THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 59, No. 4, April 1980
Printed in U.S.A.

Spectral Shaping by Simultaneous Amplitude and Frequency Modulation

By J. SALZ and J. J. WERNER

(Manuscript received November 16, 1979)

In an attempt to simplify the design of separation filters in frequency-division multiplexed digital data transmission by FSK, it was found that the energy on one side of the FSK carrier can be reduced by suitably modulating the amplitude of the FSK signal. We present a simple technique for controlling the energy on one side of the carrier of an FSK signal. The method entails modulating the amplitude of the FM wave in unison with the frequency. In other words, if the energy in the region above the carrier must be reduced, the envelope is somewhat suppressed while the higher frequency is transmitted. The proposed hybrid modulation technique is applicable in frequency-division multiplexed digital data transmission by FSK and can often ease the design of separation filters. Here we investigate this modulation technique theoretically as well as experimentally. Our theoretical results include closed-form formulas for the various power spectra of interest. We also determine the optimum amount of amplitude modulation to achieve the greatest amount of sidelobe energy suppression for a fixed amount of total average power. The experimental investigations substantiate the analysis, and we make comparisons of the technique with alternative approaches.

I. INTRODUCTION

It has been known that complete elimination of one of the sidebands in FM can be achieved by modulating the envelope by a signal related to the baseband modulation. This scheme is known as SSB-FM.[1,2] In general, beneficial spectral shaping in FM is achieved only if the amplitude modulating signal is related in some manner to the frequency modulation. Also the amplitude modulation must be positive so that the zero-crossings of the original FM wave will not be altered.

For random data, the spectrum of FSK is known to be symmetrical about the carrier frequency, and it appears intuitively clear that reducing the carrier amplitude during the period while the higher frequency is being transmitted reduces the energy of the spectrum in this frequency range.

We have observed experimentally that the amount of amplitude modulation critically affects the energy distribution. These preliminary experimental observations prompted us to examine this subject analytically and to verify the analysis by further experimentation.

Section II outlines the derivations of the AM-FM spectral density when the baseband modulation is a synchronous digital data signal. Section III repeats the derivation for asynchronous data. Section IV discusses numerical and experimental results

II. THE SPECTRUM OF AM-SYNCHRONOUS FSK

To illustrate the principal ideas and to bring out the possible benefits of AM-FSK, we use two simple models of digital FM and restrict ourselves to simple amplitude modulation mechanisms. We first outline the derivation of the spectrum for a general parameter set and then specialize to binary FM.

Consider a stream of data symbols, $a_0, a_1, \cdots$, where each a_n can assume real values independently and with equal probability. Now form the pulse train

$$D(t) = \sum_{n=0}^{\infty} a_n g(t - nT), \tag{1}$$

where $g(t)$ is a unit-height, T-second duration rectangular pulse. This signal will be used to modulate the frequency of a carrier wave. For the amplitude modulation, form another pulse train

$$\tilde{D}(t) = \sum_{n=0}^{\infty} b(a_n) g(t - nT), \tag{2}$$

where $b(x)$ is a nonnegative function. Next, apply (1) to an FM modulator and multiply the output by (2). These operations result in an AM-FM wave which is represented by

$$S(t) = \tilde{D}(t) \exp\left\{ i \left[\omega_c t + \omega_d \int_0^t D(t')\, dt' + \theta \right] \right\}, \tag{3}$$

where ω_c is the carrier angular frequency, ω_d is a constant of proportionality, and θ is a uniformly distributed phase angle on the interval $[-\pi, \pi]$. We use the complex representation of FM for calculational economy without loss of generality.

Because $\tilde{D}(t)$ is always positive, $D(t)$ can be recovered by conventional FM demodulation methods,[3] and the role of the amplitude modulation here is only to shape the spectrum of the FM wave. The manner in which spectral shaping is accomplished is our main interest and motivates our investigation of the spectral properties of (3).

Denote the spectral density of $S(t)$ by $G(\omega)$. A convenient definition of the spectrum is the limit

$$\lim_{N\to\infty} \frac{1}{N} E\{|Z(\omega, N)|^2\}, \tag{4}$$

where

$$Z(\omega, N) = \int_0^{NT} S(t)e^{-i\omega t}\, dt$$

and where $E\{\cdot\}$ denotes the ensemble average.

