2

Causes of Signal Impairment

S. Tirró

I. Introduction

The quality of the received signal may be impaired by many causes, such as noise, equipment distortions and/or mismatching, spectrum truncation, interference, propagation delay, and echo. All these causes are discussed extensively in this chapter, whereas other causes impairing the transmission of digital signals, such as modem imperfections, will be discussed in Chapter 10.

The noise generated inside the electronic equipment is called internal, as opposed to the noise received by the antenna from the surrounding space, which is called external. Internal and external noise are discussed respectively in Sections II and III. They may be mathematically modeled in the same way, using the Cartesian or polar noise representations discussed in Section IV.

The noise generated in the analog-to-digital conversion, which is called quantizing noise and shows different features, will be considered in Section V.

The subjective effect of baseband noise varies significantly with frequency and signal type. Therefore, different noise weighting curves have been specified, as is extensively discussed in Chapter 5 concerning signal quality.

The effects of equipment linear and nonlinear distortions are analyzed respectively in Sections VI and VII, particular emphasis being placed on generation of RF intermodulation products by nonlinear amplifiers.

Sections VIII to X deal respectively with spectrum truncation, intelligible crosstalk, and equipment mismatching.

The various causes of interference are discussed in Section XI, and Section XII deals with a phenomenon peculiar to satellite communications using the geostationary orbit, namely the echo due to the long propagation delay.

S. Tirró • Space Engineering S.r.l., Via dei Berio 91, 00155 Roma, Italy.

Satellite Communication Systems Design, edited by S. Tirró. Plenum Press, New York, 1993.

II. Internal Noise

Three types of internal noise exist, namely low-frequency noise, shot noise, and thermal noise. Low-frequency noise is due to irregularities in contact surface of the semiconductors and in cathodes of electron tubes; its power spectral density varies inversely with frequency. Usually this type of noise may be neglected; therefore the following discussion will concentrate on the other two types.

A. Shot Noise

Shot noise is due to the discrete structure of electricity and is generated when the current is due to rapid movement of relatively few electrons (Schottky effect). This noise contribution is therefore significant in tubes or semiconductors, whereas it may be neglected with respect to thermal noise in conductors (see next section). Shot noise had practical importance in the past only in tunnel diode amplifiers, which became obsolete due to the success of parameteric amplifiers and, later, FET amplifiers. Shot noise is no longer relevant in any part of a satellite communication system, but in the future it may again become important in the optical equipment used for intersatellite links (see Chapter 15).

Shot noise power is proportional to the expression

$$i^2 = 2eI\,\Delta f \tag{1}$$

where e = electron charge = 1.8×10^{-19} C
I = continuous current circulating in the tube or in the semiconductor
Δf = bandwidth in hertz

B. Thermal Noise and Noise Temperature

The random movement of electrons in any resistor generates noise (Johnson effect).[1,2] The available power spectral density of the noise generator equivalent to this phenomenon is given by the Nyquist formula

$$\frac{dW}{df} = \frac{hf}{e^{hf/kT_0} - 1} \tag{2}$$

where f = frequency (Hz)
h = Planck constant = 6.62×10^{-34} J · s
K = Boltzmann constant = 1.374×10^{-23} J/K
T_0 = physical temperature of resistor (K)

A more complete expression of the resistor noise temperature is

$$\frac{dW}{df} = \frac{hf}{e^{hf/kT_0} - 1} + hf \tag{2'}$$

where the last term is the lowest energy a harmonic oscillator can reach. This term is never zero, even at 0 K; therefore it is called *zero-level energy*. The role of this term becomes significant only at submillimeter wavelengths, and is very important at optical frequencies, especially when heterodyne or homodyne detection is employed.

In the radioelectric domain $hf/kT_0 \ll 1$ and (2) and (2′) simplify to

$$\frac{dW}{df} \approx KT_0 = \text{const}$$

whereas at the optical frequencies used in optical intersatellite links ($\lambda \approx 1$ micron) the complete expression (2′) must be used.

For radioelectric frequencies the generated noise voltage spectral density is therefore constant and of value

$$\overline{V_N^2(f)} = 4KT_0R_s \quad \text{V}^2$$

where K = Boltzmann constant = 1.374×10^{-23} W/Hz · K
T_0 = resistor physical temperature in kelvins
R_s = resistance value in ohms

Since the generated noise is zero if $T_0 = 0\,\text{K}$, this type of noise is called thermal, and since its spectral density is constant the noise is called white, in analogy with white light, which contains all colors with equal intensities.

Maximum power transfer occurs when the load resistance R_L equals the internal resistance of the generator, R_s. Therefore, half of the generator voltage may be obtained at the load resistance, with maximum power transfer of

$$N = \int_B \overline{V_N^2(f)} \frac{1}{4R_s} df = KT_0B \quad \text{W}$$

where B is the bandwidth of interest. This is the available noise power from any matched resistance and depends only on the physical temperature and on the bandwidth.

Any existing circuit or equipment generates thermal noise, because both resistive components and physical temperature of operation can never reach zero. The noise power generated inside the equipment and delivered at its output in a band B can always be written in the form

$$N = KT_{\text{eq}}B \tag{3}$$

where T_{eq} is the equivalent noise temperature, i.e., the physical temperature of a resistor generating the same available noise power in the same bandwidth. If the equipment has a power gain G, the equivalent noise temperature at the equipment input is defined as T_{eq}/G.

C. Noise of an Attenuator

It may be easily demonstrated that the equivalent noise temperature internally produced in a purely resistive component attenuating the input power in the $\alpha{:}1$ ratio ($\alpha > 1$), is

$$(\alpha - 1)T_0 \qquad \text{at the attenuator input} \tag{4}$$

$$\frac{\alpha - 1}{\alpha}T_0 \qquad \text{at the attenuator output} \tag{5}$$

If the attenuator input is closed on a matched resistor feeding the input with a T_0 noise temperature, the total output noise temperature is therefore

$$\frac{T_0}{\alpha} + \frac{\alpha - 1}{\alpha} T_0 = T_0$$

and the corresponding total input noise temperature is

$$T_0 + (\alpha - 1)T_0 = \alpha T_0$$

Therefore, in a purely attenuative equipment, the output noise temperature can never exceed the working temperature if the input noise temperature is lower than or equal to the working temperature. It is also easily seen that, for any noise temperature feeding the attenuator input, the output noise temperature will approach T_0 for very large values of α.

D. Noise Figure

The equipment noise figure F is defined as the ratio between the total output noise power and the output power obtained when the input is closed on a matched resistor working at the conventional room temperature of 290 K.[3] This figure has a constant value and is a real figure of merit of the equipment because the physical temperature of the resistor has been arbitrarily fixed; otherwise F would depend not only on the equipment noise performance but also on the level of the input noise.

Now let G be the equipment power gain and N_e the part of the total output noise generated inside the equipment. From the above definition it follows that

$$F = \frac{290GKB + N_e}{290GKB} = 1 + \frac{N_e}{290GKB} \tag{6}$$

If the noise temperature feeding the equipment input is really 290 K, F provides a measurement of the carrier-to-noise ratio deterioration from the equipment input to its output, i.e., a very convenient figure of merit of the equipment from the noise performance viewpoint.

III. External Noise

The noise received by the antenna is called external. If all the space surrounding the antenna is at physical temperature T, the antenna receives noise power

$$W = KT\,\Delta f \tag{7}$$

with the usual meaning of symbols.

Since the temperature is not constant, and considering also the variability of the antenna gain in the various directions, the antenna noise temperature can be written as

$$T_A = \frac{1}{4\pi} \iint G(\theta, \lambda) T(\theta, \lambda)\, d\Omega \tag{8}$$

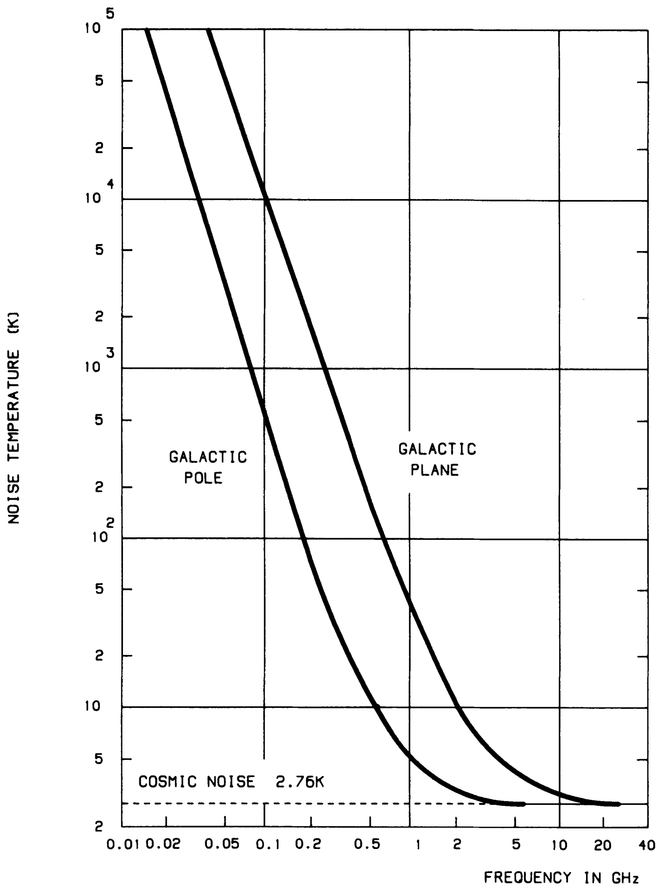


Fig. 1. Cosmic and galactic noise temperature.

where θ, λ define a direction in space and $d\Omega$ is the solid angle element around this direction.

The external noise sources may be classified as (1) extraterrestrial (cosmos, galaxy, sun, radiostars, planets) and (2) terrestrial (the earth itself, the atmosphere, industrial activities (man-made noise)).

Due to the typical working elevation angle, the earth antenna main radiation lobe and the first sidelobes are generally oriented toward the sky, and receive from this region cosmic noise (2.76 K) and galactic noise, which at the frequencies of interest for satellite communications are generally negligible (see Fig. 1).

The noise received from the sun, radiostars, and planets is generally negligible. It is important, however, to remember the interest of some radiosources for accurate antenna gain measurements and the outages caused by sun interference for a small part of the year (see Section VII E in Chapter 7).

The earth is generally considered to radiate a noise temperature of 290 K. An earth antenna receives this temperature with its far sidelobes (5 to 20 dB below the isotropic gain, depending on the antenna design). Therefore, the earth contribution to the antenna noise temperature may be 3–100 K. A satellite antenna will receive this temperature with its main lobe; however, the temperature received from the earth must be decreased by 10% for irregular land surface and by 50–70% for a regular wet soil or calm sea.

