Convention Record

of the I·R·E

1953 NATIONAL CONVENTION

Part 2—Antennas and Communications

SESSIONS ON . . .

Antennas I — General

Antennas II — Microwave

Antennas III — Propagation

Symposium: Trends in Mobile Communications

Communications Systems

SPONSORED BY

IRE PROFESSIONAL GROUPS ON . . .

Antennas and Propagation

Communications Systems

Vehicular Communications

Presented at the IRE National Convention, New York, N.Y., March 23 - 26, 1953

The Institute of Radio Engineers

CONVENTION RECORD OF THE I.R.E.

1953 NATIONAL CONVENTION

PART 2 - ANTENNAS AND COMMUNICATIONS

TABLE OF CONTENTS

THE MEASUREMENT OF HIGHLY DIRECTIVE ANTENNA PATTERNS AND OVER-ALL SENSITIVITY OF A RECEIVING SYSTEM BY SOLAR AND COSMIC NOISE

Jules Aarons
Upper Air Laboratory, Geophysics Research Directorate
Air Force Cambridge Research Center
Cambridge 39, Mass.

Summary--Utilizing the more intense of the over 100 celestial sources of radio frequency energy and known data about the effective diameter of the sun at various ranges of radio frequencies, a technique for plotting the directional characteristics of large antennas is outlined. Over-all system sensitivities (receiver, antenna, and transmission line) are checked by using values already obtained for sky temperatures. The general receiver characteristics necessary for such measurements are outlined. Patterns of the sun and cosmic sources are illustrated and their analysis evaluated.

I. Background

Basic research in astronomy has, within the past five years, used two new tools extensively, the low noise stable radio receiver and the highly directive antenna. A new field has been opened, i.e., radio astronomy. Both the sun and the stars have been explored extensively with this tool.

The greatest amount of energy, as recorded by observations, from both of these sources is in the low radio frequencies 18 mc to 300 mc. The cosmic sources have been explored almost wholly in this frequency region. The sun maintains its irregular or disturbed component basically in the low frequencies and has a rather quiet radiation at the microwave end of the radio spectrum.

Equipment used to pick up these sources, while in most cases specifically designed for the task, is not unique in the field of electronics, utilizing only good basic principles to produce exceedingly stable units.

Several workers in the field now use previously established facts about the sun and the cosmic sources to calibrate their antennas. For other antennas which have a large effective area, the cosmic and solar sources might prove to be useful. These sources are never, of course, in the near field; they can be located precisely at all times using astronomical tables; they present a means for checking the over-all system sensitivity over long periods of time in case of deterioration of equipment. This technique of sensitivity calibration would aid in maintaining similar sets of identical equipment at constant levels by providing a "primary" noise source standard.

II. Receiver Principles Used in the Design of Radio Telescopes

For these measurements three basic factors are important.

1. The most vital consideration in over-all design of the receiver is its stability. The Australian scientists[1] in their cosmic noise instrumentation have calculated that in a typical 100 mc receiver a change of one part in 3,000 in the B plus voltage or one part in 1500 of the filament voltage produces changes of one part in 1,000 in second detector voltage. The equipment they have designed achieves stability of one part in 10,000 for the B plus voltage, and one part in 6,000 for the filament voltage. They, therefore, were enabled to obtain changes in sky temperatures of .3 degrees Kelvin. A full wave square wave rectifier is used to detect the noise output.[2] This represents the optimum type of rectifier[2]. Bucking voltage is set up to rid the receiver of the steady state noise and a milliammeter or microammeter records the signal. Other laboratories, such as the Cornell University Radio Astronomy Laboratory, have achieved somewhat similar orders of magnitude for their equipment.

2. In addition to constant supply voltages and DC filament voltage (in some cases), these low noise receiving systems use long time constants in their recorders and DC amplifiers.

The minimum signal which can be detected is proportional to the square root of the bandwidth-time constant product.

$$\frac{P_{abs}}{P_n} = \frac{1}{\sqrt{BC}} \qquad (1)$$

P_{abs} = minimum detectable signal power

P_n = power generated in the receiver and delivered to the receiver

B = bandwidth in cycles per second

C = time constant

From the analysis suggested by Dicke[3], the output is the sum of a large number of successive random pulses, each of average duration $1/B$. In a time interval ΔT (the time constant of the output system) $B\Delta T$ pulses are received. Solar noise bursts have been of the order of one half second or longer, and time constants have been maintained in

that range. Bandwidths have varied, although in many cases bandwidths have been below radar or television bandwidth needs.

3. Atmospheric noise, ignition noise, etc., in the vicinity of the antennas is an important factor. When the design conditions outlined above have been satisfied, this third factor takes over as the predominant minimum noise level. Use of quiet periods of the day when this noise is at its minimum helps, but it does not overcome the predominance of this type of noise.

III. The Sources

a. The Sun

The sun is a variable source of radiation emitting on top of its normal "quiet" level signal an enhanced base or long term level plus short-lived bursts or outbursts. The radio stars are all constant sources of radiation, although in at least one case an ionospheric fluctuation is superimposed on the steady signal.

Angular width and general activity vary on different frequencies. (Fig. 1) The width of the photosphere, the optical width of 31 minutes, 59.3 seconds, approaches the radio frequency width at 1 cm. However, at 90% of the emergent intensity of its central ray at 60 mc the sun appears as a source of 47 minutes 11 seconds. Essentially, the radio disk is determined by the particular layers of the solar atmosphere which radiate energy predominantly in definite portions of the radio frequency spectrum.

The general activity at the low end of the radio frequency spectrum is much greater than in the microwave region. For the former, during a disturbed day, several hundred bursts may occur within an hour at 200 mc. Cases occur when the total power (bursts plus base level) exceeds quiet sun base level power by a factor of one hundred. Quiet days without any bursts occur, too. Even during a moderately disturbed day, a quiet period can occur long enough to run through an antenna sweep and, therefore, obtain a useable pattern.

At wavelengths of the order of 10 cm and 3 cm, the picture is quite different. At 10 cm, bursts of energy lasting up to 9 minutes were recorded occasionally.[4] These bursts are in addition to a longer time constant component or base level. "Sudden increases of radiation at 3.18 cm caused by disturbed conditions in the sun, were found to be rare."[5] It was found that observable changes in magnitude occurred for a total of approximately fifteen minutes in an observing time of 230 hours. The increase in intensity ranged from six percent to over thirty percent and lasted for periods of the order of a minute.

b. Cosmic Noise

Much exploratory work has been done in the radio frequency band of 18 mc to 400 mc in cosmic noise sources. At 1200 mc, Cygnus A, the most intense "point" source is rather diffuse.[6] At 3,000 mc no point sources have been located.

In the lower frequency range, well over 100 of these radio stars have been located accurately, the widths of the sources ranging from one minute of arc to thirty minutes of arc. The intensity varies from 128×10^{-24} watts/sq meter/cycle/second/steradian downward at a frequency of 100 mc.[7] The amount of the investigation of the lower valued sources has not been large enough to warrant certifying many of them as bona fide sources.

A few of the sources, depending on viewing angle, have fluctuating components. One of these, Cygnus A, had variations of about 10% at 200 mc. These fluctuations decrease with frequency. The origin of fluctuations is definitely of ionospheric nature, being correlated with an index of F2 activity. The superimposition on the steady state signal of this ionospheric twinkling is due to the low angle of elevation, the closeness of the meridian.

IV. Measurements

a. Technique

Taking measurements with this technique must be done with the aid of radio astronomy celestial maps. At the meter wavelengths the solar source cannot be used if the milky way is in the background. The same holds true for radio stars. Those sources which are near the galactic center will stretch out the antenna pattern. Since we now have a gaussian distribution to contend with, the equation for actual antenna angle is

$$\theta_t = \sqrt{\theta_m^2 - \theta_s^2} \qquad (2)$$

θ_t = the true antenna angle

θ_m = the measured antenna angle

θ_s = the gaussian curve of the contours through which the antenna is sweeping

For the case of the point source which is sharp and above the intensity of neighboring contours, the particular curve must be considered in subtracting source size from measured size in order to obtain the true antenna angle. In the case of the sun, the intensity may be considered rectangular for a good approximation at the shorter wavelengths but only a fair approximation at the longer wavelengths. The sides are sloping, but the top is flat. In this case, then, the angle subtended by the sun may be subtracted from the measured antenna angle and the true angle obtained.

When the sun is quiet, a sweep across the sun can be taken in a short period if the antenna is movable in altitude and azimuth. Since the

rotation of the earth with respect to the sun is slow (moving 15 minutes of arc in one minute) the time of sweep would be determined by the time constants of the circuit and the physical means of rotating the antenna.

If the stars are considered as points on the inner surface of a rotating sphere, there lies within this sphere another globe, the earth. The axes of rotation of the two spheres are inclined towards one another; they are not parallel. This failure to "line up" will produce an incorrect pattern for an antenna which is fixed in azimuth and the sources allowed to sweep across it. These patterns can be measured reasonably well if the beam angle is small and the source, therefore, takes only a small amount of time to cross the antenna pattern.

After taking the pattern, the correction factor of sky contours must be applied. Isothermal lines at various frequencies have roughly the same shape. At 18.3 mc the contours are similar to 100 mc.[8] A field technique which might be put to use is to make measurements on a particular antenna under laboratory controlled conditions, reducing losses to a minimum, and achieving an adequate test pattern. If an antenna test pattern is then obtained of the sun or a cosmic source, a standard can be set up for pattern and sensitivity comparison.

b. Measurements at Cornell University

Examples of the measurement of antenna beam angles are simple to analyze for a first approximation of the pattern. Figure 2 illustrates a swing of the antenna across the sun on 16 October 1952 on 200 mc by the Cornell University solar noise equipment. A peak value of 37.5 units was subtracted from a background value of 7 units. Therefore, the half power points are 18 degrees apart. Subtracted from this is the value of the width of the source (at 90% down the emergent intensity) or 38 minutes. The resulting value is 17 degrees 22 minutes, assuming the solar source as a rectangle. Figure 3 illustrates measurements taken of the side lobe patterns. This was done by increasing the gain of the equipment and concentrating on this particular portion of the pattern.

On the night of 29 September 1952 the Cornell antenna was set to the declination of Cygnus and an antenna pattern taken. A peak of 46 units and a minimum of 7 were obtained. The half power points were 1 hour 45.86 minutes apart.

$$\frac{4 \text{ minutes}}{\cos \text{ declination}} = 1 \text{ degree}$$

A value of 20 degrees is obtained for the half power points. Since Cygnus lies in one of the most "congested" areas of the sky from the cosmic noise point of view, this is the maximum error to be obtained before correction by means of contours. Although the Cygnus source is very small (one minute to one minute 30 seconds), the area in which it is situated has a high energy level.

On interferometer patterns, or patterns where maximum and minimum lie close together, shallow minima show an extended source, while sharp deep minima indicate a small source. In the case of the Cornell University interferometer, at 200 mc (Figure 4) the pattern is similar to that of sine waves. Each cycle measures 1 degree 7 minutes. Cygnus was used as the source. For Cassiopeia we obtained a figure of 1 degree 20 minutes for each cycle.

V. Sensitivity

a. Over-all Sensitivity Measurements

A possibility exists for measuring changes in the sensitivity of the receiving equipment from the antenna to the recorder. Since the essential criterion of the proposed system will rely on its ability to distinguish between noise levels, let us first consider some of the noise equations.

The available noise from a resistor[9] is:

$$P_n = \frac{e_n^2}{4R} = KT_aB \qquad (3)$$

P_n = available noise power in watts

e_n = RMS voltage

K = Boltzmann's constant 1.37×10^{-23} watt-sec/degree K

T_a = "ambient" temperature in degrees Kelvin usually taken at Friis' standard of 290 degrees K or 17 degrees C

$KT = 4 \times 10^{-21}$ watt seconds

The noise figure (F) of a receiver under test is the ratio of the actual noise power output to the noise power output of an ideal network.

The component of the noise figure due to internally generated noise (all noise other than the input resistor noise) is equal to F-1.

In order to detect a small amount of energy being radiated from external sources, the power absorbed by the antenna and delivered to the receiver must be of the same magnitude or larger than the fluctuations in the internally generated noise of the receiver. If the internally generated noise were perfectly steady in its amplitude, it would be theoretically possible to detect any increase caused by external power being absorbed. The variable component of the noise is a function of the bandwidth and the time constant of the circuit, provided the supplies are stable and no outside interference takes place.

Equation 1 shows that for a bandwidth of one megacycle and a time constant of one second in the recording apparatus it is theoretically possible to detect signals of the order of 1/1000 of the internal noise.

$$P_{abs} = \frac{KT_aB\ (F-1)}{\sqrt{BC}} \qquad (4)$$

P_{abs} = K T B (for the case of homogeneous areas of emission - see Section V. c.) (5)

$$T_c = \frac{T_a\ (F-1)}{\sqrt{BC}} \qquad (6)$$

T_c = Minimum detectable sky temperature

With the measurement of P_{abs} (the actual amount of power which the antenna must absorb before the power output rises above the internal noise power), and with the measurement of the noise figure, the bandwidth, and the time constant, a new criterion for receiver evaluation is available, i.e., that of stability.

The ratio of measured power absorbed (before a change of level is noted) to theoretical power absorbed by the antenna yields the stability factor M. A value of one, just as in the case of the noise figure, is obtained when the unit is perfectly stable.

$$\frac{T_{cm}}{T_{ct}} = M \qquad (7)$$

T_{cm} is the measured temperature of the homogeneous sky source which yields a level which can be distinguished above the fluctuations of receiver noise.

$$T_{ct} = \frac{T_a\ (F-1)}{\sqrt{BC}}$$

A change of the stability factor as evidenced by inability to detect certain radiating areas of the sky will be caused by an increase in F, a change in the system stability due to power supply fluctuations (assuming bandwidth and time constant are fixed), or a deterioration in the antenna system (transmission line, mismatching, etc.). Antenna gain considerations do not enter into the equation.

In the case of receiving systems dependent on comparison techniques, the stability can be an important consideration. A simple means of detecting increase in the noise factor or other parameters of the receiver is afforded by the technique of pointing the antenna at the sky.

As a routine check on the stability factor (M), one technique is to attempt to distinguish between two sets of contours or two areas. Instead of using the absolute power from one area, the difference in power between two areas with a temperature difference of ΔT can be utilized. The receiver should be able to distinguish between two areas which have

$$\Delta T_{cm} = \Delta T_{ct}\ M$$

Figure 5 represents Cornell University's 200 mc stabilization data which are taken daily. Here a reference resistor at 290 degrees represents a zero point. The levels above and below this are for portions of the sky centered at the location given. Utilizing Equation 6

F = 5.9

B = 5mc

C = 1/2 second

a ΔT_{ct} of .9 degrees K is obtained.

In one radio astronomy project designed for somewhat different purposes, an M of 2 was obtained. For normally stabilized radio receivers it is estimated that an M of 100 would result.

b. $\theta_{Antenna} > \theta_{Energy\ Sources}$

The energy density of black body radiation in a wavelength increment $d\lambda$ is expressed by Planck's law.

$$\psi_\lambda d\lambda = \frac{8\pi ch}{\lambda^5} \quad \frac{1}{e^{ch/\lambda KT}-1} \qquad (8)$$

ψ_λ = energy density

c = velocity of light

In the region of interest to this paper (microwave and meter radiation at solar temperatures) KT >> hf

since $h = 6 \times 10^{-34}$ watt sec^2

$k = 1.37 \times 10^{-23}$ watt sec/deg K

After summing into total emissive power and converting the equation into frequency form, the power absorbed by an antenna with the energy source smaller than the beam angle[10]

$$P_{abs} = \frac{G_0 K \Omega_S T_S}{4\pi} B \qquad (9)$$

P_{abs} = power absorbed by the radiation resistance of the antenna in watts

Ω_S = solid angle subtended by the optical disk of the sun

T_S = equivalent black body temperature of the sun in degrees Kelvin

G = the gain in the maximum direction

Figure 6 gives the value of $k\Omega_S T_S$ for the sun as a function of frequency.

While the solid angle subtended by the sun at various frequencies is known (although some details

are in dispute), the values of the widths of the celestial point sources are not as reliable. For rough estimates they can be used in the equations.

c. $\theta_{\text{Energy Source}} > \theta_{\text{Antenna}}$

If the entire area of the antenna's main lobe is homogeneously filled with radiation, the directional characteristics of the antenna are not related to the equation.

$$P_{abs} = k\ T\ B$$

For a large antenna, the measurements involving portions of the sky which are constant could be used to obtain a field intensity in microvolts/meter.

No contours are available for the microwave region due to the low temperature differences involved. For the galaxy, the effective temperature varies approximately as λ^{-2}. At 4.7 meters, Hey, Parsons, and Phillips (1948) obtained a maximum of 18,000 degrees K.[11] Thirty degrees K was obtained from Reber's data on 60 cm for the same region.

VI. Conclusion

The possibilities for the use of this technique lie as a field tool for checking patterns and sensitivity. The inaccuracies involved in evaluating background radiation, contours, and solar energy distribution are such that a very accurate laboratory field set-up would be a better tool. However, for a check on field instruments as to their noise level, stability, antenna, pattern, and terrain fluctuations, the technique can be of use.

VII. Acknowledgment

I would like to thank S. M. Colbert of the Cornell University Radio Astronomy Laboratory for his assistance in obtaining the data for this report. I would also like to thank Prof. Charles Burrows for his cooperation with the work. Miss Martha Henissart assisted in preparing the graphs and the data and in many of the problems of interpretation.

References

1. Allen, C. W. and Gum, C. S. "Survey of Galactic Radio-Noise at 200 Mc/S", Aus. J. Sci. Res. A. (1950) v. 3, p. 224

2. Mayer, H.F., "Radio Noise Considerations"; Solar Noise Technical Report No. 1, AMC Contract W19-122 ac-41, Cornell University, Ithaca, N. Y.

3. Dicke, R.H. "The Measurement of Thermal Radiation at Microwave Frequencies", Rev. Sci. Instr., (1946) v. 17, p. 268

4. Covington, A.E. "Solar Noise Observations on 10.7 Cm", Proc. IRE (1948) v. 36, p. 454-7

5. Minnett, H.C. and Labrum, N.R. "Solar Radiation at a Wavelength of 3.18 cm", Aus. J. Sci. Res. A. (1950) v. 3, p. 60

6. Piddington, J. H. and Minnett, H.C. "Radio Frequency Radiation from the Constellation of Cygnus", Aus. J. Sci. Res. A. (1952) v.5, p. 17

7. Stanley, G.J. and Slee, O.B. "Galactic Radiation at Radio Frequencies", Aus. J. Sci. Res. A. (1950) v. 3, p. 234

8. Shain, C.A. "Galactic Radiation at 18.3 Mc", Aus. J. Sci. Res. A. (1951) v. 4, p. 258

9. Friis, H.T. "Noise Figures of Radio Receivers", Proc. IRE, (July, 1944) v. 32, p. 419

10. Van de Hulst, H.C. "A Course in Radio Astronomy", Leiden Observatory, (1951)

11. Hey, J.S. and Parsons, S.S. and Phillips, J.W. "An Investigation of Galactic Radiation in the Radio Spectrum", Proc. Roy. Soc. (1948) v. 192, p. 425

12. Smerd, S.F. "Radio Frequency Radiation from the Quiet Sun", Aus. J. Sci. Res. A. (1950) v. 3, p. 34

13. Hagen, J.P. "Temperature Gradient in the Sun's Atmosphere Measured at Radio Frequencies", Astro. J. (1951) v. 113, p. 547

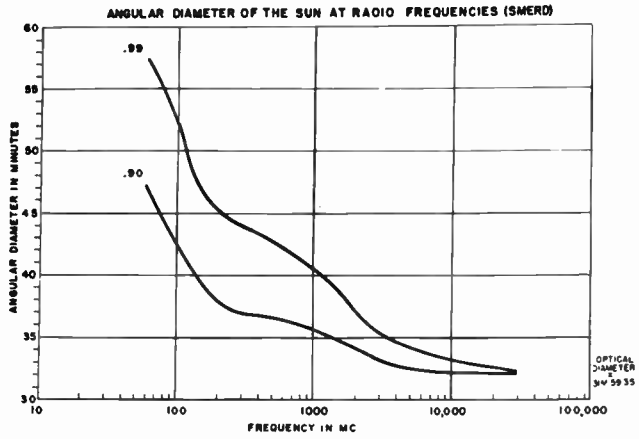


Fig. 1 - Angular diameter of the sun at radio frequencies (SMERD).

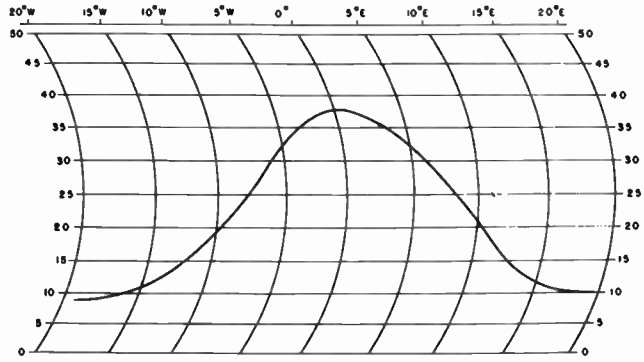


Fig. 2 - Sweep across sun on Oct. 16, 1952. Main lobe, 200-mc antenna.

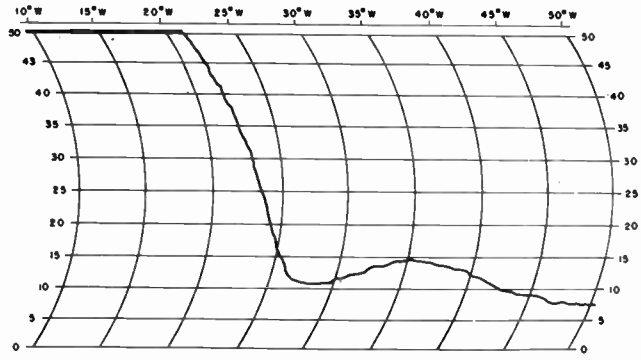


Fig. 3 - Sweep across sun on Oct. 16, 1952. Side lobe, 200-mc antenna.

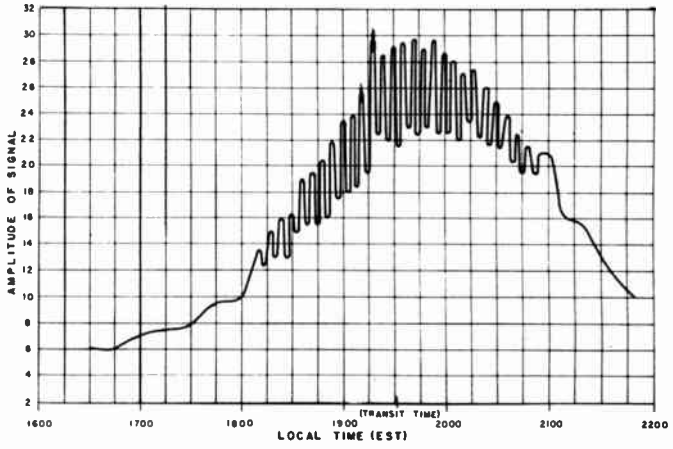


Fig. 4
200-mc interferometer pattern
of R Cygnus, Sept. 29, 1952.

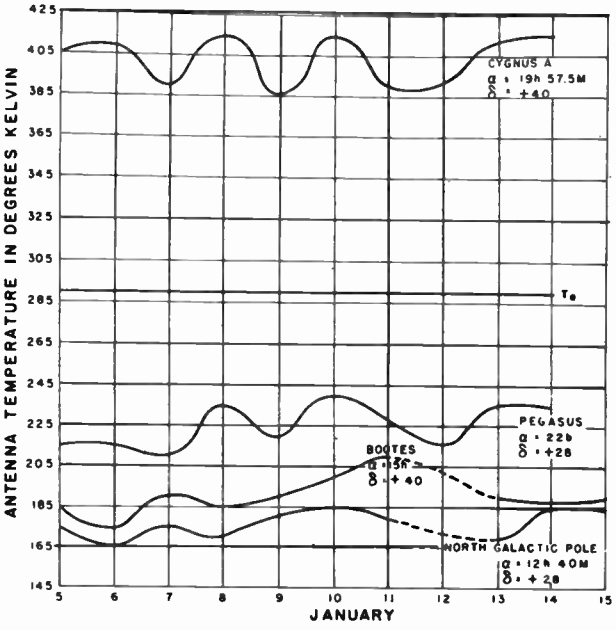


Fig. 5
Stability measurements by means of
sky noise, Jan. 5-15, 1953.

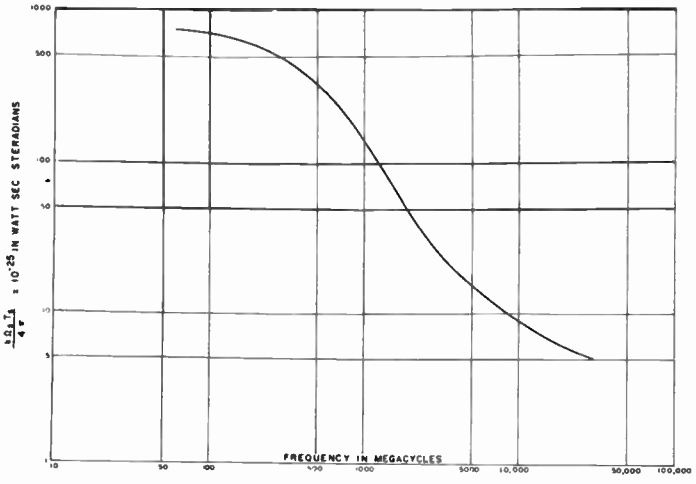


Fig. 6
Effective solar temperature as a
function of frequency (Hagen).

RADIATION PATTERNS FOR APERTURE ANTENNAS WITH NON-LINEAR PHASE DISTRIBUTIONS

Charles C. Allen
General Electric Company
Schenectady, New York

I. Introduction

Two dimensional antenna patterns have been determined for several aperture distributions having symmetrical amplitude and non-linear phase by a method of calculation employing automatic punch-card machines to perform numerical integrations. These patterns have been arranged in sets which permit one to predict the pattern characteristics expected from a given aperture phase distribution.

Spencer and Austin[1] have evaluated the pattern integrals for quadratic and cubic phase distributions with both uniform and cosine amplitude distributions. The method used approximates the pattern by the first few terms of a series which involves derivatives of the pattern obtained with a uniform phase distribution.

Friis and Lewis[2] have used a substitution resulting in Fresnel integrals to evaluate the pattern for quadratic phase and uniform amplitude distributions. The first few terms of a series expansion of the integrand were used to determine the pattern for cubic phase and uniform amplitude distributions.

The method of calculation used here permits direct evaluation of the pattern integral with negligible error for any phase and amplitude distributions. The results presented extend previous work to the cases of quadratic and cubic phase distributions combined in various proportions together with uniform, tapered, and cosine amplitude distributions. Phase patterns, which permit determination of the effective aperture phase center, are presented as well as the magnitude patterns.

II. Method of Calculation

Let the aperture distribution along a one dimensional aperture of length 2a be linearly polarized and given in amplitude and phase by

$$\hat{F}(\xi) = \hat{f}(\xi)\, e^{j\hat{\Psi}(\xi)} \tag{1}$$

where ξ = aperture coordinate in actual units. If 2a is several wavelengths and the aperture phase deviation, $\hat{\Psi}(\xi)$, is not excessive, the far-field antenna pattern is given by[3]

$$\hat{E}(\theta) = \int_{-a}^{a} \hat{F}(\xi)\, e^{j\beta\xi \sin\theta}\, d\xi \tag{2}$$

where $\beta = 2\pi/\lambda$, θ = angle from normal to aperture, and λ = wavelength.

Let the aperture coordinate, ξ, and the angle, θ, be normalized as follows:

$$\left.\begin{aligned} x &= \xi/a \\ u &= \beta a \sin\theta \end{aligned}\right\} \tag{3}$$

Equations (1) and (2) then become

$$F(x) = f(x)\, e^{j\Psi(x)} \tag{4}$$

and

$$E(u) = a\int_{-1}^{1} F(x)\, e^{jux}\, dx = a\int_{-1}^{1} f(x)\, e^{j[ux + \Psi(x)]}\, dx \tag{5}$$

Equation (5) may be written

$$E(u) = a\int_{-1}^{1} f(x)\,[\cos\Phi(x) + j\sin\Phi(x)]\, dx = p(u) + jq(u) = g(u)\, e^{j\alpha(u)} \tag{6}$$

where $\Phi(x) = ux + \Psi(x)$. The magnitude pattern is given by g(u) and the phase pattern by $\alpha(u)$. The angle $\alpha(u)$ represents the deviation of the equi-phase contour of the far field from a spherical wavefront having its center at the origin in the aperture.

For numerical integration on automatic punch-card machines, equation (6) is written

$$E(u) \doteq a\sum_{i=1}^{n} f(x_i)[\cos\Phi(x_i) + j\sin\Phi(x_i)]\,\Delta x \tag{7}$$

where n = number of increments in the aperture, Δx = width of each increment, and x_i = value of x at the center of the i^{th} increment. It is seen that the approximation is in the integral rather than the integrand and can be made as close as desired by selecting a sufficient number of increments in the aperture.

The aperture gain factor relative to that for uniform illumination of the aperture is given by[4]

$$\mathcal{G} = \frac{1}{2}\,\frac{|E(u)|^2_{max}}{aP} \tag{8}$$

where

$$P = a\int_{-1}^{1}[f(x)]^2\, dx \doteq a\sum_{i=1}^{n}[f(x_i)]^2\,\Delta x \tag{9}$$

It can be shown that for a given amplitude distribution, f(x), the greatest maximum magnitude of E(u) is obtained with a uniform phase distribution, $\Psi(x) = 0$. A non-linear

phase distribution reduces the maximum magnitude of E(u) and results in a decreased gain factor.

III. Sets of Patterns

A. Arrangement.

The patterns have been arranged in sets such that each set is for one phase distribution and three amplitude distributions. The amplitude distributions are of the form

$$f(x) = M + N \cos \frac{\pi}{2} x \qquad (10)$$

with M and N selected as in the following table.

Amplitude Distribution	M	N
Uniform	1.000	0.000
10 DB Tapered	0.316	0.684
Cosine	0.000	1.000

The set of reference patterns given in Fig. 1 is for a uniform phase distribution, $\Psi(x) = 0$. The peak pattern magnitude is taken as zero decibels in the reference patterns for each amplitude distribution. The magnitude patterns for other phase distributions are plotted relative to the zero decibel level of the reference patterns.

Ten sets of magnitude patterns are given in Figs. 2 through 11 for phase distributions of the form

$$\Psi(x) = -\delta(Ax^2 + Bx^3) \qquad (11)$$

Two values of δ equal to $\pi/4$ and $\pi/2$ are used with five pairs of values of A and B as given in the following table.

