

FREQUENCY MODULATION ERROR CALCULATION

D. L. NYRE

**Douglas Aircraft Company
Missiles & Space Systems Division
Santa Monica, California**

Summary In a frequency modulation process, small amplitude sidebands near the deviation limits are sometimes distorted or lost. The problems and expense of preserving these lesser sidebands are often great, and spectrum space must be reserved for them. To date there appears to have been no clear statement of their importance in terms of data accuracy.

This paper presents a numerical calculation of the errors accrued in a demodulated signal after truncation of the FM spectrum obtained with sinusoidal modulation. It is shown that the error increases for decreasing bandwidth. Error is evaluated for several bandwidths, and relationships to power outside the band are noted. Applications include specification of bandwidth for a given error in demodulated signal. The use of a limiter is assumed.

Introduction The purpose of this paper is to describe the procedure and results of a study that reveals the error in sharply bandpass-limited FM. The procedure consists of simulating each step in an FM process by means of frequency analysis and comparing the demodulated output with the input. The RMS error has been plotted for various signal amplitudes, frequencies, modulation indices, and bandpass filters. The method developed is general in nature and can be used with any set of data.

Figure 1 shows a block diagram of an FM system. Each block in this system is simulated as follows.

Cosine Oscillator and Voltage-Controlled Oscillator The modulated signal is taken to be a cosine wave as follows:

$$e_m = E_m \cos 2 \pi F_m t \quad (1)$$

The commonly accepted model for a sine wave carrier which in turn is frequency modulated by a cosine wave is: (1)

$$e = E_c \sin \left[2 \pi F_c t + \left(\frac{E_m \Delta F_c}{F_m} \right) \sin 2 \pi F_m t \right] \quad (2)$$

where:

- e = instantaneous value of the modulated carrier
- E_c = peak value of the carrier
- F_m = modulating frequency
- F_c = carrier frequency
- t = time
- E_m = peak value of modulating signal (0 to 1.0)
- e_m = instantaneous value of modulating signal
- ΔF_c = peak carrier deviation

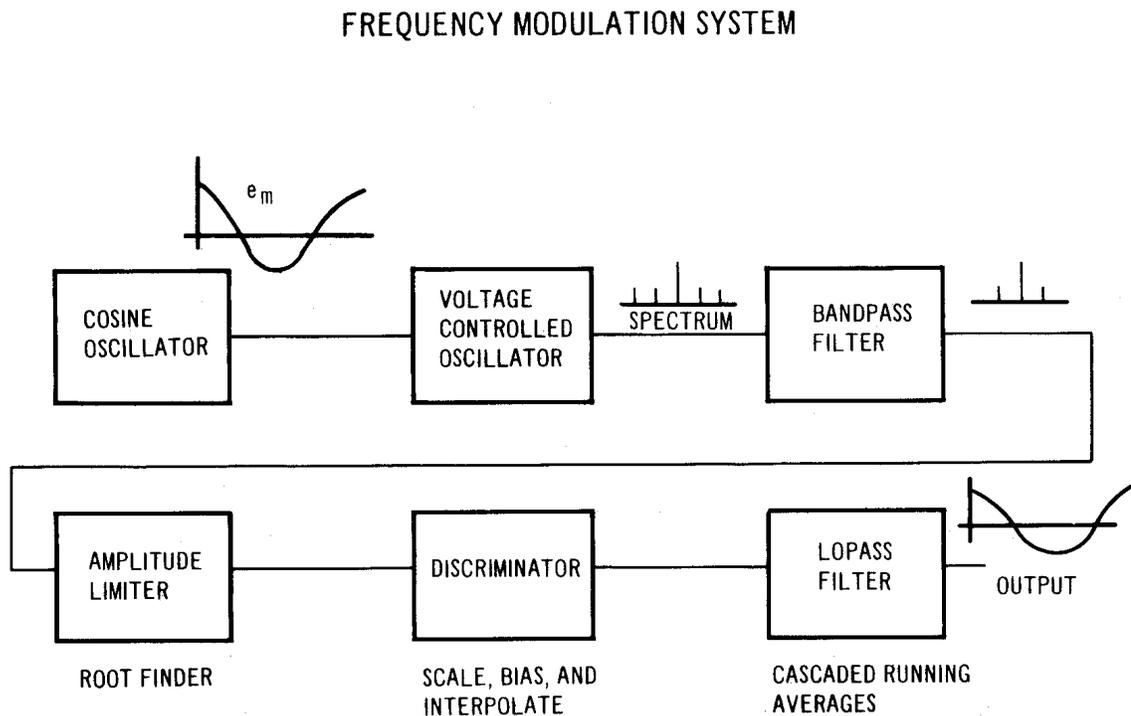


FIGURE 1

The modulated carrier can be expanded into an infinite sum of sine waves as follows:

$$\begin{aligned}
 e = E_c \{ & J_0(B) \sin \omega_c t + J_1(B) [\sin(\omega_c + \omega_m)t - \sin(\omega_c - \omega_m)t] \\
 & + J_2(B) [\sin(\omega_c + 2\omega_m)t + \sin(\omega_c - 2\omega_m)t] \\
 & + J_3(B) [\sin(\omega_c + 3\omega_m)t - \sin(\omega_c - 3\omega_m)t] \\
 & \quad | \quad | \quad | \\
 & \quad | \quad | \quad | \\
 & \quad | \quad | \quad | \\
 & + J_n(B) [\sin(\omega_c + n\omega_m)t + (-1)^n \sin(\omega_c - n\omega_m)t] \}
 \end{aligned}$$

where:

- $B = E_m \Delta F_c / F_m =$ the modulation index (Beta)
- $\omega_c = 2\pi F_c$
- $\omega_m = 2\pi F_m$
- $J_n(B) =$ Bessel function of argument Beta and order $n = 0, 1, 2, 3, \text{ etc.}$
- $n =$ sideband number. It goes from zero to infinity.