The sequence of squared magnitudes of $Z(\omega, N)$ is calculated in a straightforward manner:

$$|Z(\omega, N)|^2 = \left| \sum_{k=0}^{N-1} b(a_k) \int_{kT}^{(k+1)T} \exp\left\{ i\left[(\omega_c - \omega)t + \omega_d \sum_{n=0}^{\infty} a_n \cdot \int_{-nT}^{t-nT} g(t')\, dt' \right]\right\} \right|^2. \tag{5}$$

By recognizing that the phase angle,

$$\psi(t) = (\omega - \omega_c)t - \omega_d \sum_{n=0}^{\infty} a_n \int_0^{t-nT} g(t')\, dt',$$

must be a linear function of time in the interval $kT \leq t \leq (k+1)T$, it is possible to put (5) into the form

$$|Z(\omega, N)|^2 = \sum_{k,s=0}^{N-1} b(a_k)b(a_s)e^{it\nu(s-k)}F(\nu - a_k\omega_d)F^*(\nu - a_s\omega_d) \cdot \exp\left\{ i\alpha\left[\sum_{n=0}^{k-1} a_n - \sum_{n=0}^{s-1} a_n \right]\right\}, \tag{6}$$

where $\alpha = \omega_d T$, $\nu = \omega - \omega_c$, $_*$ is the complex conjugate, and

$$F(x) = \int_0^T e^{-ix\xi}\, d\xi.$$

[See Ref. 4 for detailed manipulations leading to (6).]

To get the spectral density, (6) is first averaged with respect to the identically distributed random variables, $a_0, a_1, a_2, \cdots$, and then the limit in (4) evaluated. Manipulations similar to those carried out in Ref. 4 lead to the desired result, namely,

$$G(\omega) = \lim_{N\to\infty} \frac{1}{N} E\,|\,Z(\omega, N)\,|^2$$

$$= E_a\left[b(a)S_i\left(\frac{\nu T - a\alpha}{2}\right)\right]^2$$

$$+ 2\,\mathrm{Re}\left\{\frac{e^{-i\nu T}}{1 - e^{-i\nu T}C(\alpha)}\left(E_a\left[e^{\frac{ia\alpha}{2}}S_i\left(\frac{\nu T - \alpha a}{2}\right)b(a)\right]\right)^2\right\}, \quad (7)$$

where a is a data symbol, $C(\alpha) = E_a(e^{ia\alpha})$, and $S_i(x) = \sin x/x$. The above result for $G(\omega)$ holds only when $|\,C(\alpha)\,| < 1$. When $|\,C(\alpha)\,| = 1$, the spectrum contains delta functions. These can be determined without too much difficulty,[4] but, since in this application $|\,C\,| < 1$, we do not need the most general results.

In our application, the data are binary so that $a = \pm 1$. We also specialize to $\alpha = \omega_d T = \pi/2$. For this special case, the spectral density formula reduces to a very simple form.

Let

$$b(1) = A$$

and

$$b(-1) = B.$$

Now evaluate explicitly the averages specified in (7).

$$2G(\omega) = A^2S_i^2\left(\frac{\alpha - T\nu}{2}\right) + B^2S_i^2\left(\frac{\alpha + T\nu}{2}\right)$$

$$+ \mathrm{Re}\left\{e^{-iT\nu}\left[\sqrt{-1}\,A^2S_i^2\left(\frac{\alpha - T\nu}{2}\right) - \sqrt{-1}\,B^2S_i^2\left(\frac{\alpha + T\nu}{2}\right)\right.\right.$$

$$\left.\left. + 2ABS_i\left(\frac{\alpha - T\nu}{2}\right)S_i\left(\frac{\alpha + T\nu}{2}\right)\right]\right\}. \quad (8)$$

Making use of the identities

$$2\left(\frac{\alpha - T\nu}{2}\right)^2 S_i^2\left(\frac{\alpha - T\nu}{2}\right) = 1 - \sin \nu T,$$

$$2\left(\frac{\alpha + T\nu}{2}\right)^2 S_i^2\left(\frac{\alpha + T\nu}{2}\right) = 1 + \sin \nu T,$$

$$(\alpha = \pi/2),$$

eq. (8) becomes

$$G(\nu) = K\left[\frac{A}{\alpha - T\nu} + \frac{B}{\alpha + T\nu}\right]^2 \cos^2 \nu T, \quad (9)$$

where K is an unimportant normalization constant.