Phase Distribution	A	B
Quadratic	1.0	0.0
Quadratic + Cubic	1.0	0.5
Quadratic + Cubic	1.0	1.0
Quadratic + Cubic	0.5	1.0
Cubic	0.0	1.0

As will be seen in the discussion which follows, the phase pattern for a cubic phase distribution is alternately 0° and 180°. Only eight sets of phase patterns are given, therefore, in Figs. 12 through 19. The phase patterns are given only for the main beam of the magnitude patterns and determine the phase slope of the far field in the direction of peak intensity.

B. Pattern Characteristics.

In this discussion of pattern characteristics, a symmetrical aperture amplitude distribution is assumed in all cases; that is, f(x) is an even function of x.

For a uniform phase distribution, $\Phi(x) = ux$. Since this is an odd angle function of x, q(u) in equation (6) is zero for all values of u, and the pattern, E(u) = p(u), is real. The pattern lobes have alternating signs and zeroes in between; hence, $\alpha(u)$ is alternately 0° and 180°.

Consider now a quadratic phase distribution giving $\Phi(x) = ux - \delta x^2$. Since this is not an odd angle function of x, both p(u) and q(u) exist, and the pattern, E(u), is complex. The nulls of the magnitude pattern, g(u), are filled in by q(u) which is not zero at the zeroes of p(u); therefore, the phase pattern is an odd multiple of 90° near the nulls of the magnitude pattern. Since $\Phi(x)$ for a given value of u equals $\Phi(-x)$ for -u, p(-u) = p(u) and q(-u) = q(u), or E(-u) = E(u). Both the magnitude pattern and the phase pattern have even symmetry about u = 0, therefore, as seen in Figs. 2, 7, 12, and 16. The main effects of a quadratic phase distribution are a reduction in the maximum pattern magnitude and gain factor, an increase in side lobe levels, and filled in nulls.

Consider next a cubic phase distribution giving $\Phi(x) = ux - \delta x^3$. Since this is an odd angle function of x, q(u) is zero for all values of u, and the pattern, E(u) = p(u), is real. Since $\Phi(x)$ for a given value of u does not equal $\Phi(-x)$ for -u, p(-u) does not equal p(u) in general and the pattern is asymmetrical as seen in Figs. 6 and 11. The cubic phase distribution tilts the main beam so that the peak intensity occurs at the value of u for which $\Phi(x)$, when properly weighted by the amplitude distribution, most nearly approximates uniform phase. The side lobe levels are increased in the direction of beam tilt and are decreased on the other side of u = 0. The first side lobe of increased level is termed the "coma" lobe and is generally considered to occur toward the direction u = 0 from the tilted main beam. This is true in practice because a cubic phase distribution is generally accompanied by a linear phase distribution of opposite sign which tilts the entire pattern in the opposite direction without changing its shape.

The combined effects of quadratic and cubic phase distributions are seen in varying degrees in Figs. 3, 4, 5, 8, 9, and 10. The quadratic phase term has filled in the nulls and made the patterns complex. The corresponding phase patterns are in Figs. 13, 14, 15, 17, 18, and 19. The cubic phase term has tilted the pattern, increased the side lobe levels in the direction of beam tilt, and reduced the side lobe levels on the other side of the main beam.

C. Change of Signs.

The sets of patterns given in Figs. 2 through 19 can be used for any combination of signs of A and B in equation (11) by making the appropriate change in the sign of u or $\alpha(u)$. Consider first a change in the sign of $\Psi(x)$; that is, let both A and B be negative. For a given $\Phi(x) = ux + \Psi(x)$, changing the sign of $\Psi(x)$ and of u gives $-\Phi(x) = -ux - \Psi(x)$. Since $\cos[-\Phi(x)] = \cos\Phi(x)$ and $\sin[-\Phi(x)] = -\sin\Phi(x)$, p(u) and q(u) for $-\Psi(x)$ equal respectively p(-u) and -q(-u) for $+\Psi(x)$. The pattern for $-\Psi(x)$ is, therefore, the conjugate of the pattern for $+\Psi(x)$ with the sign of u reversed and may be summarized

as follows

$$g(u)_{-\psi} = g(-u)_{+\psi}$$

$$\alpha(u)_{-\psi} = -\alpha(-u)_{+\psi}$$

or, $$E(u)_{-\psi} = E^*(-u)_{+\psi}$$

Consider next a change in the sign of only B. For a given $\bar{\Phi}(x) = ux - \delta(Ax^2 + Bx^3)$, changing the signs of B and u gives $\bar{\Phi}(-x) = -ux - \delta(Ax^2 - Bx^3)$. Since the integration in equation (6) gives the same result for $\bar{\Phi}(-x)$ as for $\bar{\Phi}(x)$, p(u) and q(u) for -B equal respectively p(-u) and q(-u) for +B. The pattern for -B is, therefore, the same as the pattern for +B with the sign of u reversed and may be summarized as follows

$$g(u)_{-B} = g(-u)_{+B}$$

$$\alpha(u)_{-B} = \alpha(-u)_{+B}$$

or, $$E(u)_{-B} = E(-u)_{+B}$$

Consider last a change in the sign of only A. For a given $\bar{\Phi}(x) = ux - \delta(Ax^2 + Bx^3)$, changing the sign A gives $-\bar{\Phi}(-x) = ux - \delta(-Ax^2 + Bx^3)$. Since the integration of $\cos[-\bar{\Phi}(-x)]$ and of $\cos\bar{\Phi}(x)$ are the same, and the integration of $\sin[-\bar{\Phi}(-x)]$ and of $\sin\bar{\Phi}(x)$ only differ in sign, p(u) and q(u) for -A equal respectively p(u) and -q(u) for +A. The pattern for -A is, therefore, the conjugate of the pattern for +A and may be summarized as follows

$$g(u)_{-A} = g(u)_{+A}$$

$$\alpha(u)_{-A} = -\alpha(u)_{+A}$$

or, $$E(u)_{-A} = E^*(u)_{+A}$$

IV. Pattern Synthesis

While arbitrary criteria for the allowable magnitudes of the quadratic and cubic phase distributions have been given for some cases, such as those given by Ruze[5], the pattern requirements on which such criteria should be based vary appreciably with the particular application of an antenna. By using the sets of patterns presented here for semi-quantitative pattern synthesis, the suitability of a given aperture distribution for a particular application can be determined.

A. Method.

First, a systematic method of comparing an actual phase distribution with the theoretical set of distributions will be discussed. Consider the arbitrary phase distribution $\psi'(x)$ in Fig. 20(a). Let this be represented by a power series

$$\psi'(x) = K' + L'x + Cx^2 + Dx^3 + Ex^4 + \cdots \quad (12)$$

Subtract the straight line, K + Lx, joining the ends of $\psi'(x)$ from $\psi'(x)$ to obtain

$$\psi(x) = \psi'(x) - (K + Lx)$$

$$= A + Bx + Cx^2 + Dx^3 + Ex^4 + \cdots \quad (13)$$

where $A = K' - K$ and $B = L' - L$. This requires that $\psi(x) = 0$ at $x = \pm 1$ as shown in Fig. 20(b). If $\psi(x)$ is now approximated by the first four terms of equation (13) subject to the conditions that $\psi(x) = 0$ at $x = \pm 1$, it is necessary that $C = -A$ and $D = -B$ which gives

$$\psi(x) \cong A + Bx - Ax^2 - Bx^3 \quad (14)$$

A family of curves is given in Fig. 21 for equation (14) with A = 1 for B between 0 and 1, and B = 1 for A between 1 and 0. Other curves of the family can be obtained by interpolation. The complete family of curves represents all possible shapes of $\psi(x)$ given by equation (14).

In comparing an actual phase distribution with the family of curves in Fig. 21 to determine the best values of A and B to use in equation (14), the phase is first reduced to zero at $x = \pm 1$ as discussed above. The magnitude of $\psi(x)$ is then adjusted to obtain the best fit to a member of the family of curves by multiplying by a constant, $1/\delta$. The values of A and B for that member are used in the right hand side of equation (14) which is then multiplied by the factor δ to obtain the correct phase magnitude. This gives

$$\psi(x) \cong \delta(A + Bx - Ax^2 - Bx^3) \quad (15)$$

The constant term, $K' = K + \delta A$, merely adds a constant phase angle to $\alpha(u)$ for all values of u. The linear term, $L'x = (L + \delta B)x$, merely tilts the pattern by an amount $u = -(L + \delta B)$.

Since the constant and linear terms of $\psi(x)$ can be handled separately, the actual phase distribution as approximated by equation (15) is further reduced for use with the sets of patterns to

$$\psi(x) \cong -\delta(Ax^2 + Bx^3) \quad (16)$$

A three way interpolation can then be performed with the sets of patterns which are selected to properly bracket $\psi(x)$ as regards the values of A, B, and δ. For each set, the patterns are interpolated for the amplitude distribution taper being used with the given phase distribution. The interpolated patterns from each set are then interpolated as regards A and B (or δ), and the last interpolation is with respect to δ (or A and B).

The quantities obtained from these interpolations are the side lobe levels, the null levels, the beam width, the beam tilt, and the gain factor. In cases where the amplitude distribution being used is in fair agreement with equation (10), these quantities can be used directly. Where this is not the case, it is probably more correct to determine the changes in these quantities from their values for a uniform phase distribution, $\psi(x) = 0$. These changes can then be considered as applying to the pattern obtained from the actual amplitude

distribution and a uniform phase distribution.

B. Example.

To illustrate the method of pattern synthesis, consider the amplitude distribution, f(x), and the phase distribution, $\psi(x)$, given in Fig. 22. These distributions are typical of those obtained with a parabolic reflector having a feed displaced from the focus perpendicular to the axis. f(x) is tapered to about 25 db at x = -1 and 16.5 db at x = 1 and is nearly cosine in shape; therefore, a taper of 20 db is assumed for interpolation.

The straight line connecting the ends of $\psi(x)$ is given by

$$K + Lx = -0.91 + 14.54x \qquad (17)$$

A plot of $\psi(x) - K - Lx$ when multiplied by 1/1.26 interpolates into the standard curves of Fig. 21 at A = 0.06 and B = 1.0. The approximated phase distribution is therefore

$$\psi(x) \cong 1.26\ (0.06 + x - 0.06x^2 - x^3) \qquad (18)$$

For interpolation with the sets of patterns, the constant and linear terms are dropped giving

$$\psi(x) \cong -1.6\frac{\pi}{4}\ (0.06x^2 + x^3) \qquad (19)$$

The sets of patterns which bracket $\psi(x)$ are Figs. 5, 6, 10, and 11. For each set, the pattern for 20 db taper is obtained. These patterns are then interpolated for A = 0.06 and $\delta = 1.6\frac{\pi}{4}$. The resulting synthesized pattern is given by the dotted line in Fig. 23. The interpolated pattern based on equation (19) has maximum intensity at u = 0.50 and is plotted in Fig. 23 after adding the tilt caused by the linear term of equation (18) which gives maximum intensity at u = 0.50 - 1.26 = -0.76. To this must be added the tilt caused by the linear term of equation (17) which gives $u_o = -0.76 - 14.54 = -15.3$. The gain factor, $\mathcal{G}$, obtained by interpolation of the values given on the sets of patterns is 0.83.

The solid line in Fig. 23 is the pattern obtained by the method of calculation outlined in section II after subtracting equation (17) from the phase distribution given in Fig. 22. The direction of maximum intensity is u = -0.90; therefore, $u_o = -0.90 - 14.54 = -15.4$. The calculated gain factor is 0.80. Some of the differences between the calculated and the synthesized patterns, such as the filled in nulls and lower gain factor, can be attributed to the slight asymmetry in f(x), and others to the higher order terms in $\psi(x)$.

V. Conclusion

Sets of patterns have been presented which are useful in obtaining semi-quantitative information regarding the characteristics of the pattern expected from an aperture antenna having a non-linear phase distribution. This method has been found helpful in predicting the suitability of aperture distributions for particular applications.

Acknowledgements

The author wishes to express his appreciation to Mr. G.E. Feiker for his encouragement during the progress of this work and to Mr. R. Habermann, Jr. who supervised the work on the automatic punch card machines.

This method of analysis was first used in conjunction with a Signal Corps contract.

References

1. R.C. Spencer and P.M. Austin, "Tables and Methods of Calculation for Lines Sources", RL Report No. 762-2; March 30, 1946.

2. H.T. Friis and W.D. Lewis, "Radar Antennas", Bell System Tech. Jour., Vol. 26, pp 219-247; April, 1947.

3. S. Silver, "Microwave Antenna Theory and Design", McGraw-Hill Book Co., p. 183.

4. S. Silver, op. cit., p. 178 and 183.

5. J. Ruze, "Wide-Angle Metal-Plate Optics", Proc IRE, Vol. 38, p. 57; January, 1950.

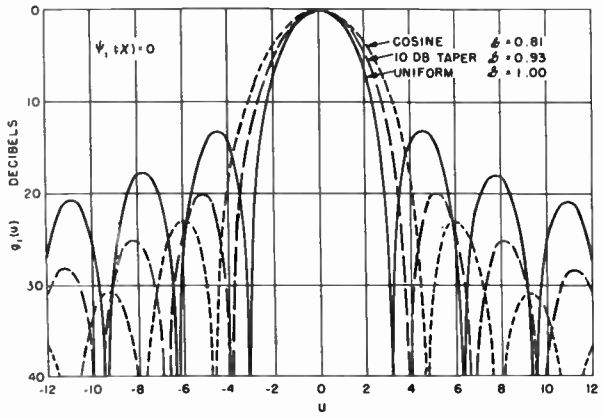


Fig. 1

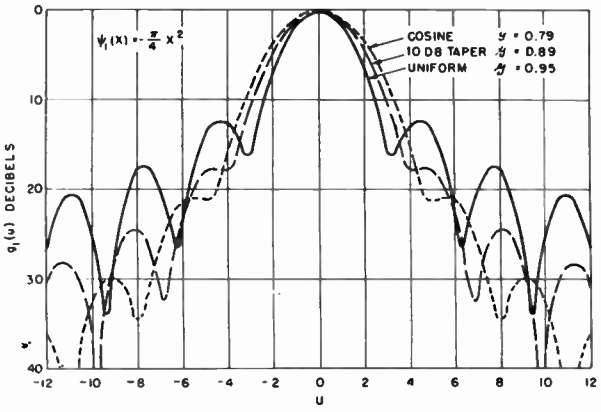


Fig. 2

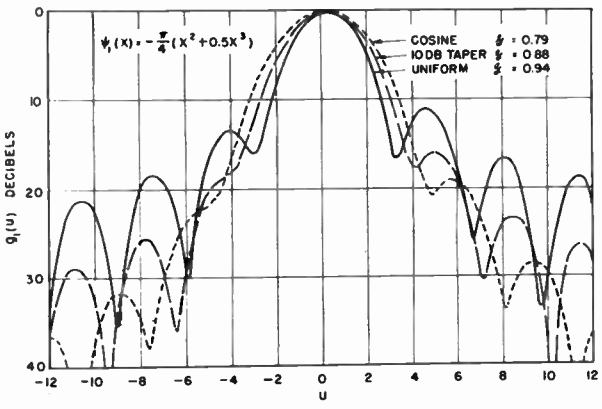


Fig. 3

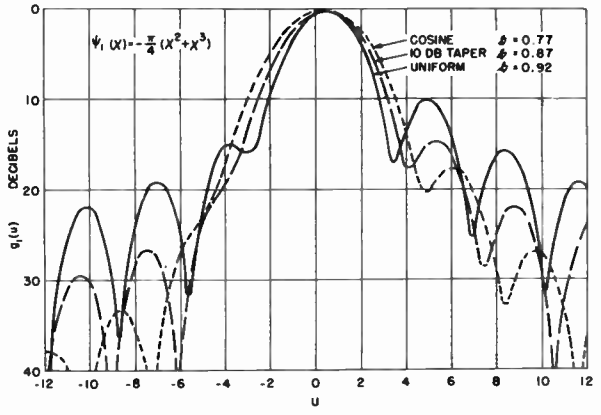


Fig. 4

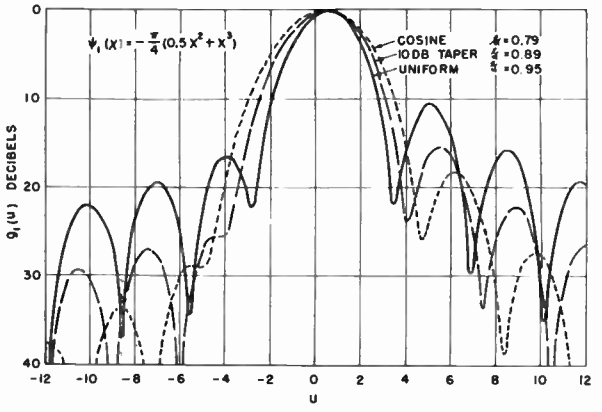


Fig. 5

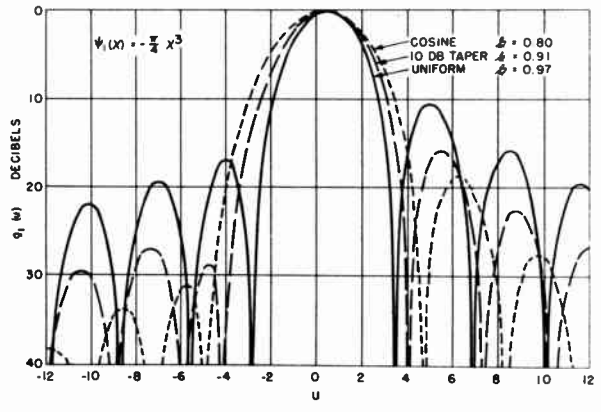


Fig. 6

Magnitude patterns.

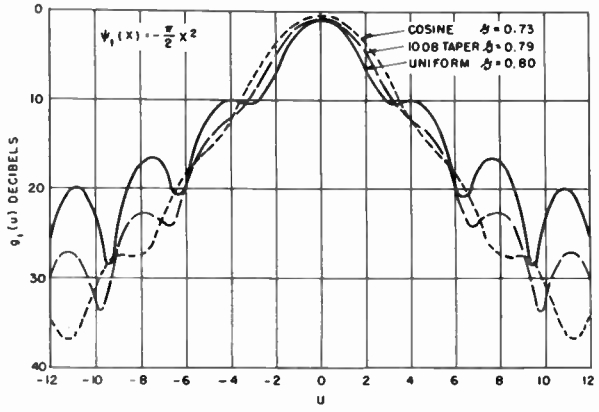


Fig. 7

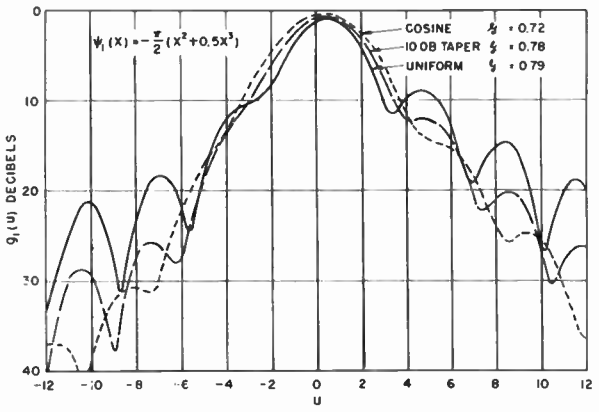


Fig. 8

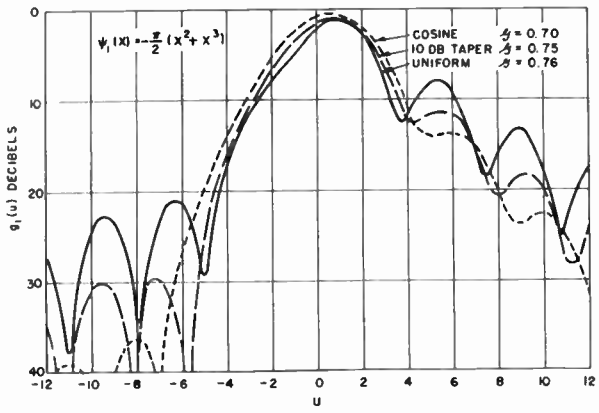


Fig. 9

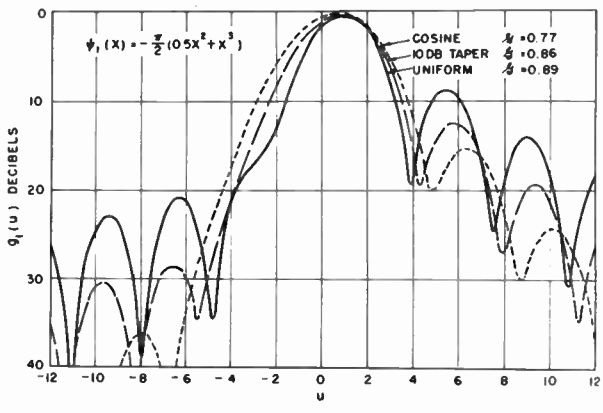


Fig. 10

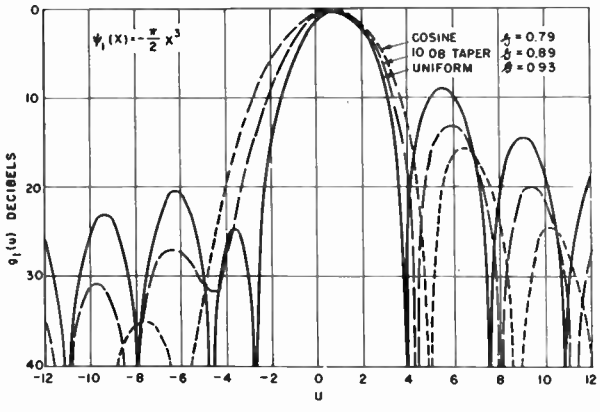


Fig. 11

Magnitude patterns.

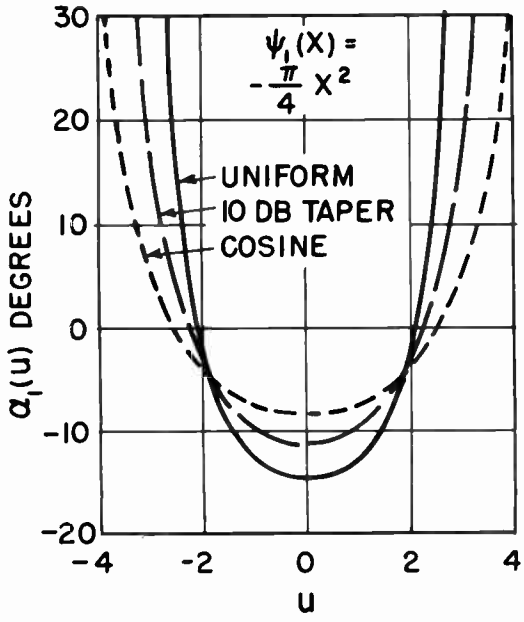


Fig. 12

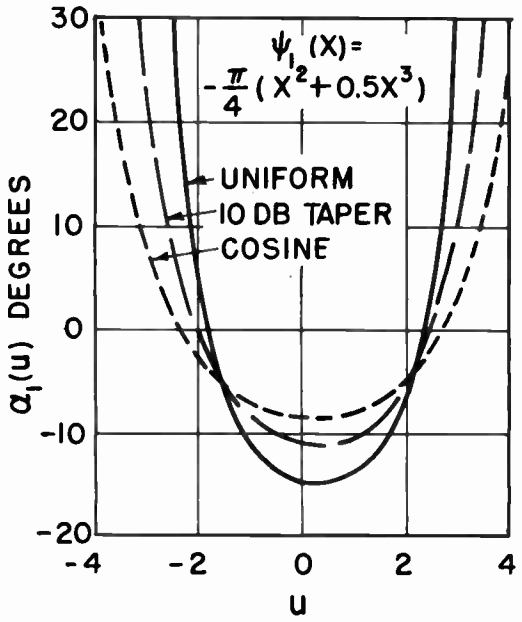


Fig. 13

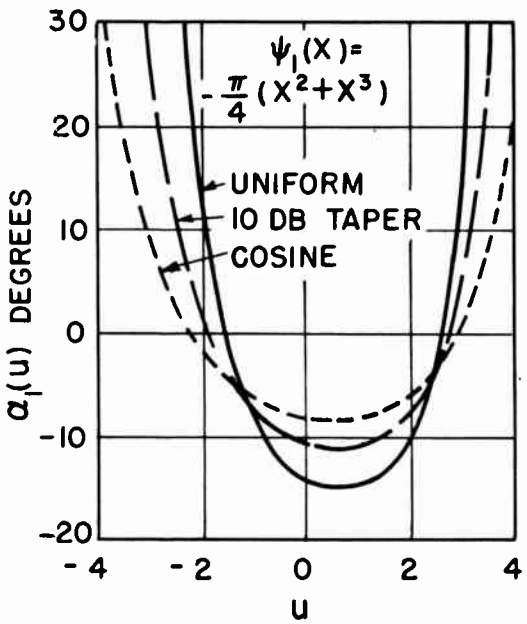


Fig. 14

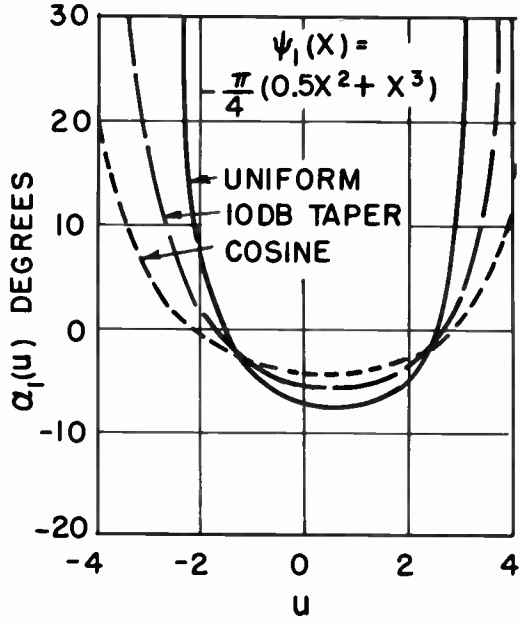


Fig. 15

Phase patterns.

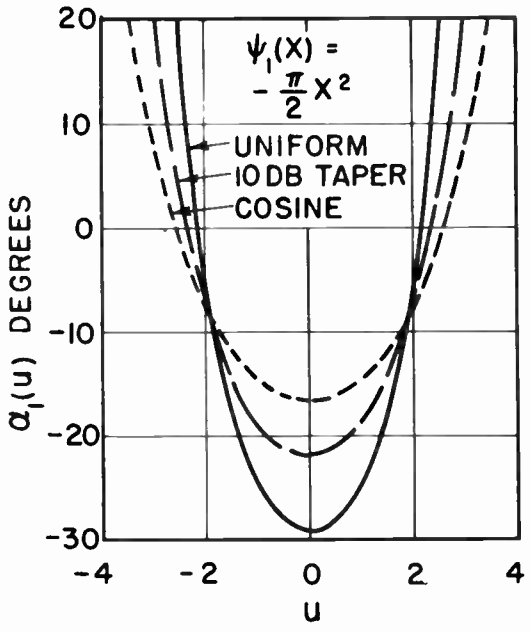


Fig. 16

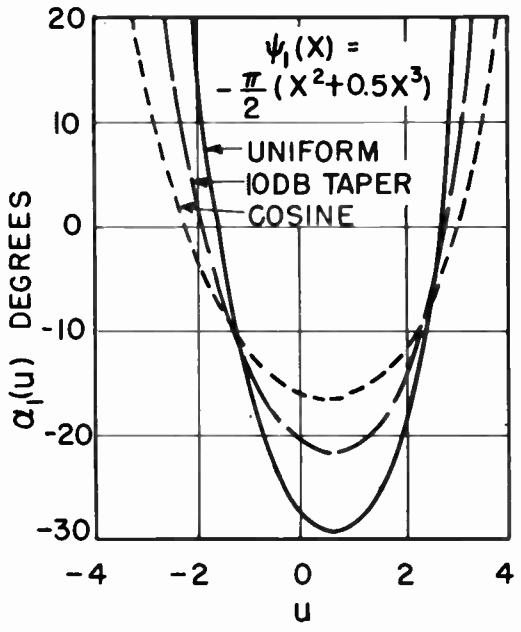


Fig. 17

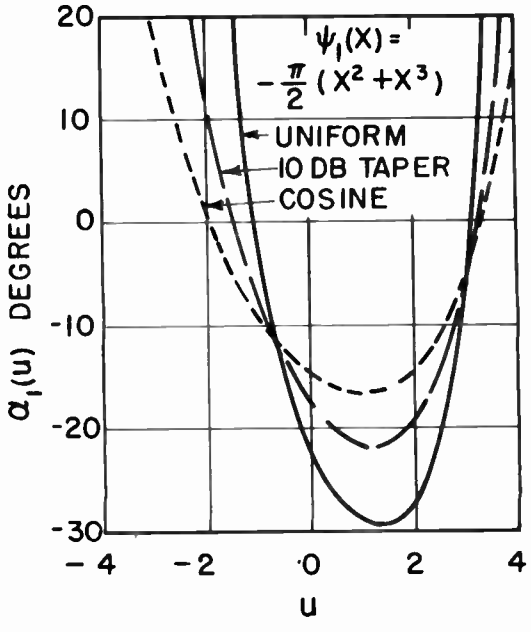


Fig. 18

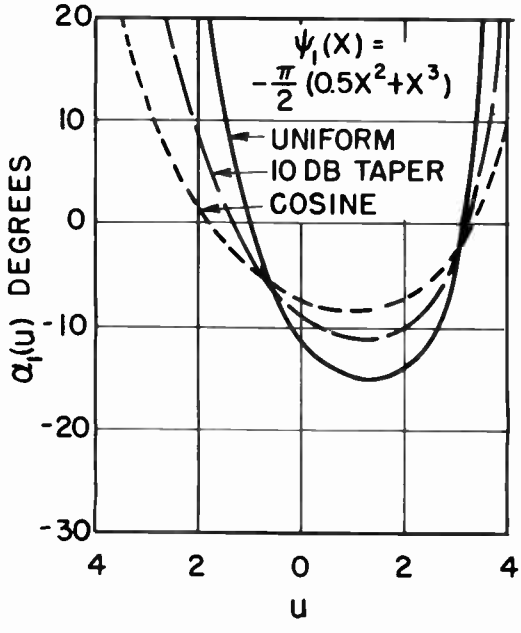


Fig. 19

Phase patterns.