The spectrum of the modulated carrier then, is simply a sum of sine waves, and its waveform can be simulated at any time, t , by adding up the products of individual sine wave values times the proper Bessel coefficients. The Bessel values can be computed by the methods of Holt,² or they can be obtained from published tables³ provided the initial conditions of the problem are selected to fit the tables.

Bandpass Filter An ideal bandpass filter is simulated by using only those sine waves which lie within the passband. At this point, the as yet unknown amount of data distortion is introduced by cutting off an arbitrary and known amount of sideband power.

For the Gaussian bandpass case, in addition to truncating the spectrum at some arbitrary sideband number, n , those sidebands which get through are modified in amplitude by the gain of a bandpass Gaussian filter whose transfer function is:

$$\text{GAIN} = -\text{EXP} \left[0.35 \left(\frac{F - F_c}{\Delta F_c} \right)^2 \right] \quad (4)$$

Amplitude Limiter The amplitude limiter clamps the amplitude of the modulated wave and preserves only the axis crossings. It destroys all characteristics of the carrier waveform except polarity and time of occurrence of polarity changes. In order to simulate the amplitude limiter it is necessary to find the axis crossings, and this is done in the following manner:

At some time t , calculate e , and observe its polarity. Then make additional trial calculations with additional search steps and note that whenever a root is crossed the polarity changes. When a root is crossed, cut the search step in half and reverse the direction of the search.

This procedure is portrayed in Figure 2, and it is expedited by high speed computing.

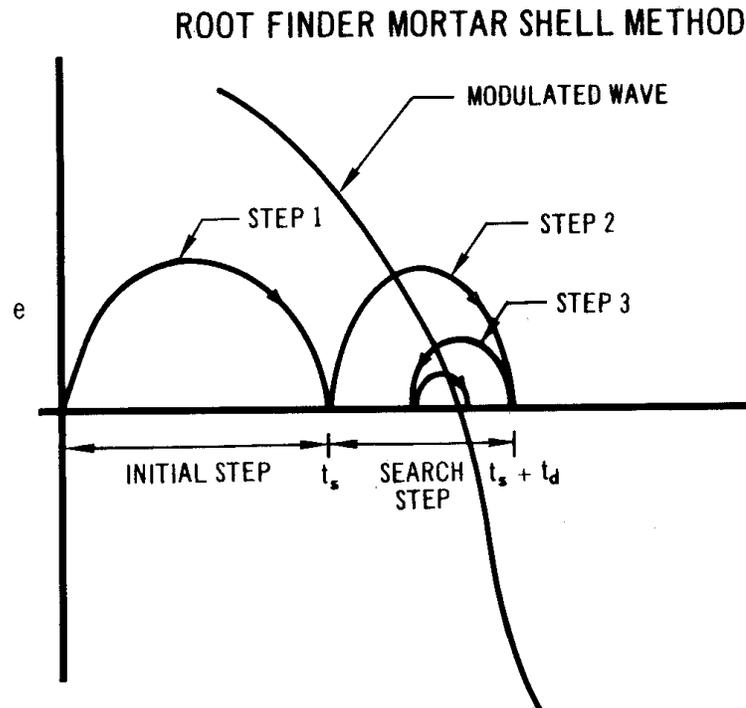


FIGURE 2

Reducing the search step to one half its previous value each time the root is crossed, gives a sort of “binary convergence.” Each additional trial calculation doubles the accuracy to which the location of the root is known. This procedure may be called the “Mortar Shell Method” after artillery practices.

Digital Discrimination Upon finding the roots of the modulated wave, the output of the limiter has been essentially simulated, and it is necessary to show that the modulating cosine wave can be extracted from this information.

Returning briefly to the FM wave before limiting, we remember that the equation for it is:

$$\frac{e}{E_c} = \sin \left[2 \pi F_c t + \left(\frac{E_m \Delta F_c}{F_m} \right) \sin 2 \pi F_m t \right] \quad (5)$$

An ideal discriminator would be one which continuously found the arcsine of e/E_c , and solved for the modulating signal as follows:

$$2\pi F_c t + \frac{E_m \Delta F_c}{F_m} \sin 2\pi F_m t = \arcsin \left[\frac{e(t)}{E_c} \right] \quad (6)$$

Knowing that the instantaneous carrier angle is the angle swept out at the constant carrier rate plus a term proportional to the integral of the modulating function we see that:

$$2\pi \Delta F_c \int e_m dt = \frac{E_m \Delta F_c}{F_m} \sin 2\pi F_m t = \arcsin \left[\frac{e}{E_c} \right] - 2\pi F_c t \quad (7)$$

and differentiating to expose e_m , the modulating signal is found as follows:

$$e_m = E_m \cos 2\pi F_m t = \frac{1}{2\pi \Delta F_c} \frac{d}{dt} \left[\arcsin \left(\frac{e}{E_c} \right) \right] - \left(\frac{F_c}{\Delta F_c} \right) \quad (8)$$

However, since the only things known about an FM wave after limiting are its axis crossings and polarity, it is necessary to find e_m in terms of this. Referring to Figure 3, the modulated carrier, e , passes through zero at odd multiples of Δt measured from time t_i .