The atmospheric noise is generally the most important part of the earth station antenna noise and is produced by the atmospheric attenuation of the radioelectric wave. Therefore, the atmosphere can be modeled as an attenuator working at a mean physical temperature of about 270 K. The atmospheric attenuation value strongly depends on the local weather conditions. In clear weather, oxygen and water vapor effects prevail, whereas in bad weather the attenuation due to clouds, fog, and especially rain prevail. Section III in Chapter 8 discusses in detail the effects of the earth atmosphere on a radioelectric wave propagating through it.

Industrial or man-made noise is due to human activities on the planet surface, and is significant in big cities or industrialized areas. Measurements performed long ago in New York City have given the following results: the city may be modeled as a flat surface producing a noise temperature varying inversely to the frequency with a slope of 7.9 dB/octave and with a value of about 10,000 K at 1 GHz. Since the earth station antenna sees this noise source within a small solid angle and with the far sidelobes (due to the working elevation angle), the man-made noise contribution to the antenna noise temperature is generally small. At 4 GHz this contribution ranges between 1 and 2 K at 90° elevation and 4 and 6 K at 5° elevation for a 30-m-diameter antenna.

IV. Modeling of Internal and External Noise

Throughout this book the hypothesis will be maintained that predetector noise is of thermal origin or that its characteristics are similar to those of thermal noise. In other words, the noise is

- *Stationary*—its statistical characteristics do not vary in time.
- *Ergodic*—its statistical characteristics may be deducted from a single implementation of the noise process in time.
- *Gaussian*—its amplitude probability distribution is Gaussian.
- *White*—its power is uniformly distributed at all frequencies of interest.

These hypotheses are generally verified by any type of external or internal noise.

The following hypothesis will be added: the noise is of relatively narrow band. Since the noise bandwidth $2\,\Delta f$ of the filters intentionally placed behind the demodulators is much smaller than the carrier frequency f_c, this hypothesis is always well verified in real demodulating systems.

To build a simple mathematical model of the noise, a basic theorem of statistical mathematics will be used, stating that the sum of n statistically independent random variables tends to assume a Gaussian behavior when n approaches infinity, whatever the probability distribution of each random variable may be (central limit theorem). The proof of this theorem may be found, for instance, in Ref. 4. This result is very important because it is possible to build for the Gaussian noise an exact mathematical representation by using the sum of infinite sinusoidal components $A_K \sin(\omega_K t + \theta_K)$, with phases θ_K having rectangular probability distribution in $[-\pi + \pi]$ and arbitrary amplitudes A_K;[5] i.e.

$$n(t) = \sum_{K=1}^{\infty} A_K \cos(\omega_K t + \theta_K) \tag{9}$$

The correct amplitudes A_K of this representation can be derived by a nonrigorous procedure that is easy to understand. The $2\,\Delta f$ bandwidth can be subdivided into infinite bands of infinitesimal width δf, where the continuous noise spectrum can be replaced with a sinusoid of equal power. If N_0 is the noise power spectral density, it follows that the sinusoid amplitude must be such as to verify the equation

$$\frac{A_K^2}{2} = N_0 \delta f$$

therefore,

$$A_K = \sqrt{2N_0\delta f} \tag{10}$$

Since the noise is white, the amplitude will be equal for all sinusoids, so the noise expression may be written as

$$\begin{aligned} n(t) &= \sum_{K=1}^{\infty} \sqrt{2N_0\delta f}\cos(\omega_K t + \theta_K) \\ &= \sum_{K=1}^{\infty} \sqrt{2N_0\delta f}\cos\{[(\omega_K - \omega_c)t + \theta_K)] + \omega_c t\} \end{aligned}$$

where $\omega_c = 2\pi f_c$.

Using the cosine addition formula and defining,

$$x_p(t) = \sum_{1}^{\infty} \sqrt{2N_0\delta f}\cos[(\omega_K - \omega_c)t + \theta_K]$$

$$x_q(t) = \sum_{1}^{\infty} \sqrt{2N_0\delta f}\sin[(\omega_K - \omega_c)t + \theta_K]$$

we obtain

$$n(t) = x_p(t)\cos\omega_c t - x_q(t)\sin\omega_c t \tag{11}$$

where $x_p(t)$ and $x_q(t)$ are noise components in phase and in quadrature, respectively, with the carrier.

Also, $x_p(t)$ and $x_q(t)$ are the sum of infinite sinusoidal components with rectangular probability phases, so they are Gaussian. In addition, the total power of $x_p(t)$ and $x_q(t)$ equals the power of the noise $n(t)$. In fact,

$$\overline{x_p^2(t)} = \overline{x_q^2(t)} = \sum_{1}^{\infty} \frac{1}{2}(\sqrt{2N_0\delta f})^2 = N_0 \sum_{1}^{\infty} \delta f = \sigma^2 \tag{12}$$

while

$$\overline{n^2(t)} = \overline{x_p^2(t)\cos^2\omega_c t} + \overline{x_q^2(t)\sin^2\omega_c t} = \frac{\overline{x_p^2}}{2} + \frac{\overline{x_q^2}}{2} = \sigma^2 \tag{13}$$

The probability distribution of $x_p(t)$ has the Gaussian form

$$p(x_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2\sigma^2}\right) \tag{14}$$

and a similar formula may be written for $p(x_q)$.

Since $x_p(t)$ and $x_q(t)$ are composed by pairs of sinusoidal functions of equal amplitudes and quadrature phases, they will also have equal amplitudes and be in quadrature to each other; i.e.,

$$x_p(t) = V(t)\cos\phi(t)$$
$$x_q(t) = V(t)\sin\phi(t)$$

where $V(t)$ and $\phi(t)$ are two slowly varying time functions, since for the narrowband hypothesis, $|\omega_k - \omega_c| \ll \omega_c$ for every value of k. The noise expression may therefore be rewritten as

$$n(t) = V(t)\cos[\omega_c t + \phi(t)] \tag{15}$$

i.e., a narrowband noise may also be represented by a sinusoid having both amplitude and phase slowly varying in time. This new representation is in polar coordinates, while the previous one was in Cartesian coordinates.

It can be demonstrated that V and ϕ are independent statistical variables and their probability distributions are as follows:

$$\begin{aligned} q(V) &= \begin{cases} \dfrac{V}{\sigma^2} e^{-V^2/2\sigma^2} & \text{for } V \geqq 0 \\ 0 & \text{for } V < 0 \end{cases} \\ q(\phi) &= \begin{cases} \dfrac{1}{2\pi} & \text{for } 0 \leqq \phi \leqq 2\pi \\ 0 & \text{elsewhere} \end{cases} \end{aligned} \tag{16}$$

Note that $q(V)$ is a Rayleigh distribution.

Both the Cartesian and the polar noise representations will be used in this book.

V. Quantizing Noise and Digital Companding

Quantizing noise is created when an originally analog signal is converted into digital form. This type of noise is zero only when the amplitude of each signal sample is exactly equal to one of the values represented by the available codewords. To be coded, the signal is first sampled and then quantized; however, there are often important reasons to compress the signal prior to coding it, using a suitable $y = F(x)$ law (see Fig. 2). The quantized signal will be obtained by choosing, for each sample value $y(t)$, the closest existing value $\bar{y}(t)$ in the available set of codewords. On the receiving side, in the hypothesis of ideal transmission (no errors), after application of the appropriate expansion function

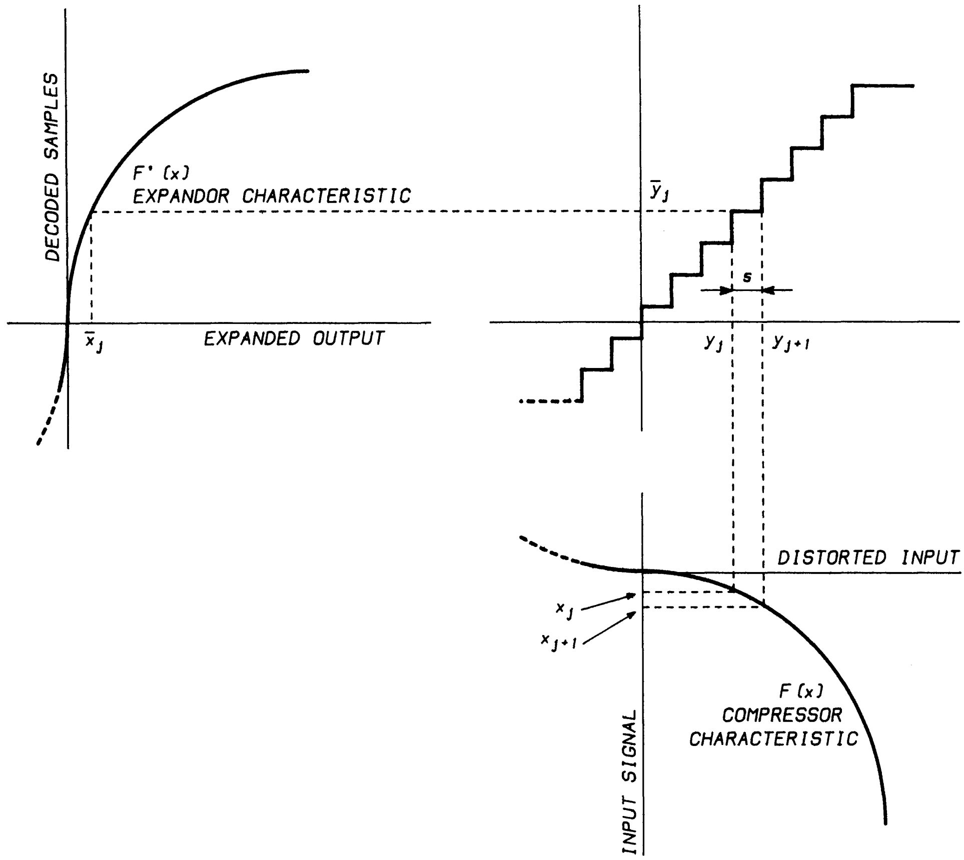


Fig. 2. Nonuniform codec using a compandor. (Reprinted from Ref. 11, with permission of AT&T, © 1982 AT&T.)

$x = F^{-1}(y)$, a received signal $\bar{x}(t)$ is obtained. The difference

$$\bar{x}(t) - x(t) = n_q(t) \tag{17}$$

is the quantizing noise.

Now let the signal be defined in the conventional interval $(-1, +1)$ and the available quantization levels suitably distributed there (in other words, no overload is possible). If n is the number of bits per codeword, there will be $N = 2^n$ available codewords.