Several checks reveal the physical reasonableness of (9). When $A = B$, we have no amplitude modulation and hence $S(t)$ becomes a constant-envelope, continuous-phase FM wave. In this case,

$$G(\nu) \backsim \left[\frac{1}{(\nu T)^2 - \alpha^2}\right]^2,$$

which is a well-known result.[5]

Notice that, in this case, because of phase continuity, the spectrum behaves like $1/\omega^4$ for large ω. When $A \neq B$, the spectrum behaves only as $1/\omega^2$, as ω becomes large. In the extreme situation when $A = 0$, we have 100-percent amplitude modulation and, in fact, no frequency modulation. The only tone transmitted is at $\omega_c - \omega_d$, and $S(t)$ is an on-off sine wave with frequency $\omega_c - \omega_d$. The spectrum of this wave is the well-known $(\sin x/x)^2$. Put $A = 0$ in (9) and, with $\alpha = \omega_d T = \pi/2$, obtain

$$G(\omega) \backsim \left[\frac{\sin[\omega - (\omega_c - \omega_d))T]}{(\omega - (\omega_c - \omega_d))T}\right]^2,$$

which is in agreement with physical intuition.

III. THE SPECTRUM OF AM-ASYNCHRONOUS FSK

The baseband asynchronous FSK wave is constructed as follows. On the interval $[0, T]$, pick a set of points at random and arrange them such that

$$0 = t_0 < t_1 < t_2 < \cdots t_N = T. \tag{10}$$

Define a set of functions

$$g_{\Delta_n}(t - t_n) = \begin{cases} 1, & t_n \leq t \leq t_{n+1} \\ 0, & \text{otherwise,} \end{cases} \tag{11}$$

where

$$\Delta_n = t_{n+1} - t_n. \tag{12}$$

In terms of (10) and (11), construct the baseband signal $X(t)$ as the following time series

$$X(t) = \sum_{n=0}^{N-1} a_n g_{\Delta_n}(t - t_n), \tag{13}$$

where $\vec{a} = (a_0, a_1, \cdots, a_{N-1})$ is an arbitrary set of identically distributed random variables.

The instantaneous phase $\psi(t)$ in this case is

$$\begin{aligned} \psi(t) &= \omega_c t + \omega_d \int_0^t x(t')\, dt' + \theta \\ &= \psi_1(t) + \theta, \quad 0 \leq t \leq T. \end{aligned} \tag{14}$$

The constants ω_c and ω_d have the same definition as in the previous section.

As before, let $b(x)$ be a nonnegative function and construct the AM-asynchronous FM wave as follows:

$$S(t) = \sum_{n=0}^{N-1} b(a_n) g_{\Delta_n}(t - t_n) e^{i\psi_1(t) + i\theta}. \tag{15}$$

Now define the truncated Fourier transform

$$\begin{aligned} Z(\omega, T) &= \int_0^T S(t) e^{-i\omega t}\, dt \\ &= e^{i\theta} \sum_{k=0}^{N-1} b(a_k) \int_{t_k}^{t_{k+1}} \exp[i(\psi_1(t) - \omega t)]\, dt \\ &= e^{i\theta} \sum_{k=0}^{N-1} \frac{b(a_k)}{i\lambda_k} \left[\exp\left\{ i \sum_{n=0}^{k} \lambda_n \Delta_n \right\} - \exp\left\{ i \sum_{n=0}^{k-1} \lambda_n \Delta_n \right\} \right], \end{aligned} \tag{16}$$

where

$$\lambda_n = \omega_d a_n - \omega + \omega_c.$$

To proceed further, the magnitude squared of (16) must be averaged with respect to the set of random vectors

$$\vec{\Delta} = (\Delta_1, \Delta_2, \cdots, \Delta_N) \quad \text{and} \quad \vec{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_N).$$

In (15), the Δ_n's are the intervals between transitions. We shall assume that these intervals are independent and identically distributed. On the other hand, the random variables λ_n, which are linear functions of the amplitudes a_n, are not independent if only actual transitions are considered. Clearly, for observable transitions $a_n \neq a_{n-1}$ and therefore adjacent amplitudes cannot be independent.

This difficulty in the analysis has been overcome in Ref. 6 where asynchronous FM without amplitude modulation is treated. Our problem here is a slight modification of the pure FM case, and identical reasoning and techniques to those exhibited in that reference can be used to obtain a closed-form formula for the power spectrum with amplitude modulation. Since the algebra leading from (16) to the final result is rather involved, we shall not repeat it here and will give only the final formula.