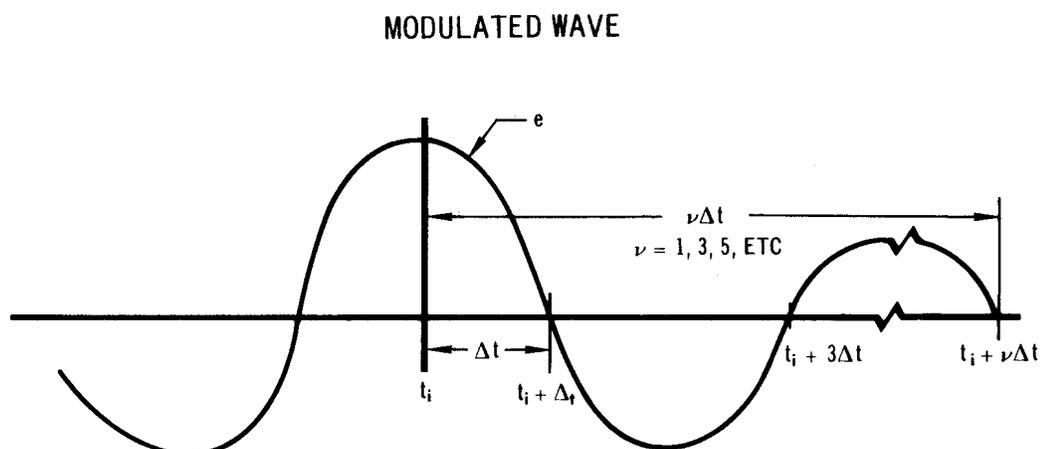


FIGURE 3

Writing the argument of the modulated wave at axis crossings to the right and left of t_i ,

$$2\pi F_c(t_i + \nu \Delta t) + \frac{E_m \Delta F_c}{F_m} \sin 2\pi F_m(t_i + \nu \Delta t) = \arcsin \left[\frac{e(t_i + \nu \Delta t)}{E_c} \right] \quad (9)$$

$$2\pi F_c(t_i - \nu \Delta t) + \frac{E_m \Delta F_c}{F_m} \sin 2\pi F_m(t_i - \nu \Delta t) = \arcsin \left[\frac{e(t_i - \nu \Delta t)}{E_c} \right] \quad (10)$$

The difference between root crossing is $\nu\pi$ radians. Subtracting the 2nd of the previous two equations, from the first:

$$4\pi F_c \nu \Delta t + \frac{E_m \Delta F_c}{F_m} [\sin(2\pi F_m t_j + 2\pi F_m \nu \Delta t) - \sin(2\pi F_m t_j - 2\pi F_m \nu \Delta t)] = \nu\pi \quad (11)$$

and using the trigonometric sum and difference formula:

$$\sin(A + B) - \sin(A - B) = 2 \cos A \cdot \sin B \quad (12)$$

write:

$$E_m \left(\frac{\sin 2\pi F_m \nu \Delta t}{2\pi F_m \nu \Delta t} \right) \cos 2\pi F_m t_j = \frac{1}{4\Delta t \Delta F_c} - \frac{F_c}{\Delta F_c} \quad (13)$$

Let the measured time between axis crossings taken be T_ν . Then: $T_\nu = 2\nu\Delta t$ and $\Delta t = T_\nu/2\nu$. Substituting into the previous equation gives:

$$E_m \left(\frac{\sin \pi F_m T_\nu}{\pi F_m T_\nu} \right) \cos 2\pi F_m t_j = \frac{1}{T_\nu} \left(\frac{\nu}{2\Delta F_c} \right) - \frac{F_c}{\Delta F_c} \quad (14)$$

This is recognized as a possible demodulation scaling equation from which the modulating signal can be calculated in terms of the times between axis crossings, T_ν , and system constants ν , ΔF_c , and F_c where:

T_ν = measured time between crossings

ν = number of crossing intervals taken at one time

$E_m \cos(2\pi F_m t_j)$ = the demodulated cosine wave value e_m at time t_j

and:

$(\sin \pi F_m T_\nu) / \pi F_m T_\nu$ = an aperture term

The aperture term is a sort of frequency varying quasi-linear operator which reduces the amplitude of the output. This is a built-in characteristic of limited FM systems regardless of the amount of bandwidth allowed or the type of discriminator used.⁴ It varies slightly with signal amplitude, which gives rise also to a small apparent dc shift. The error becomes negligible for large carrier to modulating frequency ratios and small ν . It amounts to about 0.1 per cent for a frequency ratio of 13 and a ν of 1.

