NBS

Technical Note 103

EQUIPMENT CHARACTERISTICS AND THEIR RELATION TO SYSTEM PERFORMANCE FOR TROPOSPHERIC COMMUNICATION CIRCUITS

A. F. BARGHAUSEN, F. O. GUIRAUD, R. E. McGAVIN,
S. MURAHATA, AND R. W. WILBER

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

THE NATIONAL BUREAU OF STANDARDS

Functions and Activities

The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended by Congress in Public Law 619, 1950. These include the development and maintenance of the national standards of measurement and the provision of means and methods for making measurements consistent with these standards; the determination of physical constants and properties of materials; the development of methods and instruments for testing materials, devices, and structures; advisory services to government agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; and the development of standard practices, codes, and specifications. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services, and various consultation and information services. Research projects are also performed for other government agencies when the work relates to and supplements the basic program of the Bureau or when the Bureau's unique competence is required. The scope of activities is suggested by the listing of divisions and sections on the inside of the back cover.

Publications

The results of the Bureau's research are published either in the Bureau's own series of publications or in the journals of professional and scientific societies. The Bureau itself publishes three periodicals available from the Government Printing Office: The Journal of Research, published in four separate sections, presents complete scientific and technical papers; the Technical News Bulletin presents summary and preliminary reports on work in progress; and Basic Radio Propagation Predictions provides data for determining the best frequencies to use for radio communications throughout the world. There are also five series of nonperiodical publications: Monographs, Applied Mathematics Series, Handbooks, Miscellaneous Publications, and Technical Notes.

A complete listing of the Bureau's publications can be found in National Bureau of Standards Circular 460, Publications of the National Bureau of Standards, 1901 to June 1947 ($1.25), and the Supplement to National Bureau of Standards Circular 460, July 1947 to June 1957 ($1.50), and Miscellaneous Publication 240, July 1957 to June 1960 (Includes Titles of Papers Published in Outside Journals 1950 to 1959) ($2.25); available from the Superintendent of Documents, Government Printing Office, Washington 25, D. C.

NATIONAL BUREAU OF STANDARDS

Technical Note

103

JANUARY 15, 1963

EQUIPMENT CHARACTERISTICS AND THEIR RELATION TO SYSTEM PERFORMANCE FOR TROPOSPHERIC COMMUNICATION CIRCUITS

A. F. Barghausen, F. O. Guiraud, R. E. McGavin,
S. Murahata, and R. W. Wilber
Boulder Laboratories
Boulder, Colorado

NBS Technical Notes are designed to supplement the Bureau's regular publications program. They provide a means for making available scientific data that are of transient or limited interest. Technical Notes may be listed or referred to in the open literature.

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington 25, D.C. - Price $1.00

TABLE OF CONTENTS

ABSTRACT

The performance of a tropospheric communications system, either within the line of sight or beyond the line of sight, is directly dependent on the operating characteristics of the equipment.

Performance predictions of a communications system are made on the basis that equipment will operate in a prescribed manner. The degree of success of the communications system will depend largely upon how well these predicted values correspond to the actual operating values.

Consideration is given to those portions of the equipment that have definite effect upon the operating performance. Specific items of equipment and methods for determining their performance are considered. Representative results in light of the present state of the art permits an evaluation of an actual system in terms of realizing an "optimum" system.

In systems that do not have the "optimum" characteristics desired, consideration is given to laboratory devices which may alleviate these deficiencies. Future systems should consider incorporating these devices as development permits.

EQUIPMENT CHARACTERISTICS AND THEIR RELATION TO SYSTEM PERFORMANCE FOR TROPOSPHERIC COMMUNICATION CIRCUITS

by

A. F. Barghausen, F. O. Guiraud, R. E. McGavin
S. Murahata and R. W. Wilber

1. INTRODUCTION

The performance to be expected on a tropospheric scatter path can be predicted according to the methods outlined by Barsis, Norton, Rice, and Elder [1961] and Florman and Tary [1962]. However, these performance predictions are made on the assumption that the equipment used on a particular path will perform in an expected manner. Transmitter power, antenna gains, line losses, receiver noise figure, and a variety of other parameters are assumed to have certain values. The degree of success of operation will depend largely upon how well the correspondence between the assumed values and the actual values can be achieved.

This report considers these aspects of the equipment which affect the performance of a tropospheric circuit. While this report is divided into various sections to permit evaluation of specific parts, it should be emphasized that the performance of the total system is the desired result. Consequently, the assumption is that each part operates the same after incorporation in the system as it did before incorporation.

Each section considers both the purpose of the specific part or item, and methods for determining its performance.

In general, tropospheric circuits can be divided into two types:

Those which are concerned with the propagation of radio energy to a point far beyond the radio line of sight.

Those where propagation is over relatively short distances where the receiving terminal is within the radio line of sight of the transmitting terminal.

The same principles generally apply whether beyond line of sight or within line of sight is considered. The requirements and tolerances recommended are more stringent for long paths, whereas some latitude may be allowed for the shorter path. In this connection, it may be assumed that, unless otherwise stated, no distinction is being made between the two, and application, in so far as it is practical, will be uniform.

2. THE ANTENNA SYSTEM

The antenna system is here defined as the entire transmission system, which includes the antenna and the transmission lines. The antenna system is the transformer between the medium carrying the radio energy and the terminals, the transmitting or receiving equipment. The efficiency of the system is dependent upon how well this transformation takes place.

The antenna itself acts as a converter between the energy in a transmission line and the energy in a free-space wave. According to the princple of reciprocity, the characteristics of the antenna will be the same whether used as a transmitting or a receiving antenna. In point-to-point communication, directional antennas are used since energy directed to other than the intended terminal is wasted. The type of antenna to be used depends upon the type of path considered. Within-line-of-sight paths do not require high-gain antennas such as used on the long paths; however, very often a high-gain antenna will be used to take advantage of the narrow beamwidth to eliminate multipath reception and interference on a line-of-sight path. In the absence of such necessity, line-of-sight paths utilize low-gain antennas, the most common types being Yagis, dipoles, folded dipoles, horns, corner reflectors, and small parabolic antennas. Due to the power requirements for beyond line of sight, high-gain narrow-beam antennas are used, and in tropospheric propagation, the parabolic antenna fed by a horn primary antenna is in almost universal use.

The most important parameters affecting the efficiency of an antenna system are:

The line losses or attenuation of the transmission line, from transmitter to transmitting antenna.

The impedance matching of the components of the antenna system.

The radiation pattern of the antenna.

In some cases, the polarization is an important factor.

With the exception of the radiation pattern, the evaluation of the antenna system characteristics is the same, whether the path is within or beyond the line of sight.

However, it should be realized that the impedance and the radiation pattern are not necessarily independent quantities, and the adjustment to correct one possibly will have an effect upon the other. None the less, in considering these two parameters and the measurement of them, it is assumed, for convenience, that they are independent, it being understood that the goal to be achieved is an antenna having the proper radiation pattern as well as the proper impedance.

2.1. The Transmission Line

Before the antenna is measured for its characteristics, it is important to determine whether the characteristics of the transmission line by itself are such as to introduce errors into the measurement. Two parameters are of interest here: the VSWR of the transmission line which is a measure of the relationship between the transmission line characteristic impedance and its terminating impedance, and the line losses or attenuation. A transmission line calculator, as described by Smith [1944], is recommended for this work.

a. The Voltage Standing Wave Ratio (VSWR)

If an ideal transmission line is terminated in its characteristic impedance, the VSWR, as measured from the sending end, will be unity. However, if the transmission line is poorly assembled or if there are discontinuities due either to non-uniformity or to injury to the transmission line or to the presence of foreign matter in the transmission line, then a VSWR of unity would not be measured in all sections of the line even though it were terminated in its nominal characteristic impedance. It is important, therefore, to measure the VSWR of the transmission line. This measurement consists of terminating the transmission line with a load known to be equal to the characteristic impedance of the transmission line and measuring the VSWR at the sending end of the transmission. Such terminations for any size and type of line are available commercially and should be used for this measurement. The VSWR of the transmission line should be measured with this termination in place. If the VSWR should be found to be significantly different from one (e.g., in excess of 1.05:1), then the transmission line should be checked section by section to insure that all flanges are clean and undamaged, that all connections are proper and tight, and that the line is free from all foreign matter. All faulty sections of line should be replaced. Unless these precautions are taken, any further measurements should be made with caution.

Several methods can be employed to measure the VSWR of the transmission line. Fundamentally, all methods can be resolved into the two common types: measurement of E_{max} and E_{min} by use of a slotted line, and measurement of the impedance or admittance at some point along the transmission line.

Two types of slotted lines are available: slotted coaxial transmission line, and slotted waveguide. The same principles of operation apply whether coaxial line or waveguide is used for the slotted section. In essence, the slotted waveguide line is merely a section of waveguide in which a slot is milled in the center of the broad wall so that the disruption of the surface current flow is a minimum, and its presence does not appreciably modify the original field pattern. It is usually long enough so that several minima and maxima can be observed. A sliding probe is loosely coupled through the slot to the electric field which induces a voltage in the probe proportional to the probe depth. This voltage is then fed through associated circuitry and appears as a reading on an output meter. The relationship between this reading and the value of the field at the point being measured depends on the model of slotted line used. The manufacturer's recommendations should be consulted. The probe is moved through points of maximum field and minimum field, the ratio of which is the VSWR.

It must be remembered that the accuracy of any measurement is dependent upon the accuracy of the instruments used. Consequently, the slotted line should have a very good VSWR of its own if it is not to impair the accuracy of these measurements. Slotted lines having a VSWR of 1.05 are available. If the slotted line has a VSWR of 1.05, then it is capable of use in matching the VSWR of any device to within 1.10 in the most pessimistic configuration. A slotted line is a precision instrument and must be handled with care. The manufacturer's recommendations should be consulted to insure no damage due to excessive currents, to insure that the detector is used in its calibrated range, and to insure negligible error due to excessive probe depth. The slotted line should be protected from water or dust if its accuracy is to be maintained.

The second method, that of measuring the impedance or admittance at some point in the line, is usually used at the lower frequencies. The use of the slotted line is not recommended below 300 Mc/s, and even 400 Mc/s might prove to be inconvenient due to the wavelength. Essentially, this method utilizes a bridge circuit and null detector. The bridge is supplied with a known impedance which is used to balance out the impedance of the device under measurement. At balance, a null output from the bridge arrives at the detector. The bridge is then calibrated in either $R + jX$, $Z \underline{/\theta}$, or $G + jB$.

b. The Attenuation or Line Loss

It is not good practice to assume that the manufacturer's specifications are sufficient to determine the line loss of a system, especially where long sections of transmission line are used. Discontinuities and departures from a perfect transmission line can cause attenuation due

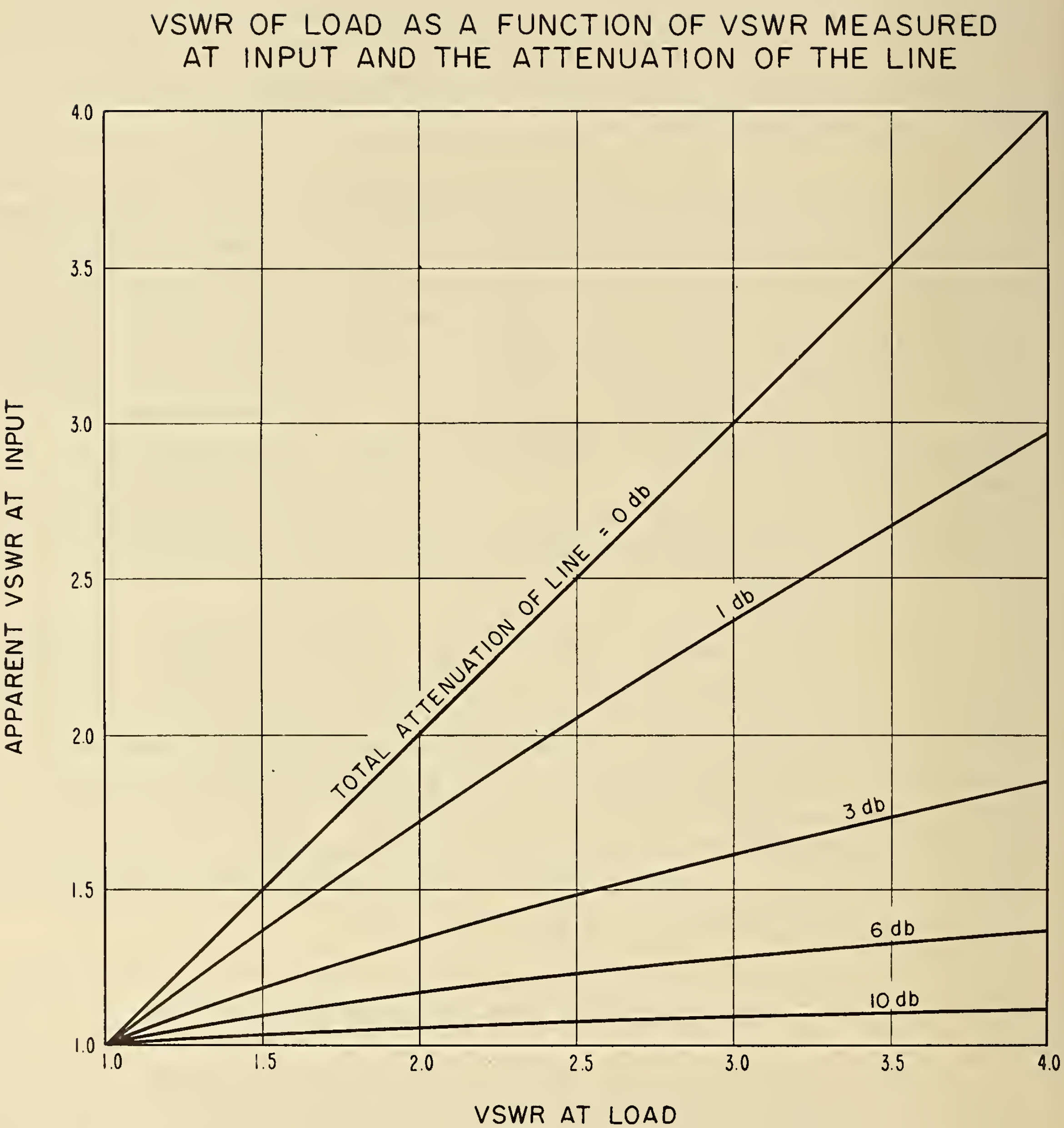

VSWR OF LOAD AS A FUNCTION OF VSWR MEASURED AT INPUT AND THE ATTENUATION OF THE LINE
APPARENT VSWR AT INPUT
VSWR AT LOAD
1.0
1.5
2.0
2.5
3.0
3.5
4.0
TOTAL ATTENUATION OF LINE = 0 db
1 db
3 db
6 db
10 db

Figure 1

to reflection within the transmission line. These reflections will not necessarily modify the VSWR measured at the end of a long transmission line, since, in theory, if the transmission line is infinite in length, a VSWR of one will be measured for any load terminating the transmission line. Consequently, the longer the transmission line, the more optimistic will be the measurement of the VSWR at the sending end of the line. For an accurate determination of the VSWR at the termination, but measured at the sending end of the line, a knowledge of the attenuation through the transmission line is required (see figure 1).

One accurate method of measuring the attenuation in a transmission line can be described as follows. The transmission line to be measured is made a part of a longer transmission line having the same characteristic impedance. Then the power loss through the combined system at the operating frequency is compared to the power loss which occurs in the system without the transmission line under test in the system. The difference in these two power losses is the power loss resulting from the attenuation of the transmission line under test. The receiving end of a large rigid transmission line would be terminated with a reducer from the transmission line down to some convenient flexible coaxial cable having the same characteristic impedance. It should be emphasized that the reducer (e.g., from 3-1/8 or 6-1/8 inch line down to RG8/U or RG17/U) must be of good quality, especially as regards the impedance match through the reducer. The flexible line is then connected between the reducer and the detector, which can be either a receiver or a bolometer bridge. If the transmission line is short (50 meters or less in length), it is not necessary to move the detector out to the end of the transmission line, and the flexible line can be long enough to reach back to the input of the transmission line to be tested. However, if the transmission line to be tested is long (greater than 50 meters), the attenuation in the flexible cable may be large, especially at the higher frequencies. Then the attenuation of the combined system may include only a very small additional attenuation due to the transmission line under test. The flexible line may be sufficiently unstable that the changes in attenuation rate due to flexing and heat change can be greater than the attenuation of the transmission line being tested. Therefore, under conditions such as these, it is recommended that the detector be placed at the end of the transmission line and be connected through a short piece

ATTENUATION RATES OF
STANDARD TRANSMISSION LINES

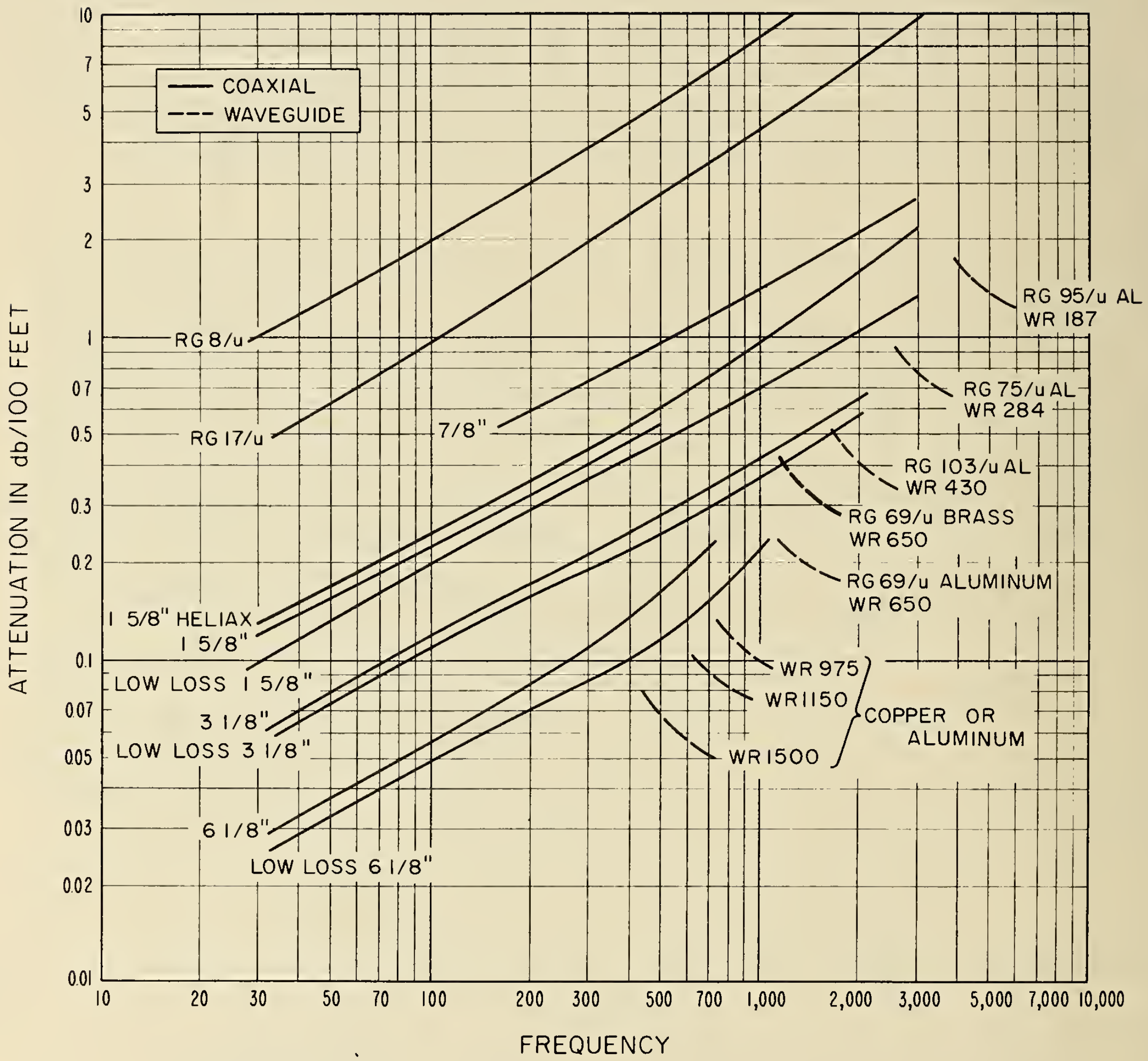


Figure 2

of flexible line. The second measurement with the transmission line out of the circuit will then require that the detector be moved to the input of the transmission line. (To insure accuracy in this measurement, the detector must be stable and should be powered from a regulated voltage source; if regulated a-c is not available at the end of the transmission line, a voltage regulator should accompany the detector). When the transmission line under test is taken out of the system from the generator, the signal is fed only through the flexible coaxial cable from the signal source to the detector or recorder and the input level reduced until the same output level is obtained. The difference in the two power levels is then the attenuation which was introduced by the transmission line under test.

The db attenuation can be determined directly if a precision variable attenuator calibrated in db is inserted in the line and remains in the system for both measurements. It is essential that the impedance match of the attenuator be constant irrespective of the value of attenuation inserted into the system. The measurements with the attenuator, the transmission line, and the flexible coaxial line are then compared to the same system minus the transmission line under test. The original level is produced in the second measurement by inserting attenuation from the variable attenuator until the level received is identical to the previous measurement. The value of the attenuation necessary to restore the level to the previous level is the amount of attenuation that the transmission line introduces into the system. If the measured attenuation is greater than that as found from figure 2 by more than several percent, then the line attenuation may be excessive, and the line should be checked carefully for causes of the departure. Other methods have been proposed from measuring the attenuation of the transmission line and its effect upon measurements of other parameters of the system. However, most of these methods assume a uniform rate of attenuation down the line. This is not necessarily encountered in practice, especially with rigid transmission line which is made up of many sections. Discontinuities where sections are bolted together can give rise to a serious problem both as to attenuation and VSWR. This is not reflected in an assumption of a uniform rate of attenuation. Consequently, the direct method of inserting and removing the transmission line under test in an overall transmission system is recommended.

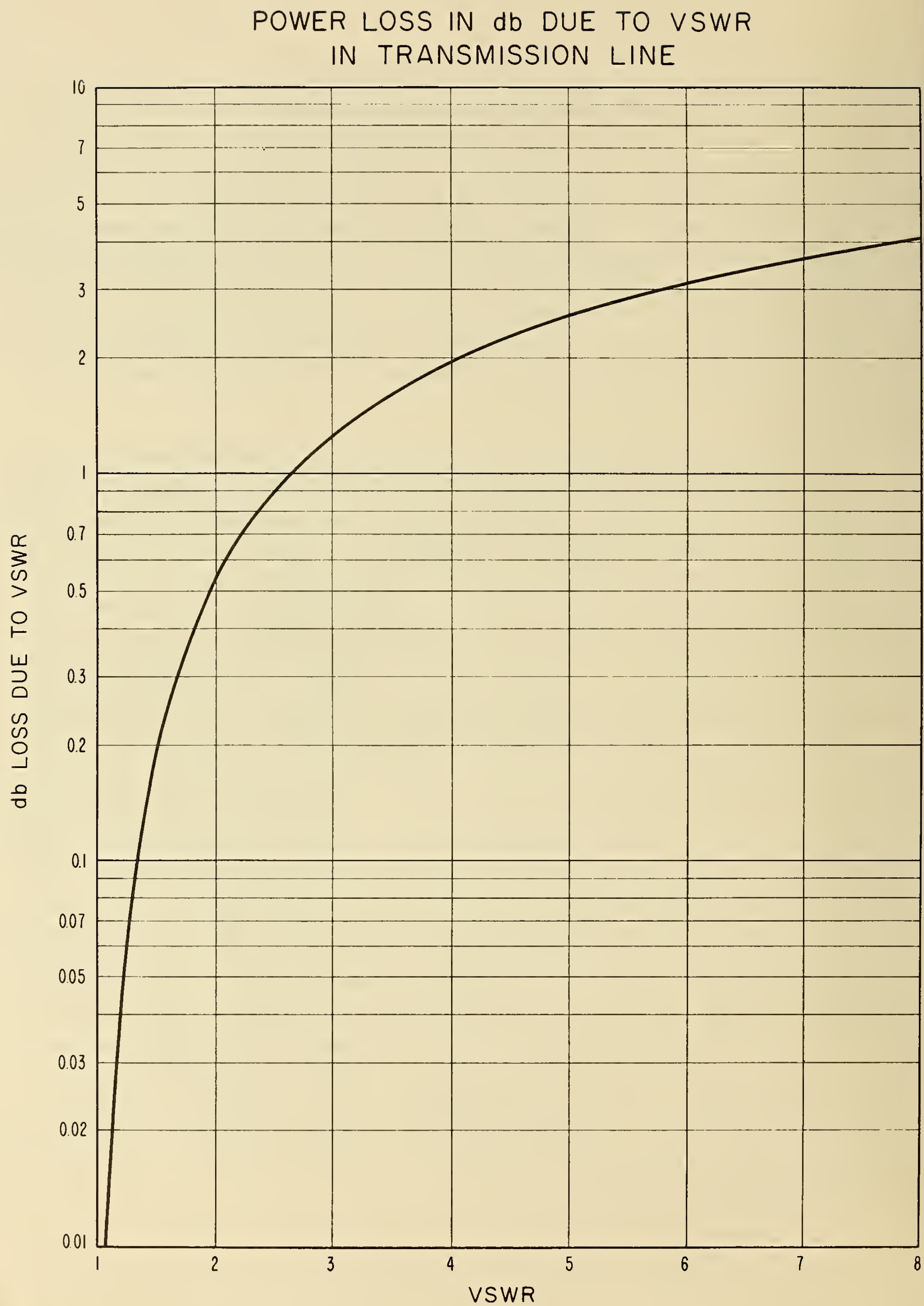

POWER LOSS IN db DUE TO VSWR
IN TRANSMISSION LINE
db LOSS DUE TO VSWR
10
7
5
3
2
1
0.7
0.5
0.3
0.2
0.1
0.07
0.05
0.03
0.02
0.01
1
2
3
4
5
6
7
8
VSWR

Figure 3

Additional references for recommended study may be found in King et al. [1945], Skilling [1951], and Wind and Rapaport [1955].

2.2. The Antenna

One of the most important parts of a communication system is the antenna. Therefore, a reliable and accurate determination of its operating characteristics is imperative to evaluate the system performance. The primary technical aspects of the antenna can be treated in the following categories. While the mechanical aspects of the antenna are not listed, they are of utmost importance, but are beyond the scope of this report.

a. The Voltage Standing Wave Ratio (VSWR)

The procedure for measuring the VSWR of the antenna and the transmission line is the same as the method employed to measure the VSWR of the transmission line itself. In this case the antenna replaces the termination. A properly designed and erected antenna will have a low VSWR. Without any major adjustments to the antenna, VSWR of 1.3:1 should be easily realized in practice.

This VSWR is the measure of the match of the antenna to the transmission line. From the standpoint of radiation efficiency, this may or may not be a real problem. If the plate circuit of the transmitter can be tuned to match the impedance at the input end of a lossless transmission line, then all the power from the transmitter will be delivered to the antenna and be radiated. However, if the transmitter cannot be matched to the impedance at the input end of the transmission line, which the antenna terminates, some of the power is reflected into the plate circuit to be reradiated or to be absorbed there. The reradiated power can cause distortion due to its phase lag and the absorbed power can cause a heating problem at the transmitter. In any case, power is lost and the full transmitter power is not utilized. Figure 3 illustrates the loss in power that can be expected if the antenna system has a mismatch which is not compensated by transmitter tuning. One of the chief disadvantages of large standing waves on a line is the degradation of the power-carrying capacity of the line. An excessive VSWR will increase the possibility of voltage flashover.

The power dissipation capabilities are not vital at the receiver input. However, the VSWR is still of importance. Although it can be shown that the optimum operation of the receiver is obtained when the receiver is mismatched to the transmission line, the conditions for this mismatch for optimum operation are not arbitrary. Consequently, as far as the transmission line is concerned, it is more advantageous to maintain a matched or flat line in the direction of the antenna and to produce the proper mismatch condition at the input stages of the receiver. This question is considered in detail later in this report where the noise figure of the receiver is considered. Therefore, as far as the antenna system is concerned, VSWR's which approach unity are desirable in all cases.

Methods are available to correct for large VSWR's. However, a properly designed antenna system should not exhibit an excessive VSWR when the system is optimized; a VSWR in excess of 3:1 should be rare. Where parabolic antennas are used, matching devices are necessary to optimize the VSWR since in most cases the uncorrected VSWR will be limited due to the energy from the reflector that intercepts the transmitting horn. The energy from the reflector acts as a reflected wave and sets up a standing wave. A proper mismatch must then be introduced to re-reflect this energy back to be radiated. Such provisions usually are incorporated in the transmitting horn in the form of tuning stubs.

b. The Radiation Pattern and Gain

The radiation pattern of an antenna is the graphical representation of the radiation intensity [Schelkunoff et al., 1952]. The isotropic antenna, a fictitious ideal antenna, is a radiating point source. Other antennas do not radiate equally in all directions, and hence do not have the spherical pattern of the isotropic antenna. With direction antennas, the energy is concentrated in one direction, so that the power density along one direction is much greater than that in any other direction. If the antenna is located in free space, then the ratio of the field strength experienced at a point in the direction of maximum radiation to that which would be experienced if the same power were radiated from an isotropic antenna is the free-space gain of the antenna.

THEORETICAL GAIN OF PARABOLIC ANTENNAS (54% SURFACE EFFICIENCY ASSUMED)

$$G_{db} = 20 \log_{10} f_{mc} + 20 \log_{10} D_{ft} - 52.6$$

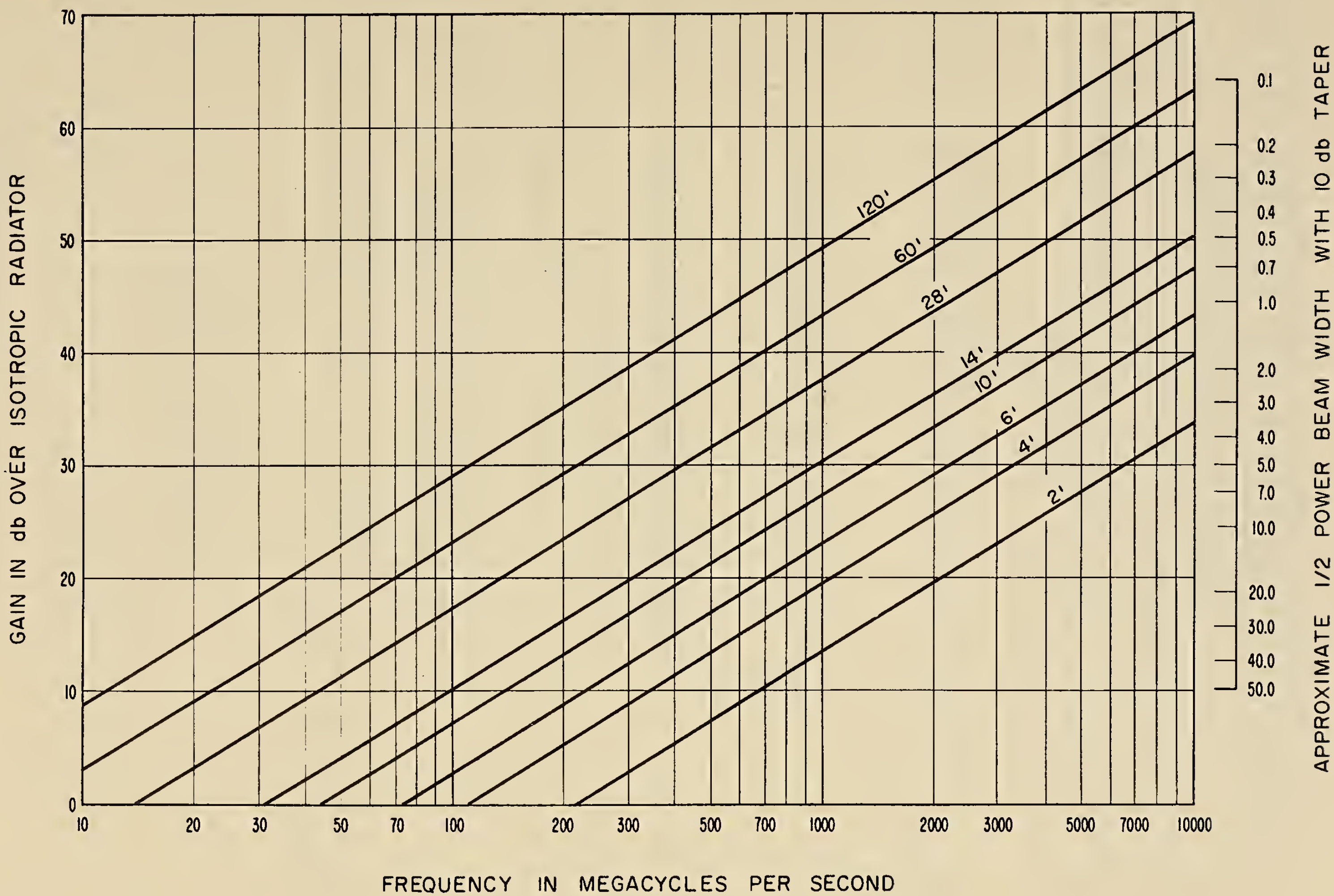


Figure 4

Since the isotropic antenna is an ideal antenna, the comparison is made relative to an antenna whose gain with respect to an isotropic antenna is known. For example, a vertical dipole has a gain of 2.15 db over an isotropic antenna. An additional parameter closely related to the gain of the antenna is the beamwidth. The 3-db (or half-power) beamwidth of the antenna is defined as the width of the cone of energy measured from a point on each side of the axis of maximum radiation where the power density is decreased to one-half. This total width is the half-power beamwidth of the antenna.

The gain to be expected from a parabolic antenna as shown in figure 4 can be estimated from the following expresion from FTR Handbook [1949]:

$$G_{db} = 20 \log_{10} f_{Mc} + 20 \log_{10} D_{ft} - 52.6 \qquad (1)$$

where

f_{Mc}	=	frequency in megacycles per second
D_{ft}	=	diameter of the reflector in feet
G_{db}	=	gain in decibels over an isotropic antenna
N.B.		These calculations assume a reflector efficiency of 54%, a reasonably attain figure.

If the parabolic reflector is illuminated uniformly, the approximate relation between the half-power beamwidth and the gain is expressed by the following:

$$\phi \cong \frac{142}{\sqrt{G}} \text{ degrees} \qquad (2)$$

where

ϕ = plane angle between half-power points of the radiation pattern of the antenna

$$G = \text{antilog}\left(\frac{G_{db}}{10}\right)$$

$$\sqrt{G} = \text{antilog}\left(\frac{G_{db}}{20}\right)$$

Example

f_{Mc} = 1000 megacycles

D_{ft} = 60 feet

Gain

$$G_{db} = 20\ \log_{10}(1000) + 20\ \log_{10}(60) - 52.6$$

$$= \quad 60 \quad + 35.6 \quad - 52.6$$

$$G_{db} = 43 \text{ decibel}$$

Beamwidth

$$\phi \cong \frac{142}{\sqrt{G}}$$

$$\sqrt{G} = \text{antilog}\left(\frac{G_{db}}{20}\right)$$

$$\sqrt{G} = \text{antilog}\left(\frac{43}{20}\right) = \text{antilog } 2.15$$

$$\sqrt{G} = 141.3$$

$$\phi \cong \frac{142}{141.3} \cong 1.0^{\circ}$$

However, since uniform illumination is not often used because side-lobe suppression is better achieved with a tapered form of reflector illumination (i.e., the energy incident at the edge of the reflector is reduced in a particular manner below that incident at the center of the antenna), the beamwidth is usually somewhat wider. A convenient approximation for a 10-db tapered illumination is

$$\phi \cong 1.15\ \phi_2 \tag{3}$$

where

ϕ_1 = half-power beamwidth with a 10-db taper

ϕ_2 = half-power beamwidth with uniform illumination

It should be evident that if the antenna beamwidth is very narrow, considerable accuracy is required in the pointing of the antenna. An example of the importance of the proper alignment can be given by comparing the beamwidth of typical parabolic antennas of certain gain figures.

For example, the relationship between gain and beamwidth can be illustrated as follows:

TABLE 1

Gain	Half-power beamwidth
30 db	5°
35 db	3°
$>$ 44 db	$< 1^{\circ}$

Small departures of the true direction can cause a considerable power loss due to misalignment. Therefore, two parameters are of exceptional importance: the gain of the antenna and the direction of the beam.

Of all the parameters of a tropospheric scatter system, probably the most difficult to measure is the actual pattern and gain of the antenna. In the practical case, a true determination of the gain and pattern of the antenna is all but impossible, and any method will yield only an estimate of the true condition.

The gain figure used in the transmission loss formula is the free-space gain or at least an estimate of it. The free-space gain of an antenna is the ratio of the signal received at a distance from the specified antenna to that which would be received from an isotropic antenna if both antennas were located in free space. However, the antenna to be measured is not located in free space, and the pattern will be affected by the ground in the foreground of the antenna. In the general case, the effect of the energy reflected by the ground is very difficult to analyze and still more difficult to eliminate. With extremely large antennas, it is not practical to move them to a site where free-space conditions can be approximated, and the determination of the pattern and gain will depend upon the ability to determine the effects of the ground reflection.

In free space the radiation from a parabolic antenna is a plane wave. If the same antenna is placed in close proximity to the ground, the radiation from the ground in the region close to the antenna can distort the wave front such that it is no longer a plane wave. This also distorts the free-space pattern. In most cases, especially where large parabolic antennas are used, this distortion is a second-order effect. However, this effect should be considered where the antenna is close to the ground and where the immediate foreground is rough and broken.

Even if the wave front is not affected by the ground, the radiation pattern will be distorted because of ground reflection. Any point in space will receive energy both directly from the transmitter and indirectly through reflection from the ground in front of the antenna. This produces a familiar lobe structure. If the ground were perfectly smooth, the effect would be amenable to analysis. In the practical case, however, the ground is seldom smooth. In general, the exact effects of the reflection from the ground in the vicinity of the antenna is difficult to determine, and under conditions usually encountered in ordinary field measurements, exact determinations can be considered to be impossible.

MINIMUM DISTANCE REQUIRED BETWEEN THE ANTENNA UNDER TEST AND THE RECEIVING ANTENNA FOR AN ERROR NO GREATER THAN I PERCENT

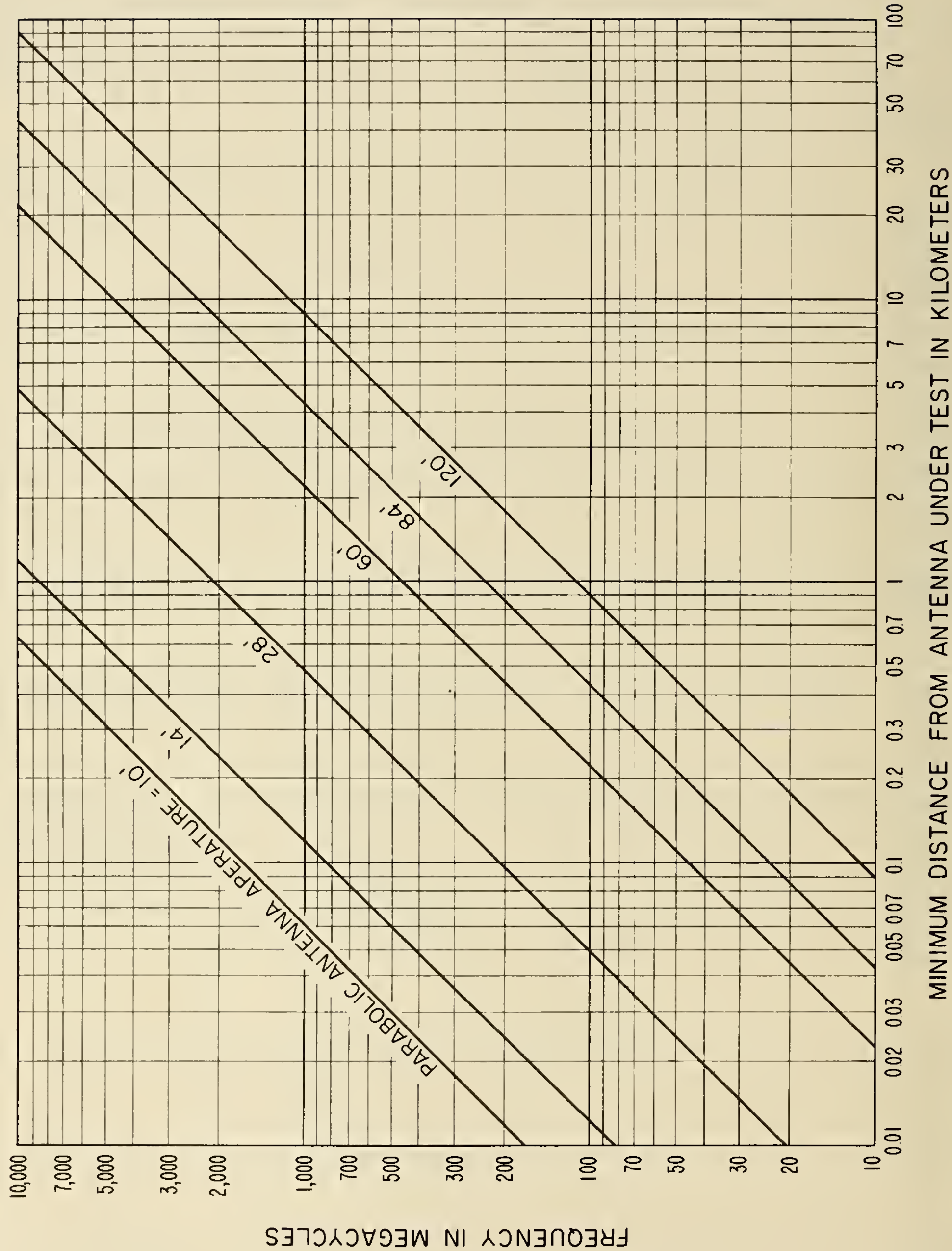


Figure 5

However, an "estimate" of the free-space gain can be determined. The degree of accuracy of this estimate is related to the terrain over which the antenna is used. If the terrain is extremely rough and broken with significant prominences along the transmission path, or if there are structures in the near vicinity of the transmission path, then the estimate probably will be poor. If the transmission path directly in front of the antenna is approximately smooth and free of obstructions (including the area immediately surrounding the antenna) the estimate will be accurate, i.e., within several db of the actual free-space gain.

The measurement consists of transmission to a receiving terminal established along the transmission path some distance in front of the transmitting antenna. A comparison is made between signals received at the receiving location from the antenna under test and from a standard antenna. Before such a measurement can be undertaken, possibilities of further error must be evaluated. Two important considerations are necessary: the minimum distance required between the two antennas, and the evaluation of the effect of ground reflection.

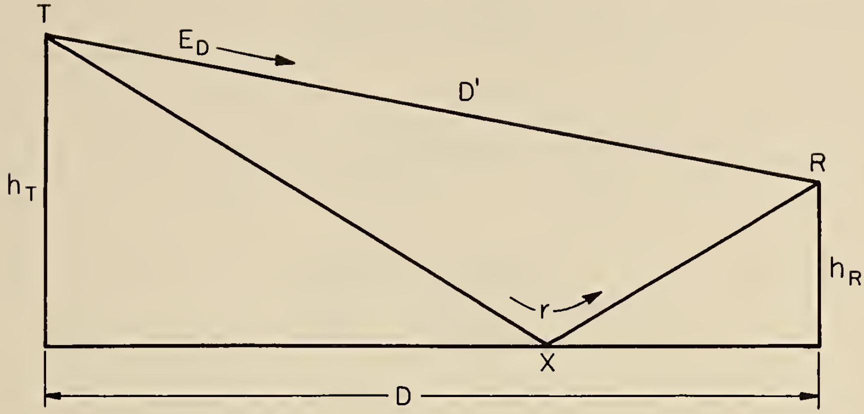
Even if the two antennas were located in free space, error would be present in the measurement of the gain of the antenna under test due to the finite distance between the two antennas. Although a usual determination assumes that only plane waves are considered, where the antennas are separated by a finite distance, the wave front arriving at the receiving antenna from a point source is not a plane wave but has a radius of curvature. This phenomenon can most easily be understood if we consider the large antenna to be the receiving antenna. If the wave front which arrives at the large antenna in free space is not a plane wave, then the field at the edge of the antenna differs in phase from the field at the center of the antenna by a small amount, Δ. If an accurate pattern is to be measured, Δ must be kept negligible. A commonly used criterion given by Silver [1949] is that the difference between the spherical wave and the plane wave at the edge of the antenna reflector should be less than $\lambda/16$. By this condition, then, the minimum distance from the antenna to be tested must be greater than $2d^2/\lambda$ where d is the diameter of the antenna and λ is the wavelength of the carrier frequency as shown in figure 5.

Using this minimum distance, the measured power gain in free space would be in error by no more than 1%, with respect to a perfectly plane wave front. For example, if a 60-foot parabolic antenna is being tested at a frequency of 1000 Mc/s, then this minimum distance, such that the error will be not greater than 1%, would be 2.23 km (1.39 miles) between antennas. In comparison, if the minimum distance used is d^2/λ, the apparent gain is reduced by 5%, and the side lobes of the antenna appear to increase. If at all possible, the condition mentioned previously, that is, $2d^2/\lambda$, should be maintained. At times this might be impractical since, if the same antenna were being used at 4000 Mc/s, this minimum distance is 8.92 km (5.54 miles). If the antenna is located on a coast line, then extreme ranges such as this would necessitate the use of a ship or slow moving aircraft such as a helicopter for the target terminal.

