-end of the receiver see Fig . This filter will have a bandwidth wide enough to pass the modulated signal of interest, and is designed to restrict out-of-band noise from entering the receiver[4]. Any noise that does enter the receiver will therefore be bandpass in nature, i.e., its spectral magnitude is non-zero only for some band concentrated around the carrier frequency f_c . For example, if white noise have a PSD of $N_0/2$ is passed through such a filter, then the PSD of the noise that enters the receiver is given by

$$S_n(f) = \frac{N_0}{2} \text{ for } |f - f_c| \leq B \text{ and } 0 \text{ elsewhere}$$

and is shown in Fig.. We are now in a position to develop a representation specifically for such bandpass noise. To achieve this, we will use a simple artifice, namely, to visualize the noise as being composed of the sum of many closely spaced randomly-phased sine waves.

Consider the bandpass noise signal $n(t)$, whose PSD is given by equation and is shown in Fig. The average noise power in the frequency slices Δf at frequencies f_k and $-f_k$, is

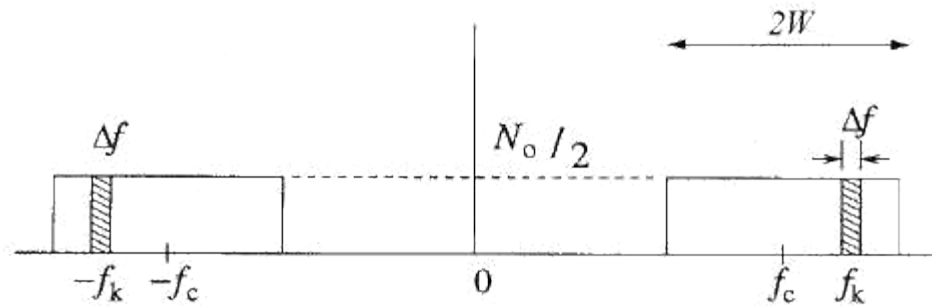


Figure. Power Spectral Density of band limited white noise process $n(t)$

found to be

$$P_k = \frac{N_0 B}{2}$$

where the factor of 2 is present because we are summing the slices at negative and positive frequencies.

For Δf small, the component associated with this frequency interval can be written $n_k(t) = a_k \cos(2\pi f_k t + \theta_k)$

where θ_k is a random phase assumed independent and uniformly distributed in the range $[0; 2\pi)$, and a_k is a random amplitude. It can be shown that the average power of the randomly-phased sinusoid is

$$P_k = E\{a_k^2\}$$

The complete bandpass noise waveform $n(t)$ can be constructed by summing up such sinusoids over the entire band, i.e.,

$$n(t) = \sum_{k=-\infty}^{\infty} a_k \cos(2\pi f_k t + \theta_k)$$

where

Now, let $f_k = (f_c + B_k) + f_c$, and using the identity for the $\cos(\cdot)$ of a sum^k we obtain the required result.

Result

$$n(t) = n_c(t) \cos 2\pi f_c t - n_s(t) \sin 2\pi f_c t$$

where

$$n_c(t) = \sum_{k=-\infty}^{\infty} a_k \cos(2\pi (f_k - f_c) t + \theta_k)$$

$$n_s(t) = \sum_{k=-\infty}^{\infty} a_k \sin(2\pi (f_k - f_c) t + \theta_k)$$

From equation we see that

$\Delta f = B$. Hence, $n_c(t)$ and $n_s(t)$ are baseband signals. The

representation for $n(t)$ given by equation is the representation we seek, and is referred to as the *bandpass representation*. Although we have derived it for the specific case of a bandlimited white noise process, it is actually a very general representation that can be used for *any* bandpass signal.

Average power and power spectral density

If this representation of bandpass noise is to be of use in our later analyses, we must find suitable

statistical descriptions[3]. Hence, we will now derive the average power in $n(t)$, together with the average power and PSD for both $n_s(t)$ and $n_c(t)$.

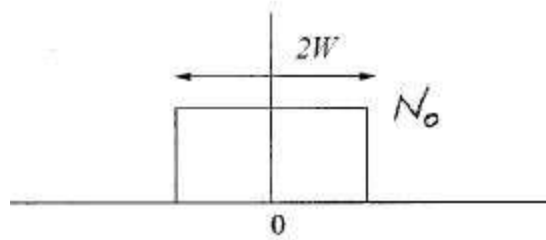


Figure: Power spectral density of each of the baseband noise processes $n_c(t)$ and $n_s(t)$

The average power in $n(t)$ is $P_n = E\{n^2(t)\}$. Recall that for a zero-mean Gaussian process the average power is equal to the variance

$$P_n = E\{n^2(t)\} = E\left\{\sum_{k=-\infty}^{\infty} \left[\frac{1}{\sqrt{2}} \left(a_k \cos(\omega_k t + \phi_k) + a_{-k} \cos(\omega_{-k} t + \phi_{-k}) \right) \right]^2 \right\}$$

Since we have assumed the phase terms are independent, it follows that

$$E\{\cos(\omega_k t + \phi_k) \cos(\omega_l t + \phi_l)\} = 0 \text{ for } k \neq l$$

and

$$E\{\cos^2(\omega_k t + \phi_k)\} = \frac{1}{2} \text{ for } k = l$$

Hence,

$$P = E\{n^2(t)\} = \sum_{k=-\infty}^{\infty} \frac{1}{2} a_k^2 = \frac{1}{2} \sum_{k=-\infty}^{\infty} a_k^2$$

This is what you should intuitively expect to obtain, given in equation. A similar derivation for each of $n_c(t)$ and $n_s(t)$ reveals that

$$P_c = E\{n_c^2(t)\} = \frac{1}{2} \sum_{k=-\infty}^{\infty} a_k^2 = \frac{1}{2} \sum_{k=-\infty}^{\infty} a_k^2$$

and

$$P_s = E\{n_s^2(t)\} = \frac{1}{2} \sum_{k=-\infty}^{\infty} a_k^2 = \frac{1}{2} \sum_{k=-\infty}^{\infty} a_k^2$$

Thus, the average power in *each* of the baseband waveforms $n_c(t)$ and $n_s(t)$ is *identical* to the average power in the bandpass noise waveform $n(t)$.