Linear Interpolation Data samples recovered from the digital discriminator occur at a rate set by the instantaneous frequency of the carrier, which in turn is a function of the modulating signal amplitude. With a ν of 1 and $E_m = 1$, for example, across one cycle of the modulating signal, the sample rate varies from twice the carrier frequency, up to

twice upper bandedge frequency, back to twice center, down to twice lower bandedge, and back to twice center. As distortion occurs, the instantaneous variations may be somewhat different but the general pattern remains.

The non-constant sample rate is made constant by interpolating data values between points as follows:

$$AD = \left[\frac{AM - AMM}{0.5(TR - TBB)} \right] (TS - TB) + \left[\frac{AM + AMM}{2} \right] \quad (15)$$

where the Fortran symbols used for this program are:

- AD = the interpolated value at time TS
- AM = last calculated modulation value
- AMM = previous calculated modulation value
- TR = last root of modulated carrier found
- TB = next previous root found before TR
- TBB = next previous root found before TB
- TS = ideal sample time for equi-spaced samples (normally twice the unmodulated carrier frequency)

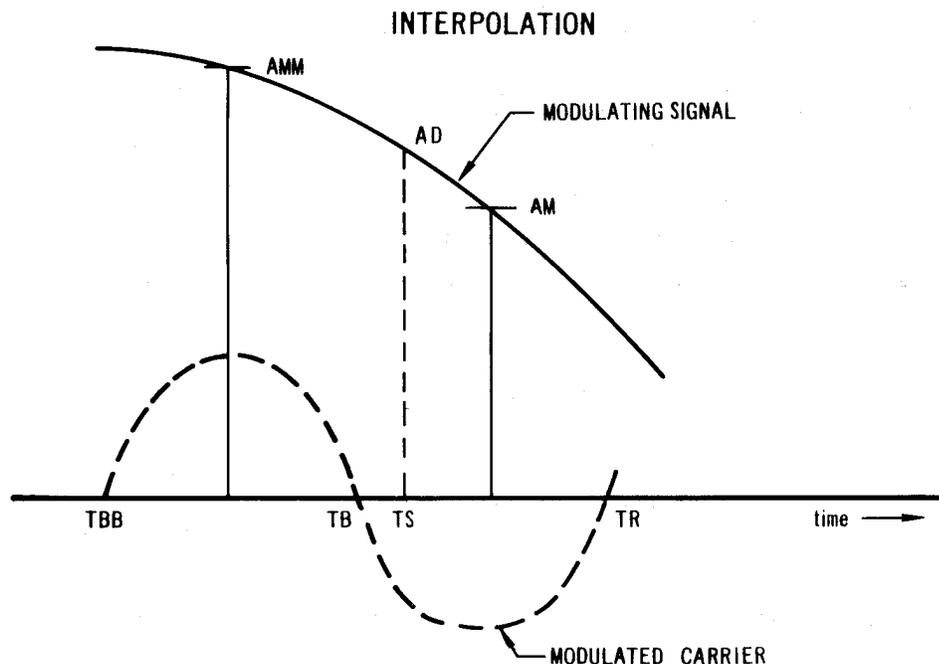


FIGURE 4

Lopass Filter A cascaded running average digital filter is used to get rid of harmonics and any other above-band distortion in the discriminator output. It is simple and effective for this application because it has zero gain at some fractional values of the sampling rates. By placing these zeros at or near the expected harmonic frequencies, a distorted sine wave can be quite thoroughly cleaned up. Such is the procedure followed here.

The transfer function for a running average filter can be derived by referring to Figure 5, and writing the smoothed value of a sampled sine wave as a sum.

$$E_s = \frac{\sum_{k=0}^s E \sin [2\pi F (t_n - k\Delta t) + \phi]}{m} \quad (16)$$

where:

- E_s = smoothed value
- m = number of samples averaged in a row
- t_n = sample time now
- k = index of summing over m values
- Δt = sampling interval
- F_s = sampling rate = $1/\Delta t$
- ϕ = initial phase angle of sine wave
- s = $(m-1)$ = number of samples averaged minus one
- F = frequency
- E = peak value

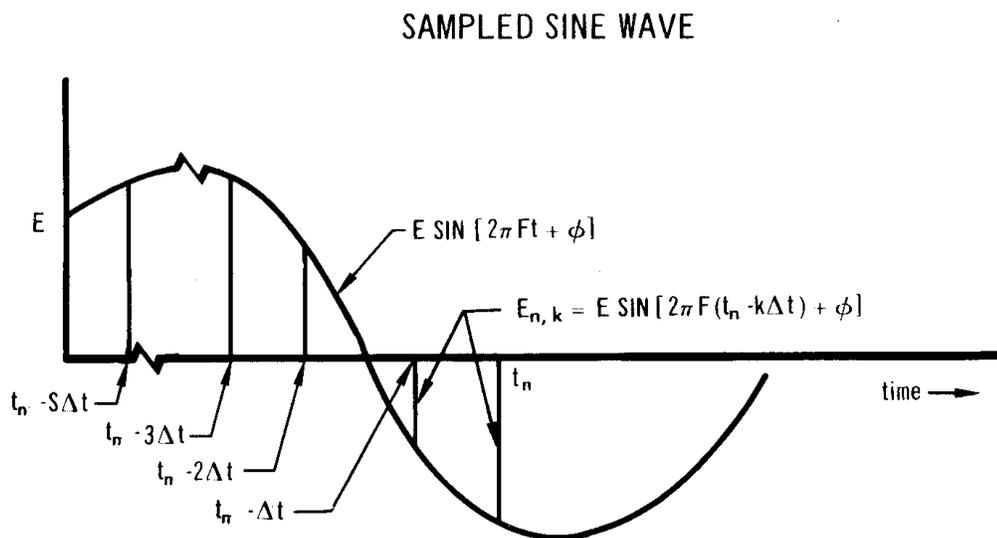


FIGURE 5

Rearranging this summation, writing in closed form, and separating the non-time varying phase and gain terms from the time varying part, write:^{5,6}