As has been pointed out previously, the presence of the ground renders the measurement of the radiation pattern difficult. If the immediate foreground of the parabolic antenna is reasonably smooth and if the immediate area around the antenna is free from obstructions, the assumption of a plane-wave emission from the parabolic antenna is at least a reasonably approximation.

The second effect, that of the distortion of the pattern due to the addition of the direct radiation and the ground reflected radiation, can be evaluated if the ground within sight of the antenna is not extremely rough and broken. If the ground were perfectly smooth, a regular lobe structure would be formed where a field maximum occurs when the two components are in phase and where a field minimum occurs when the two components are out of phase. Where the ground in the sight of the antenna is broken, no easy method of analysis is available except insofar as an approximation to a smooth surface is valid. An exhaustive analysis of the surface, point by point, would be necessary to determine the actual ground reflection. If the roughness is not too excessive and is fairly regular, then uniform reflection can be assumed. The amplitude of the reflection coefficient could then be assumed to have a mean value over most of the reflecting region.

A better appreciation of the problem can be had from a brief consideration of phase addition of the two components when propagation is between two terminals over a smooth earth.



E_D = field due to the wave arriving over path D',

E_R = field due to the wave arriving over path R,

E = resultant field at R,

r = reflected path distance,

h_T = transmitting antenna height,

h_R = receiving antenna height,

X = point of reflection,

D = flat earth distance from base h_T to h_R,

and D' = distance of transmitting antenna to receiving antenna.

Assuming that the reflecting surface is reasonably flat, we have by simple geometry:

$$D'^2 = (h_T - h_R)^2 + D^2 \tag{4}$$

$$r^2 = (h_T + h_R)^2 + D^2 \tag{5}$$

$$r^2 - D'^2 = 4\, h_T h_R \tag{6}$$

If

$$h_T \text{ and } h_R << D, \text{ then } R + D' \approx 2\, D$$

$$r - D' = \frac{4h_T h_R}{r + D'} \approx \frac{2h_T h_R}{D} \tag{7}$$

the difference in path length is approximately equal to $2h_T h_R/D$.

If horizontal polarization is used, a phase shift of 180° occurs upon reflection at x for any path consideration. If vertical polarization is used, a phase shift of 180° occurs only for regions where the angle between the reflecting surface and the reflected ray is small, a condition almost universally found in this type of measurement. Therefore, if the path length difference is λ, then the two signals arrive out of phase and cancel.

If $2h_T h_R/D = 2n\lambda/2$, where $n = 1, 2, 3 \ldots.$, a minimum occurs. (8)

If $2h_T h_R/D = (2n - 1)\, \lambda/2$, where $n = 1, 2, 3 \ldots.$, a maximum occurs. (9)

If this path length difference is then converted to phase difference between the two paths,

$$\theta = 4\, \pi\, h_T h_R/D\lambda \text{ radians.} \tag{10}$$

The energy reflected from the ground is a function of the ground constants, the dielectric constant, and the conductivity. In general, not all the energy is reflected in a specular fashion nor is the phase angle undisturbed. (The change in the phase angle has been considered above.) The amplitude of the reflection coefficient ρ will always be less than one. Therefore, the amplitude of the reflected wave will be the amplitude of the wave striking the reflecting surface multiplied by the reflection coefficient. In addition, the antenna pattern must be considered, which can be represented by $f(\psi)$, the power density in any given azimuth or elevation angle. In this text this will be referred to as the antenna function. Especially where narrow-beam antennas are concerned, the power density in the direction of the reflecting region will be less than that along the direct path. The amplitude of the reflected wave must be multiplied also by this antenna function. Then the amplitude of the reflected wave will be

$$E_R = \rho f(\psi) E_D \tag{11}$$

Using the law of cosines to determine the vector sum of the direct and the ground-reflected waves yields

$$E^2 = E_D^2 \left[1 + (\rho f(\psi))^2 - 2 \rho f(\psi) \cos \theta\right] \tag{12}$$

where θ is the angle between the direct and the ground-reflected waves The values of maximum and minimum field strength caused by alternate reinforcement and cancellation of the direct and reflected waves can be determined by the expressions

$$E_{max} = E_D \left[1 + \rho f(\psi)\right] \tag{13}$$

$$E_{min} = E_D \left[1 - \rho f(\psi)\right] \tag{14}$$

That is, the value of the measured field can vary between these two limits if the attenuation with distance is ignored.

From the composition of the lobe structure determined by the above equations, it can be shown that the minimum fields are very sharp, whereas the maximum fields are approached through a shallow gradient. This indicates that the maximums are the more useful since positioning of the antenna is less critical, and the integration of the field over the aperture of the antenna approaches the integration of a constant. It can be seen that there is a point where the measured field will be equal to the field produced by the direct wave E_D, which is the free-space field. But this point has the disadvantage of being located on a rather steep part of the curve of the lobe structure and is undesirable because of the above reasons.

Then the estimate of the gain of the antenna should be determined by comparing the field measured with the antenna under test to the field measured with a standard antenna, where the measurement is made at the maximum of the field pattern at the receiving antenna. This measurement is based on several assumptions: one, that the reflection coefficient is the same for both the test antenna and the standard antenna and, two, that the antenna pattern functions are approximately the same. The former assumption is reasonably since both antennas observe the same terrain; the latter assumption may be questioned especially where very narrow-beam antennas are being used. The effect of the latter assumption may be reduced by moving the receiving site farther out along the transmission path. (This does have the disadvantage, however, in that the height of the first maximum is increased.) The degree of error introduced due to the antenna function can be evaluated according to the following:

Let

E_{max_1} = the field measured with the antenna under test.

E_{max_2} = the field measured with the standard antenna.

Then

$$\frac{E_{max_1}}{E_{max_2}} = \frac{E_{D_1}}{E_{D_2}} \left[\frac{1 + \rho f_1 (\psi)}{1 + \rho f_2 (\psi)} \right] \tag{15}$$

The maximum error due to a difference in the antenna functions of the two antennas will be 3 db, this occurring when the quantities within the braces is equal to either 2 or 1/2. This, of course, is a limit and will never be encountered in practice. It will be rare when an error as great as 2 db is introduced due to this condition.

Example:

Assume h_T or transmitter antenna height of 10.7 m (35 feet)

D = 2.5 km (1.55 miles)

λ = .3 meter, i.e., f = 1000 Mc/s

Then the maximum fields occur where

$$\frac{2\,h_T h_R}{D} = (2n - 1)\,\frac{\lambda}{2}$$

The first maximum occurs where

$$\frac{2h_T h_R}{D} = [\,2(1) - 1\,]\,\frac{\lambda}{2}$$

$$h_R = \frac{D\lambda}{4\,h_T}$$

$$h_R = \frac{2.5\,(10^3)\,(.3)}{4\,(10.7)} = 17.5 \text{ meters (57.4 feet)}$$

Then under these conditions the maximum field at the receiving location should be measured at approximately 18 meters above the ground.

c. Measurement of Antenna Beamwidth and Gain

The method proposed is based on a comparison between the signal received from a standard antenna whose gain is known and the antenna under test. The standard antenna can be one of many types. Standard horns are available commercially having known gains and radiation patterns, and should be considered especially for the higher frequencies. However, for the lower frequencies a standard horn is quite large and heavy, and the use of corner reflectors is recommended as described by Cottony and Wilson [1958]. Corner reflectors have the advantage that they are relatively light even at the lower frequencies, can be collapsed to lie flat, and can be so constructed that they can maintain their characteristics through many assemblies and disassemblies.

It is assumed that the receiving location can be seen from the transmitting location. A theodolite or transit set up at the base of the antenna to be tested should be pointed along the transmission path. Either radio communications or visual signals should be used between the two sites. A mobile unit should proceed out to the proper distance and should be positioned on the radial by an operator at the transit. A signal is radiated from the standard antenna. The receiving antenna is raised, passed through the maximum field, and then returned to the position where the maximum field is measured. This height should be close to that calculated. It will be noted that the field pattern will not be a smooth curve but will in general display a certain amount of variability due to the roughness of the reflecting surface. In the region of the maximum this variability should concentrate about a mean value, which mean value should be taken as the value of the maximum. If the reflecting surface is inordinately rough, a lobe structure will be difficult to detect, and the field will vary with no apparent predicability. (If such a condition exists, the accuracy of this measurement may be poor.) After the value of the field produced by the radiation from the standard antenna has been logged, the standard antenna is lowered and the signal is applied to the antenna under test. The above procedure is repeated and the value logged. The maximum field in the second measurement should occur at the same antenna height that it did in the first measurement. The above two readings of received fields are then corrected taking into account the differences in line losses, and the ratio of these values is the estimate of the gain of the antenna. This ratio is usually converted to db.

The gain of the antenna to be tested may be 20 db over the gain of the standard antenna, and the dynamic range of the receiver may not permit an accurate comparison of signals as much as 20 db apart.

It is suggested that accurate attenuators be inserted at the input to the receiver when measuring the high-gain antenna. The attenuation thus added must be taken as part of the apparent gain of the antenna under test.

An estimate of the field to be expected at the receiving terminal from the radiation from the standard antenna can be accomplished by calculating the free-space field to be expected at the distance D using the methods described by Rice, Longley, and Norton [1962].

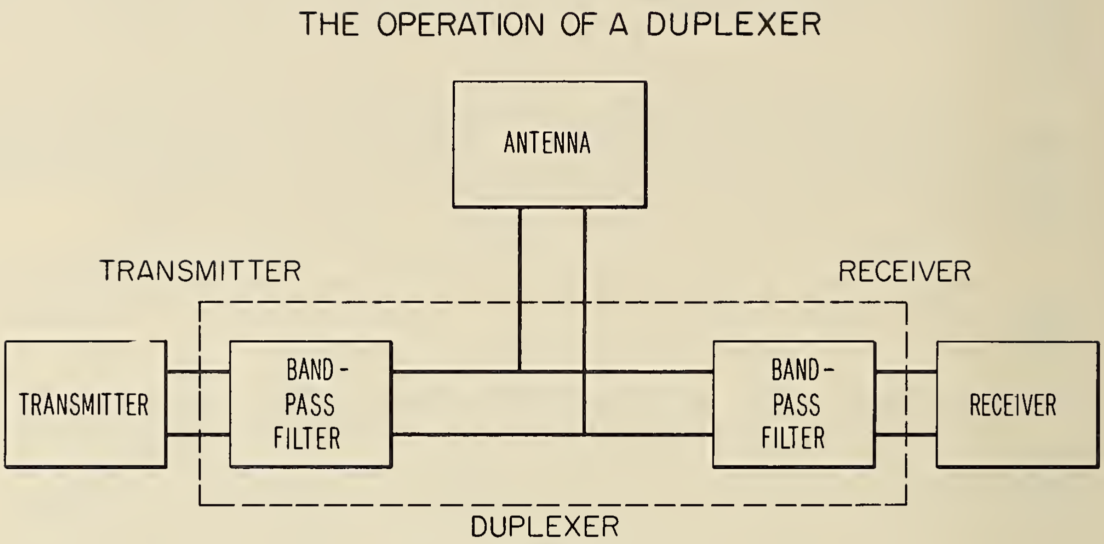
This value is then used as the value of the field due to the direct wave, E_D.

An estimate of the beamwidth between half-power points in the horizontal or vertical plane patterns can be made by repeating the above measurements on each side of the radial over a sufficient angle to incorporate at least twice the theoretical beamwidth of the antenna. The limitations on this measurement are the same as those for the original measurement. If the lobe structure of the antenna is difficult to detect, the determination of the beamwidth will be a poor estimate. The value of the gain loss should be equal on each side of the transmission path. If these two measurements are significantly different, then two additional points, midway between the points already measured, should be used to produce two additional gain determinations. These five points should fall on a relatively smooth curve from which estimates of the beamwidth and pointing accuracy can be determined.

The reader is referred to Kraus [1950], Schelkunoff and Friis [1952], and other standard reference books for a more detailed treatment on the subject of antennas.

3. DUPLEXER

A duplexer is a device which makes possible the use of a single antenna for simultaneous transmission and reception. Where pulse techniques are used and the requirement for simultaneous transmission and reception from the same antenna is not present, electronic switching devices may be used for the transmitter and receiver circuits, respectively, as described by Smullen and Montgomery [1949]. This manner


THE OPERATION OF A DUPLEXER
ANTENNA
TRANSMITTER
RECEIVER
TRANSMITTER
BAND-PASS FILTER
BAND-PASS FILTER
RECEIVER
DUPLEXER

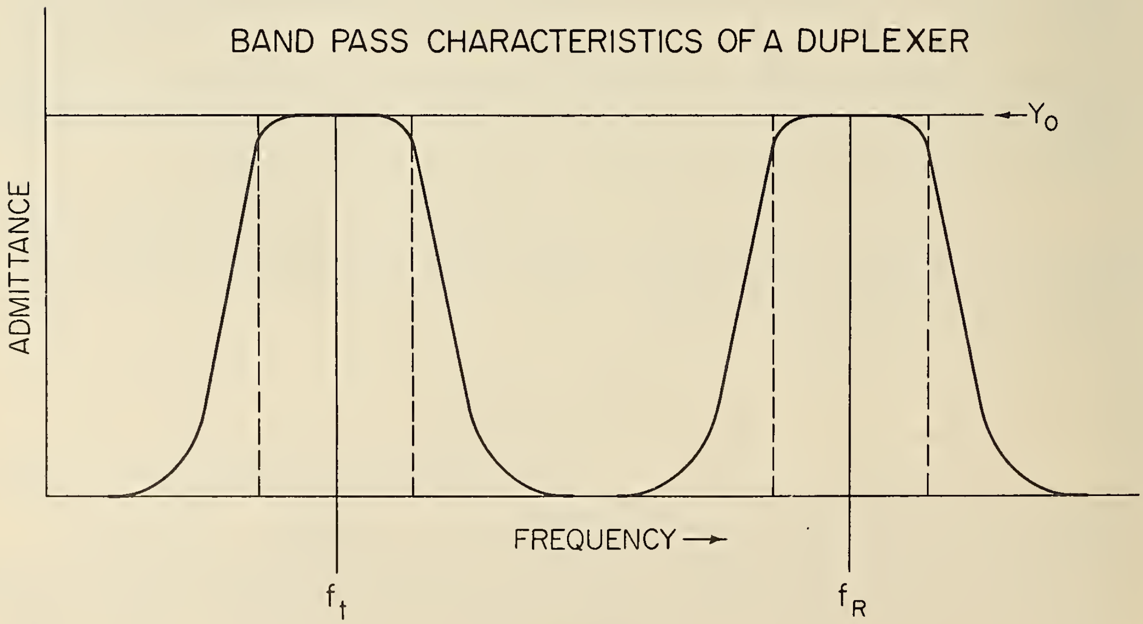

BAND PASS CHARACTERISTICS OF A DUPLEXER
ADMITTANCE
Y_0
FREQUENCY
f_t
f_R

Figure 6

of operation presupposes that the antenna is used for transmitting at one instant of time, whereas it is used for receiving at another instant of time. However, a cw system imposes more restrictive conditions on the duplexer than would a system involving pulse transmission. Where cw is used, a duplexer must effectively keep the transmitter isolated from the receiver at all times and yet maintain maximum coupling between the transmitting and receiving antenna circuits. If the transmitted and received signals on the same frequency and polarization, it can be shown that a lossless 3-terminal network cannot satisfy these requirements. A 4-terminal network, however, can satisfy these requirements, such a network being in the nature of a magic T. However, the use of this device requires that the fourth terminal, the one not connected either to the transmitter, receiver, or antenna must be terminated in a matched load. This matched load dissipates half the transmitted power and half the received power, indicating that if the device is essentially lossless, then there is still an insertion loss of 3 db whether the transmitted direction or the received direction is considered. Not only is this loss undesirable from the aspect of circuit efficiency, but this loss imposes severe requirements on a terminating load, especially where high-powered transmitters are being used.

To avoid these restrictions, the transmitted signal and the received signal are usually separated in frequency, and the frequency difference between these two signals makes possible their isolation one from the other while maintaining a good coupling factor to the antenna. Essentially, the device is so tuned that the receiver leg appears to have an admittance approaching zero at transmitting frequency whereas the transmitter leg has an admittance approaching zero at the receiving frequency. The antenna must be sufficiently broad-banded such that both the transmitted frequency and the received frequency can be matched with a low VSWR.

Figure 6 illustrates the principle of operation of such a duplexer. The transmitter and receiver legs are connected in parallel to the antenna. The bandpass filter on the receiver leg is so constructed that it presents the characteristic admittance of the line at the center frequency of the received signal, yet presents an admittance approaching zero at the transmitted frequency. Likewise, the bandpass filter in the transmitter leg is so constructed that at the junction it exhibits the characteristic admittance of the line at the center frequency of the transmitted signal, yet presents an admittance of essentially zero at the received frequency. Therefore, practically speaking, the receiver leg does not exist as far as the transmitter frequency is concerned nor does the transmitter leg exist as far as the received frequency is concerned.

Figure 6 is the parallel combination; a series coupling is also possible but is not often used.

It is evident that the isolation between transmitter and receiver is directly related to the quality of the bandpass filter used in each leg. From the standpoint of admittance matching, the object of the design is to keep the susceptance of the off resonant branch as small as possible as well as having the conductance essentially zero. With respect to the resonant frequency of each branch, the insertion loss is desired to be maintained at a minimum over a wide band, which the attenuation of the resonant frequency of the opposite leg be maintained at a maximum. It can be seen that the bandpass filters are required to have a fairly flat top as well as having very good selectivity, that is, sharp skirts to the response curve.

Essentially, then, the duplexer is merely two bandpass filters terminated in the transmitter and receiver, respectively, and connected to the antenna in parallel. The coupling between the transmitter leg and the receiver leg is therefore a function of the characteristic of these band-pass filters and the separation in frequency between the transmitter leg and the receiver leg.

This type of duplexing can be extended to include almost any arbitrary number of transmitters and receivers operating in this fashion, with each addition to the junction being characterized by its own particular band-pass filter. Systems called multicouplers as described by Cline and Schiffman [1959] have been devised where as many as 20 inputs, transmitters, and/or receivers, can be coupled into the same antenna. The method is exactly the same except that the antenna must be capable of more and more broad-banded operations as the number of inputs increase. Where many of these inputs are required, a simple junction as illustrated in figure 6 will not suffice, and these inputs usually are distributed along a transmission line. This requires more complicated matching since the sections of transmission line between the inputs can be critical as far as length is concerned with respect to the operating frequency of each input. Line stretchers can be used between the inputs to tune the overall complex for the optimum operation over the frequency ranges desired.

Attention will be concentrated on the simple duplexer whereby a single transmitter and a single receiver are coupled to the antenna. Since, in the usual case, the duplexer is not field tunable and in manufacturing they are fixed tuned to operate over the required frequency band, it is not necessary to consider the different methods whereby this goal is achieved. It is sufficient to refer to the manufacturer's instruction book to ascertain the methods used to achieve this result. A duplexer can be made in different ways. Probably the most common utilizes coaxial transmission line for the bandpass filters. From the aspect of simplicity of design, the duplexer utilizing waveguide is advantageous since the filters can be built into the waveguide by means of irises. Irrespective of the type considered, the principles are exactly the same; therefore, what is considered with respect to waveguide duplexers applies equally well to coaxial duplexers. From the aspect of measuring the characteristics of such a duplexer, it can be considered to be a 3-port black box with an antenna terminal, a receiver terminal, and a transmitter terminal.

Other similar duplexers are described by Cohn and Coale [1956] and Lewis and Tillotson [1948].

3.1. The Operating Characteristics of the Duplexer

a. The Insertion Loss or Coupler Factor

The measurement of the coupling factor between the antenna and the transmitter and receiver ports, respectively, is a measure of insertion loss of the attenuation factor of the duplexer. A method for measuring this insertion loss or coupling factor will take into consideration any directive qualities which the duplexer may have inherent in its design. Essentially such a method consists of the following: a monitored signal is fed from the antenna into the duplexer and is monitored at the receiver port. Conversely, a signal can be injected into the duplexer at the transmitter port and monitored at the antenna port. The difference between the input and output signals is a measure of the attenuation or the coupling factor of the duplexer. A receiver whose power level can be monitored is used as a detector on the receiving leg, and a signal generator tuned to the center frequency of the receiving circuit is connected to the antenna terminal. A signal at the center frequency of the receiving system whose level is accurately monitored at the signal generator is then fed through the antenna port and received at the receiving port. The level of the signal is then recorded at the receiver. If an output meter is not available at the receiver, then some provision can be made to measure the AGC voltage, and this can be used as the level determining reading. The signal generator is then connected directly to the receiver, and the circuit remains the same with the exception that the duplexer is removed from it. Again the same monitored signal would be received at the receiver, and its level would then be noted. The ratio of these two levels would then be the insertion loss or the coupling factor for the duplexer. The same method would be used on the other leg of the duplexer with a signal generator connected to the transmitter port and the receiver connected to the antenna port. If an insertion loss greater than 1/2 db is experienced, the duplexer may not be in proper operating condition. Since no provision is present to adjust the duplexer, then an excessive insertion loss could be interpreted that the duplexer was not well designed. It should be remembered that for this measurement the port which is not being measured must be terminated in the characteristic impedance of the line.

b. The Bandpass Characteristics

The method of measuring the insertion loss of the dupelxer may be extended to measure the bandpass characteristics of the duplexer if a variable frequency receiver in the range of operation is available. Since the usual 1/2-db bandwidth of the duplexer should be known, a number of points within the region of the center frequency should be sufficient to determine the flatness of the curve over its desired bandwidth. The signal generator could be tuned to several discrete frequencies and the receiver tuned to meet these frequencies. A comparison of the output of the receiver is then graphed to determine the bandwidth characteristics of the duplexer. It is recommended that these measurements include frequencies other than those contained within the 1/2-db bandwidth. The nature of the selectivity of the filter can be well estimated if the 20-db bandwidth is measured. This would entail measuring several additional points where the receiver response may be down 20 db below the response experienced at the center frequency.

RELATIVE PASS BANDS OF THE TRANSMITTER AND RECEIVER LEGS OF A DUPLEXER

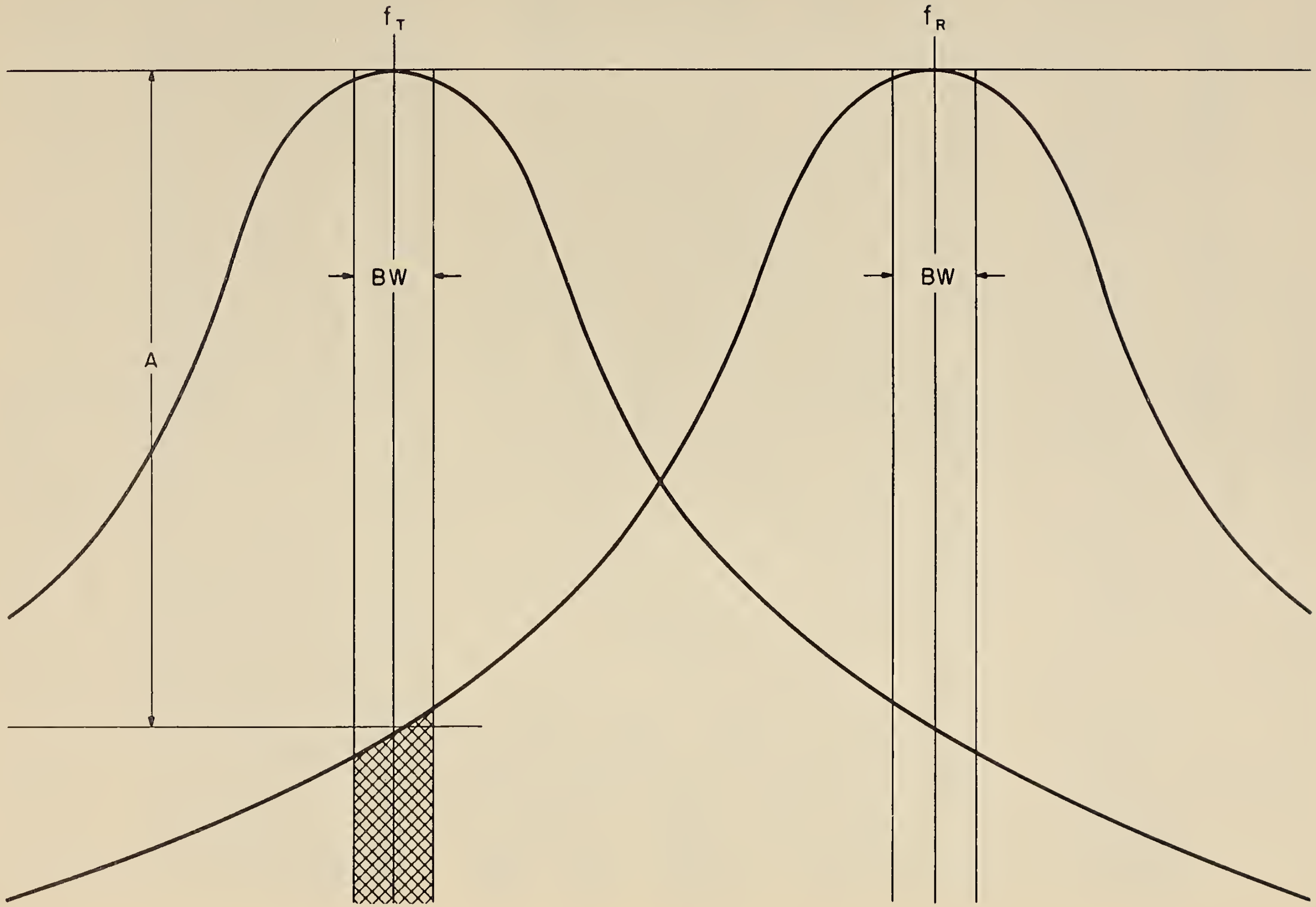


Figure 7

c. The Isolation

Since the attenuation necessary between the transmitter and receiver leg is extremely high, that is, in excess of 100 db, ordinary methods of measuring this isolation will not be adequate. Performance of the duplexer can best be determined while it is in the actual operating circuit. The necessity for a good degree of isolation can be illustrated by consulting figure 7. The response curve to the left is the idealized frequency response curve of the transmitter leg. The response curve to the right is the idealized frequency response curve of the receiving leg. It can be seen that the response of the receiver leg to the transmitter frequency should be maintained at a minimum. The response of the receiver leg to the transmitted energy is illustrated by the shaded portion in the transmitter bandpass. The amount of area under this curve is indicative of the amount of energy from the transmitter that appears in the receiver leg. The isolation of the receiver leg from the transmitter leg is indicated by the distance A, the attenuation from the peak power of the transmitter. This attenuation should be maximum. Two effects will follow if this attenuation A is not sufficient. One is that increased noise will appear in the bandpass of the receiver and, secondly, if the energy associated with this unwanted signal in the receiver leg is sufficiently high, it can effectively block the receiver at the rf stages. This situation is idealized in that it assumes all the energy from the transmitter will be contained within the indicated bandwidth, which, of course, is not necessarily true.

The method to be utilized here will be an approximate determination of the change in noise level and in minimum detectable signal caused by the duplexer. The receiver will be connected to the receiver port, the transmitter to the transmitter port, and the antenna to the antenna port of the duplexer. Then, with the transmitter off, a recording of the noise output of the receiver should be made; that is, the receiver, after the initial warm-up time has elapsed, should be allowed to operate merely on the noise from the system. Then the transmitter should be turned on and should be adjusted to maximum operating power with full modulation. The noise level of the receiver should then be again recorded for comparison with the first recording. If no discernible difference (i.e., less than 1 db) is evident between the two recordings, then it can be assumed that the isolation between the transmitter and the receiver is sufficient.

Blockage of the receiver due to excessive energy from the transmitter will be very apparent in that the receiver will tend to saturate, causing the noise level of the receiver to change markedly, indicating the duplexer to be inadequate.

This measurement of the output of the receiver should be made immediately after the IF amplifier or immediately after the detector. Then the degree of isolation is not only a function of the duplexer but is also a function of the selectivity of the receiver.

If a significant difference is noticed between the two recordings of the noise level of the receiver, then it is still to be determined whether or not this increased noise in the receiver on the second reading was due to the lack of isolation of the duplexer or due to the leakage and radiation of the transmitter. A method to determine this can be outlined as follows. The receiver is disconnected from the duplexer, and both the receiver and the duplexer are terminated in good quality matched terminations. With the receiver input then terminated with the matched load, a record of the noise level is obtained. The transmitter is then turned on and adjusted to full power and full modulation. A second record of the noise level of the receiver is obtained. If there is any difference between these two readings, i.e., greater than 1 db, then it can be assumed that stray coupling between the transmitter and the receiver is excessive, and the determination of the performance of the duplexer would be in error. It is extremely important for the accuracy of this determination that grounds be maintained between the duplexer and the transmission line to the receiver such that the paths of stray coupling experienced while the duplexer was in the system will not be disturbed. If excessive coupling is noted between transmitter and receiver when they are <u>not</u> connected through the duplexer, then steps must be taken to eliminate this stray coupling problem before the duplexer can be measured. After the problem of stray coupling has been eliminated, the measurement of the isolation, as outlined above, can be accomplished. If the increase in the noise level due to operation through the duplexer is in excess of 1 db, the isolation should be considered inadequate.

4. COMPARISON OF DIFFERENT TRANSMISSION SYSTEMS

Generally, most tropospheric scatter communications systems will employ either frequency modulation (FM) or single sideband suppressed carrier, amplitude modulated (SSBSC-AM). While other transmission systems are used such as pulse code modulation (PCM), only FM and AM modulation systems will be considered.

4.1. Methods of Modulation

Information can be transmitted on a carrier by varying any of the parameters of a sinusoidal wave in accordance with a modulating voltage. Thus, a carrier is described by

$$M(t) = A_c \cos(\omega_c t + \phi_c) \tag{16}$$

where

$M(t)$ = instantaneous value of the carrier at time t

A_c = amplitude of the carrier

ω_c = the angular frequency of the carrier = $2\pi f$ where f is the frequency

ϕ_c = the phase angle of the carrier when $t = 0$

a. Amplitude Modulation (AM)

Amplitude modulation is accomplished by letting the modulating signal vary the amplitude of the carrier wave about its mean value. Thus, A_c (equation (16)) is caused to vary in accordance with the modulating signal. If we represent the modulating signal as a time varying function V(t), then the amplitude modulated carrier is

$$M(t) = A_c [1 + K V(t)] \cos(\omega_c t + \phi_c) \tag{17}$$

where

K = a constant determined by the modulator

If a sinusoidal modulating wave is assumed, that is,

$$K\,V(t) = a \cos \rho t \tag{18}$$

where

a = amplitude of the modulating signal whose maximum value is 1.0

then

$$M(t) = A_c \left[1 + a \cos \rho t\right] \cos (\omega_c t + \phi_c) \tag{19}$$

b. Single-Sideband Modulation (SSB)

If (19) is expanded, it will be noted that three terms appear.

$$M(t) = A_c \cos (\omega_c t + \phi_c) + \frac{aA_c}{2} \cos (\omega_c t + \phi_c + \rho t) + \frac{aA_c}{2} \cos (\omega_c t + \phi_c - \rho t) \tag{20}$$

The first term is the carrier, and the other two terms are the sidebands. Single-sideband (SSB) systems suppress the carrier and one of the sidebands such that the output contains but one of these components. The carrier can be suppressed by use of a balanced modulator such that the carrier does not appear in the output. The two sidebands are then passed through a special SSB filter which passes one and rejects the other.

Then (20) becomes

$$M(t) = \frac{aA_c}{2} \cos \left[(\omega_c + \rho)\, t + \phi_c\right] \tag{21}$$

c. Angle Modulation

Angle modulation has two principal subdivisions: phase modulation and frequency modulation. Since true phase modulation does not utilize the spectrum as effectively as frequency modulation, phase modulation, today, is used primarily as a means of obtaining frequency modulation.

d. Phase Modulation

Phase modulation is defined as angle modulation in which the instantaneous phase angle of the carrier is varied in accordance with the amplitude of a modulating wave, and the frequency deviation, F_d, or the change in the carrier frequency, is proportional to both the amplitude and frequency of the modulating wave. Thus, the phase modulated carrier is given by

$$M(t) = A_c \cos [\omega_c t + \phi_c K V(t)] \tag{22}$$

where

$V(t)$ = time varying modulating wave

K = a factor of proportionality determined by the design of the modulating system

If the modulating signal is a sine wave, then (22) becomes

$$M(t) = A_c \cos [\omega_c t + a \phi_d \cos \rho t] \tag{23}$$

where

ϕ_d = phase shift produced by the maximum amplitude of the modulating signal

a = relative amplitude of the modulating signal, which can vary from 0 to 1.

It can be seen that the instantaneous frequency deviation is a function of both the amplitude and frequency of the modulating signal.

e. Frequency Modulation (FM)

By contrast, in FM, the instantaneous frequency of the carrier, the time derivative of the instantaneous phase angle $\theta(t)$ is varied by the amplitude of the modulating wave. The frequency deviation is proportional only to the amplitude of the modulating wave. The frequency of the modulating F_m determines the rate of change of the carrier frequency. Frequency modulation requires that the instantaneous carrier frequency be varied about its average value $f_o = \omega/2\pi$ in proportion to the amplitude of the modulating signal, and between the limits of $f_o + F_d$ and $f_o - F_d$ is the frequency deviation or peak swing of the carrier produced by the maximum values of the modulating signal. If a sinusoidal wave with an angular frequency ρ is used to modulate the carrier, the instantaneous carrier frequency is

$$f = f_o + a F_d \cos \rho t \tag{24}$$

where

a = the relative amplitude of the modulating wave which may have any value between 0 and 1.

The carrier can be described by $M(t) = A_c \cos(\omega_c t + \phi_c)$, where the value of the argument is dependent upon the following integration:

$$\phi_c = 2\pi \int a F_d \cos \rho t \, d t \tag{25}$$

$$\phi_c = 2\pi a F_d t \frac{\sin \rho t}{\rho t} \tag{26}$$

$$\phi_c = a \frac{F_d}{F_m} \sin \rho t \tag{27}$$

where F_m = frequency of the modulating wave = $\rho/2\pi$.

Then substituting (27) into (16), the equation of the frequency modulated wave is

$$M(t) = A_c \cos (\omega_c t + a \frac{F_d}{F_m} \sin \rho t) \tag{28}$$

Thus, an FM wave varies in instantaneous frequency above and below the center (carrier) frequency. Taking the limit of (26) as t approaches 0, the instantaneous carrier frequency $\omega_c t + 2\pi a F_d t$ shows that only the amplitude of the modulating signal determines how the FM wave varies from the center frequency. Equation (28) shows that the frequency of the modulating wave determines the rate of change of the carrier frequency.

As an example, if the amplitude of the modulating signal of one volt caused 100 kc/s change in frequency, the FM wave would swing from a maximum frequency of 100.1 Mc/s to a minimum frequency of 99.9 Mc/s. If the frequency of the modulating signal is 200 kc/s, this would happen at the rate of 200,000 times per second.

The amount by which the frequency differs from the center frequency is proportional to the amplitude of the signal. If the amplitude of the signal is doubled, the amount of frequency swing will double.

If the peak amplitude remains at one volt and the modulation frequency is doubled, the frequency will still vary between the limits of 100.1 and 99.9 Mc/s, but the frequency rate would increase to 400,000 times per second.

The frequency difference between the modulated wave and the center frequency is called the frequency deviation of the wave. The maximum amplitude of the modulating signal determines the maximum frequency deviation of an FM wave. In the final result the amplitude remains constant; only the frequency changes.

The reader is referred to Fibbs and Johnstone [1956], and Hund [1942] and Arguimbau and Stuart [1956] for a more comprehensive treatment of the processes and applications of frequency modulation.

4.2. Comparison of the Different Systems of Modulation

The three most common types of modulation, AM, SSB, and FM, can be compared. The comparison is made here not from the aspect of propagation but from the aspect of the equipment; the evaluation of systems of modulation with respect to propagation characteristics is considered by Florman and Tary [1962]. The comparison can be made on the basis of bandwidth, signal-to-noise ratio, frequency stability requirements, power requirements, distortion, efficiency, and simplicity of equipment.

a. Bandwidth Requirements

Bandwidth reduction is one of the most significant advantages of SSB over any other system of modulation. For example, two SSB voice circuits will occupy the same spectrum space as a single conventional AM voice circuit. Narrow-band PM and FM will require considerably more spectrum space than does AM.

An AM system occupies a bandwidth equal to twice the highest modulating frequency. For example, in an AM multiplex system of 24 channels plus 4-order wire channels, each channel requiring 4 kc/s, then the total bandwidth required is 2 x 28 x 4 = 224 kc/s.

Since the bandwidth of a single sideband system is only half that of AM, the required bandwidth for a comparable SSB system would be 112 kc/s.

The bandwidth required for an FM system is a function of the sidebands necessary to convey the desired intelligence with an acceptable amount of distortion. Even with a sinusoidal modulating signal, the expansion of (28) into the Bessel function is

$$M(t) = A_c \sum_{n=-\infty}^{\infty} J_n(m_\rho) \cos(\omega_c + n_\rho)\, t \qquad (29)$$

where

$$m_\rho = a \frac{F_d}{F_m}$$

$J_n(m_\rho)$ = Bessel function of the first kind, of order n for the argument m_ρ

(For mathematical simplicity, both positive and negative n must be considered for each sideband.)

The sidebands of an FM signal extend over an infinite bandwidth. All of these sidebands are not required. Practice has shown the distortion caused by clipping of sidebands is well within acceptable limits if only those sidebands with an amplitude greater than 1% of the unmodulated carrier amplitude are considered. The number of sidebands increase with the modulation index (m), where

$$m = \frac{F_d}{F_m} \tag{30}$$

From (30) it can be shown that as the modulation frequency (F_m) increases the sidebands decrease. Alternately, the sidebands increase as the frequency deviation (F_d) increases. The frequency deviation (F_d) is determined solely by the amplitude of the modulating signal.

For a modulating signal of constant amplitude, the bandwidth occupied is equal to twice the frequency of the highest significant sideband which is given by the number of significant sidebands multiplied by the modulating signal frequency. Thus, as the number of significant sidebands decrease with increasing frequency, the bandwidth tends to remain constant. The bandwidth is a function of the modulation index, and it can be shown that

$$B\,W = 2\,F_d \text{ if } F_{d(max)} >> F_m$$

$$B\,W = 2\,F_m \text{ if } F_{d(max)} << F_m$$

BANDWIDTH OCCUPANCY VERSUS MODULATION INDEX FOR FREQUENCY MODULATION

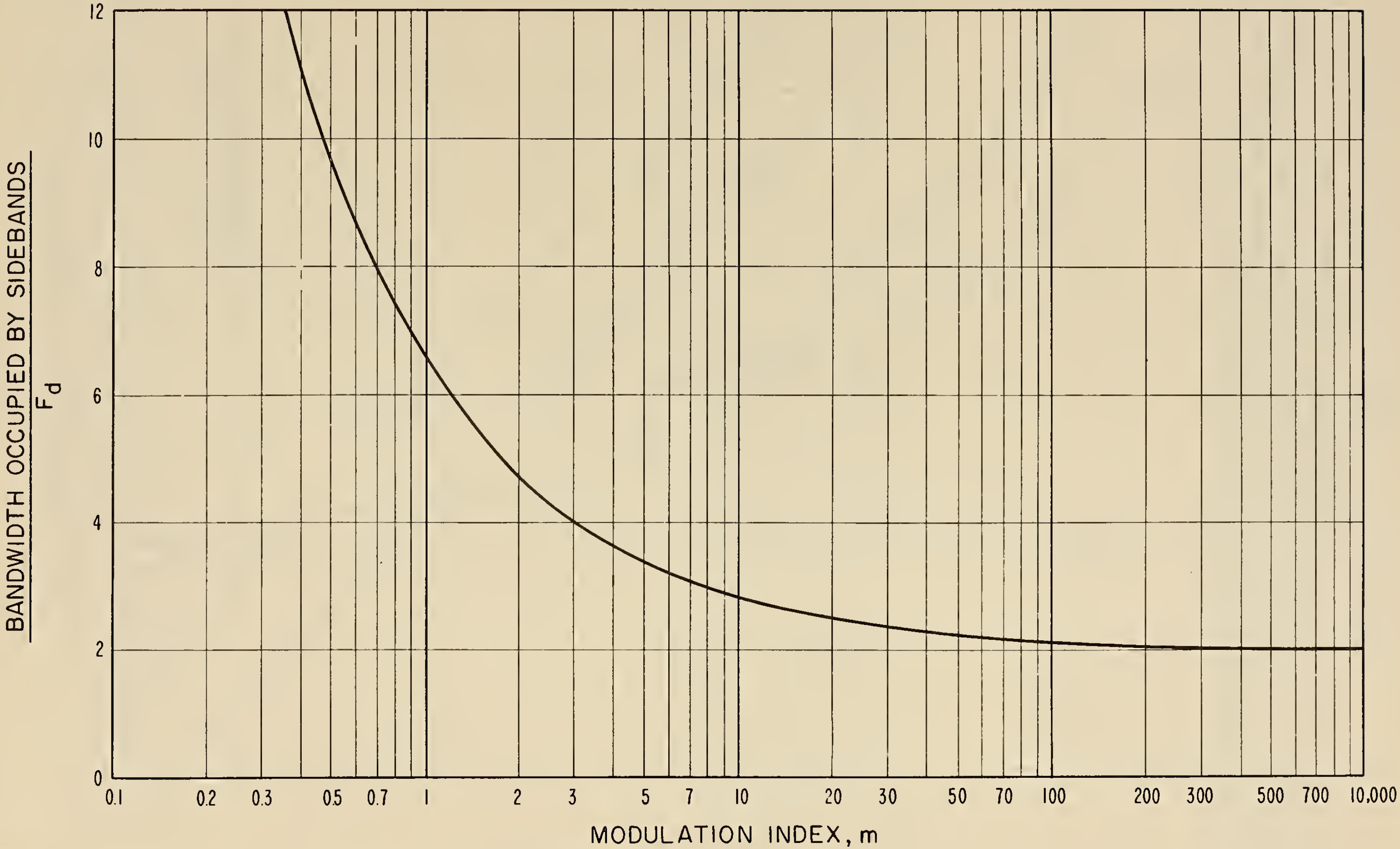


Figure 8

RELATIVE SIGNAL-TO-NOISE RATIOS OF
AM AND FM SYSTEMS
AS A FUNCTION OF FIELD STRENGTH

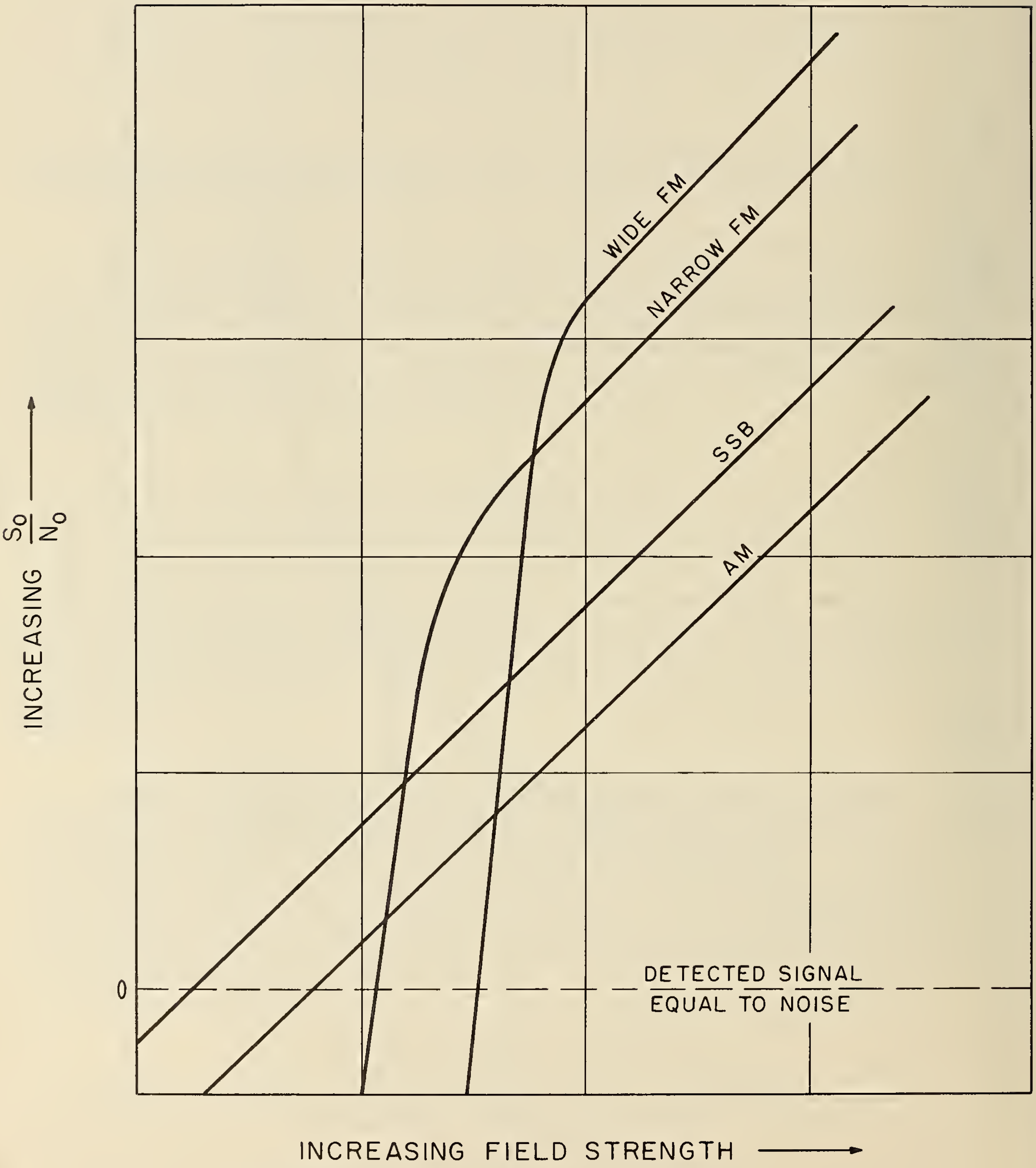


Figure 9

Figure 8 is an illustration of the relationship between the bandwidth, the frequency deviation, and modulating frequency. For example, if an FM multiplex system of 24 channels plus 4-order wire channels were desired, then for 4-kc/s channel bandwidths, the following can be found from figure 8:

m	=	2	m	=	0.6
F_d	=	224 kc/s	F_d	=	67.2 kc/s
B W	=	1052 kc/s	B W	=	590 kc/s

b. Signal-to-Noise Ratios

One of the primary advantages of FM over other forms of modulation is the improvement of the output signal-to-noise ratio for a given input as illustrated in figure 9. This advantage is accomplished only if the signal level is greater than the FM threshold level. An FM system requires a larger signal input for a minimum output than does the aforementioned systems. Black [1953] has shown that once this threshold level is exceeded, the output signal-to-noise ratio of an FM system can be expressed as

$$\left[\frac{S_o}{N_o}\right]_{FM} = 3\left[\frac{F_d}{F_m}\right]^2 \frac{P_c}{2F_m n} = 3m^2 \frac{P_c}{2F_m n} \tag{31}$$

where

P_c	=	the average carrier power
n	=	noise power density at the input to the detector in watts/unit bandwidth
F_d	=	frequency deviation
F_m	=	the modulating frequency

$2F_m n$ is then the noise power at the detector input. While it may seem that the signal-to-noise ratio could be made to increase indefinitely merely by increasing the frequency deviation (F_d), this is not so, since wider bandwidths imply increased power to maintain the signal level above the detector threshold.

