THE EFFECTS OF SYSTEM NONLINEARITIES ON SYSTEM NOISE STATISTICS

Follow-on Contract to NAS 9-8166

20 October 1971

Prepared for
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
MANNED SPACECRAFT CENTER
HOUSTON, TEXAS

Prepared by
Electronic Systems Engineering Department
Electronics Systems Laboratory

17618-H223-R0-00
Page intentionally left blank
ABSTRACT

This report studies the effects of nonlinearities in a baseline communications system on the system noise amplitude statistics. So that a meaningful identification of system nonlinearities can be made, the baseline system is assumed to transmit a single biphase-modulated signal through a relay satellite to the receiving equipment. The significant nonlinearities thus identified include square-law or product devices (e.g., in the carrier reference recovery loops in the receivers), bandpass limiters, and traveling wave tube amplifiers (TWTA).

When considered alone, a nonlinear device can be expected to produce a non-gaussian output from a gaussian input. However, if the nonlinearity is followed by a linear filter, the resultant amplitude statistics can be restored to nearly gaussian form, particularly if the filter has a sufficiently narrow bandwidth. In the case of angle-modulated signals, the filter should have linear phase characteristics in addition to linearity in the usual sense of linear transformations. If the filters in the communication system have these linear characteristics and their bandwidths are narrow enough, the noise will remain essentially gaussian throughout the system.

Thus, if the system is sufficiently narrowband, the nonlinearities in the system will have a negligible effect on the amplitude statistics of the noise. The primary effects of the system nonlinearities, then, are intermodulation products and altered power spectra.
Page intentionally left blank
**CONTENTS**

1. INTRODUCTION ............................................. 1-1  
   1.1 Definition of a Baseline Communication System ........ 1-1  
   1.2 Typical Nonlinearity ............................... 1-3  
   1.3 Content of Study .................................. 1-6  

2. SUMMARY AND CONCLUSIONS ................................. 2-1  

3. IDENTIFICATION OF SYSTEM NONLINEARITIES ............... 3-1  
   3.1 Block Diagram Description .......................... 3-1  
      3.1.1 Transmitting Section ......................... 3-3  
      3.1.2 Receiving Section ............................ 3-3  
      3.1.3 Reference Frequencies Section ............... 3-4  
      3.1.4 Relay Satellite Configurations .............. 3-5  
   3.2 System Nonlinearities ............................. 3-6  
      3.2.1 Receiving Station ............................ 3-7  
      3.2.2 Transmission Medium .......................... 3-10  
      3.2.3 Relay Satellite .............................. 3-11  
      3.2.4 Transmitting station ......................... 3-12  

4. NOISE CONSIDERATIONS .................................... 4-1  
   4.1 General Considerations ............................ 4-1  
   4.2 Examples ........................................... 4-2  
      4.2.1 Square-Law Nonlinearity ....................... 4-3  
      4.2.2 Bandpass Limiter ............................. 4-6  
      4.2.3 Nth-Order Nonlinearity ....................... 4-10  

REFERENCES ................................................. R-1
<table>
<thead>
<tr>
<th>Illustration</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Baseline Communication System</td>
<td>1-2</td>
</tr>
<tr>
<td>1-2</td>
<td>Square-Law Statistics</td>
<td>1-4</td>
</tr>
<tr>
<td>1-3</td>
<td>Square-Law Power Spectra</td>
<td>1-5</td>
</tr>
<tr>
<td>3-1</td>
<td>Transmitting/Relay/Receiving Station Composite Block Diagram</td>
<td>3-2</td>
</tr>
<tr>
<td>4-1</td>
<td>Filtered Square-Law Statistics</td>
<td>4-4</td>
</tr>
<tr>
<td>4-2</td>
<td>Limiter Statistics</td>
<td>4-7</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

The predictions of improved performance for a coded digital communications system are generally based on the assumption that only gaussian noise is present in the system. This assumption may not be valid if the signal encounters significant nonlinear processing in the communication link. System nonlinearities can produce interference which is non-gaussian in character due to 1) nonlinear processing of thermal noise, 2) intermodulation of data, voice, and synchronizing signals, 3) intermodulation between the antenna tracking and communications functions, or 4) noisy recovery of the bi-phase demodulator carrier reference signal.

1.1 DEFINITION OF A BASELINE COMMUNICATION SYSTEM

The identification of specific points in the overall communications link where nonlinearities are present requires that a baseline communication system be defined, such as in Figure 1-1. In this figure the baseline communication system is divided into three segments: Transmitting station, relay satellite, and receiving station. In general, the communication system will be two-way, but the return link would be identical in form to that shown and thus add no new information concerning system nonlinearities. The link shown may be from ground to spacecraft, spacecraft to ground, or spacecraft to spacecraft. A relay configuration has been defined because it poses the largest number of problem areas.

The transmitting station contains the source of the data to be communicated, a modulator, and a transmitter. For the purposes of this study, direct bi-phase modulation of the data onto an S-band or other carrier is assumed.

A fairly general configuration has been chosen for the relay satellite. For the frequency translation configuration, the incoming spectrum is simply shifted to a new carrier frequency and retransmitted. This is a versatile arrangement capable of accommodating a wide variety of...
Figure 1-1. Baseline Communication System

RECEIVING STATION

DATA PROCESSOR

TRANSMITTER

DATA SOURCE

TRANSMITTING STATION

DATA PROCESSOR

TRANSMITTER

TRANSMITTER

RELAY SATELLITE

* FREQUENCY TRANSLATOR OR DEMODULATOR/MODULATOR OR DEMODULATOR/BIT SYNCHRONIZER/MODULATOR

TRANSMITTER

TRANSMITTER

* FREQUENCY TRANSLATOR OR DEMODULATOR/MODULATOR OR DEMODULATOR/BIT SYNCHRONIZER/MODULATOR

RELAY SATELLITE

TRANSMITTER

TRANSMITTER

TRANSMITTER
spectral formats. But it offers little or nothing in the way of signal-to-noise ratio (SNR) improvement because most of the received noise (that in the spectral passband) is retransmitted. A demodulation/remodulation configuration requires additional complexity for the same spectral versatility and lack of SNR improvement but offers the possibility of maintaining the transmitted power independent of the input power such as through the use of automatic gain control. The demodulation/bit reconstruction/remodulation configuration involves even more complexity and requires a specific knowledge of the incoming spectrum. However, a significant improvement in SNR can be achieved, limited primarily by the (noisy) recovery of the received carrier and the resultant bit error rate (BER) of the reconstructed data.