$$E_s = E \left[\frac{\sin m \pi \frac{F}{F_s}}{m \sin \pi \frac{F}{F_s}} \right] \sin \left[2 \pi F t_n + \phi - \pi (m - 1) \frac{F}{F_s} \right] \quad (17)$$

$$\text{FILTER GAIN} = \frac{\sin m \pi \frac{F}{F_s}}{m \sin \pi \frac{F}{F_s}} \quad (18)$$

$$\text{FILTER PHASE} = -\pi (m - 1) \frac{F}{F_s}$$

The phase in this type of filter is linear, and the gain goes to zero at:

$$F_{\text{ZERO}} = Z \left(\frac{F_s}{m} \right) \text{ cps} \quad (19)$$

$$Z = 1, 2, 3, 4, \text{ ETC.}$$

By cascading several running averages, and placing the zeros of transmission at the harmonic frequencies of the modulating sine wave, the distorted discriminator output is reduced to the sine wave that it was at the input to the voltage controlled oscillator. Only the amplitude is different, and this is the error.

The amplitude of the sine wave out is subtracted from the amplitude of the sine wave in, to give the error term caused by band-limiting the sideband spectrum.

Reporting the Error The input sine wave has a peak value of EM. It is allowed to vary, for different runs, between 0 and 1.0 The instantaneous input, e_m , varies between $+E_m$ and $-E_m$ since it is a sine wave. The output of the simulator is also a sine wave, and by proper choice of modulating frequency and lopass filter cutoff, the harmonics introduced into the data by the sideband distortion are removed. This makes reporting the error simple, although there still remains several ways to go about it.

It is not the purpose here to discuss the relative merits of various definitions of error, but rather to select one method, describe it carefully, and follow it consistently. It is always possible to translate the error term later into any of many possible forms. The relationships are usually simple factors such as: 2, square root of 2, . 636, 20 log ratio, 10 log ratio, error squared, etc.

Referring to Figure 6, column 1, the input is a sine wave whose peak amplitude is allowed to vary in the range of 0 to 1.0. The system full scale, then, is 2.0. That is the dc full scale too. It corresponds to the familiar 5 volt scale range of telemetry oscillators into which a 2.5 volt sine wave could be fed, a 5.0 peak-to-peak sine wave, or a 1.75 volt RMS sine wave.

ILLUSTRATION OF ERRORS

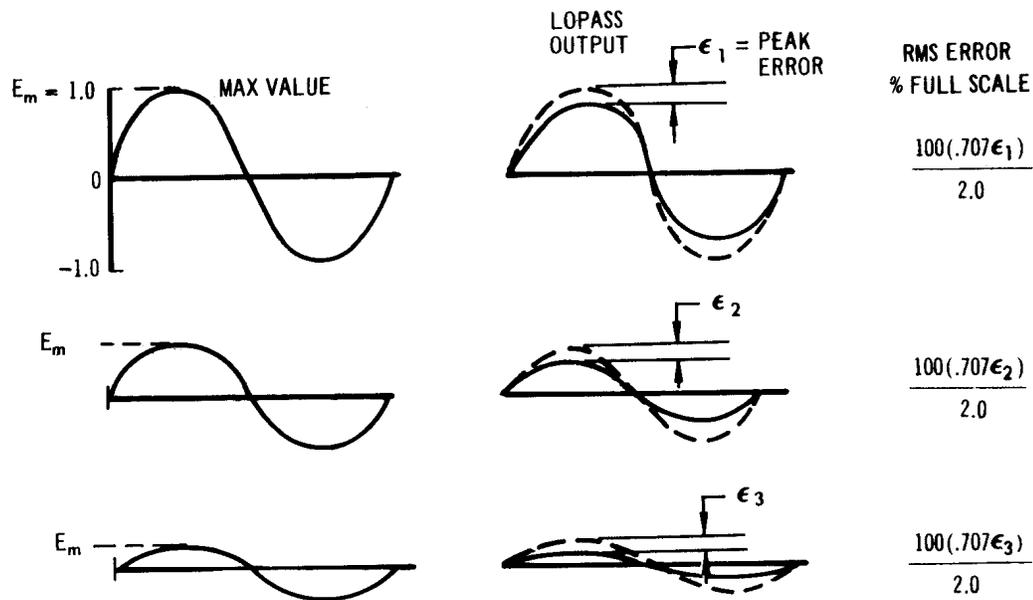


FIGURE 6

The outputs corresponding to these inputs are shown in column 2. The difference between the input and the output is the error. It too is a sine wave with peak value ϵ .