The comparable signal-to-noise ratio for an AM system can be determined by considering an AM wave (equation (19)) applied to an ideal detector. The output signal will be proportional to the modulation. The output signal power can be expressed

$$S_o = c^2 a^2 P_c \tag{32}$$

where

c = a proportionality constant

a = the amplitude of the modulating wave

P_c = the average carrier power

The noise present in an AM system assuming an ideal low-pass filter of bandwidth B, and a purely random character for the noise is

$$N_o = 2c^2 Bn \tag{33}$$

where

n = the noise power density at the input to the detector in watts/unit bandwidth

B = bandwidth

Then the signal-to-noise ratio can be expressed

$$\left[\frac{S_o}{N_o}\right]_{AM} = \frac{a^2 P_c}{2B\,n} \tag{34}$$

If 100% modulation is assumed, a will be equal to one and

$$\left[\frac{S_o}{N_o}\right]_{AM} = \frac{P_c}{2B\,n} \tag{35}$$

Assuming an ideal detector, equal average carrier power, equal noise density at the detector, and with 100% AM modulation.

$$\text{FM versus AM} \quad \frac{\left[\dfrac{S_o}{N_o}\right]_{FM}}{\left[\dfrac{S_o}{N_o}\right]_{AM}} = 3 \, \frac{F_d^2}{F_m^2} = 3 \, m^2 \tag{36}$$

If the modulation index m is in excess of 0.6, FM systems are superior to AM, and the rate of improvement is proportional to the square of the modulation index.

FM provides a 4-to-1 advantage beyond that expressed above if peak powers are considered rather than average powers. Consequently, when the comparison is made between FM and SSB, this further FM enhancement must be considered since only peak power comparison is reasonable where SSB is concerned. This does not mean that FM enjoys this total increase over SSB, since SSB enjoys an 8-to-1 advantage over conventional AM. This indicates that the enhancement of FM or SSB is only half that which is enjoyed over AM.

$$\text{FM versus SSB} \quad \frac{\left[\dfrac{S_o}{N_o}\right]_{FM}}{\left[\dfrac{S_o}{N_o}\right]_{SSB}} = \frac{3}{2} \, \frac{F_d^2}{F_m^2} = \frac{3}{2} \, m^2 \tag{37}$$

If the modulation index m is in excess of 0.8, FM systems are superior to SSB, and the rate of improvement is again proportional to the square of the modulation index. However, this apparent enhancement should be viewed with caution. It is emphasized that this comparison is valid only when the signal is well above the threshold of the FM system. Especially in tropospheric scatter circuits where a continuously variable signal is present, the power requirement may be prohibitive if the received signal is going to be maintained above the FM threshold 100% of the time. This FM advantage must be considered in the economy of the particular circuit under consideration. If the signal is expected to fade below the threshold of the FM system, then this advantage vanishes and the performance of the FM system deteriorates rapidly to below that expected of other systems.

c. Frequency Stability Requirements

The requirements for frequency stability are quite stringent where SSB systems are used. These requirements arise due to the suppressed carrier. At the receiver the carrier must be reinserted for demodulation. It is very important that the exact relationship between the carrier and the sidebands be preserved. This places narrow limits upon the frequency variations of the oscillators both at the transmitter and at the receiver. For voice communications, a deviation of as much as 50 c/s can cause loss of intelligibility under conditions of poor signal strength. As the frequency of the transmitted signal increases, the percentage tolerances on the oscillator frequency become increasingly difficult to maintain. To overcome this difficulty, many SSB systems transmit a reduced carrier which furnishes the control for the frequency of the oscillator at the receiver.

Neither conventional AM nor FM has this tight frequency tolerance. The only requirement in the FM case is that the oscillator frequency variations must be small with respect to frequency changes caused by the modulating signal. A frequency stability of .002% is usually adequate for an FM system.

d. Power Requirements

Since in SSB only one of the sidebands is transmitted, there is a great saving in the amount of power necessary at the transmitter. This is equivalent to an increase in the effective power gain. Under comparable conditions, the SSB system provides a power gain of up to 9 db over a conventional AM system. Furthermore, an improvement in the signal-to-noise ratio can be expected since SSB requires lesser bandwidths.

The FM system is more difficult to analyze because good performance can be expected only as long as the input signal at the receiver is in excess of the threshold level. Near or below this level the performance deteriorates. Hence, for a fading signal, it can be expected that SSB will have approximately a 10-db average power gain over the FM system.

e. Distortion

Distortion in the high-power amplifier can become a very serious and limiting factor in the method of modulation. In all systems harmonic distortion will be present, but due to the resonant circuits in the output of the amplifier a minimum of these distortion products will be present in the output. A far more serious type of distortion is that which is produced when a linear amplifier is driven out of its linear range, as would be that case in SSB. Distortion products are produced which fall close to the desired signal and cannot be filtered out. These unwanted signals can destroy intelligibility. As a compromise, the tube is usually derated and operated over a smaller range, thus reducing its efficiency. FM, PM, and AM do not suffer from this deficiency, since the final amplifier in these systems does not require these linear characteristics.

f. Final Amplifier Efficiency

In general, AM tropospheric scatter circuits will utilize high level modulation. This places greater demands on the modulator because the modulator will be required to produce one-half of the power in the carrier. There is some compensation, however, since the final amplifier can then be operated in a class C condition, or, if the final amplifier utilizes a klystron, it can operate in the saturated region. This has the advantage that the efficiency of the final amplifier is at a maximum, and the amplifier can be operated at rated continuous carrier output. The efficiency of a saturated klystron power amplifier may be in the range of 40%.

In the case of SSB, the amplifier must be operated only over its linear range. In the case of the klystron, the peak power is usually limited to about 80% of the maximum saturated output, thus reducing excessive distortion. The efficiency will be dependent on the character of the modulating signal. For example, two tones, not synchronized in phase, will have an average-to-peak power ratio of 50%. However,

BASIC FM TRANSMITTER

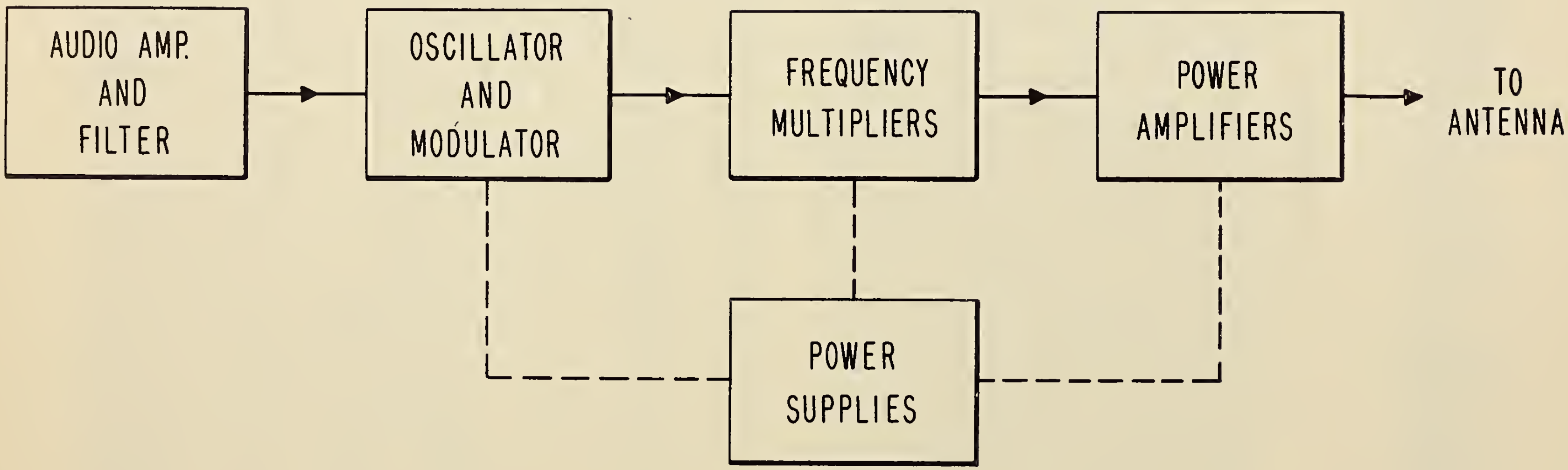


Figure 10

as the modulating signal becomes more complex, this ratio decreases. Gaussian noise, which more closely approaches voice, reduces this ratio to approximately 15% for the 1% peaks.

Where FM and PM systems are used, the final amplifier is usually operated at saturation. However, the maximum efficiency may be reduced if the klystron is broad-banded to accommodate the wider bandwidths required in FM. In general, under comparable conditions FM will display a 6-db advantage over AM if the peak power output of the amplifier is the limiting factor, and a 3-db advantage if the limitation is in terms of the rms power output.

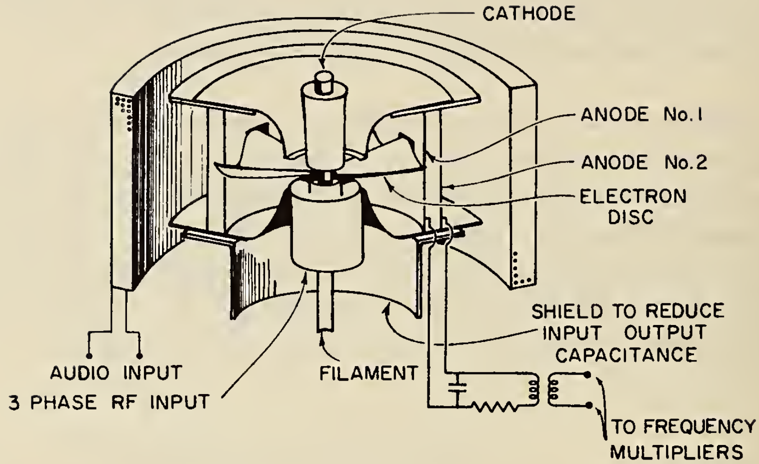
5. THE FM TRANSMITTER

In order to transmit information as electromagnetic wave it is necessary to convert intelligence to a modulating signal, to superimpose this modulation on a carrier wave, and to deliver this modulated wave to the antenna system. FM transmitters are largely conventional and consist of three basic parts: the oscillator-modulator, the frequency multiplier, and the power amplifier. In the first section the carrier frequency is generated and frequency modulated, in the multiplier section the carrier is increased in frequency to that of the operating frequency, and in the amplifier section the power is increased to the desired operating level as shown in figure 10.

5.1. Oscillator and Modulator Sections

Frequency modulated waves are produced by varying the carrier frequency at the modulating signal rate above and below the unmodulated carrier frequency. This can be achieved by using a reactance tube or a voltage variable capacitor to alter the frequency of the transmitter master oscillator. The reactance tube or voltage variable capacitor is controlled by the frequency of the modulating signal. Phase modulation is produced by allowing the modulating signal to shift the phase of the carrier, the change of the phase angle being proportional to the instantaneous amplitude of the modulating signal. Phase modulation, as such, is seldom used since it does not utilize a given frequency channel as effectively as does FM transmission. However, phase modulation is important as a means of obtaining frequency modulation,

PHASITRON



PHASE DEVIATION OF THE OUTPUT VOLTAGE
WITH RESPECT TO FIELD STRENGTH OF THE MODULATING COIL IN GAUSS

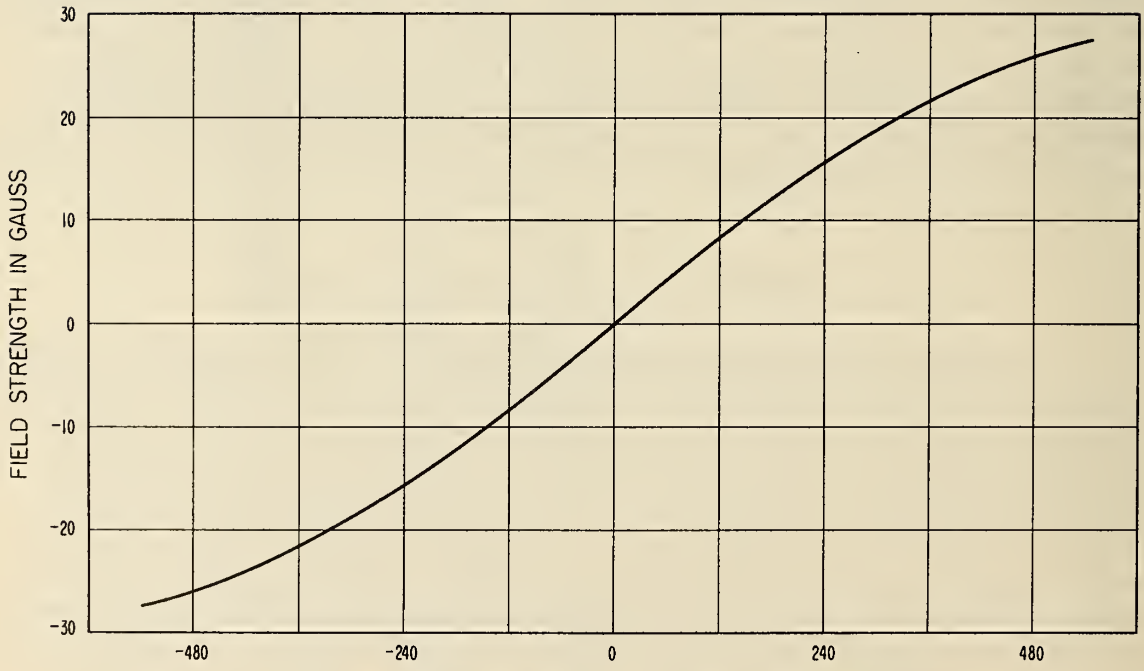


Figure 11

the phase modulator requiring an integration circuit or a pre-emphasis network to remove the effect of the frequency of the modulating signal on the total frequency deviation and produce frequency modulation.

A direct method of generating FM is to control the frequency of an ordinary oscillator such as the hartley oscillator by a modulating voltage. The most common method of doing this is by means of a reactance tube, although the voltage variable capacitor has a promising future in such applications. If a reactance tube operates on a crystal-controlled oscillator, the resulting modulating is closer to phase modulation than to frequency modulation because the frequency deviation that can be secured by varying the tuning of a crystal oscillator is quite small. Although a reactance tube operating on a free-running oscillator is capable of producing wide deviation, the frequency of such a transmitter is susceptible both to variations in power supply voltages and to changes in parameters due to aging and temperature. Such susceptibility results in very poor frequency stability. This can be improved through the use of push-pull reactance modulators, but in the end the desired stability must be achieved through the use of automatic frequency control in which the average transmitter frequency is compared against a crystal-controlled oscillator. This results in considerable added complexity to the transmitter. The sensitivity of the reactance tube modulator depends on the transconductance of the modulator tube and certain circuit factors, but in general this modulator is capable of causing a frequency deviation of several hundred cycles with a distortion of a few hundredths of a percent at the oscillator frequencies commonly used.

This problem of frequency stability can be avoided through the use of a fixed frequency oscillator and appropriate phase shifting network on the output of the oscillator controlled by the modulation. One such method utilizes a specially designed tube called the Phasitron to introduce comparatively wide phase variations in the crystal-controlled RF carrier, as described by Bailey and Thomas [1946]. The Phasitron is capable of producing phase deviations of up to $\pm 450^{\circ}$ with a distortion of 1.35% at 50 c/s (see figure 11). Phase deviations of $\pm 200^{\circ}$ or 3.5 radians are commonly used. With this type of service, distortion is about 1% and the frequency swing $\pm$ 175 c/s, with a modulating frequency of 50 c/s. The noise is more than 78 db below the signal at full deviations (100% modulation).

SERRASOID MODULATOR

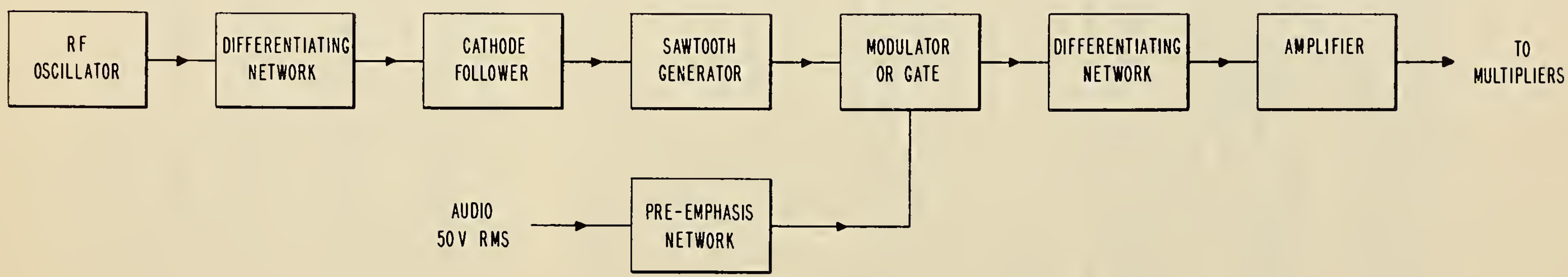


Figure 12

The Phasitron is so constructed that a sinusoidally varying electron disc is formed. At the periphery of this disc are concentric anodes The inner of the two anodes is punched with holes arranged in such a manner that the rotating disc of electrons strikes the inner anode half the time, and half the time it passes through and strikes the outer anode.

Thus, if a tuned circuit is connected across the two anodes, it will be excited at the oscillator frequency. The Phasitron is surrounded by a varying magnetic field produced by the modulating frequency in a solenoid. The action of this field is such as to advance or retard the rotation of the disc about its axis and to produce in its rotational motion either a slight increase or a decrease in its peripheral velocity.

The modulating solenoid is very nearly a pure inductance with a constant voltage applied to it. The resulting current through the coil and the rotation of the electron disc are then inversely proportional to the modulating frequency. Thus a constant deviation is acquired and FM results.

The Serrasoid modulator as described by Day [1948] is widely used today. It can provide phase shifts of up to about ± 150° or a corresponding frequency excursion of ± 130 c/s with a distortion of a few tenths of a percent. The noise originating in the modulator can be kept about 80 db below that level obtained with 100% modulation. By using some harmonic of the output pulse, Serrasoid modulators have been successfully used at rf frequencies as high as 5 Mc/s and with modulating frequencies as high as 300 kc/s.

The Serrasoid modulator, (see figure 12) consists of 3 tubes which generate a stable, very linear sawtooth wave and a modulated gate which generates a pulse whose time phase is proportional to the instantaneous audio signal. The rf oscillator draws plate current during only a small part of the cycle. This signal is then differentiated and clipped by a cathode follower. These pulses are used to time the sawtooth generator. Because the linearity of the modulator depends upon the straightness of the sawtooth wave, great effort is made to achieve the desired waveform. The sawtooth generator is directly coupled to the modulators or gate tube. This tube is normally cathode-biased so that conduction begins when the sawtooth wave reaches about half its peak value. Once conduction starts the sawtooth is clipped. The bias of the modulator is varied by the modulating signal, thus varying the clipping level of the sawtooth wave and the time of conduction of the modulator tube. The negative going portion of the wave is differentiated and amplified. The phase of the resulting pulse can thus be advanced or retarded because the audio signal alters the bias on the modulator tube. These pulses are then applied to the frequency multipliers and reshaped to sinusoidal waves.

Table 2 illustrates the comparison among the different types of modulators considered here.

TABLE 2

Type of Modulator	Deviations Obtainable		Distortion	Noise in db Below Full Modulation	Comments
	Phase in degrees	Frequency in c/s			
Reactance tube acting on a free-running oscillator		± 9000	< .05%		Poor stability without complicated stabilizing circuitry, excellent deviation, usable at high rf frequencies
Reactance tube acting on an amplifier tank circuit	± 25	± 22			Poor deviation
Phasitron	± 450	± 390	< 1.5%	> 78	Good deviation, low noise
Serrasoid	± 150	± 130	< 1%	80	Low noise, good high frequency response, usable at high rf frequencies.

Most manufacturers have their own favorite modulating methods. In addition to those listed above, these may include the use of cathode-ray tubes or methods for varying the transconductance of a tube. In a specific piece of equipment, the manufacturer's instruction manual must be consulted to determine the type of modulation used and to obtain detailed information as to the action of the circuit and its adjustments. The reader is referred to Terman [1947], Camies [1959], Fibbs and Johnstone [1956], and Black [1953], among others, for additional information.

5.2. Frequency Multipliers

Modulation is accomplished at low rf power and in most cases at a comparatively low frequency. The modulator output is used to drive a series of low level, receiving tube, class C, multiplier stages. In the final stages of the multiplier the power level is increased to that necessary to drive the power amplifier.

The frequency multipliers perform one additional important service. The modulators produce frequency shifts of relatively small magnitude, and this is increased to the desired frequency deviation in the multipliers. The deviation is given by

$$F_d = Mf_d \tag{38}$$

where

M = the frequency multiplication factor of the multiplier chain

f_d = the frequency deviation produced by the modulator

5.3. Power Amplifiers

The power amplifiers used in tropospheric scatter systems are either conventional tubes or klystrons. Since linearity is not required in the power amplifiers in FM transmitters, conventional tubes are operated class C, and klystrons are operated at the peak of their mode curves with high efficiency. The power output of an FM transmitter may usually be increased at any time by the addition of larger power amplifiers to the existing equipment.

5.4. Power Supplies

The requirements of the power supplies for an FM transmitter are not particularly stringent. Aside from the oscillator and modulator section, the regulation of the power supply is not important since the load does not change with modulation. Ripple should be low (0.1% or better).

The oscillator and modulator power supplies should be well regulated and have very low ripple since poor power supply performance at these points can alter the modulation or the frequency of the transmitter.

5.5. Measurement of Frequency Deviation in an FM Transmitter

It is important to know the frequency deviation of an FM transmitter for several reasons. Perhaps one of the most important is that the rf spectrum occupied bears a direct relationship to the frequency deviation. The relationship between the modulating index, frequency deviation, and the modulating signal is given by

$$m = \frac{F_d}{F_m} \tag{39}$$

where

m = modulating index

F_d = frequency deviation

F_m = frequency of the modulating signal

Elaborate equipment has been designed for the purpose of measuring frequency deviation. If such equipment is available, it is simply a matter of following the instruction manual supplied by the manufacturer, to determine the frequency deviation. If such equipment is not available, or if the results found by the use of such equipment are to be subjected to confirmation, then the following method is recommended. The Bessel zero method will allow the determination of the frequency deviation using only an audio oscillator and a receiver tunable to the carrier frequency of the transmitter. The receiver must have an internal beat frequency oscillator (BFO).

For certain values of modulation index, specifically 2.405, 5.520, 8.654, 11.792, 14.931, etc. (where the Bessel function passes through zero), the total transmitter power is contained in the sidebands, none being present in the carrier. It is possible by increasing the amplitude of the modulating signal and thus varying m, to cause the carrier of the FM signal to disappear. The frequency deviation can then be calculated from (39) where 2.405 (corresponding to the first disappearance of the carrier) has been substituted for m.

A standard communications receiver with a BFO is required, and the receiver is tuned to an unmodulated transmission from the FM transmitter. The BFO should be adjusted until a tone such as 500 cycles per second is heard in the output of the receiver. When the frequency of this beat is set, the BFO adjustment should be fixed and the ear allowed to become accustomed to this pitch. This is extremely important as later the frequency deviation will be varied by changing the amplitude of the modulating signal, resulting in a number of side-bands at various frequencies. This will produce many tones in the output of the receiver. All of these conflicting tones, with the exception of the desired 500-cycle tone, must be ignored at all times. This may require some practice. Next, a signal F_m from the audio oscillator is fed into the FM transmitter. As the amplitude of this audio signal is increased, the sideband frequencies and their innumerable beat notes are produced. However, as the amplitude of the audio signal is increased from zero, F_d increases, and a steady attenuation of the beat note arising from the carrier will occur. At some point this beat note will disappear completely and reappear once again as the amplitude of the audio signal is further increased. If the amplitude of the audio signal is "rocked" on both sides of the apparent zero, two points will be found where the beat signal can be just perceptible. The center point between these two positions will be the desired level.

With these factors in mind, it is now possible to determine the amplitude of a given audio signal necessary to produce the maximum allowable frequency deviation. The first step is to calculate the modulating frequency that will produce the maximum allowable frequency deviation when the modulation index is 2.405. This can be obtained from (39). The audio oscillator should be set to this frequency. If the computed value of modulating frequency is too high to be within the audio bandpass of the transmitter primary stages, or if it is too high for the audio oscillator, then a value 5.520, 8.654, etc., should be used for m corresponding to the second, third, etc., occurrence of carrier disappearance.

When acceptable modulating frequency can be obtained, this is applied to the transmitter from the audio oscillator. The amplitude of the modulating signal is increased slowly until the audio beat of the carrier disappears appropriately. When this occurs, the maximum allowable frequency deviation F_d has been reached and the corresponding amplitude of the modulating signal can be read using a vacuum-tube voltmeter. This level must not be exceeded during the operation of the transmitter. If some value of m other than 2.405 such as 5.520 or 8.654 were used, the first disappearance of the carrier beat note corresponds to a relatively low frequency deviation. The amplitude of the audio signal must be increased further until the carrier has disappeared the same number of times as the order of the value of m. In other words, the carrier should disappear twice for an m value of 5.520 and three times for an m of 8.654.

If the above method requires an excessively high value for m or if a receiver tunable to the transmitter frequency is unavailable, the following method may be used. In this case, a receiver with a BFO which will tune to the lower frequency signal immediately after the point of modulation is used. A portion of this modulated signal is coupled into the receiver. This may be done by means of a small pick-up coil coupled to a transformer. Since the frequency multiplying stages of the transmitter increase both the carrier frequency and the frequency deviations by the same factor, it is possible to determine the frequency deviation at the receiver pick-off point by using (39) and substituting for F_d the maximum frequency deviation which can be tolerated at this point and for m a value of 2.405, it is possible to determine the desired modulating frequency. This audio frequency is then applied to the transmitter and the amplitude of the audio increased until the beat note disappears. This amplitude of signal then corresponds to the maximum allowable frequency deviation. Here, again, a higher order of m such as 5.520 or 8.654 may be used, remembering that the same order of disappearance of the carrier must be used as is used for m. In other words, the carrier must disappear twice for a value of 5.520 for m or three times for a value of 8.654, etc.

TYPICAL UHF SSB TRANSMITTER

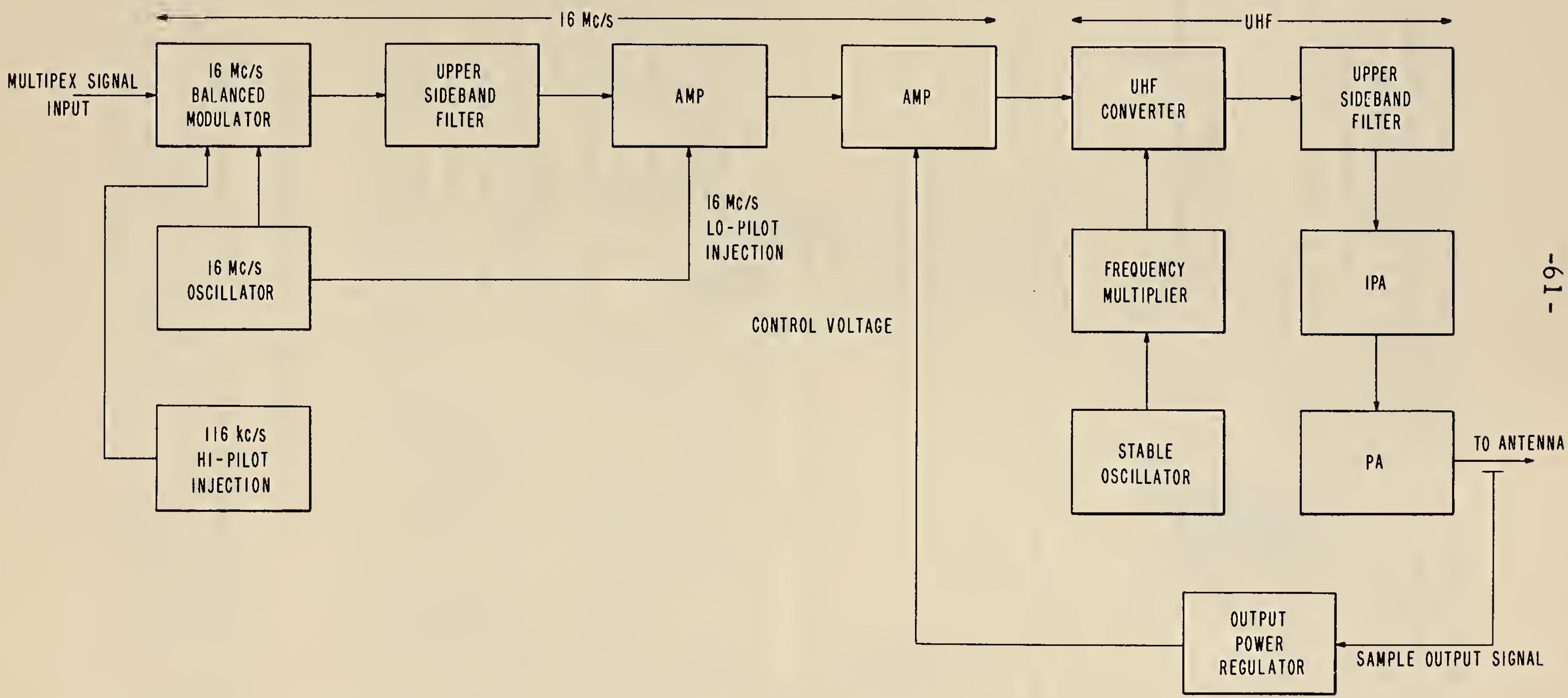


Figure 13

The previous method considers the modulating signal, both as regards amplitude and frequency necessary to produce the desired frequency deviation. It can be seen that this method can be used for many values of the modulating index m, and, consequently, care must be exercised in the selection of the value of m. The following example will illustrate the use of this method:

Example: Given an FM system whose maximum frequency deviation is to be ± 150 kc/s, the modulating frequency is to be 20 kc/s. Then using (39)

TABLE 3

m	F_d	F_m
2.405	150 kc/s	62.37 kc/s
5.520	150 kc/s	27.17 kc/s
8.654	150 kc/s	17.33 kc/s
11.792	150 kc/s	12.72 kc/s

From Table 3 it is evident that the value of m to be used must lie between 5.520 and 8.654, that is, between the second and the third order values of m. In this case the modulating frequency corresponding to m equal to 8.654 is sufficiently close that a good estimate will be found by using a single value of m. (Greater accuracy can be attained if interpolation is used.) Consequently, if the modulating signal generator is adjusted to 17.33 kc/s and if the amplitude is increased until the beat note disappears the third time, then the permissible level of the modulating signal is that signal level experienced when the carrier disappears the third time. If this signal level is not exceeded, then the frequency deviation will not exceed the imposed limits. Excessive frequency deviation can produce distortion at the receiver due to sideband clipping.

6. THE SSB TRANSMITTER

A block diagram of a typical multichannel SSB transmitter for operation in the UHF region is shown in figure 13. The transmitter consists of an exciter for translating the signal from the multiplex equipment to the UHF region and for amplifying it to a power level

BASIC RECTIFIER-MODULATORS

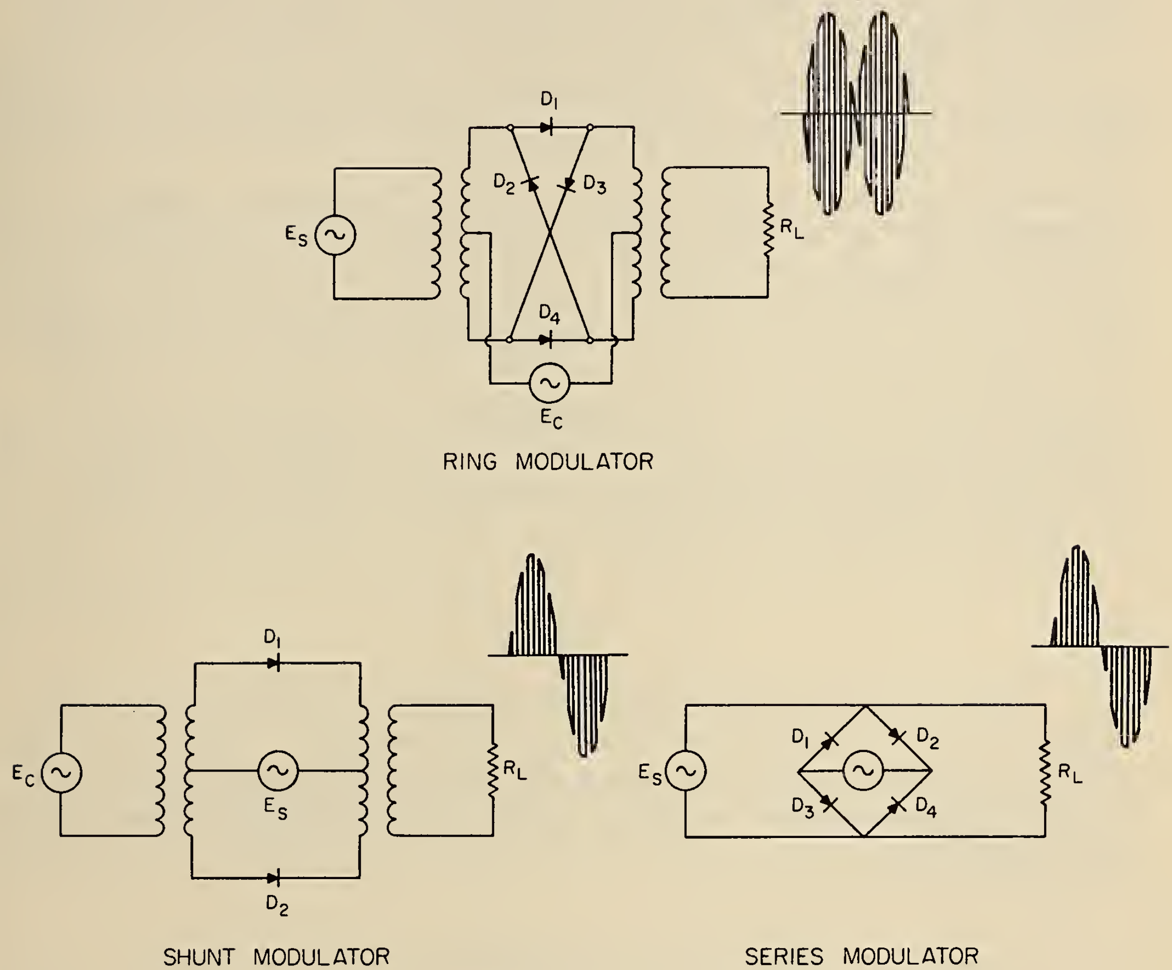


Figure 14

sufficient to drive the intermediate power amplifier. For one typical 24-channel system, the multiplexed signal occupies a baseband between 12 and 112 kc/s. A pilot tone is located at 116 kc/s. This composite signal is fed to a 16 Mc/s balanced modulator along with a 16 Mc/s carrier signal. The 116 kc/s pilot signal is used by the receiver for correction of amplitude variations due to selective fading. The desired sideband is selected by means of a high Q filter following the modulator. This output signal is amplified, and some of the 16 Mc/s carrier signal is reinserted and used in the receiver for automatic gain control (AGC), automatic frequency control (AFC), and demodulation of the SSB signal. The signal is again translated this time to the final operating frequency by the UHF converter. Tuned coaxial line resonators form the UHF filter which passes the upper sideband; additional filtering is provided in the intermediate power amplifier (IPA) and final power amplifier (PA). The UHF signal is then amplified to the desired power level in the IPA and PA. At the output of the PA, a sample of the signal is fed back to the 16 Mc/s amplifiers to provide automatic power output regulation of the SSB transmitter.

6.1. Modulators

Radio frequency sidebands are obtained by combining the audio signals and the rf carrier in an amplitude modulator. There are many modulator circuits; however, they can be grouped according to their circuit components into rectifier modulators and multi-element vacuum-tube modulators.

a. Rectifier Modulator

Rectifier modulators may be one of three general types--ring, series, or shunt, depending on the manner in which the diodes are connected in the circuit. The basic circuits for these modulators are shown in figure 14. In each circuit the rectifiers are made to work as switches by using large rf switching signals which greatly exceed the audio signal levels. These modulators are almost always connected as balanced modulators so that there is a minimum of the rf switching voltage in the modulator output. If the system is symmetrical, the carrier frequency will balance out and will be absent in the output.

The output of these modulators consists of a series of pulses whose repetition rate and polarity are determined by the carrier signal and whose amplitude is proportional to the audio signal. A spectrum analysis of these pulses would show an upper and lower sideband displaced by the modulating frequency about the carrier frequency. A similar set of sidebands will be placed about the second harmonic of the carrier and still more about the higher harmonic carrier frequencies.

Of the three types of rectifier modulators, the ring modulator has the highest efficiency where the output can be twice the output of the shunt or the series modulators. The split-ring modulator, where it is possible to balance independently two sets of diodes, can be used where carrier balance is important.

Rectifier balanced modulators have their greatest advantage in that they are more stable than vacuum-tube modulators. If they are made from good quality, well balanced components, intital carrier balances exceeding 40 db can be obtained with the level of the third-order intermodulation products holding 50 db below the desired sideband output.

b. Vacuum-Tube Modulators

Multi-element vacuum-tube modulators, whose output is a function of the instantaneous waveform of the modulating signal, fall into two types: the product modulator and the square-law modulator.

A product modulator output is proportional to the input modulating signal and to the carrier signal. An example of a product modulator is a double-grid vacuum tube where the carrier is impressed on one grid and the modulating signal on the other. Modulation takes place due to combined action of the grids and not due to any nonlinearities of the tube.

In the square-law modulator, modulation takes place as a result of the nonlinearity of the tube. An example of such a modulator is a triode in which the shape of the "grid voltage versus plate current" curve has at least second-order curvature. (This characteristic is present in all vacuum tubes and is a cause of distortion in amplifiers.)

If the carrier signal and the modulating signals are imposed on the grid of the triode, and if the tube has purely square-law characteristics, then the output signal will contain only the desired sum and difference frequencies with no other products except the second harmonics of the input signals.

Modulation also can take place by using a circuit in which the output is proportional to the polarity of the modulating signal. An example of such a modulator is a plate-modulated triode operating class C. The modulating signal is used to vary the plate voltage applied to the class C amplifier. The resulting output is a series of pulses whose repetition is that of the carrier frequency, and the amplitude is proportional to the modulating signal. A tuned circuit is then used to suppress the harmonics of the signal. A double-grid vacuum tube can also be used as a switching type of modulator by increasing the amplitude of the carrier signal applied to one of the grids. This signal can be large enough to drive the plate current of the tube to cutoff in one direction and saturation in the other, resulting in an output signal similar to the waveform of the rectifier modulator.

Vacuum-tube modulators have the advantage in that a conversion gain can be realized, whereas in the rectifier modulators a loss in gain is incurred. But vacuum-tube modulators are somewhat unstable as to gain and impedance, which makes them undesirable as balanced modulators. However, the product and square-law modulators are particularly useful as frequency changers since they generate a minimum of unwanted products.

6.2. Single-Sideband Generation

The two basic systems of generating single-sideband signals are the filter system and phase-shift system.

Both systems produce a single sideband of an amplitude modulated radio-frequency carrier, and both are capable of good overall system performance.

FILTER METHOD OF PRODUCING A SINGLE-SIDEBAND SIGNAL

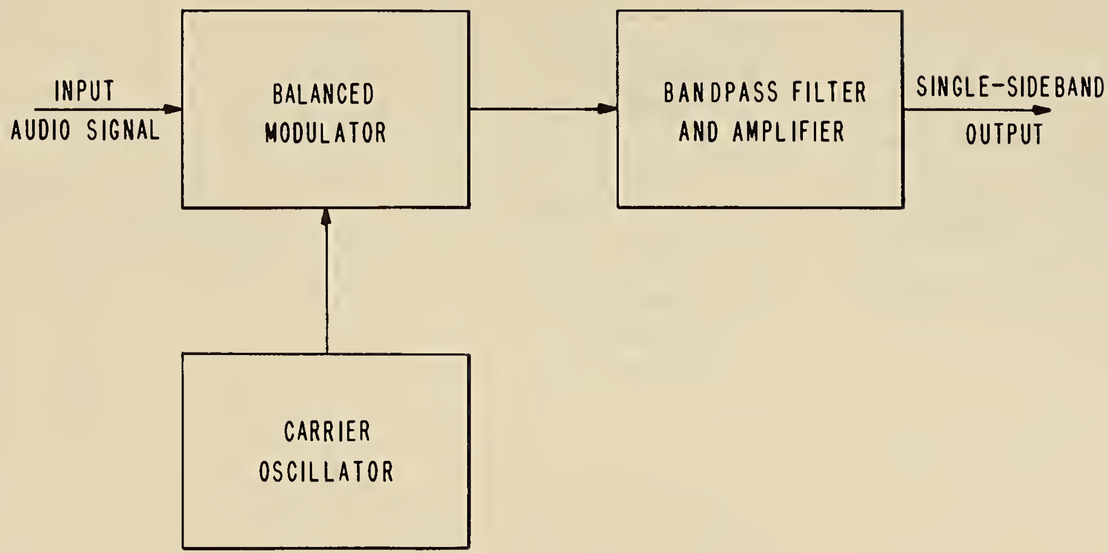


Figure 15

PHASE-SHIFT METHOD OF PRODUCING
SINGLE SIDEBAND

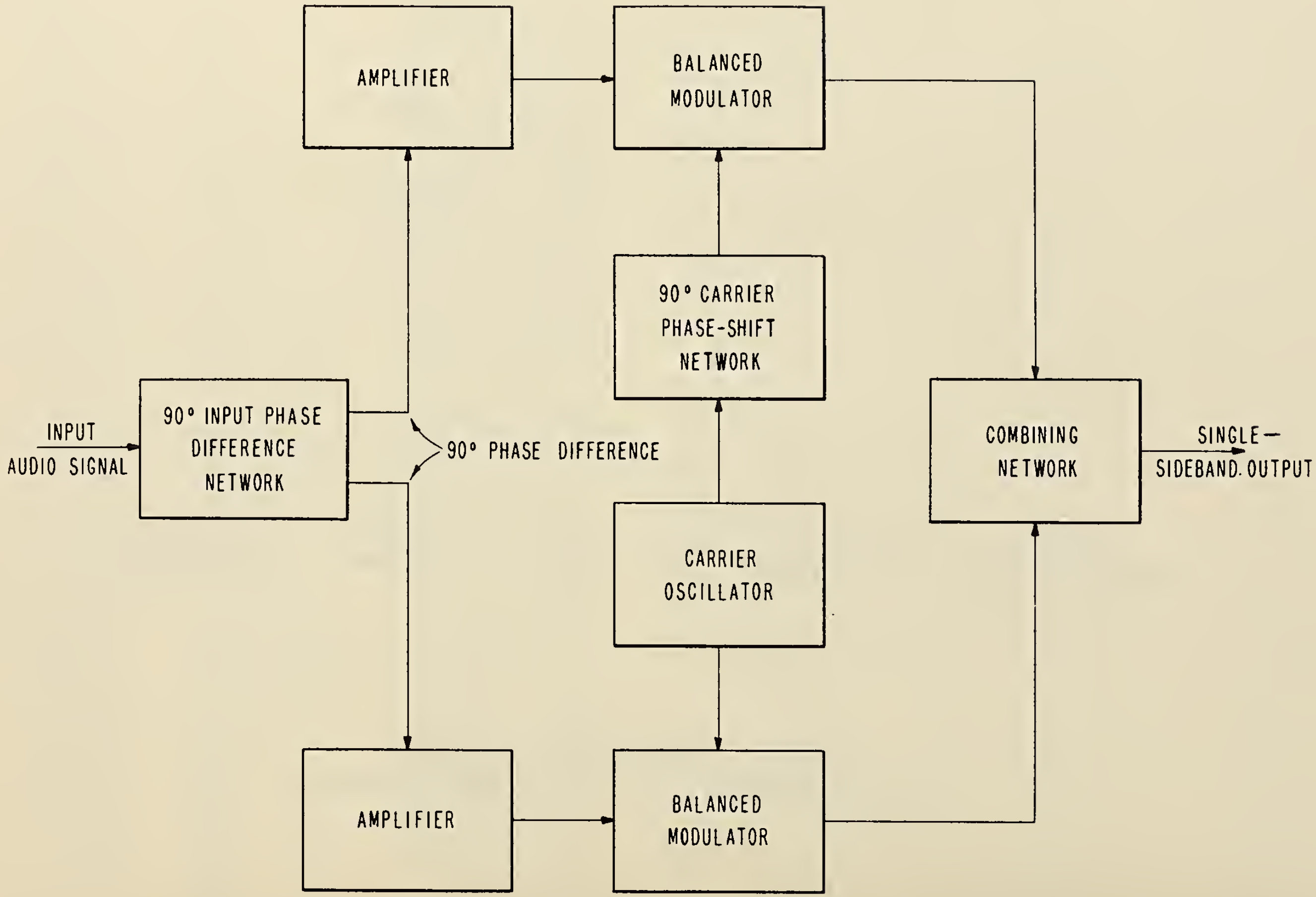


Figure 16

a. The Filter Method

The filter method of producing single sideband, which is the most familiar SSB generator, consists of a balanced modulator followed by a selective bandpass filter (see figure 15). The balanced modulator produces two sidebands. Theoretically, the rf carrier is balanced out and does not appear; however, practical design considerations determine the extent of the carrier suppression. Present balanced modulators can easily suppress the carrier to from 30 to 40 db below the peak envelope power (PEP) of the sidebands. Further suppression of the carrier by the SSB filter can result in an additional 20 db of carrier suppression. Total carrier suppression of from 50 db to 60 db can, therefore, reasonably be expected from the transmitter system.