Now, considering the PSD of $n_c(t)$ and $n_s(t)$, we note from above equations that each of these waveforms consists of a sum of closely spaced baseband sinusoids. Thus, each baseband noise waveform will have the same PSD, which is shown in Fig.. Since the average power in each of the baseband waveforms is the same as the average power in the bandpass waveform, it follows that the area under the PSD in Fig. must equal the area under the PSD in Fig. The PSD of $n_c(t)$ and $n_s(t)$ is therefore given by

$$S_c(f) = S_s(f) = \frac{1}{2} S_n(f) \quad , \quad |f| < B$$

A phasor interpretation

Finally, we will interpret the bandpass representation in another way. Notice that equation can be written

$$n(t) = \text{Re}\{g(t) e^{j2\pi f_c t}\}$$

where

$$g(t) = n_c(t) + jn_s(t)$$

and $\text{Re}\{\cdot\}$ denotes the real part. We can also write $g(t)$ in terms of magnitude and phase as

$$g(t) = r(t) e^{j\theta(t)}$$

where

$r(t)$ is the envelope and $\theta(t)$ is the phase of $g(t)$

the noise. The phasor diagram representation[2] is shown in Fig. Because of this representation,

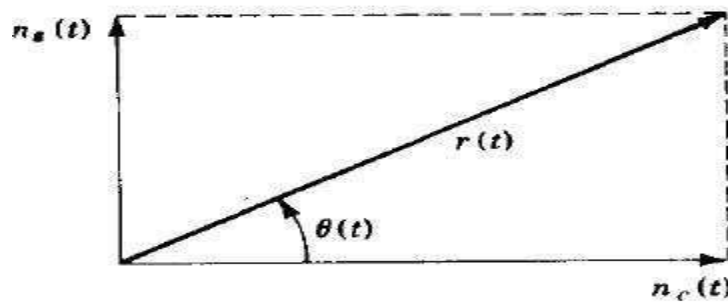


Figure. Phasor representation of bandlimited noise

tation, $n_c(t)$ is often referred to as the *in-phase* component, and $n_s(t)$ as the *quadrature-phase* component. Substituting the magnitude-phase representation for $g(t)$ gives

$$n(t) = r(t) \cos[2\pi f_c t + \theta(t)]$$

This is an intuitively satisfying result. Since the bandpass noise $n(t)$ is narrow band in the vicinity of f_c , one would expect it to be oscillating on the average at f_c . It can be shown that if $n_c(t)$ and $n_s(t)$ are Gaussian-distributed, then the magnitude $r(t)$ has a Rayleigh distribution, and the phase $\theta(t)$ is uniformly distributed.

Noise in Analog Communication Systems

Background

You have previously studied ideal analog communication systems. Our aim here is to compare the performance of different analog modulation schemes in the presence of noise[4]. The performance will be measured in terms of the signal-to-noise ratio (SNR) at the output of the receiver, defined as

$$\text{SNR}_O = \frac{\text{Signal Power}}{\text{Noise Power}}$$

Note that this measure is unambiguous if the message and noise are additive at the receiver output;
we will see that in some cases this is not so, and we need to resort to approximation methods to obtain a result.

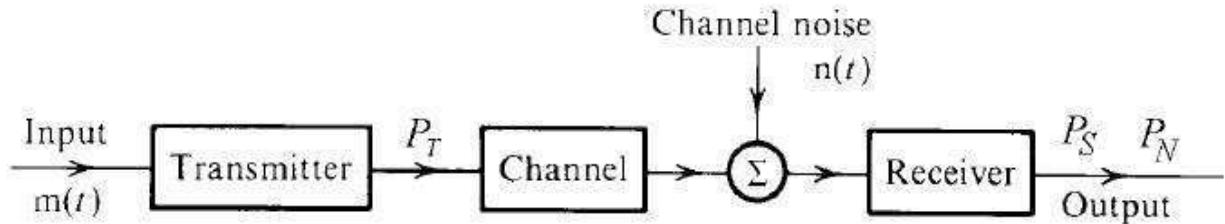


Figure. Model of an analog communication system.

A model of a typical communication system is shown in Fig. 3.1, where we assume that a modulated signal with power P_T is transmitted over a channel with additive noise. At the output of the receiver the signal and noise powers are P_S and P_N respectively, and hence, the output SNR is $\text{SNR}_O = P_S/P_N$. This ratio can be increased as much as desired simply by increasing the transmitted power. However, in practice the maximum value of P_T is limited by considerations such as transmitter cost, channel capability, interference with other channels, etc. In order to make a fair comparison between different modulation schemes, we will compare systems having the same transmitted power. Also, we need a common measurement criterion against which to compare the different modulation schemes. For this, we will use the *baseband* SNR. Recall that all modulation schemes are bandpass (i.e., the modulated signal is centered around a carrier frequency). A baseband communication system is one that does not use modulation. Such a scheme is suitable for transmission over wires, say, but is not terribly practical. As we will see, however, it does allow a direct performance comparison of different schemes.