Since the error is a sine wave, it is convenient to consider its RMS value. This is obtained by multiplying ϵ by 0.7071. Then to convert RMS error to per cent of full scale, divide by full scale and multiply his takes the form:

$$\text{RMS ERROR} = \frac{100 [.7071(\epsilon)]}{\text{FULL SCALE}} \% \quad (20)$$

$$\text{Full Scale} = 2 \text{ time the magnitude of } \hat{E}_m = 2.0$$

Variable E_m , Constant F_m , Beta = 1.0 In Figure 7 the RMS error is plotted versus E_m for two bandwidths, $2 \Delta F_c$, and $2 \Delta F_c + 2 F_m$. Als o, $F_m = \Delta F_c$, which might justify calling Figure 7 a “Beta = 1.0 system. The error varies from 0 to 3 per cent for the narrow bandpass and 0 to 0.2 per cent for the wide bandpass. Since the error is caused by loss in sideband power, it is interesting to compare the two for possible cause and effect

relationship. Sideband power lost versus EM is shown in the insert in Figure 7. It is defined as 1.0 minus the sum of the squares of all sidebands (including the $J_0(B)$ term) passed through the bandpass filter. The power lost amounts to about 3 per cent for the narrow bandpass and 0.04 per cent for the wide bandpass.

The wide bandpass $2 \Delta F_c + 2 F_m$, causes considerably less error than the narrow band filter, but the price paid requires doubling the spectrum occupancy for the Beta = 1 system.