The available quantized levels will be

$$\bar{x}_i = \frac{x_{i+1} - x_i}{2}, \qquad i = -\frac{N}{2}, \ldots, -1, 0, +1, \ldots, +\frac{N}{2}$$

where x_i are the extremes of the various quantization intervals. The mean square expected error (which equals the quantizing noise power) will be

$$\overline{e^2} = \sum_{i=-N/2}^{+N/2} \int_{x_i}^{x_{i+1}} (\bar{x}_i - x)^2 p(x)\, dx \tag{18}$$

where $p(x)$ is the probability density of the input signal. The determination of

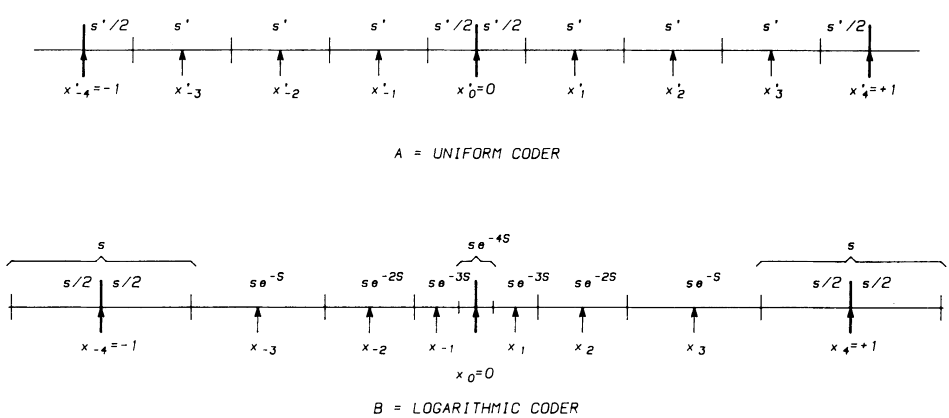


Fig. 3. Quantization intervals for $n = 3$, $N = 2^3 = 8$.

the $\bar{x}_i$ values which minimize $\overline{e^2}$ (synthesis problem) always requires knowledge of $p(x)$ for every type of coder; on the contrary, the calculation of $\overline{e^2}$ for a given set of $\bar{x}_i$ values (analysis problem) does not require such a knowledge for particular types of coders (linear, logarithmic), as will be seen shortly.

The coder is called linear (or uniform) if the quantized levels are equally spaced (see Fig. 3a); in this case the compressor is absent and the constant amplitude of the quantization intervals is given by

$$s' = \frac{2}{2^n} = \frac{2}{N}$$

Naturally, the best choice for a linear coder is to place the highest and lowest values of the codewords at a distance $s'/2$ from the highest and lowest possible signal values, respectively. Under these conditions,

$$x_i = is, \qquad i = -\frac{N}{2}, \ldots, -1, 0, +1, \ldots, +\frac{N}{2}$$

If s' is small enough (i.e., N is large), it is possible to consider $p(x) = p_i$ = constant over each interval; therefore

$$\begin{aligned} \overline{e^2} &\approx \sum_{i=-N/2}^{+N/2} p_i \int_{x_i'}^{x_{i+1}'} (x_i - x)^2\,dx = \sum_{i=-N/2}^{+N/2} \frac{(x_{i+1}' - x_i')^3}{12} p_i \\ &= \frac{s'^2}{12} \sum_{i=-N/2}^{+N/2} (x_{i+1}' - x_i')p_i = \frac{s'^2}{12} \end{aligned} \tag{19}$$

That is, knowledge of $p(x)$ is not necessary to compute the quantization noise of the linear coder, as anticipated. This result is rigorously valid for any signal with a rectangular probability density function (e.g., a sawtooth), and holds with good approximation for any regular signal if N is large enough.

Bennett[6] has demonstrated that the spectrum of the quantizing noise is approximately white in the signal bandwidth.

Assuming now that the signal is sinusoidal, with amplitude $2^n s' = 2$ peak-to-peak, the sinusoid power is

$$S = \frac{1}{2}\left(\frac{2^n s'}{2}\right)^2 = \frac{(2^n s')^2}{8}$$

Therefore, the signal-to-quantizing noise power ratio is

$$\frac{S}{Q} = \frac{S}{\overline{e^2}} = 10\,\mathrm{Log}_{10}\frac{(2^n s')^2/8}{s'^2/12} = 6n + 1.8 \quad \mathrm{dB} \tag{20}$$

Each additional bit per codeword improves therefore by 6 dB the S/Q ratio. This value of the S/Q ratio is obtained only when the signal peak-to-peak amplitude is such as to occupy all the coder dynamic range. If the signal power is decreased with respect to this maximum value, the S/Q value will proportionally decrease.

A uniform codec is therefore not convenient when the message structure is *a priori* largely unknown; this is surely the case of speech signals. It was seen in Section II F of Chapter 1 that the speech peak factor is 18.6 dB on the average, but it may reach 25 dB for particular talkers. If the variation of average volume from one talker to another is also considered, the dynamic range of the input signal may reach 40 dB. A uniform codec would therefore provide an S/Q ratio 40 dB worse for weak talkers with respect to strong talkers.

Uniformization of codec S/Q performance over a wide range of input levels is obtained by using nonuniform codecs, where the steps are of different amplitudes; in this case a compressor precedes the coder at the transmitting side, whereas an expandor follows the decoder at the receiving side. The compressor–expandor pair is called a compandor. It is important not to confuse this digital, instantaneous compandor with the analog syllabic compandor, which will be discussed in Section II of Chapter 9.

Let $y = F(x)$ be the compressor characteristic and $F'(x)$ its first derivative. If $s = y_{i+1} - y_i$ is kept constant, we have

$$s = F(x_{i+1}) - F(x_i) \approx F'(x_i)(x_{i+1} - x_i)$$

Therefore, the quantizing noise power is

$$\overline{e^2} \approx \frac{s^2}{12}\int \frac{p(x)}{[F'(x)]^2}\,dx \tag{21}$$

To obtain a constant S/Q over a wide range of input levels, it would be ideal to use a logarithmic $F(x)$; in this case $F'(x) = 1/x$ and

$$\overline{e^2} \approx \frac{s^2}{12}\int p(x)x^2\,dx = S\frac{s^2}{12} \tag{22}$$

so

$$\frac{S}{Q} = \frac{S}{\overline{e^2}} = \frac{12}{s^2} = \mathrm{const} \tag{23}$$

As anticipated, as with logarithmic coding, calculation of the quantizing noise power does not require knowing $p(x)$. The value of S/Q so obtained for logarithmic coding cannot be immediately compared with the one obtained for linear coding, since $s \neq s'$ (see Fig. 3b). It is immediately derived that

$$x_{\max} = e^0 = 1$$
$$x_{N/2} = e^{-s/2} \simeq 1 - s/2$$
$$x_{N/2-1} = e^{-s/2}e^{-s} \simeq \left(1 - \frac{s}{2}\right)e^{-s}$$
$$\vdots$$
$$x_{N/2-k} = e^{-s/2}e^{-ks} \simeq \left(1 - \frac{s}{2}\right)e^{-ks}$$
$$x_{N/2-(k+1)} \simeq \left(1 - \frac{s}{2}\right)e^{-(k+1)s}$$
$$\vdots$$
$$x_1 = x_{N/2-(N/2-1)} \simeq \left(1 - \frac{s}{2}\right)e^{-(N/2-1)s}$$

Therefore,

$$\Delta x_K = x_{N/2-K} - x_{N/2-(K+1)} = \left(1 - \frac{s}{2}\right)[e^{-Ks} - e^{-(K+1)s}]$$
$$\simeq \left(1 - \frac{s}{2}\right)se^{-Ks} \simeq se^{-Ks}$$

That is, the amplitude of the quantizing intervals decreases when x goes from 1 to 0.

Of course, the sum of all positive quantizing intervals must equal unity:

$$\frac{s}{2} + \sum_{K=0}^{N/2-2} \Delta x_K + \frac{1}{2}\Delta x_{N/2-1} = 1$$

Therefore

$$\frac{s}{2} + \sum_{K=0}^{N/2-2} se^{-Ks} + \frac{1}{2}se^{-(N/2-1)s} = 1$$

Let $N = 256$ (8 bits/sample) and recall the formula for the sum of a geometrical series; then

$$\frac{s}{2} + se^{-s}\frac{e^{-126s} - 1}{e^{-s} - 1} + \frac{1}{2}se^{-127s} = 1$$

The solution is $s \simeq 0.04$. Therefore, $S/Q \simeq 38.7$ dB, compared with 49.8 dB obtained by linear coding (where $s' = \text{const} = 1/128 = 0.0078$). Achievement of a quality constant over a very large dynamic range by logarithmic companding is therefore balanced by a significant deterioration (about 11 dB) of the quality obtainable at the highest talker level.

The analysis up till now was based on sinusoidal signals, which do not show overload, but are not representative of a real speech signal. However, they are interesting for testing the codec alignment, since they easily show incorrect placement of the quantization levels. A speech signal is better represented by a Laplacian amplitude distribution (see Section II D in Chapter I), which is basically an exponential distribution. Since the quantization noise is minimized when the quantization level frequency is proportional to the signal amplitude probability density function, the important result is obtained that a logarithmic compression law minimizes the quantization noise in the case of a speechlike signal, in addition to keeping constant the S/Q ratio over a wide dynamic range. Some deterioration of the S/Q ratio must be expected when the signal power is high, due to the significant overload probability, which is obtained in these conditions and which has been neglected in the previous analysis.

The compression laws standardized by the CCITT are modified into a linear or pseudolinear law for low signals.[7]

The compression law used in the United States and Japan is the μ-law:[8]

$$F_\mu(x) = \operatorname{sgn}(x)\frac{\ln(1 + \mu\,|x|)}{\ln(1 + \mu)}, \qquad -1 \le x \le +1 \tag{24}$$

whereas the compression law used in Europe is the A-law:[9,10]

$$F_A(x) = \begin{cases} \operatorname{sgn}(x)\dfrac{1 + \ln A\,|x|}{1 + \ln A}, & \dfrac{1}{A} \le |x| \le 1 \\ \operatorname{sgn}(x)\dfrac{A\,|x|}{1 + \ln A}, & 0 \le |x| \le \dfrac{1}{A} \end{cases} \tag{25}$$

where $\operatorname{sgn}(x)$ is the sign of x.

Since $F_A(x)$ is logarithmic for $|x| > 1/A$ and linear for $|x| < 1/A$, it provides a perfectly flat S/Q quality for high signal levels and a slightly worse performance for lower signal levels with respect to the μ-law.