The receiving station contains a receiver, a (bi-phase) demodulator, a bit synchronizer or other bit reconstruction device, and a data processor. For the purposes of this study, the output of the bit synchronizer is assumed to be an estimate (containing errors) of the original data output from the data source in the transmitting station.

1.2 TYPICAL NONLINEARITY

As an example of how a nonlinearity can alter the system noise statistics, consider the effect of a square-law nonlinearity on a narrowband gaussian random variable. As shown in Figures 1-2 and 1-3, the resultant output probability density is chi-squared, while the originally rectangular power spectrum becomes triangular (from the convolution of the input spectrum with itself), plus an impulse at zero frequency (due to the nonzero mean of the output process). A more complete discussion of this example is presented in Section 4.

The important consideration here is that system nonlinearities can cause non-gaussian statistics, which can render relatively meaningless many of the results of communication theory. For example, consider how the complementary error function definition of the bit synchronizer bit error rate (BER) performance versus input signal-to-noise ratio (SNR) would have to be reexamined if the input noise were chi-squared or exponential instead of gaussian.
Square-Law Nonlinearity:

\[ x(t) \rightarrow a(\cdot)^2 \rightarrow y(t) \]

\[ y = ax^2 \]

Input Probability Density Function: (Gaussian)

\[ p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \]

Resultant Output Probability Density: (Chi-Squared)

\[ p_Y(y) = \begin{cases} 
\frac{1}{\sqrt{2\pi\sigma y}} e^{-\frac{y}{2a\sigma^2}}, & y \geq 0 \\
0, & \text{Otherwise} 
\end{cases} \]

Figure 1-2. Square-Law Statistics
Square-Law Nonlinearity:

\[ x(t) \rightarrow a(\cdot)^2 \rightarrow y(t) \]

\[ y = ax^2 \]

Input Power Spectrum: (Narrowband Rectangular)

\[ S_x(f) = \begin{cases} 
A, & f_0 - B/2 < |f| < f_0 + B/2 \\
0, & \text{Otherwise}
\end{cases} \]

Resultant Output Power Spectrum: (Narrowband Triangular, Plus Impulse)

\[ S_y(f) = \begin{cases} 
4a^2 A^2 B^2 \delta(f), & f = 0 \\
4a^2 A^2 (B - |f|), & 0 < |f| < B \\
2a^2 A^2 (B - |f - 2f_0|), & 2f_0 - B < |f| < 2f_0 + B \\
0, & \text{Otherwise}
\end{cases} \]

Figure 1-3. Square-Law Power Spectra
1.3 CONTENT OF STUDY

Section 1 has introduced the problem of system nonlinearities possibly altering the statistics of the system noise and defined the baseline communications system from which specific nonlinearities will be identified. The overall summary and conclusions of this study are presented in Section 2. A basic system block diagram is defined in Section 3, and the associated nonlinearities are identified and discussed. Section 4 discusses the effects on the system noise statistics of the major nonlinearities identified in Section 3.

The list of references at the end of this report is rather extensive. It covers primarily the more recent work in the subject area. Many pertinent papers (particularly the older ones) have been omitted, since their results are used in the stated references and would thereby contribute to a longer list without materially adding to depth of coverage provided. Thus, references to other literature (in some cases a significant quantity) have been relegated to the bibliographies provided by the stated references.
At carrier frequencies of ≈ 100 to 500 MHz and above, the predominant noise sources are receiver front-end and galactic, both of which are gaussian. Such channel effects as atmospheric absorption and polarization losses, fading, and multipath contribute non-gaussian (predominantly Rayleigh or Rician) components which, for the purposes of this study, are assumed to be signal perturbations and hence do not affect the gaussian nature of the system noise.

The most significant system nonlinearities occur in the carrier reference recovery loops (where squaring or product operations are introduced), in limiters, and in traveling wave tube amplifiers (TWTA). These nonlinearities introduce signal distortion in the form of intermodulation products and, together with the system filters, alter the spectral shape of the system noise. A nonlinearity by itself can produce a non-gaussian output from a gaussian input; for example, the probability density of the output of a square-law detector is the chi-squared distribution when the input has a gaussian probability density. However, if the nonlinearity is followed by a sufficiently narrow linear filter, the output of this filter is (nearly) gaussian, but with a different mean and variance from those of the original gaussian input. If the original input is non-gaussian, most likely the output will remain non-gaussian, modified by both the nonlinearity and the subsequent filter.

If the filters in the communication system are sufficiently narrowband, so that the averaging times are sufficiently large compared to the correlation time(s) of the gaussian input(s), the noise will remain (nearly) gaussian throughout the system. Thus, for all practical purposes, the system nonlinearities will have negligible effect on the system noise statistics, provided the system is sufficiently narrowband. The primary effects of the system nonlinearities are intermodulation products and the resultant modified power spectra; while the noise spectral density may not remain "white," being shaped by filters and the nonlinearities, the associated probability density remains (nearly) gaussian.
3. IDENTIFICATION OF SYSTEM NONLINEARITIES

The baseline communications system defined in Figure 1-1 contains many elements that are common among the transmitting station, relay satellite, and receiving station. Consequently, the task of identifying system nonlinearities can be simplified somewhat by combining the elements of the three system segments into a single composite block diagram, such as shown in Figure 3-1. In this figure the transmitting section is shown on the left, the receiving section on the right, and the reference frequencies section in the center. Those elements enclosed by dashed boxes may or may not be included in a particular implementation, and optional connections are shown dashed. While this diagram might not correspond to some actual existing or proposed system configurations, it nevertheless contains elements found in most modern communications systems and embodies a significant number of potentially troublesome nonlinearities.

3.1 BLOCK DIAGRAM DESCRIPTION

The basic assumption in Figure 3-1 is that the data is bi-phase modulated directly onto an S-band or the carrier.