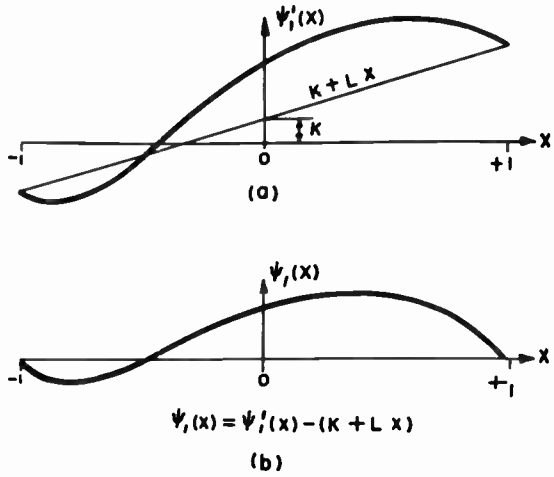


Fig. 20 - Phase distributions.

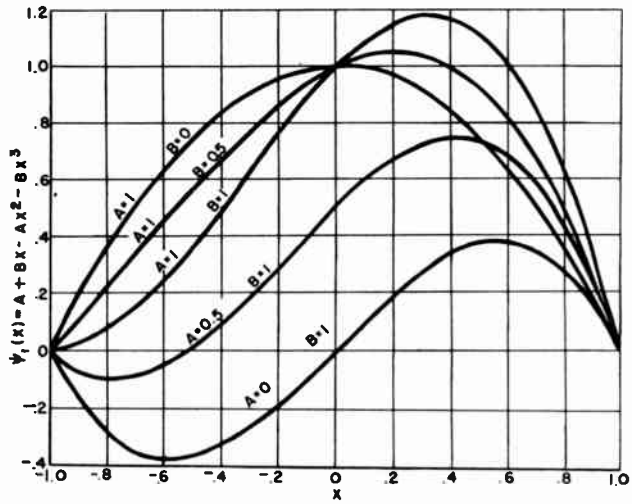


Fig. 21 - Family of curves.

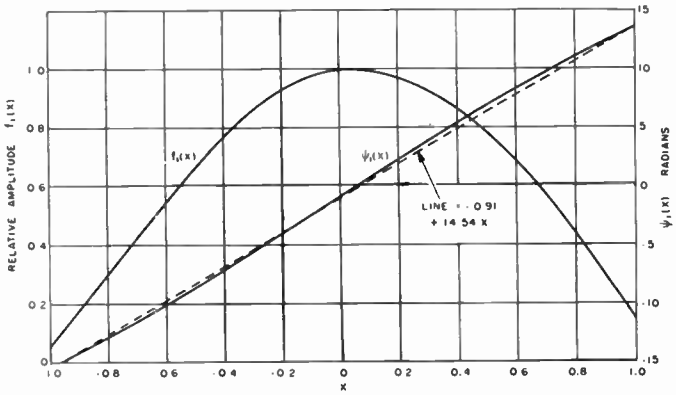


Fig. 22 - Aperture distribution.

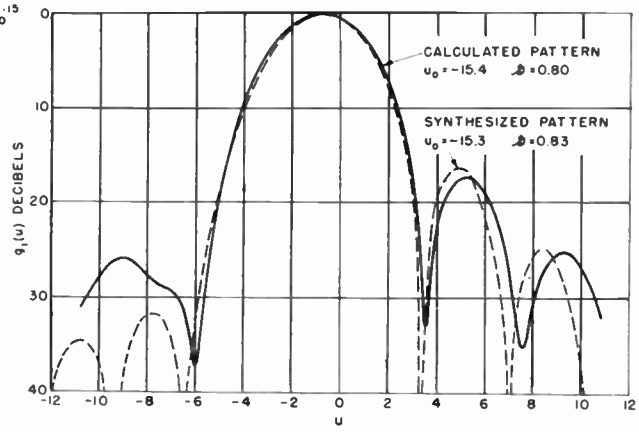


Fig. 23 - Magnitude patterns.

CORRUGATED SURFACE ANTENNAS

M. J. Ehrlich
L. L. Newkirk

Research and Development Laboratories
Hughes Aircraft Company
Culver City, California

I. Abstract

The study of the properties of corrugated surface and dielectric sheet antennas was limited in this paper to the ascertainment and development of means for control of the parameters which govern the efficiency of the feed in launching the dominant surface wave onto the corrugated or dielectric surface. A qualitative description of the mechanism of the feed interaction with the radiating surface was hypothesized, and strong experimental evidence for support of the hypothesis was obtained. Two very successful feeds were then developed; the first was an aperture feed, and the second was an end-fire slot array embedded in the corrugated surface.

II. Introduction

Present-day and contemplated airframes cannot carry radar antennas which introduce drag. The antennas must, therefore, be flush-mounted. In most instances, the problem has been solved to date by excision of a section of the skin and substitution of a plastic radome for the section removed. The radar antenna is then placed within the interior of the airframe and radiates out through the radome. However, airframe space requirements are such that the feasibility of the use of surface antennas must be investigated. A surface antenna is defined as an antenna which (a) uses a portion of the skin as the radiator, or (b) has the radiating elements embedded in the skin. A trapped surface wave radiator is an example of the former, and a two-dimensional slot array is an example of the latter.

Conceptually, it seems possible to have slot arrays on curved surfaces such that fan, pencil or shaped beams are obtained. Another approach may be the use of a corrugated or dielectric clad surface which supports a trapped wave. The surface would act as the radiator and it would require but a simple thin line source feed. The two most fundamental questions to be answered in connection with the latter antennas are (1) can beam shaping be obtained by use of a curved trapped surface wave radiator, and (2) can the wave be properly launched onto the surface?

It is obvious that the first question is more basic as regards future antenna research, but it contains within it the assumption that a satisfactory solution has been obtained to the second question. This is evident when it is appreciated that the total radiation pattern of a real trapped surface-wave antenna is composed of the vector sum of the field from the surface and the field radiated by the feed. Thus, to perform any accurate empirical studies of beam shaping by change of surface curvature and/or phase velocity, the radiated field should not contain any appreciable contribution from the feed acting as a radiating element.

Therefore, a beam-shaping study presupposes a feed in which the dominant feed mode is almost wholly transformed to the dominant TM surface mode and launched onto the guiding surface. In summation, then, the factors which led to the development of the feeds for trapped surface wave antennas discussed in this study were:

(1) Zero drag requirements of high-speed aircraft,
(2) Limited space available on the airframe for the conventional reflector antenna together with its mounted feed and radome,
(3) Desirability of exploration of feasibility of external continuous surface-distributions as antennas for furnishing shaped beams,
(4) Necessity for obtaining a satisfactory feed before the beam-shaping problem could be investigated.

III. Previous Investigations

In recent years, corrugated structures have been applied successfully to coaxial and hollow waveguides for the treatment of problems connected with traveling wave tubes[1,2,3] and linear accelerators[4,5,6,7]. The low wave velocities and special band-pass characteristics of these guides permit their use as phase control devices of smaller size and lighter weight than either dielectric field or ridge waveguides.

Of equal importance with such work has been the accompanying investigation of corrugated structures of a single surface only, considered as a single surface waveguide, or as a corrugated surface antenna. Initial theoretical studies of such structures dealt with their wave-propagating properties rather than with their radiating properties. Cutler[8], assuming only the fundamental mode of propagation along a corrugated surface and a transmission line mode within the corrugations, matched the average impedance at their common boundary to derive approximate expressions for the propagation constant as a function of the surface parameters, for both plane and cylindrical corrugated surfaces. Somewhat more accurate expressions were derived by Walkinshaw[9], who acknowledged the presence of the higher order modes and matched the external and

internal fields at their common boundary by a Fourier analysis, assuming on the basis of quasi-static approximations an appropriate form of electric field over the mouth of each corrugation. Goubau[10] analyzed electromagnetic wave propagation along a cylindrical conductor and found that the extent of the field may be controlled either by a modification of the surface into a periodic structure, or by coating the surface with a dielectric layer. Rotman[11], in a series of experiments concerning wave propagation along planar and cylindrical corrugated surfaces, found the propagation constants and the field components to be in good agreement with the theory.

The most recent advances in the application of courrugated surfaces to wave propagation and radiation problems have been made at Stanford Research Institute. There, research in this field began with a theoretical study of wave propagation along a doubly infinite corrugated surface[12]. By means of the variational principle, values were derived for the propagation constant as a function of the surface parameters. Experimental measurements of the propagation constant were in close agreement with the theory. Also, probe measurements were made of the electric field over the corrugated surface and were found to be in good agreement with the assumed field used in the theoretical analysis. A later study was made concerning the problem of reflections from a semi-infinite plane terminating an infinitely wide corrugated surface[13]. The results showed that end reflections are not serious since they may be made small by a proper choice of corrugation dimensions at the termination end. The propagation constant of a corrugated surface of finite width, b, was also determined. For 'b' equal to several wavelengths or greater, the propagation constant was found to approximate closely that of an infinitely wide corrugated surface.

IV. Radiation Pattern of a Corrugated Surface Array and Feed

On completion of study of wave propagation along a corrugated surface, the endfire radiation properties of a corrugated surface of finite length were investigated[13]. By means of a Green's function method, an analysis of the problem was made by Lucke, which led to the following approximate formulation. For the far zone magnetic field, H_Z,

$$H_Z \propto \int_0^L E_x(x) e^{\pm ikx \cos\theta}\, dx. \qquad (1)$$

The coordinate system in which this formula is valid is shown in Figure 1. H_Z is the amplitude of the far zone magnetic field, $E_x(x)$ is the tangential electric field along the corrugated surface of length, d, and k is the free space propagation constant.

On the basis of results obtained in the previous studies of wave propagation along a corrugated surface, $E_x(x)$ was taken to be the field component of a traveling mode of the corrugated surface type, $e^{-i\beta x}$, where β is the propagation constant along the corrugated surface in the x direction. Upon substitution of this expression for $E_x(x)$ into Eq. (1), subsequent integration and then solution for $F(\theta)$, one obtains the amplitude of the far zone,

$$F(\theta) \propto \frac{\sin\left(\frac{\beta L - kL\cos\theta}{2}\right)}{\left(\frac{\beta L - kL\cos\theta}{2}\right)} \qquad (2)$$

This expression is identical to that obtained in conventional end-fire theory[14] in which a wave of propagation constant β is assumed to be traveling along an array of length, L.

These first theoretical studies of the nature of the radiation pattern to be expected from a corrugated surface antenna assumed a distribution of fields over the surface. Implicit in such an assumption is the statement that the feed does not radiate into space or alter the surface distribution, but acts wholly as a mode transformer between the dominant feed mode and the propagating dominant surface mode. Subsequent measurements[15,16] of surface wave antenna radiation patterns revealed major deviations from the theoretical values. Not only were the measured deviations large, but measurement of the patterns of several surface arrays revealed a structure common to the perturbations in the several instances. These characteristic perturbations are:

(1) A reduced beam width,
(2) A minimum at approximately the theoretical half-power point,
(3) A very high first side lobe which actually may be larger than the main lobe,
(4) Large higher-order side lobes whose magnitude and position differ widely from the theoretical values.

Experimental studies[15,16] showed that these perturbations were a result of constructive and destructive interference between the radiation field arising from the surface array distribution and the radiation field of the feed structure. Calculations were made by Lucke[15] of the theoretical radiation pattern of an array composed of a single slot feed and a corrugated surface when the slot was also acting as a radiating element. This pattern thus calculated had perturbations similar to those found by measurement. It is clear that in the light of this work the large deviations between the measured radiation pattern and the theoretical pattern are to be regarded as arising from the constructive and destructive interference occuring between the field radiated by the surface and the field radiated by the feed. The total pattern may then be represented as the vector sum of the two fields. The total field is given by

$$E_T(\theta) = Af_1(\theta) + f_2(\theta)\, e^{j(kd\cos\theta - \varphi)} \qquad (3)$$

where

$f_1(\theta)$ = the space factor of the feed,

$f_2(\theta)$ = the space factor of the surface, as given in Eq. (2),

d = spacing between the centers of phase of the two elements,

φ = phase difference between the two elements,

$k = \frac{2\pi}{\lambda}$, where λ is the free space wavelength,

A = amplitude ratio of the two radiators in the common direction of maximum radiation of both elements.

The geometry of the array is shown in Figure 2.

V. Theoretical Description of the Two Source Array

The two source model was hypothesized to explain the strong interference effects manifested in the measured radiation patterns. However, not only can the gross interference effects be inferred from this model, but the detailed structure of the pattern and the parameter governing the relative phases and amplitudes of the two sources can be obtained as well.

A feed and surface array somewhat different from the slot and surface array considered by Lucke will be assumed in the following discussion.

The model considered has the corrugated surface or dielectric sheet placed well within the feed and is of zero corrugation depth or sheet thickness at the edge in the direction of the incident wave. The depth or sheet thickness is increased gradually so that at the exit aperture — real or equivalent — of the feed, the surface has the depth corresponding to the desired phase velocity of the surface array. Several such feed configurations are illustrated in Figures 3a, b, and c.

The difference between the array configurations discussed by Lucke and those considered here is the introduction of a tapered section of corrugated surface or dielectric sheet within the feed. The tapered section provides a region wherein partial transformation can be made from the dominant feed mode to the dominant surface mode. The premise is made that the longer and more slowly tapered the internal transformer, the greater the amount of mode conversion. The empirical support for this is the performance of equivalent transformers in effecting mode conversion within waveguide structures. On the basis of this premise, it follows that the relative amplitudes of the excitations of the feed and surface as separate radiators may be altered by variation of the dimension of the transformer section.

The foregoing discussion has tacitly assumed the essential separability of the fields exciting the feed aperture and the dominant traveling surface modes entering into the surface array radiation. The concept of separability is fundamental to the analysis of the two source array. The earlier studies at Stanford Research Institute had clearly demonstrated that the total radiation pattern was actually the sum of the two separate patterns, i.e., the feed radiating as an aperture antenna and the surface array as an end-fire antenna. The projected use of the internal mode transformer stems directly from these empirical data as it is argued that even though the relative amplitudes are altered by a change in transformer dimension, the two sources still radiate separately. Each source still radiates with its own spatial pattern. A change of relative phase and interphase center spacing, however, may be introduced by the change in transformer length.

Clearly, if the separability argument is allowed, the use of a mode transformer follows as a reasonable consequence. Then in the limit it is seen that as the transformer is lengthened indefinitely, a complete mode conversion would take place and the feed would cease to radiate. This cessation would be manifested by the absence of the pattern perturbations previously noted. Any differences that might be noted would arise from the diffraction of the exponentially decaying electric component of the dominant traveling surface mode at the feed aperture edge. If the height of the aperture were sufficiently great, say several wavelengths, such diffraction effects would be negligible for the range of surface phase velocities encountered in practice.

These conclusions are readily apparent for as the transformer is extended indefinitely in the -x direction and the upper surface of the feed is removed to a point sufficiently remote from the surface, the configuration is that of a semi-infinite corrugated surface with its source at infinity, and, consequently, no perturbation results.

VI. Experimental Work

The experiment that was designed to verify the hypothesis regarding the mode conversion transformer and its concomitant assumption of field separability was simple in concept. A corrugated surface was constructed so that its length, L, external to the aperture, A, of a doubly flared horn feed could be kept constant while the length, t, of an internal mode transformer was varied. The radiated pattern in the far zone was then measured for a range of transformer lengths. A line drawing of the scheme is shown in Figure 4.

The external length, L, of the array is kept constant and the internal length, t, which acts as the mode transformer is varied. The changes are achieved by cutting back the bottom wall of the horn toward the throat so as to have a longer transition section. The tapered segment, B, is moved back the same amount, and a segment, s, of untapered corrugated surface is fitted into place to fill up the gap. The bottom waveguide wall is

smoothly tapered over a length, b. In this manner, by combining segments of corrugated surface together to form the middle untapered portion, the external length is held fixed.

An experiment also was performed which used a dielectric sheet as the trapped surface wave radiator. A series of dielectric sheets were constructed with a fixed external length, L, and different lengths of transformer, t. The taper was kept constant as in the corrugated surface taper section. The dielectric sheet was then fitted into the mouth of a horn and the far zone radiation pattern was measured for a range of transformer length.

VII. Experimental Details

A. Experimental Apparatus

1. R-F Power Source and Related Equipment. A Varian X-13 klystron operating at a frequency of 9840 mc was used as the source of the r-f power. A calibrated wave meter and linear amplifier were used for determining the output frequency, and a crystal detector and linear amplifier for monitoring the power output.

2. Design of the Test Corrugated Surface. The dimensions of the corrugated surface which would support the propagation of a trapped mode at X-band were determined by use of the relation[12]

$$\cot kh = \frac{a}{a+w}\left[\left(\frac{\beta}{k}\right)^2 - 1\right]^{-\frac{1}{2}} \qquad (4)$$

where k is the free space propagation constant, h is the corrugation depth, a is the space between adjacent corrugations, w is the corrugation width, and β is the propagation constant along the corrugated surface. The formula is accurate to within a few percent for $M > 19$, where M is the number of corrugations per corrugated surface wavelength.

It was found that the corrugations were required to be 1/32 inch wide, 1/8 inch deep, and spaced 1/16 inch center to center. The surface was milled from brass sheet stock to form a corrugated surface 2-3/8 inches wide and 10 inches long. Over a length of 1-3/8 inches at each end of the corrugated surface, the corrugation depths were tapered uniformly from 1/8 inch to zero. When completed, the corrugated surface was embedded in a small brass ground plane, 12 inches by 24 inches in surface dimensions, and the complete configuration is shown in Figures 5, 6, and 7.

3. Aperture Feed. Feed horns. Four brass feed horns of different aperture size were made, being 2-3/8 inches wide in the H-plane as seen in Figure 8. The aperture height and flare angle of each horn were adjusted to produce a phase variation across its aperture equal to 0.115λ, where λ is the free space wavelength, equal to 1.2 inches at 9840 mc. The aperture heights, A, were 0.73λ, 1.14λ, 1.79λ and 2.88λ, and the respective flare angles were 17.5°, 11.2°, 7.4° and 4.5°. Both sides of each horn were equipped with thick brass flanges through which screws were passed to hold the horn in position on the corrugated surface so as to maintain continuous electrical contact over the entire extent of the surface in common.

An RG-52U rectangular X-band waveguide, open at one end, was also used as a feed to excite the corrugated surface.

4. Radiation Pattern Measurements. In order that the experimental boundary conditions conform to those assumed in the theory[13], it was necessary to obtain pattern data within the confines of a large ground plane. Accordingly, the corrugated surface and small brass ground plane were flush-mounted in a large ground plane to which was affixed a long (six feet) rotating arm as seen in Figures 9 and 10. The ground plane was four feet by seven feet in surface dimensions and made of 24 ST aluminum sheet fastened to a plywood surface. A probe was fastened securely to the rotating arm, thereby permitting vertical movement of the probe along an arc of a circle whose center coincided with the midpoint of the axis formed by the intersection of the ground plane with the aperture plane of the feed system. The probe was located 54 inches from the axis, a distance approximately equal to L^2/λ, where L is the length of the corrugated surface external to the feed system, and λ is the free space wavelength. The probe consisted of a receiving horn (0.84λ by 0.84λ aperture), bolometer, tuning stub, coaxial cable and linear amplifier. A similar experiment was performed using a dielectric sheet polyrod fed by a horn. The dielectric sheet with its tapered transformer section is shown in Figure 11 and is seen inserted in the horn in Figure 12. A series of sheets were made with varying transformer lengths and fixed external lengths.

VIII. Experimental Results

The changes in the radiation pattern as a function of the transformer length of the corrugated surface horn combination, whose geometry is shown in Figure 13, are clearly seen in Figures 14, 15, and 16.

Each aperture-surface array shows the same behavior. When the transformer is of zero length, i.e., the corrugated surface begins immediately without the aperture, the pattern is greatly distorted and possesses the characteristic perturbations previously noted. As the transformer section is introduced into the feed, the measured pattern tends more and more to the theoretical pattern of the surface by itself. This is shown for three aperture sizes; namely, $A = 0.73\lambda$, 1.14λ and 2.88λ.

In Figure 14, the pattern changes are most marked. The pattern perturbations are almost wholly removed by an increase in the transformer length. This absence of perturbation is also shown in Figure 17, which is the pattern of a surface array fed by a rectangular waveguide with a transformer 2λ in length. The total pattern is compared to the theoretical pattern of the surface alone and the agreement is quite good. The geometry of the waveguide feed is seen in Figure 18.

The differences between the measured and the

theoretical radiation patterns arise from several sources. First, of course, is the fact that the theoretical pattern is computed on the basis of the field distribution that exists on a finite section of the infinite corrugated surface array, while the actual surface distribution is tapered over the extent of the surface radiator and is not uniform as assumed in the theory. The second is the slight radiation from the feed which produces some beam sharpening in this particular instance. The phenomenum of importance is the absence of the pattern perturbations mentioned in the earlier portion of this study which arise from interference between the radiation fields of the strongly radiating feed and surface radiator.

In the study of the dielectric sheet, the same pattern changes as a function of transformer length are seen in Figure 19, although not by any means as clearly marked as in Figure 14, which showed the changes produced with the corrugated surface. However, the decrease of large lobes at large angles, as well as a decrease in structure complexity, mark the diminution of the interference effects which were produced by a reduction in the direct radiation from the feed. It is apparent that a transformer section greater than 1λ in length is required as well as a much thicker sheet for the particular horn-dielectric sheet combination used in the experiment. The purpose of the dielectric sheet experiment was to confirm the similar behavior of the two types of trapped surface wave radiators and to demonstrate that the concept of mode conversion applied to both surfaces.

The second experiment designed to verify the assumption of the separability of the aperture field distribution and surface field distribution was also quite simple in concept. If the assumption is made that the corrugated surface or dielectric sheet is terminated so as to produce negligible reflection, then variations of the surface length, L, external to the feed aperture, A, will not affect the energy distribution in the two modes. The external surface array has a theoretical far zone radiation pattern given by

$$F(\theta) \propto \frac{\sin\left(\frac{\beta L - kL\cos\theta}{2}\right)}{\frac{\beta L - kL\cos\theta}{2}} \qquad (2)$$

where $\beta = \frac{2\pi}{\lambda g}$, $k = \frac{2\pi}{\lambda}$, λg = surface wavelength,

λ = free space wavelength.

The beamwidth of an end-fire array is inversely proportional to the square root of the length expressed in wavelengths. The equi-phase surfaces surrounding the terminated corrugated surface or dielectric sheet are such that the center of radiation appears to be somewhere near the end of the radiator while the phase center of the feed is some slight distance back from the aperture. Thus, the distance between the center of phase of the two elements is comparable to the length of the surface external to the feed. The array and its equivalent were shown in Figure 2.

A small change in 'L', the length of the external array, will not affect A, $f_1(\theta)$, and will only affect $f_2(\theta)$ by a very small amount. A curve of the half-power beamwidth of $f_2(\theta)$ is given in Figure 20 as a function of L/λ, and from this, the change in θ_{hp} for a variation Δ in L is readily ascertained. An expression for the half-power beamwidth, θ_{hp}, of a uniformly illuminated end-fire array is given[14] as

$$\frac{\theta_{hp}}{4} = \sin\frac{-10.47}{(L/\lambda)^{1/2}} \qquad (5)$$

For an array of length, $L \geq 4\lambda$, a change of length of 0.5λ produces at most a four percent change in beamwidth. A change in the external length will then to a good approximation only alter 'd' the interphase center spacing. In this manner, separate control is obtained over the relative phase of the two radiators.

The external length, L, was altered by insertion of sections of the corrugated surface shown in Figure 7, and in the dielectric sheet study by insertion of different sheets.

The variations in pattern for a corrugated surface length increase of one wavelength are seen in Figures 21 and 22 for two surface lengths and aperture sizes. It is shown in both that as the interphase center spacing, d, is increased, the interference effects are increased in magnitude and occur closer to the main beam. The same phenomena occurred in the dielectric sheet study and are shown in Figures 19d and 23.

The shift in angular position gives support to the assumption that the interphase center spacing, d, is large and also demonstrates the shift in relative phase arising from the change in array length. The shift in d for a one wavelength change in the array length, L, was computed for the case of a dielectric sheet and was found to be approximately one wavelength.

A least squares analysis was performed which used the data of Figure 19d and the values of the amplitude ratio, relative phase φ, and interphase center spacing, d, were determined. These values were then used to compute the pattern of Figure 23a for the new external length, and fairly good agreement was obtained between the computed and measured values.

It now has been amply demonstrated that the distributions are separable and that the relative amplitudes and phases can be adjusted over a wide range. It is, of course, evident that as the length of the transformer sections is altered, the principal effect is to change the amplitude ratio, A, of the two sources, but in addition, the change in the transformer section length also changes the phase. This phase change is readily restored by a relatively slight alteration in the length, L, of the external array.

The relative phase and magnitude of the two radiators can now be independently adjusted. Some speculations may be made regarding the pattern improvements to be gained by direct utilization of the feed radiation in combination with the surface radiation.

Consider one of the composite aperture feed surface arrays. If the interphase center distance, d, is large compared to wavelength, relatively slight changes in frequency or dimension will markedly alter the total radiation pattern. The combination of a large surface array and a small aperture or the opposite combination may show frequency sensitivity if the broader beamwidth element radiates an appreciable amount. The array which combines a relatively short end-fire array of about three to six wavelengths in length with a feed whose aperture is about one to two wavelengths in extent possesses some interesting properties when the relative excitations are properly adjusted. An array of this type would be insensitive to frequency or dimensional variation. In such an array, there exists the possibility of obtaining a reduced beamwidth compared to the beamwidth of the surface or aperture alone. The beamwidth reduction is also accompanied by some slight improvement in side-lobe level.

Although the combination array possibilities are of interest, the most important facts are the separability of the radiated fields and the control over the relative phase and magnitude of the equivalent generators which feed the two elements.

IX. Development of a Flush-Mount Feed for a Corrugated Surface

The results of the mode transformer experiment showed conclusively that the utilization of the transformer section furnished a highly satisfactory aperture feed. The extent of the vertical aperture could be made quite small as is seen in Figures 17 and 18, where RG-52U rectangular waveguide was used as the feed structure. It seems reasonable to expect that the surface could be slightly curved in the vicinity of the feed in order that the aperture be recessed to provide the flush surface required for an airframe. However, a true flush mount feed would be preferred and development of such a feed has been undertaken.

Upon examination of the propagating modes in the -y direction in the corrugations, as shown in Figure 1, it is seen that for $w < \lambda/2$ the only propagating mode is the TEM mode with $\bar{E} = |E|\,\bar{i}_x$ and $\bar{H} = |H|\,\bar{i}_z$, where E_x and H_z are related by the factor $\sqrt{\frac{\mu_1}{\varepsilon_1}}$ and $|E_x| = \sqrt{\frac{\mu_1}{\varepsilon_1}}\,|H_z|$. The corrugations may be considered as a short-circuited section of a parallel plate transmission line of infinite extent in the $\pm Z$ directions.

A slot radiator embedded in the bottom of the corrugation would radiate into the parallel plate medium formed by the corrugation walls and thus serve to establish the desired field configuration at the surface, albeit, with somewhat altered phase relation between E and H. If the feed is composed of slots embedded in the corrugated surface and spaced $\lambda g/4$ apart along the surface, the feed creates an end-fire array at the upper face of the surface when the slots are fed by generators of proper phase and amplitude. Such generators may be realized by feeding the slot array from below by means of a transmission line with internal loading of magnitude sufficient to reduce the phase velocity to that of the corrugated surface. If, for the moment, questions of interslot mutual coupling, disturbed surface phase relations and finite source length in the infinite corrugation are ignored, a simplified description of the operation of the proposed feed is readily obtained. The feed is shown in schematic form in Figure 24. The dielectric loading is adjusted so that $\lambda g_1 = \lambda g_2$, thus ensuring the end-fire conditions of the equivalent sources at the corrugation apertures. The sources are 90° apart in time phase owing to the method of internal feeding and 90° apart in path length along the outer surface. The contributions cancel in the backward direction and add in the forward direction. It was expected that the creation of the TEM field distribution in the corrugations by the slot feed would serve to launch the proper TM surface mode and trap practically all the energy on the surface.

The construction details and complete end-fire slot array feed are shown in Figures 25, 26, and 27. The choice of the TEM mode transmission line feed for the slots was dictated by the requirement that the slots be parallel and that a reasonable degree of control of the slot excitation be readily obtained.

The excitation of slot radiators by strip transmission lines has been developed in detail[17] and some empirical data is available[18]. In brief, the characteristics of a transmission slot fed by a strip TEM line are (1) when cut in the outer face of the strip TEM line, the slot behaves as a series element and (2) the coupling may be varied over a wide range, $1 < \frac{R_{slot}}{Z_o} < 0.1$ either by rotation of the slot when it is centered with respect to the center of the stripline or by transverse displacement of the center of the slot with respect to the center of the stripline. When the slot is perpendicular to the axis of the stripline, the slots are aligned parallel in the feed so they may be embedded in the bottom of the corrugations as shown in the figures.

The amplitude of the individual element excitation is controlled by the amount of transverse displacement of the slot center with respect to the center of the strip. All the slots are cut at an angle of 90° with respect to the center conductor of the feed line.

The variation in transverse spacing from element to element that is required to compensate for the excitation change which arises from mutual coupling is seen in Figure 26. The amount of change required is computed by means of an incremental resistance argument similar to the

incremental conductance method of Watson[19].

The feed line is loaded with dielectric in order to match the velocity of propagation to that of the corrugated surface. A strip transmission line, open at one end was used as a feed system for the slot array. The outer conductor of the line, 0.525 inch wide to prevent the propagation of the TE_{10} mode, was made by inserting two brass shim plates, each 0.400 inch deep and 0.187 inch wide, into RG-52U rectangular waveguide. The center conductor, of phosphor-bronze, was 0.250 inch wide and 0.020 inch thick, and was supported by a lucite strip 0.187 inch deep and 0.063 inch wide. The purpose of the lucite strip was to reduce the phase velocity in the transmission line to the value corresponding to propagation along the corrugated surface. The phase velocity was $v_{ph} = 0.94$ C, and the amount of loading required to obtain this was not great.

The performance of the slot array feed is good and is shown in Figure 28. The pattern of the surface plus feed matches quite closely the theoretical pattern of the surface alone. The perturbations that characterize the patterns measured in previous investigations are almost wholly removed. The slight beam sharpening and higher first side lobe probably are due to a small amount of feed radiation. This can be further reduced by a slight increase in the number of slots in the feed. The radiation from 150° to 180° is caused by reflections from the matched load termination on the feed line and an incorrect value of excitation for the first and last slots of the feed. The VSWR that is obtained for the termination was 1.2 at the spot design frequency of 9840 mc. The back lobe is reduced to -14 db when correction is made for this termination reflection. The slot excitations are computed by means of an incremental resistance method[19]. Inherent in this method is the assumption that the slot is embedded in an infinite linear array of identical slots.