Figure 4 shows the μ-law compression characteristic for various values of μ; linear coding (no compression) is obtained for $\mu = 0$, while the commonly used value $\mu = 255$ provides a performance very close to that of the ideal logarithmic compression (see Fig. 5). The curves in Fig. 5 are smooth; the actual curves have a fine-grained sawtooth structure, which may be shown by speech signal Laplacian representation (Section II D, Chapter 1). Some 10 dB of dynamic range are lost in the upper region (input level higher than -10 dBm0) when a speechlike signal is considered, instead of a sinusoid, due to the increased overload probability. The behavior of the A-law is similar, with the small differences already mentioned.[11]

The quasi-logarithmic compression law is inherited from the past. The first implementation of the logarithmic compressor was analog, giving rise to divergence and ambiguity for sufficiently small value of s; in other words, there is no bijective mapping between x and y when $|x| < 1$ (see Fig. 6). Both divergence and ambiguity are avoided by using a linear law when x is close to zero. Today, thanks to VLSI digital circuits, implementation of the logarithmic compressor in a single chip, following the sampling circuit, is possible (see Fig. 7b). In this way

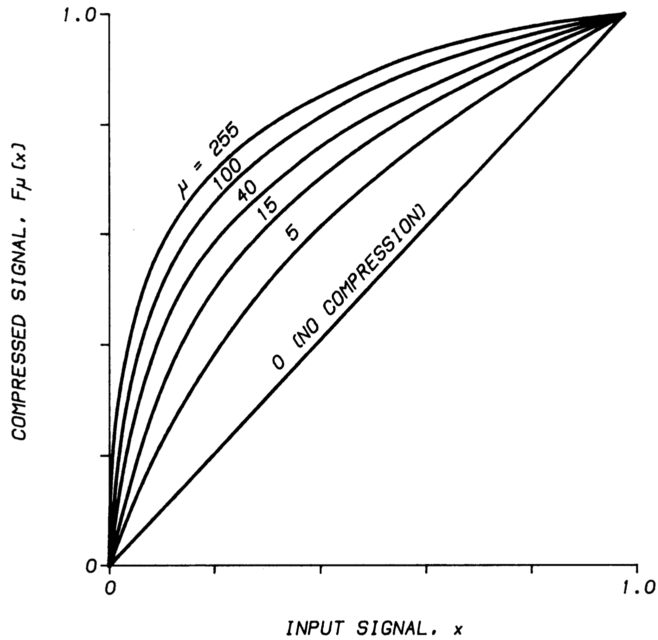


Fig. 4. Logarithmic compression characteristics (μ-law). (Reprinted from Ref. 11, with permission of AT&T, © 1982 AT&T.)

there is no passage through the y variable, no ambiguity, and no need for a linear region in the compression characteristic. For practical reasons, however, the logarithmic characteristic is still approximated by a multiple-segment law, with a segment including zero, similar to the old quasi-logarithmic laws for analog implementation. The number of segments is 13 for the A-law and 15 for the μ-law.

VI. Equipment Linear Distortions

When a system component produces an output strictly proportional to the input at a given frequency, this component is said to perform linearly. The effects

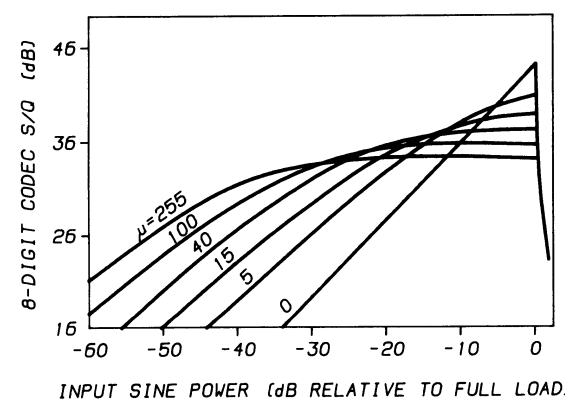


Fig. 5. S/Q performance for μ-law codecs. (Reprinted from Ref. 11, with permission of AT&T, © 1982 AT&T.)

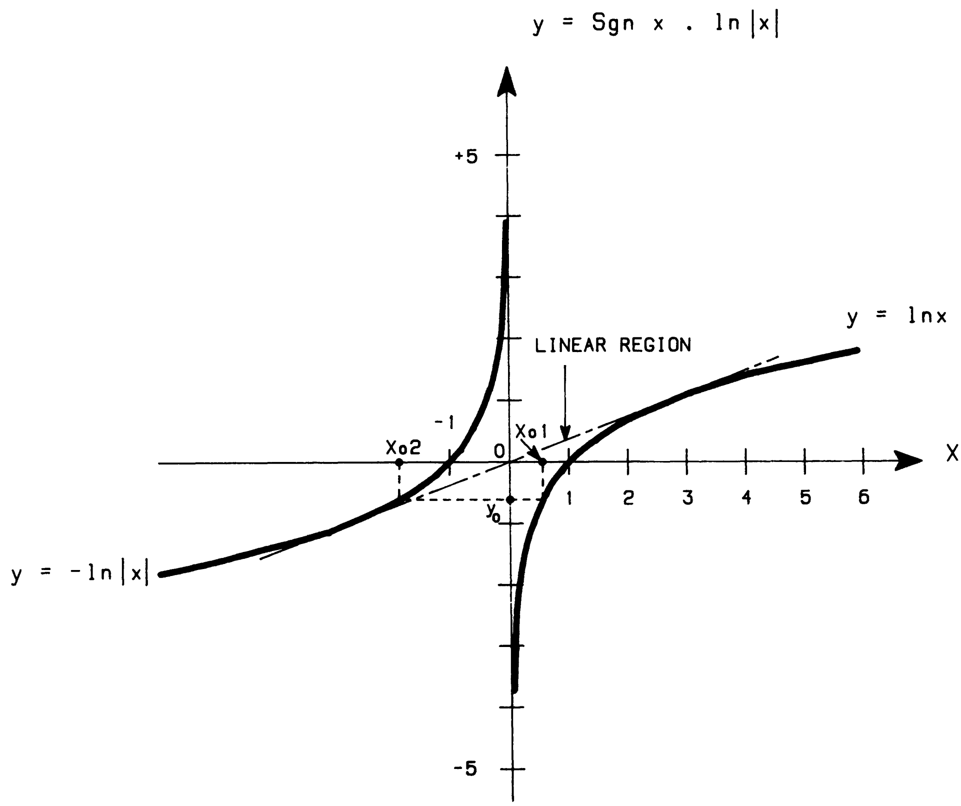


Fig. 6. The purely logarithmic law $y = \operatorname{sgn} x \cdot \ln |x|$ originates ambiguity, which may be solved by a linear region around the origin.

of nonlinearity are dealt with in Section VII. Even when performing linearly, a component may be nonideal, because its behavior may be different from one frequency to another. It is well known that a perfect transmission is obtained only if the component gain and the time delay due to the signal propagation from component input to output are both constant at all frequencies of interest. The second condition is equivalent to saying that the signal phase at the component output must vary in proportion to the frequency.

The deviations of the component gain from the ideal behavior are measured by applying to the component input a sinusoid of constant amplitude and of frequency varying in time over the range of interest; the amplitude variations measured at the component output will provide the desired information.

The deviation of the component delay from the ideal behavior is also called group delay distortion (GDD) or envelope delay distortion (EDD), and is measured by applying to the component input an amplitude-modulated carrier, which is swept over a bandwidth much larger than the modulating frequency. If subscripts c and m indicate the carrier and the modulating tone, the following relations must be verified:

$$\omega_c \gg \omega_m, \qquad A_c \gg A_m$$

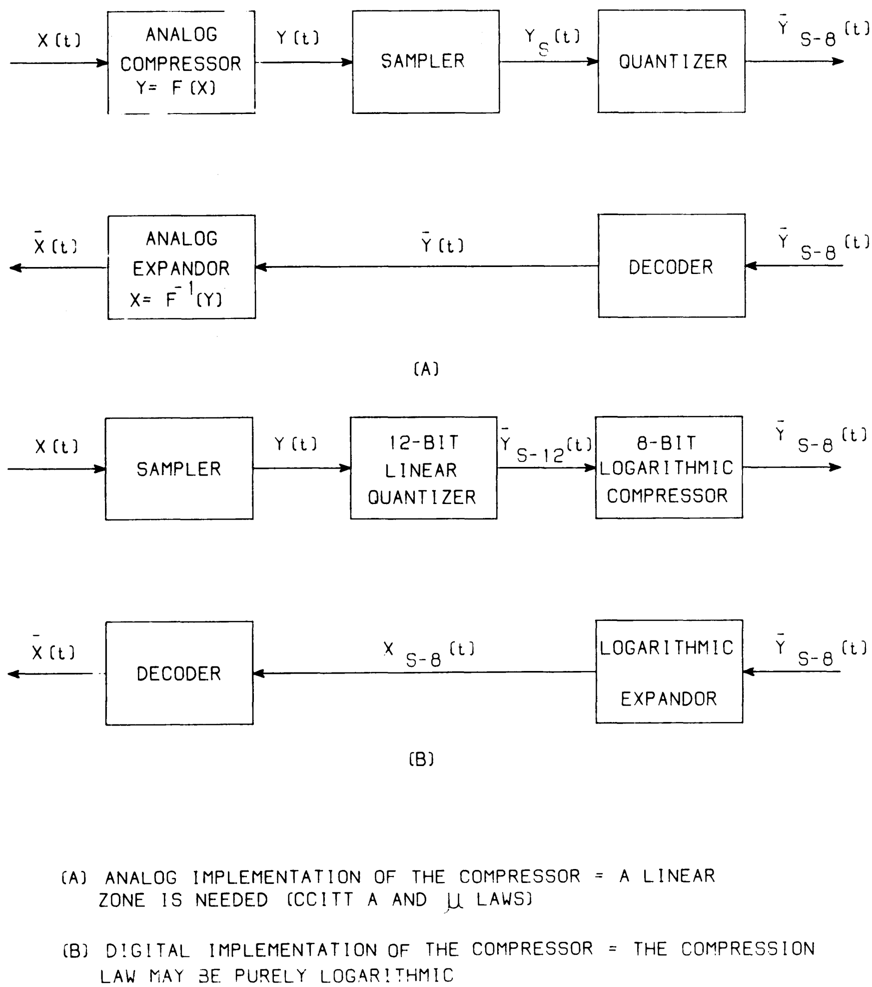


Fig. 7. Implementation of a pulse-code modulation (PCM) coder with logarithmic compression law.

Three phase-coherent frequencies are thus obtained, enabling the measurement at the component output of any deviation from linearity of the phase characteristic.

The above-defined linear distortions are not as powerful as the high-power amplifier (HPA) nonlinear distortion, which generates intermodulation products, causing interference from one carrier to another. Their effect, however, cannot be neglected, and consists of

- Production of intermodulation noise among the channels pertaining to a multichannel analog carrier (see, for instance, Section V I in Chapter 9 for frequency modulation multichannel telephony)
- Distortion of the television signal in case of analog transmission (see Section VII C in Chapter 9)
- Creation of intersymbol interference in digital transmissions (see Section III B in Chapter 10)

The amplitude response and group delay response of a component or chain of components are generally modeled by the sum of

- A linear component, or slope, which is measured in %/Hz (or, more commonly, in dB/MHz) for the gain response, and in ns/MHz for the group delay response
- A parabolic component, measured in dB/MHz^2 and in ns/MHz^2, respectively, in the two cases
- A residual ripple (having taken out the linear and parabolic components) measured in dB and nanoseconds peak-to-peak respectively in the two cases

It is common practice to define masks for the specification of amplitude and group delay distortions; in Section VI of Chapter 5 detailed information will be provided on the masks specified by Intelsat for various types of signals.