Baseband Communication System

A baseband communication system[4] is shown in Fig., where $m(t)$ is the band-limited message signal, and W is its bandwidth

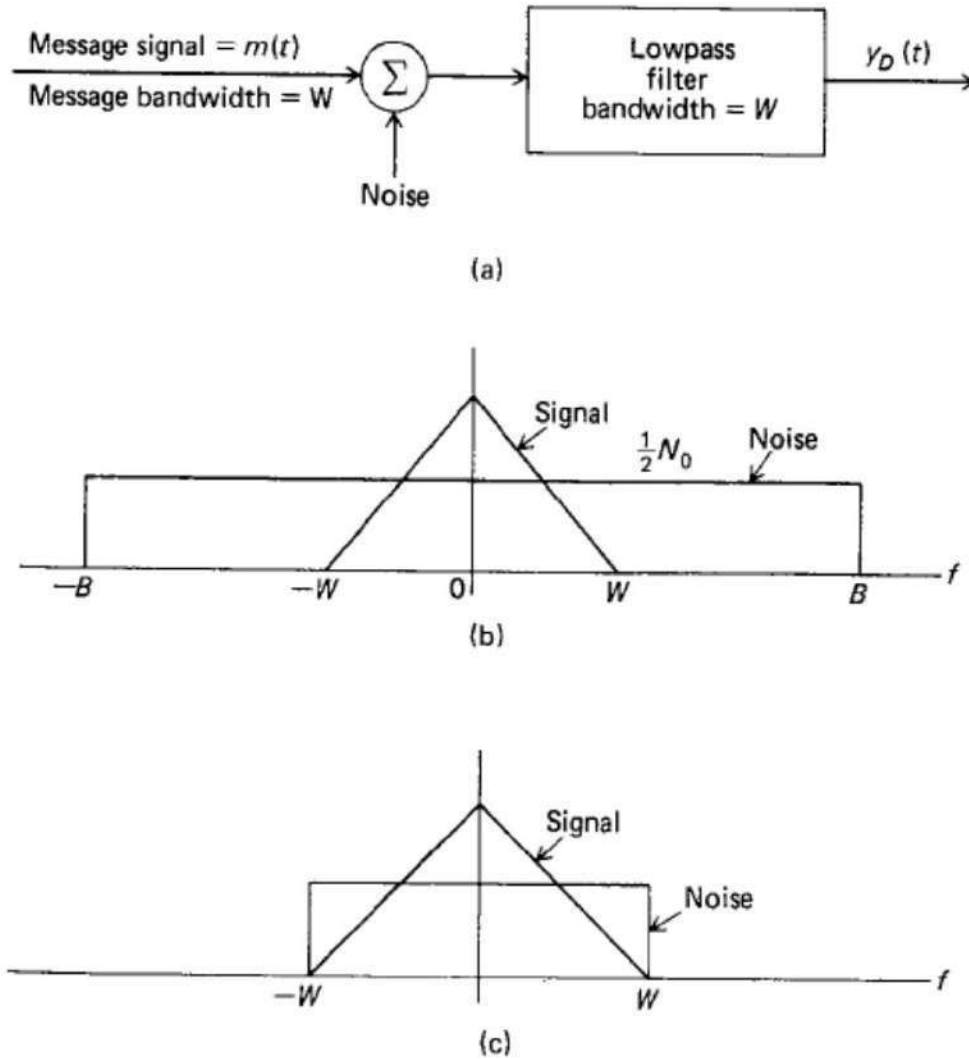


Figure. Baseband Communication System: (a)Model (b) signal spectra at filter input,and(c)signal spectra at filter output.

An example signal PSD is shown in Fig. The average signal power is given by the area under the triangular curve marked “Signal”, and we will denote it by P . We assume that the additive noise has a double-sided white PSD of $N_0/2$ over some bandwidth $B > W$, as shown in Fig. For a basic baseband system, the transmitted power is identical to the message power, i.e., $P_T = P$.

The receiver consists of a low-pass filter with a bandwidth W , whose purpose is to enhance the SNR by cutting out as much of the noise as possible. The PSD of the noise at the output of the LPF is shown in Fig. , and the average noise power is given by

$$\int_{-W}^W \frac{N_0}{2} df = N_0 W$$

Thus, the SNR at the receiver output is

$$\text{SNR}_{\text{baseband}} = \frac{P_s}{N_0 W}$$

Notice that for a baseband system we can improve the SNR by: (a) increasing the transmitted power, (b) restricting the message bandwidth, or (c) making the receiver less noisy.

Amplitude Modulation

Review

In amplitude modulation[3], the *amplitude* of a sinusoidal carrier wave is varied linearly with the message signal. The general form of an AM signal is

$$s(t)_{\text{AM}} = [A + m(t)] \cos(2\pi f_c t)$$

where A is the amplitude of the carrier, f_c is the carrier frequency, and $m(t)$ is the message signal. The *modulation index*, is defined as

$$m_p = \frac{\max |m(t)|}{A}$$

where m_p is the peak amplitude of $m(t)$, i.e., $m_p = \max |m(t)|$

m_p), then the envelope of $s(t)$ will have the same shape as the message $m(t)$, and thus, a simple envelope detector can be used to demodulate the AM wave. The availability of a particularly simple receiver is the major advantage of AM, since as we will see, its noise performance is not great.

If an envelope detector cannot be used, another form of detection known as *synchronous detection* can be used. The block diagram of a synchronous detector is shown in Fig.