Thus, for the special case where the number of t points in a fixed interval T obeys the Poisson probability law with parameter $\nu = (1/\overline{T})½$ = average number of real transition per second, we get for the power spectrum

$$G(\omega) = \lim_{T\to\infty} \frac{1}{T} E\,|\,z(\omega, T)\,|^2$$

$$= \frac{[A\overline{T}(\overline{\omega} + \omega_d) - B\overline{T}(\overline{\omega} - \omega_d)]^2}{(\overline{T\omega})^2 + (\overline{T}^2(\overline{\omega}^2 - \omega_d^2))^2}, \tag{17}$$

where

$$\overline{T} = 1/2\nu, \qquad \overline{\omega} = \omega - \omega_c$$

and

$$b(1) = A, \qquad b(-1) = B.$$

Our experimental work is based on synchronous data, and formula (17) is included only for completeness. Clearly, the spectrum for the asynchronous data exhibits similar properties to that of the synchronous case.

IV. NUMERICAL AND EXPERIMENTAL RESULTS

In Figs. 1 and 2, we have plotted several curves exhibiting the power spectrum $G(f)$ [eq. (9)] for different values of the parameter

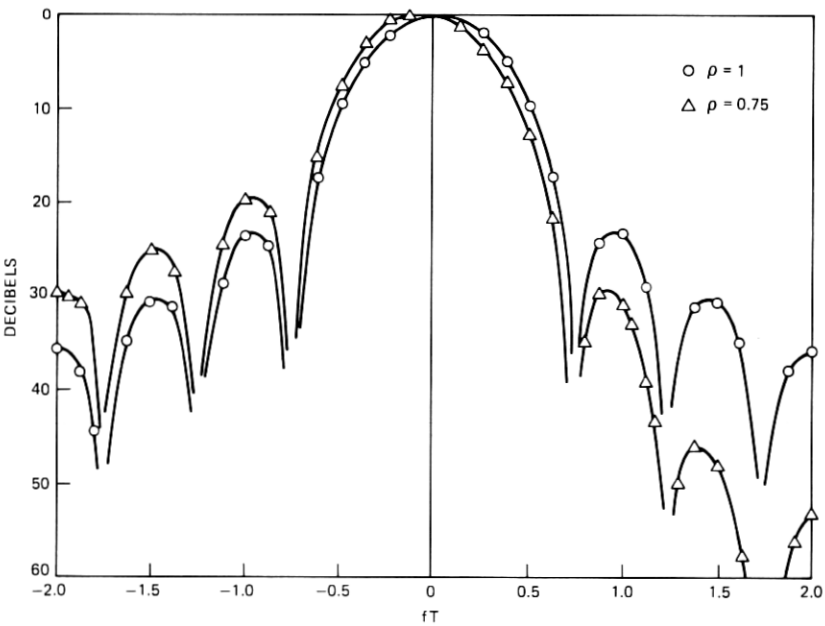


Fig. 1—Power spectral densities for FSK and AM-FSK.

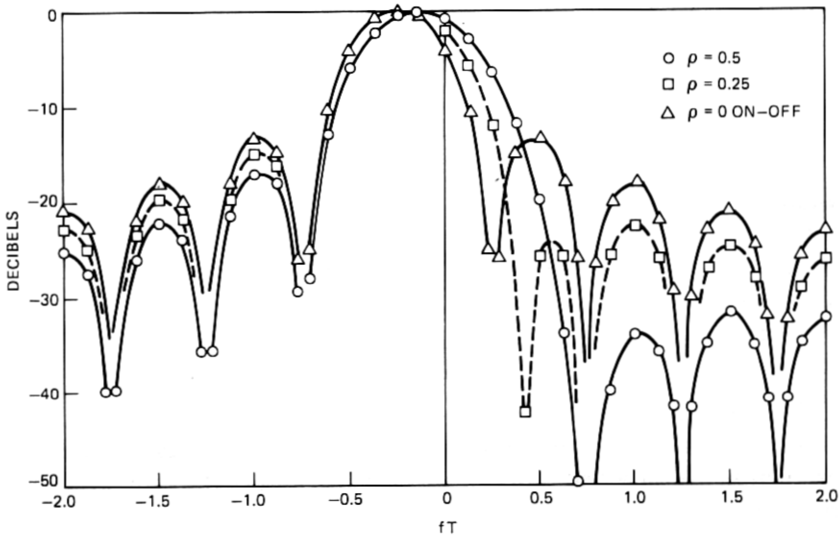


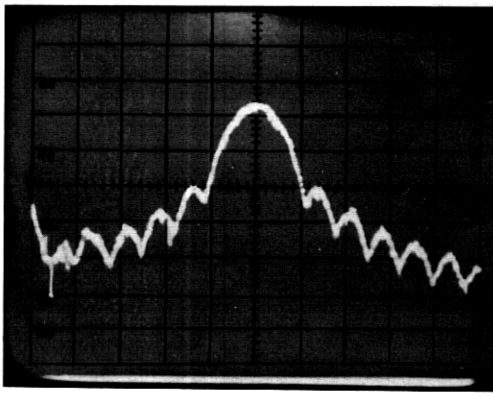
Fig. 2—Power spectral densities for AM-FSK.