This condition is not met even approximately for the end slots, and their excitations should be somewhat increased. It is felt that such a change would further reduce the back lobe.

The slot array corrugated surface combination was insensitive to a frequency change of 90 mc, and this is shown in Figure 29.

X. Conclusions

The concept of separability of the field distribution of the two source model has been amply demonstrated by the experimental results. The mode transformer, which is a logical outgrowth of this concept, furnishes a means to obtain a wide range of values of the ratio of the relative amplitude of the two sources.

In both types of trapped surface wave radiators, the relative amplitude may be varied over a wide range by proper choice of the transformer length. The exact values of the upper and lower bounds that may be arbitrarily achieved are difficult to evaluate, but examination of Figures 14, 15, 16 and 19 would give about 0.15 for the lower bound and about 3.0 for the upper bound obtained with the transformers and aperture feeds used in this experiment. The accompanying change in relative phase which is manifested by the variation of the height of the main lobe as a function of transformer length is readily changed by a slight change in the external length of the surface wave radiator.

The vertical aperture and the transformer combined constitute a simple , easily-constructed and highly satisfactory feed for trapped surface wave radiators. It is simple to extend it to a line source aperture to serve as a feed for a large two-dimensional surface. While not wholly flush-mounted, it can be made with a small vertical aperture and still give satisfactory performance. The embedded end-fire slot array is also an excellent feed for a corrugated surface as it simultaneously satisfies the requirements of producing minimum pattern perturbation and of being flush-mounted.

However, some problems may be encountered in the extension of this type of feed for large surface arrays. A more thorough study of the impedance characteristics of the slots when embedded in the corrugations would have to be made before a line feed could be built.

The simultaneous control of the relative amplitudes and phases is of interest for use in connection with beam sharpening of end-fire arrays of about one to five wavelengths in extent. In these instances, by allowing the small aperture feed to radiate, and by adjusting the relative amplitudes, a decrease of about 25 percent in the end-fire array beamwidth can be gained while some side lobe reduction is effected at the same time. These pattern improvements are purchased at the cost of adding a small, i.e., about a half-wavelength vertical aperture to the end-fire array.

XI. Recommendations

It is felt that the next step is to proceed with the problem of determining the radiation characteristics of trapped surface wave radiators embedded on curved surfaces. The objective of the proposed research is the evaluation of the feasibility of obtaining predetermined shaped beams from the surface radiators. The principal curved surfaces to be investigated would be the right circular cylinder and the elliptic cylinder, as these correspond quite closely to typical airframe contours.

XII. Acknowledgements

The work described in this report has been supported by the Airborne Antenna Division, Air Force Cambridge Research Center, through Contract No. AF 19(604)-262 FP.

XIII. References

1. Third Quarterly Progress Report, "Traveling

Wave Tube Project," Stanford University, July, 1947 through September, 1947.
2. A. W. Lines, G. R. Nicoll, A. M. Woodward, "Some Properties of Corrugated Waveguides," British TRE Report No. T2114, Telecommunications Research Establishment, Gt. Malvern, Worc., Ministry of Supply, London, August, 1948.
3. J. R. Pierce, "Traveling Wave Tubes," Bell Sys. Tech. Journ., Vol. 29, p. 189, April, 1950.
4. E. L. Chu and W. W. Hansen, "The Theory of Disk-loaded Waveguides," J. App. Physics, Vol. 18, p.996, November, 1947.
5. R. B. R. Shersby-Harrie, "Traveling Wave Linear Accelerator," Proc. Phys. Soc., Vol. 61, part 3, p. 255, September, 1948.
6. W. Walkinshaw, Theoretical Design of Linear Accelerator for Electrons," Proc. Phys. Soc., Vol.61, part 3, p. 246, September, 1948.
7. L. Brillouin, "Waveguides for Slow Waves," J. App. Phys., Vol. 19, p. 1093, November, 1948.
8. C. C. Cutler, "Electromagnetic Waves Guided by Corrugated Conducting Surfaces," Bell Telephone Labs., Inc., Whippany, N. J., October, 1941.
9. Walkinshaw, loc. cit.
10. G. Groubau, "Surface Waves and Their Application to Transmission Lines," J. App. Phys., Vol.21, p. 1119, November, 1950.
11. W. Rotman, "A Study of Single Surface Corrugated Guides," Proc. IRE, Vol. 39, No 8, p.952, 1951.
12. First and Second Quarterly Progress Reports, July, 1949-January, 1950, "Ridge and Corrugated Antenna Studies," Stanford Research Institute, Stanford, California.
13. Third Quarterly Progress Report, January, 1950-April, 1950. SRI
14. S. Schelkunoff, "Electromagnetic Waves," D. Van Nostrand Co., p. 342-300, New York, N. Y. 1943.
15. Fourth Quarterly Progress Report, SRI.
16. Fifth Quarterly Progress Report, SRI.
17. N. A. Begovich, "Recent Developments in Slot Radiators," paper presented at IRE West Coast Convention, Long Beach, California, September 15, 1950.
18. E. Strumwasser, J. Short, R. Stegen, J. Miller, "Slot Study in a Rectangular TEM Transmission Line," HAC TM No. 265, January, 1952.
19. Silver, S. "Microwave Antenna Theory and Design," M97 Radiation Lab. Serv., Vol. 12, pp.296-298, McGraw-Hill, New York, 1949.

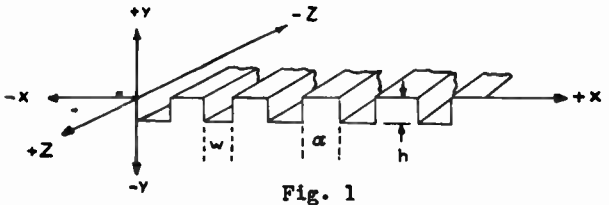


Fig. 1
Coördinate system in which Equation (1) is valid.

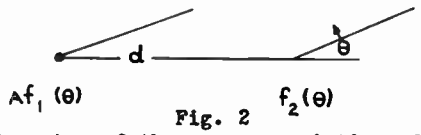


Fig. 2
Geometry of the array consisting of a single slot feed and a corrugated surface.

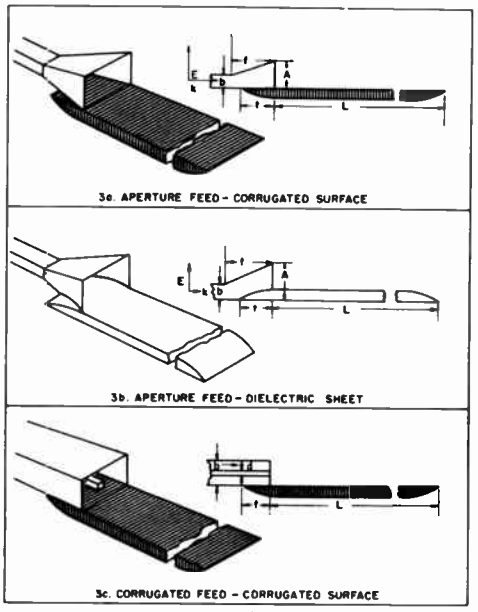


Fig. 3

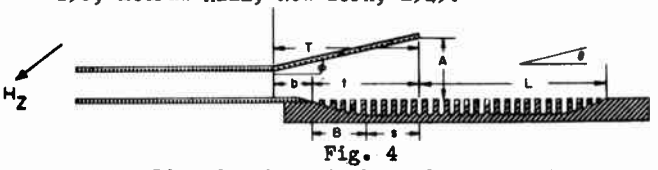


Fig. 4
Line drawing of the mode conversion transformer assuming field separability.

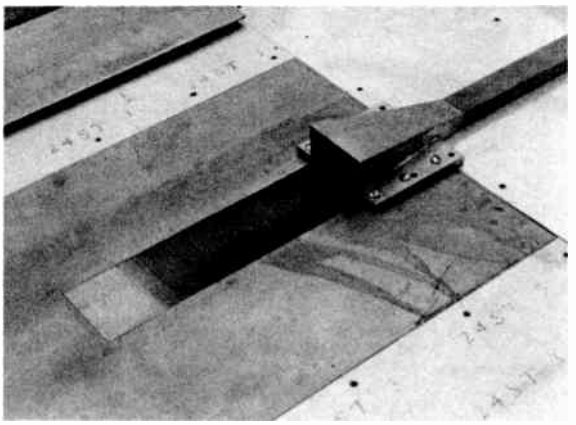


Fig. 5
Photograph showing complete configuration of the test corrugated surface.

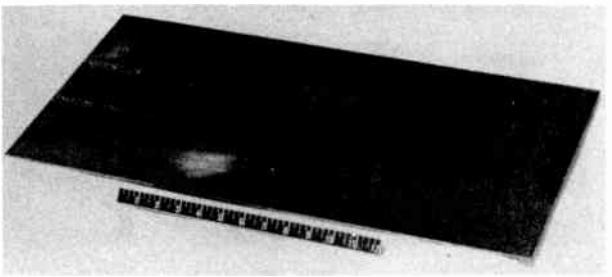


Fig. 6
Photograph showing complete configuration of the test corrugated surface.

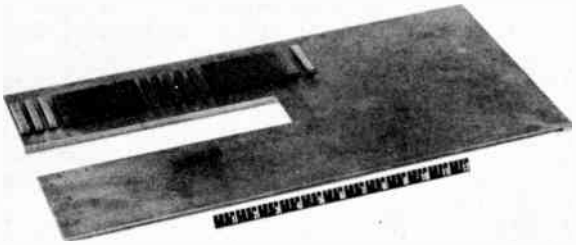


Fig. 7
Photograph showing complete configuration of the test corrugated surface.

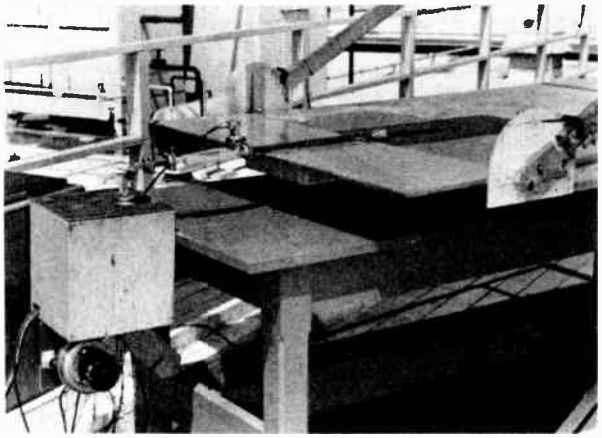


Fig. 10
Photograph of the corrugated surface and small brass ground plane flush-mounted in a large ground plane to which is attached a rotating arm.

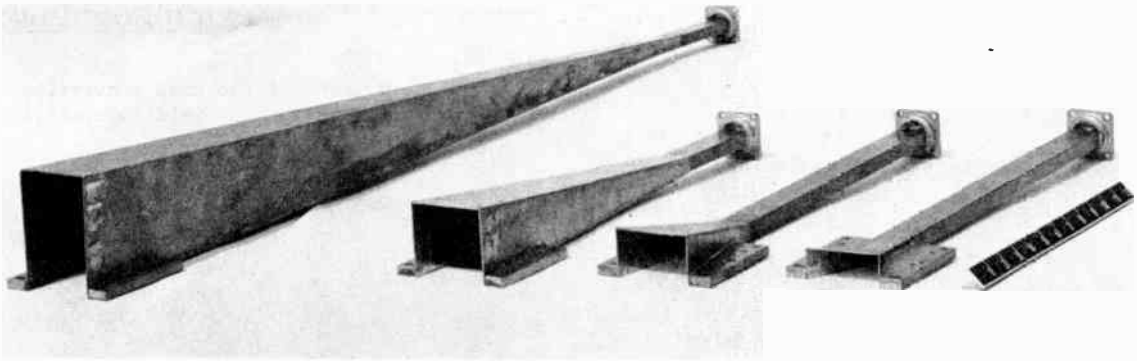


Fig. 8
Photograph of four brass feed horns of different aperture sizes.

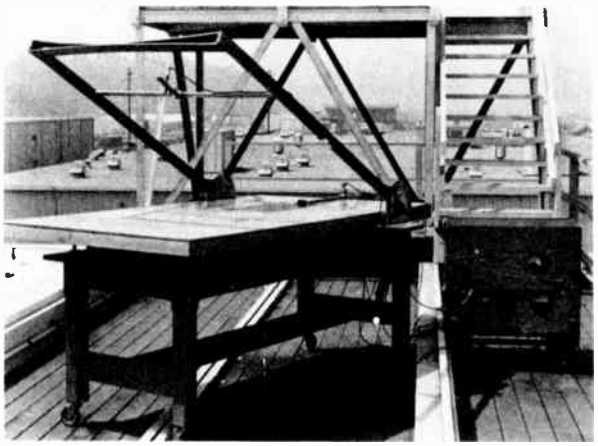


Fig. 9
Photograph of the corrugated surface and small brass ground plane flush-mounted in a large ground plane to which is attached a rotating arm.

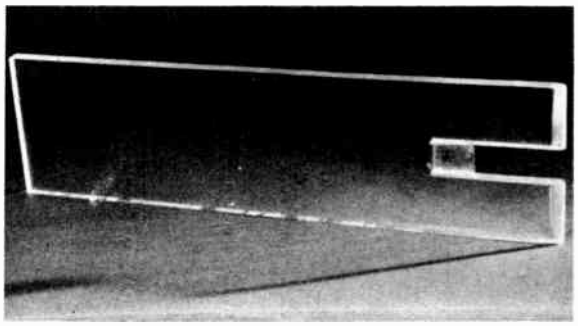


Fig. 11
Dielectric sheet and tapered transformer section.

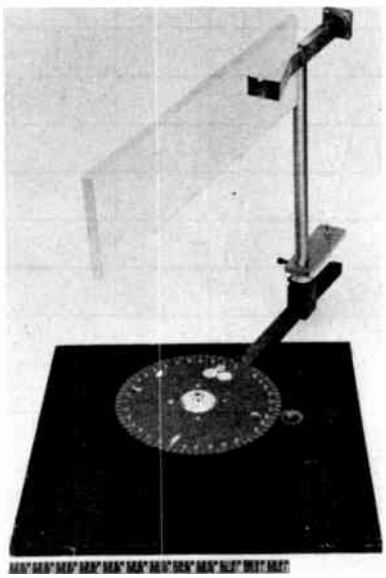


Fig. 12 (left)
Photograph of dielectric sheet and tapered transformer section inserted in horn.

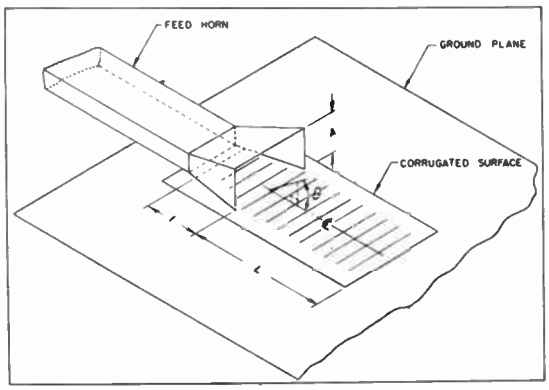


Fig. 13
Diagram showing experimental arrangement of feed horn and corrugated surface. Not to scale.

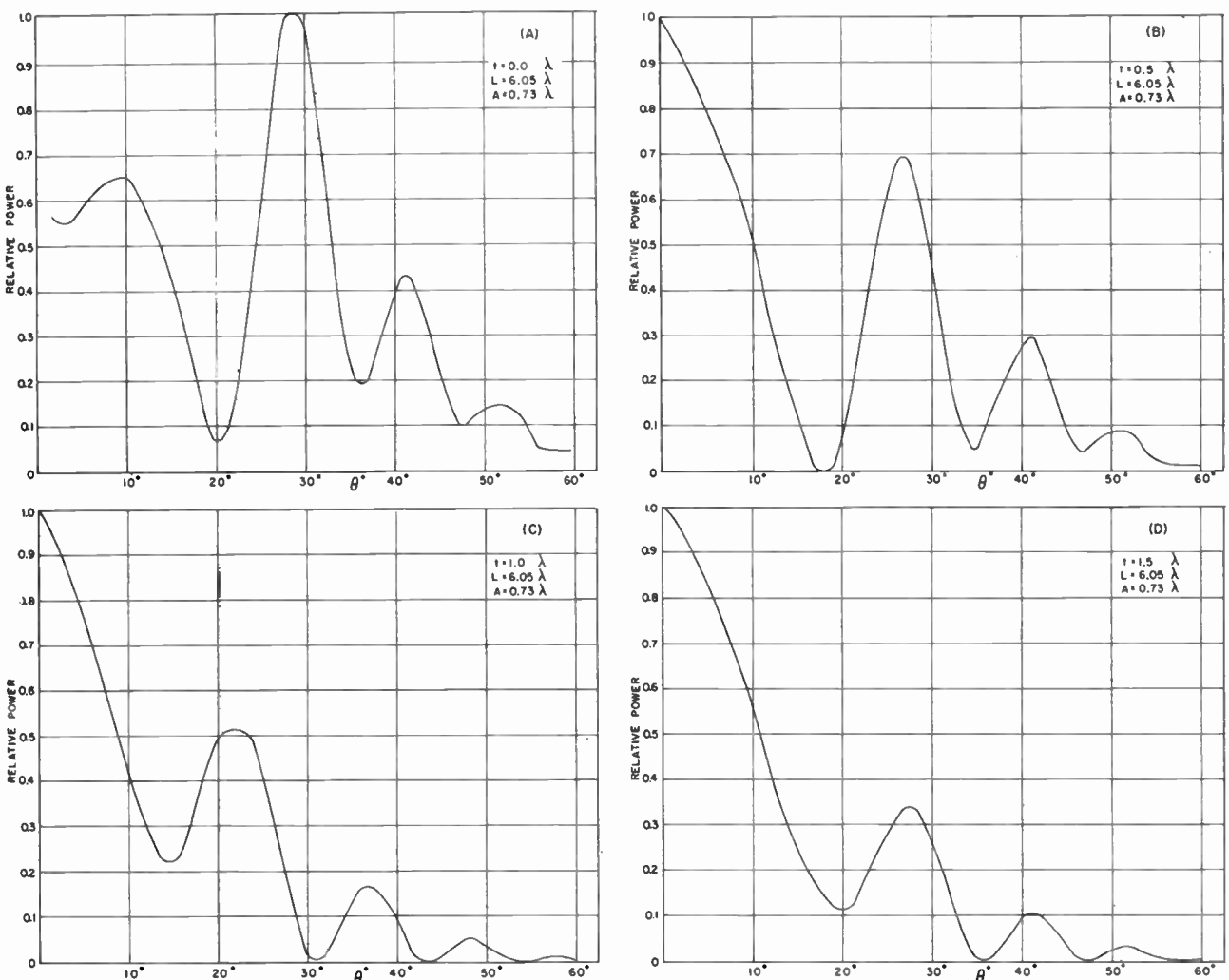


Fig. 14
Composite radiation patterns of corrugated surface excited by horn as a function of t and A for A = 0.73λ.

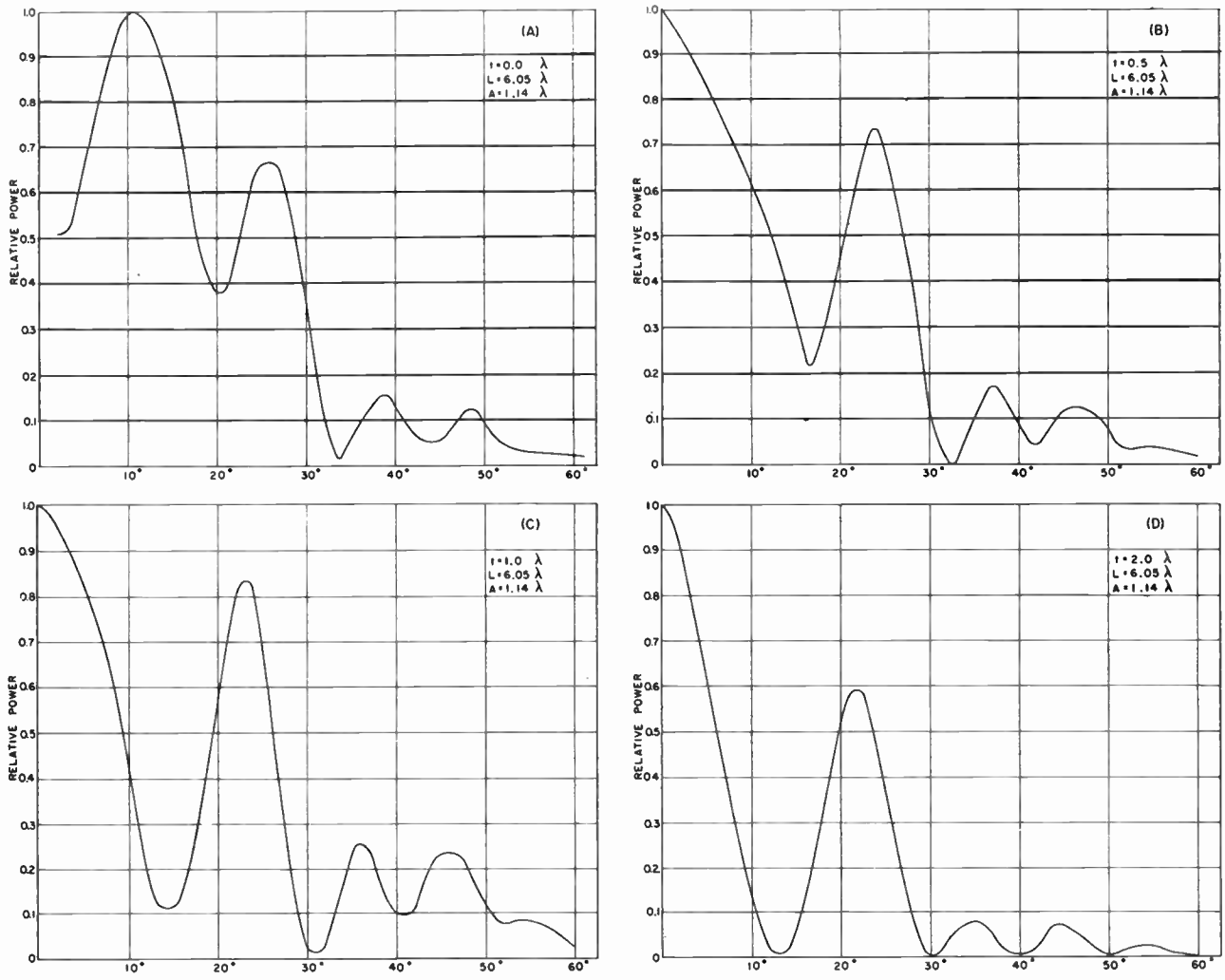


Fig. 15
Composite radiation patterns of corrugated surface excited by horn as a function of t and A for A = 1.14λ.

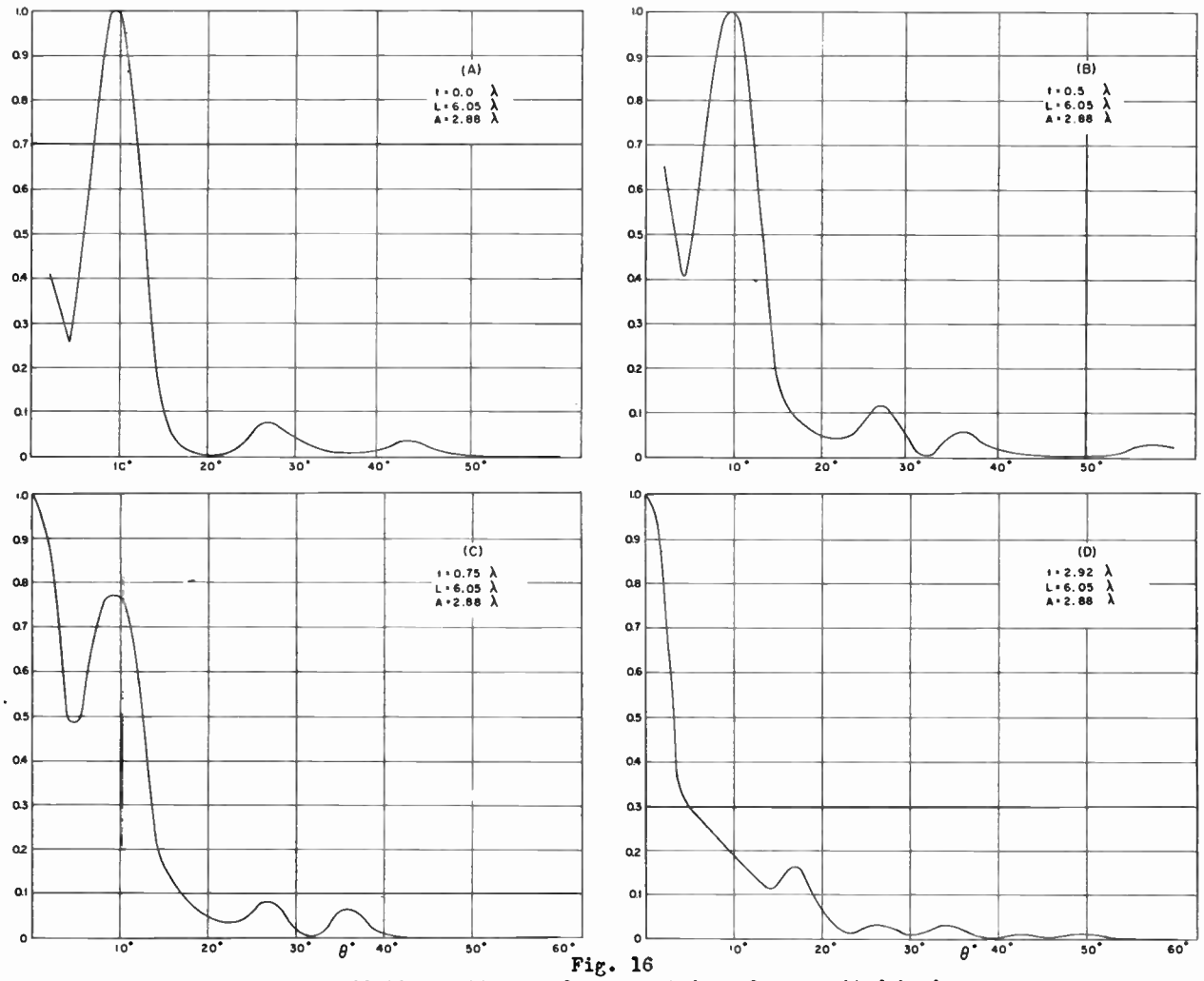


Fig. 16
Composite radiation patterns of corrugated surface excited by horn as a function of t and A for A = 2.88λ.

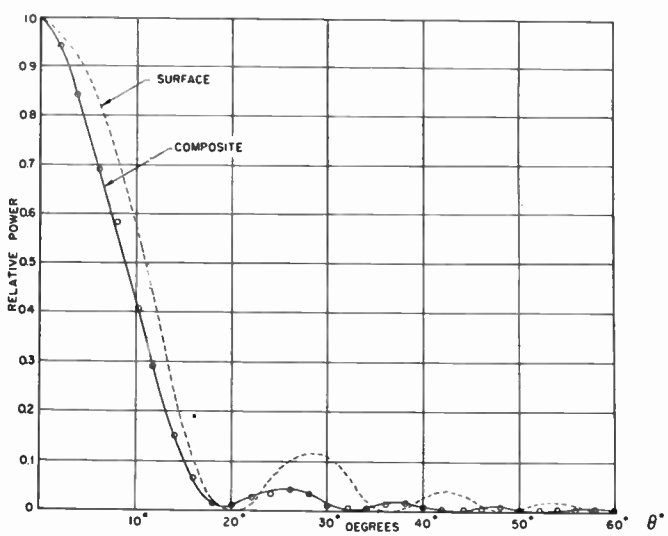


Fig. 17
Composite radiation pattern for corrugated surface of length 7.33λ excited by wave guide compared with theoretical pattern for corrugated surface.

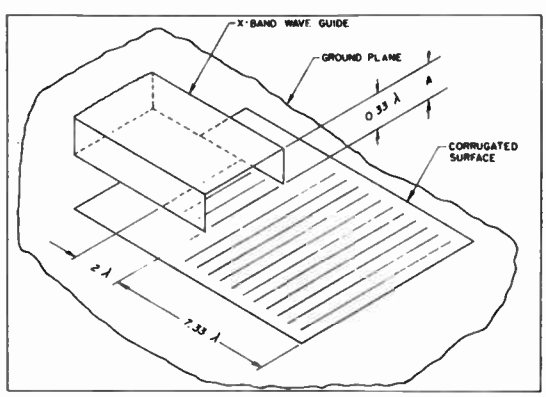


Fig. 18
Diagram showing arrangement of X-band waveguide feed and corrugated surface. Not to scale.

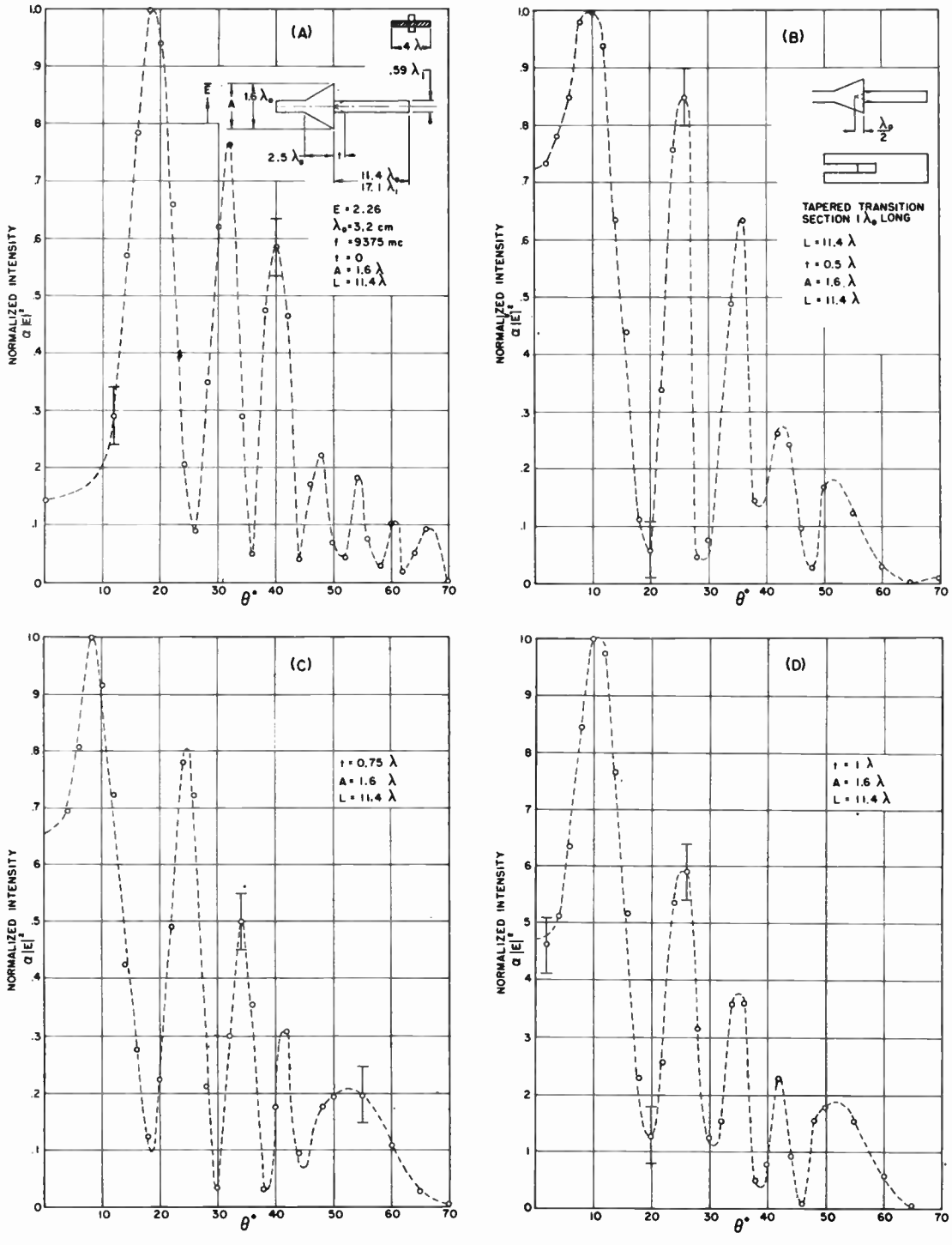


Fig. 19
Patterns of dielectric sheet as a funotion of transformer length.