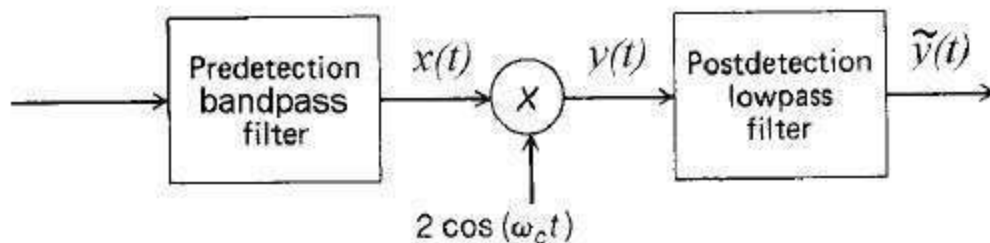


Figure. Synchronous demodulator

process involves multiplying the waveform at the receiver by a local carrier of the same frequency (and phase) as the carrier used at the transmitter. This basically replaces the $\cos(\omega_c t)$ term by a $\cos(2\omega_c t)$ term. From the identity

$$2 \cos^2(x) = 1 + \cos(2x)$$

the result is a frequency translation of the message signal, down to baseband (i.e., $f = 0$) and up to twice the carrier frequency. The low-pass filter is then used to recover the baseband message signal. As one might expect, the main disadvantage with this scheme is that it requires generation of a local carrier signal that is perfectly synchronized with the transmitted carrier.

Notice in equation that the AM signal consists of two components, the carrier $A \cos(2\pi f_c t)$ and the sidebands $m(t) \cos(2\pi f_c t)$. Since transmitting the carrier term is wasteful, another variation of AM that is of interest is one in which the carrier is suppressed. This is referred to as *double side band suppressed carrier* (DSB-SC), and is given by

$$s(t)_{\text{DSB-SC}} = A m(t) \cos(2\pi f_c t)$$

In this case the envelope of the signal looks nothing like the original message signal, and a synchronous detector must be used for demodulation.

Noise in DSB-SC

The *predetection* signal (i.e., just before the multiplier[3] in Fig.

$$x(t) = s(t) + n(t)$$

The purpose of the predetection filter is to pass only the frequencies around the carrier frequency, and thus reduce the effect of out-of-band noise. The noise signal $n(t)$ after the predetection filter is bandpass with a double-sided white PSD of $N_0/2$ over a bandwidth of $2W$

(centered on the carrier frequency), as shown in Fig. Hence, using the bandpass representation the predetection signal is

$$x(t) = [A m(t) + n_c(t)] \cos(2\pi f_c t) - n_s(t) \sin(2\pi f_c t)$$

After multiplying by $2 \cos(2\pi f_c t)$, this becomes

$$\begin{aligned} y(t) &= 2 \cos(2\pi f_c t) x(t) \\ &= A m(t) [1 + \cos(4\pi f_c t)] + n_c(t) [1 + \cos(4\pi f_c t)] - n_s(t) \sin(4\pi f_c t) \end{aligned}$$

where we have used and

$$2 \cos x \sin x = \sin(2x)$$

Low-pass filtering will remove all of the $2f_c$ frequency terms, leaving

$$y(t) = A m(t) + n_c(t)$$

The signal power at the receiver output is

where, recall, P is the power in the message signal $m(t)$. The power in the noise signal $n_c(t)$ is

since from (2.34) the PSD of $n_c(t)$ is N_0 and the bandwidth of the LPF is W . Thus, for the DSB-SC synchronous demodulator, the SNR at the receiver output is

To make a fair comparison with a baseband system, we need to calculate the transmitted power

$$P_T = E\{A^2 m^2(t) \cos^2(2\pi f_c t)\}$$

and substitution gives

$$\text{SNR}_O = \frac{P}{P_N}$$

Comparison gives

$$\text{SNR}_{\text{DSB-SC}} = \text{SNR}_{\text{baseband}}$$

We conclude that a DSB-SC[3] system provides no SNR performance gain over a baseband system.

It turns out that an SSB system also has the same SNR performance as a baseband system.

Noise in AM, Synchronous Detection

For an AM waveform, the predetection signal[3] is

$$x(t) = [A + m(t) + n_c(t)] \cos(2\pi f_c t) - j n_s(t) \sin(2\pi f_c t)$$

After multiplication by $2 \cos(2\pi f_c t)$, this becomes

$$y(t) = A[1 + \cos(4\pi f_c t)] + m(t)[1 + \cos(4\pi f_c t)] + n_c(t)[1 + \cos(4\pi f_c t)] - n_s(t) \sin(2\pi f_c t)$$

After low-pass filtering this becomes

$$y(t) = A + m(t) + n_c(t)$$

Note that the DC term A can be easily removed with a DC block (i.e., a capacitor), and most AM demodulators are not DC-coupled.

The signal power at the receiver output is

$$P_s = E\{m^2(t)\}g = P$$

and the noise power is

$$P_N = 2N_0W$$

The SNR at the receiver output is therefore

$$SNR_o = \frac{P}{2N_0W}$$

The transmitted power for an AM waveform is

and substituting this into the baseband SNR, we find that for a baseband system with the same transmitted power

$$SNR_{baseband} = \frac{P}{N_0W}$$

Thus, for an AM waveform using a synchronous demodulator we have

$$SNR_{AM} = \frac{SNR_{baseband}}{2}$$

In other words, the performance of AM is always worse than that of a baseband system. This is because of the wasted power which results from transmitting the carrier explicitly in the AM waveform.