$\rho = A/B$.† An FSK transmitter with amplitude modulation (AM-FSK) has also been implemented. The measured power spectra for different values of ρ are shown in Figs. 3 and 4. In the experiments, we have chosen a carrier frequency of 250 Hz and a bit rate of 75 b/s. The agreement between the analytical and experimental results is seen to be excellent. For $\rho = 1$, we have a conventional binary FSK signal. From Figs. 3 and 4, we see that when ρ decreases the amount of power in the right sidelobes decreases at first and then increases again. At the same time, the amount of power in the left sidelobes increases steadily but at a slower rate. Quantitative results are given in Fig. 5, where we have plotted the residual powers $P_u = \int_{f_\ell T}^{\infty} G(f)\, df$ and $P_L = -\int_{-f_\ell T}^{-\infty} G(f)\, df$ in the right and left sidelobes, respectively, as a function of the parameter ρ. For the two sets of curves, corresponding to two different values of $f_\ell T$, the minimum value of the residual power in the right lobes is obtained for a value of ρ approximately equal to 0.75. It is possible to derive a closed-form expression for the value of ρ which gives the minimum residual power as a function of $f_{\ell T}$. The derivation is given in the appendix. For $f_{\ell T} = 1$, the optimum theoretical value of ρ is 0.7623.

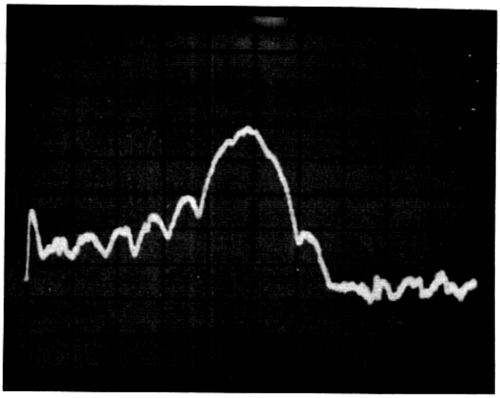
AM-FSK has potential applications in situations where the power in the right or left lobes of a conventional FSK signal might be detrimental

† In the following we always assume $\rho \leq 1$. The spectra corresponding to $\rho > 1$ are simply the mirror images with respect to the carrier frequency of the spectra corresponding to $\rho < 1$.

to the proper functioning of a contiguous neighbor. Filter requirements can also be eased if power in the interfering lobes can be reduced. Clearly, methods other than amplitude modulation exist to reduce the power in the lobes of an FSK signal. One obvious method is simply to filter some of the power. However, in a digital environment, such an approach can be rather costly. Typically, a good nonrecursive filter requires about two orders of magnitude more computation time than is required for the implementation of the amplitude modulation described here. A second possibility to decrease the amount of interference produced by the lobes of an FSK signal on another signal consists in moving away the carrier frequency of the FSK signal. To investigate this possiblity, we have computed the amount $\Delta f_c \cdot T$ by which the FSK spectrum should be shifted with respect to the AM-FSK spectrum in



$\rho = 1$



$\rho = 0.75$

Fig. 3—Measured power spectra of FSK and AM-FSK (10-dB per division along the vertical axis).