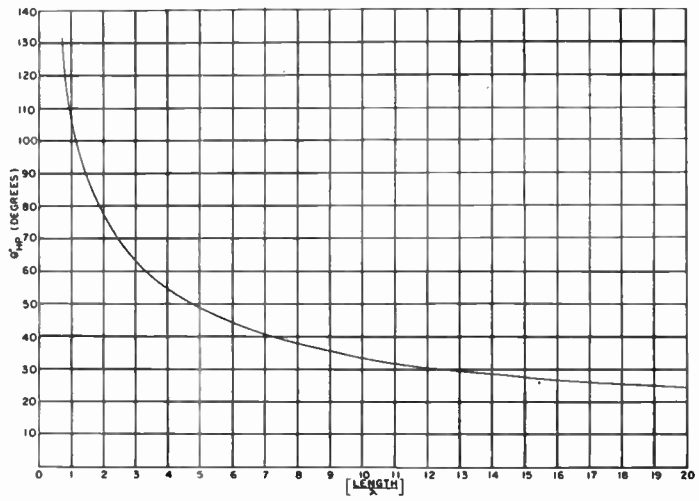


Fig. 20
Half-power beamwidth of a uniformly illuminated end-fire array as a function of array length

$$\frac{\theta}{2} = \sin^{-1}\frac{0.47}{\sqrt{\frac{L}{\lambda}}}$$

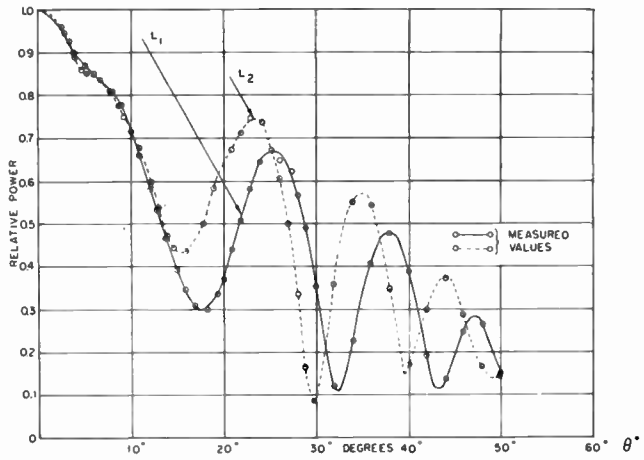


Fig. 21
Composite radiation pattern as a function of external length of corrugated surface.
$t = 1.17\lambda$
$A = 0.73\lambda$
$L_1 = 6.73\lambda$
$L_2 = 7.33\lambda$

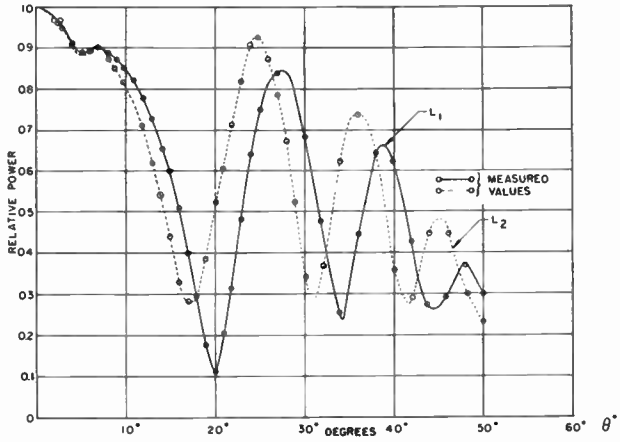


Fig. 22
Composite radiation pattern as a function of external length of corrugated surface.
$t = 0.54\lambda$
$A = 0.43\lambda$
$L_1 = 6.73\lambda$
$L_2 = 7.33\lambda$

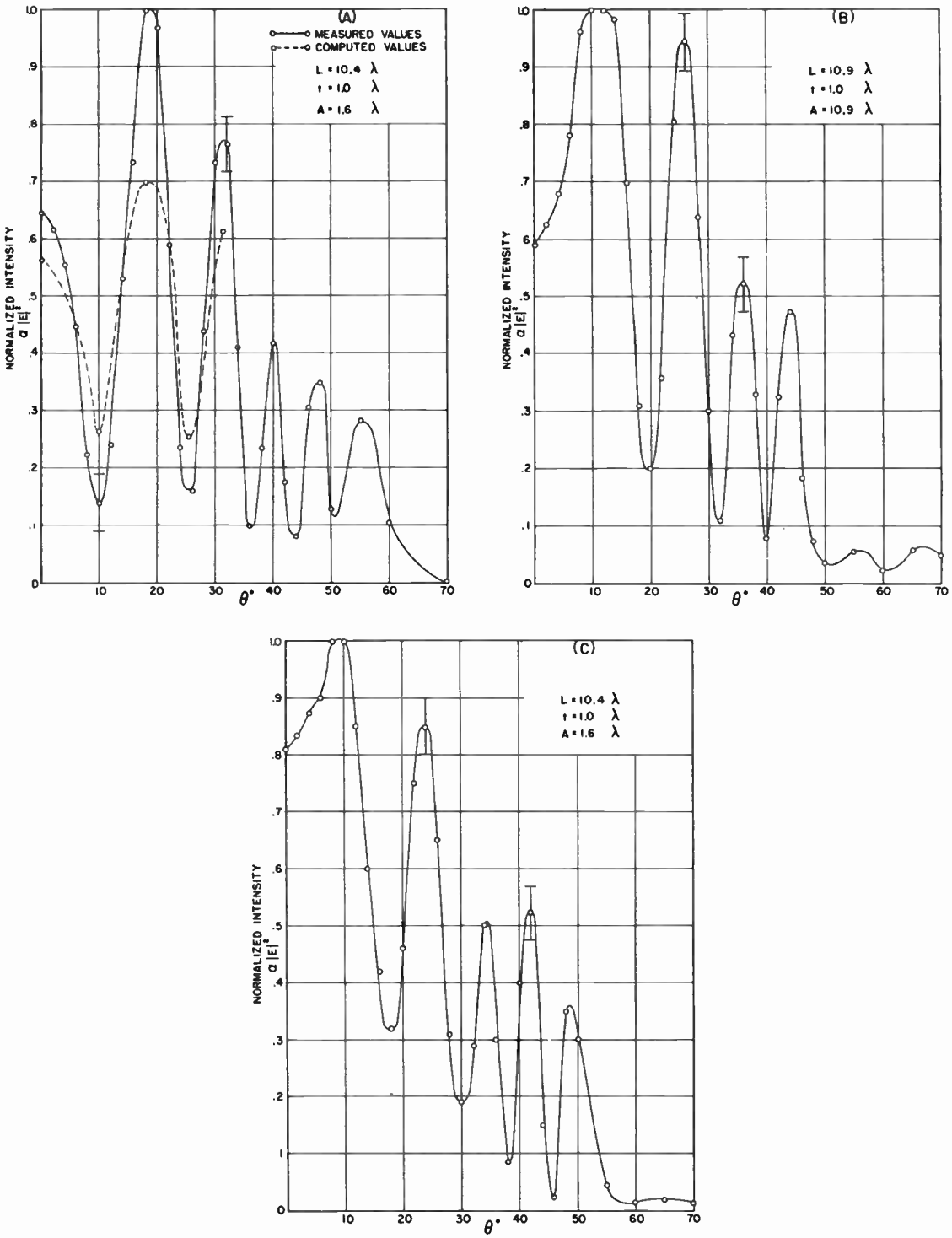


Fig. 23
Variations in pattern for a dielectric sheet.

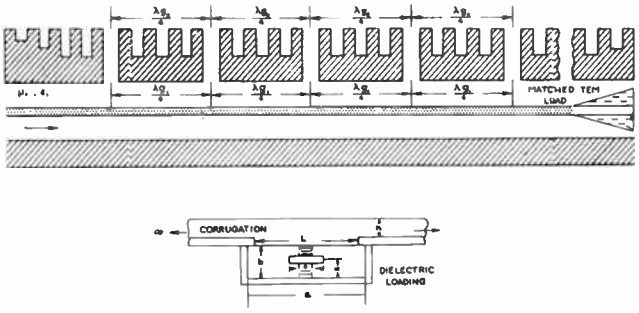


Fig. 24
Cross section of end-fire slot array embedded in corrugated surface and fed by a strip transmission line.

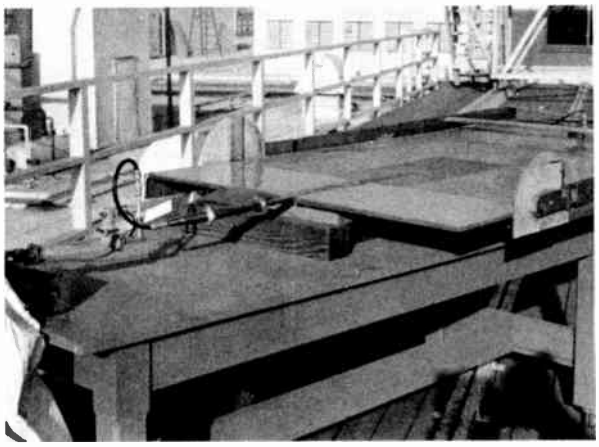


Fig. 27
Photograph of a corrugated surface fed by an end-fire slot array.

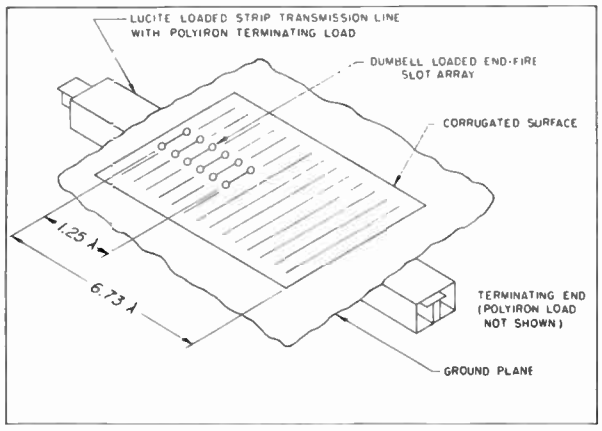


Fig. 25
Schematic diagram showing experimental arrangement of slot array feed and corrugated feed. Not to scale.

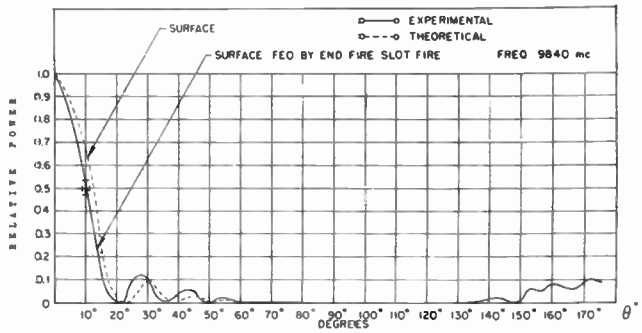


Fig. 28
Measured radiation pattern of corrugated surface of length 6.73λ excited by six-element end-fire slot array compared with theoretical pattern for corrugated surface.

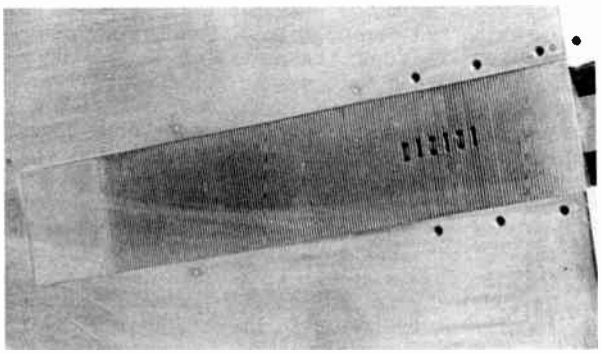


Fig. 26
Variation in transverse spacing from element to element that is required to compensate for the excitation change which arises from mutual coupling.

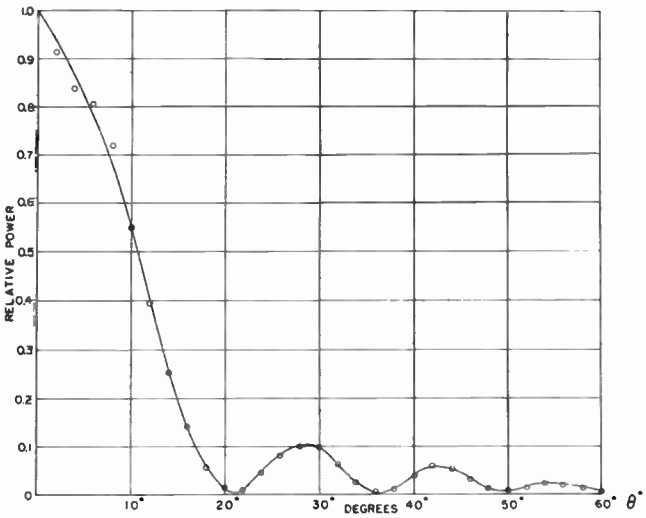


Fig. 29
Composite radiation pattern of corrugated surface of length 6.73λ excited by a six-element end-fire slot array.
f = 9930 mc

AN ANECHOIC CHAMBER MAKING USE OF A NEW BROADBAND ABSORBING MATERIAL

Alan J. Simmons and William H. Emerson
Naval Research Laboratory
Washington 25, D. C.

Introduction

In the study of antennas, a characteristic of prime interest is the radiation pattern measured when the antenna is radiating into free space. With many small microwave antennas it is feasible to measure "far-field" patterns at fairly short distances and therefore it is convenient to work indoors. This is only possible if reflections from walls, floor and ceiling do not appreciably disturb the free space conditions. Thus it is desirable to construct a special chamber whose inside surfaces are coated with non-reflecting material and which contains a minimum of antenna mounting structure.

The construction of such an anechoic chamber or darkroom has been held up by the lack of suitable absorbing material. A good material should reflect only a small percentage of the incident power over a wide range of frequency and of angle of incidence for any polarization. In addition it should be lightweight, easily mounted, and relatively low in cost. A material with these characteristics has been designed by the Absorbent Materials Branch at the Naval Research Laboratory and is now being manufactured commercially. Using this material, an anechoic chamber has been built by the Antenna Research Branch and has proven very useful over the entire microwave region.

The Absorbing Material

Design and Development

Radio frequency energy propagating in free space may be effectively absorbed by passing it into a dissipative medium so designed as to reflect only a small amount of the incident energy. This medium should have relatively low dielectric constant to facilitate impedance matching to free space and yet have as high a value of attenuation as is consistent with the low dielectric constant in order to reduce the required thickness of the medium. These characteristics suggest the use of a resistive expanded material for this purpose. In the course of development, various types were evaluated. On the basis of cost and performance the material chosen for a practical absorber was one made by impregnating a commercially available mat of curled animal hair with a mixture of conducting carbon black in neoprene to provide conduction loss. The hair material is widely used for packing and upholstering purposes because of its resiliency and light weight. It is also well suited for use as an expanded framework for supporting the dissipative material because of its large percentage volume of air (approximately 96%) and the relatively large total surface area provided by the great quantity of hairs. The lossy rubber is applied by either spraying or dipping. This process coats the individual hairs with a thin film of lossy material. Measurements showed that this conducting layer has a resistivity of approximately 8 ohm cms and at 3000 mc a dielectric constant of about 50 with a loss tangent of about 1.4.

Reflections from the air-material interface may be minimized over a wide frequency range by either geometrical shaping of the front surface or by distributing the loss with depth. In using the first method, the top surface of the treated hair material was cut into a series of pyramids of equal height. While it was shown experimentally that other shapes such as cones and wedges were similarly effective, these were less desirable from the standpoint of fabrication. In using the second method, the top surface was left flat and the amount of conducting rubber applied to the hair was increased gradually with depth in the material from a low value at the front face to a large amount at the back. The air-to-material transition was also aided here by the rough, uneven nature of the hair surface. Both these methods may be considered as utilizing lossy tapered sections to transform impedance from that of free space to that of the medium. As frequency is increased, both methods were found experimentally to start reducing

reflection in the region where the length of taper is a quarter wavelength long in air and to continue to improve in effectiveness as the electrical length of taper increases.

Experiments were performed to determine empirically the influence of factors such as density of hair, type of carbon black, amount of distribution of loss, and size and shape of pyramids. Application of the results of these experiments demonstrated that it was possible to design both flat and pyramidal types of absorbers which would meet the requirements for the anechoic chamber.

Experimental development and small scale production of these absorber types was carried out at the Naval Research Laboratory pilot plant.*

Measurement Techniques

Figure 1 shows a diagram of the apparatus used to make free space measurements in the frequency range of 2,500 to 25,000 mc. The arch allows convenient variation of the angle of incidence and polarization of the radiation. Transmitted energy is reflected from a metal plate to a receiving horn. A comparison of the amount of energy reflected from the metal plate with that detected when the plate is covered by an absorber allows a determination of the percentage of incident power reflected by the material.

At frequencies below 3000 mc a completely enclosed system is used to measure the performance. In this case a sample of the material to be evaluated is mounted at the end of a shorted section of special coaxial transmission line. A measurement of VSWR in the line is used to determine the ratio of reflected to incident power.

There is reasonable agreement between these two methods of measurement in the frequency range of overlap.

Characteristics of Commercial Materials

Since January 1952 the Sponge Rubber Products Company of Shelton, Conn. has manufactured quantities of both flat and pyramidal materials. Figure 2 shows a cross-section view of both types. The commercial pyramidal material which was used in the Naval Research Laboratory darkroom is about 3-3/4" thick including the pyramids which are 2-1/2" high and 3" wide. It weighs approximately 1.4 pounds per square foot. The flat material is somewhat thinner and more flexible than the pyramid type. It is about two inches thick and weighs about 0.4 pounds per square foot.

Figure 3 is a plot of specular reflection vs. frequency at angles near normal incidence for both types of materials. It is seen that performance is similar for both types at frequencies above 1500 mc but that, as a result of greater thickness and dissipation, the pyramid material is somewhat better at lower frequencies. The energy reflected from these materials as compared to incident energy is about 20 db down at 10 cms, 25 db down at 3.2 cms, and 30 db down at 1.2 cm for radiation at normal incidence. The high frequency limit of effectiveness has not been determined on these materials. Measurements made at a wavelength of 5 mm indicates that the performance of these materials is still continuing to improve. Above 2500 mc, the reflection from both materials is essentially the same whether they are backed by a conducting sheet or left open circuited.

Figure 4 shows how performance varies with angle of incidence and polarization for specular reflection. It is seen that the reflection is somewhat similar for both polarizations and that these materials are effective over a wider range of angles of incidence at smaller wavelengths. Both flat and pyramid types are similar in respect to performance at various angles for the wavelengths shown.

It has been established that it is the intrinsic absorbing property of the material, rather than the scattering of the incident energy over a wide solid angle, that leads to the low reflection which is observed. This was determined by comparing free space measurements where only the specularly reflected energy is detected with those made with the material in a closed horn where scattered energy is added to that directly reflected. No increase in reflected energy was found in the latter case.

*This phase of the work was under the direction of H.A.Tanner of the Chemistry Div.

Measurements were also made to determine the upper limit of power-handling ability of these materials. It was found at 1300 mc that both flat and pyramidal types could dissipate as much as one watt per square inch, and that at this power level the temperature of the hair is around 200°F.

While these materials were designed primarily for indoor use, they have been found to have sufficiently stable characteristics to withstand considerable exposure to rain and sun. This stability was well demonstrated by one sample which was kept under water in the Potomac River for 10 weeks. Upon removal and drying, the reflection from it was found to have changed by less than 3 db at 5 test frequencies above 2500 mc. Other samples which were left exposed to the weather for 11 weeks showed less than 4 db variation in reflection above 2500 mc. Samples of both types were also measured directly after withdrawal from a tank of water to determine the effects of wetness. Here the pyramid material was found to return to within 3 db of original performance within 30 seconds whereas the flat material continued to show considerable reflection until natural drainage and evaporation removed large drops of water trapped near the front surface. In view of these experiments it is considered likely that, at least in dry weather, these materials can also be used effectively to reduce reflections from obstacles near outdoor antenna test ranges.

The Anechoic Chamber

Physical Description

The framework of the chamber was built of 2" x 4" lumber inside an available room and is 20 feet long, 10 feet wide and 7 feet 8 inches high. The absorbing mats are tied to the walls and ceiling with string and laid on the floor where necessary. Two access doorways at either end are masked by screens of absorbing material on a portable framework.

The antenna under test transmits and is mounted on an L-shaped arm (Figure 5) which extends down into the chamber from a rotating mount on the roof of the chamber. The klystron oscillator is mounted on this arm and rotates with the antenna. The receiving antenna is on a rolling mount which is guided by a track down the middle of the floor. This mount contains a motor for rotating a linearly polarized antenna about its axis for automatic recording of antenna polarization characteristics.

Maximum separation between receiving and transmitting antennas is about 13 feet, making possible the study of antennas up to 10 inches in diameter at X-band (around 3.2 cm) and 18 inches at S-band (around 10.0 cm), using the criterion $R_{min} = 2D^2/\lambda$ where D is the maximum linear aperture dimension of the larger of the two antennas.[1]

Theoretical Considerations

In the measurement of directive antennas, the greatest error in pattern occurs when the transmitting antenna, whose pattern is being measured, has rotated so that it points at a spot on the wall midway between the receiving and transmitting antennas. At this angle energy is transmitted from one antenna to the other by specular reflection off the wall. If the angle of rotation is θ_r and the effect of the reflection is represented by an image antenna behind the wall, then the ratio of undesired to desired signal strength is

$$\frac{|E_u|}{|E_d|} = \frac{R_1}{R_2}\sqrt{\frac{P_I G_T(0) G_R(\theta_r)}{P_T G_T(\theta_r) G_R(0)}} \quad ,$$

where R_1 is the distance between transmitting and receiving antennas,
R_2 is the distance between image and receiving antennas,
P_T is the total power radiated by the transmitting antenna,
P_I is the total power radiated by the image antenna,
$G_T(\theta)$ is the gain of the transmitting antenna,
$G_R(\theta)$ is the gain of the receiving antenna.

If $P_I/P_T = \varphi(\theta_r)$, the power reflection coefficient of the absorbing material at the angle θ_r, and if $G_T(\theta) = G_R(\theta)$, i.e., transmitting and receiving antennas are identical, then

$$\frac{|E_u|}{|E_d|} = \frac{R_1}{R_2}\sqrt{\varphi(\theta_r)}$$

As a numerical example, if $R_1/R_2 = 1/2$ and $\rho(\theta_r) = 1\%$, as measured on single mats, then

$$\frac{|E_u|}{|E_d|} = .05$$

and the pattern could be in error by $\pm 1/2$ db at this angle θ_r. An error of this magnitude is usually tolerable.

If

$$\frac{G_R(\theta_r)}{G_R(0)} > \frac{G_T(\theta_r)}{G_R(0)} \quad ,$$

that is, the receiving antenna is less directive than the transmitting antenna, the error becomes larger, so it is obviously advisable to make the receiving antenna as directive as the transmitting antenna, except at frequencies where ρ is so small as to make this precaution unnecessary. It is not worthwhile, in general, to make the receiving antenna much more directive than the transmitting antenna, since this means increasing the separation, R_1, in accordance with $R_{min} = 2D^2/\lambda$. Increasing R_1 makes θ_r smaller, which makes the factor R_1/R_2 larger, and can make $\rho(\theta_r)$ larger. In addition, with θ_r smaller, the interference occurs at angles closer to the main beam axis where most interest in the pattern is usually centered.

Another case of interest is the non-directive beacon-type antenna. If a directive receiving antenna is used the only significant reflections come from the back wall, directly behind the omnidirectional antenna, since this is the only spot in the main beam of the receiving antenna. Thus

$$\frac{|E_u|}{|E_d|} = \frac{R_1}{R_2}\sqrt{\rho(0)} \quad .$$

This expression is independent of the receiving antenna gain as long as that gain is sufficient to discriminate appreciably against reflections from side walls. Here again if $\rho(0) \leq 1\%$ and $R_1/R_2 \leq 1/2$, the pattern is accurate to at least $\pm 1/2$ db.

Experimental Tests

In the foregoing simple theory, various assumptions have been made including the one that the wall acts as a specular reflector whose reflection coefficient may be measured on a single sample by the two-horn method mentioned previously. In order to see if these assumptions give a reliable guide to the use of the room, and in order to evaluate the room experimentally, three kinds of experiments were performed. The first and most direct was to compare patterns measured in the chamber with those measured on a fairly reliable roof top site. The second was an attempt to measure reflections off the walls directly, similar to the measurements made on individual mats. The third was an attempt to measure standing waves set up in the room by reflection off all the walls.

1. Comparison patterns for two fairly large antennas at X- and S-band are shown in Figures 6 and 7. Figure 6 shows E-plane patterns of an 8 inch double-dipole-fed paraboloid at X-band ($\lambda = 3.2$ cm). Notice that the outdoor patterns are not perfect, showing some fluctuation at wide angles due to reflections off nearby objects. It is felt that the agreement between the two patterns is good.* Figure 7 shows H-plane patterns of an S-band ($\lambda = 10$ cm) dipole-disc-fed paraboloid, and again agreement is good between indoor and outdoor patterns. To illustrate the effect of wall reflections, the patterns of the S-band antenna were repeated using a small non-directive antenna as a receiving antenna, with results shown in Figure 8. In this case, θ_r was about 35° and $\rho(\theta_r)$ was several percent, as determined by single-mat measurements, so that in the region around 35° severe pattern distortion may be noted (compare Figures 8 and 7). It is to be pointed out, however, that this distortion was effectively eliminated by use of a directive receiving antenna as shown in Figure 7.

2. Using these same antennas, a direct measurement of wall reflection was made by measuring patterns with the receiving antennas directed at various points on the side wall and noting the apparent increase in side lobe

*The filling in of the first minima was caused by phase error, not by reflection.

level due to reflection. By far the greatest reflection was found at the angle of specular reflection, indicating that non-specular scattering plays a small part in the perturbation of patterns. Results for this one angle of incidence, 55°, ($\theta_r = 35°$) which was the largest feasible in the room and thus should represent a worst case are shown in Table I.

TABLE I

WAVELENGTH	POLARIZATION WITH RESPECT TO PLANE OF INCIDENCE	MAXIMUM POWER REFLECTED (AT 55°)
3.2 cm	perpendicular	-27 db
3.2 cm	parallel	-35 db
10.0 cm	perpendicular	-13 db
10.0 cm	parallel	-23 db

These measurements are a check on the single-mat measurements. They are obtained by a slightly different method and measure reflections from a whole wall instead of a single mat. The results are in good agreement.

3. Although in almost all cases only the reflections from a single wall at a time need be considered when the perturbing effects of the room are taken into account, one experiment was performed which measured the effect of multiple-wall reflections. This was a sort of standing-wave measurement. One omnidirectional antenna, a monopole mounted on a ground plane, was swung in an arc about a similar antenna (see Figure 9) and the rapid pattern fluctuations were measured to give an indication of reflections adding in random phase to the direct signal. The fluctuations in the neighborhood of the center of the room were converted into percent undesired signal power and the results for a 4 foot separation of antennas are shown in Table II.

These measurements, although somewhat approximate, agree fairly well with what might be calculated using single-sheet reflection coefficient measurements and taking into account multiple-wall reflections in random phase.

TABLE II

λ	MAXIMUM UNDESIRED POWER %
3.2 cm	1%
5 cm	4%
7.5 cm	3%
10 cm	3%
12 cm	2-1/2%
15 cm	5%
20 cm	4%
25 cm	3%
30 cm	5%

Final Remarks

The room was designed principally for use in the range from about 3,000 to 10,000 mc. No difficulty should be observed in extending this frequency range upward since the absorbing properties of the material have been shown to remain good, at least through 60,000 mcs. A gradual increase in reflection below 3000 mc makes the room less useful, though the results of the standing-wave test and measurements on single samples indicates that the reflections at frequencies down to 1000 mc have not increased beyond the point of usefulness.

In conclusion, it seems that the results of measurement on a single sample may be applied by means of a simple image theory to give correct order of magnitude for the perturbing signals due to reflections. In our chamber, though it is relatively small, patterns of X-band antennas up to 10 inches in aperture dimension may be safely measured with results often better than those on an average roof site. The absorbing material is more than adequate at X-band. At S-band, with the precaution of using as directive a receiving antenna as the antenna under test, antennas up to 18 inches may be safely measured to an accuracy of about 1/2 db out to pattern angles where the signal strength is 20 db down. Omnidirectional antenna patterns may be measured to an accuracy of less than ±1/2 db from 3000 mc up, with slightly reduced accuracy between 1000 and 3000 mc.