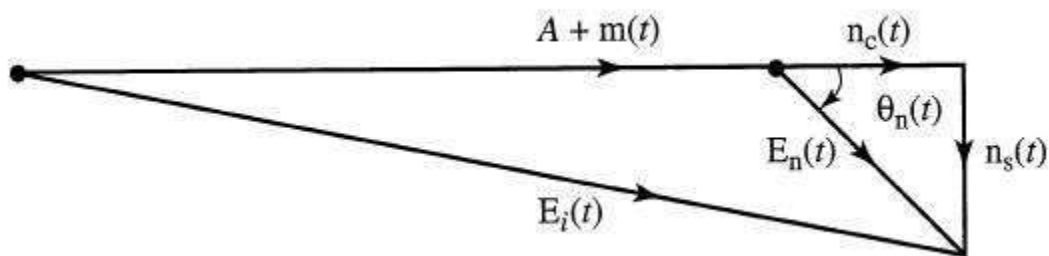


Figure. Phasor diagram of the signals present at an AM receiver

Noise in AM, Envelope Detection

Recall that an envelope detector [4] can only be used if $\gamma < 1$. An envelope detector works by detecting the *envelope* of the received signal. To get an appreciation of the effect of this, we will represent the received signal by phasors, as shown in Fig. The receiver output, denoted by “ $E(t)$ ” in the figure, will be given by

$$y(t) = \text{envelope of } x(t) = \sqrt{1 + (n(t))^2 + 2n(t)}$$

This expression is somewhat more complicated than the others we have looked at, and it is not immediately obvious how we will find the SNR at the receiver output. What we would like is an approximation to $y(t)$ in which the message and the noise are additive.

(a) Small Noise Case

The receiver output can be simplified if we assume that for almost all t the noise power is small [2], i.e., $n(t) \ll [A + m(t)]$. Hence

Then, most of the time,

which is identical to the post-detection signal in the case of synchronous detection. Thus, (ignoring the DC term A again) the output SNR is

$$SNR_O = 2$$

which can be written in terms of baseband SNR as

$$SNR_{env} = 2 + 2 \quad SNR_{baseband} = 2$$

Note that whereas $SNR_{AM} = SNR_{baseband}$ is valid always, the expression for SNR_{env} is *only* valid for small noise power.

Large Noise Case

Now consider the case where noise power is large [2], so that for almost all t we have $n(t) \gg [A + m(t)]$. Rewrite as

$$y^2(t) = [A + m(t)]^2 + 2[A + m(t)]n(t) + n^2(t) = [A + m(t)]^2 + n^2(t) + 2[A + m(t)]n(t)$$

For $n(t) \gg [A + m(t)]$, this reduces to

$$y^2(t) = n^2(t) + 2[A + m(t)]n(t) \approx \frac{[2(A + m(t))n(t)]^2}{n^2(t)} = 4(A + m(t))^2$$

Where

$$E_n(t) = \sqrt{2(A + m(t))^2} = 2(A + m(t))$$

$n(t)$ is the envelope of the noise (as described). But from the phasor diagram in Fig. , we have $n_c(t) = E_n(t) \cos(t)$, giving

$$y(t) \approx E_n(t) \cos(t)$$

Further,

$$y(t) \approx (1 + x) \cos(t) + x \cos(t) \quad \text{for } x \ll 1, \text{ so this reduces to}$$

$$y(t) \approx (1 + x) \cos(t) + x \cos(t) = (1 + x) \cos(t)$$

The main thing to note is that the output of the envelope detector contains no term that is proportional to the message $m(t)$. The term $m(t) \cos(t)$ is the message multiplied by a noise term $\cos(t)$, and is no use in recovering $m(t)$. This multiplicative effect corrupts the message to a far greater extent than the additive noise in our previous analysis; the result is that there is a complete loss of information at the receiver. This produces a *threshold effect*, in that below some carrier power level, the performance of the detector deteriorates very rapidly. Despite this threshold effect, we find that in practice it does not matter terribly. This is because the quality of a signal with an output SNR less than about 25 dB is so poor, that no-one would really want to listen to it anyway. And for such a high output SNR, we are well past the threshold level and we find that equation holds. From a practical point of view, the threshold effect is seldom of importance for envelope detectors.

Frequency Modulation

Having studied the effect of additive noise on amplitude modulation systems[1], we will now look at the SNR performance on frequency modulation systems. There is a fundamental difference between these two. In AM, the message information is contained within the amplitude of the signal, and since the noise is additive it adds directly to the modulated signal. For FM, however, it is the frequency of the modulated signal that contains the message information. Since the frequency of a signal can be described by its zero crossings, the effect of noise on an FM signal is determined by the extent to which it changes the zero crossing of the modulated signal. This suggests that the effect of noise on an FM signal will be less than that for an AM system, and we will see in this section that this is in fact the case.

Review

Consider the following general representation of a carrier waveform[1]

$$s(t) = A \cos[\theta(t)]$$

where $\theta(t)$ is the instantaneous phase angle. Comparing this with the generic waveform $A \cos(2\pi f t)$, where f is the frequency, we can define the instantaneous frequency as

For an FM system, the instantaneous frequency of the carrier is varied linearly with the message, i.e.,

$$f(t) = f_c + k_f m(t)$$

where k_f is the frequency sensitivity of the modulator. Hence, the instantaneous phase is

and the modulated signal is

There are two things to note about the FM signal: (a) the envelope is constant, and (b) the signal $s(t)$ is a non-linear function of the message signal $m(t)$.

Bandwidth of FM

Let the peak message amplitude be A_m and f_m (so that the instantaneous frequency will vary between $f_c - \Delta f$ and $f_c + \Delta f$). Denote the deviation of the instantaneous frequency from the carrier frequency as the *frequency deviation* Δf .

Define the *deviation ratio* (also called the FM modulation index in the special case of tone modulated FM) as

where W is the message bandwidth.