The transmitting section consists of a data source, a bi-phase modulator followed by a bandpass filter (BPF) or limiter (BPL), a mixer followed by a bandpass filter, a traveling wave tube power amplifier (TWTA), a diplexer or other power splitting device, and an antenna assembly. The receiving section consists of an antenna assembly (which may or may not be shared with the transmitting function), a diplexer or other directional power device, possibly a parametric amplifier (particularly in ground stations), a two-stage intermediate frequency (IF) section, a bi-phase demodulator, a bit synchronizer or other bit reconstruction device, and a data processor or end user of the communicated data. The reference frequencies section consists of possibly a master oscillator (in the originating station to which the rest of the system is frequency and/or phase locked) or auxiliary oscillator (for use in slave stations until the receiving system is in lock), a carrier recovery phase lock loop (in the receiving stations), and frequency multiplier or synthesizer chains from which the transmitting and receiving intermediate frequency (IF) and radio frequency (RF) references are obtained.
Figure 3-1. Transmitting/Relay/Receiving Station
Composite Block Diagram
3.1.1 Transmitting Section

In the transmitting section, the output of the data source is assumed to be a single binary bit stream which may be a composite of multiplexed and coded commands, telemetry, ranging code, voice, television, and/or other digital signals. The binary output of the data source is bi-phase modulated onto an IF carrier which may be either sinusoidal or, if limited, a square wave. The output of the bi-phase modulator is bandpass filtered (or, in the case of square-wave modulation, bandpass limited) to control the spectral occupancy of the modulated signal and to delete any undesired components generated by the modulation process. The modulation is performed at IF as a matter of practical expediency, since it is easier to build the circuits and control the process at IF as compared to RF.

The filtered IF output of the bi-phase modulator is mixed with an RF reference to translate or shift the modulated spectrum to the desired transmitting carrier frequency. The bandpass filter following the mixer extracts the desired mixer product term and shapes the resultant RF spectrum. Note that a mixer is required rather than a frequency multiplier because bi-phase modulation is assumed; multiplication by an even factor (e.g., x 30) would remove all of the phase modulation in addition to providing the desired RF frequency.

The RF mixing is performed at relatively low signal levels as a practical matter because of design and operation considerations. As a result, the RF signal must be amplified to the desired power level for final radiation from the antenna. While transistor power amplifiers are becoming feasible at S-band and above, much of the current technology relies on traveling wave tube amplifiers (TWTA) for this final power amplification. The amplified RF signal is coupled to the transmitting antenna through a diplexer or other isolation device.

3.1.2 Receiving Section

In the receiving section of Figure 3-1, the received signal is coupled from the antenna through a diplexer or other directional isolating device, possibly through a parametric amplifier (for boosting signal gain and improving signal-to-noise ratio, SNR), to the IF section. Two stages of IF mixing, amplification, and filtering are shown to indicate that the
incoming spectrum is usually translated from RF to IF in more than one step. In many cases the signal level in the final IF output is maintained fairly constant through the use of automatic gain control (AGC) in one or more of the IF stages.

The final IF signal is translated to baseband in a bi-phase demodulator whose output after filtering is a noisy version of the binary data stream originally output from the data source in the transmitting station. This noisy signal is reconstructed in a bit synchronizer or other device, whose output is a noise-free (but not necessarily error-free) version of the original bit stream. The bit synchronizer output is then processed as required.

3.1.3 Reference Frequencies Section

The reference frequencies section provides the necessary frequency and/or phase references for the transmitting and receiving sections. In the master or controlling transmitting station, all of the transmit references are derived from a (usually precision) master oscillator. In other stations, including the relay satellite, an auxiliary oscillator may or may not be used to provide distant stations a signal upon which to acquire, track and lock while the associated receiving section is becoming locked; for non-coherent communications such an oscillator might be used continuously as the source of the transmit references.

The receive references and in most cases the transmit references in all but the master station are derived from the output of a carrier recovery phase lock loop (PLL) operating on the final IF output of the receiver section. If AGC is not employed in the IF sections, then a limiter may precede the loop to provide a reduced signal dynamic range, enabling the loop to acquire and track more effectively.

Because the data has been bi-phase modulated, the carrier recovery PLL will most likely be some sort of Costas loop or squaring loop. As noted in Reference 1, no carrier component is present in the received signal, so the necessary reference signal must be derived from the sideband information in the received signal. The resultant noisy reference is used to provide a phase reference for the phase detector in the bi-phase demodulator in the receiving section and for the transmit and receive references.
The transmit and receive references are shown as separate lines in Figure 3-1 to indicate that the two sets of frequencies are not necessarily the same and do not necessarily share a common basic source. Since the master or auxiliary oscillator on the transmit side and the carrier recovery voltage controlled oscillator (VCO, in the PLL) on the receive side are not necessarily implemented to produce the required IF frequencies directly, frequency multiplier chains or frequency synthesizers and filters are shown between the oscillator outputs and the bi-phase modulator and second IF mixer and between the IF and RF sections. While frequency multipliers are in common use in current technology, frequency synthesizers offer two significant advantages: they can readily provide noninteger "multiplication" factors and can be programmed (by command) to provide a variety of different factors.

3.1.4 Relay Satellite Configurations

Figure 3-1 also shows possible interconnections for the three relay satellite configurations defined in Figure 1-1: frequency translation, demodulation/remodulation, and demodulation/bit reconstruction/remodulation. The frequency translation configuration can be implemented by connecting the final IF output of the receiving section to the mixer input in the transmitting section. In this case, the data source, bi-phase modulator, and bandpass filter would be deleted from the transmitting section, as well as the bi-phase demodulator, bit synchronizer, and data processor in the receiving section. The frequency translation approach is in common use in current technology because it allows the use of any of a number of uplink/downlink spectral formats. Subject to the constraints of carrier frequencies and spectral occupancies, any combination of subcarriers and modulation formats can be handled.