Reference

1 S.Silver, "Microwave Antenna Theory and Design," Rad.Lab.Series, Vol.12, McGraw-Hill 1949, p.575

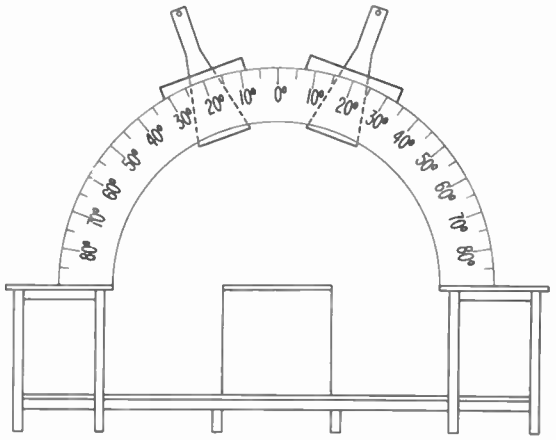


Fig. 1
Absorbing material measurement arch.

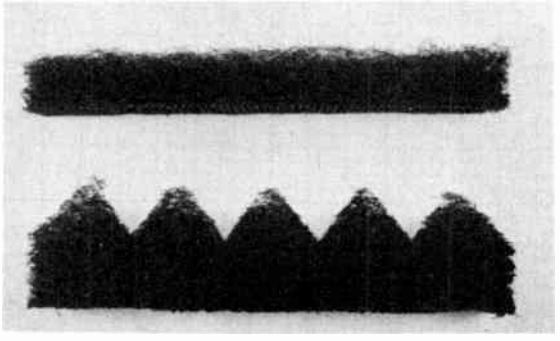


Fig. 2
Cross section view of two kinds of material.

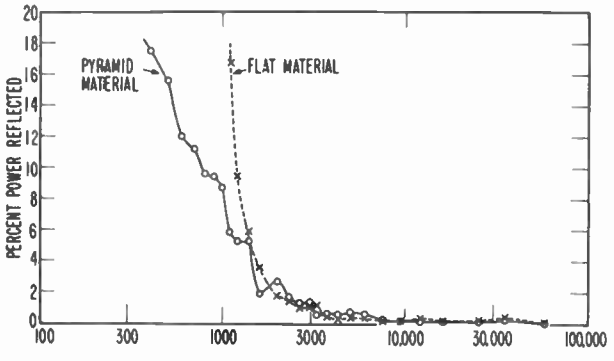


Fig. 3
Production absorbing material, performance versus frequency (normal incidence).

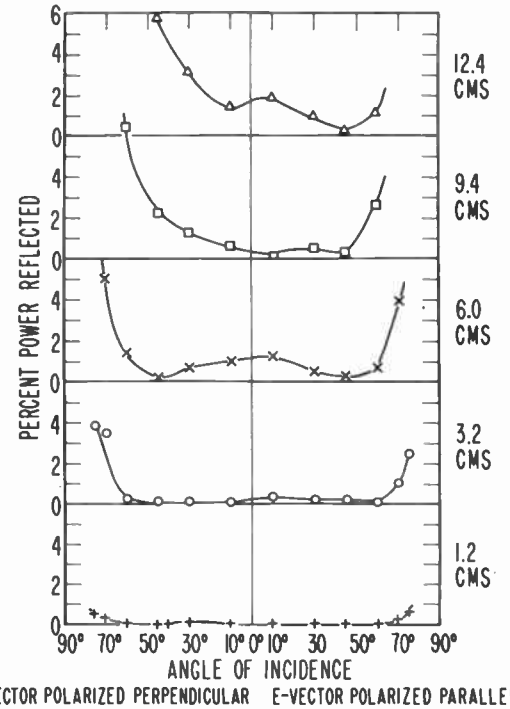


Fig. 4
Performance versus angle of incidence.

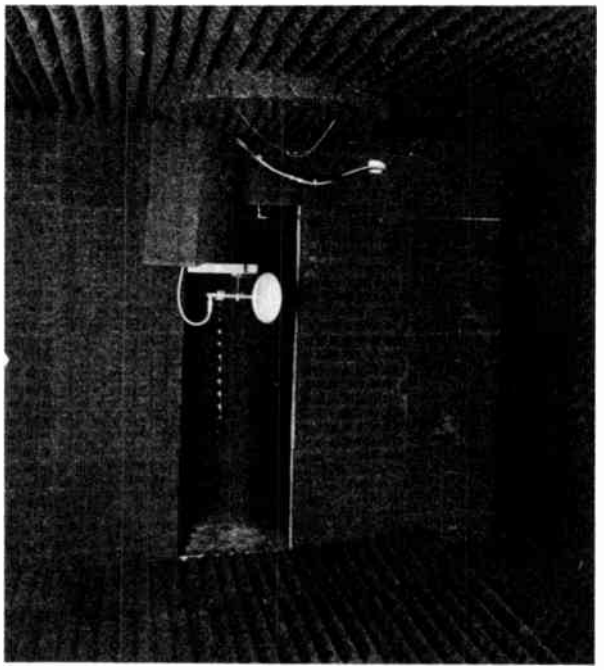
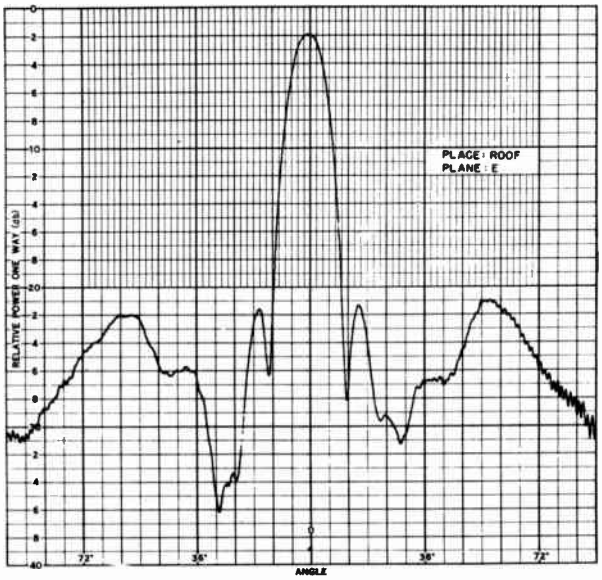


Fig. 5
Anechoic chamber showing rotating antenna mount.



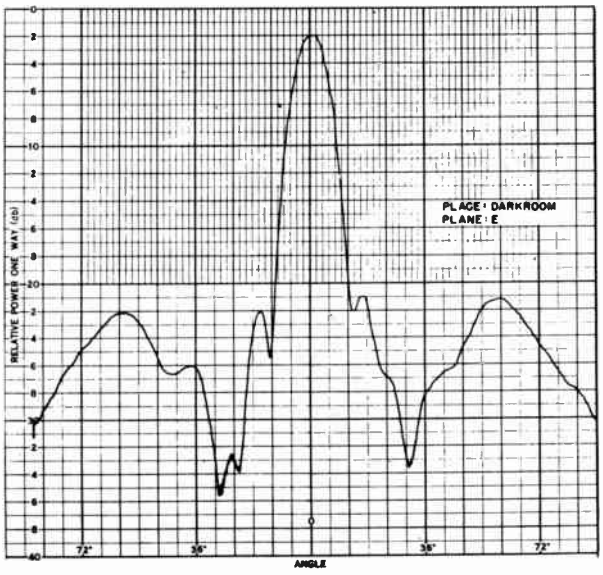
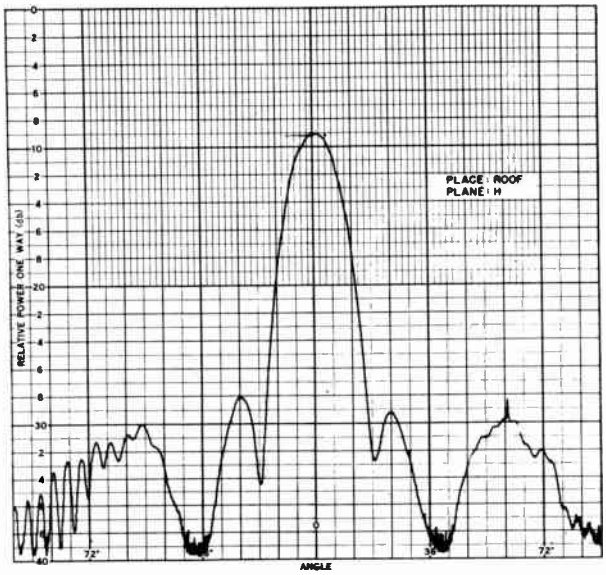


Fig. 6
Patterns of 8 inch paraboloid, $\lambda = 3.2$ cm.



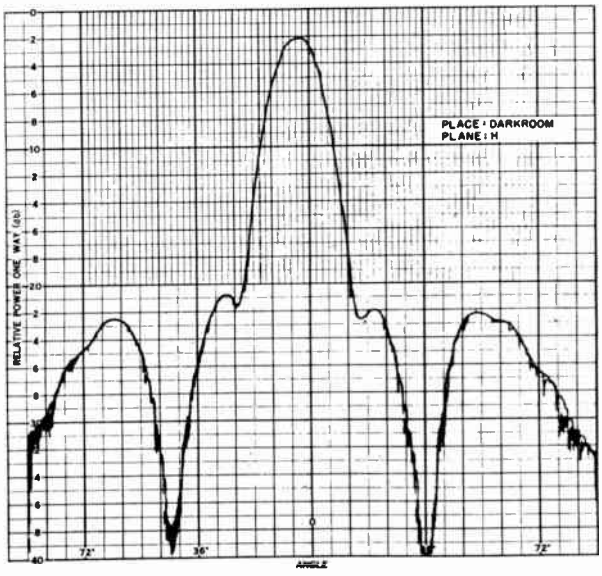


Fig. 7
Patterns of 17 1/2 inch paraboloid, $\lambda = 10.0$ cm.

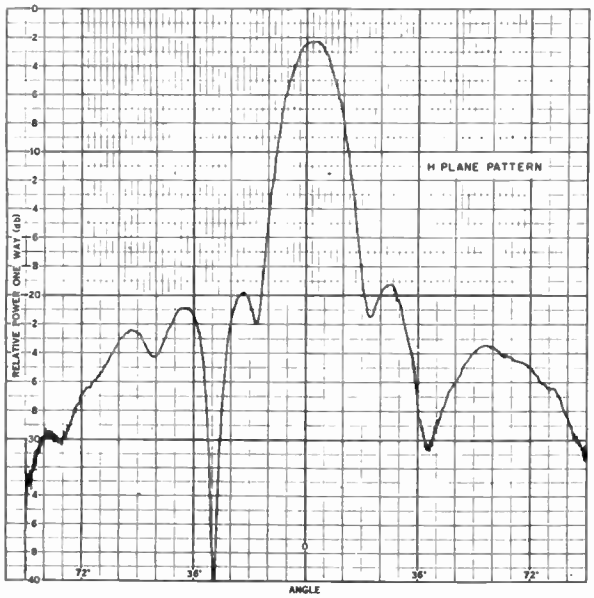


Fig. 8
Pattern of 17 1/2 inch paraboloid
showing effect of wall reflections.

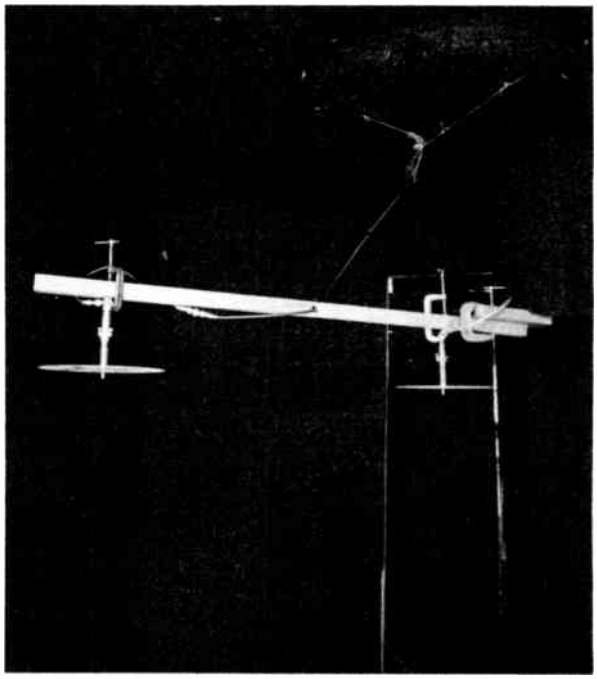


Fig. 9
Standing-wave measurement setup.

WIDE-FREQUENCY-RANGE TUNED HELICAL ANTENNAS AND CIRCUITS

A. G. Kandoian and W. Sichak
Federal Telecommunication Laboratories, Inc.
Nutley, New Jersey

Introduction

Transmission line circuit elements, or more generally circuit elements with distributed constants, are used advantageously at V.H.F. and U.H.F. At lower frequencies, except for antennas and occasionally in associated networks, their use is not common for the very good reason that such circuit elements become prohibitively bulky. In antenna applications the large size is tolerated in order to achieve a useable radiation efficiency.

This condition prevails as long as the type of distributed constant circuit elements under consideration is limited to the conventional coaxial or parallel transmission line type. There is however a class of distributed constant circuit elements, made up of helical conductor with or without cylindrical outer conductor, which can be used with considerable advantage at all radio frequencies up to and including the UHF band. Of particular interest for the present discussion is the case of the helix with a diameter small compared to wavelength.

Though this class of helical circuit elements has seen limited use in applications similar to the ones under discussion at present, it is felt that their potentialities have not been fully appreciated and utilized nor has their theory and fundamental properties been adequately discussed in technical literature.

The helical antenna, with diameter considerably less than a wavelength, radiating predominantly vertical polarization in the normal mode (peak of radiation pattern normal to the helix axis) has several advantages over a conventional dipole when the height must be considerably less than a quarter wavelength. The radiation resistance is about 50% greater than a dipole of the same height and the input impedance can be made real so that the problem of matching to a transmission line is simplified. In addition the helix can be matched by grounding the base of the helix and connecting the transmission line to the proper tap point so that no external matching networks are required.[1]

The basic idea of the short resonant helical antenna is not new; for example, Wilson[2] used a helix 2 feet high with a 6 foot whip to make an efficient 4 Mc mobile antenna. It is understood that some radio amateurs use a type similar to Wilson's or a complete helix on the 4 Mc and 14 Mc bands. However, the small diameter helix does not seem to have been discussed extensively in the literature, in contrast with the work of Kraus[3] and others on the large diameter helix.

The two general types are shown on Figure 1. Type A generally requires an external matching network and for some applications is less desirable than type B. Type B uses shunt feed so that the matching network is part of the radiating structure. In addition, type B can be made tunable over wide frequency ranges as shown in Figure 1c by using a variable tap point and a variable short circuit. The lower frequency limit is obtained when the length of wire in the helix is approximately a quarter wavelength, while the upper limit is obtained when the height is a quarter wavelength. With proper design, tuning ranges up to 100 to 1 can be obtained.

Velocity of Propagation

The phase velocity along the axis of the helical antenna can be obtained from the previous work on coaxial lines with helical inner conductors and also from the work on the helix for use in traveling wave tubes[4]. The velocity is given by

$$(C/V)^2 = 1+(M\lambda/\pi D)^2 \qquad (1)$$

M is a function of the helix diameter, frequency, and number of turns per unit length and is given on Figure 2. Also shown is the ratio (approximate) of the apparent velocity along the wire to the velocity of light, V_w/c. This ratio is close to one for the helices used in traveling wave tubes ($M \gg 1$), less than about 1.3 for the usual coaxial lines with helical inner conductors, but may be greater than two for some helical antennas. The exact expression for V_w/c is

$$V_w/c = \left[\frac{1+(N\pi D)^2}{1+(M\lambda/\pi D)^2}\right]^{1/2} \qquad (2)$$

The approximation in Figure 2 is obtained by neglecting one in comparison with the other factors.

When the helix is used with a ground plane the height must be a quarter wavelength at the velocity given by equation (1) in order to obtain a real input impedance

$$\frac{h}{\lambda} = \frac{1}{4\,c/v} = \frac{1}{4\sqrt{1+20(ND)^{2.5}(D/\lambda)^{1/2}}} \qquad (3)$$

The last expression holds when $ND^2/\lambda \leq 1/5$. Figure 3 shows measured and calculated values of height vs. number of turns/inch. The curve near the measured points was calculated using equation (3), while the lower curve was calculated assuming that $V_w/c = 1$. As can be seen from the figure, the measured height is sometimes twice the height predicted using $V_w/c = 1$.

Figure 4 shows how the calculated resonant frequency varies with the number of turns per inch while holding the wire length constant. If V_w/c were a constant, each curve would be a straight line with zero slope.

Radiation Resistance and Effective Height

The effective height[5] of a short dipole above a perfect ground is $\ell_{eff} = h/2$. The effective height of a resonant helical antenna is $\ell_{eff} = 2h/\pi$ because the current distribution is sinusoidal instead of linear. The radiation resistances are given by

$$\begin{aligned} R_r &= (20h/\lambda)^2 && \text{short dipole} \\ R_r &= (25.3h/\lambda)^2 && \text{resonant helix} \end{aligned} \qquad (4)$$

For heights near a quarter wavelength the numerical factor in the resonant helix equation should be changed to 24 in order to get 36 ohms when the height is a quarter wavelength.

Polarization

For a helix in free space the ratio of the vertically polarized field to the horizontally polarized field is given by

$$\frac{E_v}{E_h} = \frac{J_0(\pi D/\lambda)}{(N\pi D)J_1(\pi D/\lambda)} \approx \frac{\lambda}{5ND^2} \qquad (5)$$

The approximate result, good for small diameters, has been given by Wheeler[6]. If the helix is resonant, circular polarization is obtained when the overall height is about 0.9 times the diameter.

When the helix is used with a ground plane the horizontal polarization is weakened considerably if the height of the helix is small. For this case the pattern of the horizontally polarized field is

$$E'_H \approx E_H\ (2\pi h/\lambda)\ \sin\theta \cos\theta \qquad (6)$$

The maximum value of this pattern occurs at $\theta = 45^\circ$. The pattern of the vertically polarized field is

$$E'_V \approx E_V \pi \cos\theta \qquad (7)$$

The ratio of the horizontally polarized field at $\theta = 45^\circ$ to the vertically polarized field at $\theta = 0^\circ$ is

$$\left(\frac{E'_H}{E'_V}\right)_{max} \approx \frac{5ND^2h}{\lambda^2} \qquad (8)$$

Thus if circular polarization were obtained with a helix 0.1 high in free space the horizontally polarized field at $\theta = 45^\circ$ would be 20 db below the vertically polarized field at $\theta = 0^\circ$ when the helix is used with a ground plane.

Losses

Base Insulator Losses. Low efficiency may be obtained with short dipoles because of power loss in the insulator used to support the dipole. The choice of dielectrics is limited because of mechanical requirements and a power factor less than 10^{-3} is probably unobtainable at present. The equivalent circuit is shown on Figure 5. For this circuit

$$\frac{P_L}{P_R} = (wC\tan\delta)(R_r^2 + X_r^2)/R_r \qquad (9)$$

For a 35 foot whip 2 inches in diameter, below about 3 Mc, with a 40 uuf base insulator of power factor 10^{-3} and using the following

$$\begin{aligned} R_r &= (20\ h/\lambda)^2 \\ X_r &= -Z_0 \cot(2\pi h/\lambda) \qquad (10) \\ Z_0 &= 60\left[\ell n\left(\frac{4h}{d}\right) - 1\right] \end{aligned}$$

this becomes

$$P_L/P_r \approx 1.25/F^3_{mc} \qquad (11)$$

At 1 Mc the efficiency is about 50%, while at 300 KC the efficiency is about 2% (17 db loss).

In addition to the base insulator loss there will be further losses in the network used to match the antenna to a transmission line. In contrast a helix 35 feet high made of 1/2 inch wire wound on a form one foot in diameter with 1.6 turns per inch has a calculated total efficiency of about 5% (13 db loss) at 300 KC.

Helix Losses. The ohmic losses in the metal of a short dipole are ordinarily quite small but the losses in the helical antenna may be appreciable because the wire diameter is usually much smaller than the diameter of a dipole of the same height. The loss can be calculated[7] assuming that the current distribution is sinusoidal and neglecting proximity effects.

$$\frac{P_L}{P_r} = \frac{210^{-4}(V_w/c)}{d\ F_{mc}^{1/2}(h/\lambda)^2} \qquad (12)$$

This equation gives the ratio of power dissipated in the copper to the power radiated. The efficiency of the helix is therefore $1/(1+P_L/P_r)$. Figure 6 is a plot of height vs. resonant frequency for 50% efficiency assuming that $V_w/c = 1$.

Measurements were made at 100 Mc of the power radiated by helices of various heights and wire diameters compared to the power radiated by a matched quarter wavelength dipole. The base of each helix was soldered to the ground plane and the inner conductor of the 50 ohm transmission line connected to the proper tap point on the helix to get a standing wave ratio less than 1.5. The results are shown on Figure 7. The solid curves are calculated using equation 9 above. In most cases the agreement is fair, except for the thinnest wire, where the measured points are consistently low.

Q and Tap Point

The Q of an antenna is of interest because it limits the bandwidth which is to be used. The Q* of the helical antenna can be calculated if some assumptions and approximations are made.

The antenna is assumed to be equivalent to a resonant line a quarter wavelength long. The input resistance ($\lambda/4$ away from the open circuit) is[8]

$$\frac{R_{base}}{Z_0} = \frac{\pi}{4Q} \qquad (13)$$

R_{base} is the sum of radiation resistance and wire resistance, referred to the base of the antenna. For a resonant antenna $R_w = R_0 \ell/2 = R_0 (\lambda/8)(V_w/c)$

$$R_{tap} = \frac{Z_0^2 \sin^2\theta_1}{R_{base}} = \frac{4}{\pi} Q\, Z_0 \sin^2\theta_1 \qquad (14)$$

The characteristic impedance which best fits the experimental data is the helical transmission line characteristic impedance derived previously[9]. This characteristic impedance is also given by

$$Z_0 = 60 \left[\ell n(4h/D) - 1\right] \quad (C/V) \qquad (15)$$

The first factor in this equation is the familiar expression which predicts the reactance of ordinary dipoles. The (C/V) factor takes account of the fact that the axial velocity of the helical antenna is less than that of the ordinary dipole.

Figure 8 shows how the Q varies with h/λ for one of the antennas used in obtaining the data for Figure 7. The agreement is fair using equation 13. When equation 14 is used to predict the tap point θ_1, fair agreement is also obtained except when the diameter is large and/or C/V is large.

Circuit Applications

Important applications of helical transmission lines are delay lines, wide tuning range resonant circuits, and the extension of transmission line techniques to frequencies as low as 300 KC.

A delay line 1-5/8" in diameter for use at 10 Mc was built with the following characteristics: characteristic impedance, 5000 ohms; a delay of 0.2 microseconds/foot, and an attenuation of 0.9 db/microsecond.

Figure 9 shows resonant frequency, vs. linear movement of a shorting plunger for five quarter wavelength resonators. The helical inner conductors were made with different configurations--some were wound with constant pitch, some with a constantly varying pitch, and others with sections of different pitch. One of the resonators shown tuned from 50 Mc to 3500 Mc--a range of 70/1.

The tuning range is, approximately:

$$\frac{f_{max}}{f_{min}} = \frac{\lambda_{max}}{\lambda_{min}} \approx \frac{\ell}{h} \approx C/V \qquad (16)$$

Thus with proper design tuning ranges of 100/1 are possible.

Acknowledgement

Acknowledgement is due to R. A. Felsenheld, L. Juhas, and W. Huey for making the measurements reported in this paper.

Glossary

C	-	velocity of light
d	-	wire diameter, inches
D	-	mean helix diameter, inches
F_{mc}	-	frequency in megacycles
h	-	height
ℓ_{eff}	-	effective height
N	-	number of turns per inch
P_L/P_r	-	ratio of power dissipated to power radiated
R_r	-	radiation resistance
R_0	-	resistance per unit length
V	-	phase velocity along the axis
V_w	-	apparent phase velocity along the wire
w	-	2πx frequency
X	-	reactance
Z_0	-	characteristic impedance
$\tan\sigma$	-	power factor
λ	-	wavelength
θ_1	-	angular distance from short circuited end to tap point

*In this paper "Q" is the "unloaded Q". When driven by a generator with zero internal resistance the radiated power is 3 db below the maximum radiated power at two frequencies fo/Q apart. When driven by a generator with internal resistance equal to the antenna resonant resistance the radiated power is 3 db below the maximum radiated power at two frequencies 2fo/Q apart.

References

1. Morrison & Smith - "The Shunt-Excited Antenna" PROC IRE, Vol. 25, pp. 673-696, June, 1937.

2. W. R. Wilson - "A Solenoid-Whip Aerial" ELECTRONICS, Vol. 14, January, 1941, pp. 56-64.

3. J. D. Kraus, "Antennas" McGraw-Hill, New York, 1950.

4. J. R. Pierce - "Traveling Wave Tubes," Van Nostrand, New York, 1950, page 229. See also PROC IRE, Vol. 35, pp. 111-123, February, 1947.
L. L. Chu & J. D. Jackson - "Field Theory of Traveling Wave Tubes" PROC IRE, Vol. 36, pp. 853-863, July, 1948.

5. See, for example, Schelkunoff and Friis, "Antennas, Theory & Practice" John Wiley and Sons, New York, 1952, page 309.

6. H. A. Wheeler, "A Helical Antenna for Circular Polarization" PROC IRE, Vol. 35, pp. 1484-1488, December, 1947.

7. For example, Schelkunoff & Friis, page 338.

8. "Reference Data for Radio Engineers," FTR, Third Edition, 1949, page 312.

9. "Coaxial Line with Helical Inner Conductor" Submitted for publication in the PROC IRE.

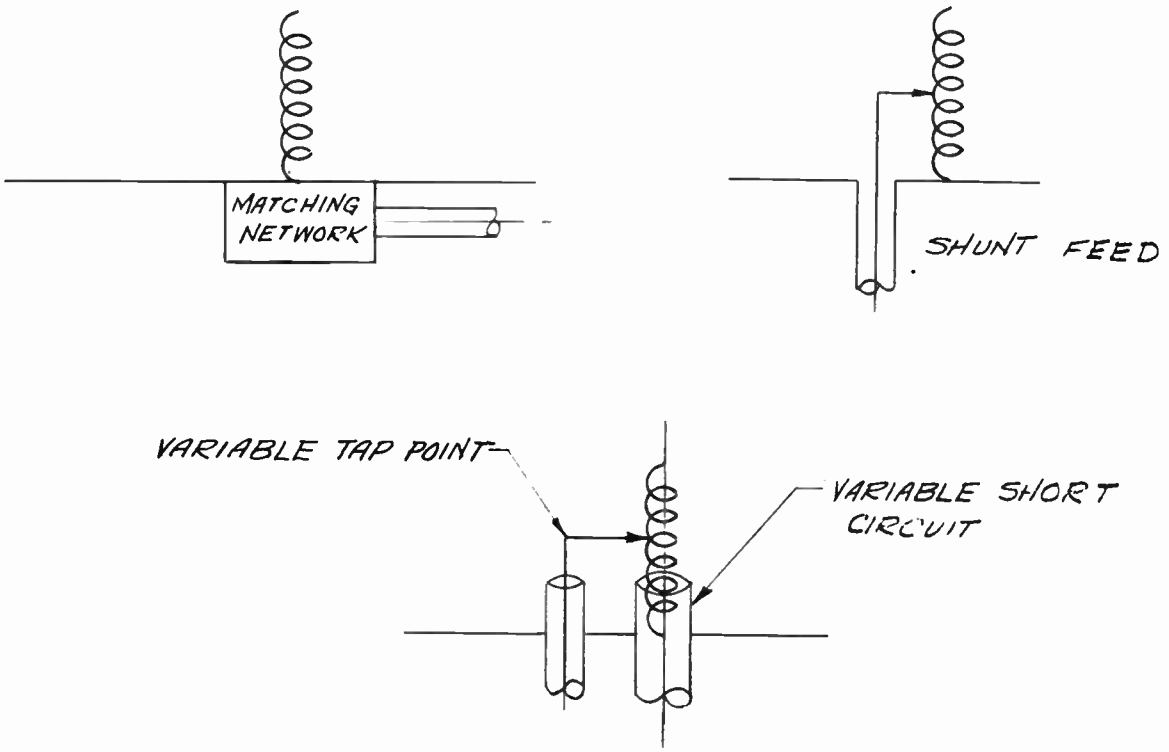


Fig. 1 - Helical antenna types.

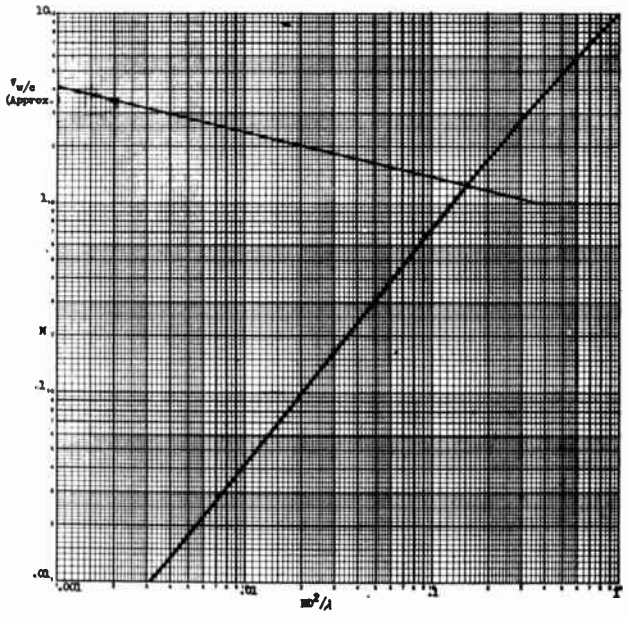


Fig. 2

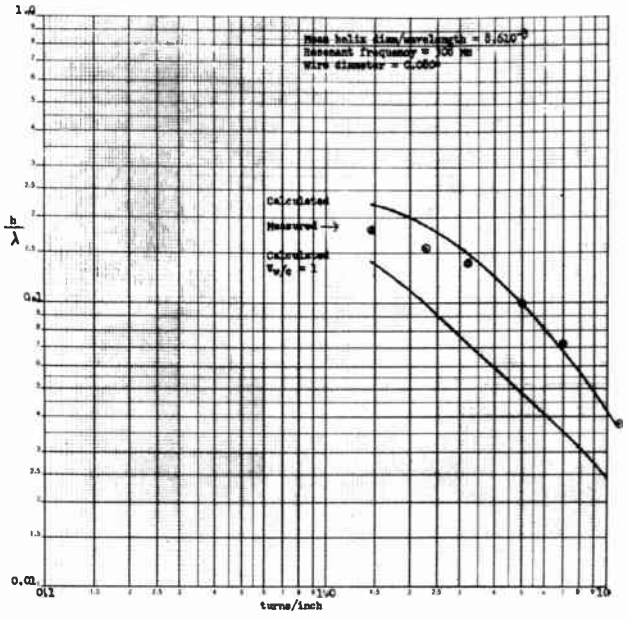


Fig. 3 - Helical antenna above ground plane.