VII. Nonlinear Distortions

A. General

When several frequencies simultaneously transit through a nonlinear component, intermodulation products arise, which must be considered as causes of signal impairment. The most complex situation occurs when each frequency is a frequency-modulated (FM) carrier; in this case intermodulation arises:

- In video amplifiers and FM modems: the voltage-frequency characteristic of the modulator and the frequency-voltage characteristic of the demodulator cannot be perfectly linear in real equipment.
- In high-power amplifiers, which show a saturation effect (i.e., the output power has a maximum value for any input power), and therefore a power response which is nonlinear when sufficiently close to saturation.

These two types of impairment sources will be called video and radiofrequency nonlinear distortion, respectively. Video nonlinear distortion generates intermodulation among baseband frequencies, so it has no impact on the predetection noise level, whereas RF nonlinear distortion generates intermodulation among radiofrequency carriers and produces intermodulation noise to be added to the thermal noise at the detector input. For this reason these two sources of impairment are considered in a completely different way:

- Baseband intermodulation noise is considered in the distortion noise budget together with baseband noise due to equipment linear distortions and to equipment mismatching (see Section X).
- RF intermodulation noise is considered in the link budget together with thermal and shot noise, and with interference.

This section concentrates on RF intermodulation noise, which is by far the more important. Noise produced by video nonlinear distortions is considered in Chapter 9.

In amplitude modulation the radiofrequency spectrum is obtained by simple translation of the baseband spectrum. Therefore, the intermodulations generated by the video nonlinearity and by the RF nonlinearity have the same nature and are considered together.

The nonlinear characteristic of the microwave power amplifier may be modeled as

$$F(a) = V(a)e^{j\phi(a)} \tag{26}$$

where a = input signal amplitude
$V(a)$ = AM–AM characteristic
$\phi(a)$ = AM–PM characteristic
(AM = amplitude modulation and PM = phase modulation).

Since power is an expensive resource, it is not usually possible to use the HPA sufficiently far from saturation (i.e., well within the linear region) so as to obtain a negligible level for the intermodulation products. Rather, the HPA operating point must be optimized by trading the impairment due to intermodulation products for the improvement due to larger power of the transmitted carrier. This trade-off is rather complex and will be discussed in Chapter 11, Section VI.

Multicarrier operation historically occurred first in satellite HPAs, due to the extensive use of frequency-division multiple access since early satellite communications. Later the amplification of several carriers in the same HPA was also used in earth stations.

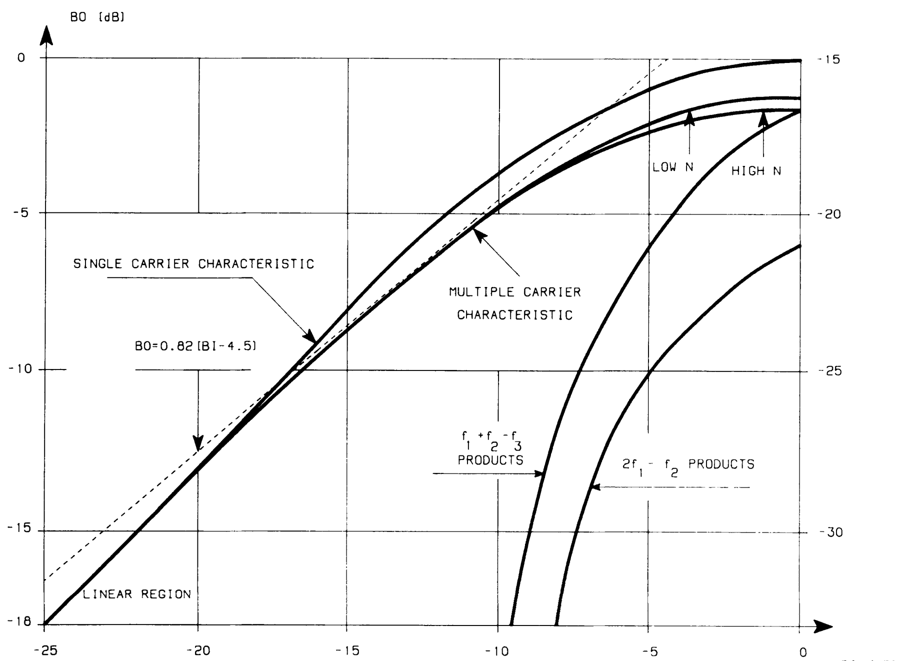


Fig. 8. Typical HPA characteristic.

Figure 8 shows a typical HPA characteristic,[12] where it may be noted that

1. Saturation always occurs for the same value of input power, regardless of the number of carriers.
2. The total useful output power is constant regardless of the number of carriers in the linear region (where the power lost in intermodulation products is negligible), while for higher power values it decreases when the number of carriers increases.
3. Input and output powers in the figure refer to the respective single-carrier saturation values; when the power value is lower than the saturation value by X dB, the HPA is usually said to operate with a back-off of X dB. In Fig. 8 the abscissa gives the HPA input back-off, simply called BI, while the corresponding ordinate gives the HPA output back-off, BO.
4. When the HPA amplifies two carriers, the total useful power at saturation decreases by about 1.2 dB with respect to the single-carrier case. This loss of useful power is called natural BO of the HPA and increases slightly to 1.5–1.6 dB for a large number of carriers.
5. Whereas BI is always intentional, BO in general is the sum of a natural BO and an intentional BO.
6. The characteristic is practically equal for any number of carriers when the tube is operated with multiple carriers sufficiently far from saturation. In particular, for BO = 5–8 dB the system is optimized for multicarrier operation (see Chapter 11, Section VI), and the tube characteristic may be approximated as

$$\mathrm{BO} = 0.82(\mathrm{BI} - 4.5) \tag{27}$$

where both back-off values are in dB.

Another important phenomenon due to amplifier nonlinearity is the *compression effect.* When two carriers of different levels are applied to the HPA input, the ratio between the two carrier powers increases from the input to the output, so the carrier of lower power is "compressed" with respect to that of higher power. The excitation levels at the HPA input must therefore be adjusted so as to obtain the desired power ratio at the output.

B. Unmodulated Carriers and Intermodulation Lines

Applying the signal

$$s(t) = \sum_{i=1}^{N} A_i \sin(2\pi f_i t - \phi_i) \tag{28}$$

which is the sum of N unmodulated sinusoids, to the input of the nonlinear component, one will obtain at the nonlinear component output, in addition to the input frequencies, frequencies, which do not exist at the input, generated by the nonlinear amplification process. These frequencies are called intermodulation products or, in the case of modulation absence, intermodulation lines. The frequency of the intermodulation product is in general

$$f_x = \sum_{i=1}^{N} m_i f_i \tag{29}$$

where f_i are the input frequencies and m_i are integer numbers, which may be positive or negative.

The order of the intermodulation product is defined as

$$K = \sum_{i=1}^{N} |m_i| \tag{30}$$

Therefore, for instance, $2f_1 + f_2 - 2f_3$ is a fifth-order product.

In satellite communications the transponder frequency is large with respect to the transponder bandwidth, so even-order intermodulation products fall out of the useful band. Only odd-order products thus have practical interest, and we need discuss only third-order and fifth-order products because, in general, the power of the intermodulation product decreases with the order of the product.

The number of intermodulation products increases very quickly with the number of input sinusoids. For instance, there may be $N(N-1)$ products of the $2f_1 - f_2$ type, and $\frac{1}{2}N(N-1)(N-2)$ products of the $f_1 + f_2 - f_3$ type, while the number of the fifth-order products is much larger. As a consequence, rigorous analyses are possible only for small values of N. Hence, the nonlinearity must be modeled as the sum of a simple amplitude nonlinearity plus a module producing conversion of amplitude modulation to phase modulation (see Fig. 9).

A simplified model may include only amplitude nonlinearity, approximated by a sum of sinusoids with real coefficients.[13] Figure 10 gives the results obtained with this model for a typical HPA when five equal sinusoids are amplified. The level variations of a carrier produce parallel level variations of all the intermodulation products generated by this carrier. In general, a 1-dB variation of the ith carrier level will cause an m_i-dB variation of the intermodulation product, where m_i is defined in (29).

From Fig. 10 one may deduct the carrier-to-third-order product power ratio for a given value of output back-off. In particular, the C/I value obtained with two equal carriers when the output back-off is 10 dB is called D_3 and is often specified by the tube manufacturer. A typically specified value of D_3 is about

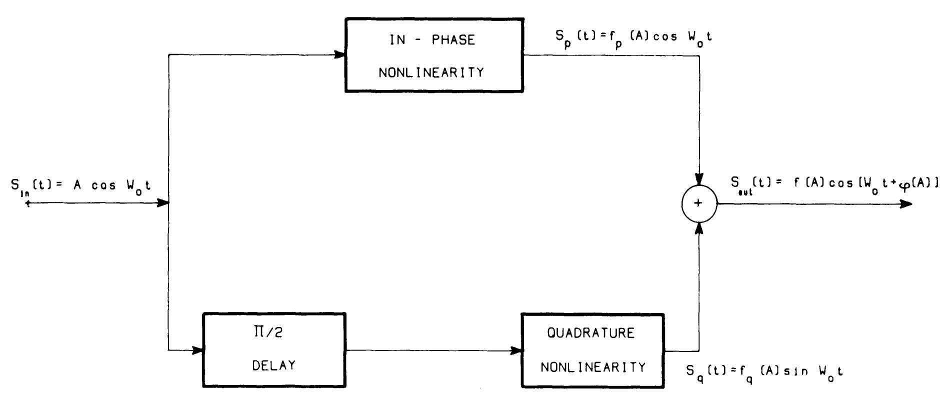


Fig. 9. TWTA modeling. $\phi(A)$ is the phase variation due to AM–PM conversion. A simplified amplitude nonlinearity model, without AM–PM conversion, is obtained by neglecting the quadrature component.