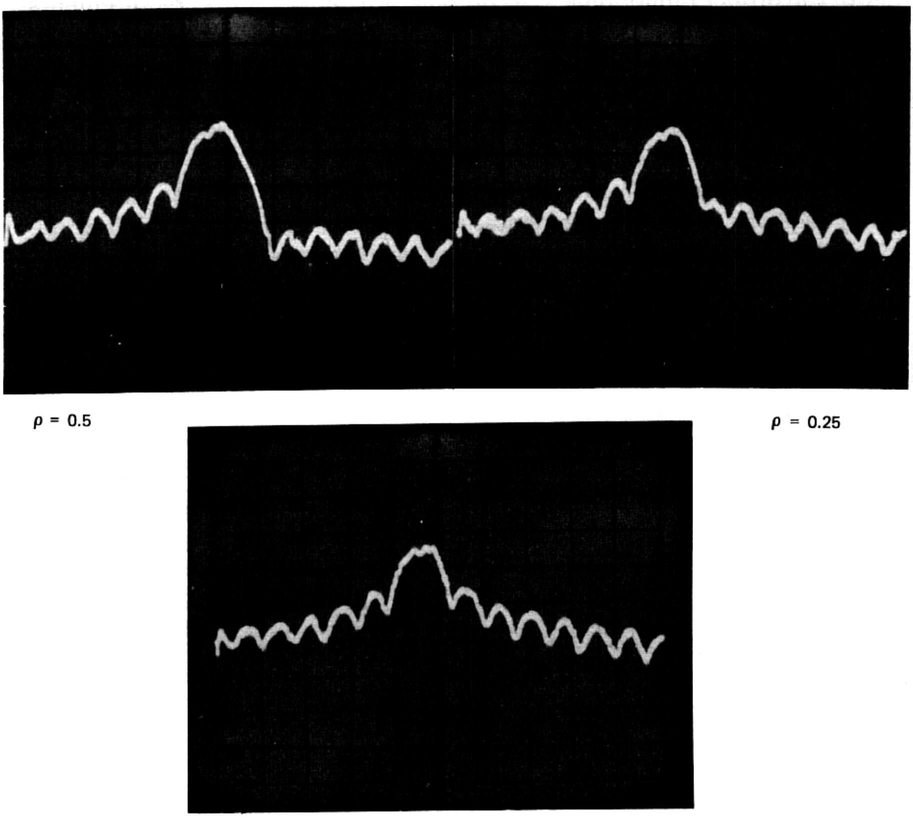


Fig. 4—Measured power spectra of AM-FSK signals.

order to get the same residual power above the point $f_cT = 1$.† The results are shown in Fig. 6. In the neighborhood of the point $\rho = 0.75$, the shifting Δf_c of the carrier frequency has to be close to $1/T$ to get the same performance for the FSK and AM-FSK signals.

Finally, a third way to reduce the power in the sidelobes of the FSK signal consists in simply decreasing the total power of the signal. In Fig. 6, the dashed curve shows by what amount the power of the FSK signal must be reduced with respect to the power of the AM-FSK signal in order to have the same residual power beyond the point $f_cT = 1$. It can be seen that around the optimum point, near $\rho = 0.75$, this is achieved when the power of the FSK signal is about one-tenth the power of the AM-FSK signal. Figure 7 gives a more detailed comparison

† The point $f_cT = 1$ corresponds to the center of the first lobe of the nonshifted signal.

between the residual power for FSK and AM-FSK as a function of $f_\ell T$ in the case where $\rho = 0.75$.

APPENDIX

The Optimum Value of ρ

The residual power in the upper sidelobes of the power spectrum is given by:

$$P_u = \int_{\nu_\ell}^{\infty} \left[\frac{A}{\alpha - T\nu} + \frac{B}{\alpha + T\nu} \right]^2 \cos^2 T\nu \, d\nu, \tag{18}$$

where $\alpha = \pi/2$ and $\nu_\ell > \alpha/T$. The functional P_u is minimized under the power constraint

$$P = A^2 + B^2 = \text{Constant}. \tag{19}$$

This variational problem can be solved in a straightforward manner by introducing a Lagrange multiplier and evaluating

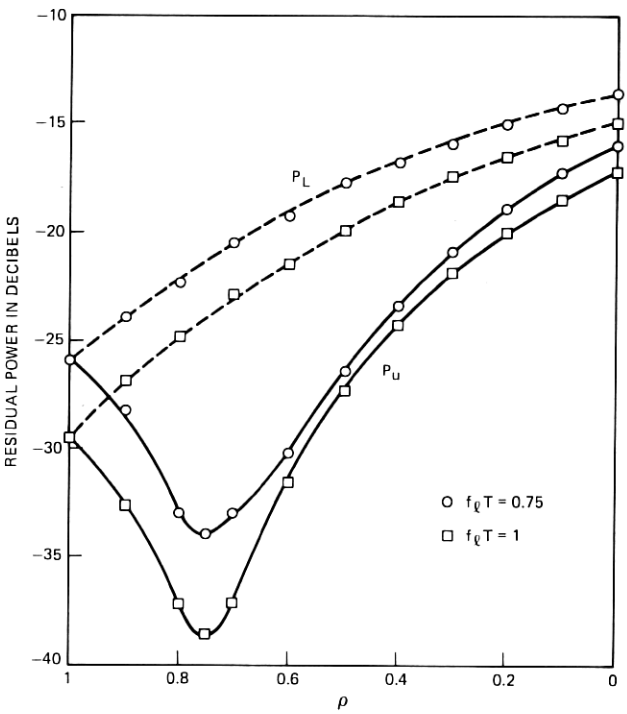


Fig. 5—Residual power in the upper and lower side lobes of AM-FSK.