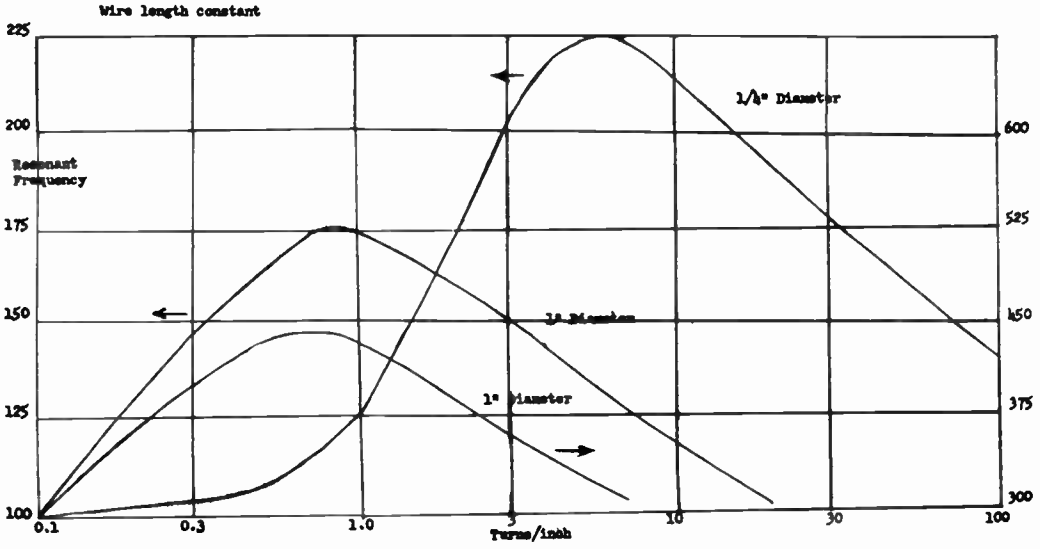


Fig. 4

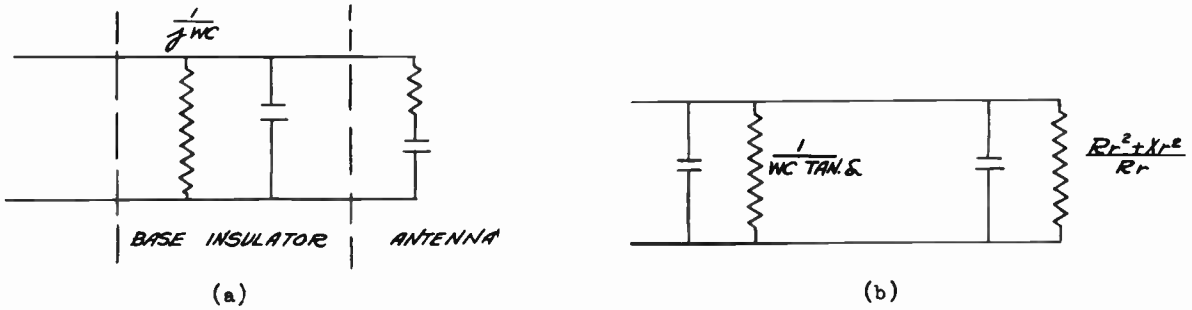


Fig. 5 - Circuits of a short dipole and base insulator. (a) actual circuit; (b) equivalent circuit.

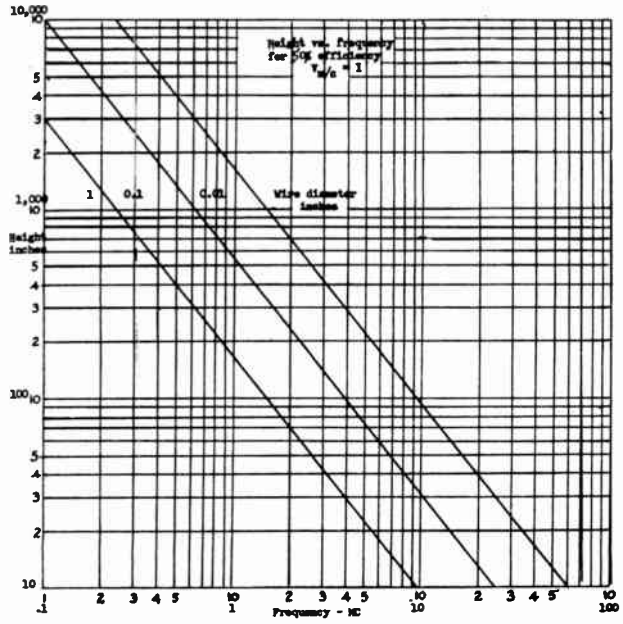


Fig. 6

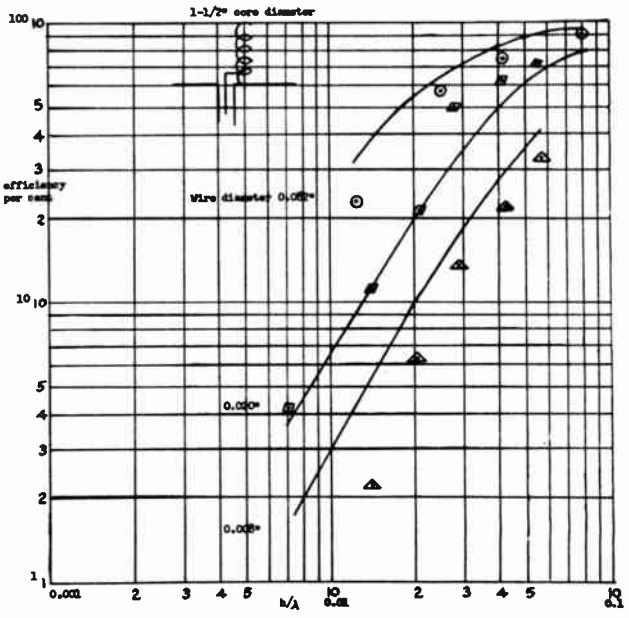


Fig. 7

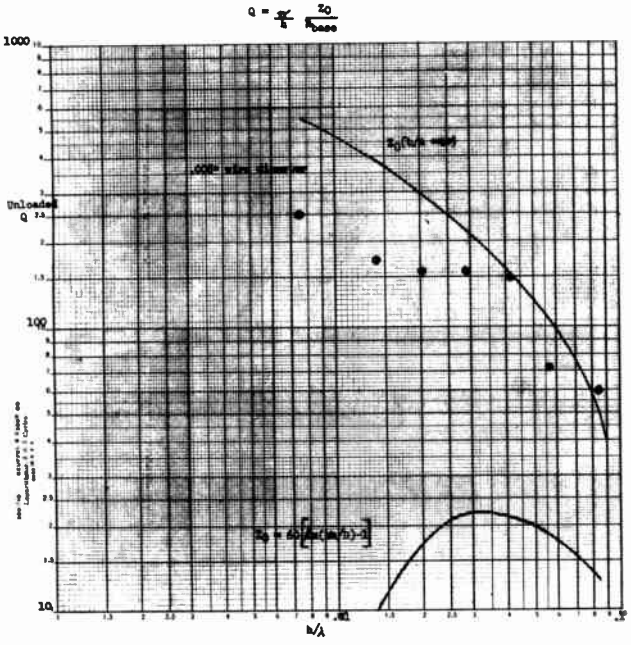


Fig. 8

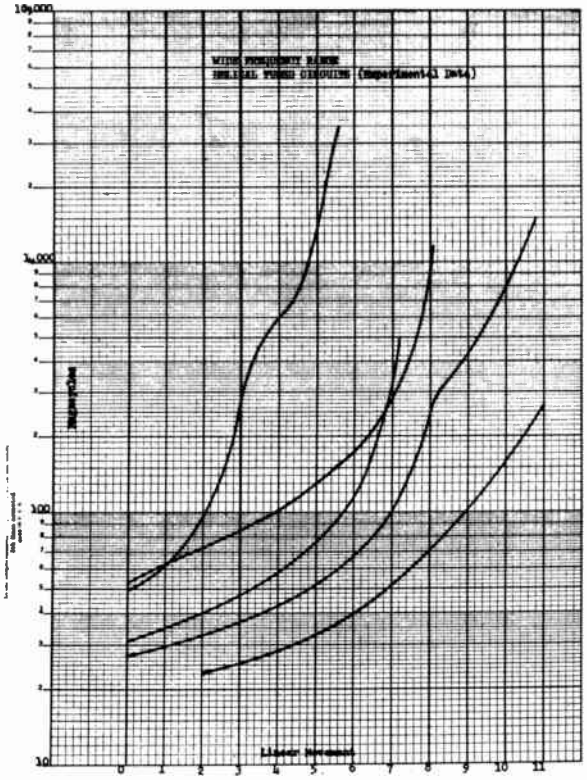


Fig. 9

ARRAYS OF FLUSH MOUNTED TRAVELING WAVE ANTENNAS

J. N. Hines, V. H. Rumsey, and T. E. Tice
Antenna Laboratory, The Ohio State University
Columbus, Ohio

This paper discusses the results of an experimental investigation into the problems that arise in the design of arrays of tapered depth traveling wave slots.

The conventional scheme used in array design employs the principle of pattern multiplication, That is,

Array Pattern = Primary Pattern x Array Factor.(1)

Pattern multiplication assumes that the primary patterns of the elements in the array are identical except for relative magnitudes. Past experience with arrays of resonant slots shows that the conventional method of array design has sometimes failed to predict the pattern of an array with sufficient accuracy for practical purposes because the primary patterns were distorted. An exact method for array design can be obtained by applying the principle of superposition.

In order to apply the principle of superposition correctly the first step is to choose the location of the "input terminals" of each element of the array. The individual primary patterns, $\underline{E}_i\ (\theta, \varphi)$, are determined by applying unit voltage to successive pairs of terminals, the remaining terminals being short circuited. Then when voltages V_i are applied to the terminals the array pattern is given by,

$$\text{Array Pattern} = \sum_i V_i\ \underline{E}_i\ (\theta, \varphi). \qquad (2)$$

The problem is then to find the distribution of voltages which gives the best approximation to the desired array pattern and design a feed system to produce these voltages.

This technique of array design is much more involved than the conventional method but it is absolutely accurate. Because of its complexity one objective of the investigation has been to determine the limitations on the use of the simpler method.

A comparison of the radiation pattern of a single slot with the pattern of the same slot in the presence of an adjacent parasitic slot (with λ separation) shows a radical change in the pattern measured in the plane of the elements. The main lobe is broadened and tipped toward the side of the antenna that contains the parasitic slot. However, the pattern in the plane normal to the slots is only slightly affected by the presence of the parasitic. Additional measurements using three and then four elements in the antenna (with λ separation) indicated the same results. A comparison of the radiation patterns of a single slot and slot with parasitic measured as a function of center-to-center spacing (0.8λ to 1.25λ) shows that the distortion due to the parasitic slot increases as the spacing in wavelengths decreases.

Despite the differences in the primary patterns caused by the presence of other radiators, the mutual impedance between input terminals was found to be negligible. This effect may be analyzed by picturing the traveling wave slot as a directional coupler. That is, the field excited in the parasitic slots is a traveling wave moving in the same direction as the field in the driven element. Since the tapered depth antenna is an efficient radiator all of the energy in the parasitic slot is radiated and none returns to the input terminals.

To determine the applicability of pattern multiplication to the design of an array of traveling wave slots the pattern of a four-element array was calculated (in the plane of the array) based first on the method of pattern multiplication and then on the method of pattern superposition. These calculations were carried out for various excitations. Pattern multiplication gives a fairly good approximation to the array pattern for some excitations; for other excitations the more rigorous technique is necessary.

TRANSIENT BUILD-UP OF THE ANTENNA PATTERN IN END-FED LINEAR ARRAYS

Norman H. Enenstein
Member of the Technical Staff
Research and Development Laboratories
Hughes Aircraft Company
Culver City, California

Introduction

Linear antenna arrays have found many applications in radar and communication systems[1-6]. When used to obtain narrow antenna beams at high frequencies, the simplest design generally results when the array is fed from one end. The energy is radiated progressively as the wave groups travel along the transmission line exciting the radiating elements. The direction of the radiated beam for end-fed arrays depends on the spacing of the radiating elements, the pathlength in the transmission line between radiating elements, and the operating frequency.

This paper is concerned with the effect on the antenna pattern attributed to the transit time of a wave group which enters the feed end and proceeds to the elements of the array. For example, when the antenna is excited with an r-f pulse, there is progressive excitation of the elements of the array as the pulse traverses the transmission line. Consequently, there is a progressive build-up of the antenna pattern. The effect of transit time imposes a physical limit on the utility of such arrays.

If an antenna of the end-fed type is used in a radar system, the effect on the angular discrimination of the system is important when the pulse duration and the total array transit time are comparable. If it is used in a communication system the distortion in the signal is important when the associated bandwidth and reciprocal of the transit time are comparable.

Analysis of the problem may be approached in two ways: the time domain and the frequency domain. The time domain approach entails the synthesis of the radiated pattern or received pattern step by step as each progressive element of the array is excited by a given wave group. The frequency domain approach treats the array as a space filter whose field response in a given direction varies as a function of frequency. Conventional techniques of circuit analysis may be used when the antenna is considered in the frequency domain. The equivalence of both approaches is related through the Fourier transform.

The observation of the transient in the pattern build-up depends on the nature of the measurement device. If the response of the instrument used in the measurement is slow compared to the transient period, the effect of the transient in the observation will be obscured. It is therefore important, in the analysis of the effect due the transient period, to discuss the observation technique. The example of the observation of radar echoes is used as an illustration.

All antenna systems are subject to transient phenomena and exhibit space filtering characteristics. Methods similar to those applied to the end-fed array case may be extended for these other systems.

Time Domain Characteristics

Consider an end-fed array (Fig. 1) of 2N+1 radiating elements spaced a distance, l, apart which is excited by a signal whose wavelength is λ. The steady state response, as expressed by the space factor of the array, $S_\lambda(\Psi)$, is[6]

$$S_\lambda(\Psi) = \sum_{n=-N}^{n=N} A_n e^{jn\left(\frac{2\pi l}{\lambda}\sin\Psi-\theta\right)}, \qquad (1)$$

where Ψ is the observation angle as measured from the broadside position, A_n is the relative illumination of the n^{th} element, and θ is the progressive phase shift between elements which is dependent on the wavelength and the nature of the feed structure.

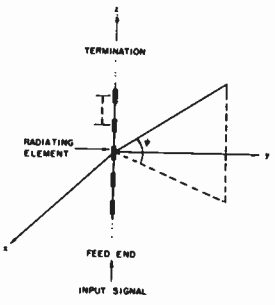


Fig. 1 - End-fed linear antenna array.

If the A_n coefficients are all real and have the same sign, then the direction of the beam maximum, Ψ_m, as a function of the wavelength will be

$$\Psi_m = \arcsin \frac{1}{\lambda} \frac{\theta}{2\pi} \quad . \tag{2}$$

Eq. (1) may be expressed as

$$S_\lambda(\Psi) = \sum_{n=-N}^{n=N} A_n \, e^{j\gamma n \ell} \quad , \tag{3}$$

where

$$\gamma = \frac{2\pi}{\lambda} (\sin \Psi - \sin \Psi_m) \quad . \tag{4}$$

The space factor may be generalized and expressed in the following form

$$S_\lambda(\Psi) = \int_{-\frac{L}{2}}^{\frac{L}{2}} A(z) \, e^{j\gamma z} \, dz \quad , \tag{5}$$

where the length of the array is $L = 2N\ell$, and the illumination function, $A(z)$, may either be a continuous or discrete function. As pointed out by Spencer[7], in this form the space factor is simply the Fourier transform of the antenna illumination function. On a steady state basis, the time dependent response is

$$S_\lambda(\Psi,t) = e^{j\omega t} \int_{-\frac{L}{2}}^{\frac{L}{2}} A(z) \, e^{j\gamma z} \, dz \quad , \tag{6}$$

where ω is the signal angular frequency.

In the transient situation, the signal first excites the elements near the feed end and then progresses to the rest of the antenna. At the time t, it may be that a wave group has excited the portion extending to

$$z = \frac{L}{T} t \quad , \tag{7}$$

where T is the total transit time of the group through the array. Since the extent of the array in space is $-\frac{L}{2} < z < \frac{L}{2}$, the time in which a wave group will be present in the array is $-\frac{T}{2} < t < \frac{T}{2}$.

In general, the array will be excited by a signal

$$e_1(t) = f_1(t) \, e^{j\omega_1 t} \quad , \tag{8}$$

where the absolute value of $f_1(t)$ is the signal amplitude, the argument of $f_1(t)$ is the phase, and ω_1 is the center frequency. The interval of the signal near $t = 0$ (Fig. 2a) proceeds through the antenna prior to some later interval near t_1. In circuit analysis, it is often convenient to invert the signal[8] (Fig. 2b) using the notation ξ for the time base and picture the signal as coming from the left ($\xi \rightarrow -\infty$) and moving to the right ($\xi \rightarrow +\infty$). In this form, the signal amplitude and phase is represented by $f_1(t-\xi)$. The signal at time, t_1, is shown in Fig. 2b. For $t<t_1$ the signal is further to the left, and for $t>t_1$ the signal has moved further to the right.

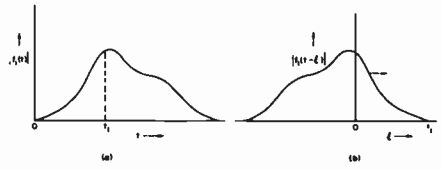


Fig. 2

Inversion of signal time dependence.

With the signal portrayed in this manner, one can visualize the signal passing through the array (Fig. 3) which is also bounded in time between $-\frac{T}{2} < \xi < \frac{T}{2}$. The illumination function is represented in the time domain in the form $A(\xi)\, u(\frac{T}{2} - \xi)\, u(\frac{T}{2} + \xi)$, where $u(t)$ is the unit step function which is defined as

$$u(t) = \begin{cases} 0 & t < 0 \\ 1 & t > 0 \end{cases} \quad . \tag{9}$$

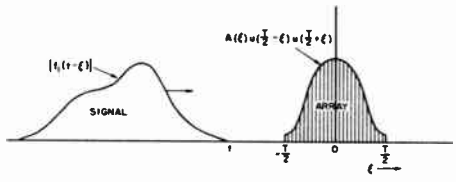


Fig. 3

Pictorialization of signal traversing array in the time domain.

When the signal traverses through the array the transient space factor as a function of time may be represented as

$$S(\Psi,t) = \int_{-\infty}^{\infty} f_1(t-\xi) e^{j\omega_1(t-\xi)} A(\xi) u(\frac{T}{2} - \xi)$$

$$U(\tfrac{T}{2}+\xi)e^{j\omega\xi}\,e^{j\Gamma\xi}\,d\xi\,, \qquad (10)$$

where

$$\Gamma = \frac{2\pi L}{\lambda T}(\sin\Psi - \sin\Psi_m)\,, \qquad (11)$$

and the relation between the time and space extent of the array is used. In the formulation of Eq. (10), it has been assumed that each radiator is instantaneously excited to the value given by $A(\xi)$. The secondary effects due to multi-reflection between the elements of the array and the finite build-up time the associated radiation field have been neglected since of prime interest here is the first order effect due to the finite transit time of a signal through the array.

Frequency Domain Characteristics

The time domain representation of the pattern build-up as given by Eq. (10) may be expressed in the frequency domain by means of the Fourier transform which gives

$$F(\Psi,\omega) = \int_{-\infty}^{\infty} S(\Psi,t)\,e^{-j\omega t}\,dt \qquad (12)$$

In extenso,

$$F(\Psi,\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_1(t-\xi)e^{j(\omega-\omega_1)\xi}\,A(\xi)\,u(\tfrac{T}{2}-\xi)u(\tfrac{T}{2}+\xi)e^{j\Gamma\xi}\,e^{-j(\omega-\omega_1)t}\,d\xi\,dt. \qquad (13)$$

Since $A(\xi)u(\frac{T}{2}-\xi)u(\frac{T}{2}+\xi)$ and $f_1(t)e^{j\omega_1 t}$ are assumed to be Fourier transformable, both integrals are absolutely convergent and the order of integration may be reversed. Changing the order of integration,

$$F(\Psi,\omega) = \int_{-\infty}^{\infty} e^{j(\omega-\omega_1)\xi}A(\xi)u(\tfrac{T}{2}-\xi)u(\tfrac{T}{2}+\xi)e^{j\Gamma\xi}\int_{-\infty}^{\infty} f_1(t-\xi)e^{-j(\omega-\omega_1)t}\,dt\,d\xi. \qquad (14)$$

Substituting $\tau = t-\xi$,

$$F(\Psi,\omega) = \left[\int_{-\infty}^{\infty} A(\xi)u(\tfrac{T}{2}-\xi)u(\tfrac{T}{2}+\xi)e^{j\Gamma\xi}\,d\xi\right]\left[\int_{-\infty}^{\infty} f_1(\tau)e^{j\omega_1\tau}\,e^{-j\omega\tau}\,d\tau\right] \qquad (15)$$

which may be expressed as

$$F(\Psi,\omega) = F_A(\Psi,\omega)\,F_1(\omega)\,, \qquad (16)$$

$F_A(\Psi,\omega)$ and $F_1(\omega)$ are the Fourier transforms of the time dependent illumination, $A(\xi)u(\frac{T}{2}-\xi)u(\frac{T}{2}+\xi)e^{j\Gamma\xi}\,e^{j\omega t}$, and the input signal $f_1(t)e^{j\omega_1 t}$, respectively. The former expression is merely the antenna pattern as expressed by Eq. (5).

$F_A(\Psi,\omega)$ can be regarded as the space transfer function of the antenna in the frequency domain. It is a very familiar fact in circuit theory that the output spectrum of a linear network, expressed in terms of the Fourier transform, is the product of the network transfer function and the input spectrum. Eq. (16) illustrates this same relationship for the output of the array. In fact, it may be said that the antenna acts as a filter whose characteristics are dependent on the direction of observation in space.

Uniformly Excited Array

As an example of the above, consider a uniformly excited array with $A(\xi) = \frac{1}{T}$. The space transfer function, $F_A(\Psi,\omega)$, as given from Eqs. (15) and (16) is

$$F_A(\Psi,\omega) = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} e^{j\Gamma\xi}\,d\xi\,, \qquad (17)$$

which when integrated yields

$$F_A(\Psi,\omega) = \frac{\sin\frac{T\Gamma}{2}}{\frac{T\Gamma}{2}} = \frac{\sin\frac{L\gamma}{2}}{\frac{L\gamma}{2}}\,. \qquad (18)$$

This, of course, is the space factor for the uniform array assuming a continuous distribution for the purpose of simplicity.

If in this example, assume that the array is fed by a transmission line carrying a TEM wave which travels in a straight line along the array direction. If the elements are successively excited 180 degrees out of phase with the adjacent elements, the successive phase shift, Θ, is

$$\theta = \frac{2\pi \ell}{\lambda} - \pi \quad . \tag{19}$$

And in this case, Eq. (4) becomes

$$\gamma = \frac{2\pi}{\lambda}\left[\sin \Psi - \left(1 - \frac{\lambda}{2\ell}\right)\right] \quad . \tag{20}$$

The beam is broadside ($\Psi_m = 0$) for the wavelength where $\lambda = \lambda_o = 1/2$ or the frequency where $\omega_o = \frac{\pi c}{\ell}$, where c is the velocity of light.

Let us take the broadside angle ($\Psi = 0$) as the angle of observation in considering the filtering characteristics. Combining Eqs. (18) and (20) we have

$$F_A(o,\omega) = \frac{\sin \frac{L}{2}\left(\frac{\omega}{c} - \frac{\pi}{\ell}\right)}{\frac{L}{2}\left(\frac{\omega}{c} - \frac{\pi}{\ell}\right)} \tag{21}$$

Figure 4 illustrates the filtering characteristics of the antenna on a steady state basis. Viewed in this manner, the antenna may be said to have a bandpass filter characteristic. In the instances more general than this end-fed array, the bandwidth will depend on the nature of the feed structure and the type of antenna concerned. Thus, of one knows the C-W characteristics (both amplitude and phase dependence) for any antenna, the transient response may be synthesized by the weighed linear superposition of the frequency components of the input signal.

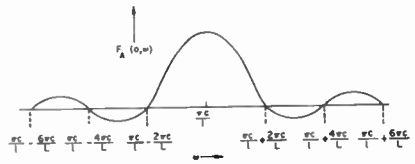


Fig. 4

Filtering characteristics of a simple uniformly illuminated end-fed linear array in the broadside direction.

Receiving Antenna

In a similar fashion, we may deal with the transient situation for a receiving antenna. The received signal measured at the antenna feed will result first from the excitation reaching the nearest radiator to it. The effect of the radiators nearer the termination will be delayed by the transit time through the feed structure. In addition, the progressive phase shift associated with the direction of arrival will modify the vector addition of the contribution of the various array elements to the signal. If the arriving signal is

$$e_2(t) = f_2(t)e^{j\omega_2 t} \quad , \tag{22}$$

and the antenna characteristics are as previously given, the transient space factor on reception, $S'(\Psi,t)$, will be

$$S'(\Psi,t) = e^{j\omega_2 t}\int_{-\infty}^{\infty} f_2(t-\xi)e^{j(\omega-\omega_2)\xi} A(\xi)u\left(\frac{T}{2} - \xi\right)u\left(\frac{T}{2} + \xi\right)e^{j\Gamma\xi}\, d\xi, \tag{23}$$

which is identical with the transmission situation as expressed by Eq. (10). This, of course, should be suspected since antennas are linear and the reciprocity theorem holds. In consequence, the space transfer function, $F_A(\Psi,\omega)$, will be the same for both transmission and reception.

Antenna in Radar System

As an illustration of the limitations imposed by the transit time, consider a radar antenna using a uniform array that is excited by a rectangular r-f pulse of duration τ. It will be of interest to consider the transient characteristics when the pulse lengths are ten times, equal to, and one-tenth the transit time through the antenna array (Fig. 5).

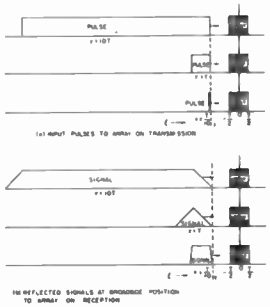


Fig. 5

Signal envelopes to uniformly excited array on transmission and reception.

For this example, we have

$$e_1(t) = u(t)u(\tau - t)\, e^{j\omega_1 t} \tag{24}$$

and

$$A(\xi) = 1 \quad . \tag{25}$$

Let the beam maximum be at broadside, $\Psi_m=0$, for $\omega=\omega_1$. Thus the transient space factor on transmission according to Eq. (10) is

$$S(\Psi,t) = e^{j\omega_1 t} \int_{-\infty}^{\infty} u(t-\xi)u(\tau-t+\xi)$$

$$u(\frac{T}{2}-\xi)u(\frac{T}{2}+\xi)e^{j\Gamma_1\xi}\, d\xi, \tag{26}$$

where

$$\Gamma_1 = \frac{2\pi L}{\lambda_1 \tau} \sin \Psi \quad . \tag{27}$$

The three cases are shown in Fig. 5a for the situation described by Eq. (26). For broadside position, the transmitted signal envelopes are as shown in Fig. 5b. The response of the array to the reflected signal is

$$S'(\Psi,t) = \int_{-\infty}^{\infty} S(\Psi,t-\xi)$$

$$u(\frac{\tau}{2}-\xi)u(\frac{\tau}{2}+\xi)\, e^{j\Gamma_1\xi}\, d\xi, \tag{28}$$

with the same angle of observation for transmission and reception assumed.

The signal envelops at various observation angles are shown for the three cases in Fig. 6-9. Illustrated are the envelops of the transmitted signal, received signal, and the output of a six-stage radar i-f amplifier. It is assumed that the radar signal is reflected from a simple target. The analysis of the response of the i-f amplifier is given in Appendix A. For the first case, $\tau = 10T$, the overall i-f bandwidth is assumed to be that which will give a sufficiently rapid response and the most favorable signal detectability for the conventional radar circumstance[9], namely $1.2/\tau$. For the second case, $\tau = T$, the i-f bandwidth is assumed to be $1.2/\tau$ or $0.6/\tau$. For the last case, $\tau = T/10$, $1.2/\tau$ and $0.12/\tau$ i-f bandwidths are assumed.

The first case (Fig. 6) illustrates the circumstance where the pulse length is appreciably greater than the transit time through the antenna. The peaks of the signals are of interest in radar systems. Hence, the signal peaks of the transient on transmission and reception are compared to the steady response as function of the observation angle in Fig. 6f. The transients at the beginning and the end of the signal are responsible for degradation of the pattern. In addition to the amplitude fluctuation during the transient period, there is a phase fluctuation. Both the amplitude and phase vary with greater rapidity as the observation angle departs from broadside. Although detrimental signal peaks may affect the antenna discrimination, the i-f filtering associated with normal practice does eliminate these peaks since the receiver response is slow compared to the changes in amplitude and phase during the transient period.

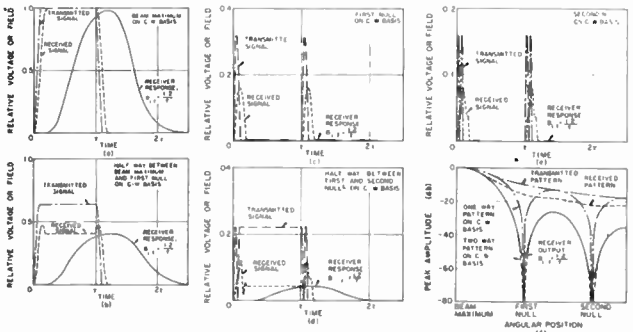


Fig. 6

Response and pattern characteristics for a uniformly illuminated array excited by an rf pulse length ten times the transit time through array.

In the second case, where the pulse length and transit time are equal, the transients as shown in Fig. 7 are similar to those of the previous case; however, the antenna pattern transient build-up phase is immediately followed by the transient decay phase. The antenna pattern has no steady state phase. The pattern of the peaks of the transmitted and received

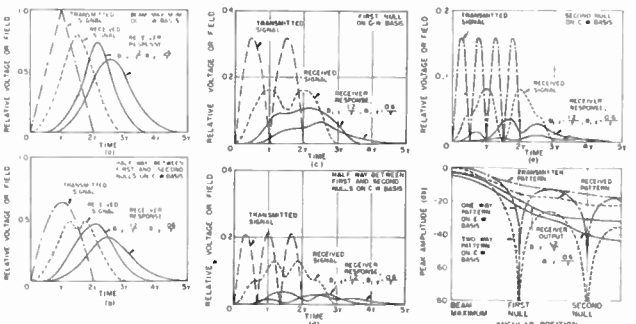


Fig. 7

Response and pattern characteristics for a uniformly illuminated array excited by an rf pulse length equal to the transit time through array.

signal (Fig. 7f) are the same as in the previous case. However, after filtering with the requisite i-f bandwidth, $1.2/\tau$, associated with a chosen pulse length, the pattern still suffers an appreciable deterioration since the receiver response is sufficiently rapid to be able to respond to the amplitude and phase fluctuations of the returning signals away from the main beam direction. Narrowing down the i-f bandwidth to $0.6/\tau$ does not improve the antenna pattern very much.