The rapid increase in attenuation required of a sideband selecting filter in order that it may adequately suppress the undesired sideband, limits the frequency at which a single sideband signal can be produced by this method. Many inductors and capacitors have low Q factors and can be successfully used for sideband filters only at relatively low carrier frequencies, up to about 50 kilocycles per second. Mechanical filters have been built to operate up to 60 kc/s for a single voice channel, and crystal filters have been made to reject sideband and image responses as close as 1 to 2% of the desired frequency. This permits a first conversion frequency of 10 to 20 Mc/s in multichannel operation. With such filters, it is possible to construct bandpass filters at 10 to 21 Mc/s which pass a desired sideband extending from 10 to 100 kc/s or more from the carrier frequency while rejecting the other sideband by 40 db or more. The signal may then be heterodyned to the desired frequency and amplified to the desired power level. The major advantage of the filter method lies in its stability and in its suppression of the unwanted sideband.

b. The Phase-Shift Method

The phase method produces SSB through the use of a balancing process. In this system, a modulating signal is applied to two balanced modulators through a wideband 90° phase difference network. The rf carrier is also applied to these modulators through a 90° phase-shifting network (see figure 16). Therefore, each of the two input signals, carrier plus modulating signal, are applied to the balanced modulators in phase quadrature. The two double-sideband suppressed carrier signals thus generated will combine in the common plate circuit of the combining network. Since the signals are in phase quadrature, one set of sidebands will add while the other set will cancel. The amount of suppression of the unwanted sideband in the phase-shift system is primarily dependent upon the design and adjustment of the wide-band phase-shift network. This network must maintain the two audio input signals exactly 90^{o} displaced in phase over the entire band, and the amplitude of the two signals must be equal.

The significant advantage of the phase-shift method is that the undesired sideband is removed without the need for expensive filters. Furthermore, this method can be used at a higher rf carrier frequency, reducing and at times eliminating the number of heterodyning steps necessary to reach the operating frequency.

6.3. Oscillator Requirements

At the receiver the detection process consists of a frequency translation of the SSB rf signal down to the baseband spectrum. To accomplish this translation process the carrier which is suppressed prior to transmission must be reconsituted with the receiver signal at a frequency close to that of the original carrier. In practice, a deviation between the local and the transmitter carrier frequencies of as little as 50 c/s can cause loss of intelligibility of signals under conditions of low signal-to-noise ratios.

This implies a rather stringent requirement on the frequency control of SSB which results in a general increase in the complexity and cost of equipment. For example, if it is assumed that half of the maximum allowable 50-c/s departure between the local and the transmitter frequencies is distributed equally between the two ends of the path, then the minimum oscillator stability that can be tolerated would be 6.75 parts in 10^8 at an operating frequency of 400 Mc/s, whereas this tolerance shrinks to 2.5 parts in 10^8 if the operating frequency is 1000 Mc/s. Furthermore, if the signal is to be relayed again by SSB, then the stability requirements become even more restrictive.

Although the development of small ruggedized frequency standards is progressing rapidly and stabilities of one part in 10^8 are commonplace, some form of AFC is recommended. The method currently in use involves the transmission of some carrier power at all times but reduced to at least 20 db below the transmitter peak power. This signal is used as a slow acting AFC to control the frequency error at the receiver. This carrier of low pilot is used at the receiver for AFC and the demodulation of the SSB signal.

6.4. The rf Power Amplifier

The most important consideration in regard to the power amplifier of an SSB system is its linearity. In an rf tuned amplifier, both the input and the output voltages tend to approximate sine waves due to the flywheel effect of the relatively high Q tank circuits. Distortion in the rf amplifier will cause distortion in the SSB envelope. Class A amplifiers are used to maintain distortion at a minimum at low power (less than 2 watts). Class A amplifiers can be operated rather easily over the linear portion of their characteristic curves. As long as the maximum output voltages are maintained at less than one-tenths of the d-c voltage, the signal-to-distortion ratio can be maintained at 50 db or more. (The signal-to-distortion ratio is the ratio in db of the desired signal to the largest distortion product.) As the power levels in the transmitter increase, class AB amplifiers are used since class A amplifiers are not practical at high power. Efficiencies of 55-70% at full output can be obtained with class AB amplifiers. Recent power amplifier developments permit adjacent channel operation, using power amplifiers with signal-to-distortion ratios of from 35 to 40 db. By using rf feedback and envelope distortion cancelling techniques, ratios approaching 50 db have been achieved with high-frequency 20-kw peak envelope power amplifiers.

The dynamic regulation of the power supplies become important when using class AB amplifiers. When the load on an unregulated supply is suddenly increased, there is an interval during which the only available energy is that stored in the output filter capacitor. Since the current cannot change rapidly through the power supply filter choke, there may be a momentary drop in the output voltage which may be longer in duration than the modulation peak causing distortion. This behavior will not be evident on a d-c meter but can be observed on an oscilloscope. In class A operation, or in linear klystron amplifiers, this requirement on the power supply is not as rigorous since the load current is constant at all times. In class AB operation, plate current is drawn only as the modulation is applied and will vary from a small no-signal current to a large value on modulation peaks.

a. Distortion

Two principal causes of distortion in the rf power amplifier are: grid current loading and non-linearity of the amplifier tube.

When grid current is drawn in class AB operation, the effective impedance of the previous stage, the driver, changes in accordance with the amount of grid current. This is equivalent to imposing a varying load on the previous stage. If distortion is to be avoided, the regulation of the previous stage, the driver, must be excellent. The driver must be capable of maintaining the waveform under conditions of varying load. The usual method used to provide better regulation consists of adding fixed resistance across the grid circuit of the driven stage, thus presenting a load to the previous stage that will be modified only slightly when grid current flows. Also, it is desirable that the grid circuit of the class AB circuit be high C; this improves regulation and, in addition, simplifies the coupling from the previous stage. A stable bias supply is also required, since it is important that the bias remain constant and independent of grid current flow.

The non-linear characteristics of the amplifier tube generate distortion products that can be derived from the dynamic characteristics of the amplifier tube. The shape of this curve, the choice of the operating point, and the level of excitation will determine the amount of distortion which will be produced by the tube.

6.5. Regulation of the Output Power

The loading of a multichannel SSB circuit depends upon the traffic on the individual incoming channels. Since not all voice channels will be busy at all times, it is desirable to pass the benefits of the full power capabilities of the system to the channels in use. To allow full utilization of the transmitter output capabilities, the output signal is

A SSB OUTPUT POWER REGULATOR

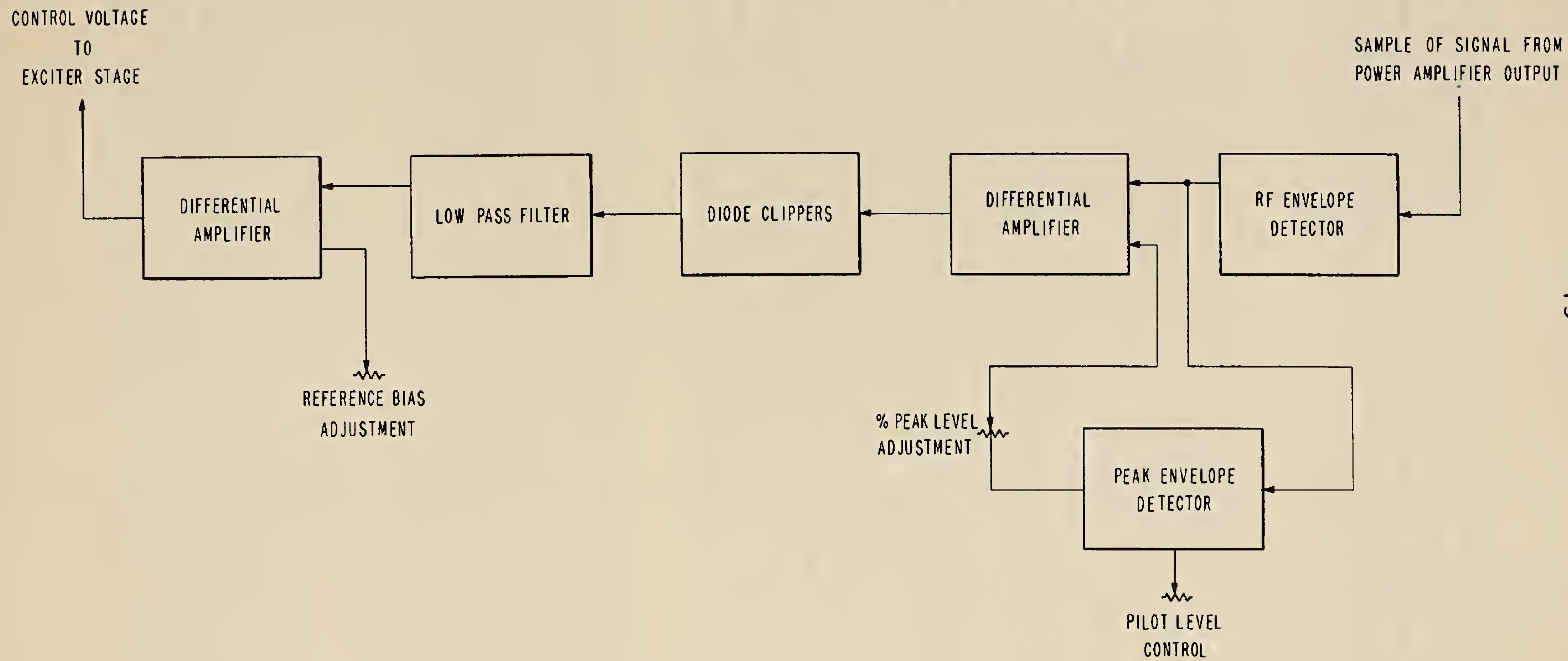


Figure 17

sampled and applied to two detectors, a peak detector and an rms detector, whose outputs are fed back as AGC to the rf amplifiers in the exciter section of the transmitter. The rms detector circuit monitors the effective output of the power amplifier and limits it to a preset level while the peak detector limits the peak power output of the final amplifier.

Figure 17 shows a block diagram of an output power regulator. A diode power detector reproduces the envelopes of a sample of the rf output of the power amplifier. The peak value of the envelope is measured by a peak detector. If the level of the input signal to the power amplifier is sufficient to drive the output of the final amplifier to saturation, the peak of the envelope will have a certain maximum value. A certain fraction of this peak provides a reference voltage to one input of the differential amplifier, while the entire envelope is fed to the other input. The diode clippers and low-pass filter produce a voltage proportional to the percentage of time the envelope exceeds the preset fraction of the peak envelope. The low-pass filter also determines the response characteristics of the regulator. The output of the filter is amplified by another differential amplifier which has a reference voltage on one input. The output of this amplifier is a negative voltage which controls the gain of one of the amplifiers in the exciter of the transmitter. The pilot-level control circuit allows the power amplifier output to be held at any desired level when no channels are in use.

A transmitter is usually considered fully loaded when its peak power capability is being exceeded 1% of the time. The distribution of the signal amplitudes will depend on the number of channels in use and the type of information on the channels. For example, with unprocessed voice, 1% peaks will approach 8.2 db as a limit above the average power as the number of voice channels increase, the figure 8.2 db being obtained after four voice channels have been combined for the composite signal.

6.6. Intermodulation Distortion

The limiting factor in power output capability of the single-sideband transmitter is intermodulation distortion generated by the nonlinear characteristics of the power amplifier. This distortion will increase as power output increases and will appear as splatter or crosstalk in and around the transmitter spectrum.

PRACTICAL SPECTRUM
TWO-TONE SINGLE-SIDEBAND SIGNAL

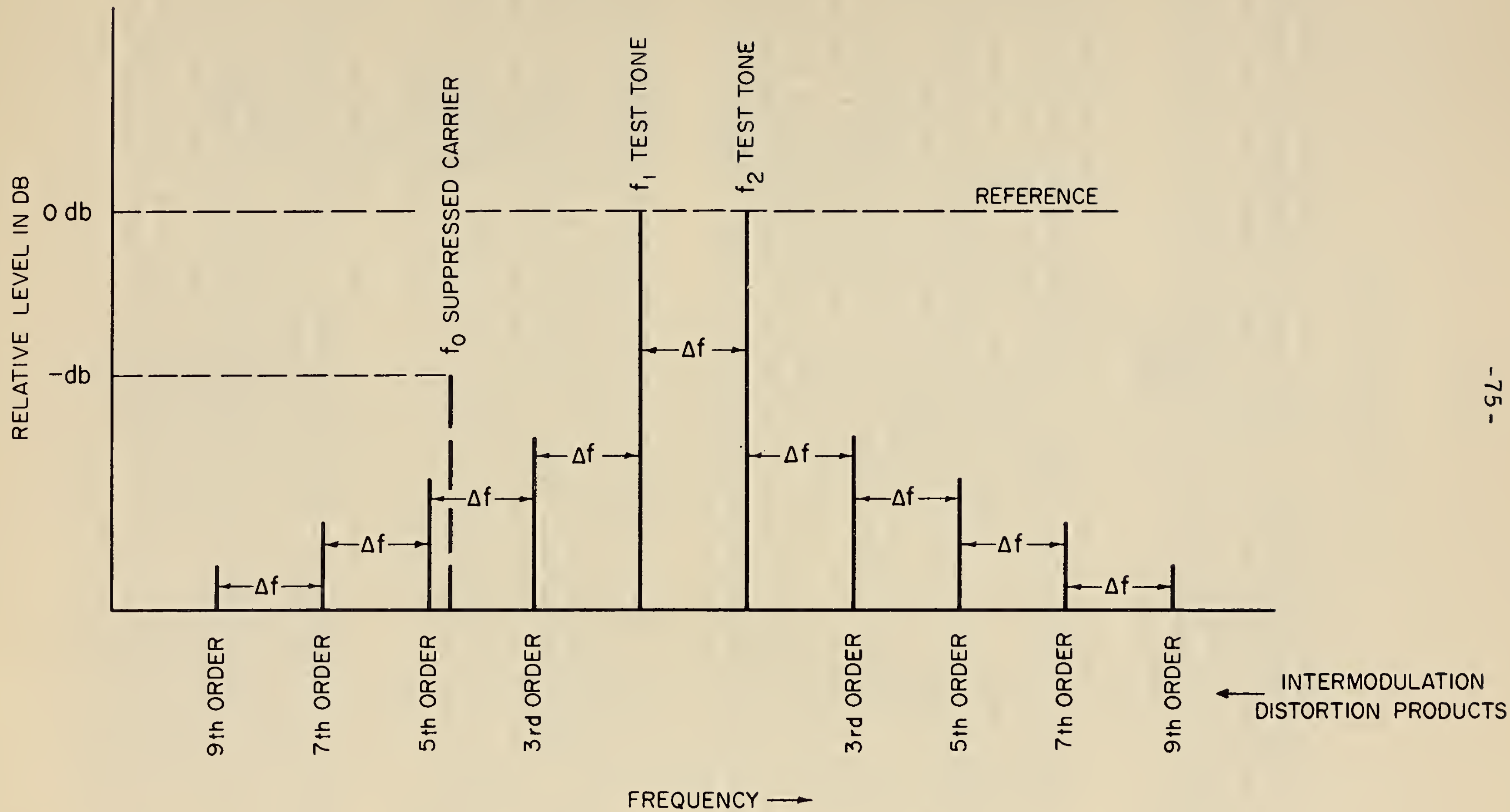


Figure 18

a. Two-Tone Test

Two test tones of equal amplitude have become a standard test for intermodulation distortion. One tone is insufficient to produce intermodulation while more than two tones produce so many intermodulation products that analysis becomes impractical. Tones of equal amplitude place more demanding requirements on the system than is likely to be encountered in normal use.

The idealized spectrum for a two-tone single sideband signal will consist of three frequencies: the suppressed carrier and the two tones translated to the desired frequency about the carrier. In practical circuits there are new frequencies generated by the nonlinearities in the amplifier. The frequencies not eliminated by the selective circuits of the transmitter are the different frequencies of odd order. The frequencies of these spurious signals are of the form $nf_1 - (n - 1) f_2$ and $nf_2 - (n - 1) f_1$, where f_1 and f_2 are the frequencies of the two test tones and where $n + (n - 1)$ is the order of the product or the order of curvature causing the product. It can be seen that many of these products will fall within the pass band of the transmitter and will be spaced along the passband by the difference between the frequencies of the two test signals as shown in figure 18.

In the two-tone method of measuring intermodulation distortion, it will be these spurious signal power levels which will be measured in comparison to the amplitude of one of the test tones at the transmitter output. The signal-to-distortion ratio then is defined as the ratio of either of the two desired test tones to the largest undesired product. This ratio is usually expressed in db. The third-order intermodulation products should be the largest unless unusual cancellation occurs. The higher-order products should become progressively smaller.

When two sinusoidal voltages of equal amplitude but unrelated in phase are applied to an SSB transmitter, the peak envelope power (PEP) will be 6 db greater than that experienced with a single tone. (PEP is the average power of the combined signal measured over one rf cycle.) The average power dissipated by the load is half the PEP when the test signal is composed of two tones of equal amplitude. This neglects the vestigal carrier power which is very small. (If the carrier is suppressed 20 db, then the power in the carrier is only 1% of the average power in one tone.)

BLOCK DIAGRAM OF TWO-TONE TEST
FOR INTERMODULATION DISTORTION

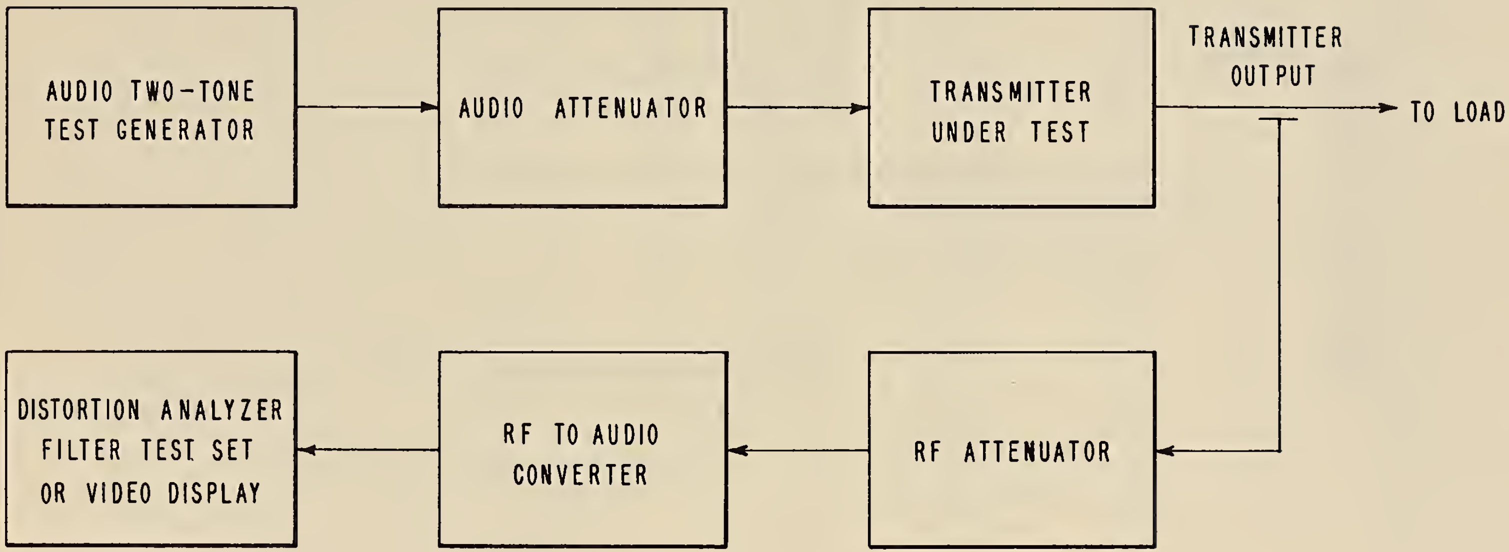


Figure 19

INTERMODULATION PRODUCTS IN A
NOISE-LOADED SYSTEM

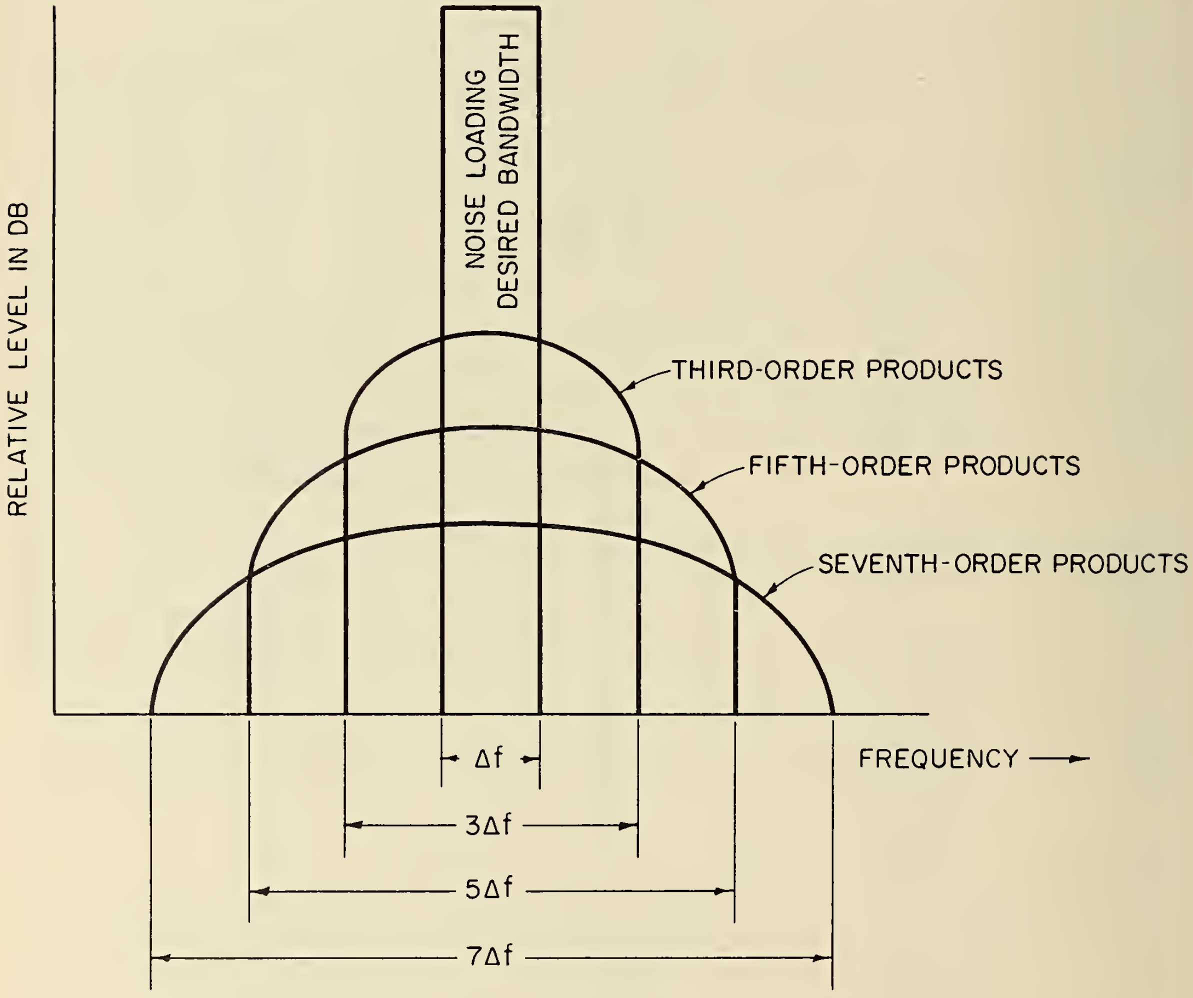


Figure 20

Any two frequencies will serve for this test; however, a 3-to-5 frequency ratio is often used as a suitable ratio to alleviate some troublesome problems occurring with many other ratios. Care should be exercised in the selection of frequencies so that:

(1) The frequency spacing is sufficient--harmonics, and the sum and difference frequencies do not interfere with the measurement.

(2) The frequency spacing is not so small that difficulty will arise in the separation of components for measurement.

(3) The frequency spacing is not so large that most of the distortion products will fall outside of the band-pass. This situation will bear no resemblance to actual operation.

Figure 19 is a block diagram of the equipment configuration for the two-tone test of intermodulation distortion.

b. Noise-Loading Test

Although the two-tone test for the measurement of intermodulation distortion is very versatile, measurements using Gaussian noise as the test signal more nearly represent the typical complex signal actually transmitted. As shown in figure 20, a transmitter loaded with noise over a discrete bandwidth B will have third-order intermodulation products that appear in a band equal to three times the desired bandwidth with the same center frequency. The fifth-order intermodulation products will fall inside a bandwidth equal to five times the desired bandwidth and have the same center frequency. The seventh-order intermodulation products will fall inside a bandwidth seven times that of the desired bandwidth and so on. The amplitudes of the products are usually in approximately inverse proprotion to the order of the product. By using results of two-tone test an approximate plot of the shape of the curves describing the amplitudes of the products may be predicted.

NOISE-LOADED INTERMODULATION TEST

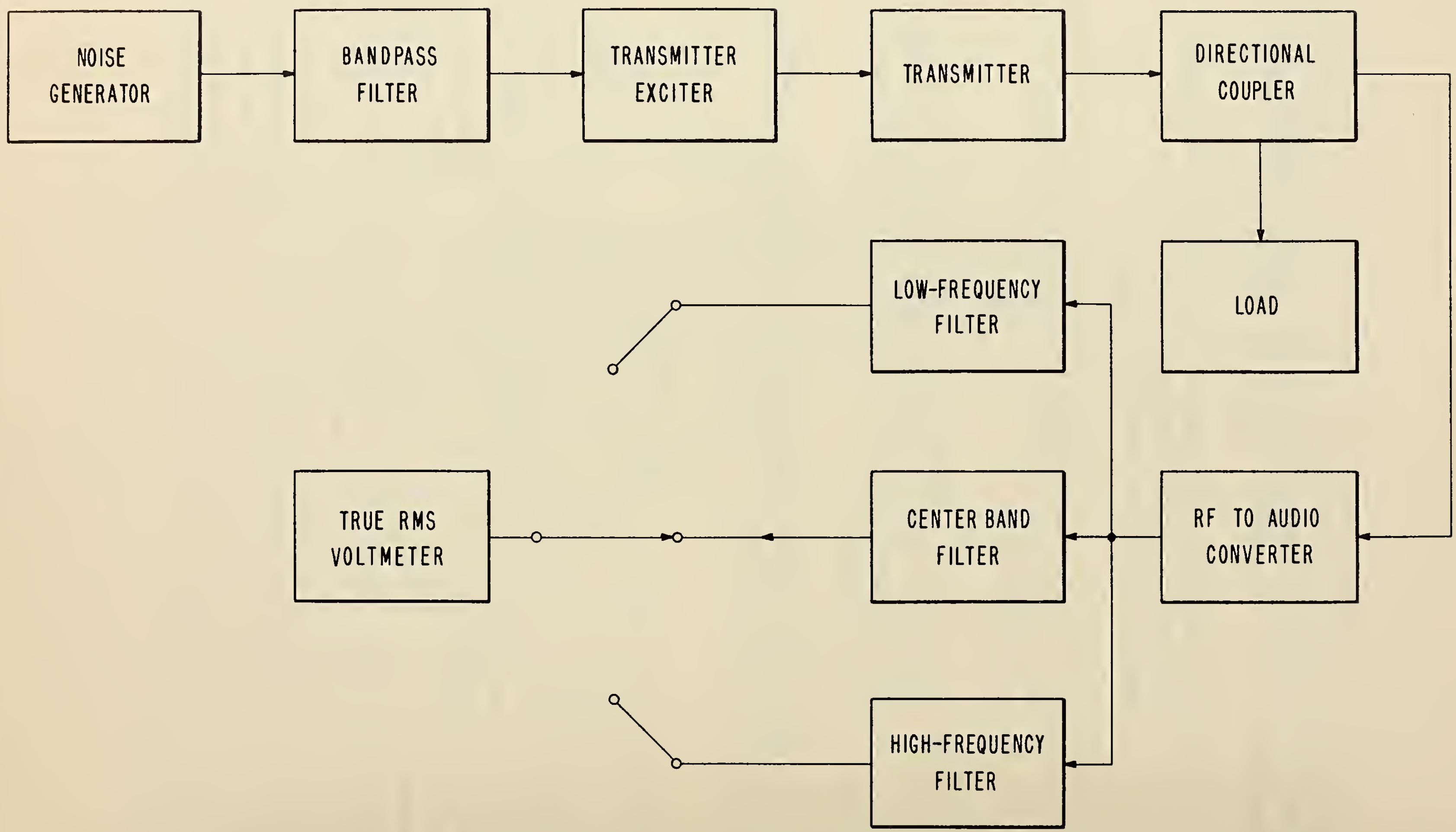


Figure 21

Using Gaussian noise as the test signal requires that the intermodulation distortion ratio be defined for equivalent bandwidths, that is, the signal-to-distortion power ratio, expressed in db, becomes the ratio of the desired noise sampled over a given bandwidth to the spurious noise in an equivalent bandwidth adjacent to the noise-loaded band.

Figure 21 shows a block diagram of a typical fixed frequency noise-loaded intermodulation test described by Icenbice and Fellhauer [1956] for a system having a baseband bandwidth of 12 to 112 kc/s and is directly applicable to single-sideband systems with other basebands by proper selection of filters.

If random noise is fed to the transmitter through a bandpass filter which passes equally all noise frequencies in the band to be tested except the low and high channel, the noise can equally load all but a few kilocycles per second of the transmission band to any degree of modulation desired. Three bandpass filters with equal bandwidth and insertion loss are used at the detection end of the system. Care should be exercised in the selection of the bandpass filters. The skirts of the distortion measuring filters should not overlap the skirts of the noise-loaded band so that an appreciable amount of noise energy will be injected directly into the measuring filter. If the overlap occurs at a point where the measuring filter is 60 db down from its peak, then little error will occur in distortion measurements as low as 50 db.

The amount of residual noise can be measured by injecting the noise signal directly into the filter and measuring its output.

One bandpass filter is chosen near the center of the noise passband for a reference signal, and the other two filters are selected to pass the spurious intermodulation products on the low and high side of the band-limited noise. The noise is measured across a load resistor with a true rms voltmeter through the reference filter. Once the reference level is established, the other two filters can be switched separately into the true rms voltmeter load circuit, and their level in db below the reference measurement is taken. The two levels in db below the reference represents the low-frequency and high-frequency intermodulation distortion generated in the loaded system.

A receiver operating at the transmitter output frequency is used for detection of the signal. A sample of the transmitter output is taken using a directional coupler. This signal may require further attenuation to avoid overloading of the receiver or detector. The output of the receiver or detector is then fed to the filters. Impedance matching must be carefully observed between receiver and filters.

7. KLYSTRON AMPLIFIERS

The klystron amplifier is particularly well suited to long distance communications circuits where high gain and high power are the principal requirements. In addition, such factors as bandwidth, circuit simplicity, and ease of adjustment all combine to render this type of tube superior to conventional three- or four-element tubes.

7.1. Power Output Efficiency

The power output of a klystron is coupled from the output cavity of the tube into an unbalanced coaxial transmission line or waveguide by means of either a coupling loop or capacitive probe or, in some cases, a combination of both. Usually, in the latter case, the coupling probe consists of an open-end loop adjusted so that the capacity between the open end of the loop and the wall of the cavity is sufficient to resonate the output coupling circuit. In the case of external type cavities, care should be taken to be sure the output coupling probe is not damaged or deformed prior to its installation.

A curve of power output when plotted against rf power input is a Bessel function of the first order and the first kind. Since the first part of this curve is practically linear, the klystron can be used as a linear modulated amplifier in single-sideband transmitters. Under these conditions, operation of the tube must be limited to the linear portion of the Bessel curve in order to prevent modulation distortion. If this distortion is to be kept at an acceptable level for a multichannel system, linearity should be maintained to within about 1%. This limitation reduces the peak power output of the tube for SSB service to approximately 80% of its saturated output capabilities. When speaking of klystrons, the term "saturated output" is the maximum power output

POWER OUTPUT AS A FUNCTION OF DRIVING POWER
KLYSTRON 3K 20,000 LK
MIDDLE CAVITY DETUNED TO f_0+f AND ALL OTHER CAVITIES TUNED TO f_0

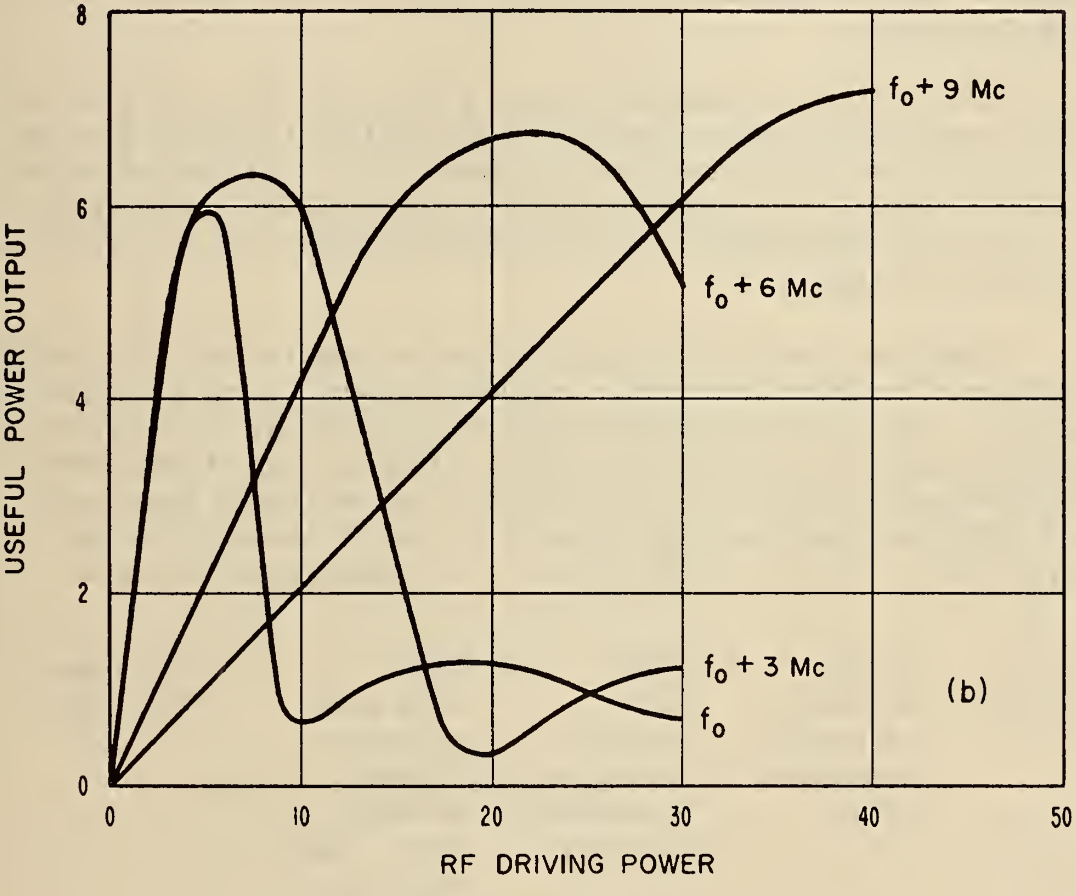


Figure 22

which can be obtained from a tube for any given beam voltage and current. Figure 22 shows the relationship between useful power output and excitation for a typical tube. The reader is referred to articles by Bruene [1956] and Icenbice and Fellhauer [1956] for additional information on klystron linearity.

The efficiency of a large klystron may range from about 15% to over 50% depending upon bandwidth requirements and the system in which it is used.

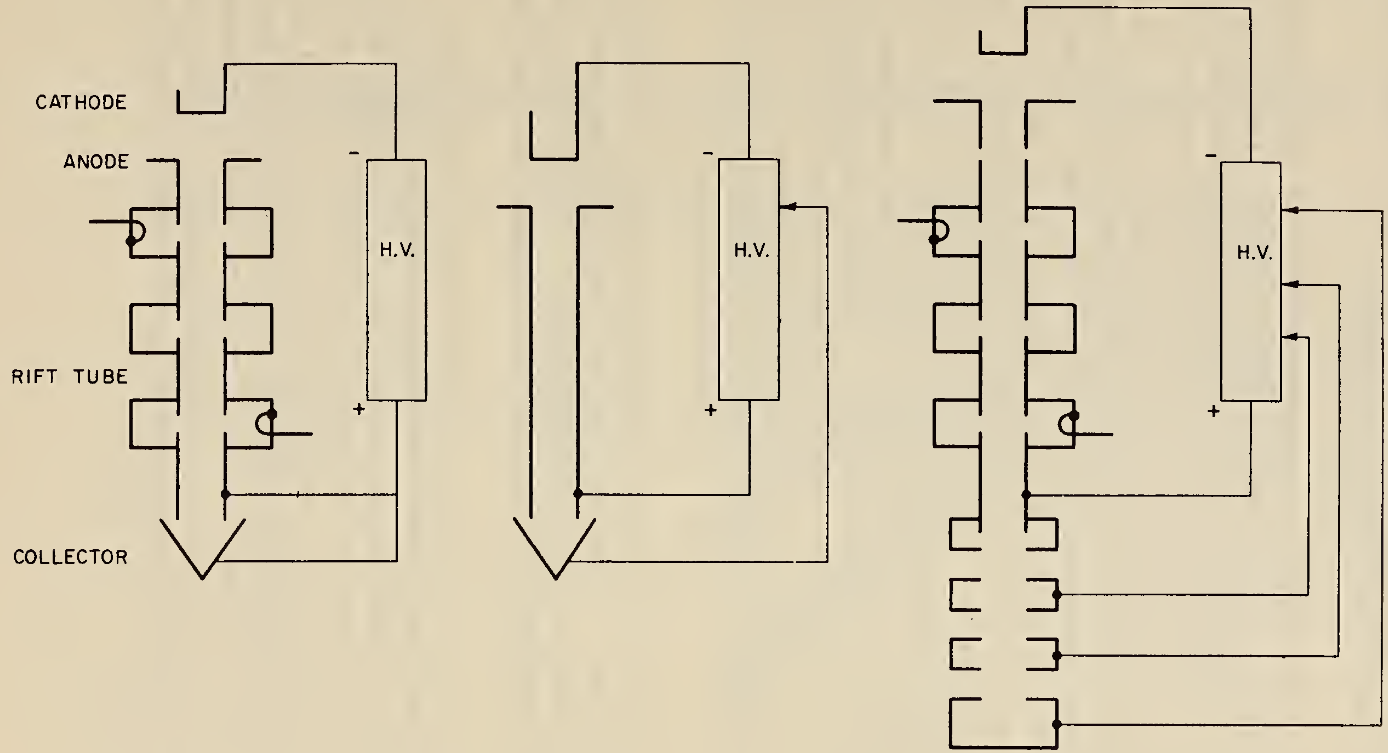
When used as an amplifier in an FM transmitter, the average carrier output level will be confined to the region around the peak of the curves. The actual departure on either side of the saturation peak will depend upon the required bandwidth. For this type of operation, the tube conversion efficiency can be quite high since the average power output level is high.

When the conventional type of klystron tube is used as a modulated linear amplifier in a single sideband system, its efficiency is quite low. If its average power output is measured by the accepted method of applying two sinusoidal audio voltages of equal amplitude simultaneously to the balanced modulator, the measured output will equal 50% of the peak envelope power. If now all factors are considered, the expected overall efficiency for SSB service would be:

Average measured output	=	(40%) efficiency of tube at saturation level	x	(80%) peak excursion limitation to maintain linearity	x	(50%) ratio average power to peak envelope power, under conditions specified	=	16% of beam power input

Actually, in a multichannel voice communication system, the ratio of PEP to average power may be more on the order of 6 or 7 to 1 rather than 2 to 1 which will further reduce the overall long-term average conversion efficiency. Since this is the efficiency of the amplifier only, and since additional power is required to operate the exciter,

TYPICAL KLYSTRONS



(A)

CONVENTIONAL KLYSTRON CONNECTION TO BEAM SUPPLY

(B)

CONNECTION FOR COLLECTOR DEPRESSION

(C)

CONNECTION FOR SEGMENTED COLLECTOR

Figure 23

final amplifier filament or bombarder supply, heat exchanger and losses within the high voltage beam supply, such as a transmitter could conceivably have an overall efficiency of 5%.

A number of ideas have been considered for improving the efficiency of a klystron when used for SSB service. The first is that of using a special connection for the collector known as the depressed collector. This is shown in figure 23 (B). The objective is to reduce the collector potential and thereby lower the energy of the electron beam by the decelerating action of the field at the collector gap. The reduction of collector potential below that of the beam voltage is limited by the allowable increase in drift tube current. A further refinement of this idea is demonstrated in Eimac's type X625B klystron where the collector is segmented. This is shown in the drawing of figure 23 (C). Each segment is at a different potential with minimum collector voltage at the collector end of the tube. Such an arrangement will reduce the beam power input to the tube, thereby improving its conversion efficiency. The idea of the depressed collector has proven quite effective, particularly in low-powered klystrons. In larger tubes where the beam potential is quite high, the depressed collector may not be as effective due to the high velocity of the electrons.

Other schemes for achieving high efficiencies in SSB service have been investigated. Those having the most promise were conducted with tubes which contained a modulating anode. Such arragnements are not yet being employed in present day commercial equipment.

Efficiency is an important consideration. For example, in FM service, the Varian type VA 842 tube will provide an average power output of 75 kw over a bandwidth of 3 Mc/s at an efficiency of approximately 43%. This amounts to an input of about 175 kw. It should be pointed out that only the input to the final tube is considered and that other power requirements will raise the primary power needed to well over 175 kw. Such a transmitter would require a heat exchanger capable of dissipating the full 175 kw of power since, in the event excitation is removed, the full beam power input to the tube must be dissipated by the collector.

THE POWER OUTPUT AND EFFICIENCY OF AN EIMAC KLYSTRON 3K 20,000 LK

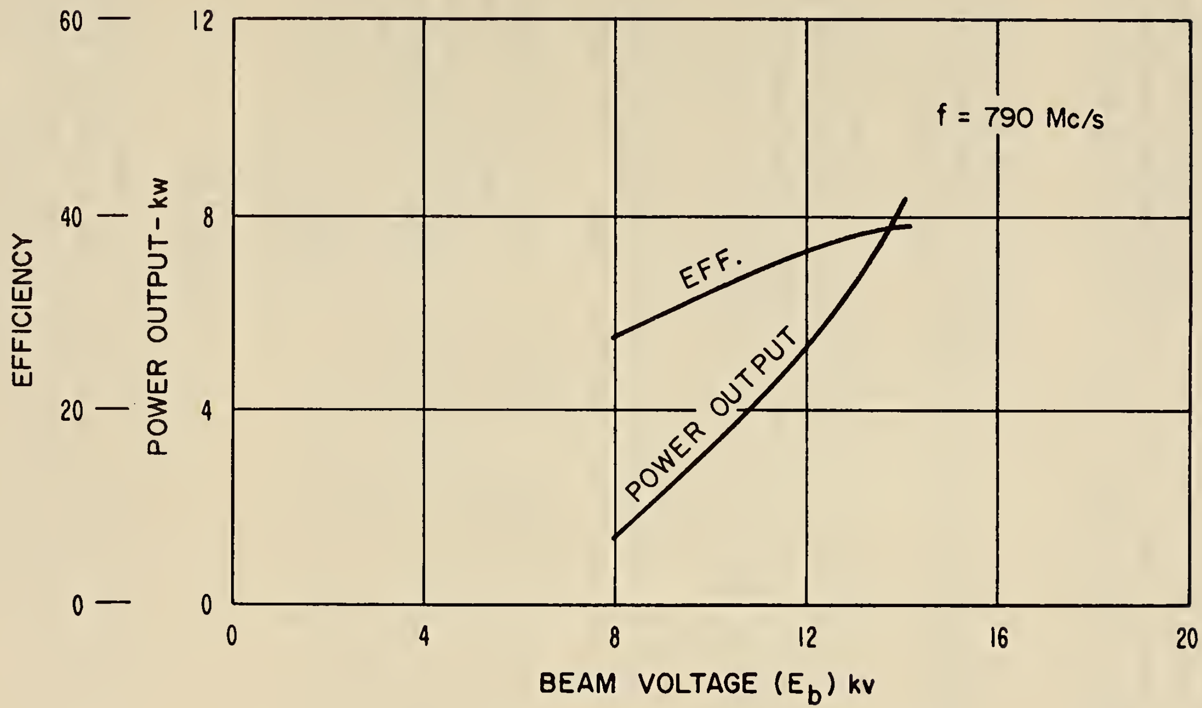


Figure 24

It will be noted from the curves of figure 24 that the efficiency and power output of a klystron are functions of beam voltage. There is an optimum voltage for best efficiency and power output for a given tube since voltage determines the velocity of the electron bunches through the drift tubes and cavities. Because the bunching action of the electrons takes time, for a given frequency, there is an optimum balance between the velocity of the electron bunches and the distance which they must travel if they are to give up their energy in the output cavity at the optimum moment.