Unlike AM, the bandwidth of FM is not dependent simply on the message bandwidth. For Small β the FM bandwidth is approximately twice the message bandwidth (referred to as narrowband FM). But for large β (referred to as wide-band FM) the bandwidth can be much larger than this. A useful rule-of-thumb for determining the transmission bandwidth of an FM signal is *Carson's rule*:

Observe that for $\beta \ll 1$, $\approx 2W$ (as is the case in AM). At the other extreme, for $\beta \gg 1$, $\approx 2\Delta f$, which is independent of W .

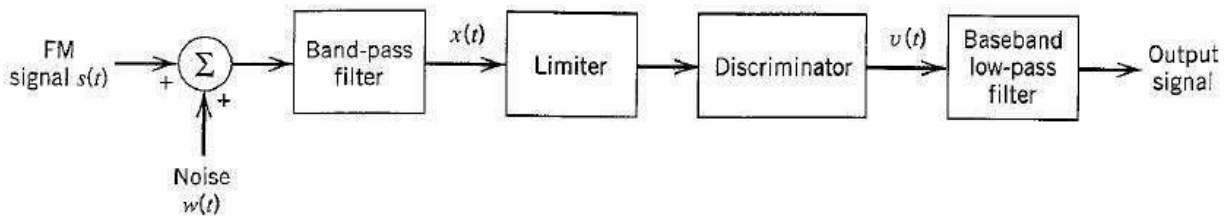


Figure: Model of an FM receiver.

Noise in FM

The model of an FM receiver[1] is shown in Fig., where $S(t)$ is the FM signal, and $w(t)$ is white Gaussian noise with power spectral density $N_0/2$. The bandpass filter is used to remove any signals outside the bandwidth of $\pm B$, and thus, the predetection noise at the receiver is bandpass with a bandwidth of $2B$. Since an FM signal has a constant envelope, the limiter is used to remove any amplitude variations. The discriminator is a device whose output is proportional to the deviation in the instantaneous frequency (i.e., it recovers the message signal), and the final baseband low-pass filter has a bandwidth of W and thus passes the message signal and removes out-of-band noise.

The predetection signal is

$$x(t) = A \cos(2\pi f_c t + \theta(t))$$

First, let us consider the signal power at the receiver output. When the predetection SNR is high, it can be shown that the noise does not affect the power of the signal at the output.

Thus, ignoring the noise, the instantaneous frequency of the input signal is

$$f_i = f_c + \frac{1}{2\pi} \frac{d\theta(t)}{dt}$$

and the output of the discriminator (which is designed to simply return the deviation of the instantaneous frequency away from the carrier frequency) is

The output signal power is therefore

$$P_s = \frac{A^2}{2}$$

where

P is the average power of the message signal.

Now, to calculate the noise power at the receiver output, it turns out that for high predetection SNR the noise output is approximately independent of the message signal. In this case, we only have the carrier and noise signals present. Thus,

The phasor diagram of this is shown in Fig. From this diagram, we see that the instantaneous phase is

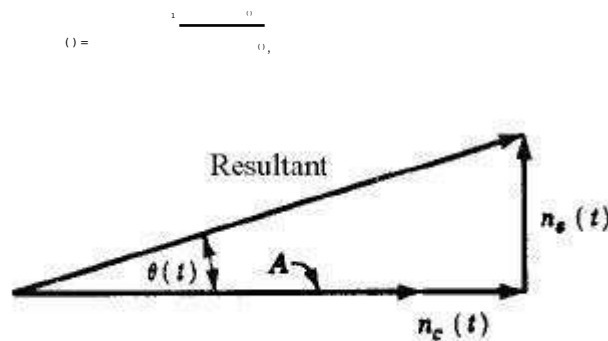


Figure: Phasor diagram of the FM carrier and noise signals.

For large carrier power, then most of the time

$$\theta(t) \approx \frac{1}{2\pi} \frac{d\theta(t)}{dt}$$

where the last line follows from $\tan \epsilon \approx \epsilon$ for small ϵ . But, the discriminator output is the instantaneous frequency, given by

$$f_i = f_c + \frac{1}{2\pi} \frac{d\theta(t)}{dt}$$

We know the PSD of $\theta(t)$ shown in Fig., but what is the PSD of $d\theta(t)/dt$?

Fourier theory tells us that:

(if $x(t) \rightarrow X(f)$)
then

$$\frac{d}{dt} x(t) \rightarrow j2\pi f X(f) \quad (1)$$

In other words, differentiation with respect to time is the same as passing the signal through a system having a transfer function of $H(f) = j2\pi f$. It can be shown that if a signal with PSD

$S_x(f)$ is passed through a linear system with transfer function $H(f)$, then the PSD at the output of the system is

$$S_y(f) = |H(f)|^2 S_x(f)$$

If the PSD of $x(t)$ has a value of N_0 within the band $\pm B_T/2$ as shown in Fig, then

$d(t)$ has a PSD of $S_{N_d}(f)$. The PSD of $d(t)$ before and after the baseband LFP

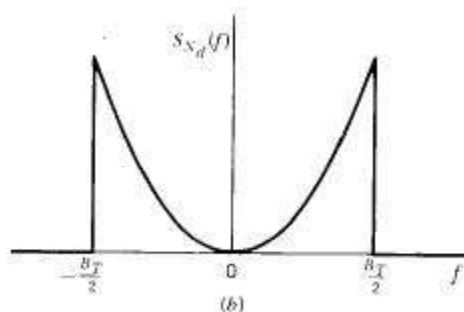
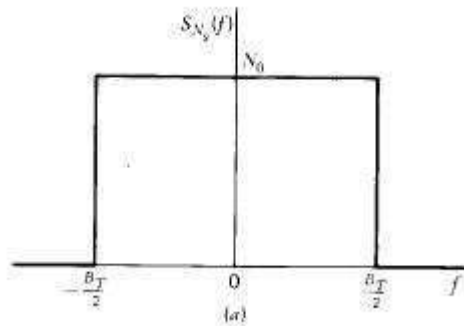
is shown in Fig. (b) and (c) respectively.