The demodulation/remodulation configuration can be realized by deleting the data source, bit synchronizer, and data processor and by connecting the output of the demodulator to the input of the modulator, possibly inserting filtering or other signal processing in the connection. If the spectral versatility of frequency translation is to be retained, then a linear phase demodulator and phase modulator are required. This permits the inclusion of baseband filtering, which is somewhat easier to implement than at IF. In this manner it is possible to provide an output
power that is essentially independent of the input power (for angle modulated carriers). If a bi-phase demodulator is used, only a bi-phase modulated spectrum can be processed, and a linear phase modulator will be required if binary reconstruction is not provided.

A demodulation/bit reconstruction/remodulation configuration can be obtained by deleting the data source and data processor and by connecting the output of the bit synchronizer to the input of the bi-phase modulator. This approach requires the most complexity and places the most restrictions on the allowable uplink/downlink format. But it offers the possibly significant advantages of being able to reconstruct the data at a lower signal-to-noise ratio (SNR) than would otherwise be available at the receiving station and to transmit a "clean" signal perturbed only by the noisy recovered carrier reference (and, of course, by the bit error rate of the bit synchronizer). Under the proper conditions, the resultant net bit error rate at the receiving station will be much lower than if one of the other two relay techniques or no relay at all were used.

3.2 SYSTEM NONLINEARITIES

Except for the data source and data processor in Figure 3-1, which for the purposes of this study are assumed to be noise-free, every element in the communications system presents candidate nonlinearities and/or perturbations. The identification and discussion of system nonlinearities will start with the bit synchronizer in the receiving station and will work backward through the system from the receiver through the transmission medium to the relay satellite and then to the transmitter. This order will put the presentation in proper perspective, since in many respects the final stages in the overall communication system are the most critical.

The following discussions will be fairly short and qualitative in nature, since in Section 4 will be presented the fact that the noise remains (nearly) gaussian in a sufficiently narrowband system containing nonlinearities. The most severe system nonlinearities, bandpass limiters, square law or product devices in the carrier recovery loop, and traveling wave tube amplifiers (TWTA), are discussed in more detail in Section 4 to illustrate how the noise remains gaussian.
3.2.1 Receiving Station

The discussion of the receiving station will begin with the bit synchronizer, followed by the bi-phase demodulator, the carrier recovery loop, the receive and transmit reference frequency chains, the RF-IF section, and the antenna and first amplification section.

3.2.1.1 Bit Synchronizer

As noted in Reference 2, bit synchronizer bit error rate (BER) performance is degraded primarily by prefiltering of the input signal to the bit synchronizer, by timing or synchronization errors (both internal and external to the bit synchronizer), and by baseline offsets (e.g., non-zero average input signal, both internal and external). Bit synchronizers usually employ integrate and dump, zero threshold matched filter bit detectors (which are "matched" for rectangular pulses). Prior filtering can cause the input signal to be distorted due to bandwidth restriction, which causes the filter to no longer be matched, as well as intersymbol interference. Prefiltering also shapes the input noise spectrum but, being a linear process, not the noise statistics.

Within the bit synchronizer itself imperfect (nonlinear) integration characteristics will also cause the filter to be mismatched, but this is largely a signal "perturbation" rather than a noise effect. A potentially more serious problem is in the reconstruction of the bit synchronization (phase) reference which controls the integrate and dump functions. This noisy phase reference contributes to timing errors which in turn introduce or accentuate intersymbol interference. As will be considered in the discussion of the carrier recovery loop, the phase error process can be considered gaussian if the signal-to-noise ratio (SNR) in the bit synchronization phase lock loop is high enough or, equivalently, if the loop bandwidth is sufficiently narrow. Consequently, the timing error problem is also a signal rather than a noise perturbation.

Finally, if the incoming data is on a split-phase (SP) or other phase-shift-keyed (PSK) subcarrier, the bit synchronizer will contain a decommutator for converting the input signal to a non-return-to-zero (NRZ) format prior to matched filter detection. If the decommutator has a spurious or nonlinear
response, the resultant NRZ signal can be seriously distorted, contributing to a degraded BER. However, except for very low input SNR's this will also be predominantly a signal effect.

3.2.1.2 Bi-Phase Demodulator

Bi-phase demodulation is essentially a mixing operation and can have spurious and nonlinear responses. As will be noted in the discussion of mixers, spurious and nonlinear responses can be controlled to some degree by the proper choice of the waveform and drive level of the reference signal. Reference 1 shows that the filtered output of an (ideal) demodulator with a noisy input and reference signals can be expressed as a perturbed signal component plus narrowband gaussian noise.

3.2.1.3 Carrier Recovery Loop (Including Bandpass Limiter)

Associated with the carrier recovery loop are significant nonlinearities contributed by the bandpass limiter and the square-law (in the squaring loop) or product (in the Costas loop) devices used to remove the data modulation from the loop error signal. These nonlinearities are major when compared to other effects and are treated separately in Section 4. Other contributions within the loop are spurious and nonlinear mixer responses in the loop phase detector(s), nonlinear phase response in the loop filter, and spurious frequency and control-voltage-to-output-frequency nonlinearity in the loop VCO.

Mixer response is discussed elsewhere. Nonlinear phase response can cause the loop filter to be dispersive, since it would not treat all frequencies the same. However, this filter is part of a closed-loop feedback control system whose function it is to reduce the (phase) tracking error to zero (see chapters 3 and 4 of Reference 3). The filter is thus chosen to provide the desired closed-loop transfer function, and the only problems that arise are when the phase response differs significantly from that intended. Another way of considering this filter is that it is very narrowband compared to the rest of the system and has an output signal (the VCO control voltage) that is being servoed to zero. In other words, the filter mainly controls the dynamic behavior of the loop and has little or no affect on the statistics of the noise in the loop. Similar comments apply to the possible VCO control nonlinearity. The statement that the phase
detector, loop filter, and VCO do not appreciably affect the gaussian character of the noise within the loop is supported by the observation in Reference 1 that the probability density for the loop phase error resembles a gaussian function for large loop SNR's or, equivalently, narrow loop bandwidths.

3.2.1.4 Reference Frequency Chains

The frequency multipliers or synthesizers in Figure 3-1 can have spurious and nonlinear responses. Usually these circuits are very carefully designed so that the stability of the resultant RF and IF references is directly related to that of the base reference(s). As is the loop filter in the carrier recovery loop, the bandpass filters used in the reference frequency chains can be expected to have a negligible degrading effect on system performance.