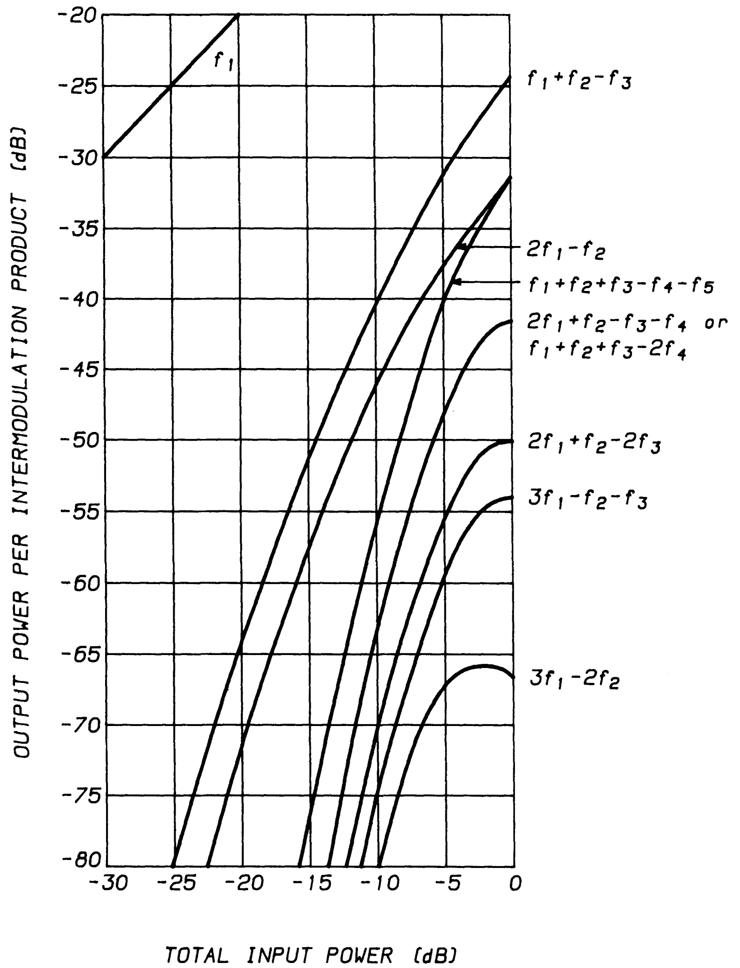


Fig. 10. Intermodulation with five carriers. (Reprinted with permission from Ref. 16.)

26 dB, whereas typical measured values are 29–30 dB. These values are significantly worse than the value provided by Fig. 10, due to the higher power per carrier (two frequencies instead of five share the tube power) and to the simplifications in the model, which does not consider the effects of the AM–PM conversion (see next section). Once D_3 is known, it is possible to compute the power of the $2f_i - f_j$ product by using the following approximate formula, valid for conditions sufficiently close to those used for measuring D_3:

$$IM_{ij} = -D_3 - 2P_0 + 2P_i + P_j \tag{31}$$

where P_0 = reference power used for D_3 measurement (dBW)
P_i = power of the ith carrier (dBW)
P_j = power of the jth carrier (dBW)

When more than two carriers are used, third-order products of the type $f_1 + f_2 - f_3$ arise, the power of which may be evaluated by the approximate formula

$$IM_{1,2,3} = -D_3 - 2P_0 + P_1 + P_2 + P_3 + 6 \tag{32}$$

This type of product is therefore dominant for more than two carriers, due to the 6-dB higher power value (see, in particular, Fig. 8 for the case of three equal carriers).

According to (32), if the power of all carriers is reduced by 1 dB the intermodulation products level is decreased by 3 dB, with a consequent improvement of the C/I ratio of 2 dB. However, this is only true in the linear region, since in the saturation region 1 dB of increase of the input back-off causes an IM reduction smaller than 3 dB and an output power reduction smaller than 1 dB. The overall result is that

- The IM reduction is larger than 3 dB/dB of increase in the output back-off.
- The C/I improvement is larger than 2 dB/dB of increase in the output back-off.

Of course, the precise values of the C/I slope will depend on the selected TWTA input–output characteristic. For the TWTA defined in Fig. 8 the C/I slope will be about 3 dB/dB for BO = 3.5–7.5 dB, and about 4.5 dB/dB close to saturation (see Fig. 11).

The analysis of the AP–PM conversion is much more difficult and requires expansion of the nonlinearity in a series of Bessel functions with complex coefficients.[14] Comparison of the calculations with available measurements shows that the nonlinearity is memoryless (i.e., its behaviour does not depend on the signal frequency) for the helix TWTAs presently used onboard communication satellites. Also the high-power cavity TWTAs used in earth stations may be considered memoryless in a narrow bandwidth.[14]

C. Modulated Carriers and Intermodulation Noise

When the N carriers are frequency-modulated with large modulation index and Gaussian modulating signal (which is a good approximation of a multiplex telephone signal), their spectrum is well approximated by a Gaussian spectrum with variance σ_i^2 (see Chapter 9, Section IV C). The spectrum of the intermodulation product f_x will also be Gaussian, with variance

$$\sigma_x^2 = \sum_{i=1}^{N} (m_i \sigma_i)^2 \tag{33}$$

The sum of all intermodulation products spectra will produce a total spectrum of intermodulation noise. The intermodulation noise must be considered together with downlink and uplink thermal noise in the satellite system design, and has therefore received much attention since the early days of satellite communications. The first measurements were performed in 1967 by Westcott at the British Post Office.[15]

The level of the intermodulation noise on the useful carrier may be minimized by proper frequency plan selection. An interesting example is reported in the CCIR Handbook[16] for 10 FDM–FM carriers occupying a total bandwidth

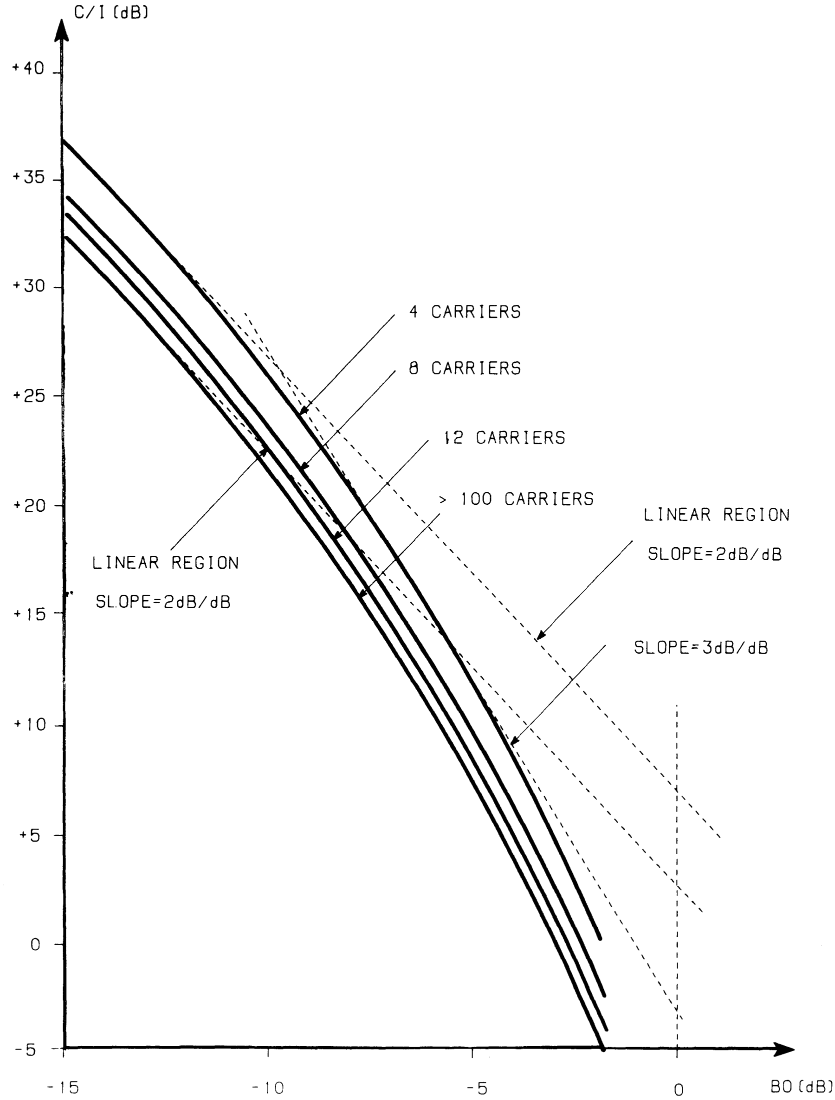


Fig. 11. Carrier-to-intermodulation noise power ratio vs. satellite TWTA output back-off and number of active carriers for the most disturbed carrier, located at the center of the transponder band.

of about 90 MHz. The highest-power carriers have a 132-channel capacity and occupy the external positions in order to push out of the useful band most of the intermodulation noise. Immediately after we find two 60-channel carriers, and six 24-channel carriers occupy the inner positions. Figure 12 shows the carrier-to-intermodulation noise power ratio due, for this example, to the amplitude nonlinearity alone (dashed curves) and to the combined effect of amplitude nonlinearity and AM–PM conversion (solid curves). These results are typical of a satellite helix TWTA, with a 1–1.5°/dB AM–PM conversion coefficient, and

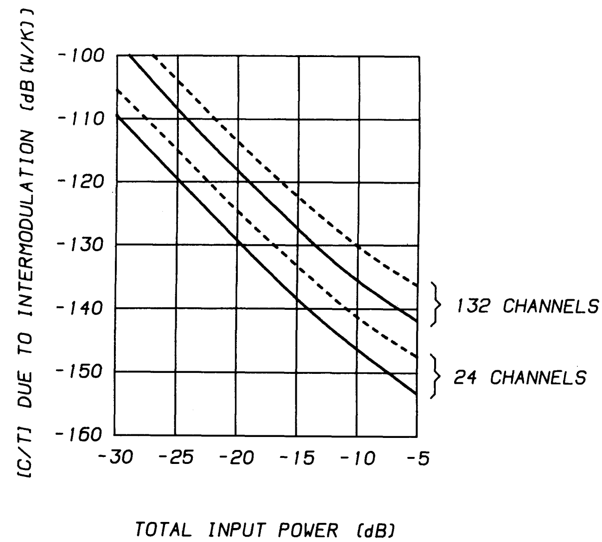


Fig. 12. Intermodulation with 10 carriers: (——) without AM–PM conversion; (-----) with AM–PM conversion. (Reprinted with permission from Reference 16).

show a deterioration of 3–6 dB due to AM–PM conversion with respect to the simple amplitude nonlinearity effects.

Neglecting the noise due to AM–PM conversion, we concentrate on the contribution due to the amplitude nonlinearity alone, for 4, 12, or more than 100 carriers of equal power.

The most important intermodulation products are those of the third order, which may be of the form $f_i \pm f_j \pm f_k$ or $2f_i \pm f_j$, and may fall in the useful band when the signs are properly selected; more precisely, only part of the products $f_i + f_j - f_k$ and $2f_i - f_j$ fall in the band occupied by the useful signals, thereby causing disturbances.