It will also be noted from figure 22 that maximum output and efficiency do not occur under the synchronously tuned condition but when the middle cavity is detuned to the high side of resonance.

The efficiency of a klystron is affected to some extent by the electron beam configuration which is controlled by adjustment of the magnetic focus coil currents. Slight adjustment of the individual coil currents will affect both the output and tube body current. Improper adjustment of the beam focus will cause excessive body current and reduced output. This is due to the fact that many electrons are allowed to leave the confinements of the main beam and strike the body of the tube, thereby never reaching the output cavity and collector.

The beam power input to a klystron is constant regardless of whether it is operating as a linear amplifier in a single-sideband transmitter or as an amplifier for an FM transmitter. For this reason, regulation of the d-c beam power supply is not at all critical. In contrast to this, if a conventional three- or four-element tube is used as a linear amplifier in SSB service, regulation of the plate supply must be very good to prevent distortion on modulation peaks.

7.2. Bandwidth

The bandwidth of a klystron is usually specified under one of two headings, namely, "synchronously tuned" or "stagger tuned." The term "synchronously tuned" refers to the condition where all cavities are tuned or resonated at the driven frequency. The term "stagger tuned" describes the method whereby the cavities are not all resonated at the operating frequency.

HYBRID NETWORK FOR PARALLELING THE OUTPUT OF TWO TRANSMITTERS

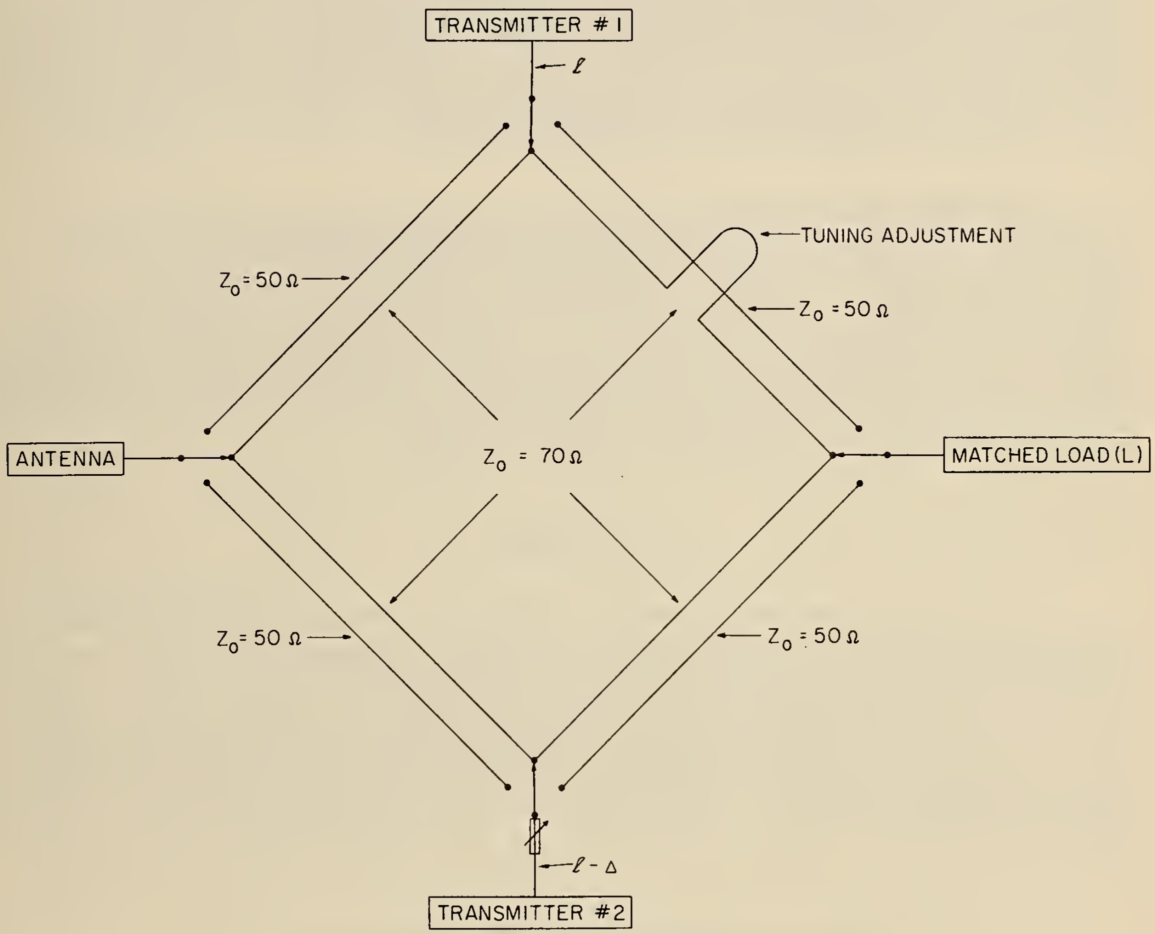


Figure 25

Where the bandwidth requirements do not exceed two or three Mc/s, the most common and efficient means of achieving this is to detune the center cavity, or cavities, slightly to the high side of resonance. The curve of figure 22 showsthat such an adjustment will increase the bandwidth and at the same time provide more useful power output at slightly less gain. In the event this technique will not provide adequate bandwidth, a further broadening of the passband can be had by loading the center cavity by means of an external resistance. This is not always practical since the technique requires that the cavity be provided with a coupling probe.

Another effective means of increasing the passband of the tube is that of detuning the input cavity to the low-frequency side of resonance and, when necessary, loading this circuit by means of an external resistance. This method should not be used unless the first suggested method proves inadequate. Any detuning or loaded operation will increase the excitation requirements. Such techniques will usually serve to increase the bandwidth substantially.

7.3. Gain

In general, the power gain of a klystron amplifier is quite good, ranging between 10 and about 60 db. Gain is dependent upon the number of cavities contained within the tube, how they are tuned, and whether they are resistance loaded for increased bandwidth. Most tubes used in forward scatter communications transmitters contain three cavities which have power gains ranging between 25 and 40 db even under broad banded operation.

8. POWER COMBINERS FOR TRANSMITTERS

One of the first requirements of a transmitter in a forward scatter communications system is that it be capable of generating the necessary power to assure, for a given percentage of time, a predetermined grade of service at the distant receiving terminal. Such a prerequisite can be met either by use of a single transmitter large enough to supply the full power requirement or by two or more similar transmitters, each capable of developing its share of the required power. In the latter case the output of the transmitters must be combined by means of a hybrid network shown in figure 25 and similar to that described by Tyminski and Hylas [1953].

The hybrid ring network is an rf coaxial containing three equal arms of approximately one-quarter wavelength each and a fourth arm approximately three-quarters of a wave long. The latter usually contains an adjustable phasing section for use in balancing the bridge at the operating frequency. The four junctions of the bridge are terminated as shown. The matched load L should be capable of dissipating the full output of one transmitter. The characteristic impedance of the four arms of the bridge is usually 70 ohms where 50-ohm terminations are employed. This provides a suitable load impedance for the two transmitters.

If it is assumed that the bridge is in balance and that the power output of the two transmitters is equal and in phase at their respective bridge junctions, the combined output from the two transmitters, less slight losses in the bridge elements, will be dissipated in the antenna since voltages at the antenna junction of the bridge are in phase. Because of the added half-wave of line in one leg of the bridge between the transmitters and the load L, phase cancellation occurs at that junction. Consequently there is no power dissipated in L. Any power appearing at L would result from unequal power levels or phase angles at the two transmitter junctions of the bridge or from an improperly balanced bridge.

Failure of either transmitter will result in a 6-db reduction in power at the antenna. Half of the output of the operating transmitter will be expended in load L since phase cancellation no longer occurs across its junction terminals. (A 3-db loss in power due to the failure of one transmitter plus an added 3-db loss due to half of that transmitter's power output being dissipated by load L accounts for the 6 db.)

In practice, a means is provided whereby the hybrid ring can be removed from the circuit. In the event one transmitter fails, the operator can choose a period of low traffic to switch the transmitter to the termination L for trouble-shooting while the operating transmitter can be switched directly to the antenna, thereby raising the power level at the antenna by 3 db over the loss encountered if the hybrid ring remained in the circuit.

8.1. Tuning Procedure

If the combiner is to function properly and efficiently it is important that the bridge be carefully balanced at the operating frequency. This may have been done at the factory. If so, it is probable that the three-quarter wave element cannot be adjusted in the field. If this is not the case, and the element length is variable, the following procedure can be followed:

(1) T_1 and T_2 should be disconnected at their respective bridge junction and the junctions left unterminated.

(2) A small amount of power at the operating frequency is supplied to the bridge at the antenna junction.

(3) A voltmeter or power meter is connected to the terminal load L.

(4) The variable arm of the bridge is then adjusted for minimum voltage or power across load L. This will approach zero as the bridge is brought into balance.

Connections are then restored to normal following a check of the bridge balance. If the transmitter characteristics are similar the bridge will remain in balance with the transmitters connected for normal operation.

There may or may not be provisions for adjusting the phase of one transmitter with respect to the other. In some cases the phase characteristics of the two transmitters are so nearly alike that a control is considered unnecessary. Where a control is provided, the transmitter output levels should be equalized and the phase-matching section adjusted for minimum power at the termination L. By making adjustments to the individual power levels and phase control, the output can be reduced to practically zero.

8.2. Circuit Reliability

When considering the reliability of a system, there are several obvious advantages to be gained by using the combined output of two smaller transmitters rather than one large transmitter.

First, if the power requirement is substantial, the reliability of large amplifier tubes and their necessary high-voltage power supplies and components is much less than for smaller tubes and correspondingly smaller components.

Second, should one transmitter fail, the circuit is not completely interrupted as would be the case where one larger transmitter was involved. As was pointed out earlier, the loss of one transmitter would reduce the transmitted power by 6 db which may or may not seriously impair circuit performance, depending on circuit loading and propagation conditions. This loss in output could be reduced to 3 db by removing the power combiner from the circuit.

Another point of interest is the possibility that during periods of high fields which occur most frequently during summer months, it would be possible to use a single transmitter with the second transmitter serving as a standby. This will reduce the primary power requirements below that which would be required by operating a single large tube at reduced plate or beam voltage.

This method of combining the power output of two transmitters can be extended to four transmitters if the requirements warrant it. Such an arrangement would contain three hybrid rings similar to that shown in figure 25.

9. THE RECEIVER

Many of the aspects of performance with regard to the receiver have been considered by Florman and Tary [1962]. It is recommended that the information contained in this report be used in conjunction with the above reference since a complete critique of the performance of the receiver demands a consideration of the standards given therein.

The typical communications receiver may be viewed as an instrument composed of a variety of basic sections so interrelated as to perform an overall reception-detection function. These basic sections of the receiver differ considerably in components, circuitry, and frequency of operation. Therefore, the overall performance characteristics of the whole receiver are directly dependent upon the performance characteristics of the individual basic sections.

THRESHOLD EFFECT FM VERSUS SSB

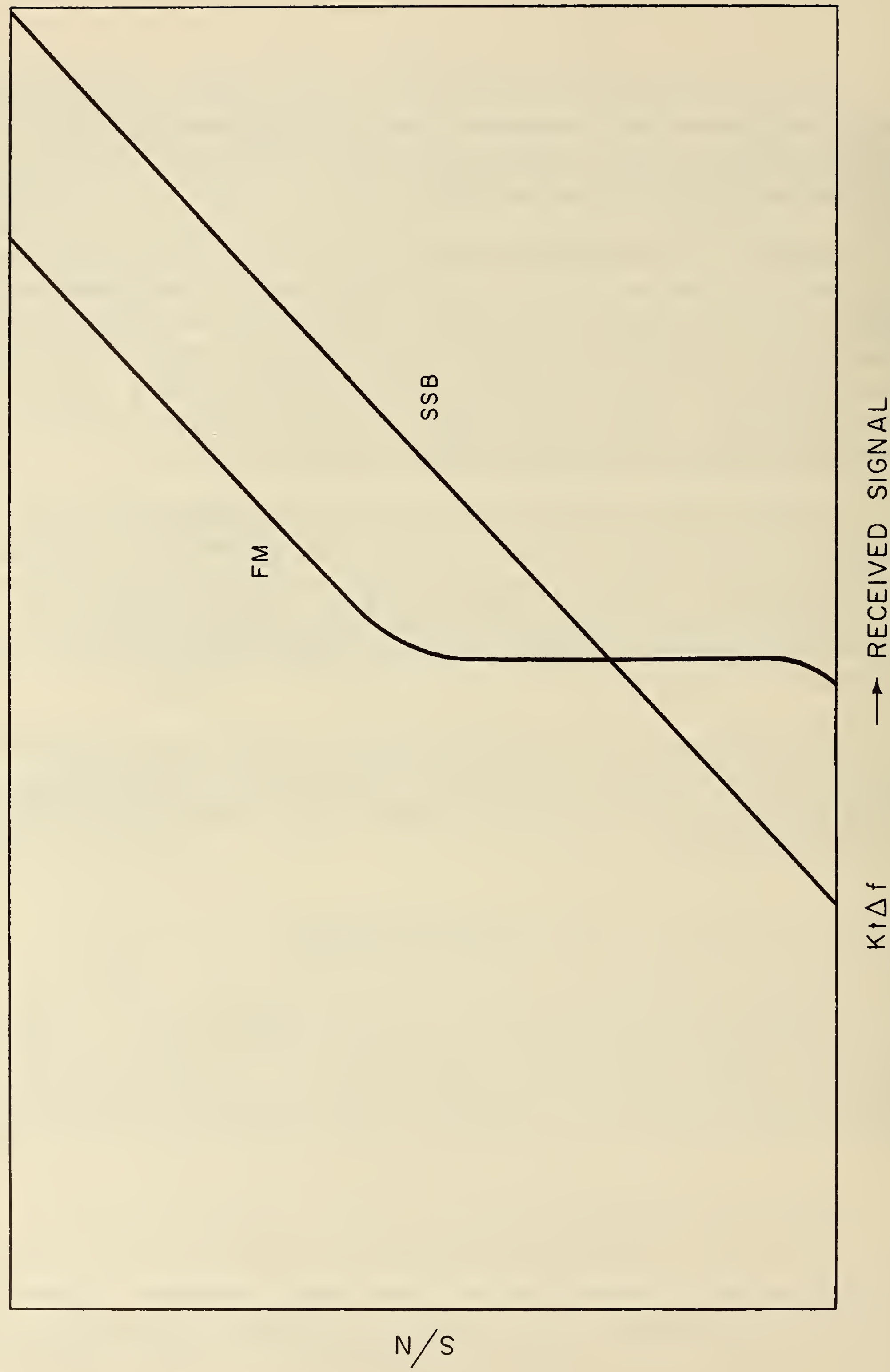


Figure 26

It is well known that the ultimate usefulness of the receiver is limited by its ability to discriminate between the weakest signals which must be detected and noise. Thus, one of the most important parameters which can be determined for the individual basic sections of the receiver is the noise figure, where it expresses the relative ability of the receiver to amplify a weak signal without introducing appreciable noise. This, in turn, is related to the threshold effect in FM systems.

A graphical illustration of the threshold effect in FM systems as contrasted to SSB is shown in figure 26. As illustrated, the weakest signal in an FM system produces nothing but noise in the receiver output. As the signal level is raised, the output continues to be noise until a transition point or threshold is reached. At or above this level, the signal becomes available at a fairly good signal-to-noise ratio, while signals below this level are unusable. This level is dependent on the noise figure and bandwidth of the particular receiver involved. In any typical tropospheric scatter communications system, the threshold level increases as the number of channels increases.

It has been demonstrated that scatter signals are reliable since they are always present. However, the amplitude of the received signal is constantly fluctuating over a fairly large range, the amount of fluctuation being dependent upon the particular path and carrier frequency. When FM is used over a scatter path, information is completely lost at the point at which the signal fades below the threshold; thus, the system must be operated with a minimum received median signal sufficiently above the threshold so that information will not be lost during deep fades.

When diversity signals are continuously combined, whether in FM or SSB, according to the proper operating mode of the combining method used, the signal-to-noise ratio of the combined signal will be better than that of the best signal. A combination of four signals all having the same signal-to-noise may result in a combined signal-to-noise ratio 6 db better than that of any one signal by itself. By controlling the amount of any given signal that is allowed through its combiner in accordance with its own signal-to-noise ratio, optimum combination can at all times be achieved.

9.1. The rf Amplifier

Almost all rf voltage amplifiers are operated under class A conditions. The major factors on which voltage gain per stage are dependent are the specific characteristics of the tube used, the input and output impedance, and the bandwidth to which the stage is designed. In some circuits, noise figures also must be considered as a limiting factor in the determination of voltage gain. However, beyond the first or second stage in cascade amplifiers, the noise is usually low enough in relation to the signal amplitude to be neglected.

At frequencies up to 250 Mc/s, circuits similar to those used at lower frequencies are practical. The voltage gain in amplifiers with a bandwidth of 2 or 3 Mc/s will range from 40 to 50 at 30 Mc/s and 8 to 10 at 215 Mc/s Triodes can be used successfully in this frequency range, but it is more common to use pentodes such as the 6AK5 or 6CB6. The voltage gain of a triode is considerably lower than a pentode, making it necessary to use a greater number of amplifier stages to obtain a given total gain. A factor limiting the use of pentodes at high frequencies is their noise figure. In general, pentodes operating at frequencies up to 250 Mc/s give greater gain than triodes. Because of their poor noise figure, pentodes are not used in the upper part of this frequency range where the noise level is approximately that of the desired signal.

In the frequency range from 300 Mc/s to 1000 Mc/s, the use of triodes and distributed-property circuit element becomes essential. Reasonably good gain and noise figure can be obtained with triodes. Above 500 Mc/s, concentric-line circuits and special lighthouse, rocket, pencil, or disk-seal planar type triodes are used to obtain voltage gain with an acceptable noise figure. One of the features of the concentric-line circuit is the ease with which a shielding and isolating effect can be obtained in grounded grid configurations. No practical pentodes have yet been developed for use as voltage amplifiers in such circuits. The voltage gain expected with triodes in circuits containing concentric-line sections usually does not exceed 3 to 12. Although higher gain can be obtained at the lower end of the frequency range, the gain declines rapidly as the frequency increases because of the lowered circuit Q and decreasing tube efficiency.

There are not many electron tubes that will operate satisfactorily at 1000 Mc/s and above as rf amplifiers. Therefore, the use of rf pre-selection and a crystal mixer has been commonly employed. The rf pre-selector is a tunable filter with low insertion loss to the operating frequency and high loss at undesired frequencies.

a. Noise in the rf Amplifier

The limiting factor in the operation of rf amplifiers, especially at frequencies above 300 Mc/s, is the noise generated by the amplifier. Consequently, low noise devices such as parametric amplifiers and masers are being considered as rf amplifiers. Since the significant part of the noise generated in a cascaded operation is contributed by the first several stages, the use of such low noise devices is equivalent to improving the sensitivity of the receiver, thus permitting the detection of extremely weak signals.

Probably the only parameters of significance affecting the performance of rf amplifiers are the noise figure and gain. Throughout this report all discussion relating to the noise figure assumes that the input noise is equal to KT_oB and that this input noise is the only noise contributed by the testing instrument or device--where the testing instrument output impedance is at room temperature (T_o = 290°K).

Although all parts of the receiver contribute noise to the output, the initial stages are the chief contributors. Unless the mixers and IF amplifiers have exceptionally high noise figures, the rf amplifier can be considered to be the determining factor. This is clearly indicated by the following expression for the total noise figure of multistate networks connected in cascade:

$$F_{1---n} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 \; G_2} + ---- + \frac{F_n - 1}{G_1 \; G_2 --- G_{n-1}} \qquad (41)$$

where F_1 represents the noise figure of the first stage,

F_2 the second stage, etc.;

G_1 represents the gain of the first stage, and

G_2 the gain of the second stage, etc.

Thus it is evident from the expression that after any high-gain network the overall noise figure is not influenced to a great extent by additional networks even though their individual noise figures are relatively high.

Consequently, for this measurement it is sometimes assumed that the rf amplifier is the major contributor and that the noise figure reflects the noise added by the amplifier alone. Such an approximation may not be valid, especially where parametric amplifiers and masers are used, since their contributions to the noise of the overall receiver are very small with respect to the mixer and the IF amplifier.

9.2. Noise Figure Measurement

Utilizing the concept of available power, the noise figure F of a network is defined as the ratio of the available signal power at the input to kTB, divided by the available signal-to-noise ratio at the output, which, written in equation form, is

$$F = \frac{S_i/kTB}{S_o/N_o} \tag{42}$$

where S_o = available signal power at output,

N_o = available noise power at output,

S_i = available signal power at input,

kTB = available noise power of signal generator.

Simply stated, the noise figure is the ratio of the signal-to-noise power at the input divided by the signal-to-noise power ratio at the output, as given by Wind and Rapaport [1958]. Thus, if the network were ideal, no noise would be added and the output noise would be the input noise from the signal generator increased only by the gain of the device, resulting in a noise figure of 1 and a perfect receiver since we now have the same available signal-to-noise ratio at the output as we have at the input.

Since the noise power per cycle available from a source at absolute temperature T is kT, where k is Boltzmann's constant, the noise power N_o, available at the output of an ideal receiver, is

$$N_o = kT \int_0^\infty G(f)\, df \tag{43}$$

where G(f) = the gain function of frequency f.

The noise figure can be expressed as the power ratio in db of the actual output noise power N available to that which would be available if the amplifier merely amplified the available thermal noise of the source at a reference temperature. This can be expressed by the following equation:

$$F = \frac{N}{N_o} = \frac{N}{kT \int_0^\infty G(f)\, df} \tag{44}$$

which, converted to db, is $F = 10 \log_{10} N/N_o$.

The reference source temperature T is taken equal to 290° Kelvin (room temperature), and k is equal to 1.38×10^{-23}. Therefore $kT = 4.00 \times 10^{-21}$ watts per cycle of bandwidth. If a receiver is operated with a source at this temperature, and if thermal noise alone corrupts the signal available from the source, then the noise figure gives directly the degradation in signal-to-noise ratio produced by excess noise in the receiver.

For receivers using conventional front ends, the measurement is conducted using the entire receiver. The output is measured either immediately before or immediately after the detector. Since power is to be measured, a square-law detector is advantageous. In the absence of a square-law or approximately square-law detector, the measurement can be made with a precision 3-db attenuator between the IF and the detector.

a. Signal Generator Method

The noise figure of a receiver may be measured using a conventional signal generator capable of delivering a known variable amount of power to the input of the receiver. It is important that the output impedance of the source generator be the same as the impedance of the transmission line from the antenna.

With the signal generator connected to the input terminals of the receiver and the power output at zero, the total noise power developed by the receiver will be

$$N = FkT \int_0^{\infty} G(f)\, df \tag{45}$$

as measured by a square-law detector connected to the output terminals of the IF amplifier. A square-law detector is defined as a detector in which the output voltage is proportional to the rf power or to the square of the rf voltage applied at the input.

The signal generator is now tuned to the frequency f_0, and the output available power level, S is adjusted until the total power output of the receiver P is equal to 2N. This is accomplished by doubling the initial reading of the square-law detector. In the absence of a square-law detector, the measurement can be accomplished using a 3-db precision attenuator between the IF amplifier and the detector. The initial power measurement is made without the attenuator. The second measurement is made with the attenuator between the IF amplifier and the detector, and the initial power level at the detector is restored by increasing the input power.

In equation form this can be represented by

$$P = FkT \int_0^{\infty} G(f)\, df + SG(f_0) \tag{46}$$

where
P = total output power,
F = noise figure,
k = Boltzmann's constant,
T = source temperature,
$G(f_0)$ = receiver power gain,
S = available signal power from signal generator.

By adjusting the power level such that P = 2N, where N is defined by (45),

$$FkT \int_0^\infty G(f)\, df = SG\,(f_0) \tag{47}$$

which reduces to

$$F = \frac{S}{kT \dfrac{1}{G(f_0)} \displaystyle\int_0^\infty G(f)\, df} = \frac{S}{kTB} \tag{48}$$

since

$$B = \frac{1}{G(f_0)} \int_0^\infty G(f)\, df = \text{effective noise bandwidth.} \tag{49}$$

Thus by (48) we can measure the noise figure of a receiver and determine the value of B in (49) by measuring the power gain as a function of frequency and integrating the curve graphically.

Extreme care must be exercised in the selection of a signal generator to insure that the available power output S is an accurate value at the low levels required. However, with the advent of improved circuitry and reliable power standards, such generators are available through commercial manufacturers.

b. Noise-Generator Method

The signal generator method of noise-figure measurement is considered to be reliable and is recommended as the most accurate method. However, it is time consuming, and other methods are availabel which should suffice to determine the noise figure for circuit operation requirements.

Essentially, this method consists of a comparison of normal noise with the noise occuring when the generator is artificially changed to a new temperature.

A temperature-limited diode is employed for this purpose at VHF and UHF frequencies, while the argon discharge tube in a wave-guide is used at microwave frequencies. The requirements of the noise source are that its noise spectrum is flat with frequency over the passband of the receiver and that a known amount of noise power is generated. The latter is accomplished by associated circuitry.

Therefore, the noise power output of the receiver N_1, with the noise source off, and the noise power output with the source on, N_2, are used to determine the noise figure.

Allowing for the contribution due to noise sources in the receiver alone which add to the total available output noise, (44) can be rewritten in the form

$$N_1 = kT \int_0^\infty G(f)\, df + (F - 1)\, kT \int_0^\infty G(f)\, df \tag{50}$$

where the first term is the amplified source noise and the second term is the amplified excess receiver noise. If X is equal to the ratio of the available power per cycle from the noise source to kT, then, when the noise source is turned on, the following equation results:

$$N_2 = X\, kt \int_0^\infty G(f)\, df + (F - 1)\, kT \int_0^\infty G(f)\, df \tag{51}$$

This is represented in ratio form

$$\frac{N_2}{N_1} = \frac{X + F - 1}{F} = 1 + \frac{X - 1}{F} \tag{52}$$

which reduces to

$$F = \frac{X - 1}{\frac{N_2}{N_1} - 1} \tag{53}$$

It is evident that this is a much simpler relationship to determine than the values needed for the signal generator method.

Here only the value of X for the source and the observed output power ratio N_2/N_1 are needed.

Precaution is necessary in two areas, namely, the noise source must present nearly the same impedance to the receiver when turned on or off, and the image frequency rejection of the receiver must be adequate.

The impedance must be matched closely; otherwise variations of receiver output noise with source impedance will affect N_2/N_1.

Most present-day superheterodyne receivers employ preselector cavities or bandpass filters to reject the unwanted energy at the image frequency. However, if a receiver is used where precautions are not taken to reject the response at the image frequency, such as wide-band radar receiver, the measurement of the noise figure obtained by the noise diode device may be 3 db lower than that obtained by the signal-generator method. To avoid this condition it is recommended that a pre-selector cavity or bandpass filter be inserted at the receiver input to attenuate the response at the image frequency.

Thus for this measurement the noise generator is connected to the receiver input and a suitable output indicator, such as a true rms voltmeter, is used at the IF output. The rms voltmeter reading is obtained with the noise source turned off. The noise source is turned on and increased until the voltmeter reading is increased by 3 db. The noise output necessary to increase the rms voltmeter reading by 3 db is a measure of the noise figure of the receiver. Noise generators are available with meters calibrated directly in db of noise power output.

Basically, the above discussion is related to the work by Friis [1944] and Oliver [1958]. Further work has been published by Haus and Adler [1957] and Brodzinsky and MacPherson [1959] with relation to the problems presented by negative resistance devices and low noise MASER's, but such studies are considered to be beyond the scope of this report.

9.3. The Mixer

Since all receivers used in either FM or SSB are usually of the heterodyning type, one or more mixers will be encountered. The mixer or converter may be defined as the device which converts the rf signal frequency to the intermediate frequency. The mixer combines the input signal with a local oscillator signal to yield the IF.

At the lower frequencies tube-type mixers are used. At HF a single tube can be used for both the oscillator and the mixer, whereas at VHF such a technique is not practical due to the inefficiency of the oscillator portion of such a converter. Above HF, especially where frequency stability is required, a crystal oscillator plus multipliers are required to produce the local oscillator signal. Tubes are used as mixers up to frequencies of 700 Mc/s. At frequencies above 700 Mc/s, crystal mixers are used since they have lower noise figures than the tube type.

The diode-, triode-, or pentode-type tubes are most often used as mixers or converters at frequencies up to approximately 700 Mc/s.

The diode mixer gives noise figures which are appreciably higher than those obtained with other types of tubes. However, the simplicity of the circuitry has advantages, and the resulting noise figure will be sufficiently good for many purposes. On particular disadvantage with the diode mixer is that, since the average impedance is low, the local oscillator must furnish a considerable amount of power which results in a problem of local oscillator stability and possible unwanted interaction between the local oscillator and mixer.

Triode and pentode mixers will result in considerably better gain and noise figures than the diode mixer. Essentially all of the characteristics which apply to these tubes at VHF frequencies as conventional amplifiers will also apply for their operation as mixers. The important difference is that the power gain and noise figure which can be obtained from a given tube as a mixer can never be as good as the power gain and noise figure of the same tube operating as an amplifier.

The output impedance is higher for the pentode mixer than for the triode mixer, and consequently the maximum power gain of the pentode mixer is higher than the maximum power gain of the triode mixer.

For these tube-type mixers, the problem of local oscillator injection must be considered and is different for each type. In the diode mixer optimum performance is obtained by biasing the tube beyond cutoff and then using a large local oscillator signal to swing the tube into the conducting region. For the triode and pentode mixers the optimum local oscillator voltage on the input grid is about seven-tenths as large as the cutoff bias. For the triode mixer this operating bias is usually obtained through a high-resistance grid leak. However, if for any reason the local oscillator fails, excessive mixer current will result, and damage may occur to the tube. In the pentode mixer where the operating bias is obtained by a high-resistance grid leak, a screen-dropping resistor is used to prevent damage due to this excessive mixer current if the local oscillator fails.

On parameter of interest especially from design considerations is the conversion transconductance, which is comparable to the transconductance of a tube in the conventional sense. Conversion transconductance is defined

$$g = \frac{\partial i_p}{\partial e_s} \tag{57}$$

∂i_p = change in plate current

∂e_s = change in signal voltage

That is the change in IF current output for a change in input signal voltage. If signal voltage is small and the local oscillator voltage large, the conversion transconductance of the mixer will be a function only of the local oscillator voltage.

a. Conversion Gain Measurement

Tube-type mixers are capable of conversion gains greater than one, but crystal mixers have conversion gains less than one since no amplification takes place in the crystal itself. The conversion gain or loss is a function of local oscillator injection. For tube mixers, it is difficult to specify the relationship between the conversion gain and the local oscillator injection since it will differ for each type of tube used as a mixer.

THE DC INCREMENTAL METHOD FOR MEASURING CRYSTAL CONVERSION LOSS

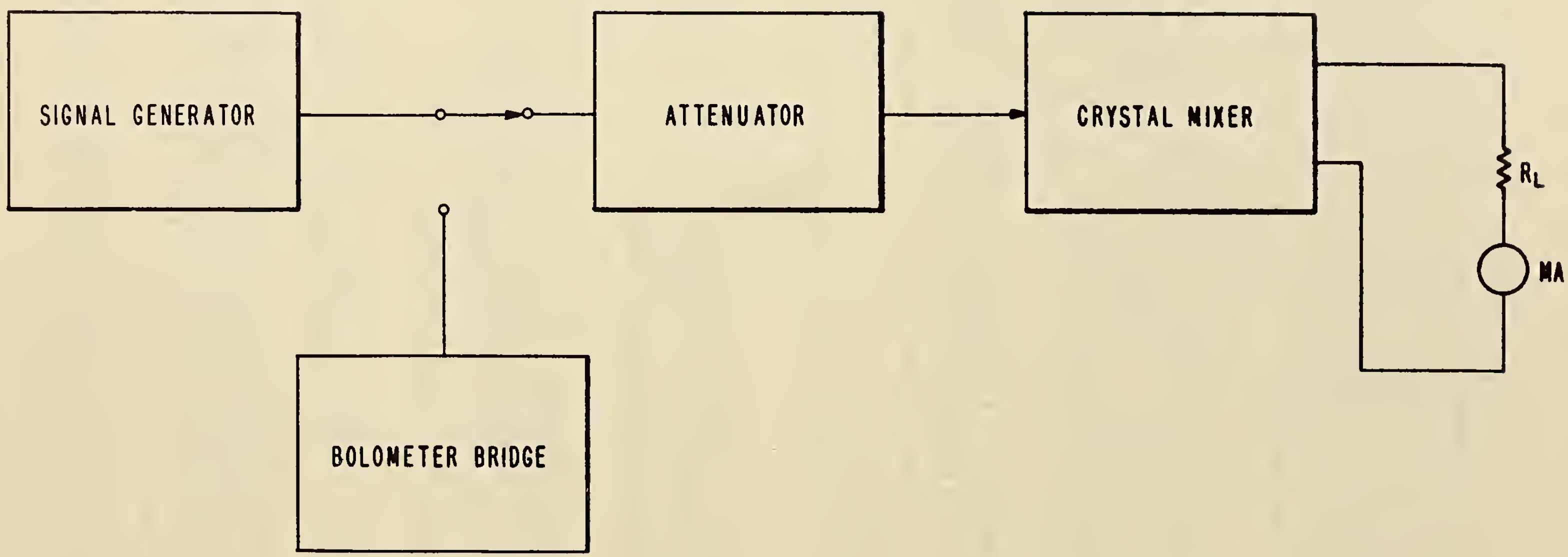


Figure 27

The conversion gain is an important parameter in determining the noise figure of a mixer. The measurement of the conversion gain of tube-type mixers is a simple procedure requiring only a calibrated rf signal input with a method of measuring the output IF current through a resistive terminating load.

In a crystal mixer the conversion loss is a function of the local oscillator power, measured in terms of the d-c crystal current. This relationship is especially sensitive where the local oscillator power is low. Consequently it is necessary that the proper level of excitation be used. If the power level is considerably different from the recommended value, biasing will occur which will vitiate the measurement. Table 4 should be consulted for the recommended local oscillator excitation for silicon crystals. (Although germanium crystals are used for mixers, silicon is far superior. Noise figures differ by as much as 20 db between the two types.) The output impedance of the crystal is another important characteristic. The crystal mixer should be terminated in its characteristic impedance (sometimes called the IF admittance). Table 4, page 109, includes the recommended value of the terminating resistance.

The d-c incremental method of measuring the conversion loss can be expressed as follows: an amplitude modulated signal from an rf signal generator is supplied through a calibrated attenuator to the crystal mixer. The attenuator must be matched to the crystal mixer. The output of the signal generator is set to the proper power level as measured with a bolometer bridge. The mixer is terminated with its proper characteristic impedance in series with a milliammeter (see figure 27). The input power is then varied a small amount, and the resulting change in the output current is noted. The conversion loss is calculated from the following expression:

$$L_{db} = 10 \log_{10} \frac{(\Delta P)^2}{2 P R_L (\Delta I)^2} \qquad (58)$$

where

L_{db} = the conversion loss in db,

ΔP = the change in the input power from the local oscillator,

ΔI = the change in crystal current,

P = initial power from the local oscillator (table 4),

R_L = the load resistance (table 4).

For example, assume

P = 1 mw,

ΔP = 0.1 mw,

R_L = 400 ohms,

ΔI = 0.2 ma,

then L_{db} = 5.06 db.

Table 4 contains the value to be expected from a number of silicon diodes used as mixers. If conversion losses are in excess of these values, that is, if the measured values are greater by more than 1 db from those in the table, the crystal should be replaced.

b. Noise Figure

Especially where a receiver is to be operated without rf amplification prior to the mixer, it is important to determine the effect of the mixer on the overall noise figure. Where tube-type mixers are used, the noise figure can be defined as

$$F = F_1 + \frac{F_2 - 1}{G_1} \tag{59}$$

where

F_1 = the best noise figure at the rf frequency,

F_2 = the noise figure of the IF amplifier,

G_1 = the conversion power gain of the mixer (less than one).

This relationship is valid exactly only for diode mixers. The noise figure of receivers using triode and pentode mixers is also a function of the loading of the mixer. This expression, however, is a good approximation for any kind of tube-type mixer.

TABLE 4

Crystal Type	Local Oscillator Excitation (milliwatts)	Terminating Impedance (ohms)	Test Data: Max. Conv. Loss db	Test Data: At Test Freq. (Mc/s)
		S-Band (300 - 4000 Mc/s)		
1N21	0.5	400	8.5	3060
1N21A	0.5	400	7.5	3060
1N21B	0.5	400	6.5	3060
1N21C	0.5	400	5.5	3060
		X-Band (4-12 Gc/s with ± 1.0 db nf)		
1N23	1.0	400	10.0	9375
1N23A	1.0	400	8.0	9375
1N23B	1.0	400	6.5	9375
1N23C	1.0	400	6.0	9375
		L-Band		
1N25	1.25	200	8.0	1000
		K-Band (16-25 Gc/s ± 1.0 db nf)		
1N26	1.0	500	8.5	23948
1N26A	1.0	500	7.5	24000
1N26B	1.0	500	7.5	24000
		Ku-Band (7-20 Gc/s with ± 1.0 db nf)		
1N78	1.0	500	7.5	16000
1N78A	1.0	500	7.0	16000
1N78B	1.0	500	6.5	16000
1N78C	1.0	500	6.0	16000

Where a crystal mixer is used, the expression for the overall noise figure involves the concept of noise temperature as follows:

$$F = \frac{t + t_1 + F_2 - 1}{G} \tag{60}$$

where

- t = ratio of the effective noise temperature of the crystal to the ambient temperature (assumes local oscillator to be noiseless),
- t_1 = the ratio of the increase of the effective noise temperature to the ambient temperature (assumes a practical local oscillator with some noise),
- F_2 = the noise figure of the IF amplifier,
- G = the conversion gain of the mixer (less than one).

The methods for measuring the noise figure of devices using mixers as the front end is identical to that used in the preceding section on the measurement of the noise figure of rf amplifier. A pre-selector should be added prior to the mixer for image rejection. It should be apparent that a receiver having a mixer with less-than-unit conversion gain as the front end is not efficient with regard to optimum noise figures; the noise figure of the IF amplifier becomes a significant contributor to the overall noise figure. Radiofrequency amplifiers, especially of the low noise varieties, will offer significant improvement in the noise figure, especially at UHF and above.

c. Excitation Requirements

The excitation requirements of a mixer are important. In order that the mixer may work properly, it is necessary that ample local oscillator power be available. However, too much local oscillator power can be harmful since the noise figure is not a linear function of the local oscillator power. For example, in the case of the crystal mixer, a minimum noise figure occurs where the crystal current equals 0.3 to 0.5 ma in the general case. In general, if the crystal current is maintained within these limits, the noise figure will tend to be optimum. Note that some compensation must be made between the optimum noise figure and the optimum conversion loss. However, the local oscillator should be capable of considerably more power than is required to provide optimum mixer operation. This permits more decoupling between local oscillator and the mixer, providing increased local oscillator frequency stability.

9.4. The Local Oscillator

The local oscillator (LO) supplies the signal to convert the rf signal frequency to the IF frequency. In some cases the frequency is generated in a single stage. At the low frequencies the problem is somewhat simple, and conventional negative grid tubes can be used. Above 1000 Mc/s a lighthouse tube or a similar type is used, and above 4000 Mc/s reflex klystrons are in almost universal use. However, if extreme frequency stability is required, the crystal oscillator is probably the best source of the primary frequency. Since the upper frequency limit of crystal oscillators is relatively low, multipliers are almost always used with them to attain the desired frequency of the local oscillator. Where crystal oscillators and multipliers are used, the reader is referred to the section on FM transmitters for a consideration of frequency multipliers.

The frequency stability is probably the most important parameter when considering the local oscillator. This is as true of the receiver as it is of the transmitters, where tolerances are held close regardless of the type of transmission. In normal AM reception the oscillator may drift as much as several kc/s with little noticeable effect except for a small decrease in the gain. The modulation intelligence will remain the same because the sidebands still retain the same relationship to the carrier. In SSB transmission, however, the requirements for frequency stability are somewhat stringent. It has been pointed out that a 50 c/s difference between the transmitter and receiver carrier can cause a loss in intelligibility, which implies a frequency change no greater than 25 c/s at the receiver. This would imply a frequency stability of the order of one part in 10^8 at 400 Mc/s. Present FM systems can operate satisfactorily with stabilities of the order of one part in 10^6.

If extremely close frequency tolerances are desired, a crystal oscillator should be considered. Crystal oscillators using premium components, good mechanical construction, and stable crystal ovens with a high degree of temperature compensation can yield stabilities of 5-8 parts in 10^{10} per day. Such oscillators are commercially available.

Even if reflex klystrons are used to supply the actual local oscillator signal, a crystal oscillator can be used to control the frequency of the klystron. A highly stable oscillator will maintain sufficient accuracy to satisfy the requirements of FM and conventional AM. Only in the case of SSB is there a danger of lack of adequacy. The use of a crystal oscillator also renders the system somewhat inflexible as far as operating frequencies are concerned.

Many receivers are not fixed frequency and the first local oscillator must track with the incoming signal. A highly stable variable oscillator must be used. Such a system may employ automatic frequency control (AFC). High degree of stability is possible using a good AFC circuit which is controlled by a pilot transmitted with the signal. The advantage of AFC is that it can compensate for transmitter drift. Its disadvantage is that it depends on the existence of the carrier at all times especially where a suppressed carrier is used. If the carrier is absent in times of very low signal, a time constant can be employed in the AFC circuit to eliminate, at least partially, this disadvantage.

9.5. Automatic Frequency Control (AFC)

Automatic frequency control is used to stabilize the frequency of the receiver. Since the chief cause of frequency instability is that attributable to the local oscillator, the control is usually accomplished in terms of the control of the frequency of the local oscillator.

The most common method of achieving AFC is the difference frequency method which utilizes the difference between the carrier and the local oscillator frequencies to provide a correction signal to the local oscillator thus maintaining a constant IF frequency. A loop is completed between the output of a discriminator and the local oscillator. The frequency of the local oscillator is adjusted to maintain

AUTOMATIC FREQUENCY CONTROL

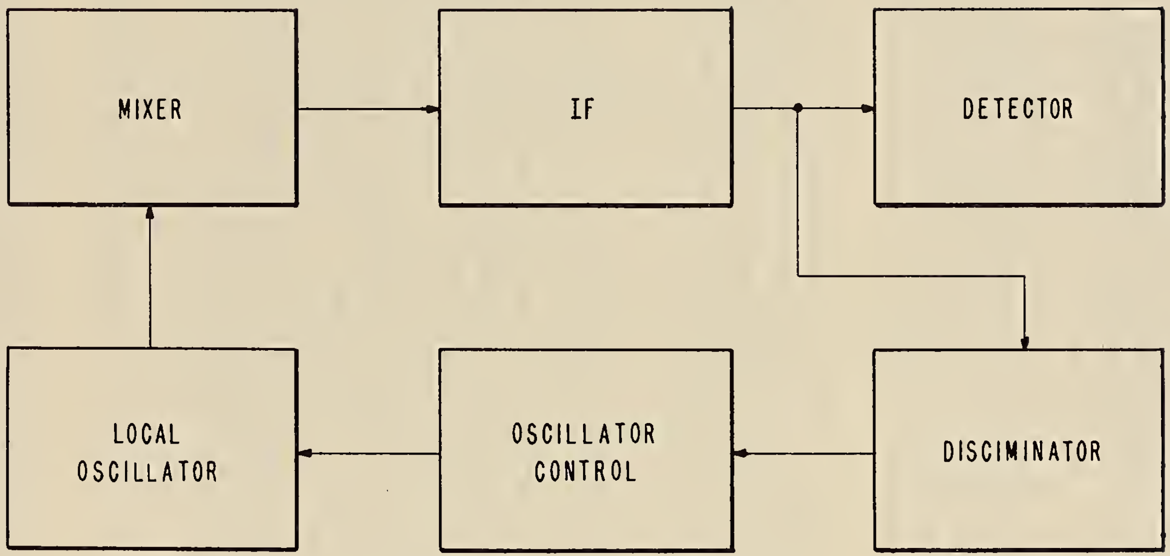


Figure 28

the frequency of the IF amplifier. The correction voltage is developed when the change in the IF frequency is due to either the local oscillator or the transmitted carrier.