3.2.1.5 RF-IF Mixers, Amplifiers, and Filters

In addition to possible spurious and nonlinear responses, the RF and IF mixers, being multiplying devices, will have noise conversion products, as well as signal/noise and signal/signal products. Some mixers also have a noise figure that is nonlinear with input level (e.g., due to a nonlinear gain distribution with input level). The conversion products can be controlled to a great extent by the proper selection of the mixing reference frequency and by filtering. As noted in Reference 4, the spurious response can be controlled somewhat by the choice of waveform and drive level of the reference signal.

The primary problems associated with IF amplifiers are saturation effects related to the dynamic range over which they will operate. This is particularly true when AGC is employed. However, these units are usually carefully designed to avoid saturation and hence are quite linear in the frequency range of interest.

Filtering is usually considered as a linear process which does not affect the gaussian nature of the system noise; it merely shapes the noise spectrum. However, as noted above, if the filter has a significant nonlinear phase response, it can be dispersive, and considerable problems may be introduced in an angle-modulated system. As a result, the in-line bandpass filters are usually designed to have reasonably linear phase characteristics.
3.2.1.6 Antenna and First Amplification

Figure 3-1 shows the possibility of a parametric amplifier preceding the first mixer. Whether or not such amplification is provided, the signal at the output of the antenna assembly is at its weakest level, so the initial stages of amplification in the receiver can contribute significant noise into the system. As noted in chapter 5 of Reference 5 and in Reference 6, the primary noise sources are thermal noise (due to resistors, having a "white" spectrum) and shot noise (due to transistors or tubes, having a $1/f$ spectrum), both of which have gaussian statistics and are additive.* Furthermore, these references also state that the noise generated in the early stages of an amplifier contributes most to the noise appearing at the output, especially if the amplifier has a high gain; this also applies to amplifiers in cascade.

Thus, a significant source of system noise is thermal, generated in the receiver front-end. A common practice in current technology is to use low noise figure (low noise temperature) amplifiers and mixers in the receiver front-end to reduce this noise contribution as much as possible. Cooled or uncooled parametric amplifiers are sometimes used for this purpose (particularly in ground stations), which also provide some improvement in signal-to-noise ratio due to their "pumping" action.

The antenna circuitry is usually quite linear, as are the initial stages of amplification, so the noise statistics in this region can be expected to be gaussian.

3.2.2 Transmission Medium

The transmission medium itself presents a wealth of noise sources and nonlinearities. Within the atmosphere there are losses due to weather (Reference 7) and absorption in the ionosphere (References 8 and 9), as well as polarization effects and fading due to multipath (see References 10 through 17).** And then there is (unintentional) interference from ground-

---

* Note, however, that the thermal noise will predominate at the frequencies of interest, the shot noise being negligible.

** Fading is usually defined in terms of Rayleigh or Rician statistics.
based, and possibly spaceborne, transmitters. For the most part, these effects can be considered to be signal perturbing rather than noise contributing.

The foregoing effects will be experienced by ground stations and by spaceborne stations whose antennas are directed primarily at the earth. For spaceborne antennas looking away from the earth, the predominant interference will be galactic or cosmic noise. Since this noise is contributed by a large number of independent sources, by the central limit theorem it will be gaussian; as shown in References 18 and 19, this noise has a $1/f$ spectral distribution.

3.2.3 Relay Satellite

A relay satellite will contain most of the elements found in the transmitting and receiving stations. The most significant nonlinearity in the relay is the traveling wave tube amplifier (TWTA), as noted in References 20 and 21. These amplifiers have a nonlinear input-output power characteristic and a nonlinear phase shift with input power and frequency. Deviations from linear phase vs. frequency can be reduced with careful design. When the amplifier is operated near saturation, the effects of nonlinear phase shift with input power are negligible compared to the limiting effects that occur in this region. However, as the TWTA is backed off from saturation, the effects of nonlinear phase shift (primarily AM-PM conversion) become significant and, at low drive levels, may become the dominant nonlinearity.

By far the most significant problem associated with the nonlinear input-output power relationship is the introduction of intermodulation (IM) distortion when more than one carrier or an amplitude-modulated (AM) carrier is amplified by the TWTA (see References 20 through 28). IM products set an upper bound on the quality of the relayed data, independent of thermal noise considerations. The highest-level IM products will be those of third-order and fifth-order. Since the system noise can be considered to be the result of many closely-spaced, independent carriers, there will be a considerable amount of intermodulation due to noise alone.
Note carefully, however, that multi-carrier intermodulation is a signal perturbation and therefore does not affect the noise statistics. Also, intermodulation is primarily a spectrum-altering effect, so the noise amplitude statistics are not affected even due to noise intermodulation.

Finally, the effects of intermodulation can be reduced by avoiding TWT saturation (or by going from a "hard" to a "soft" limiter), by increasing the spacing between adjacent signals, and by providing nonuniform spacing. An alternative approach would be to use encoding to produce highly random bit streams having few prominent spectral components and hence very low IM product levels. Also, the use of error-correcting codes on the relay link would permit a reduction in SNR to the point where IM "noise" in the repeater would not be a major factor.

3.2.4 Transmitting Station

The TWT power amplifier has been discussed in paragraph 3.4 and the mixer in paragraph 3.2.1.5. That leaves the bi-phase modulator and master or auxiliary oscillator. Bi-phase or linear phase modulators can have spurious and nonlinear responses which for the most part will be eliminated or controlled by the subsequent filtering. The master or auxiliary oscillator will have some frequency and/or phase instability which ultimately will perturb the operation of the entire system. However, good design can ensure that the short-term fluctuations are small enough to be negligible compared to other system perturbations. The long-term fluctuations will be tracked-out by the system, affecting doppler measurements to some degree but not the system noise statistics. Similar comments apply to the frequency multiplier or synthesizer chains in the transmitting station: any spurious responses that are not eliminated by filtering will perturb the signal but not seriously affect the noise statistics.
4. NOISE CONSIDERATIONS

Section 3 has introduced the basic elements in the assumed baseline communications system and has identified and discussed the associated nonlinearities and/or perturbations. The most significant system nonlinearities occur in bandpass limiters, square-law or product devices (in the carrier recovery PLL), and traveling wave tube amplifiers (TWTA). The primary effects of these and other nonlinearities are to perturb and distort the signal and to alter the spectral shape of the noise. This section will present an argument that the system noise amplitude statistics are essentially unaffected by these nonlinearities and will support this assertion with some examples.