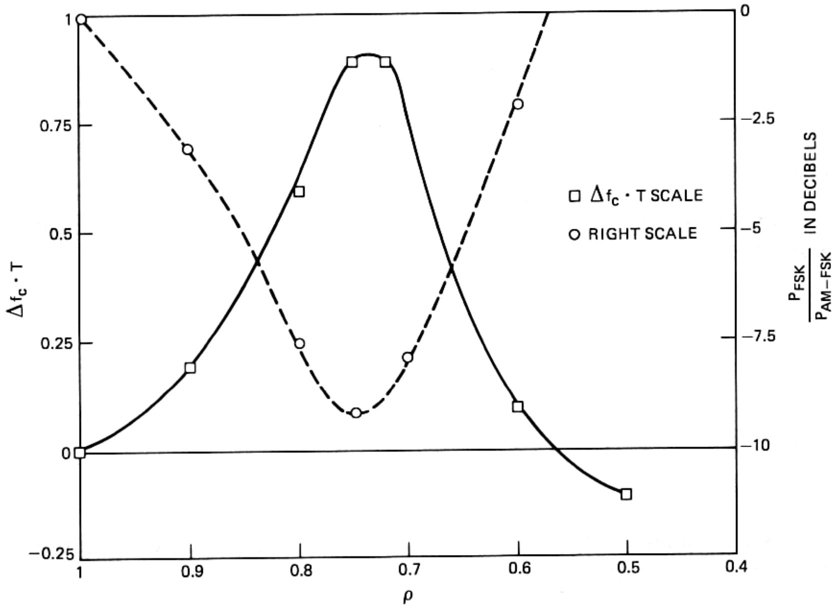


Fig. 6—Curves showing by how much the FSK signal should be shifted or attenuated in order to have the same residual power as AM-FSK.

$$\frac{\partial P_u}{\partial A} + \lambda \frac{\partial P}{\partial A} = 0 \tag{20}$$

$$\frac{\partial P_u}{\partial B} + \lambda \frac{\partial P}{\partial B} = 0. \tag{21}$$

Replacing P_u and P by their expressions and letting $\rho = A/B$, we have

$$\int_{\nu_\ell}^{\infty} \frac{1}{\alpha - T\nu}\left[\frac{\rho}{\alpha - T\nu} + \frac{1}{\alpha + T\nu}\right]\cos^2 \nu T \, d\nu + \lambda\rho = 0 \tag{22}$$

$$\int_{\nu_\ell}^{\infty} \frac{1}{\alpha + T\nu}\left[\frac{\rho}{\alpha - T\nu} + \frac{1}{\alpha + T\nu}\right]\cos^2 \nu T \, d\nu + \lambda = 0. \tag{23}$$