Returning to equation, that the PSD of $d(t)$ is known, we can calculate the average

noise power at the receiver output. It is given by

$$P_{N_d} = \int_{-B_T/2}^{B_T/2} S_{N_d}(f) df \quad (2)$$

where f is the PSD of the noise component at the discriminator output (i.e., the PSD of $f_i(t)$ in equation); the limits of integration are taken between $-W$ and W to reflect the fact that the output



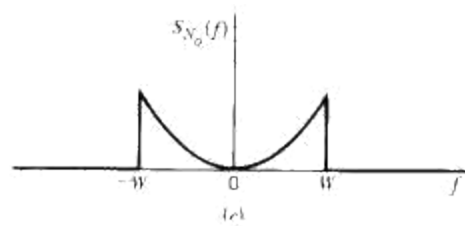


Figure: Power spectral densities for FM noise analysis: (a) $n_s(t)$ (b) $dn_s(t)/dt$ and (c) noise at the receiver output,

signal is low-pass filtered. Thus,

This expression is quite important, since it shows that the average noise power at the output of a FM receiver is inversely proportional to the carrier power $P_c/2$. Hence, increasing the carrier power has a *noise quieting* effect. This is one of the major advantages of FM systems. Finally, we have that at the output the SNR is

Since the transmitted power of an FM waveform is

$$P_T = \frac{P_c}{2}$$

substitution into earlier equation gives

$$SNR_{FM} = \frac{3}{2} \frac{P_c}{P_n} = \frac{3}{2} \frac{P_T}{P_n}$$

The SNR expression[2] is based on the assumption that the carrier power is large compared to the noise power. It is found that, like an AM envelope detector, the FM detector exhibits a

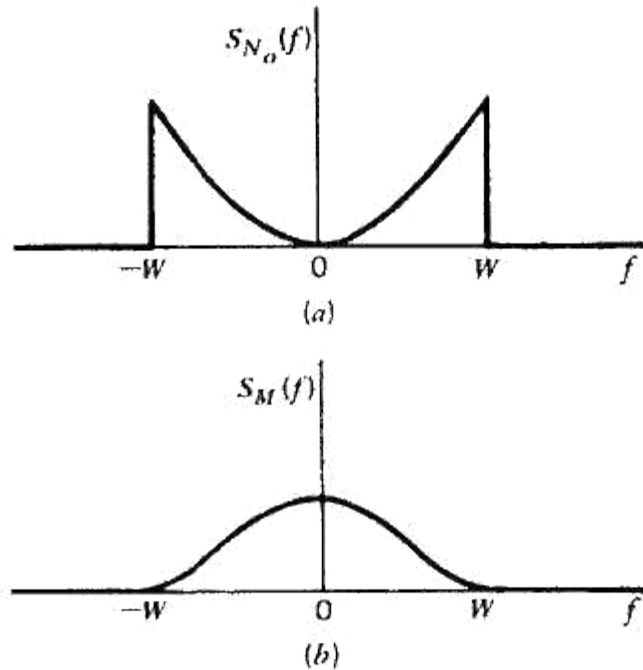


Figure : Power spectral densities of: (a) noise at the output of FM receiver, and (b) a typical message signal

threshold effect. As the carrier power decreases, the FM receiver breaks, as Haykin describes: “At first, individual clicks are heard in the receiver output, and as the carrier-to-noise ratio decreases still further, the clicks rapidly merge into a crackling or sputtering sound”. Experimental studies indicate that this noise mutilation is negligible in most cases if the predetection SNR (i.e., just after the receiver bandpass filter) is above 10. In other words, the threshold point occurs around

$$\frac{P_c}{P_n} = 10$$

Where, recall, $P_n = 2W(\beta + 1)$. For predetection SNRs above this value, the output SNR is given by equation

One should note that whereas equation suggests that output SNR for an FM system can be increased arbitrarily by increasing β while keeping the signal power fixed, inspection of equation shows this not to be strictly true. The reason is that if β increases too far, the condition of equation that we are above threshold may no longer be true, meaning that equation no longer provides an expression for the true SNR.

3.4.3 Pre-emphasis and De-emphasis

There is another way in which the SNR of an FM system may be increased [2][3]. We saw in the previous subsection that the PSD of the noise at the detector output has a square-law dependence on frequency. On the other hand, the PSD of a typical message source is not uniform, and typically rolls off at around 6 dB per decade (see Fig.). We note that at high frequencies the

relative message power is quite low, whereas the noise power is quite high (and is rapidly increasing). It is possible that this situation could be improved by reducing the bandwidth of the transmitted message (and the corresponding cutoff frequency of the baseband LPF in the receiver), thus rejecting a large amount of the out-of-band noise. In practice, however, the distortion introduced by low-pass filtering the message signal is unsatisfactory [2].