4.1 GENERAL CONSIDERATIONS

Paragraphs 3.2.1.6 and 3.3 have shown that the system noise is gaussian, generated primarily in the receiver front-ends in the relay satellite and receiving station. Much of communication theory assumes gaussian noise throughout the system, up to and including the input to the matched filter detector in the bit synchronizer. At the same time, much of this theory also assumes small signal, linear processes which do not alter the gaussian nature of the noise statistics. But what happens in a realistic system, containing many possibly troublesome nonlinearities? The square-law example in Section 1 has shown that nonlinearities can and do produce non-gaussian statistics from gaussian inputs.

As documented in the list of References, much of the current literature is unconcerned with probability density considerations; there appears to be an implicit assumption that the noise statistics remain gaussian, so that signal distortions and power spectra are the important considerations. This lack of concern can be attributed to the work of earlier investigators, which is well documented in Middleton (Reference 29). In Section 14.3-3 a most compelling reason is given: If the input to a zero-memory nonlinearity (e.g., rectification, clipping, etc.) is gaussian, and if this nonlinearity is followed by a sufficiently narrowband linear
filter, the output of this filter can be regarded as once again (asymptotically) normal, but (possibly) with a different mean and variance from those of the original input. That this is intuitively reasonable will be discussed in the examples below.

This statement can be extended throughout the communications system: If the system filters are spectrally narrow enough, so that the averaging times are sufficiently large compared to the correlation time(s) of the gaussian input(s), the noise in the system will remain (nearly) gaussian. So, for all practical purposes, the system nonlinearities will have negligible effect on the system noise amplitude statistics, provided the system is sufficiently narrowband.

As a final concluding remark, before considering some specific examples, note that the requirement for linear filters refers to linearity in the usual operator sense: linear transformation, superposition, etc. It does not require that the filters have linear phase vs. frequency characteristics. However, in an angle modulated system, nonlinear phase shift can alter the statistics of noise that has been modulated into the signal phase (e.g., when the signal is processed through a relay satellite employing a demodulation/remodulation configuration). Thus, the filters in an angle-modulated system also should have linear phase characteristics in addition to being linear in the usual sense.

4.2 EXAMPLES

The preceding discussions of system nonlinearities and their effects on system noise statistics have been largely qualitative. Further insight into these effects, plus an understanding of how narrowband linear filters can restore the gaussian nature of the noise, can be gained by considering some specific examples. The examples discussed will include the square-law device, the bandpass limiter, and a general nonlinearity expressed as an N-term power series. These nonlinearities are typical of the most significant nonlinearities identified in Section 3.
4.2.1 Square-Law Nonlinearity

A full-wave square-law detector is typical of the major nonlinearity in the frequency doubling circuits of a squaring loop (c.f., Reference 1). It consists of a square-law device with the transfer characteristic

\[ y = ax^2, \tag{4-1} \]

where \( a \) is a scaling constant, followed by a low-pass or averaging filter (Figure 4-1). Such a device is treated in Chapter 17 of Reference 29, Chapter 12 of Reference 30, and Chapter 5 of Reference 31.

Referring to Figures 1-2 and 1-3, if the input to the square-law device is gaussian with probability density

\[ p_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}, \tag{4-2} \]

the resultant output is chi-squared with probability density

\[ p_y(y) = \begin{cases} \frac{1}{\sqrt{2\pi ay}} e^{-\frac{y}{2\sigma^2}}, & y > 0 \\ 0, & \text{otherwise}. \end{cases} \tag{4-3} \]

If the input power spectrum is narrowband rectangular:

\[ S_x(f) = \begin{cases} A, & f_0 - B/2 < |f| < f_0 + B/2 \\ 0, & \text{otherwise}, \end{cases} \tag{4-4} \]

the resultant output power spectrum is triangular, plus an impulse at zero frequency to account for the non-zero mean of the output process:

\[ S_y(f) = \begin{cases} 4a^2A^2B^2\delta(f), & f = 0 \\ 4a^2A^2(B - |f|), & 0 < |f| < B \\ 2a^2A^2(B - |f| - 2f_0), & 2f_0 - B < |f| < 2f_0 + B \\ 0, & \text{otherwise}. \end{cases} \tag{4-5} \]
Square-Law Detector:

\[ x(t) = V(t) \cos[\omega_0 t + \phi(t)] \]  
(narrowband gaussian)

\[ y(t) = ax^2(t) = \frac{aV^2(t)}{2} + \frac{aV^2(t)}{2} \cos[2\omega_0 t + 2\phi(t)] \]  
(chi-squared)

\[ z(t) = \frac{aV^2(t)}{2} \]  
(2B filtering; exponential)

Envelope Probability Density Function:  (Rayleigh)

\[ p_V(V) = \begin{cases} 
\frac{V}{\sigma^2} e^{-\frac{V^2}{2\sigma^2}}, & V > 0 \\
0, & \text{otherwise} 
\end{cases} \]

Resultant (Filtered) Output Probability Density:

1. 2B Zonal Filtering:  (Exponential)

\[ p_Z(z) = \begin{cases} 
\frac{1}{a\sigma^2} e^{-\frac{z}{a\sigma^2}}, & z > 0 \\
0, & \text{otherwise} 
\end{cases} \]

2. < 2B Filtering:  (=Rayleigh)

3. << 2B Filtering:  (=Gaussian)

* Compare with Figures 1-2 and 1-3.