Figure 28 illustrates a typical AFC loop, which employs the frequency difference method. The output of the IF amplifier is applied to a discriminator, whose output becomes the correction voltage for the local oscillator. The slope of the discriminator curve determines the correction voltage as a function of the frequency deviation. The correction voltage is either applied directly to the local oscillator or amplified and then applied. In most cases the correction voltage is applied to a reactance tube, acting as a variable capacitor in the oscillator tank circuit. In some applications a servo system is used to operate a variable capacitor or inductor. Recently variable capacitance semi-conductors have proven useful in AFC circuits.

In the case of SSB the difference between the carrier frequency as found from the pilot is compared to the frequency of a locally generated carrier. The same general principle applies. The two carriers are applied to a phase discriminator, the output of which is used to control the frequency of the local oscillator.

In some cases sophisticated circuits are used to search for the carrier when the transmitted carrier drifts out of range of the normally operating AFC circuit. These are used only when extensive carrier drift is to be expected or where poor quality oscillators are used on either end of the transmission path. Such circuits become operative whenever the carrier of the local oscillator drifts out of the "lock range" of the AFC. One disadvantage to such circuits is encountered where the fading is occurring near the threshold level of the receiver. When the carrier is unavailable, the search circuit begins to hunt, causing erratic action. This action can be improved by insertion of a delay to cause the search circuit to become operative only after a specified interval of carrier absence.

9.6. The Intermediate Frequency (IF) Amplifier

The intermediate frequency amplifier is a high-gain circuit located between the mixer and detector or demodulator. The main purpose of the IF amplifier is to provide gain and to restrict bandwidth.

The number of stages required is governed by the gain and bandwidth requirements. Gain requirements are dictated by the signal available and the output required at the detector or demodulator.

Bandwidth requirements vary with the type of system. FM receivers require broad bandpass characteristics, AM requires narrower bandwidth, and SSB requires the least of all. For example, FM systems require approximately 24 to 40 kc/s per voice channel as compared to 8 kc/s for AM, and 4 kc/s for SSB.

The choice of intermediate frequency is a compromise. A common rule for selecting the IF frequency is that it should be no more than 10% of the operating frequency. This allows for an optimum balance between IF selectivity and image response.

9.7. The Automatic Gain Control (AGC)

Since a scatter communication path is characterized by a signal that fades in intensity over a considerable range, it is especially important that adequate automatic gain control (AGC) be utilized. The primary function of the AGC is to prevent overloading of the IF amplifier at times of strong signal without losing the sensitivity at times of weak signal. In most receivers AGC is usually applied to the first IF amplifier. Since the purpose of AGC is to prevent distortion, care should be taken in applying this feedback voltage to the IF amplifier. If the feedback is applied to an adequate number of stages, strong signals may produce so much bias that the amplifier will be essentially at cutoff. This action causes distortion. When AGC bias is applied only to the first one or two amplifier stages, the signal-to-noise ratio may be impaired when signals are weak and noise is controlling the AGC. On the other hand, control must be applied at a sufficient early point in the amplifier to preclude the possibility of overloading or limiting ahead of the control point. The optimum arrangement allows a signal to obtain a reasonable level, then holds this signal level within a linear region until the final amplification to the desired level for demodulation.

In some diversity systems the AGC accomplishes an additional objective. The combiner in a diversity system depends upon the output of several receivers. For proper combining action, it is necessary that proper balance be maintained among the different receivers. In times of high signal, when the signal-to-noise ratio is high, the error rate is low even if the balance between receivers is not optimum. However, when the signal is low, the balance between the receivers is very important. Also, the energy in one of the passbands may consist of more noise than signal and the AGC is operated chiefly on noise; this results in AGC action to maintain the noise level constant. Consequently, it is advantageous to maintain the gains of the receivers such that their noise outputs are equal.

Due to individual differences between receivers, each receiver should be adjusted so that the AGC action of several receivers are identical. The noise level of the receiver may then be set to approximately the same value. The signal-to-noise ratio of weak signals is more sensitive to the setting of the AGC than it is to the setting of the gain. Small departures from identical gain characteristics usually have little effect on the diversity action and, consequently, little effect upon the signal error rate.

9.8. The Limiter in FM Receivers

Since the IF stages amplify both FM and AM signals, the limiter stage has as its prime function the elimination of amplitude-modulated noise pulses that are superimposed on the FM signal prior to detection. Complete limiting requires the removal of these noise peaks from both the positive and negative portions of the signal voltage. The most widely used method of obtaining this limiting action uses grid-leak bias and low screen and plate voltage to limit plate current on the positive and negative peaks of incoming signals.

Incomplete limiting is experienced when the signal level drops below the threshold. This can be observed when AM noise pulses appear at the detector output with a drop in the input signal level.

9.9. Demodulators

The function of the demodulator of the receiver is to recover the amplitude and frequency of the modulating wave that was impressed on the carrier at the transmitter. Since the treatment here has been confined to FM and SSB operation, only these systems will be considered.

SINGLE — SIDEBAND SIGNAL DETECTION

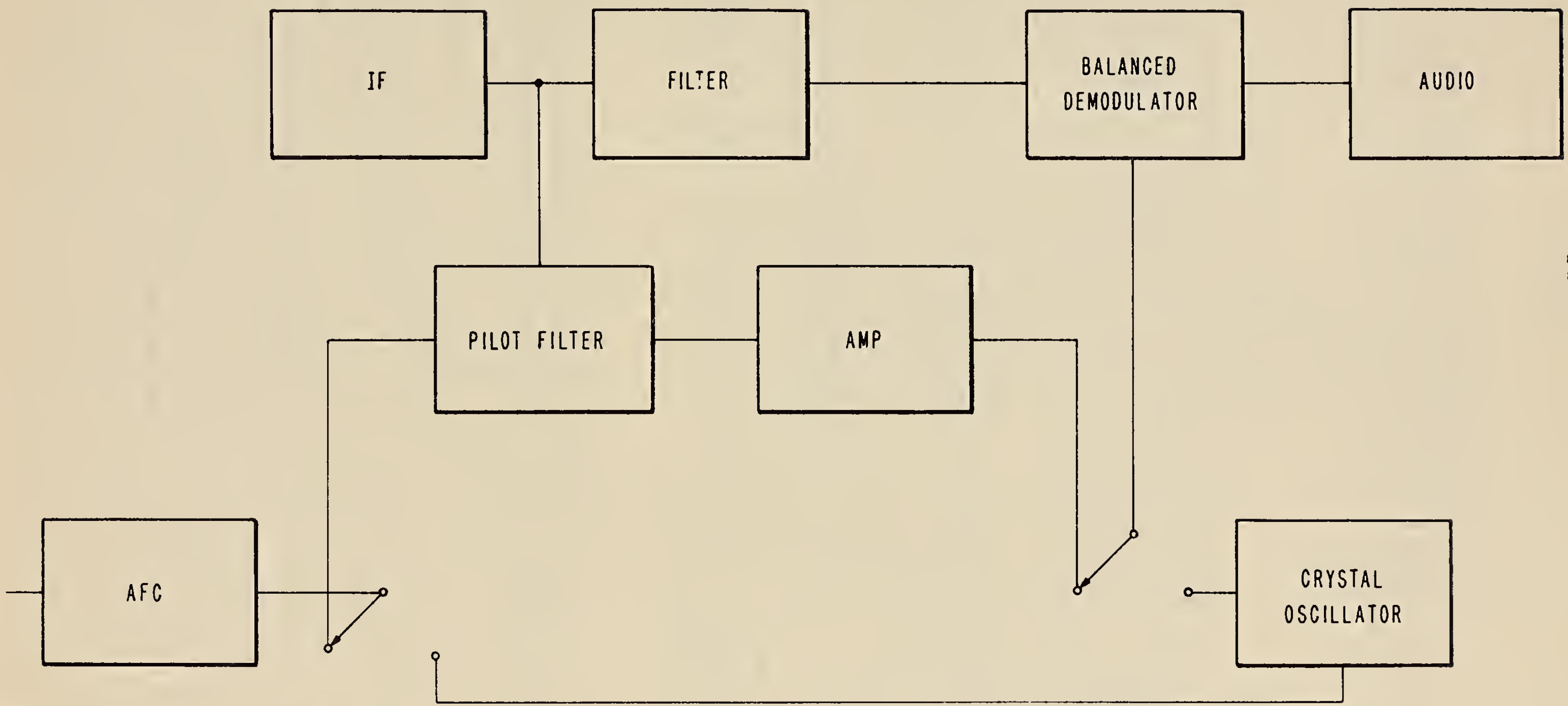


Figure 29

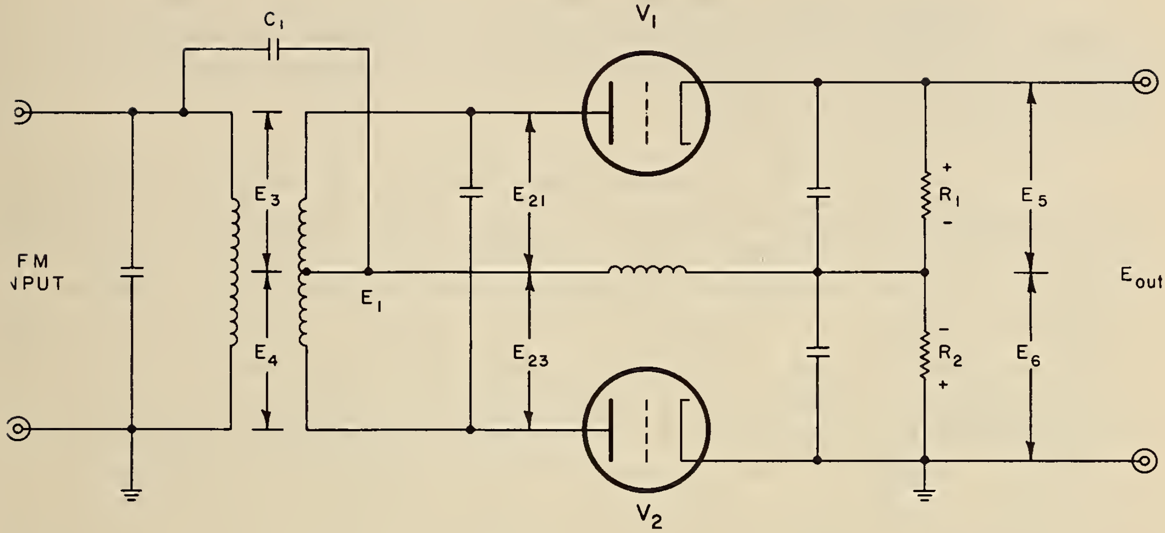
The standards of performance for this part of the receiver are considered extensively by Florman and Tary [1962] . This section will review the methods of demodulation and the general principles involved.

a. SSB Demodulation

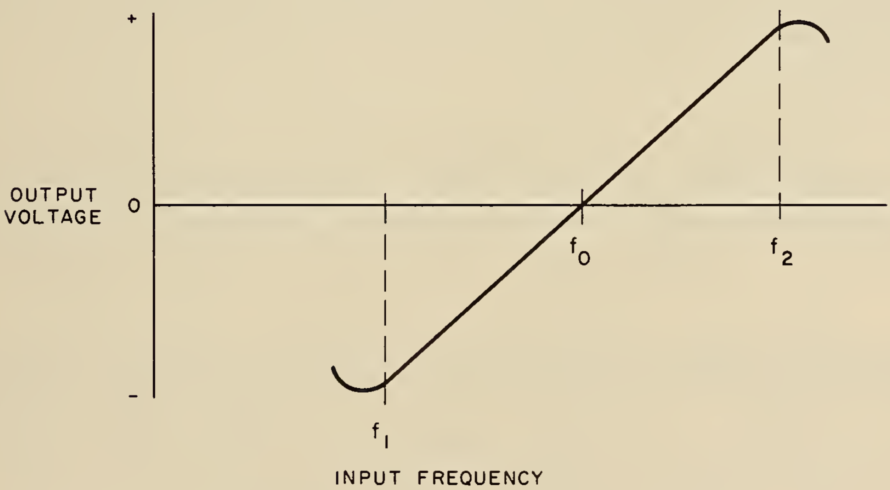
To recover the information on a SSB signal, it is customary to reconvert it to an AM signal and demodulate it in conventional fashion. This demands that a carrier be resupplied to the SSB signal. This insertion of the carrier occurs at the demodulator. This carrier can be generated locally from an oscillator which is independent of the transmitter carrier, or it can be the received suppressed carrier or pilot signal. The former method has the advantage that it does not require the presence of the pilot signal to accomplish demodulation; its disadvantage is that it is incapable of compensating for any carrier drift, and hence may be incapable of maintaining the necessary frequency relationship between sidebands and carrier. The latter method has neither the advantage nor the disadvantage of the former, and requires the presence of the carrier at all times. The carrier is usually suppressed below the level of the sidebands, and selective fading may at times affect the carrier to a greater degree than it does the sidebands. There may be times when the sidebands are available and the carrier is not.

It is practical to provide both methods, one to be used during times of strong signal and the other available for times of severe fading (see figure 29). The output of the IF amplifier is fed through a narrow-band pilot filter. The output of the filter is applied to a balanced demodulator through an amplifier which raises the level of the pilot carrier to a level roughly ten times the level of the sidebands. A balanced demodulator may be used for detection to reduce the possibilities of intermodulation distortion. The amplification of the carrier is necessary to prevent distortion due to overmodulation. Companion to this circuit would be an oscillator operating onthe carrier frequency having exceptionally good frequency stability. Since it was pointed out that a 50-c/s variation between the locally generated carrier and the transmitter carrier may cause serious distortion, the frequency stability of this oscillator must be exceptionally good. Stabilities less than 5 or 6 parts in 10^9 day are not usually considered adequate for frequencies of 1000 Mc/s or more. An AFC circuit can be incorporated in a system such as this as illustrated in figure 29.

THE PHASE-SHIFT DISCRIMINATOR



(A) SIMPLIFIED CIRCUIT



(B) THE OUTPUT VOLTAGE AS A FUNCTION OF INPUT FREQUENCY

Figure 30

b. FM Demodulation

The phase-shift discriminator is the most widely used of all forms of FM detectors in communication systems. Its chief advantage is its linearity over the total range of frequency deviation encountered in FM systems.

The phase-shift discriminator operates on the principal that the inductive voltage induced in a resonant circuit by another resonant circuit will vary in phase as the frequency of the inducing voltage varies. Figure 30(A) is a typical simplified diagram of the phase-shift discriminator. Changes in input frequency vary the phase angles between the inductively coupled voltages and the primary voltage which is capacitively coupled. The phase-shift discriminator converts these variations of phase into variations of amplitude. The input voltage is coupled to the center tap of the transformer through a coupling capacitor C_1. This capacitor must be large enough to provide effective coupling so that E_1 is always present at the center tap.

The phase-shift discriminator thus converts an FM input of constant amplitude and changing frequency to an output which varies in amplitude in accordance with the input variation in frequency. Figure 30(B) is a typical characteristic curve for the phase-shift discriminator.

To give undistorted output the input to the phase-shift discriminator must be free of any variations in amplitude. Therefore, in all FM systems the input to the phase-shift discriminator is preceded by one or more limiting stages to eliminate variations in amplitude.

The need for limiters in some types of FM receivers is eliminated by using a variation of the phase-shift discriminator called a ratio detector. The basic difference between the phase-shift discriminator and the ratio detector is the fact that in the ratio detector the polarity of the two diodes are arranged so that their load voltages are additive. Thus, the output voltage is proportional to the ratio of the voltages at the diodes. Amplitude variations in the signal input will increase or decrease these voltages but their ratio will remain constant for a particular frequency.

9.10. The Squelch

Squelch is employed in some systems to disable the output of a receiver when the signal fades into the noise. The squelch circuit, as ordinarily used, is merely a tube that is biased to cut off so that nothing can pass through when no signal is being received. As soon as a signal is received, part of the voltage built up across the diode load resistor is used to remove the blocking bias, so that a signal can be passed. Since the voltage due to the signal is higher than that due to noise, the squelch tube can be so biased that it will not operate on noise voltage but will go into action as soon as the higher signal voltage is received.

Over-bias setting will desensitize the receiver so that it will require a large signal to remove the blocking voltage. Under-bias will permit noise peaks to trigger the squelch circuit so that bursts of noise will ride through.

10. LOW NOISE DEVICES

With increasing requirements for reliable communications over long-distance tropospheric scatter circuits, attempts have been made to improve the performance of the receivers operating on these circuits. In this way definite improvement in the noise figure of the receiver will show a marked improvement in the signal-to-noise ratio of the entire communications system which includes the antenna and transmission line as well as the entire receiver itself. Thus, since the ultimate objective is the overall system performance, improvements in the individual components, to such an extent that they are no longer the predominant source of noise, is important.

Such devices which will accomplish this for the receivers are now available in the form of parametric amplifiers, MASERS, and most recently tunnel diodes.

10.1. Parametric Amplifiers

In general, parametric amplifiers or reactance amplifiers may be classified as regenerative, up-converters, traveling-wave diodes or electron-beam amplifiers. Many of these devices possess common

PENDULUM ANALOGY TO
PARAMETRIC AMPLIFICATION

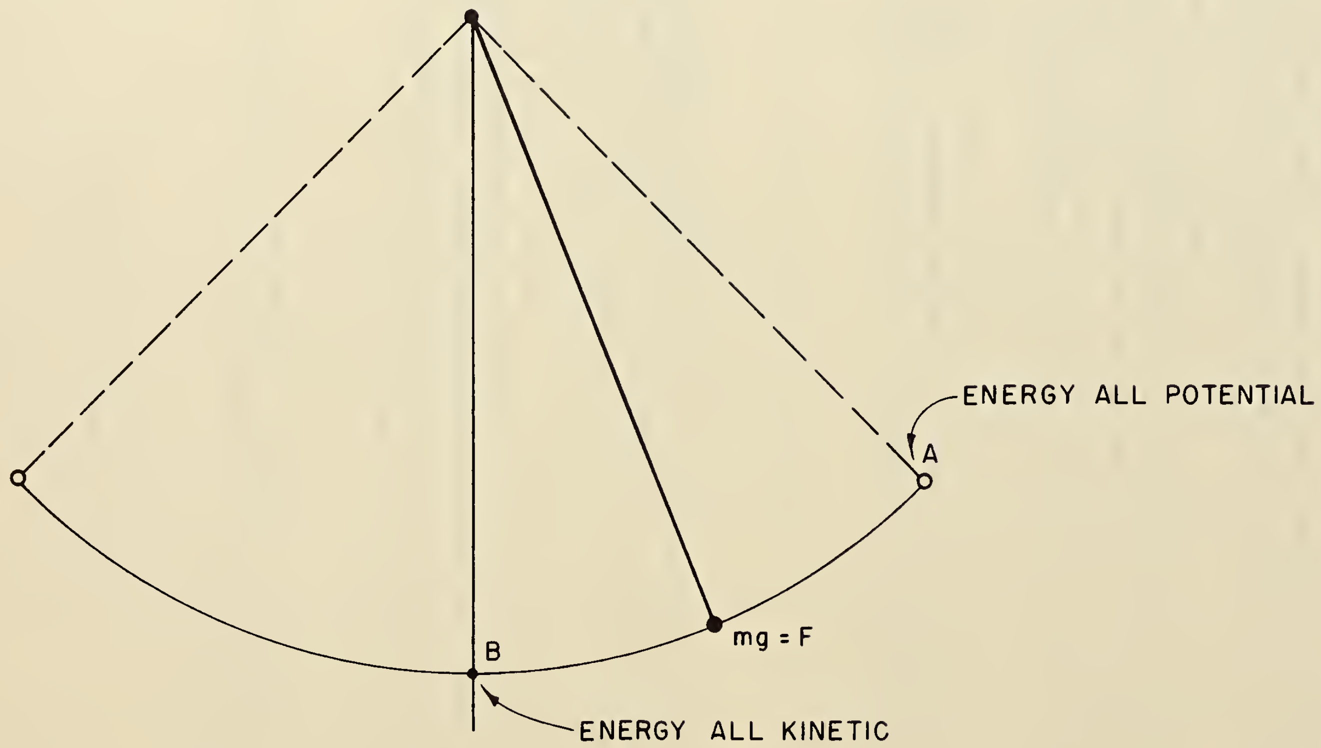


Figure 31

RLC CIRCUIT ANALOGY TO PARAMETRIC AMPLIFICATION AND RESULTING WAVEFORMS

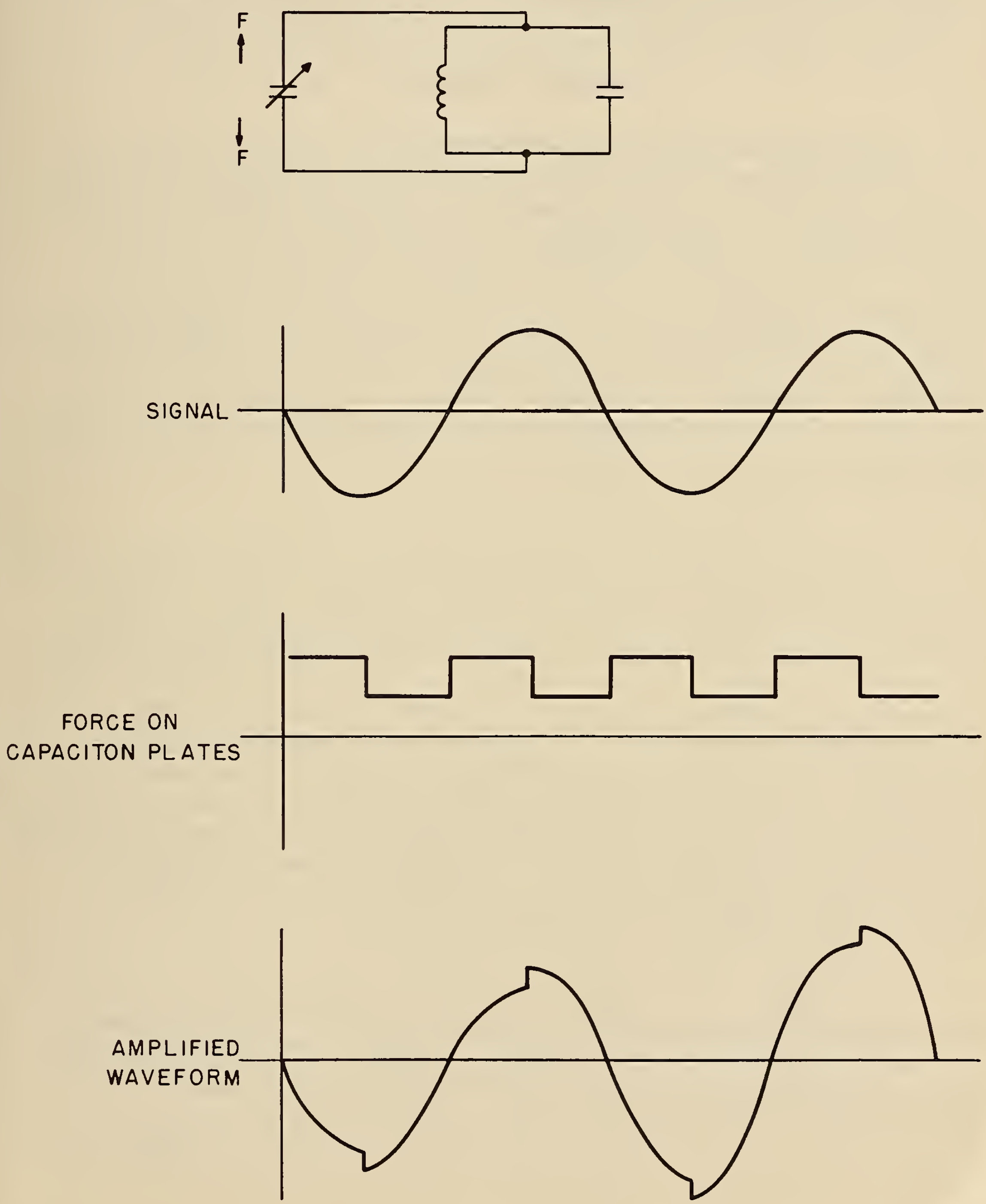


Figure 32

features in that they do not obtain power from a d-c supply from a time-varying rf signal supply. The signal output is obtained from this rf drive by means of a nonlinear time-varying element termed the pump.

a. Theory of Operation

The physical operation of a parametric device can be related to the swinging pendulum as described by Russell [1960]. Referring to figure 31, the ball at the end of the string has its highest potential energy at point A, while at point B, its lowest point, all energy is kinetic. If the string is shortened while the ball is at point A, the parametric concept takes place, in that the potential energy is increased and work is being done on the system. The string is returned to its normal length when the ball reaches its lowest position (point B) and potential energy is converted to kinetic energy with the result that the velocity of the ball is greater at this point than it was before. Repeating this process will result in an increase in the overall amplitude with more energy being transferred to the system.

A simple method used by Reed [1959] to illustrate the principle of variable-capacitance parametric amplification is shown in figure 32. If a small signal voltage s at the resonant frequency is applied to the circuit, a sinusoidal variation of voltage and charge is set up across the capacitor. Since it is desired to amplify the signal voltage and since the capacitor voltage is inversely proportional to the capacitance by the expression $V = \frac{q}{c}$, separation of the capacitor plates will increase the capacitor voltage for a given charge. If the plates are separated at the maximum sinusoidal displacement and returned at the time when there is no charge on the plates (when the wave crosses the axis), work will be transferred to the signal as shown in figure 32. For maximum energy transfer the capacitance must be varied or pumped at exactly twice the signal frequency.

In practice, it is difficult to maintain an exact phase relationship between the uncontrolled signal and pump oscillator frequency. An exact condition is not necessary if something less than maximum energy transfer is acceptable. If the frequency of the signal f_s is different from half the pump frequency $f_p/2$, conditions for maximum energy transfer are no longer satisfied and the pump and signal will now vary in their phase relationship. However, there will be a net flow of energy from the pump to the signal resulting in an amplified signal exhibiting a beat phenomenon and modulated waveform. This type of waveform consists of the original signal f_s and a lower and upper sideband equal to the sum $f_p + f_s$ and difference $f_p - f_s$ of the pump and signal frequency. These sidebands are usually referred to as the idler, image, or difference frequency.

The idler is unavoidable and cannot be suppressed without also suppressing the entire amplification. Also, the closer the idler is to the signal, the more difficult it is to separate it from the signal by filtering. In the practical amplifier it is desirable to keep the signal frequency a few megacycles away from half the pump frequency. This means that the parametric amplifier must be broadband to contain the signal and idler or image bands. Therefore, the amplifier must provide twice the bandwidth and amplify twice the noise associated with conventional amplifiers. However, by proper filtering action either the signal or idler frequencies can be separated.

b. The Semiconductor Variable Capacitance

A semiconductor has certain properties which enable it to act as a time-varying capacitance whose value is a function of the voltage across it. The voltage or pump source is in the microwave frequency range.

The P section of a diode has a high density of free holes or positive charge carriers, while the N section is rich in free electrons or negative charge carriers. With the P and N sections in contact, electrons diffuse from a region of high to low electron density while the holes diffuse across the junction from the P to N material. The resulting increase of negative charge of the P section and increase of positive charge in the N material causes a potential gradient or electric field at the interface so as to oppose and stop further diffusion of electrons and holes across the junction. As a result, a narrow region at the interface called the depletion layer will be swept free of mobile charge carriers. This is a non-conducting or dielectric region, bounded on both sides by conducting regions. This is the criterion for a parallel plate capacitor having a separation equal to the depletion layer width.

EQUIVALENT CIRCUIT OF A VARIABLE-CAPACITANCE DIODE

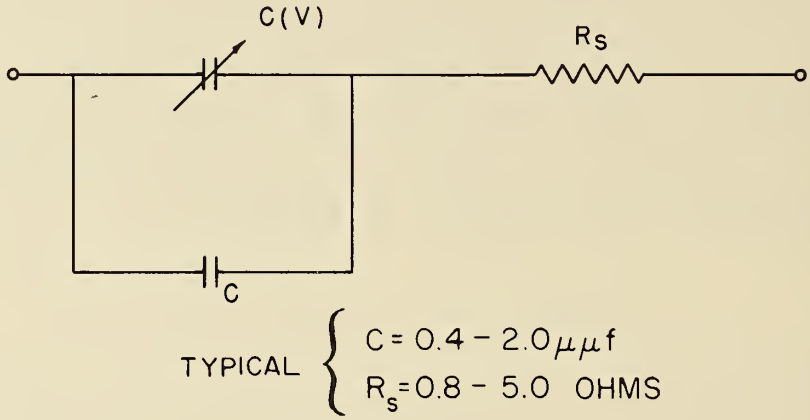


Figure 33

SINGLE-TANK NEGATIVE-RESISTANCE AMPLIFIER

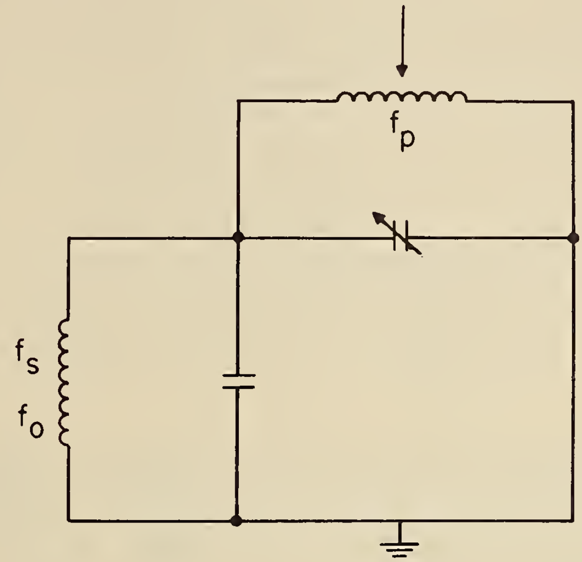


Figure 34

TWO-TANK NEGATIVE-RESISTANCE AMPLIFIER

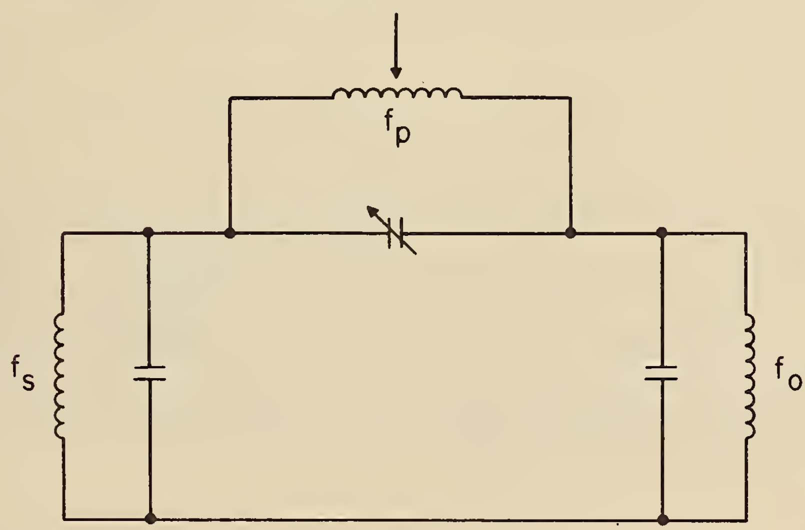


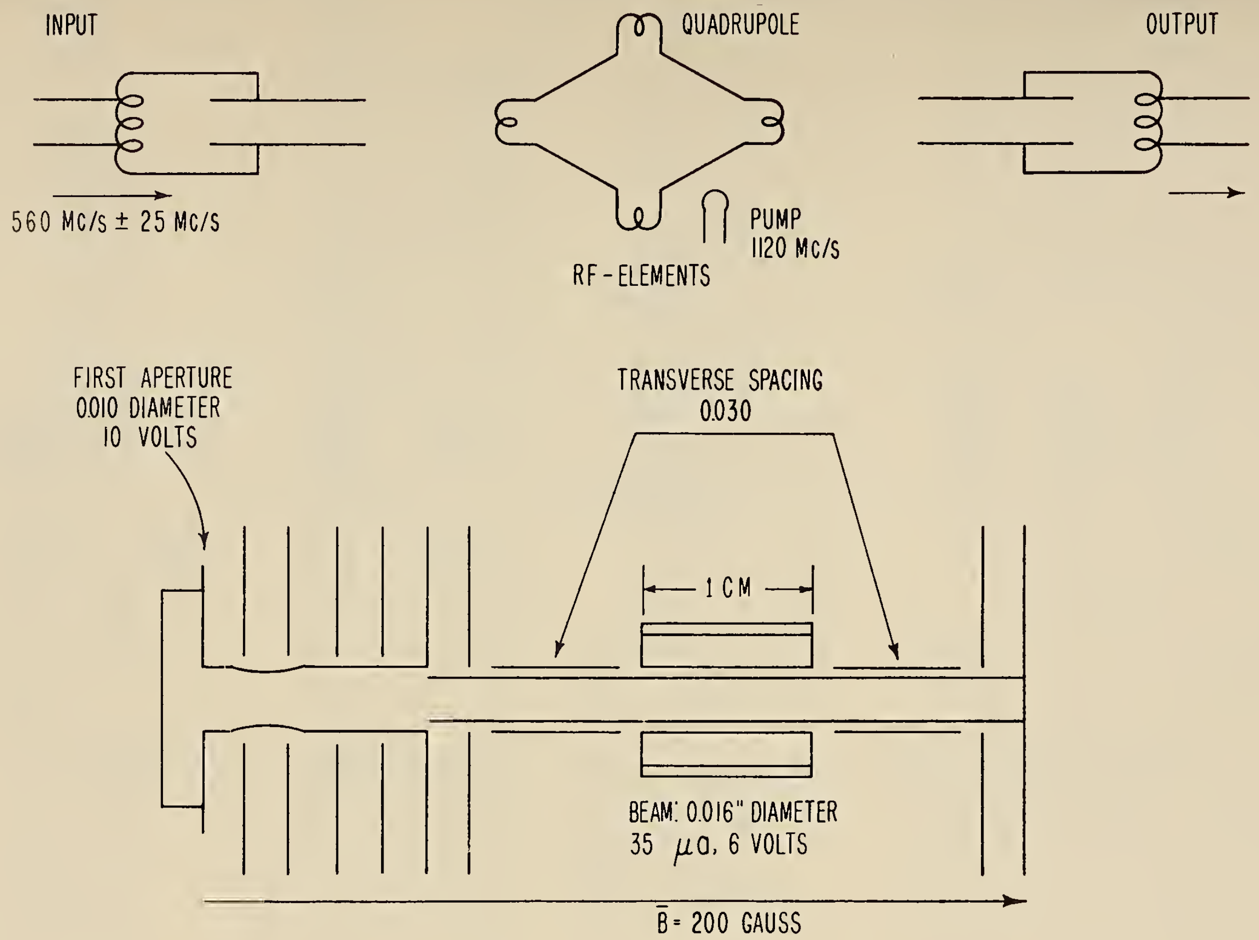
Figure 35

If the junction is now given a reverse bias, which means the negative battery terminal is connected to the P section and the positive terminal to the N section, then the electrons will move to the positive terminal while the holes move to the negative terminal, causing an increase in potential difference across the junction. As a result, the depletion layer increases in width, resulting in a decrease of capacitance and an increase in terminal voltage. Similarly, a forward bias will force the majority carrier distributions toward each other, the depletion layer will shrink, and the capacitance will increase. Thus, we have a capacitor whose terminal capacitance will vary with the applied voltage. The variable capacitance diode is an excellent device for both high-frequency operation and low noise while the transistor, consisting of similar materials, does not excel to the same extent.

Frequency limiation of the variable-capacitance diode as described by Uhlir [1958] is nearly unlimited since the change in width of the depletion layer under an applied voltage is so small that transit-time effects are negligible up to high frequencies. Before transit-time effects are reached, two limitations arise as is shown in the equivalent circuit in figure 33. C(v) is the variable capacitance while C represents the fixed capacitance which is a function of the contact area depletion-layer width, the applied bias, and the type of encapsulation. R_S represents the diode resistance due to the impedance of the semiconductor material. Consequently, at very high frequencies C can shunt out the variable-capacitance voltage with a resultant power loss and increase in the thermal noise in the resistor R_S.

c. Negative Resistance Amplifier

Figure 34 shows a generalized scheme in which straight-through amplification can be achieved, as described by Bateman and Bain [1958-1959]. One tank is tuned to an external pump oscillator frequency. When power is applied to the pump tank the diode capacitance will vary at the pump frequency rate around the capacitance value associated with the particular value of applied reverse bias voltage. The varying capacitance allows some of the pump power to be released to the signal and output tanks.



AN EXPERIMENTAL ELECTRON-BEAM AMPLIFIER
(ZENITH CORPORATION)

Figure 36

With the pump frequency approximately equal to twice the signal frequency, an amplified version of the signal f_s and the difference or idler frequency $f_p - f_s = f_o$ appears in the single tank. The circuit may then be used as an amplifier or converter depending on whether f_s or f_o is utilized. Both f_s and f_o have equal amplitude and simutaneously appear at the single tank output terminals. In wideband operation, it is recommended that f_p be located at a higher frequency, thus preventing f_s and f_o from being within the passband. This will also move the image of idler frequency f_o up with a result of no gain until the idler is separated by tuning another tank circuit to f_o, thus effectively discarding this source of energy. The schematic of the two tank circuits is also shown in figure 35.

d. The Up-Converter

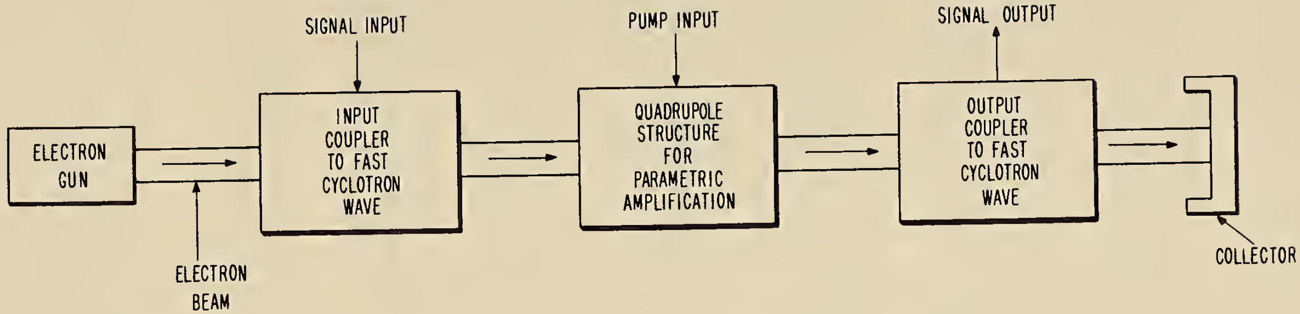
The up-conversion or sum-frequency amplifier differs from the negative-resistance amplifier in that the output frequency is the upper sideband $f_p + f_s$ where p is the pump frequency and s the signal frequency. In contrast, the negative-resistance amplifier utilizes the lower sideband $f_p - f_s$ only.

Since the gain is proportional to the frequency ratio $(f_p + f_s)/f_s$, the pump frequency must be many times greater than the signal frequency. In the up-converter, the signal frequency is shifted in the amplification process while the negative-resistance amplifier may be used at the original frequency or at a lower sideband frequency. The up-converter has unconditional stability whereas the negative-resistance amplifier requires either exceptionally good matches or a circulator for stable operation.

e. The Electron-Beam Amplifier

The electron-beam amplifier is sometimes referred to as a quadrupole amplifier since the principle of operation is based on a tube in which quadrupole fields act parametrically on the fast cyclotron wave of an electron beam to produce amplification.

Figure 36 is a block diagram showing the operating principle of the electron-beam parametric amplifier developed by R. Adler [1958] of the Zenith Radio Corporation. The beam is surrounded by



OPERATING PRINCIPLE OF THE ELECTRON-BEAM AMPLIFIER

Figure 37

a uniform magnetic field and flows parallel to the flux lines. The magnetic field intensity produced within the tube is such that the corresponding cyclotron frequency is approximately equal to the signal frequency to be amplified.

The input coupler extracts noise and inserts signal simultaneously by demodulating the noise components appearing on the electron beam and modulating the electron beam with a new fast wave equal to the input signal. The modulated electron beam then passes through the quadrupole structure which is fed by a pump signal equal to twice the cyclotron frequency where parametric action of the quadrupole field on the modulated beam results in amplification. The output coupler merely extracts the amplified signal from the electron beam.

There are several similarities and differences between electron-beam and diode-type parametric amplifiers which should be pointed out. In the electron-beam amplifier if the pump frequency is not synchronous with twice the signal frequency, an idler frequency is formed which is very close to the desired signal frequency and will appear as a beat on the desired amplified signal. This beat frequency may range from a few cycles per second to several megacycles per second, depending on the idler frequency generated by the pump and signal frequencies. For communications purposes this unwanted beat frequency is eliminated by what is called synchronous pumping where the pump frequency is exactly twice the signal frequency and optimum phase control between the two frequencies is maintained. However, experiments by Adler et al. [1958] have shown that the electron-beam amplifier is a wideband device with good gain stable qualities when compared with many cavity-type, solid-state amplifiers.

An experimental type electron-beam amplifier constructed by the Zenith Corporation and described by Adler et al. [1959] is diagrammatically shown in figure 37.

The signal frequency was 560 Mc/s, and the pump frequency was 1120 Mc/s. The rf elements are shown together with a schematic of the tube structure. The electron gun consists of a series of small apertures where the first aperture selects a small portion of the electron stream, and the potentials between succeeding apertures cause this beam to diverge and become parallel to the axis. The last two apertures reduce the beam to a diameter of 0.016 inches. The beam voltage was 6 volts at 35 microamperes to obtain 6 orbits in the coupler and 4 orbits in the quadrupole. The magnetic field was 200 gauss corresponding to a cyclotron frequency of 560 Mc/s. The amplifier has the following characteristics:

Gain	=	30 db
Bandwidth	=	40 Mc/s
Noise figure	=	1.5 db

f. General Considerations

The operation of the up-conversion amplifier may be considered analogous to the Manley-Rowe theorem [1956] describing the fundamental power relation for nonlinear (and lossless) reactance. This theorem states that when a strong pump frequency p and a weak signal of frequency s are simultaneously impressed on a nonlinear reactance, the following power relationship result:

$$\frac{P_{f_s}}{f_s} + \frac{P_{(f_p + f_s)}}{f_p + f_s} - \frac{P_{(f_p - f_s)}}{f_p - f_s} = 0 \qquad (54)$$

where only the signal and the two lowest sidebands, $f_p - f_s$ and $f_p + f_s$, are considered, and values of P_{f_s}, $P_{(f_p + f_s)}$ and $P_{(f_p - f_s)}$ are the powers at the frequencies f_s, $f_p + f_s$ and $f_p - f_s$. A positive value of P denotes net power leaving the amplifier and a negative value denotes power absorbed by the amplifier.

For the negative-resistance amplifier, the circuit does not respond to the upper sideband $f_p + f_s$; therefore the middle term of (54) drops out, and we have the following relationship:

$$\frac{P_{f_s}}{f_s} = \frac{P_{(f_p - f_s)}}{f_p - f_s} \qquad (55)$$

Therefore, P_{f_s} must be positive in order to have amplification at the signal frequency f_s, and $P_{(f_p - f_s)}$ must also be positive, showing that if amplified signal power leaves the amplifier, image power or lower sideband power must also leave the amplifier.

From (55) the image power is related to the signal power by

$$P_{(f_p - f_s)} = \frac{f_p - f_s}{f_s} P_{f_s} \tag{56}$$

This suggests another mode of operation of the negative-resistance amplifier where the amount of gain is reduced at the signal frequency, and the overall gain is increased by extracting the useful signal at frequency $f_p - f_s$, which makes use of the frequency-ratio gain $(f_p - f_s)/f_s$. This has the effect that the amplifier stability and gain are less critically dependent on the quality of matches, and the image noise is reduced.

A negative-resistance amp ifier of this type has been built by G. Herrmann [1958] of the Bell Telephone Laboratories and is diagrammatically shown in figure 38. In this amplifier a certain amount of up-conversion gain is used in addition to the negative-resistance gain, and the lower sideband is used as the useful signal output. The amplifier has the following characteristics:

Gain	=	18 db
Bandwidth	=	5 Mc/s
Noise figure	=	1.2 db at 90°K

Another similar amplifier was constructed at the Federal Telecommunications Laboratories. The amplifier operates at a signal frequency of 900 Mc/s, pump frequency at 9900 Mc/s, and the useful lower sideband frequency of 9000 Mc/s with a gain of 18 db and a noise figure of 1 db.