The number of products of type $2f_i - f_j$ falling on the mth carrier (where $1 \leq m \leq N$) is

$$P' = \left\lfloor \frac{N}{2} \right\rfloor + \lambda - 1 \tag{34}$$

where $\lfloor\ \rfloor$ = entire part of the operand

$$\lambda = \begin{cases} 1 & \text{if } m \text{ and } N \text{ are odd} \\ 0 & \text{otherwise} \end{cases}$$

The number of $f_i + f_j - f_k$ products falling on the mth carrier is[17]

$$P''(m) = P''(N - m + 1) = \sum_{i=1}^{m-1} \{N - i - |i - m| + \min(m - i - 1, i - 1)\} + \sum_{i=m+1}^{i_{\max}} (N - i - |i - m|) \tag{35}$$

where $|\ |$ = absolute value of the operand
$\min(a, b)$ = smaller of the values a or b
$i_{\max} = \lfloor (N + m - 1)/2 \rfloor$

If N is a multiple of 4, Eq. (34) simplifies to

$$P' = \frac{N}{2} - 1 \tag{34'}$$

whereas for a central carrier (i.e., $m = N/2$), $i_{\max} = 3N/4 - 1$ and (35) simplifies to

$$P''\left(\frac{N}{2}\right) = \frac{3}{8}N^2 - \frac{5}{4}N + 1 \tag{35'}$$

It is also easy to show that

$$\begin{aligned}
P''_{N/2-1} &= P''_{N/2} - 1 \\
P''_{N/2-2} &= P''_{N/2-1} - 2 \\
P''_{N/2-3} &= P''_{N/2-2} - 3 \\
&\vdots \\
P''_{N/2-k} &= P''_{N/2-k+1} - K \\
&\vdots \\
P''_1 &= P''_2 - (N/2 - 1)
\end{aligned}$$

Therefore, adding all these equations we obtain the number of $f_i + f_j - f_k$ products falling on the side carriers:

$$P''_1 = P''_N = P''_{N/2} - \sum_{i=1}^{N/2-1} i = P''_{N/2} - \frac{N}{4}\left(\frac{N}{2} - 1\right) \tag{35''}$$

These data are summarized in Table I, which also provides the total number of third-order products falling on a central carrier and on a side carrier, and the total number of third-order products falling on all N carriers, therefore causing disturbance. It is easily seen that the ratio between the number of third-order products falling on a central line and on a side line is 1.5.

Products of type $f_i + f_j - f_k$ are of higher level (see Fig. 8), and for large N they are more numerous than $2f_i - f_i$ products. Hence, for large N the band center carrier-to-intermodulation noise power ratio (C/I) is about 2 dB better than the band edge C/I.

Table II shows the situation which occurs for 4 or 12 carriers.

Table I. Number of Third-Order Intermodulation Products vs. Number of Carriers N

Number of products / Type of product	Total	If N is multiple of 4: Falling on a central carrier	Falling on a side carrier	Total in the useful bandwidth
$f_i + f_j - f_k$	$\frac{1}{2}N(N-1)(N-2)$	$\frac{3}{8}N^2 - \frac{5}{4}N + 1$	$\frac{1}{4}N^2 - N + 1$	$\frac{3}{8}N^3 - \frac{5}{4}N^2 + N - \sum_{i=1}^{N/2-1} i(i+1)$
$2f_i - f_j$	$N(N-1)$	$\frac{N}{2} - 1$	$\frac{N}{2} - 1$	$N\left(\frac{N}{2} - 1\right)$
Total	$\frac{1}{2}N^2(N-1)$	$\frac{3}{4}N\left(\frac{N}{2} - 1\right)$	$\frac{1}{2}N\left(\frac{N}{2} - 1\right)$	$\frac{3}{4}N^2\left(\frac{N}{2} - 1\right) - \sum_{i=1}^{N/2-1} i(i+1)$

Table II. Distribution of the Third-Order Intermodulation Products on the Various Carriers for $N = 4$ and $N = 12$

Type of product \ Carrier number	1	2	3	4	Total in the useful bandwidth	Grand total
$f_i + f_j - f_k$	1	2	2	1	6	12
$2f_i - f_j$	1	1	1	1	4	12
Total	2	3	3	2	10	24

Type of product \ Carrier number	1	2	3	4	5	6	7	8	9	10	11	12	Total in the useful bandwidth	Grand total
$f_i + f_j - f_k$	25	30	34	37	39	40	40	39	37	34	30	25	410	660
$2f_i - f_j$	5	5	5	5	5	5	5	5	5	5	5	5	60	132
Total	30	35	39	42	44	45	45	44	42	39	35	30	470	792

When the number of carriers is increased from 4 to 12, the number of $f_i + f_j - f_k$ products falling on one of the two central carriers increases by 20, whereas their individual level decreases by $(12/4)^3 = 27$ times if the output back-off is kept constant. The carrier power decreases in these conditions by three times; therefore, the C/I decreases by $10 \operatorname{Log}_{10} [(20/27) \cdot 3] = 3.4$ dB when the number of carriers is increased from 4 to 12 at constant output back-off. It can be similarly shown that C/I decreases by 1 dB when the number of carriers becomes very large.

Varying the output back-off for a fixed number of carriers, the carrier power is decreased proportionally to the BO, whereas the intermodulation noise power is decreased proportionally to the third power of the BO; therefore, C/I increases with the square of the BO.

The precise value of the C/I ratio depends on the satellite HPA input–output characteristic; in particular, if the characteristic (27) is assumed, it is extrapolated by computer calculations[16] that for a great number of carriers the worst C/I value obtained at band center when the TWTA works in the quasi-linear region is

$$\frac{C}{I} = 2.6 + 2\,\mathrm{BO} \quad \mathrm{dB} \tag{36}$$

From this expression the C/N_{0i} may be calculated from

$$\frac{C}{I} = \frac{C}{N_{0i}(0.9B/N)}$$

where N_{0i} = intermodulation noise power density
B = transponder bandwidth
N = number of active carriers in the transponder
0.9 = coefficient taking into account a 10% guardband

Taking logs gives

$$\frac{C}{I} = \frac{C}{N_{0i}} + 10\,\mathrm{Log}_{10} N - 10\,\mathrm{Log}_{10} B + 0.45$$

Letting $B = 36 \times 10^3$ kHz and substituting for C/I expression (36) one obtains

$$\frac{C}{N_{0i}} = 47.7 - 10\,\mathrm{Log}_{10} N + 2\,\mathrm{BO} \tag{37}$$

This value must be increased as shown in Fig. 11 for a small number of carriers.

A large number of carriers occurs in practice when a different carrier is used for the transmission of each telephone channel (single-channel-per-carrier systems, SCPC). In this case several hundred uniformly spaced carriers are amplified in the same satellite TWTA and the intermodulation noise becomes practically white. The C/N_{0i} value varies, however, with the frequency considered and may be up to 2 dB higher than the value provided by (37) for carriers located at the edge of the transponder band.

VIII. Spectrum Truncation

Most modulation systems are nonlinear and produce, as a consequence, the occupation of an infinite bandwidth (see Chapter 6, Section XII and Chapters 9 and 10). It is impossible, in practice, to devote a channel of infinite bandwidth to the transmission of a single carrier, and it is necessary to limit the channel bandwidth to allow an efficient use of the frequency resource. The optimum bandwidth occupation must result from the trade-off between the bandwidth efficiency (measured by the information quantity sent per unit bandwidth) and the signal distortion due to spectrum truncation. The elimination of spectral lines due to filtering is equivalent to the modulation of the infinite bandwidth carrier by the filtered-out lines of the spectrum. The signal distortion effects may be evaluated by following this approach.

For analog frequency modulation of a carrier by an FDM telephone signal, the distortion noise caused by spectrum truncation is called intermodulation noise, since it comes from the nonlinear interaction of the baseband channels. Intermodulation noise may be unweighted or weighted, and measured in dBm0 or dBm0*p* respectively, with a level which will be discussed in Chapter 9, Section V K.

For a carrier frequency-modulated by a television signal, the quasi-deterministic structure of the signal makes necessary the control of the peak distortion of some signal characteristics, as discussed in Section VII B of Chapter 9.

Finally, for digital transmissions, spectrum truncation effects can be completely eliminated. Because of the nature of digital transmissions, which require a decision at regular time intervals as to which symbol (out of a predefined alphabet) has been received, it is sufficient to control the distortion produced at the decision instants in order to eliminate spectrum truncation effects. Therefore, whereas the shape of the individual signal pulse is distorted by the limitation of the transmission bandwidth, so that the extension in time of each pulse becomes unlimited, an appropriate design of the filtering characteristics of the transmission channel can reduce to zero the instantaneous voltage value of each symbol corresponding to the decision instants of all other symbols. Under these conditions there is no intersymbol interference (ISI), and the bit error probability (BEP) is not degraded with respect to the infinite bandwidth case. The channel filtering design is discussed in Sections III and VII of Chapter 10.

IX. Intelligible Crosstalk

When a frequency-modulation transmitting chain is not ideal, a spurious amplitude modulation may be generated, such that the modulating signal information is carried both by FM and by AM. This deviation of the transmitting chain from the ideal is called the gain-slope phenomenon and is measured in dB/MHz, i.e., in dB of AM versus MHz of FM.

Suppose now that several FM carriers, each showing the spurious AM effect, transit through a TWTA producing AM–PM conversion. The spurious AM of a carrier is converted into phase modulation of all other carriers and is detected at the receiving end, giving rise to intelligible crosstalk. Since it is very difficult to reduce the TWTA AM–PM conversion factor, it is necessary in general to minimize the transmitting chain gain-slope in order to reduce the X-talk level.

The intelligible X-talk power level is referred to the test tone level by the equation

$$\mathrm{IXTR} = 20\,\mathrm{Log}_{10}\left\{\frac{180}{\pi}\cdot\frac{1}{K_p g f}\cdot\frac{\Delta f_{\mathrm{TT}}}{\Delta f_{\mathrm{rms}}}\cdot\frac{P_t}{P_i}\right\} \tag{38}$$

where IXTR = intelligible X-talk power ratio (dB)

K_p = AM–PM conversion coefficient (deg/dB)

g = gain-slope (dB/MHz)

f = frequency of X-talk (MHz)

Δf_{rms} = rms frequency deviation of the interfering carrier (Hz)

Δf_{TT} = test-tone deviation for the desired carrier (Hz)

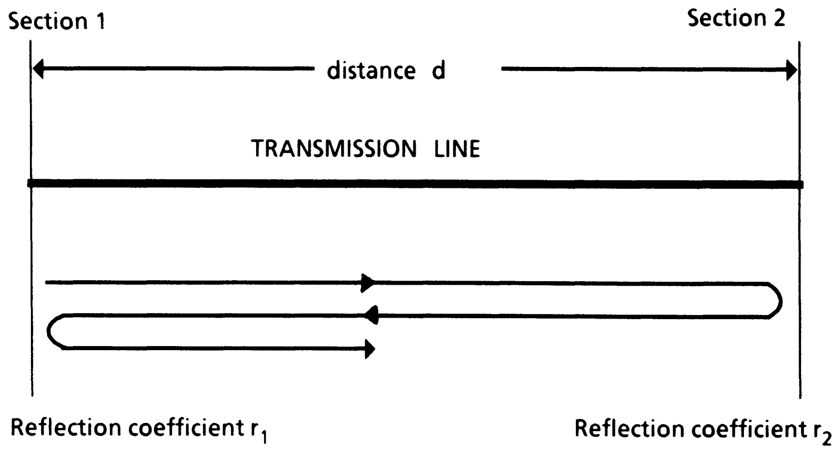
P_t = total power amplified in the TWTA (W)

P_i = power of the interfering carrier (W)

A value of IXTR typically specified for Intelsat earth stations is 58 dB.