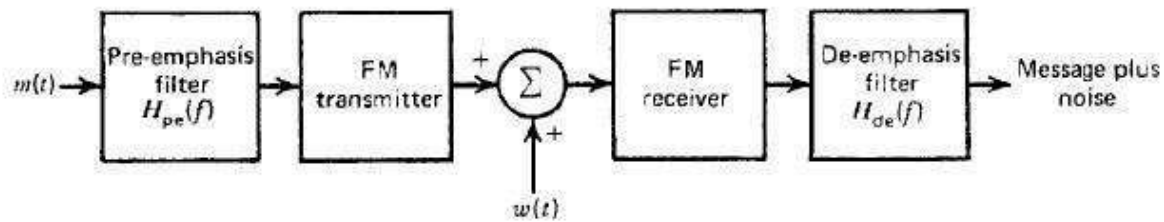
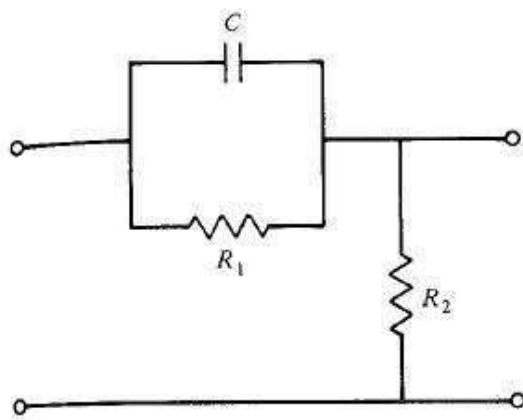
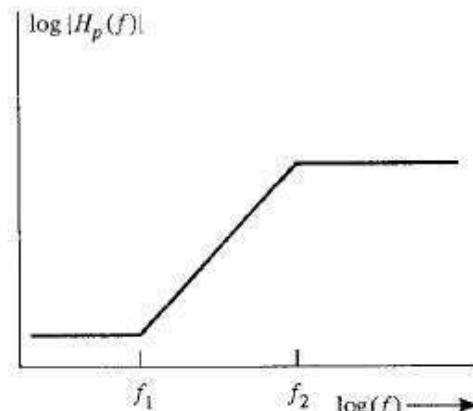


Figure: Pre-emphasis and de-emphasis in an FM system.



$$H_p(f) = K \frac{1 + j(f/f_1)}{1 + j(f/f_2)} \quad \text{where } f_1 = \frac{1}{2\pi R_1 C}, f_2 = \frac{1}{2\pi R_2} = \frac{R_1 + R_2}{2\pi R_1 R_2 C}$$

(a) Preemphasis Filter



(b) Bode Plot of Preemphasis Frequency Response

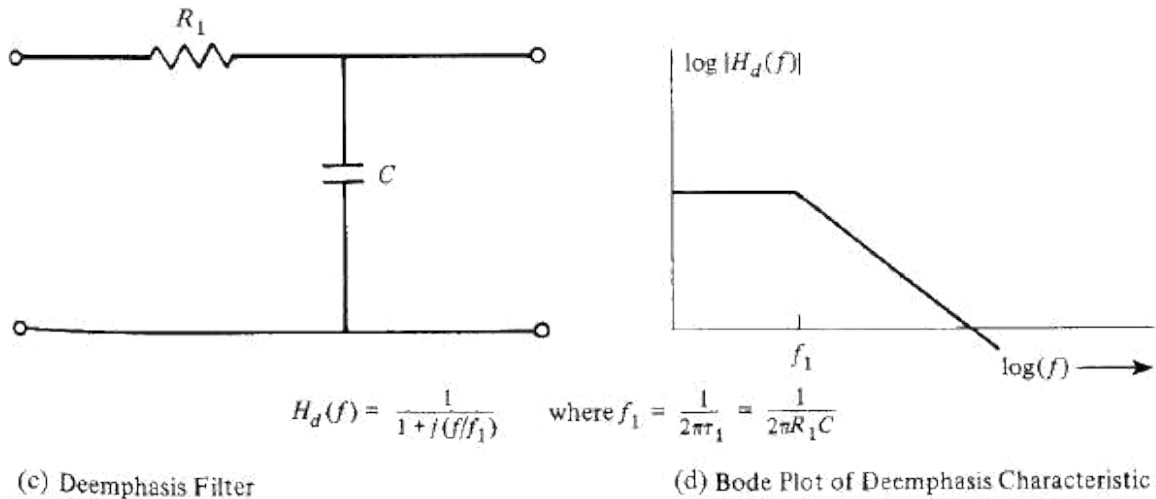


Figure: Simple linear pre-emphasis and de-emphasis circuits.

A better solution is obtained by using the *pre-emphasis* and *de-emphasis*[1] stages shown in Fig. . The intention of this scheme is that (f) is used to artificially emphasize the high frequency components of the message prior to modulation, and hence, before noise is introduced. This serves to effectively equalize the low- and high-frequency portions of the message PSD such that the message more fully utilizes the bandwidth available to it. At the receiver, (f) performs the inverse operation by de-emphasizing the high frequency components, thereby restoring the original PSD of the message signal.

Simple circuits that perform pre- and de-emphasis are shown in Fig. , along with their respective frequency responses. Haykin shows that these circuits can improve the output SNR by around 13 dB. In closing this section, we also note that Dolby noise reduction uses an analogous pre-emphasis technique to reduce the effects of noise.

REFERENCES:

[1] H. Taub, D. L Schilling, G. Saha; “Principles of Communication System,” 3rd Edition; 2008, Tata McGraw Hill, India; ISBN: 0070648115.

[2] Modern Digital and Analogue Communication Systems by “B.P. Lathi and Z Ding,” 4th Edition 2010, Oxford University Press, New Delhi.

[3] Communication Systems by “Simon Haykin,” 4th Edition, John Wiley & Sons, Inc.

[4] Communication systems Analog and Digital by sanjay Sharma ,4th edition, katson publications