If the circuit responds to the signal upper sideband only, then from the previous power expression $P_{(f_p - f_s)} = 0$ and then $P_{(f_p + f_s)} = \frac{f_p + f_s}{f_s} - P_{f_s}$. The negative sign shows that power out at one frequency

EXPERIMENTAL NEGATIVE-RESISTANCE AMPLIFIER

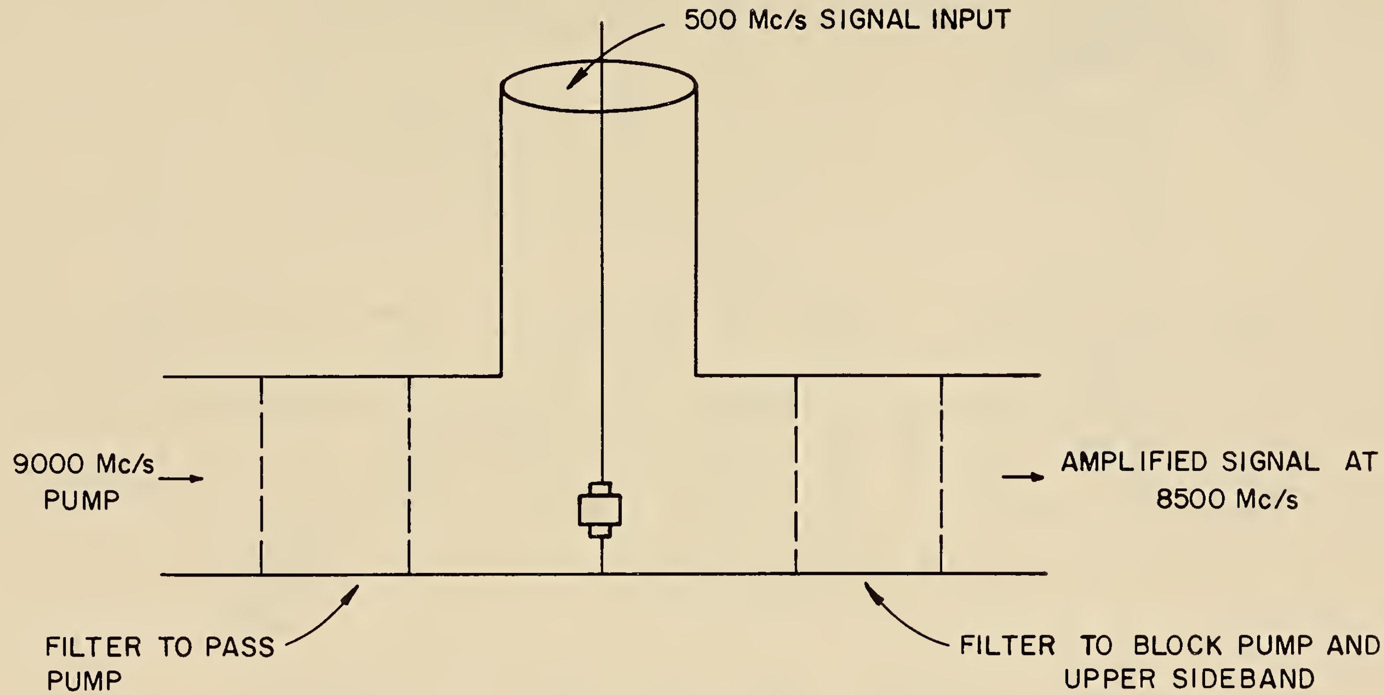


Figure 38

BLOCK DIAGRAM OF RECEIVER USING UP—CONVERTER PARAMETRIC AMPLIFIER WITH CRYSTAL MIXER

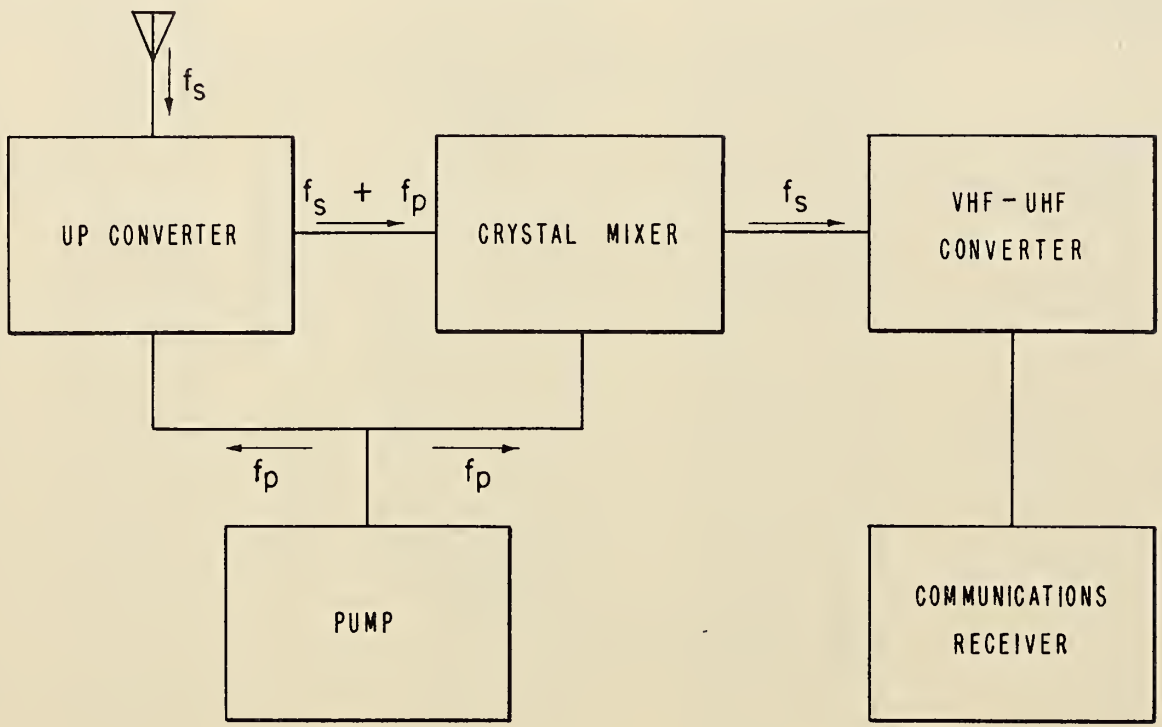


Figure 39

BLOCK DIAGRAM OF RECEIVER
USING DOWN — CONVERTER
PARAMETRIC AMPLIFIER

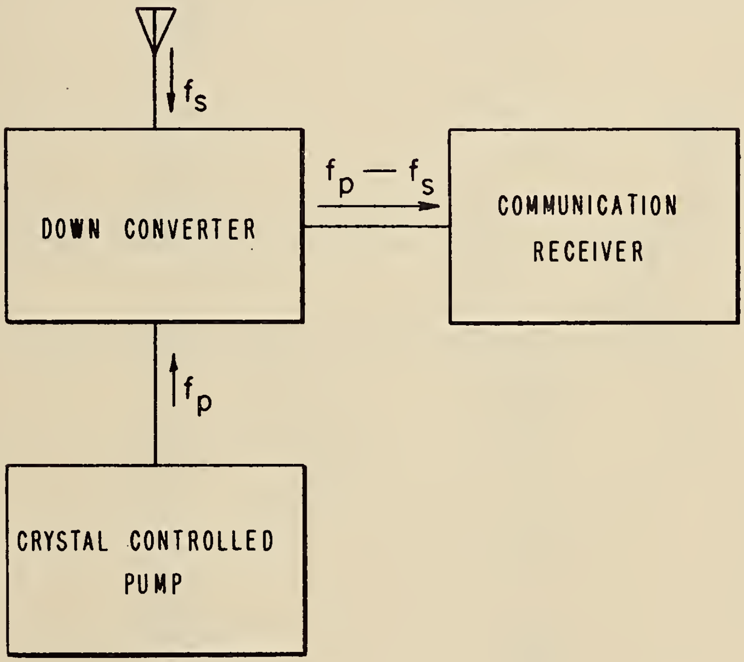


Figure 40

A FORM OF NEGATIVE-RESISTANCE AMPLIFIER
USING A VARACTOR DIODE
(HARRIS AMPLIFIER)

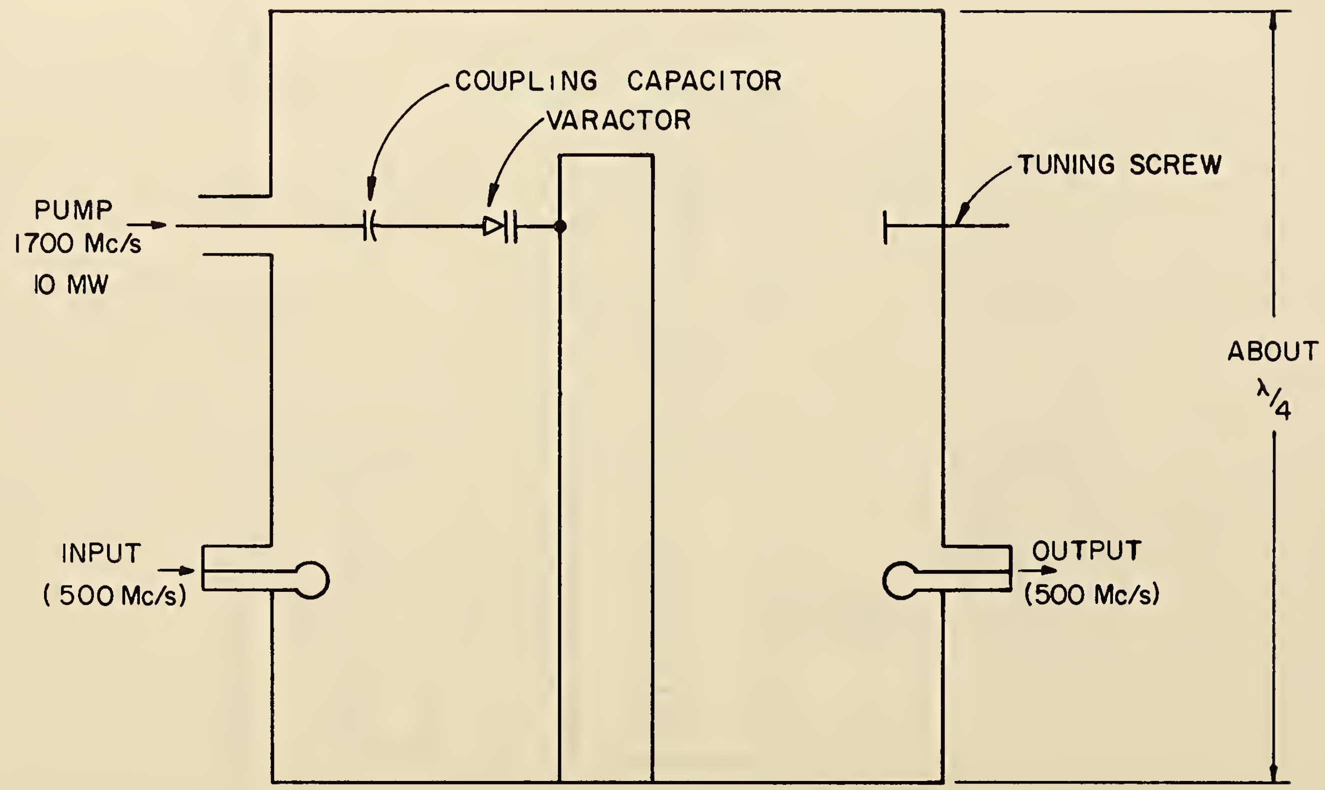


Figure 41

BLOCK DIAGRAM OF A REGENERATIVE AMPLIFIER
USING A VARACTOR DIODE

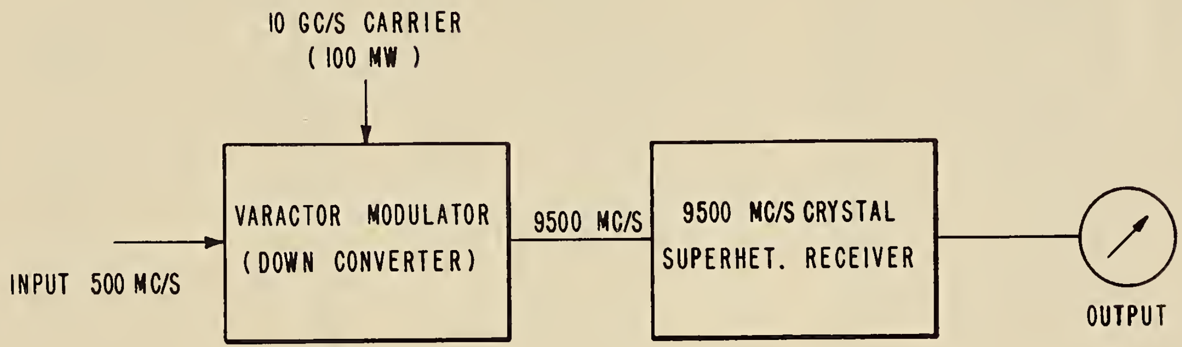


Figure 42

is desired, and power must be absorbed by the system at the other frequency. If Pf_s is put into the system, the $(f_p + f_s)/f_s$ times as much power can be extracted at frequency $f_p + f_s$.

Figure 39 shows a practical configuration using an up-converter and a crystal mixer in which the output appears at the signal frequency. The combination converter and mixer performs the same function as a straight-through amplifier. It has the desirable feature of permitting a self-controlled oscillator to be used for the pump without introducing frequency variations in the output. It has one disadvantage in that the overall noise figure will be somewhat greater than that which can be obtained from a two-tank amplifier using the same diode and pump frequency.

Figure 40 is a block diagram of a receiving system using a regenerative amplifier. As can be seen, few components are required for a complete system. In a regenerative amplifier, f_o is smaller than f_s; thus the ratio f_s/f_o may be quite large. To compensate for large values of f_s/f_o, it is necessary to have extremely tight coupling between the load and output tank. A further disadvantage is that very high regeneration is required to give useful gains. When f_o/f_s is small, a narrow bandwidth will result and the device will tend to have very poor gain stability. As a result, regenerative amplifiers with a high ratio of f_o/f_s may be difficult to adjust and maintain.

A low-noise negative-resistance amplifier now in production is shown in figure 41. It was developed and tested by F. S. Harris [1958] of Microwave Associates for use in the 400-500 Mc/s region. This amplifier uses one varactor diode mounted in a coaxial cavity. The pump frequency is several times the signal frequency. Separate coupling loops are provided for the input and output. The output frequency is exactly equal to the input frequency regardless of drifts in the pump frequency. Measured noise figures of approximately 1 db have been obtained in this frequency range.

A block diagram of a regenerative amplifier is shown in figure 42. The signal to be amplified (500 Mc/s) is mixed with several hundred milliwatts of 10 Gc/s pump or carrier frequency in the varactor diode. Sidebands are generated at 10,500 Mc/s and 9,500 Mc/s. With 500 Mc/s inputs upper sideband up-converter have noise figures of 2-3 db while regenerative amplifiers using the lower sideband have overall noise figures of 1 db.

10.2. MASER

The development of the MASER (Microwave Amplification by Stimulated Emission of Radiation) and parametric amplifiers has led to important contributions in the advancement of microwave engineering. Although the MASER in its present form is only a few years old and much research is yet required, it shows great promise in its application to such fields as radio scatter systems, satellite communications, and radio astronomy, where extremely low noise amplification is important.

These new amplifiers differ in principle and operation from the more familiar klystron and traveling-wave tube amplifiers in that MASERS utilize the property of bound electrons instead of free electrons, and radiofrequency energy rather than direct-current energy is used as a source of power.

Since the amount of noise produced in a device is a direct result of the temperature of the working substance and since no electrons are obtained from thermionic emission from a hot cathode, MASERS are among the lowest noise producing devices thus far developed.

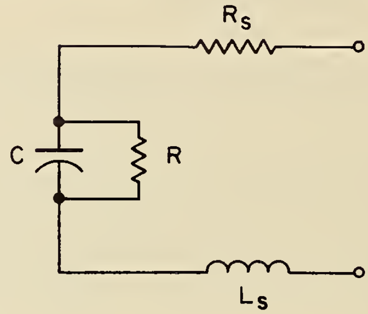
10.3. Tunnel Diode Amplifiers

Recent experimental research by Esaki [1958], Hall [1960] and Summers [1959] has shown that the tunnel diode exhibits properties which may render it quite useful as a low noise rf amplifier in scatter communications circuits.

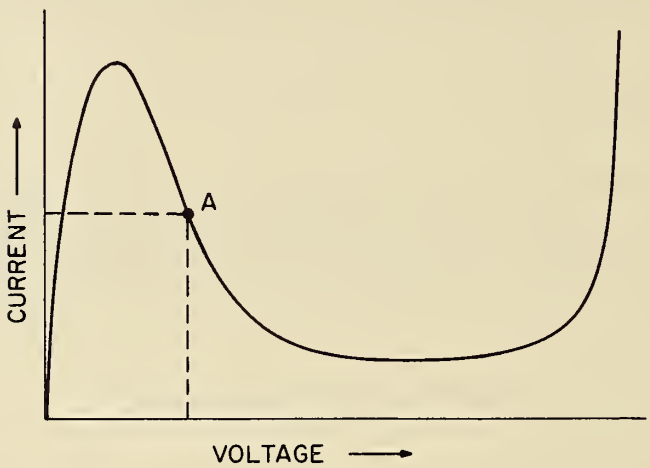
a. Theory of Operation

The current-voltage characteristics of the tunnel diode shows that a negative-resistance region exists in the direction of forward bias. This negative resistance, according to theoretical and experimental results, extends well into the high microwave region. The linear portion of the negative-resistance characteristic is utilized to construct linear rf amplifiers. Basically, this negative-resistance characteristic is identical to that found in some types of parametric amplifiers where the negative resistance is developed by the pump signal.

TUNNEL DIODE EQUIVALENT CIRCUIT



VOLTAGE – CURRENT CHARACTERISTICS



CIRCULATOR COUPLED TUNNEL DIODE AMPLIFIER

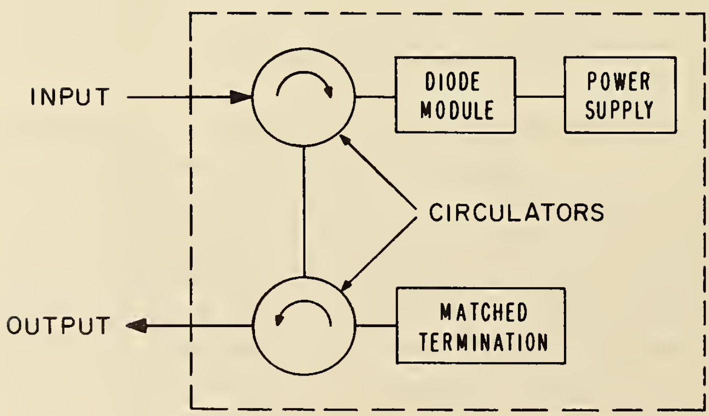


Figure 43

The equivalent circuit for the tunnel diode and the representative current-voltage characteristic curve is shown in figure 43. R is that resistance existing across the function for a given forward bias point indicated by A on the voltage-current characteristic curve. The ohmic losses in the leads and the diode itself are represented by R_S, and the associated lead inductance is indicated by L_S. The diode junction capacitance which shunts the negative resistance R is indicated by C. Therefore, the design of a linear rf amplifier is based on tuning the reactive elements of the tunnel diode with an external circuit similar to the equivalent circuit shown in figure 43. Amplification takes place when the coupling to the source and load are such that the equivalent positive conductances exceed, by a slight amount, the negative conductance of the diode.

The above description is oversimplified but should suffice for the purposes of this report. Further study of the theoretical and quantum mechanical phenomena of tunneling may be obtained from the above references and publications by Hines [1960], Van der Ziel [1960], Tieman [1960], and Neilsen [1960].

b. The Reflection-Type Amplifier

Recently the Micro State Electronics Corporation has developed a tunnel diode amplifier employing a reflection technique to overcome the coupling factors between the diode and the source and load impedances.

The basic requirement of this type of amplifier is a suitable network to separate the incident wave and amplifier reflected wave. This may be accomplished by a circulator on hybrid network. Shown in figure 43 is a block diagram of a complete tunnel diode amplifier using a four-port circulator (two three-port circulators combined) which has been manufactured by the Micro State Electronics Corporation and is commercially available. Experimental models have indicated noise figures of 3 to 6 db in the 1000 to 6000 Mc/s range with a gain of 17 to 20 db and bandwidths on the order of 10% in the lower range and 5% in the upper range.

c. General Considerations

The tunnel diode amplifier offers certain advantages over other low noise devices in their application to scatter communication circuits.

Since the tunnel diode amplifier is a completely solid-state unit, it has exceptional reliability requiring only d-c activation for operation. In addition, because of its small size and light weight it can be readily mounted at the antenna terminals, thereby eliminating the excess noise contributed to the system by the transmission line or waveguide.

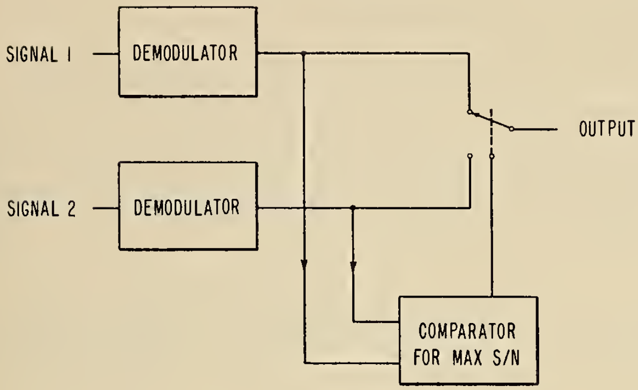
One of the main disadvantages of the tunnel diode amplifier is its stability versus environmental temperature change. Ambient temperature changes on the order of ± 20°C can cause major variations in the gain and noise figure of the device. Consequently, extreme care must be exercised in controlling its operating environment, and the unit must be temperature compensated in some way.

11. COMBINERS

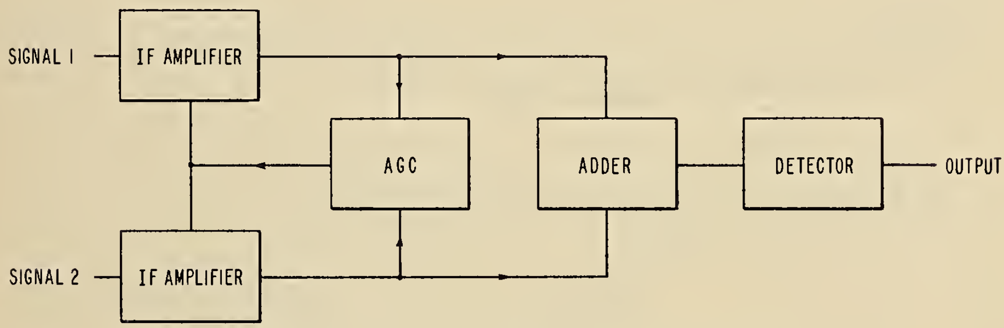
Diversity techniques are applied to tropospheric scatter systems to overcome the continual rapid fading encountered. Diversity techniques are based on the statistical character of scatter signals, and make use of the phenomenon that two or more signals arriving over different paths or at different frequencies do not fade in unison. The relative merits of space diversity (signals arriving over different paths) and frequency diversity (signals arriving at different frequencies) are considered in another part of this manual. This section will be confined to the consideration of the methods of combining diversity signals and the measurement of the performance of the combiners.

There are various ways in which diversity combining can be accomplished. Most notable are combining either immediately before or after detection, where the signals are either selected by switching or they are added together.

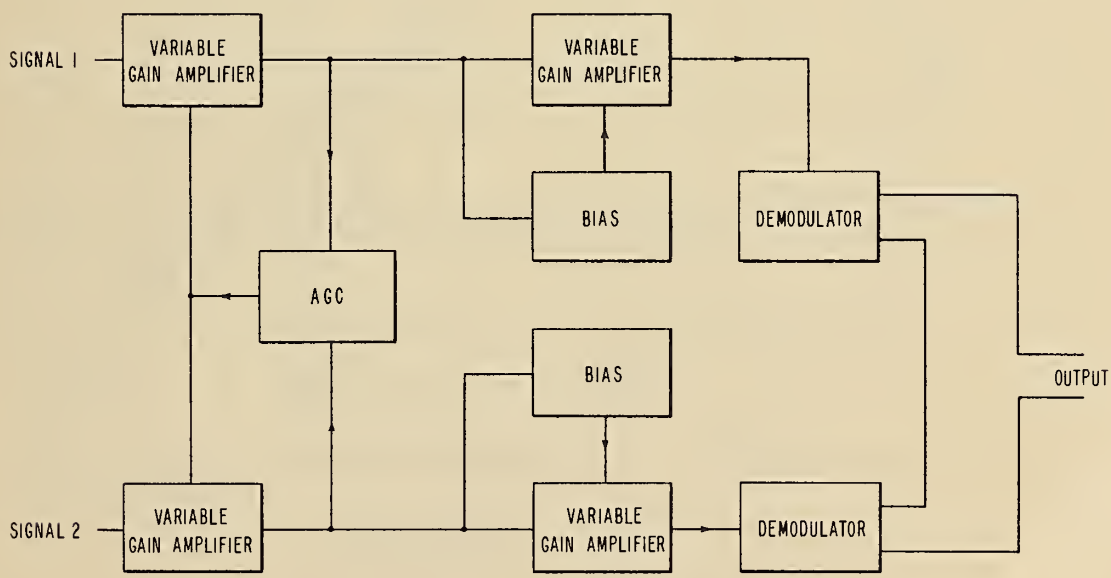
COMMON COMBINERS (SIMPLIFIED DIAGRAMS)



(a) SELECTION COMBINER (POST-DETECTION)



(b) EQUAL GAIN COMBINER (PRE-DETECTION)



(c) MAXIMAL RATIO COMBINER (PRE-DETECTION)

Figure 44

COMBINER PERFORMANCE CHARACTERISTICS

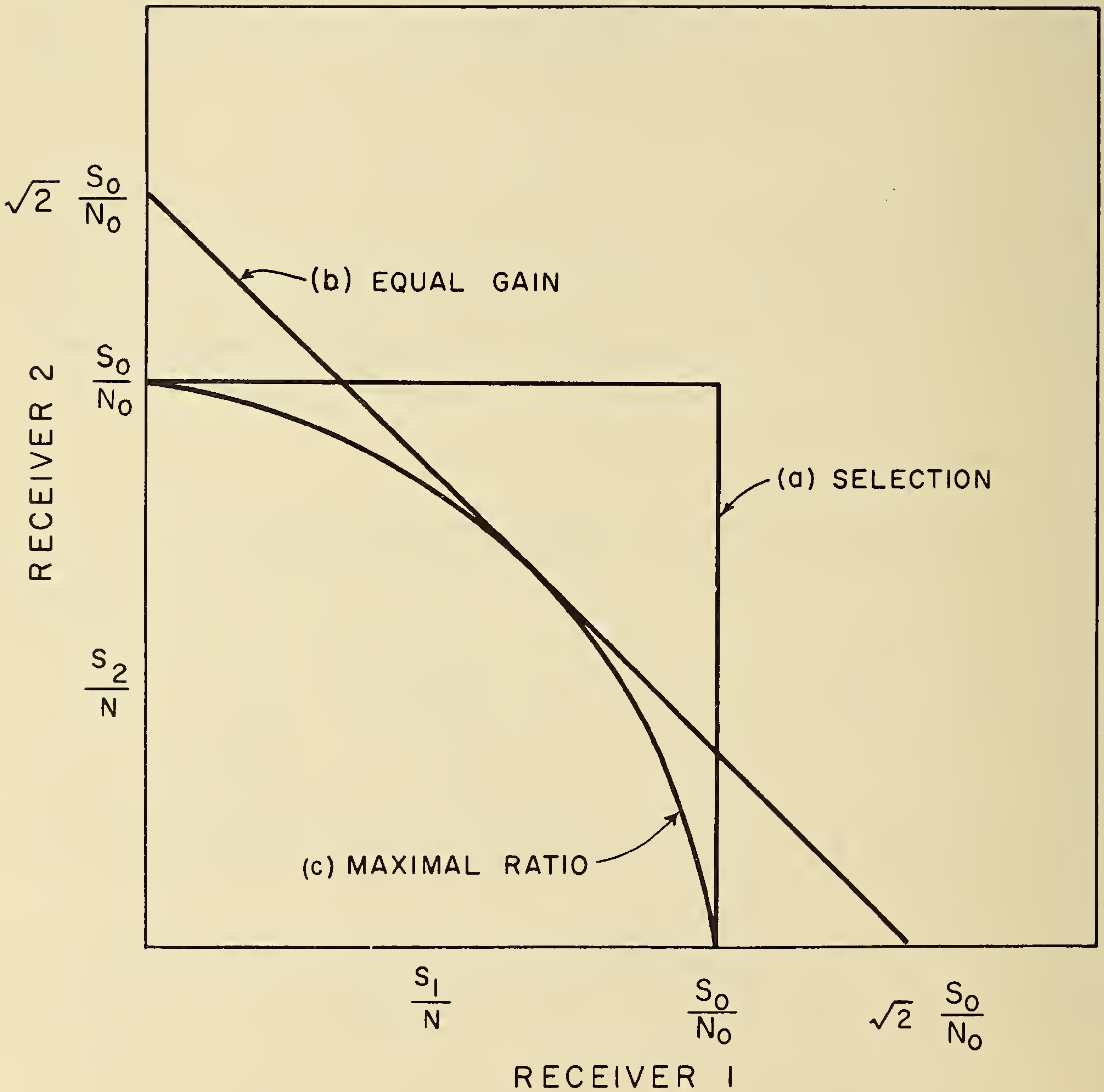


Figure 45

11.1. Methods of Combining

Classified according to the method of combining, there are three common types of combiners: the selection combiner, the equal combiner, and the maximal ratio combiner. Figure 44 shows a simplified block diagram of each of the three methods in a typical receiving system.

A selection combiner determines the larger input and connects this to the output. The equal gain or linear adder combiner sums the signals and applies the sum to the output. The maximal ratio or ratio squarer combiner squares the signals before addition and applies the sum of the squares to the output.

A comparison of the three methods of combining can be done graphically in a manner described by Altman and Sichak [1956] and illustrated in figure 45. For this comparison it must be assumed that

(1) Signals add linearly, noise adds in an rms fashion,

(2) All receivers have equal gain,

(3) All receivers have equal noise outputs, and the noise is random in character,

(4) The desired output signal-to-noise power ratio S_o/N_o is a constant.

a. The Selection Combiner

The selection combiner utilizes but one receiver at a time.

The output signal-to-noise ratio is equal to the input signal-to-noise ratio of the receiver in use at the time. This is illustrated by curve (a) in figure 45.

b. The Equal Gain Combiner

The equal gain combiner merely adds the different inputs. If two receivers are in use,

$$S_o/N_o = \frac{S_1 + S_2}{\sqrt{N_1^2 + N_2^2}} = \frac{1}{\sqrt{2}} \quad \frac{S_1 + S_2}{N} \tag{61}$$

This is illustrated by curve (b) in figure 45.

c. The Maximal Ratio Combiner

The maximal ratio or ratio squarer combiner utilizes a relative gain change between the output signals in use. For example, if the stronger signal has unity output and the weaker signal has an output proportional to G, then

$$S_o/N_o = \frac{S_1 + GS_2}{\sqrt{N_1^2 + G^2 N_2^2}} = \frac{S_1 + GS_2}{N\sqrt{1 + G^2}} \tag{62}$$

Maximizing the above expression by differentiating and equating to zero yields

$$G = S_2/S_1$$

i.e., this signal gain is adjusted to be proportional to the ratio of the input signals. Then

$$(S_o/N_o)_{max} = \sqrt{S_1^2 + S_2^2}\ /N \tag{63}$$

$$(S_o/N_o)^2_{max} = (S_1/N)^2 + (S_2/N)^2 \tag{64}$$

which is the equation of a circle. This is illustrated in figure 45 by the curve (c).

SIGNAL-TO-NOISE IMPROVEMENT IN A DIVERSITY SYSTEM

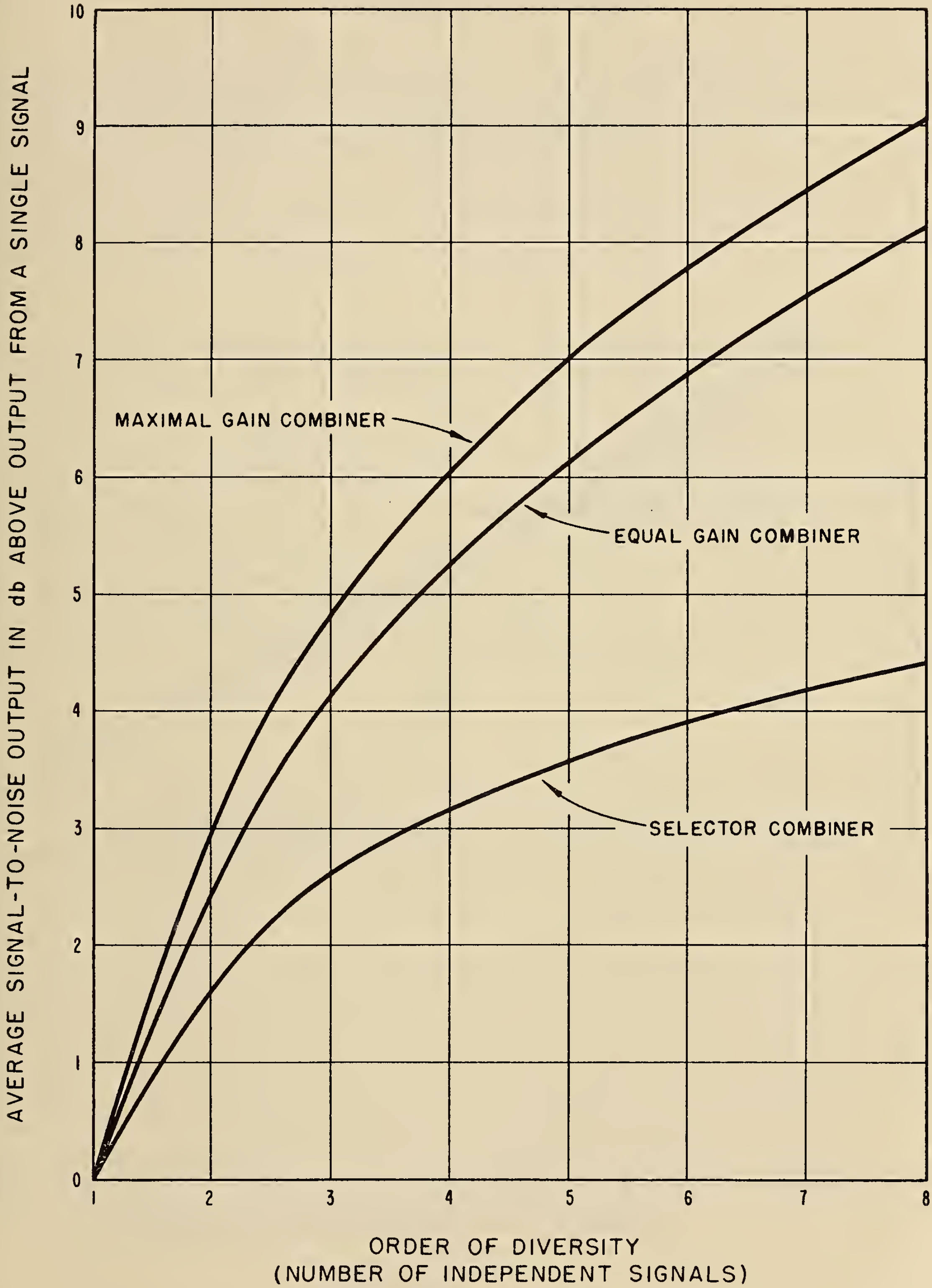


Figure 46

COMPARISON OF THE TIME DISTRIBUTION OF THE OF THREE TYPES OF COMBINERS USING DUAL DIVERSITY

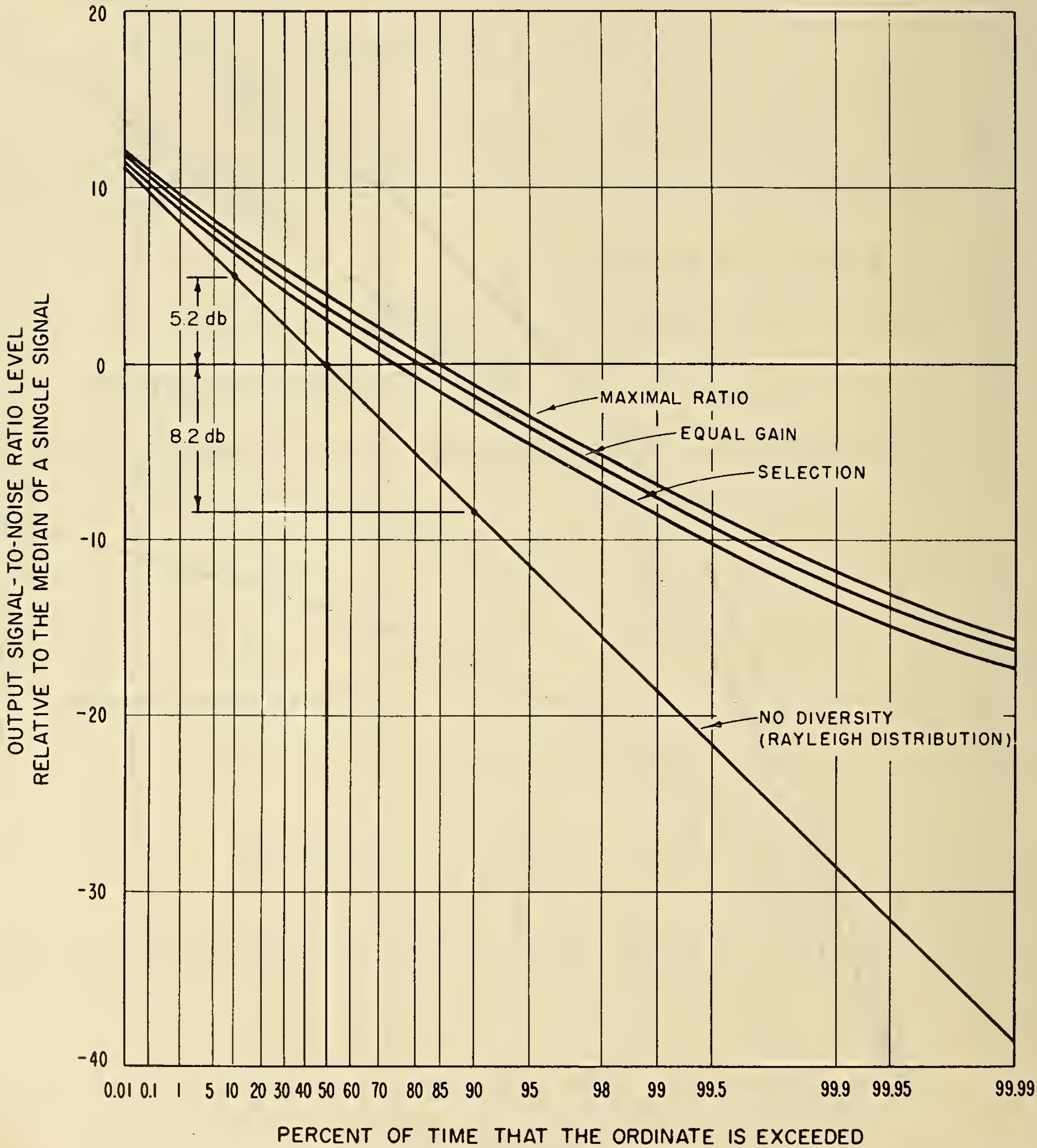


Figure 47

11.2. Comparison of Combiners

It can be seen that the maximal ratio combiner utilizes the best features of the two previous combiners. When one signal is zero, the combiner acts as a selector combiner; when both signals are equal, the combiner acts as an equal gain combiner. It can also be seen that the maximal ratio combiner yields the best output for any combination of signal-to-noise ratios since the curve of operation indicates that the optimum output occurs for lower values of input signal-to-noise ratios as evidence by its proximity to the origin.

Another method of comparison as described by Brennan [1959] is based on the difference in output to be expected from these three methods as the number of independent signals increase. If each of the signals are assumed to be Rayleigh distributed, then a statistical analysis of the output from each of these combiners results in the comparison illustrated in figure 46. It can be seen that as the number of independent signals increases, the superiority of the maximal ratio combiner increases. Where quadruple diversity if used (four independent signals) the average signal-to-noise ratio of the output using the selection combiner is only a little better than 3 db superior to no diversity, whereas the equal gain combiner results in a 5.25-db improvement with the maximal ratio combiner yielding a 6-db improvement.

If the operation of the combiners are compared, not on the basis of average values but on the basis of a time distribution of the output signal-to-noise ratio, the real significance of the combiner becomes evident. A time distribution is merely the percentage of an interval of time where the signal amplitude exceeds particular signal levels. As can be seen from figure 47, the Rayleigh distribution has a time distribution where the signal amplitude is approximately 5.2 db above the median value for 10% of the time intervals of the measurement, whereas, for 90% of that interval, the signal amplitude exceeded a level that is 8.2 db below the median. (This does not yield any information on how these time intervals are distributed over the measuring intervals. The distribution is actually a summation of time increments over the entire measuring intervals.) Where Rayleigh distributed signals are combined according to the principles utilized in the above three types of combiners, the output can be predicted as a function of the types of combiners, the output can be predicted as a function of the

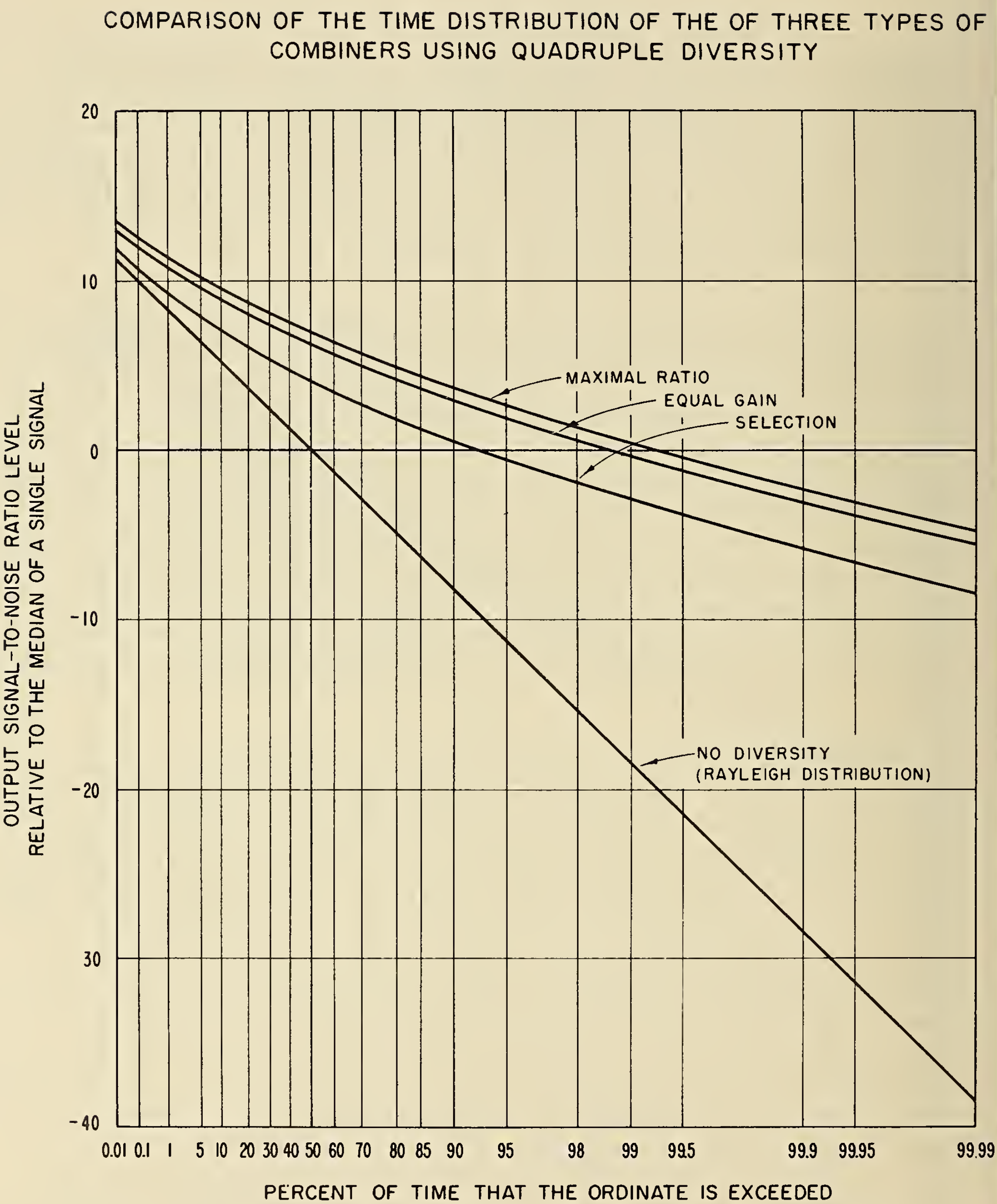

COMPARISON OF THE TIME DISTRIBUTION OF THE OF THREE TYPES OF COMBINERS USING QUADRUPLE DIVERSITY
OUTPUT SIGNAL-TO-NOISE RATIO LEVEL RELATIVE TO THE MEDIAN OF A SINGLE SIGNAL
20
10
0
-10
-20
30
-40
MAXIMAL RATIO
EQUAL GAIN
SELECTION
NO DIVERSITY (RAYLEIGH DISTRIBUTION)
0.01 0.1 1 5 10 20 30 40 50 60 70 80 85 90 95 98 99 99.5 99.9 99.95 99.99
PERCENT OF TIME THAT THE ORDINATE IS EXCEEDED

Figure 48

order of diversity and the type of combiner. This is illustrated in figure 47 for dual diversity and in figure 48 for quadruple diversity. It should be noted that although there is enhancement at the median (50%), the real enhancement occurs at the high percentage (low level) signal-to-noise ratios. The signal-to-noise ratio exceeded 99.9% of the time is improved by at least 23 db when quadruple diversity is employed.

Under conditions of Rayleigh fading there is but small improvement of the maximal ratio combiner over the equal gain combiner. The best improvement will not be greater than 1.05 db. If the fading lacks the depths characterizing the Rayleigh distribution, the equal-gain combiner approaches the maximal ratio combiner. (The maximal ratio combiner is independent of the type of distribution, whereas the equal-gain combiner is not.) But, if the fading is such that exceptionally deep fading is experienced, the operation of the equal-gain combiner, in the limit, can become inferior to the selection combiner. Hence the equal-gain combiner should not be used if exceptionally deep fading is expected.