Eliminating λ between (22) and (23) and redefining some variables, we get

$$\rho^2 - \rho \frac{K_2 - K_3}{K_1} - 1 = 0, \tag{24}$$

where

$$K_1 = \int_{\nu_\ell T}^{\infty} \frac{\cos^2 x}{\alpha^2 - x^2}\, dx$$

$$K_2 = \int_{\nu_\ell T}^{\infty} \frac{\cos^2 x}{(\alpha - x)^2}\, dx$$

$$K_3 = \int_{\nu_\ell T}^{\infty} \frac{\cos^2 x}{(\alpha + x)^2}\, dx.$$

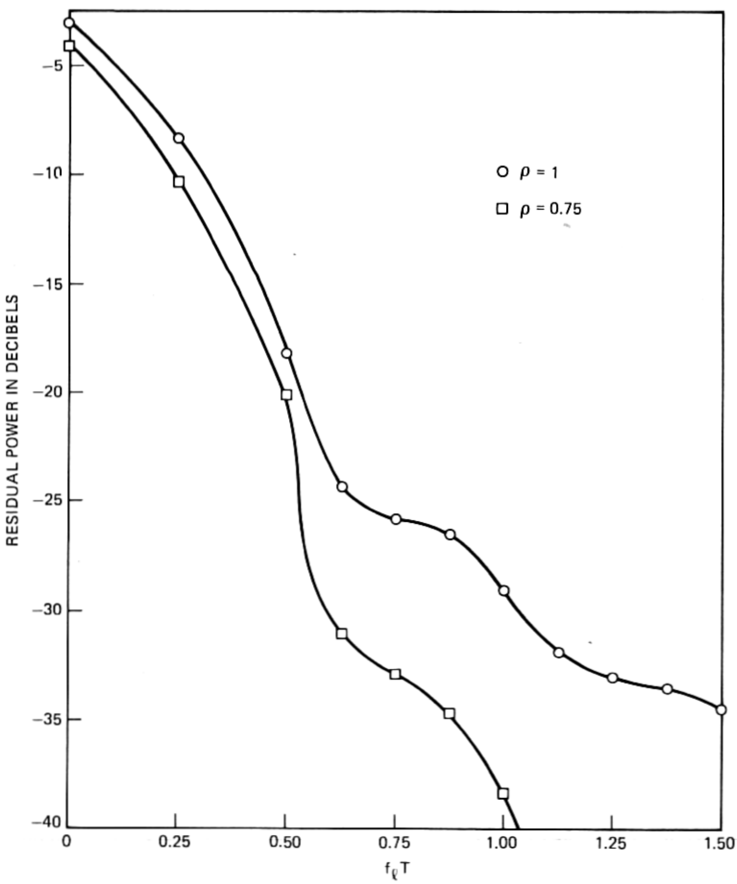


Fig. 7—Residual power in the upper sidelobes of FSK and AM-FSK.

These integrals can be expressed in terms of tabulated sine and cosine integrals.

After some algebra, we get

$$K_1 = \frac{1}{\pi} \ln \frac{2\nu_\ell T - \pi}{2\nu_\ell T + \pi} + \frac{1}{\pi} \operatorname{ci}(2\nu_\ell T - \pi) - \frac{1}{\pi} \operatorname{ci}(2\nu_\ell T + \pi)$$

$$K_2 = \frac{2}{2\nu_\ell T - \pi} (1 + \cos 2\nu_\ell T) - 2 \operatorname{si}(2\nu_\ell T - \pi)$$

$$K_3 = \frac{2}{2\nu_\ell T + \pi} (1 + \cos 2\nu_\ell T) - 2 \operatorname{si}(2\nu_\ell T + \pi),$$

where

$$\operatorname{si}(z) = -\int_z^\infty \frac{\sin x}{x}\, dx$$

and

$$\operatorname{ci}(z) = -\int_z^\infty \frac{\cos x}{x}\, dx.$$

Tables for the si(z) and ci(z) functions can be found in Ref. 7.

The optimal values of ρ are the roots of the quadratic equation (24). The root that is positive and less than one is the solution to our problem. It is given by

$$\rho_{\text{opt}} = \frac{K_2 - K_3}{2K_1} + \sqrt{\left(\frac{K_2 - K_3}{2K_1}\right)^2 + 1}. \tag{25}$$

REFERENCES

1. F. Bedrosian, "The Analytic Signal Representation of Modulated Waveforms," Proc. IRE, *50* (1962), pp. 2071–2076.
2. J. E. Mazo and J. Salz, "Spectral Properties of Single-Sideband Angle Modulation," IEEE Trans. Commun. Technol., *COM-16* (1968), pp. 52–61.
3. R. W. Lucky, J. Salz, and E. J. Weldon Jr., *Principles of Data Communication*, New York: McGraw-Hill, 1968, pp. 200–245.
4. J. Salz, "Spectral Density Function of Multilevel Continuous—Phase FM," IEEE Trans. Inform. Theory, *IT-11* (July 1965), pp. 429–433.
5. W. R. Bennett and S. O. Rice, "Spectral Density and Autocorrelation Functions Associated with Binary Frequency Shift Keying," B.S.T.J., *42* (September 1963), pp. 2355–2385.
6. R. R. Anderson, J. E. Mazo, and J. Salz, "Spectra for a Class of Asynchronous FM Waves," B.S.T.J., *44* (November 1965), pp. 2149–2163.
7. M. Abbramowitz and I. A. Stegun, "Handbook of Mathematical Functions," NBS Applied Math. Series, 55, June 1964.