X. Echo due to Equipment Mismatching

When a transmission line connects two devices with impedance mismatching at both ends of the connection, a double reflection takes place, such that a



τ = time difference between main signal and reflected signal $= \dfrac{2d}{\text{velocity}}$

α = attenuation constant of the transmission line

main signal $= A_c\, e^{j[\omega_0 t + \lambda(t)]}$

reflected signal $= r_1 r_2 e^{-2\alpha d} A_c\, e^{j[\omega_0 (t-\tau) + \lambda(t-\tau)]}$

echo coefficient $\rho = r_1 r_2 e^{-2\alpha d}$

amplitude ripple $= 20\mathrm{Log}_{10} \dfrac{1+\rho}{1-\rho}$ (dB peak-to-peak)

phase distortion $= \varphi(\omega) = \rho \sin(\tau\omega + \varphi_o)$

$\mathrm{GDD} = \tau(\omega) = \dfrac{d\varphi(\omega)}{d\omega} = \rho\tau\cos(\tau\omega + \varphi_o)$

GDD ripple $= 2\rho\tau$ (nsec peak-to-peak)

Fig. 13. Echo effect due to equipment mismatching and related ripple.

reflected signal is generated and added to the main signal (see Fig. 13). An amplitude maximum is obtained when the main and reflected signals are in phase i.e. for

$$\omega_0 \tau = 2\pi n$$

and the difference between two adjacent maxima in the frequency domain is

$$\Delta f = \frac{1}{\tau} \tag{39}$$

The part of the earth station where the mismatching problem is most severe is the connection of the HPA to the antenna, since it is difficult to obtain good wideband matching at high power levels.

In conclusion, equipment mismatching will cause an amplitude ripple and a phase ripple with equal frequencies. The effects of this ripple in FM multichannel telephony and in FM television will be discussed in Chapter 9. It is important to note that when the ripple frequency becomes smaller than the signal baseband the reflected signal is totally uncorrelated with the main signal. Thus, the noise produced by equipment mismatching must be computed as for normal interference.

XI. Interferences

Interference to a satellite communication system carrier may be originated from

- Terrestrial microwave links
- Another satellite system
- Adjacent channel carriers inside the same satellite system
- Cochannel carriers inside the same satellite system
- Long echo due to equipment mismatching in the earth station

Appendixes 1–3 will discuss the problem of keeping under control the interferences originated externally to the considered satellite communication system, while Chapters 9 and 10 will deal with the effects of interference on analog and digital transmission systems respectively. Usually one is successful, through appropriate system coordination and system design, in keeping small the fraction of overall baseband noise due to interference. This is also due to the absence of intentional interference (also called jamming) in civilian satellite systems. In some cases, however, one may be forced to withstand a severe interference environment and to use, as a consequence, an antijamming technique such as spread-spectrum modulation. An interesting example of this type is discussed in Section IV B of Chapter 14.

XII. Propagation Delay and Echo

A. General

Due to the finite speed of propagation of electromagnetic signals, the transport of information from one link extreme to the other requires a time different from zero, which is called propagation delay. This delay, generally negligible in terrestrial links, is significant in space links implemented by geostationary satellites. Since 35,786 km is the geostationary altitude, about $\frac{1}{4}$ s is needed to deliver the transmitted information to the receiving station, while about $\frac{1}{2}$ s is required to get the answer (if any). The effects of propagation delay may have different importance in different services, as discussed in Section XII B and in Chapter 13. Section XII C deals instead with the echo phenomenon, which is particularly important for telephony and requires the use of dedicated equipment such as echo suppressors or echo cancelers.

B. The Effects of Propagation Delay

In telephony, if the echo problem is perfectly solved, only the propagation delay will impair the conversation quality. However, experience shows that users generally consider satisfactory the conversation quality obtained in these conditions when the circuit is implemented by a single hop (500-ms round-trip delay), whereas a two-hop circuit (1-s round-trip delay) is considered unacceptable.[18] Extensive field trials have been performed in the U.S. domestic system operated by AT&T, showing unacceptable quality for circuits served by echo suppressors (impairment from both propagation delay and echo) and a quality practically equivalent to that of terrestrial circuits when echo cancelers are used (impairment from propagation delay only).[19] Telephony may therefore withstand a single-hop delay, whereas multiple hops are unacceptable. The situation is completely different for data transmission, where multiple hops may be acceptable, but one has to modify terrestrial network protocols also in the case of a single hop to obtain an acceptable channel throughput (see Section VIII D of Chapter 13).

As to the transmission of telephone signaling, only the single-hop case is relevant (due to the conversation requirements), and the related data transmission protocols show peculiar features (see Chapter 13, Sections VIII A and B).

C. The Echo Phenomenon

Figure 14 shows a simplified representation of a telephone circuit connecting user to user. Generally the user's loop uses only two wires, for economic reasons,

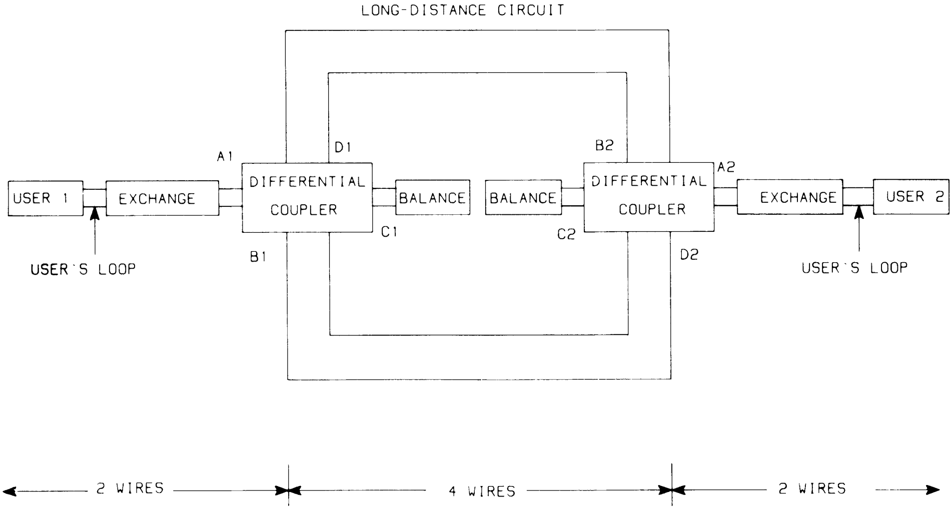


Fig. 14. Simplified block diagram of end-to-end telephone circuit.

whereas the long-distance section of the circuit, used to connect the terminal exchanges, uses four wires. The echo phenomenon exists due to verification of two conditions:

1. Nonzero propagation delay
2. Implementation of one part of the circuit by two wires and of another part of the circuit by four wires

The transition from two to four wires, or *vice versa,* requires the use of a differential coupler, which is the source of echo signals in case of impedance mismatch.

If perfect balance exists, the signal entering port A should reappear entirely at port D, whereas the signal entering port B should reappear entirely at port A. However, if impedances at ports A and C are not perfectly balanced, part of the signal entering port B will reach port D and be sent back to the signal originator, who will perceive it as an echo. In practice it is not possible to match the differential coupler to all possible user's loops, so a mean value of impedance is used, and an echo will always be present. The echo, however, will produce a significant impairment only when the propagation delay is large. For distances shorter than 2500 km the signal originator will not be able to detect the existence of an echo, whereas for geostationary satellites links it will be necessary to use special equipment (echo suppressors or echo cancelers) to recover acceptable conversation quality.

References

[1] J. B. Johnson, "Thermal agitation of electricity in conductors," *Phys. Rev.*, vol. 32, pp. 97–109, 1928.

[2] H. Nyquist, "Thermal agitation of electrical charge in conductors," *Phys. Rev.*, vol. 32, pp. 110–113, (1928).

[3] IEEE Standard 161-1971 (reaffirmed 1980), "Standard definitions on electron tubes," 1980.

[4] W. B. Davenport and W. L. Root, *An Introduction to the Theory of Random Signals and Noise,* New York: McGraw-Hill, 1958, pp. 81–82.

[5] S. O. Rice, "Mathematical analysis of random noise," *Bell Syst. Tech. J.* **23**, 282–332 (1944) and **24**, 46–156 (1945).

[6] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.* **27**, 446–472 (1948).

[7] CCITT Recommendation G. 711, "Pulse code modulation (PCM) of voice frequencies," *Red Book,* Vol. III, Fasc. III.3, Geneva, 1985.

[8] B. Smith, "Instantaneous companding of quantized signals," *Bell Syst. Tech. J.*, May 1957, vol. 36, pp. 653–709.

[9] R. F. Purton, "Survey of telephone speech-signal statistics and their significance in the choice of a PCM companding law," *Proc. IEE* **109**, 60–66 (1962).

[10] K. W. Cattermole, "Discussion on the above paper by Purton," *Proc. IEE* **109**, 485–487 (1962).

[11] Bell Laboratories, *Transmission Systems for Communications,* 1982, pp. 621–623.

[12] M. Imbeaux, CNET, private communication.

[13] A. Berman and E. Podraczky, "Experimental determination of intermodulation distortion produced in a wideband communication repeater," *IEEE Int. Convention Record* **15** (Part 2), 69–88 (1967).

[14] J. C. Fuenzalida, O. Shimbo and W. L. Cook, "Time-domain analysis of intermodulation effects caused by non-linear amplifiers," *Comsat Tech.* **3**, 89–141 (1973).

[15] R. J. Westcott, "Investigation of multiple FM/FDM carriers through a satellite TWT operating near to saturation," *Proc. IEE* **144**, 726–740 (1967).
[16] CCIR, *Handbook on Satellite Communications (Fixed-Satellite Service)*, Geneva, 1985.
[17] G. Zanotti, *Number of 3rd-Order Intermodulation Products Generated in a Nonlinear Amplifier*, Telespazio internal report, Feb. 1988.
[18] CCITT Recommendation G.114, "Mean one-way propagation time," *Red Book*, Vol. III, Fasc. III.1, Geneva, 1985.
[19] T. H. Curtis *et al.*, "Use of a digital echo canceller in the AT&T Domsat intertoll network," *Fifth Int. Conf. on Digital Satellite Communications*, Genoa, March 1981, pp. 227–234.