FREQUENCY MODULATION
 ERROR CAUSED BY CUTTING OFF SIDEBANDS

BETA = 1 SYSTEM
 IDEAL FILTER

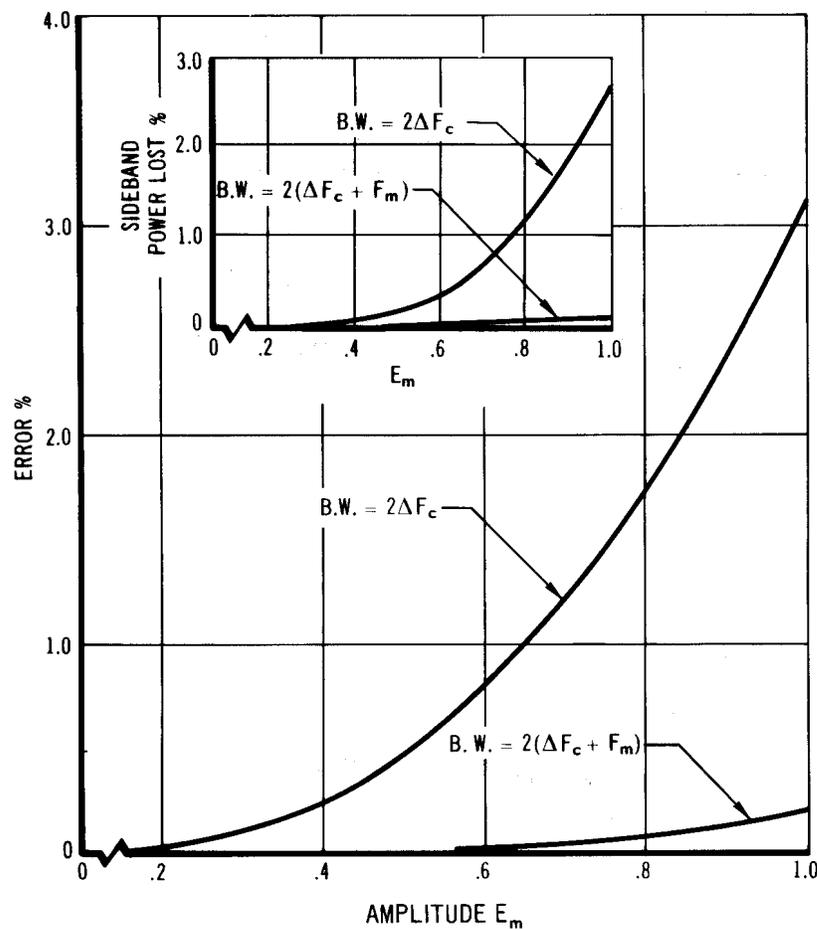


FIGURE 7

FREQUENCY MODULATION
ERROR CAUSED BY CUTTING OFF SIDEBANDS

BETA = 5 SYSTEM
IDEAL FILTER

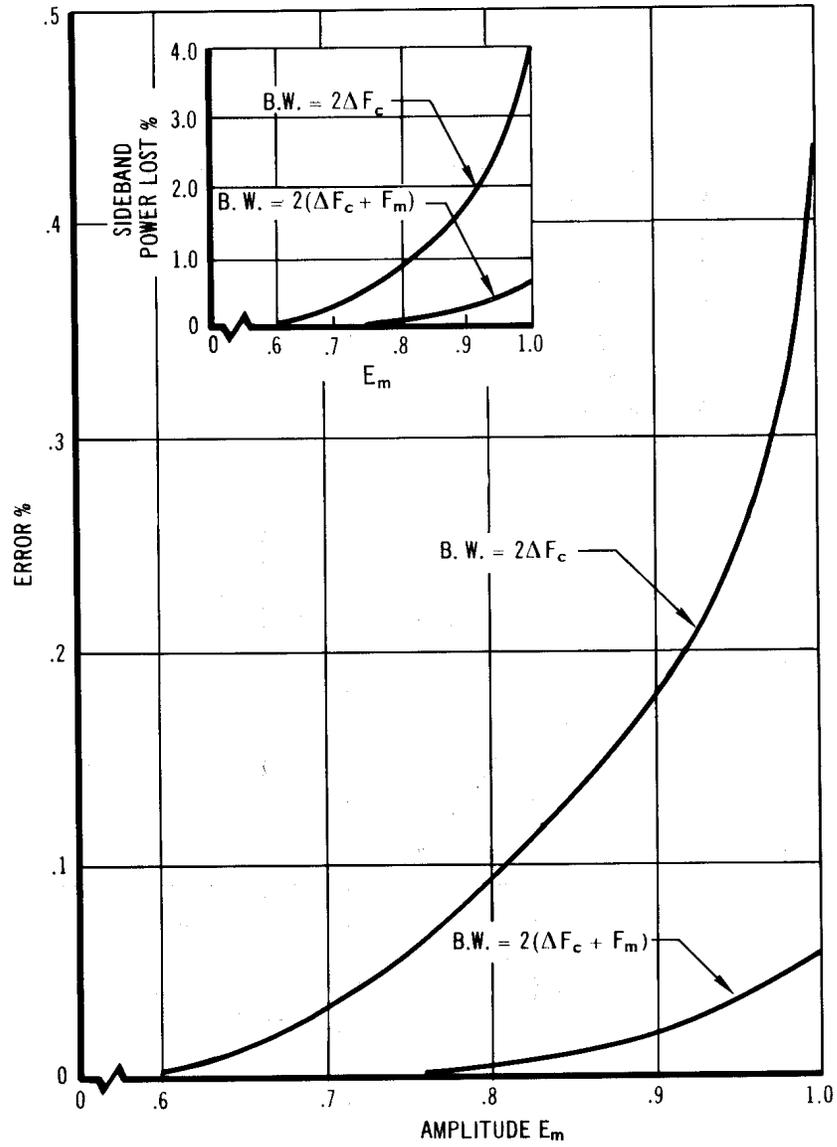


FIGURE 8

**FREQUENCY MODULATION
ERROR CAUSED BY CUTTING OFF SIDEBANDS**

BETA = 5 SYSTEM
B.W. = $2\Delta F_c$
 $E_m = 1.0$

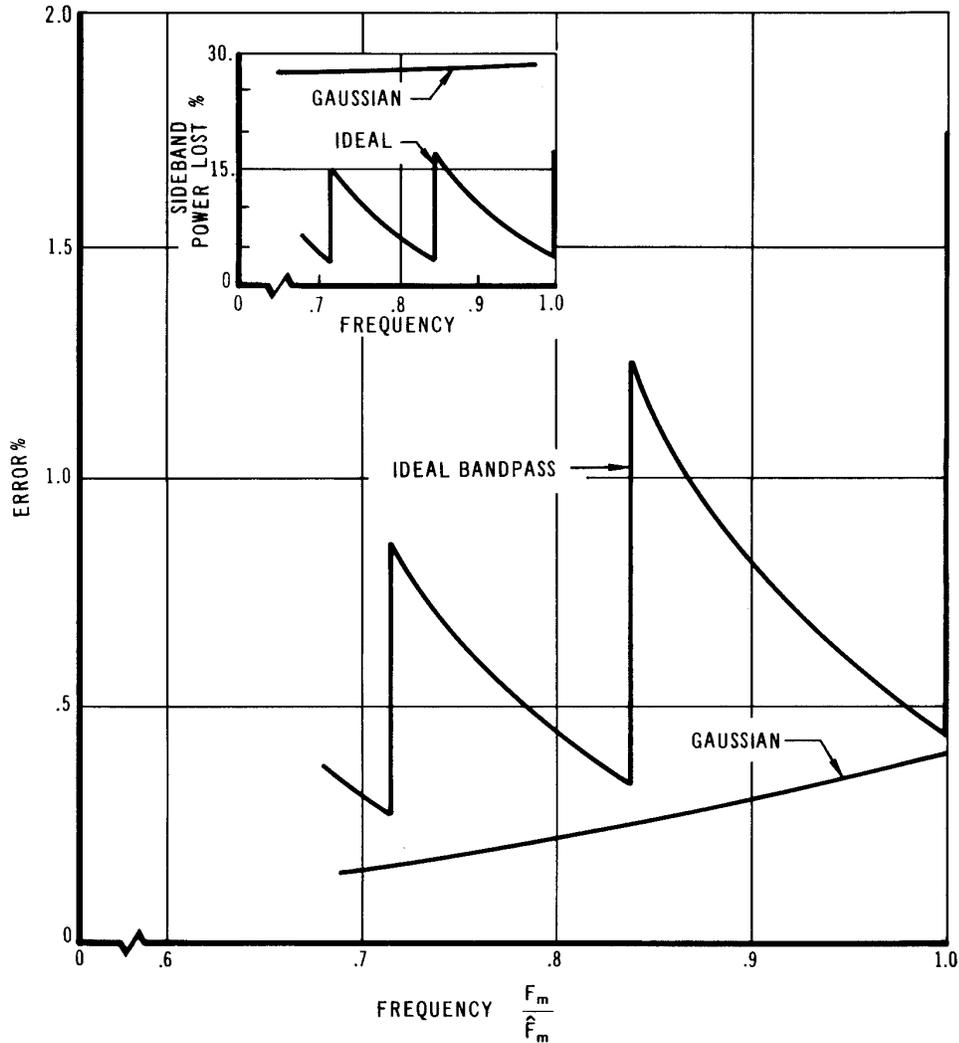


FIGURE 9

Variable E_m , Constant F_m , Beta = 5.0 The Beta = 5.0 case is plotted in Figure 8, with $\Delta F_c = 5 F_m$. RMS error varies from 0 to 0.42 per cent as E_m goes from 0 to 1.0. The maximum sideband power lost is 4 per cent.

Variable F_m , Constant E_m , Beta = 5.0 The signal amplitude, E_m is held constant at full scale while modulating frequency is varied from maximum channel response (Beta = 5.0) down to dc (Beta approaches infinity.) Error is plotted versus frequency in Figure 9

for (1) an ideal filter of bandwidth $2 \Delta F_c$, and (2) a Gaussian bandpass filter with upper and lower cutoffs (Gain 0.7) placed at $F_c + \Delta F_c$, and $F_c - \Delta F_c$ respectively.

Note the step changes in error as the additional sidebands come within range of the ideal filter. The steps do not exist with the Gaussian filter.

The effect, then, of introducing non-perfectly square filters into the bandpass can be generalized. Referring to Figure 9, note that the error changes abruptly at the point where an extra sideband gets lost or picked up. Any filter that is less than ideal will have the effect of rounding off the corners of the error discontinuities as the Gaussian

filter clearly demonstrates. Figure 10 shows qualitatively the relationship's between gain slopes and error slopes.

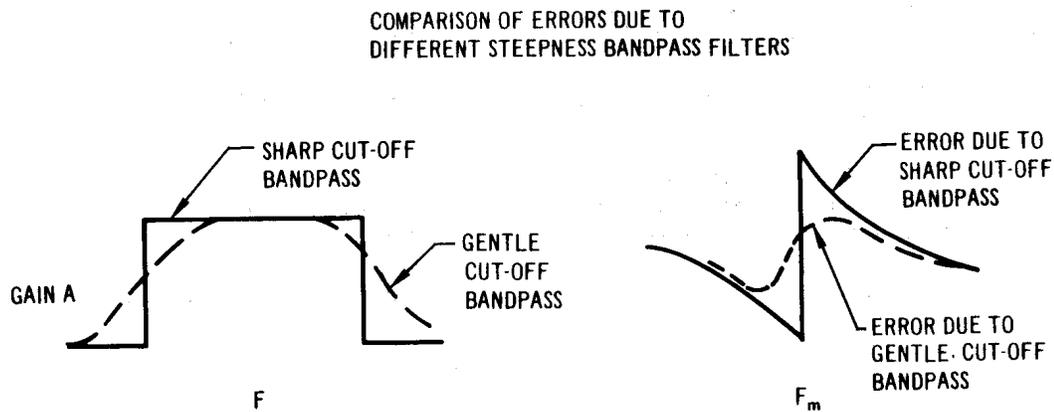


FIGURE 10