This analysis of combining action is based on certain assumptions. It was assumed that independent signals were being combined, and it was assumed that the noise is completely random. The assumption of independence requires that the signals received on the different receivers are unrelated at least in an instantaneous sense; the variations in one will yield no information concerning the variations in any other. The relationship between such signals can be expressed in terms of a correlation coefficient, where a correlation coefficient of one indicates total dependence or a one-to-one relationship at each instant of time, and where a coefficient of zero indicates total independence. If the correlation coefficient is zero, the above applies; if the correlation coefficient is one, no diversity occurs. If a degree of correlation between the extremes occurs, diversity is possible but degraded. About half the diversity efficiency is experienced for correlation coefficients of 0.8, whereas practically full efficiency is experienced for coefficients less than 0.3. This concept of diversity efficiency is considered by Florman and Tary [1962]. Of importance here are the relative merits of the three types of combiners under these conditions. As the correlation increases the selection combiner approaches the nondiversity condition; the equal-gain combiner, although degraded will approach the maximal ratio combiner which maintains its original relationship to the selection combiner.

COMBINER OUTPUT FOR RAYLEIGH DISTRIBUTED INPUTS WHEN THE CORRELATION COEFFICIENT, ρ, IS EQUAL TO ONE

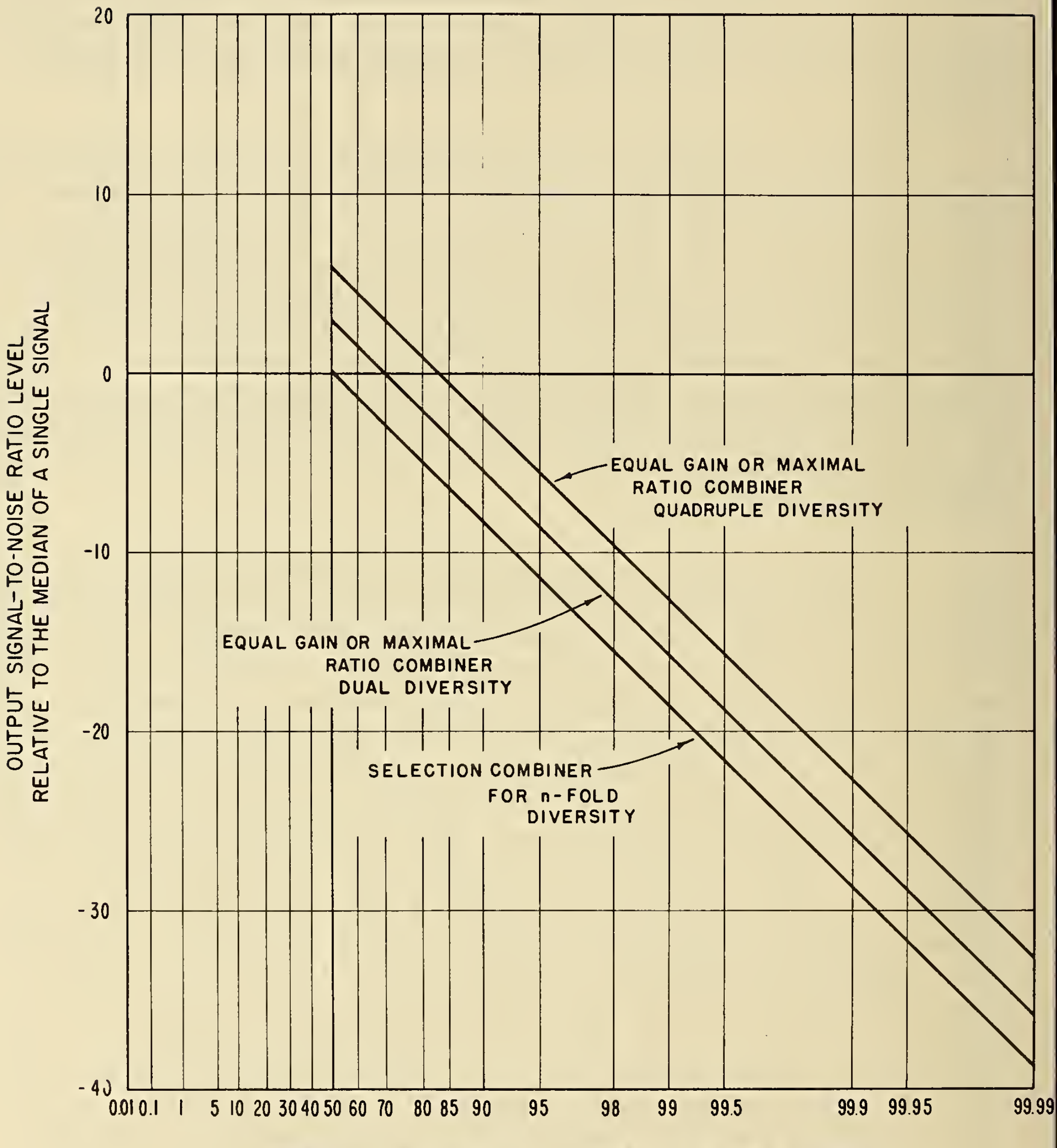


Figure 49

COMPARISON OF COMBINER OUTPUT WHEN INPUT SIGNAL HAS CORRELATION COEFFICIENTS, ρ, OF UNITY AND ZERO

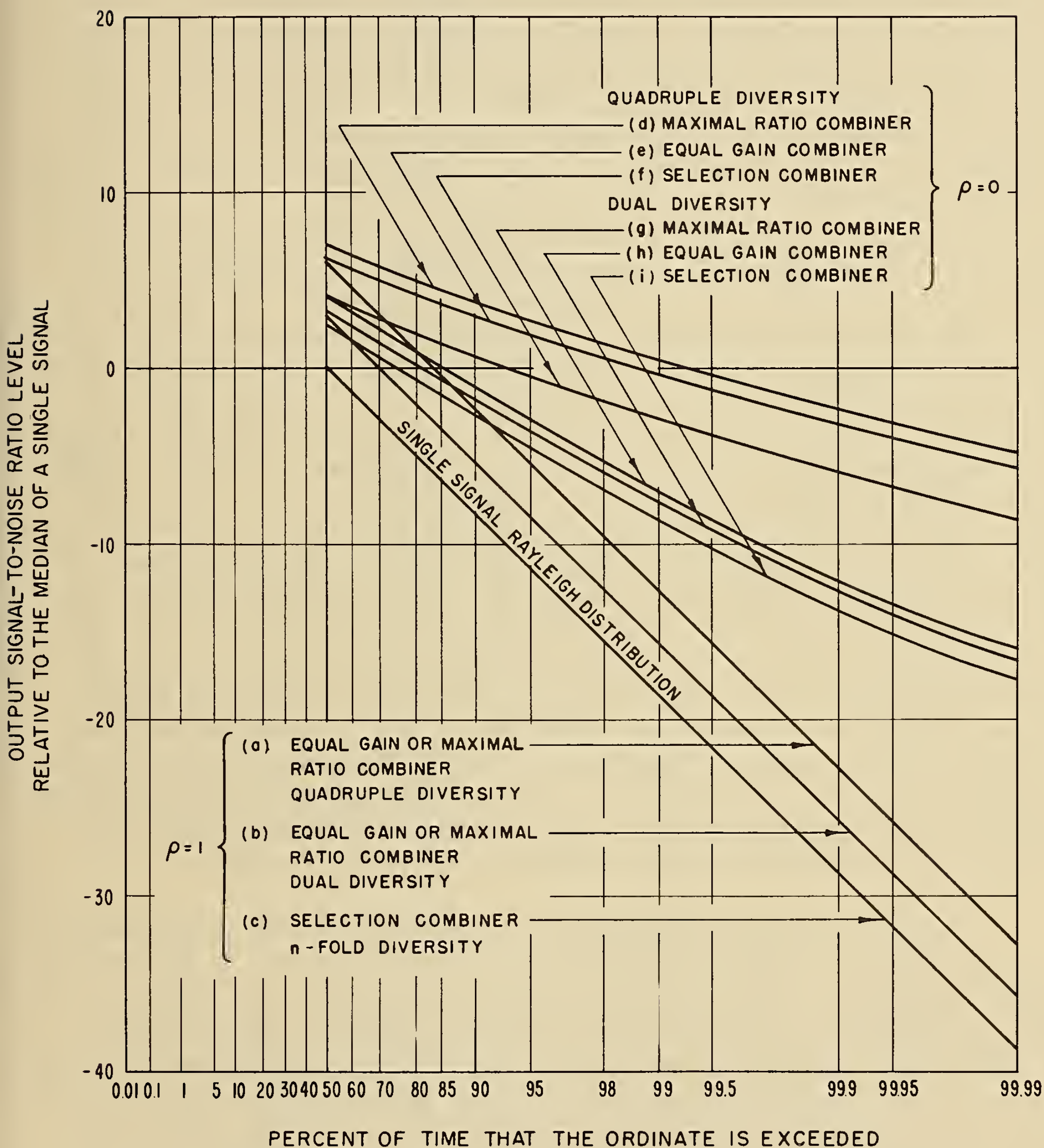


Figure 50

If the correlation coefficient is one, then the selection combiner should have no effect upon the character of the signal. Since the selection combiner utilizes only the strongest signal, the output will be the same as the one signal from that input that happened to be active at the starting time. The action of the equal-gain and maximal-ratio combiners will be different. With a correlation coefficient of one, the equal-gain combiner coincides with the maximal ratio combiner, and either combiner will yield an improvement over a signal equal to the average improvement (as illustrated for the maximal ratio combiner in figure 46) for all percentage of time greater than the median. Then for dual diversity the output will be Rayleigh distributed but displaced by 3 db above the input signal; for quadruple diversity the replacement will be 6 db (see figure 49). Hence, if the identical Rayleigh distributed signal is applied to the two inputs of a maximal ratio combiner used in dual diversity, the time distribution of the output will follow curve (b) in figure 50; whereas, if two independent Rayleigh distributed signals having equal means are applied to the same inputs, then the output will follow curve (g). Therefore, the improvement in the signal level exceeded 99% of the time that occurs when the correlation coefficient changes from unity to zero is approximately 8.7 db. It is also well to note that when the correlation coefficient is zero, the output of the maximal ratio combiner is approximately 11.7 db better than a single input at the 99% point. If the same comparison were made in a system where a selection combiner was used, the results would be a bit different. The output of the combiner would follow curve (c) in figure 50 when the correlation coefficient is unity, and would follow curve (i) when the correlation coefficient is zero. At the signal level exceeded 99% of the time, lack of correlation improves the selection combiner output by approximately 10 db. This concept of uncorrelated signals is an important factor to be considered when determining the separation distance between antennas in a diversity system.

Another assumption required in the consideration of combiner action is that concerned with the character of the noise. It is assumed that the noise is completely random and uncorrelated. Correlation between the noise occurring in the individual inputs to the combiner can exist especially when the noise includes effects from external interference. It can be shown that the presence of high correlation between the noise voltages has unfortunate results with respect to combining. Selection combining would be unaffected since but one signal is in use at any one time. However, with both equal gain and maximal ratio

PRE-DETECTION COMBINING

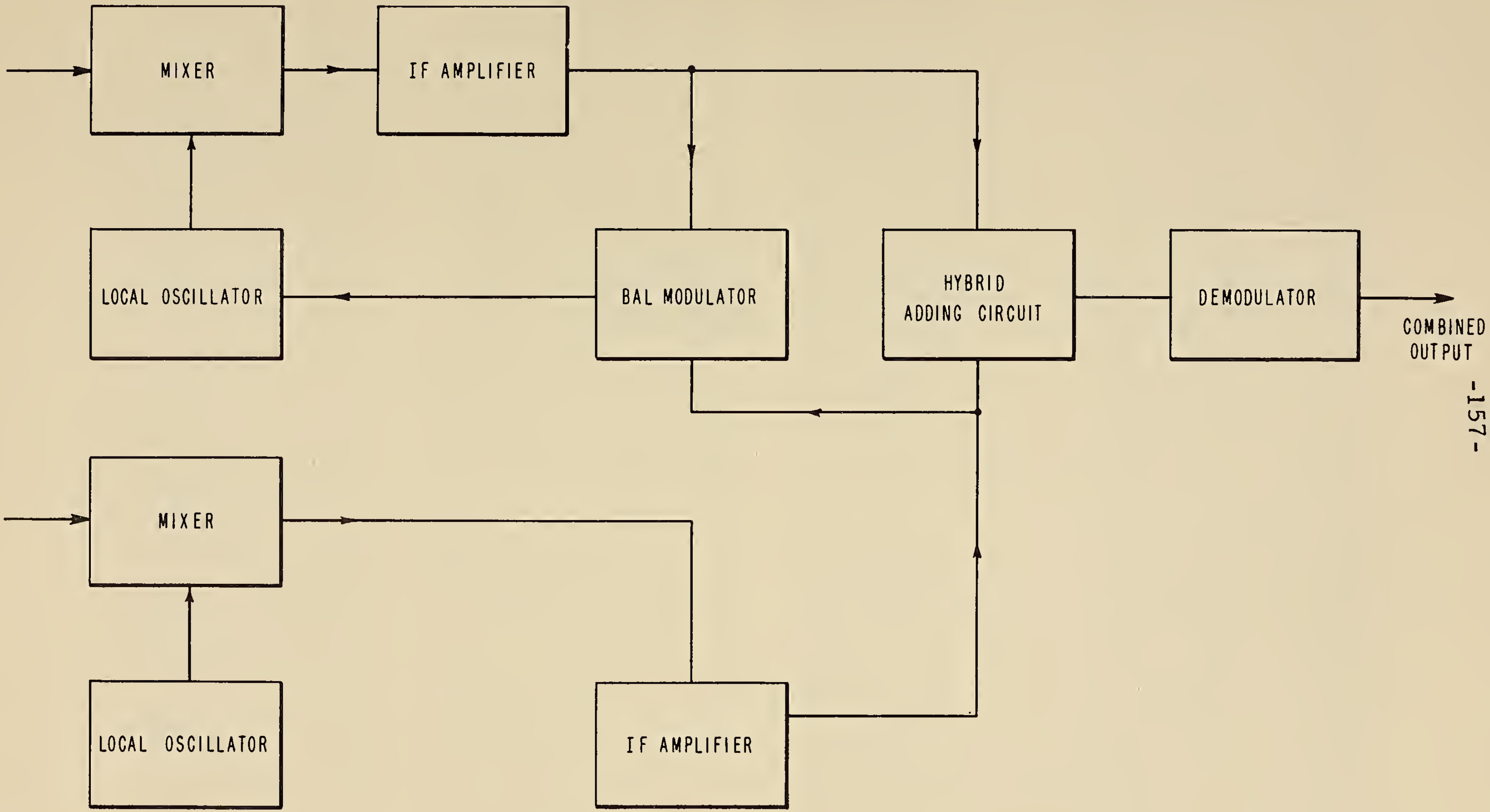


Figure 51

combiners, such is not the case. It is possible that the output signal-to-noise ratio using either of these two combiners under such circumstances will be less than that utilizing a single signal. In addition, the performance worsens as the order of diversity (number of independent signals) increases. Therefore, where highly correlated noise inputs are to be expected, the use of any combiner other than the selection combiner will probably degrade rather than enhance a single signal reception system. Since correlated noise is usually produced external to the system, condition such as these can be foreseen, at least to a degree, especially when the noise is man-made.

11.3. Pre-Detection and Post-Detection Combining

Classified according to where in the reception process the combining takes place, combiners commonly used are pre-detection IF combiners and post-detection baseband combiners. The former accomplishes combining before demodulation at the intermediate frequency; the latter accomplishes combining after detection at the baseband frequencies.

If pre-detection combining is considered, phase control circuitry is required to establish the coherency of the signals at the intermediate frequency. This control is unnecessary in the case of selection combining since only one signal at a time is used; however, it is indispensable in the case of the other two combiners (see figure 51). At the present time pre-detection combining is not used extensively because of the difficulty of establishing this phase control. Although this is difficult to achieve in practice, the principles are straightforward. Figure 51 illustrates one method of achieving the phase control as described by Adams and Mindes [1958].

Separate receiver front ends are used up to the IF amplifier output. The relative signal phase is measured at this point by a balanced modulator (phase detector), and the phase is controlled by supplying a corrective d-c voltage to one local oscillator, causing it to gain or lose phase with respect to the other local oscillator as required to establish additive signal phase at the combining point. The resulting amplitudes of the co-phased signals are added in a hybrid, and the combined signal is applied to the demodulator. In

POST-DETECTION FM COMBINER

(BASEBAND COMBINER)

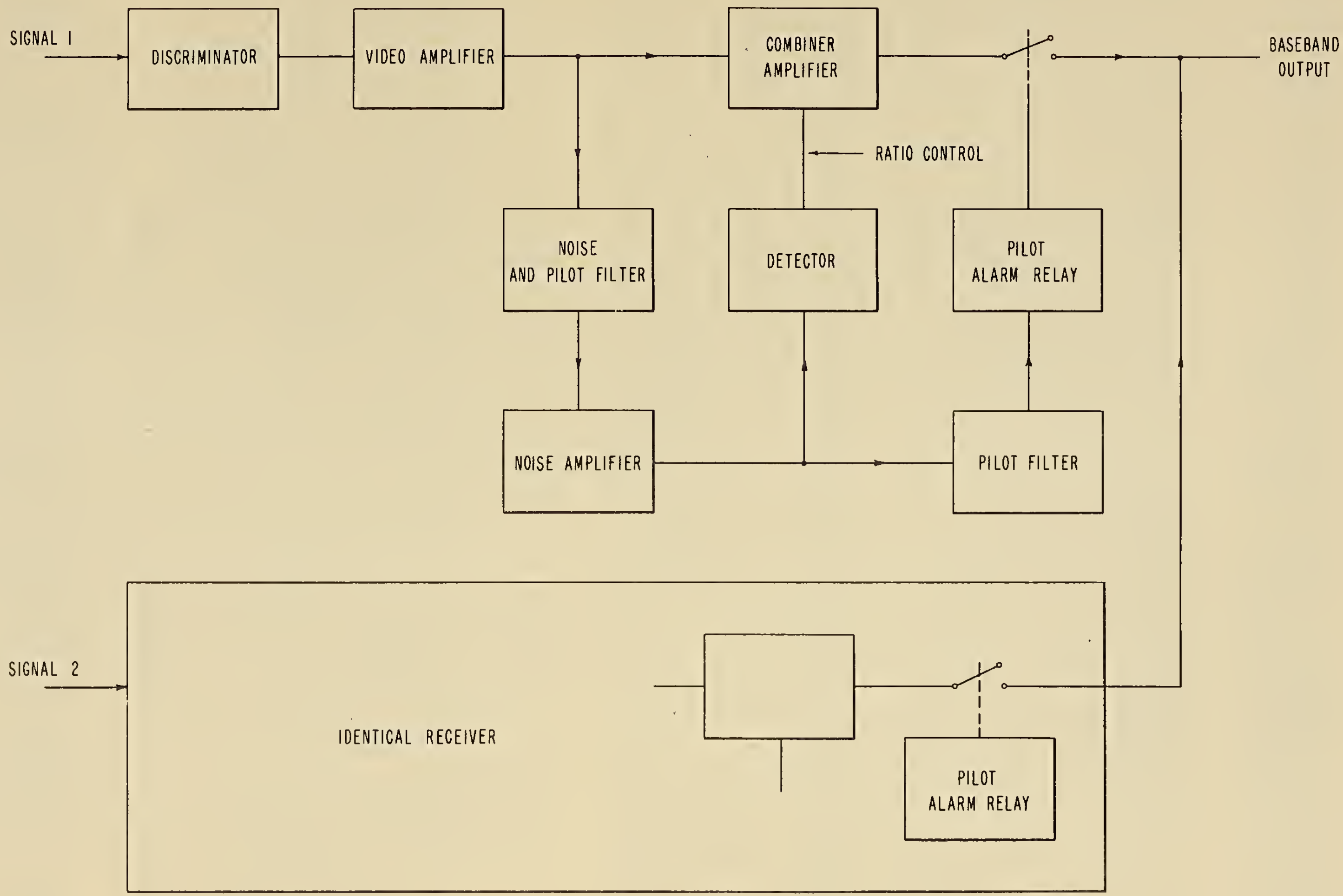


Figure 52

this system failure at any point in either receiver does not interfere with normal operation of the remaining receiver; the phase control circuit becomes inoperative at this time.

Where AM or SSB are concerned, the operation of combiners is the same whether pre-detection or post-detection is used. There is some advantage to pre-detection combining where selection diversity is used in that the effect of the transients due to switching is reduced. However, where FM is in use, some doubt exists as to the advisability of post-detection. D. G. Brennan [1959] shows that the statistical consideration of the combining of FM signals favors the pre-detection method over the post-detection method. The argument favoring the pre-detection combiner is based on the fact that the input and not the output of an FM demodulator is Rayleigh distributed. Consequently, the combining of several of these signals will not yield the results as outlined above, and the comparison between the relative merits of the various combiners will not apply. In post-detection, FM combining, the equal gain combiner is not used, and the maximal ratio combiner will yield performance little different from selection combining. (It would appear that pre-detection combiners in FM service could provide additional performance with respect to distortion, FM quieting, less circuit complexity, and better reliability near the threshold level.)

11.4. General Considerations

Most present-day systems employ post-detection or baseband combining, utilizing either maximal ratio or equal-gain combiners. This equal gain of linear adder combiner is impractical as an FM post-detection combiner. This is because in deep fades the noise output of the conventional FM receiver rises sharply to a level comparable to full signal output. Therefore, if weak FM signals are to be combined after detection, a close approximation to ratio squared combining must be provided over a wide range of signal ratios if sufficient rejection of the greatly increased noise is to be provided. Hence, maximal ratio combiners are usually used in FM post-detection.

Figure 52 is a simplified block diagram of a typical FM receiver using the baseband combiner [Adams and Mindes, 1958]. Referring to the circuit diagram, each receiver selects the noise components above the normal signal frequency, and these are amplified and detected to produce a control voltage that is applied to the corresponding combiner tube to control the combining ratio. Differences in control voltage will vary the combining ratio favoring the signal having the lower noise.

For protection against failure of either signal channel or a failure in the noise amplifier, a pilot tone is added at the transmitter point. This pilot tone is received with the signal passed through the noise amplifiers for protection against failure and finally used to operate an alarm relay that disconnects a receiver when the pilot tone is absent, thus permitting the remaining receiver to function normally. In most cases, dual pilot tones and an alarm circuit at the transmitter are used to prevent accidental loss of the receivers due to failure of the pilot tone.

Additional work in this area has been performed by Hausman [1954], Kahn [1954], Kelley [1957-1958], Long and Weeks [1957], and Mack [1955].

12. POWER SUPPLIES

12.1. Unregulated Power Supplies

No circuitry is used in the unregulated supply to maintain a constant output voltage with varying input line or variation in load. As input line voltage varies, the output voltage will vary the same percentage and in the same direction. Also, as load is applied, the internal voltage drop increases, and the output voltage will drop accordingly. A load regulation of 10 to 25% is typical. The main advantages of this type of supply are its simplicity, low cost, and high reliability. However, in all but the most basic systems, variation in power supply output may cause poor or erratic system performance.

12.2. Resonant Regulator Power Supplies

Constant voltage can be readily maintained despite input line voltage variations by incorporating magnetic voltage regulating circuitry ahead of the rectifiers and filters in the unregulated d-c supply. Since no tubes or transistors are used, the unit is extremely rugged and reliable. Load regulation is not good, although some load compensation is accomplished by the magnetic circuitry so that it is better than an equivalent unregulated supply. Load regulation for these types runs between 5 and 10% .

Momentary or continuous output short circuits cannot damage this type of supply, since the current is limited to approximately 125% to 150% of maximum rating by the resonant circuitry. Since the variable auto transformers used in these supplies increase internal impedance, this tends to offset the advantage of low-cost regulation. This type is suitable in systems requiring fixed output voltages.

Excellent line voltage regulation and transient response for line and load pulse makes these units suitable for supplying power for transistorized circuitry. Direct-current output voltage typically remains within the regulation limits for line voltage surges up to 10%. An output voltage change of 1.5% for each 1% change in input frequency will occur due to the frequency sensitive nature of this type of unit.

12.3. Magnetic Amplifier and Saturable-Core-Reactor Power Supplies

Precise regulation of output voltage for both line and load variations are obtained from these supplies, typically holding ± 0.21% regulation. A closed-loop circuit is incorporated in both types. They are insensitive to frequency variation, although slight degradation in performance may be expected when they are continuously operated at some frequency other than the nominal design frequency. When operated at lower power-line frequencies, there will be an increase in ripple. No derating of output of voltage or current is usually necessary for the electronic type, but the tubeless type may have to be derated for continuous performance. Although line transients are somewhat attenuated, maximum excursions of 5% will be experienced for load transients of 20%.

A smaller, lighter, and more reliable supply than the electronic type is the tubeless magnetic amplifier power circuit and transistor power reference type. This type has the advantage of the elimination of vacuum tubes, and hence a corresponding associated reduction in maintenance. Other advantages are a wider range of output adjustments and somewhat better dynamic characteristic. The response time of this type is faster than the electronic type but not fast enough to prevent line transients from coming through. Maximum excursions of 4% of output voltage may be experienced for load transients of 20%.

12.4. Transistorized d-c Power Supplies

Best performance such as regulation, response time, output impedance, and ripple are obtained from the fully transistorized supplies. For low and medium power units, these are substantially smaller in size and lower in weight. Some of the other features of a transistorized supply are:

(1) Fast response time and reduced ripple

(2) Low output impedance

(3) Regulation of ± .25% to ± 0.05%

(4) Ripple 0.5 millivolts to 5 millivolts rms

(5) Unaffected in input frequency variations (can be applicable for nominal 50, 60, or 80 c/s input)

Transistorized supplies have excellent dynamic performance, and "transient free" output voltage can be obtained.

12.5. Series-Tube Regulated Power Supplies

The same excellent performance characteristics of transistorized supplies can be obtained from series-tube regulated power supplies with voltage ranges from 0 to 1000 volts. Vacuum tubes are able to withstand higher voltages, and are therefore better utilized in the 75- to 1000-volt supplies. The maximum current from these supplies is ordinarily of the order of 2 amperes.

13. REPEATERS

Repeaters are used to transmit energy over relatively short paths. In some cases the path is such that the scatter mode or diffraction mode is used; however, in general, repeaters are considered to be a series of line-of-sight paths. It is this latter category that is of interest here.

The simplest combination is that where the transmission path can be divided into two free-space paths with the repeater joining the two segments. In such a case the basic transmission loss of the repeater, within line of sight of both path terminals, is given as the sum of two free-space basic transmission losses less the repeater gain. The basic transmission loss is given by Norton [1953]. That is, the transmission loss over the entire path is equal to the sum of the transmission loss on each path individually decreased by the gain contributed by the repeater. The expression for the free-space transmission loss over a path with a single repeater can be expressed as follows:

$$L_b = 64.89 + 20 \log_{10} d_1 + 20 \log_{10} d_2 + 20 \log_{10} f_{Mc1} + 20 \log_{10} f_{Mc2} - G_1 - G_2 - K \tag{65}$$

where d_1 and d_2 = the distances in kilometers from the repeater site to each of the terminals of the path. (If d_1 and d_2 are in miles, then the constant 64.89 is replaced by 73.16.)

f_{Mc1} and f_{Mc2} = the frequency in megacycles per second of the signals employed on the two legs of the path

G_1 and G_2 = the gains of the two repeater antenna in db relative to an isotropic antenna

K = the repeater gain in db

N.B. This expression is independent of the gains of the transmitting and receiving antennas. In the practical case, the gains of these antennas must be subtracted from the transmission loss.

With active repeaters, K is normally positive; with passive repeaters, K is normally negative or zero.

If more than one repeater is used, the above expression (65) is changed by adding the path loss of each additional leg. The above expression will have the following added to it for each additional leg:

$$32.45 + 20 \log_{10} d_3 + 20 \log_{10} f_{Mc3} - G_3 - G_4 - K_2 \qquad (66)$$

13.1. Active Repeaters

The above expression is valid for both active and passive repeater operation. It is especially useful in this form when active repeaters are considered. To reduce interference effects on multiple repeater paths and to reduce coupling between the transmitting and receiving antennas at the repeater, the carrier may undergo a change in frequency at each active repeater. The modulation information arrives from the transmitting location on one carrier frequency, is recovered, and superimposed upon a carrier of another frequency which is radiated either to the receiving location or to the next repeater location.

13.2. Passive Repeaters

Passive repeaters as described by Greenquist [1954] and Jakes [1953] do not amplify the signal transmitted over the paths, but merely reflect, scatter, or reradiate a portion of the original signal to the desired receiving location. This can be done either by receiving the signal on one antenna and reradiating it from another, or the signal can be reflected or scattered one or more times by means of large reflectors. The path loss formula (65) can be simplified slightly since the frequency is a constant, and since in most cases identical antennas are used at the repeater.

a. Back-to-Back Directional Antennas

When a back-to-back antenna system is used as the repeater, power is received from the transmitting location on one antenna and is fed through a short piece of transmission line to another antenna which radiates the energy toward the receiver. Parabolic antennas at the repeater are individually pointed to the transmitter and receiver, respectively. Each path then (the path from the transmitter to the repeater and the path from the repeater to the receiver) can be considered as independent line-of-sight paths.

Since the path can be dismembered into two or more segments, the equipment requirements can be considered for each segment individually. The above general statement implies that the actual gains of the antennas are known. In a previous section it was pointed out that the effective gain of the antenna is conditioned by such parameters as the minimum distance between antennas and ground reflections. Each segment of the path should be analyzed to determine whether or not these two parameters are affecting the apparent gains realized over the path.

The power requirements for the path are calculated for the effective antenna heights and the gain of the antennas. Mechanical alignment of the antennas or reflectors is necessary to utilize the path to best advantage. This is accomplished by transmitting a signal from the transmitter and adjusting the alignment in azimuth and elevation until a maximum signal is obtained. It may be necessary to adjust the heights of the antennas relative to the ground to achieve the optimum field resulting from the phase addition of the direct and ground-reflected waves. In the case of the active repeater, this maximizing of the amplitude of the signal can be accomplished at the repeater receiver, usually at the intermediate frequency. If the repeater is a passive repeater and the signal is fed directly from one antenna to the other through a section of transmission line, then the signal at the repeater location may be monitored on a field-intensity meter. (Note: It is assumed that radio communications are available between the transmitter location and the repeater location.) After the transmitter-to-repeater leg is adjusted, the repeater receiving antenna can be connected to the transmitting antenna, and the same procedure can be followed on the repeater-to-receiver leg where the signal is monitored at the receiver location.

b. Single Flat Reflectors

Where a single flat reflector is used as a passive reflector, the aforementioned expression can be used where K is assumed to be 0, and G is the gain of the reflector. For a passive flat reflector, the gain is proportional to the effective area of the reflector

$$G = 10 \log_{10} A + 20 \log_{10} f_{Mc} - 38.54 \tag{67}$$

where A is the effective area in square meters of the reflecting sheet. The effective area of the antenna is the projected area normal to the transmission path in the direction of either the transmitter or the receiver. Therefore, if the angle at which the signal from the transmitting antenna strikes the passive reflector is too great, the effective area of the reflector becomes quite small. It can be seen that for a high-gain flat reflector, considerable area is necessary. This type of repeater has the advantage of eliminating the line loss which is found in the back-to-back repeater. The problem of contour tolerance is more easily solved in the case of flat reflector in contrast to the large parabolic antennas. A much better surface efficiency is possible. (The aperture efficiency of the parabolic antenna is taken to be 54%.)

c. Multiple Flat Reflectors

To reduce the incident angle that the energy from the transmitter experiences when directed upon a passive flat reflector, two reflectors can be used, and this incident angle can be maintained to less than 45° If the projected areas along the transmission paths are assumed to be square, and if the geometric mean of the areas of the two reflectors is large with respect to $\sqrt{2d_3}$, where d_3 is the separation between the two reflectors, then (65) applies in a three-path determination.

$$\begin{aligned} L_b = {} & 97.34 + 20 \log_{10} d_1 + 20 \log_{10} d_2 + 20 \log_{10} d_3 \\ & + 60 \log_{10} f_{mc} - 2G_1 - 2G_2 \end{aligned} \tag{68}$$

where G_1 and G_2 are determined from (67).

The disadvantage of using flat reflectors as passive repeaters is in their size and the difficulty in adjusting them. Practical 7-Gc/s systems using 10-foot parabolic antennas for transmitting and receiving, and using antennas having acceptable gains at the repeater would require reflectors in the neighborhood of 55 square meters. If it is desired to maximize the gain on a path such as this, the difficulty of moving the repeaters as suggested in the section on using back-to-back parabolic antennas can be easily imagined. To maximize a system using multiple flat passive reflectors would require that signal be transmitted from the transmitting location and be received at the receiving location, and by radio communications between transmitter, repeater, and receiving sites, the position of these large reflectors could be adjusted until the signal is a maximum at the receiving location. Nevertheless, the practicality of a system using multiple flat reflectors has been demonstrated in practice as described by Cappuccini and Gasparini [1958] and Young [1957].

It is evident from the equations that the maximum transmission loss occurs when the repeater is located midway between the transmitter and the receiver, and minimum transmission loss occurs when the repeater is located close to one end of the path. This arises from the fact that the transmission loss is a function of the product of d_1 and d_2.

ACKNOWLEDGMENTS

The authors are indebted to many members of the staff of Central Radio Propagation Laboratory for the advice and assistance. In particular, recognition is due J. W. Herbstreit for his continuous help and advice during the preparation of the manuscript.

Acknowledgment is due J. C. Harman for the drafting of the figures, C. M. Benedict for the photography, and Miss S. J. Bush who typed the entire manuscript.

REFERENCES

Adams, R. T., and B. M. Mindes, Evaluation of IF and baseband diversity combining receivers, IRE Trans. on Commun. Systems CS-42, No. 1, p. 8 (June 1958).

Adler, R., G. Hrbek, and G. Wade, A low-noise electron-beam parametric amplifier, Proc. IRE 46, p. 1756 (Oct. 1958).

Adler, R., G. Hrbek, and G. Wade, The quadrupole amplifier, a low-noise parametric device, Proc. IRE 47, p. 1713 (Oct. 1959).

Altman, F. J., and W. Sichak, Simplified diversity communication system for beyond-the-horizon links, Electrical Communication, Technical Journal of ITT 33, P. 151 (June 1956).

Arguimbau, L. B., and R. D. Stuart, Frequency modulation (John Wiley and Sons, New York, N. Y., 1956).

Bailey, F. M., and H. P. Thomas, Phasitron F-M transmitter, Electronics 19, 108-112 (Oct. 1946).

Barsis, A. P., K. A. Norton, P. L. Rice and P. H. Elder, Performance predictions for single tropospheric communications links and for several links in tandem, NBS Technical Note No. 102, (Aug. 1961).

Black, H. S., Modulation theory (D. Van Nostrand Co., Inc., New York, N. Y., 1953).

Brennan, D. G., Linear diversity combining techniques, Proc. IRE 46, p. 1075 (June, 1959).

Brodzinsky, A., and A. C. MacPherson, Noise measurement of negative resistance amplifier, IRE Trans. on Instrumentation I-8, p. 44 (Sept. 1959).

Bruene, W. B., Linear power amplifier design, Proc. IRE 44, 1754-1759 (Dec. 1956).

Camies, B. S., Principles of frequency modulation (John Rider and Sons, New York, N. Y., 1959).

Cappuccini, F., and F. Gasparini, Passive repaters using double flat reflectors, Proc. IRE 46, p. 784 (Apr. 1958).

Cline, J. F., and B. M. Schiffman, Tunable passive multicouplers employing minimum loss filters, IRE Trans. on Microwave Theory Tech. MTT-7, No. 1, 121-127 (Jan. 1959).

Cohn, S. B., and F. S. Coale, Directional channel separation filters, Proc. IRE 44, No. 8, 1018-1024 (Aug. 1956).

Cottony, H. V., and A. C. Wilson, Technique of accurate measurement of antenna gain, IRE Trans. on Ant. and Prop. AP-6, No. 4, p. 366 (Oct. 1958).

Day, J. R., Serrasoid F-M modulator, Electronics 21, 72-76 (Oct. 1948).

Esaki, L., New phenomenon in narrow G_e p-n junctions, Physical Review 109, p. 603 (Jan. 1958).

Fibbs, C. E., and G. G. Johnstone, Frequency modulating engineering, (John Wiley and Sons, New York, N. Y., 1956).

Florman, E. F., and J. J. Tary, Required signal-to-noise ratios, RF signal power and bandwidth to achieve a given performance of multichannel radio communication systems, NBS Technical Note 100 (Jan. 1962).

Friis, H. T., Noise figures of radio receivers, Proc. IRE 32, p. 419 (July 1944).

Greenquist, R. E., An analysis of passive reflector antenna systems, Proc. IRE 42, p. 1173 (July 1954).

Hall, R. N., Tunnel diodes, IRE Trans. on Electron Devices ED-7, No. 1, p. 1 (Jan. 1960).

Harris, F. S., The parametric amplifier, CQ 14, No. 11, p. 74 (Nov. 1958).

Haus, H. A., and R. B. Adler, An extension of the noise figure definition, Proc. IRE 45, p. 1517 (1957).

Hausman, A. H., An analysis of dual diversity receiving systems, Proc. IRE 42, p. 944 (June 1954).

Herrmann, G. F., Low noise up-conversion from 500 mc to 10 kmc using variable capacitance diodes, presented at the 16th Annual Conference on Electron Tube Research, Quebec, Canada (June 1958).

Hines, M. E., High frequency negative resistance circuit principles for Esaki diode applications, Bell System Technical Journal 34, p. 477 (May 1960).

Hund, A., Frequency modulation, 1st Ed. (McGraw-Hill, New York, N.Y., 1942).

Icenbice, P. J., and H. E. Fellhauer, Linearity testing techniques for single sideband equipment, Proc. IRE 44, 1775-1782 (Dec. 1956).

International Telephone and Telegraph Corp., Federal Telephone and Radio Handbook (Nov. 1960).

Jakes, W. C., A theoretical study of an antenna-reflector problem, Proc. IRE 41, p. 273 (Feb. 1953).

Kahn, L. R., Ratio squarer, Proc. IRE 42, p. 1704 (Nov. 1954).

Kelley, G. J., Choosing the optimum type modulation, a comparison of several communications sytems, IRE Trans. on Commun. Systems CS-6, No. 1, p. 14 (June 1958).

King, R. W., H. R. Mimno, and A. H. Wing, Transmission lines, antennas and waveguide (McGraw-Hill, New York, N. Y., 1945).

Kraus, J. D., Antennas (McGraw-Hill, New York, N.Y., 1950).

Lewis, W. D., and L. C. Tillotson, A non-reflecting branching filter for microwaves, Bell System Technical Journal 27, No. 1, 83-95 (Jan. 1948).

Long, W. G., and R. R. Weeks, Quadrupole diversity tropospheric scatter systems, IRE Trans. on Commun. Systems CS-5, No. 3, p. 8 (Dec. 1957).

Manley, J. M., and H. E. Rowe, Some general properties of non-linear elements, Proc. IRE 44, p. 904 (July 1956).

Neilsen, E. G., Noise performance of tunnel diode amplifiers, Proc. IRE 48, p. 1903 (Nov. 1960).

Norton, K. A., Transmission loss in radio propagation, Proc. IRE 41, 146-152 (1953).

Oliver, B. M., Noise figure and its measurement, Hewlett-Packard Journal 9, No. 5, p. 3 (Jan. 1958).

Reed, E. D., The variable capacitance parametric amplifier, IRE Trans. on Electron Devices ED-6, No. 1, p. 216 (Apr. 1959).

Rice, P. L., A. G. Longley, and K. A. Norton, Tropospheric wave transmission loss prediction, To be published as NBS Technical Note 101 (1962).

Russell, C. R., The parametric amplifier, British Communications and Electronics 7, No. 2, 94-98 (Feb. 1960).

Schelkunoff, S. A., and H. T. Friis, Antennas: Theory and practice (John Wiley and Sons, New York, N.Y., 1952).

Silver, S., Microwave antenna theory and design, Radiation Laboratory Series 12, (McGraw-Hill, New York, N. Y., 1949).

Skilling, H. H., Electric transmission line (McGraw-Hill, New York, N. Y., 1951).

Smith, P. H., An improved transmission line calculator, Electronics 17, p. 130 (Jan. 1944).

Smullin, L. D., and C. G. Montgomery, Microwave duplexers, Radiation Laboratories Series 14 (McGraw-Hill, New York, N. Y. 1949).

Summers, H. S., Tunnel diodes as high-frequency devices, Proc. IRE 46, p. 1201 (July 1959).

Terman, F. E., Radio engineering (McGraw-Hill, New York, N. Y., 1947).

Tieman, J. J., Shot noise in tunnel diode amplifiers, Proc. IRE 48, p. 1418 (Aug. 1960).

Tyminski, W. Y., and A. E. Hylas, A. wide band hybrid ring for UHF, Proc. IRE 41, p. 81 (Jan. 1953).

Uhlir, A., The potential of semiconductor diodes in high frequency communications, Proc. IRE 46, No. 6, p. 1099 (June 1958).

Van der Ziel, A., Noise measure of lossy tunnel diode amplifiers, Proc. IRE 48, p. 1321 (July 1960).

Wind, H., and H. Rapaport, Handbook of microwave measurement, 2nd Ed. (Polytechnic Press Interscience Publishers, New York, N. Y., 1955).

Young, R. F. H., Passive repeater using double flat reflectors, IRE National Convention Record 5, Part 1, p. 36 (Mar. 1957).

☆ U.S. GOVERNMENT PRINTING OFFICE : 1963 O—670950

U. S. DEPARTMENT OF COMMERCE
Luther H. Hodges, *Secretary*

NATIONAL BUREAU OF STANDARDS
A. V. Astin, *Director*

THE NATIONAL BUREAU OF STANDARDS

The scope of activities of the National Bureau of Standards at its major laboratories in Washington, D.C., and Boulder, Colorado, is suggested in the following listing of the divisions and sections engaged in technical work. In general, each section carries out specialized research, development, and engineering in the field indicated by its title. A brief description of the activities, and of the resultant publications, appears on the inside of the front cover.

WASHINGTON, D. C.

Electricity. Resistance and Reactance. Electrochemistry. Electrical Instruments. Magnetic Measurements. Dielectrics. High Voltage.

Metrology. Photometry and Colorimetry. Refractometry. Photographic Research. Length. Engineering Metrology. Mass and Scale. Volumetry and Densimetry.

Heat. Temperature Physics. Heat Measurements. Cryogenic Physics. Equation of State. Statistical Physics.

Radiation Physics. X-ray. Radioactivity. Radiation Theory. High Energy Radiation. Radiological Equipment. Nucleonic Instrumentation. Neutron Physics.

Analytical and Inorganic Chemistry. Pure Substances. Spectrochemistry. Solution Chemistry. Standard Reference Materials. Applied Analytical Research. Crystal Chemistry.

Mechanics. Sound. Pressure and Vacuum. Fluid Mechanics. Engineering Mechanics. Rheology. Combustion Controls.

Polymers. Macromolecules: Synthesis and Structure. Polymer Chemistry. Polymer Physics. Polymer Characterization. Polymer Evaluation and Testing. Applied Polymer Standards and Research. Dental Research.

Metallurgy. Engineering Metallurgy. Microscopy and Diffraction. Metal Reactions. Metal Physics. Electrolysis and Metal Deposition.

Inorganic Solids. Engineering Ceramics. Glass. Solid State Chemistry. Crystal Growth. Physical Properties. Crystallography.

Building Research. Structural Engineering. Fire Research. Mechanical Systems. Organic Building Materials. Codes and Safety Standards. Heat Transfer. Inorganic Building Materials. Metallic Building Materials.

Applied Mathematics. Numerical Analysis. Computation. Statistical Engineering. Mathematical Physics. Operations Research.

Data Processing Systems. Components and Techniques. Computer Technology. Measurements Automation. Engineering Applications. Systems Analysis.

Atomic Physics. Spectroscopy. Infrared Spectroscopy. Far Ultraviolet Physics. Solid State Physics. Electron Physics. Atomic Physics. Plasma Spectroscopy.

Instrumentation. Engineering Electronics. Electron Devices. Electronic Instrumentation. Mechanical Instruments. Basic Instrumentation.

Physical Chemistry. Thermochemistry. Surface Chemistry. Organic Chemistry. Molecular Spectroscopy. Elementary Processes. Mass Spectrometry. Photochemistry and Radiation Chemistry.

Office of Weights and Measures.

BOULDER, COLO.

Cryogenic Engineering Laboratory. Cryogenic Equipment. Cryogenic Processes. Properties of Materials. Cryogenic Technical Services.

CENTRAL RADIO PROPAGATION LABORATORY

Ionosphere Research and Propagation. Low Frequency and Very Low Frequency Research. Ionosphere Research. Prediction Services. Sun-Earth Relationships. Field Engineering. Radio Warning Services. Vertical Soundings Research.

Radio Propagation Engineering. Data Reduction Instrumentation. Radio Noise. Tropospheric Measurements. Tropospheric Analysis. Propagation-Terrain Effects. Radio-Meteorology. Lower Atmosphere Physics.

Radio Systems. Applied Electromagnetic Theory. High Frequency and Very High Frequency Research. Frequency Utilization. Modulation Research. Antenna Research. Radiodetermination.

Upper Atmosphere and Space Physics. Upper Atmosphere and Plasma Physics. High Latitude Ionosphere Physics. Ionosphere and Exosphere Scatter. Airglow and Aurora. Ionospheric Radio Astronomy.

RADIO STANDARDS LABORATORY

Radio Physics. Radio Broadcast Service. Radio and Microwave Materials. Atomic Frequency and Time-Interval Standards. Radio Plasma. Millimeter-Wave Research.

Circuit Standards. High Frequency Electrical Standards. High Frequency Calibration Services. High Frequency Impedance Standards. Microwave Calibration Services. Microwave Circuit Standards. Low Frequency Calibration Services.

NBS

No. 104 not issued

No. 105 not issued