Figure 4-1. Filtered Square-Law Statistics
Since the original input is a narrowband gaussian random process, it can be expressed as

\[ x(t) = V(t) \cos [\omega_0 t + \phi(t)] , \quad (4-6) \]

where \( \omega_0 = \frac{\omega_o}{2\pi} \) is the center frequency of the input spectral density, \( V(t) \geq 0 \) is the envelope of the input, and \( 0 \leq \phi(t) \leq 2\pi \) (uniformly distributed) is the input phase. The output of the square-law nonlinearity is thus

\[ y(t) = ay^2(t) = \frac{aV^2(t)}{2} + \frac{aV^2(t)}{2} \cos [2\omega_0 t + 2\phi(t)] , \quad (4-7) \]

where the first term has spectral components (Figure 1-3) centered at zero frequency and the second term has components centered at \( 2\omega_0 \). If the bandwidth \( B \) is narrow compared to the center frequency \( \omega_0 \), these two spectra will not overlap, as shown in Figure 1-3 (compare with Figure 12-4, Reference 30).

If the square-law nonlinearity is followed by an ideal low-pass filter (also referred to as a low-pass zonal filter) with two-sided bandwidth \( 2B \) to \( 4\omega_0 - 2B \), the filter output will be

\[ z(t) = \frac{aV^2(t)}{2} , \quad (4-8) \]

whose spectrum is the central portion of the output spectrum in Figure 1-3 (from \(-B\) to \(+B\)). Since \( V(t) \) is the envelope of a narrowband gaussian random variable, it has a Rayleigh probability density function: (Figure 4-1)

\[ p_V(V) = \begin{cases} \frac{V}{\sigma^2} e^{-\frac{V^2}{2\sigma^2}} , & V \geq 0 \\ 0 , & \text{otherwise} \end{cases} \quad (4-9) \]
and the filter output has an exponential probability density:

$$p_z(z) = \begin{cases} 
\frac{1}{\sigma^2} \exp \left(-\frac{z}{\sigma^2}\right), & z \geq 0 \\
0 & \text{otherwise.}
\end{cases}$$

(4-10)

As shown in Figure 1-2, the output of the square-law nonlinearity without filtering has a chi-squared probability density $p_y(y)$. In Figure 4-1, with 2B zonal filtering the output has an exponential probability density $p_z(z)$ (compare with Figure 17.1, Reference 29 and Figure 12-3, Reference 30). With an even narrower filter, the output $p_z(z)$ probability density shown in Figure 4-1 approaches a form similar to the Rayleigh function (compare with Figure 17.3, Reference 29). Finally, as the ideal low-pass filter is made narrower and narrower, this output probability density approaches a gaussian form, with non-zero mean (compare with Figure 17.4, Reference 29).

Strictly speaking, this final filtered output process cannot be truly gaussian, since $p_z(z)$ must vanish for negative values of $z$. But, since the variance of the output decreases with the filter bandwidth and since the output mean is non-zero, for all practical purposes the output can be considered to be (nearly) gaussian. Thus, when the postrectification filter is spectrally narrow enough (compared to the bandwidth and correlation time of the input), the final output process is essentially gaussian, but with a different mean and variance from those of the original input.*

### 4.2.2 Bandpass Limiter

Perhaps an even more striking nonlinearity is a limiter, some characteristics of which are shown in Figure 4-2. The limiter falls in the class of full-wave (odd) $\lambda$th-law devices discussed in chapter 13, Reference 30. The hard limiter (referred to as a "super" limiter in Chapter 2, Reference 29) has the transfer characteristic (compare with Example 5-4, Reference 31)

* See Chapters 12 and 13, Reference 30 for treatment of signal plus noise by both the direct and transform methods, as well as Chapter 17, Reference 29 for additional theory on this subject. Reference 32 treats the square-law nonlinearity as a frequency doubler.
Bandpass Limiter:

\[ x(t) \rightarrow \text{Limiter} \rightarrow y(t) \rightarrow \text{Band-Pass Filter} \rightarrow z(t) \]

"Hard" or "Super" Limiter:

\[ y = \begin{cases} A, & x > 0 \\ -A, & x \leq 0 \end{cases} \]

\[ P(y = -A) = P(x \leq 0) = P_x(0) \]

\[ P(y = A) = P(x > 0) = 1 - P_x(0) \]

\[ p_y(y) = \int_{-\infty}^{\infty} p_y(y) dy \]

\[ P_y(y) = \begin{cases} 0, & y < -A \\ P_x(0), & -A \leq y < A \\ 1, & y \geq A \end{cases} \]

"Linear" Limiter:

\[ y = \begin{cases} A, & x \geq A \\ x, & -A < x < A \\ -A, & x \leq -A \end{cases} \]

\[ p_y(y) = P_X(y), \quad |y| < A \]

\[ P_y(y) = \begin{cases} 0, & y < -A \\ P_x(y), & -A \leq y < A \\ 1, & y \geq A \end{cases} \]

"Smooth" Limiter:

\[ p_y(y) \text{ and } P_y(y) \text{ similar to } "Linear" \text{ Limiter} \]

Figure 4-2. Limiter Statistics

4-7
\[ y = g(x) = \begin{cases} 
A, & x > 0 \\
-A, & x \leq 0,
\end{cases} \quad (4-11) \]

where the random variable \( y \) only takes on the two values \( \pm A \). As a result, the output probability density \( p_y(y) \) consists of two impulses (at \( \pm A \)), no matter what form the input probability density \( p_x(x) \) has. The impulse in \( p_y(y) \) at \( y = -A \) has an area which depends on the input probability distribution function:

\[
P\{y = -A\} = P\{x \leq 0\} = P_X(0) = \int_{-\infty}^{0} p_x(x) \, dx. \quad (4-12)
\]

Similarly, the area of the impulse at \( y = +A \) is

\[
P\{y = +A\} = P\{x > 0\} = 1 - P_X(0). \quad (4-13)
\]

The associated output probability distribution function \( P_y(y) \) is

\[
P_y(y) = \int_{-\infty}^{y} p_y(y) \, dy = \begin{cases} 
0, & y < -A \\
\int_{-\infty}^{-A} p_x(x) \, dx, & -A \leq y < A \\
1, & y \geq A.
\end{cases} \quad (4-14)
\]

In the case of a zero mean gaussian input, \( P_X(0) = 0.5 \), so the area under each impulse in the density function is 0.5.