The true bandwidth of the Gaussian filter is somewhat greater than $2\Delta F_c$ depending on one's definition of "significant sideband." The Gaussian filter might be considered a spectrum window which forces convergence of the Bessel coefficients more gradually than the abrupt ideal filter.

Conclusions Error has been calculated under conditions which it is believed are severe. There may be worse cases, but it is felt that the systems used for transmission of actual data operate under conditions that are less severe. It is unlikely that the modulating signal would be a continuous single sine wave of near maximum system frequency and occupying the full scale range. Usually the modulating signal will be more complex, and it can be described by a sum of sine waves.

Calculations of FM spectra due to multi-tone modulation have shown that the bandwidth does not increase and usually becomes narrower as the number of modulating

components (whether commensurable or incommensurable) increases, the algebraic sum of all sine waves being held constant.

The maximum error found in this study was about 3 per cent at a bandwidth of $2\Delta F_c$ and Beta of 1. Such narrow bandwidths could be used for those cases where that much error can be tolerated.

REFERENCES

1. Seely, Samuel, Electron Tube Circuits, McGraw-Hill, 1950, page 365.
2. Holt, J. C., Bessel Functions Calculation and Table, Douglas Aircraft Company Report SM-43603, dated April 9, 1963.
3. Reference Data for Radio Engineers, 4th Edition, IT&T, pages 1118 thru 1120.
4. Duncan, P. H., Limitations on Maximum Information Frequency in FM Systems, Douglas Aircraft Company Internal Memorandum A2-260-DE 3-28, dated January 23, 1963.
5. Hamming, R. W., Numerical Methods for Scientists and Engineers, McGraw-Hill, 1962, page 314.
6. Jolley, Summation of Series, Dover, 1961, page 78.
7. Holt, J. C., Frequency Spectrum of A Complex FM Wave, Douglas Aircraft Company Internal Memorandum A2-260-A343-293, dated August 20, 1963.