A "linear" limiter (Figure 4-2) has a 1:1 linear region between the \( \pm A \) limits: (compare with Example 5-7, Reference 31)

\[
y = g(x) = \begin{cases} 
A, & x \geq A \\
x, & -A < x < A \\
-A, & x \leq -A
\end{cases} \quad (4-15)
\]

In this case, the output probability density function \( p_y(y) \) equals the input probability density \( p_x(y) \) for \( |y| < A \), and \( p_y(y) = 0 \) for \( |y| > A \). It contains two impulses at \( y = \pm A \) whose areas are

\[
P\{y = -A\} = P_X(-A) \quad (4-16)
\]

and

\[
P\{y = A\} = 1 - P_X(A),
\]

respectively. The associated probability distribution function is

\[
P_y(y) = \begin{cases} 
0, & y < -A \\
\int_{-\infty}^{-A} p_x(x) \, dx, & -A \leq y < A \\
1, & y \geq A.
\end{cases} \quad (4-17)
\]
For a zero mean gaussian input, the areas under the density function impulses are equal, since by symmetry $P_X(-A) = 1 - P_X(A)$.

A "smooth" limiter is similar in many respects to the "linear" limiter, except that its transfer characteristics are smooth and therefore require more complicated formulations. The formalism might be as a $v$th-law device (e.g., in Chapter 13, Reference 30, or in References 25 and 33), or as an exponential-law device (References 34 and 35). The precise form of the output probability density depends, of course, on the input probability density and the formalism used for the limiter characteristics. Nevertheless, the statistic will have a form similar to that of the "linear" limiter.

Much of the current literature on limiters is concerned with intermodulation distortion (e.g., References 20 through 27 and 33 through 35)*, signal-to-noise ratios (e.g., References 26 and 38 through 42)**, autocorrelation (e.g., References 25, 26, 33, and 43), or output spectra (e.g., References 20 through 22, 24, 34, 35, and 44).*** Aside from the fact that Middleton (Reference 29) has shown that the noise statistics remain gaussian with sufficiently narrowband filtering, the lack of interest shown in the noise probability density in these references can be attributed to the fact that signal perturbations due to intermodulation are the more serious problems, signal-to-noise considerations can be calculated on an rms basis, and spectrum investigations do not have to be related to the noise amplitude statistics. However, a brief consideration of the effects of a bandpass limiter on the resultant noise probability density can lend some insight into the overall process.

The discussion of Figure 4-2 has shown that the unfiltered output of a limiter does not have gaussian statistics. But consider the net effect of this limiter on the statistics of an angle modulated signal (following demodulation): The signal phase information is contained essentially in the zero-crossings of the signal. The limiters in Figure 4-2 preserve the

---

* Related interference topics are discussed in References 36 and 37.
** Reference 42 also considers output probability densities.
*** Other limiter discussions can be found in References 1, 28 through 31, 45, and 46.

4-9
zero-crossings, and the phase information will be retained intact*. If the subsequent demodulator has a perfect reference, its output will be influenced primarily by the phase information, so the resultant output probability density will be the same as if the limiter were not in the circuit. Of course, if the demodulator reference is in error (due to noise or other effects), the demodulator output will also be influenced to some degree by the input signal amplitude; in this case, the presence of a limiter in the signal path may alter the resultant output from that which would be obtained without the limiter.

4.2.3 Nth-Order Nonlinearity

The previous examples have considered nonlinearities which belong to the class of $v$th-law devices (Reference 30) whose transfer characteristic can be written as

$$y = a x^v,$$  \hspace{1cm} (4-18)

where $a$ is a scaling constant, and $v$ is some non-negative real number (including fractions). The analysis of such devices is quite laborious, as evidenced by the treatment in Chapter 13, Reference 30. An even more complicated, $N$th-order nonlinearity can be considered whose transfer characteristics can be described in terms of a power series:

$$y = \sum_{i=1}^{N} a_i x^i.$$  \hspace{1cm} (4-19)

Since this represents a nonlinear transformation, simple superposition techniques cannot be used, where the responses due to each individual term could be summed to determine the overall response. Some procedure similar to convolution may be required (recall that the probability density of a sum of independent random variables is the convolution of their respective probability densities; see Chapter 7, Reference 31); but the individual terms in (4-19) are hardly independent, so even convolution of the densities may not work. At the present time, this problem is sufficiently formidable to be beyond the scope of this study.

* As noted at the end of Section 13-2, Reference 30, "When the input to an ideal band-pass limiter is a narrow-band wave, the output is a purely phase-modulated wave; the phase modulation of the output is identical to that of the input."
A rather extensive analysis of the distortion in a third-order nonlinearity is given in Reference 47. Here, the nonlinearity is assumed to have the form

\[ e_{out} = a_1 e_{in} + a_2 e_{in}^2 + a_3 e_{in}^3, \]  

(4-20)

and the input is assumed to contain three sinusoidal components:

\[ e_{in} = A \cos \omega_a t + B \cos \omega_b t + C \cos \omega_c t. \]  

(4-21)

In this case, there are three first-order components resulting from the linear \(a_1\) term. As a result of the second-order \(a_2\) term there are three dc components, six sum and difference beat components, and three second-harmonic components. From the third-order \(a_3\) term there are three third-harmonic components, twelve \((2a+b)\) beat components, four triple \((abc)\) beat components, three components causing self-compression (gain for that component decreases as the input signal increases) when \(a_3\) is negative or self-expansion (gain increases with signal) when \(a_3\) is positive, and six components causing cross compression (decrease in gain at one frequency due to an increasing input at another frequency) when \(a_3\) is negative or cross expansion (gain increases at one frequency with signal at another) when \(a_3\) is positive. These results assume unmodulated input signals and therefore do not include cross-modulation components (the transfer of modulation from one channel to another). However, the mechanism causing cross compression/expansion will also cause cross modulation when the input signals are modulated.

Since in one respect system noise can be considered to be the result of many independent, closely spaced sinusoids, the above example gives some insight to the complexity of the nth-order nonlinearity problem. Fortunately, for most system elements such as amplifiers, the higher-order coefficients \(a_i, i \geq 2\) are usually much smaller than the first-order coefficient \(a_1\), so that for all practical purposes these elements can be considered to be linear.
REFERENCES


REFERENCES (Continued)