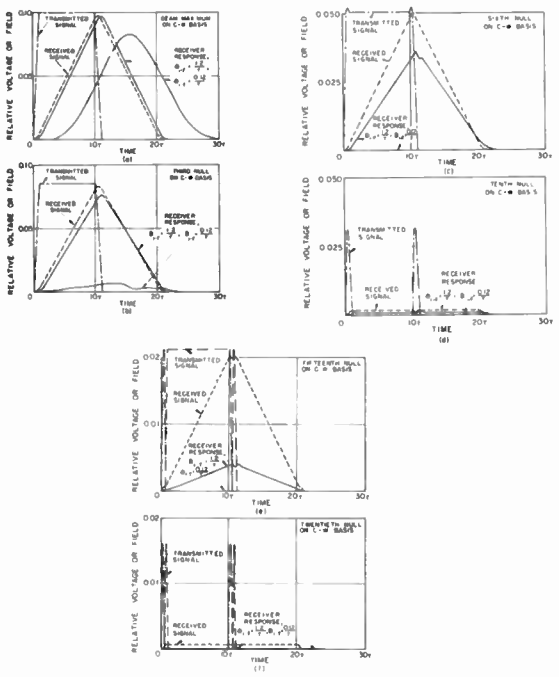


Fig. 8

Response characteristics for a uniformly illuminated array excited by an rf pulse length 1/10 the transit time through array.

Fig. 8 illustrates the response of the antenna and receiver for the case where the transit time through the array is ten times the transmitter pulse length. At no time during the pulse is the complete array excited and hence the full gain of the antenna is never realized. The transient peak of the two-way pattern in the direction of the main beam is reduced 20db with respect to the pattern on a C-W basis. In addition to this loss of gain, the pulse duration is stretched out to be more nearly that of the transit time. The overall pattern of the peaks (Fig. 9) is highly degraded for the transmitted and received signals as well as the output of the i-f amplifier whose bandwidth is $1.2/\tau$. The effective beamwidth using the $1.2/\tau$ receiver is about ten times that of the antenna on a C-W basis. Narrowing the receiver bandwidth to $0.12/\tau$, considerably improves the pattern without very much affecting the received pulse length. It is quite obvious that use of a pulse of this length originally would have given a comparable pattern without the loss in antenna gain.

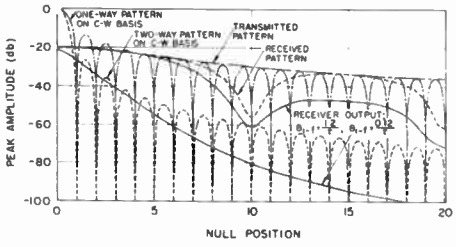


Fig. 9

Pattern characteristics of a uniformly illuminated array excited by an rf pulse length 1/10 the transit time through array.

The above situations may be represented in the frequency domain by the application of the results expressed in Eq. (18). For the r-f pulse, the spectrum is

$$F_1(\omega) = \tau\, e^{-j\frac{\tau}{2}(\omega-\omega_1)}\, \frac{\sin\frac{\tau}{2}(\omega-\omega_1)}{\frac{\tau}{2}(\omega-\omega_1)} . \quad (29)$$

Combining Eq. (29) with the transfer function of the antenna, Eq. (18), the transmitted spectrum is

$$F(\Psi,\omega) = \left(\frac{\sin\frac{L\gamma}{2}}{\frac{L\gamma}{2}}\right)\tau\, e^{-j\frac{\tau}{2}(\omega-\omega_1)}\, \frac{\sin\frac{\tau}{2}(\omega-\omega_1)}{\frac{\tau}{2}(\omega-\omega_1)} . \quad (30)$$

The received spectrum is

$$F'(\Psi,\omega) = \left(\frac{\sin\frac{L\gamma}{2}}{\frac{L\gamma}{2}}\right)^2\tau\, e^{-j\frac{\tau}{2}(\omega-\omega_1)}\, \frac{\sin\frac{\tau}{2}(\omega-\omega_1)}{\frac{\tau}{2}(\omega-\omega_1)} , \quad (31)$$

with $\dot{\gamma}$ given by Eq. (20). The center frequency of the i-f amplifier will differ from the center frequency of the transmitter pulse. However, the filtering characteristic may be shifted along the frequency axis just as the mixing process does this in the equipment. On this basis the output spectrum after the i-f amplifier may be found by multiplying $F'(\Psi,\omega)$ by its transfer function.

Viewed from the frequency domain, it can be seen how the bandpass characteristics of the measurement device (receiver) limits the region of interest of the transfer characteristics of the antenna. As the pulse length becomes smaller relative to the transit time, its spectrum width becomes larger relative to an antenna lobe on the frequency scale. Where the pulse length is less than the transit time, the antenna transfer characteristic will have a narrower spectrum than that of the pulse and significantly modifies the pulse shape in the direction of the beam.

Appreciation is expressed to Dr. Nicholas A. Begovich, and Edward J. Rhoad for the many helpful discussions on this subject and to the Central Group of mathematicians of the Laboratories for the numerical results of this paper.

Appendix A

The response of an i-f amplifier to a unit impulse, $g(t)$, may be given in the form

$$g(t) = g_o(t)\, e^{j\omega_o t} . \qquad (A1)$$

If the amplifier is excited by a signal

$$e(t) = f_o(t)\, e^{j\omega_o t} , \qquad (A2)$$

the output of the network will be given by[8]

$$h(t) = \int_{-\infty}^{t} e(t-\xi)g(\xi)\, d\xi$$

$$= e^{j\omega_o t} \int_{-\infty}^{t} f_o(t-\xi)g_o(\xi)\, d\xi . \qquad (A3)$$

For an n-stage synchronously tuned amplifier, the impulse response may be approximated by

$$g(t) \cong \frac{\left(\frac{\omega_B t}{2}\right)^{n-1}}{(n-1)!}\, e^{-\frac{\omega_B t}{2}}\, e^{j\omega_o t} , \qquad (A4)$$

where ω_B is the bandwidth in radians per second for a single stage of the amplifier and related to the overall i-f bandwidth, $\omega_{i\text{-}f}$, by

$$\omega_B \cong 1.2\sqrt{n}\ \ \omega_{i\text{-}f}. \qquad (A5)$$

When Eq. (A4) is substituted in Eq. (A3) we have a form that may be applied to the received signal

$$h(t) \cong e^{j\omega_o t} \int_{-\infty}^{t} f_o(t-\xi)\, \frac{\left(\frac{\omega_B \xi}{2}\right)^{n-1}}{(n-1)!}\, e^{-\frac{\omega_B \xi}{2}}\, d\xi . \qquad (A6)$$

References

1. H. H. Beverage and H. O. Peterson, "Diversity Receiving System of RCA Communication Inc. for Radiotelegraphy," Proc. I.R.E., vol. 19, pp 531-562; April, 1931.

2. W. H. Watson, "The Physical Principles of Waveguide Transmission and Antenna Systems," Oxford University Press, 1947.

3. Ralph Grimm, "The Comb Antenna," Proc. I.R.E., vol. 36; pp 359-362, March, 1948.

4. S. Silver, "Microwave Antenna Theory and Design" Vol. 12, M.I.T. Rad. Lab. Series, Chap. 9, McGraw-Hill, 1949.

5. O. O. Fiet, "Ultra-High Frequency Antenna and System for Television Transmission," R.C.A. Rev., vol. 11, pp 212-227; June, 1950.

6. S. A. Schelkunoff and H. T. Friis, "Antennas Theory and Practice," John Wiley, 1952.

7. R. C. Spencer, "Fourier Integral Methods of Pattern Analysis," M.I.T. Rad. Lab. Report 762-1; Jan. 21, 1946.

8. Gardner and Barnes, "Transients in Linear Systems," John Wiley, 1942, p 231.

9. J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals," M.I.T. Rad. Lab. Series Vol. 24, McGraw-Hill, 1950; p 201

A NEW MICROWAVE REFLECTOR

K. S. Kelleher
Antenna Research Branch
Naval Research Laboratory
Washington 25, D. C.

The Reflector

The reflector under consideration has been designated as a parabolic torus. It is a part of a surface of revolution obtained by rotating an arc of a parabola about a line parallel to the latus rectum as shown in Figure 1. Only the upper half of the parabola is used here in order to avoid the usual problems of feed blocking and mismatch.

Since the surface was formed by rotating the arc about the line AB, it is reasonable to expect no change in performance as a directive feed is rotated about that line. If it were possible to produce a good radiation pattern at one feed position, the same characteristics would be obtained at other positions. This surface then has a property not possessed by the standard paraboloids, whose performance deteriorates when the feed is moved from the focus.

Reflector Types

There are two general types of parabolic torus reflectors. The first is obtained from a rotation of the parabolic arc through about 120 to 180 degrees and is used whenever feed motion through angles of as much as 90 degrees is desired. In this case only a portion of the reflector is illuminated at any one time. The second type of torus employs a rotation of the parabolic arc through about 70 degrees. The entire reflector is illuminated at all times and so is used efficiently. The angle of scan in this case is limited by the increase in "coma" lobe in a manner similar to that found for the paraboloid and sphere[1].

Figure 2 shows photographs of the two reflector types. In both cases, the parabolic arc has been rotated about a horizontal axis. In the case of the fabricated reflector, each screen support member takes the shape of the parabolic arc. This reflector has a rotation angle of 130 degrees while the fibreglas reflector has an angle of 68 degrees.

1 K.S.Kelleher and H.P.Coleman, "Off-Axis Characteristics of the Paraboloidal Reflector", NRL Report #4088, December 31, 1952

Wide Angle System

Experimental Results

The wavefront reflected from the parabolic torus is not plane, but can be made to approach a plane by a proper choice of the ratio of the focal length, f, to the radius of the system, R. The wide angle reflector shown in Figure 2 had a focal length of 9" and a radius of 19.6" (f/R = 0.46). Patterns obtained at X-band from this reflector are shown in Figure 3 where they are compared to patterns from a paraboloid. It is evident that the parabolic torus is far superior to the paraboloid as a scanning reflector. The decrease in gain at the position of 40 degrees off center is due to the fact that for this position some fraction of the energy from the feed horn was not reflected by the torus. If greater coverage than the ± 40 degrees shown here is required, the angle of rotation should be increased by a corresponding amount. For example, a 90 degree coverage would be obtained from a reflector whose rotation angle was 140 degrees.

Applications

Besides the obvious application of this reflector as a wide-angle scanning-antenna, it appears to be useful as a "beam bender" in a microwave link. The present systems, which might be used for this purpose, consist of a plane reflector or of a reflector which is a section of a paraboloid. They appear to be limited either by the fact that the angle of bending is always 90 degrees or that different sections of a paraboloid might be required in order to obtain different angles of bend. With a parabolic torus reflector and two feed horns, the angle of bending can be made arbitrary. The angle between the incident and reflected beam is equal to the angle formed between the feed horn axes.

A second application is found when it is desired that a single control station should communicate with several substations. The parabolic torus would be mounted so that its circular section was in the horizontal plane. The several feed horns, corresponding

to the several substations, would be mounted at proper positions in that plane. Through this means, a group of reflectors is replaced by a single reflector with a group of feeds.

Medium Angle Systems

Experimental Results

The medium angle (fibreglas) reflector of Figure 2 had a focal length of 13" and a radius of 29" (f/R = 0.45). The angle of rotation was 68 degrees.

Patterns obtained at X-band are shown in Figure 3 where they are compared with patterns from a paraboloid of approximately equal f/D ratio. The torus is somewhat superior off-axis although its beamwidth is slightly broader. When the torus is scanned to a greater angle than shown here, the side lobe on the left ("coma" lobe) increases in intensity.

For longer f/D ratios the paraboloid and the parabolic torus become almost identical so that a point is reached at which there is little advantage in the use of the torus.

Applications

This type of parabolic torus can be used in applications where a scan somewhat better than a paraboloid is desired. One possible use occurs in marine-navigation radar. Here the vertical beamwidth of present systems is deliberately broadened so that a surface target is covered for any position of ship's roll. With a parabolic torus reflector, a narrow vertical beam, with resulting increase in system gain, could be used. The feed horn would be synchronized with the ship's motion so that it always assumed a position to hold the narrow beam on the surface target. Through use of the increased gain, such a system might trade electronic complexity for mechanical complexity and so be more acceptable to the average ship's engineer.

Parabolic Torus Limitations

At the present time, after 18 months work, two fundamental limitations of the parabolic torus have been found. They are both traced to the fact that a perfect plane wave cannot be obtained.

The first of these concerns the beamwidth in the plane of scan. For beams narrower than 3 degrees, the relationship between beamwidth and aperture size is no longer linear. There is a decrease in aperture efficiency as the beamwidth decreases. For this reason, it is considered impractical to obtain a beamwidth of the order of 1 degree in the plan of scan.

The second limitation involves the pattern structure. Energy at levels below about 15 db does not drop off rapidly as in the case of the paraboloid. This is not evident in the principal planes of the pattern as much as in other regions. This effect can best be seen by reference to the contour plot of a pattern given in Figure 5. This contour corresponds to the 0 degree pattern of Figure 4. Each scale division represents one degree. The figures along the abscissa gives the values of the contours in decibels. The abscissa represents the plane of scan, while the ordinate corresponds to the other principal plane. It can be seen that the contours after 12 db down do not fall closer and closer together as in the case of the paraboloid. This energy is at a low enough level that it does not greatly affect the gain of the antenna, but it does give a broadening of the "base" of the beam which may be harmful in some applications.

Summary

This paper describes a new microwave reflector which appears to have application to scanning problems and problems in which multiple beams from a single reflector are necessary. The wide angle type of reflector is far superior to existing reflectors while the medium angle type shows sufficient superiority to warrant its consideration.

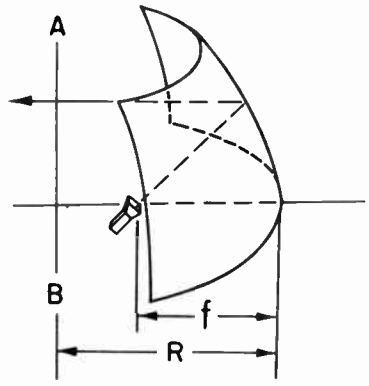
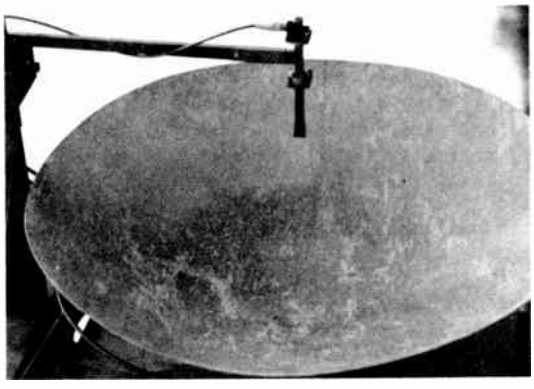


Fig. 1
Geometry of parabolic torus.



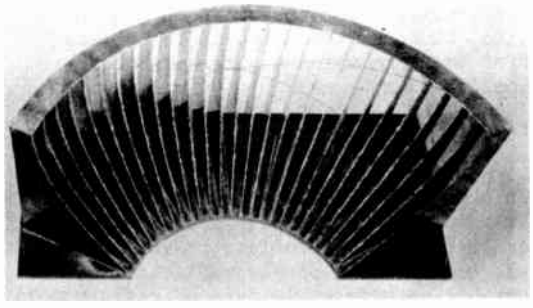


Fig. 2
Two types of parabolic torus reflectors.

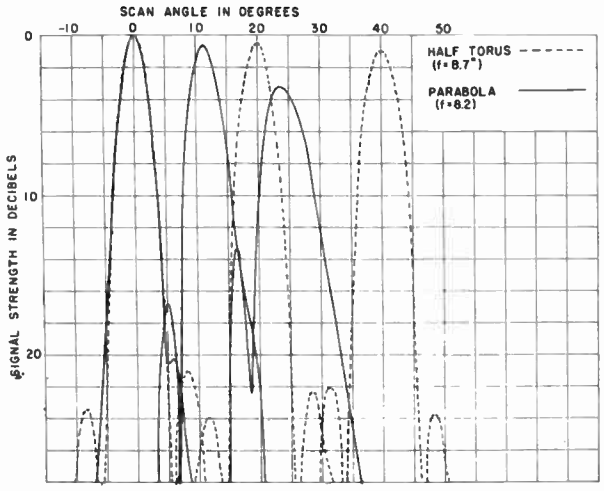


Fig. 3
Patterns of paraboloid and wide-angle parabolic torus.

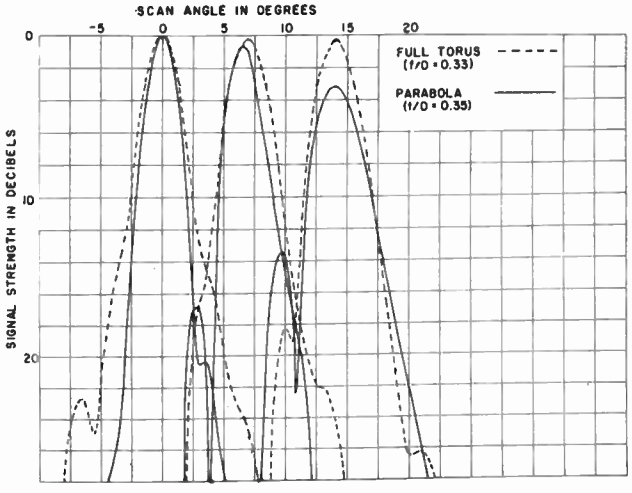


Fig. 4
Patterns of paraboloid and medium-angle parabolic torus.

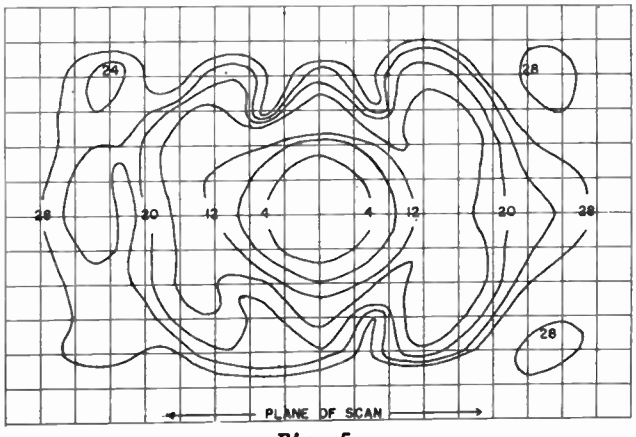


Fig. 5
Contour plot of medium-angle parabolic-torus pattern.

CROSSTALK IN RADIO SYSTEMS CAUSED BY FOREGROUND REFLECTIONS

H. W. Evans
Bell Telephone Laboratories, Inc.

In radio systems, just as in wire systems, one of the major limitations to transmission performance is crosstalk. In wire systems crosstalk is for the most part man-made, in that it is a characteristic of the wire lines and cables employed. But in radio systems, crosstalk sometimes involves nature.

Two of the major paths of crosstalk in radio relay systems utilizing a two-frequency allocation such as the Bell System's TD-2 Radio Relay System are shown in Figure 1. These paths cause signals from one direction of transmission to enter the opposite direction of transmission, and result in crosstalk appearing as noise in the message circuits or television circuit transmitted on the radio channel. These unwanted signals may arrive via the true front-to-back ratio of the antenna, or via reflections from geographical features in the antenna foreground. True front-to-back ratio is taken to mean the discrimination of the antenna between wanted signals arriving on the main forward directive lobe and unwanted signals arriving from the rear of the antenna. These two sources in combination form what is called for convenience "effective front-to-back ratio".

Thus both interferences are weaker in amplitude than the wanted signal by the effective front-to-back ratio, and the lower the effective front-to-back ratio, the greater the resulting noise. This noise cannot be reduced by increasing transmitter power or by decreasing receiver noise figure, since it is a characteristic of the antenna and its site, so these crosstalk paths place a definite limit on the level differences that can be allowed to exist across a repeater. Front-to-back ratio is therefore an important factor in specifying such fundamental radio system parameters as radio-frequency power, noise figure, maximum length of path, etc., and is a critical factor in choosing between a two-frequency and a four-frequency allocation plan.

In laying out any radiotelephone system, the total noise objective per repeater is allocated among all the various known noise sources, such as fluctuation noise, intermodulation noise, noise increases caused by fading, noise caused by such transmission imperfections as impedance discontinuities in wave-guide and cables, and noise resulting from radio-frequency crosstalk. This total noise is divided among the known noise contributors in such a manner as to maintain a practical and economic balance between the various noise sources so that any one source does not control the overall noise performance of the system. Consideration of any specific means of system improvement, such as higher power radio transmitters or lower noise radio receivers, disturbs the original balance, and a critical reexamination of the other noise contributors is indicated. Accordingly, when a program of improvement of the Bell System's TD-2 Radio Relay System was contemplated, it became essential to examine more critically the noise contribution resulting from radio-frequency crosstalk.

Measurements of a prototype antenna of the TD-2 Radio Relay System under idealized conditions showed the directional characteristic appearing in Figure 2. It will be seen that its response to signals arriving from its rear is 68 db below its response to signals arriving directly on its main lobe. Thus its true front-to-back ratio is 68 db. Actually its rearward response is not smooth, but is composed of a veritable chrysanthemum of narrow lobes, so that some antennas might be expected to operate at the 68 db envelope of the peaks of the lobes, while others might operate at higher front-to-back ratios, depending on the angle between the radio paths in front of and behind the antennas.

But the front-to-back ratio of a working antenna may be quite different from that of a prototype antenna measured under ideal conditions. Objects in the foreground of the antenna cause reflections which interfere with the determination of the directional properties of the antenna itself, and it was to circumvent this source of error that A. B. Crawford and W. M. Sharpless of the Holmdel Radio Laboratory first applied the radar principle to antenna testing. It was found that objects such as buildings and trees affected the apparent directional characteristic even when they were some thousands of feet away. Similar reflections from foreground objects are present in a chain of repeaters, and it is not enough to know the free-space characteristic of the isolated antenna. Serious crosstalk can be caused by the surrounding terrain.

Measurements of the effective front-to-back ratio of 19 antenna of the New York-Chicago TD-2 Radio Relay System were made by c-w techniques. At the time the measurements were made, there were three radio channels on this route operating in the 3700 and 4200 mc band, and each antenna was measured at the frequency of each radio channel, so that some 57 measurements were made. The results of these measurements are shown in arithmetical distribution in Figure 3. When an antenna is measured in place by c-w techniques, it is impossible to separate true front-to-back ratio from the contribution of foreground reflection, so that the values shown are a combination of true front-to-back ratio and foreground reflection, or effective front-to-back ratio. Furthermore the

signal received from the antenna's rear was not constant, but varied constantly over a 10 db range in a manner judged to be normal law. The root-mean-square front-to-back ratio over a period of time is thought to be about 5 db greater than the value shown.

It will be seen that 25 per cent of the antennas measured higher than the 68 db observed on the prototype antenna, with some measuring as high as 78 db, indicating that these antennas were probably working on the interstices of the lobes composing the rearward response with little degradation from foreground reflection. But some of the antennas measured as low as 54 db effective front-to-back ratio, and indeed the rms effective front-to-back ratio of all the antennas measured including the 5 db factor mentioned above was 64 db, 4 db poorer than that of the prototype. No significant differences were found between the front-to-back ratios measured at the three different frequencies on each antenna, and, as was expected, no significant differences were found between transmitting and receiving antennas, which are mounted side by side on the repeater stations as shown in Figure 4. No significant difference was observed in the short term variation in front-to-back ratio between antennas with high and antennas with low front-to-back ratios.

The poorest of the 19 antennas was selected for further investigation, and a series of tests established that no spurious crosstalk paths existed within the repeater station, such as leaky wave-guide joints, crosstalk between the radio equipments, and leakage around the access door to the interior of the antenna. It was concluded that the signal received from the rear of the antenna was indeed impinging on the antenna face, and that foreground reflection was its source.

Accordingly pulse-type measurements were made on the three poorest of the 19 antennas, using a 3700 mc radar transmitter producing one-half megawatt pulses 1.5 microseconds long. Of course it was necessary to remove all service from the route while these measurements were made. The radio repeater equipment itself was arranged as a radar-type receiver, with display of the received pulses on a TS-239 Oscilloscope. It was found that foreground reflection was indeed the cause of the degradation of front-to-back ratio, and the magnitude of the observed reflections accurately confirmed the c-w measurements.

But although all three antennas were degraded by foreground reflection, the reflecting objects were, interestingly enough, quite dissimilar.

Thus Figure 5 shows the reflections observed on an antenna at Buckingham, Penna. Photograph 19 shows rapidly fluctuating reflections from trees up to some 500 feet in front of the antenna; although the photograph shows reflections some 65 db below the normal signal, these reflections actually peaked at about 58 db below the normal signal, representing a front-to-back ratio contribution of about 58 db. Photograph 21 shows steady reflections from a tree-covered ridge one mile away. These reflections were 56 db below the normal signal, and represent the major contribution to the measured total effective front-to-back ratio of 54 db. When the receiver gain was increased, other lesser reflections were observed, as shown in Photograph 22. When the time scale was increased, as in Photograph 28, other reflections two miles distant were observed.

Study of the geography near the affected antenna from topographical survey maps produced Figure 6, in which the crosshatched areas are those known to be illuminated by the transmitting antenna at Wyndmoor, near Philadelphia, and within the main area of response of the affected antenna at Buckingham. The small numbers show the combined discriminations of the transmitting and receiving antennas. This is shown more clearly in Figure 7. Thus the main reflecting areas are those along the path towards Martinsville. The ridge near the Buckingham station and along the path is actually saddle shaped in contour. The nearby reflections, up to 500 feet from Buckingham station, come from high trees immediately in front of the antenna, but some 20 to 30 feet below the actual beam. The reflections observed at one mile distance come from the far end of the saddle shaped ridge. The reflections some two miles distant come from the small area just inside the three mile radius circle.

Similar observations on an antenna at Wyndmoor, Penna. are shown in Figure 8. These reflections, which were quite steady, come from residential buildings and radio broadcast towers in the neighborhood of the Wyndmoor station.

At New Holland, Penna., the major reflection was found to be from Welsh Mountain, approximately two miles from the affected antenna, as shown in Photograph 51, Figure 9. This reflecting point was confirmed by rotating the transmitting antenna directly towards the reflecting area, whereupon the reflected signal rose 12 db. Although Welsh Mountain's highest elevation is almost 500 feet above the elevation of the path, it is interesting to note that because of vertical directivity of the affected antenna the reflection originates in that area of the mountain which is at the elevation of the path, and when Photograph 51 is compared with topographical survey maps it is found that the elevation of the reflecting area is from about 50 feet below the path to 50 feet above the path.

Photograph 49 in Figure 9 shows the true front-to-back ratio of the New Holland antenna as 75 db. The three small pips at the right are a house, wagon-shed and barn, respectively, on one of the numerous farms in the Susquehanna Valley near New Holland repeater station.

The effect of cross polarization was studied on two of the three antennas tested. The interfering transmitting antenna was changed from its normal vertical polarization to horizontal polarization, and the effect on the reflections was noted. At Buckingham, the nearby reflections were reduced by 3 db, and the more distant reflections by 8 db. At New Holland, the distant reflection was reduced 10 db. These low numbers are in sharp contrast to the 30 odd db cross polarization discrimination that can be realized in direct transmission at these frequencies.

So this study calls to attention another limitation on selection of sites for radio relay repeater stations, already beset by a multitude of interlocking requirements. Reflection troubles can be minimized by carefully scrutinizing foreground geographies of proposed repeater sites for possible reflecting agencies, and by so arranging the transmission paths as to minimize illumination of possible reflectors.

Many members of the Bell Telephone Laboratories technical staff participated in the measurements described here. Particular mention should go to Miss M. K. Corr and Messrs. H. E. Curtis, E. J. Henley, R. L. Robbins, N. F. Schlaack and W. C. Warman.

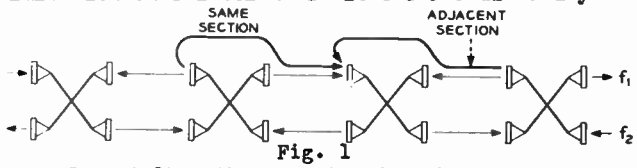


Fig. 1
Crosstalk paths associated with antenna front-to-back ratio in relay systems using two-frequency allocation.

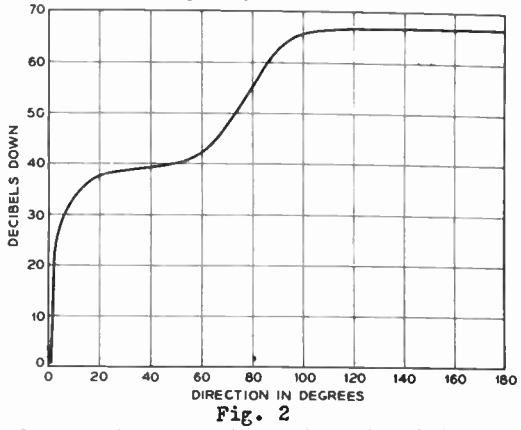


Fig. 2
Envelope of horizontal angular discrimination of horn-lens antenna used in TD-2 radio relay system.

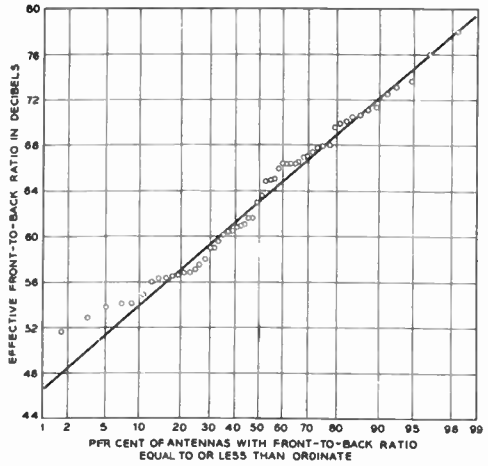


Fig. 3
Distribution of minimum peak values of effective front-to-back ratio of 19 TD-2 antennas.

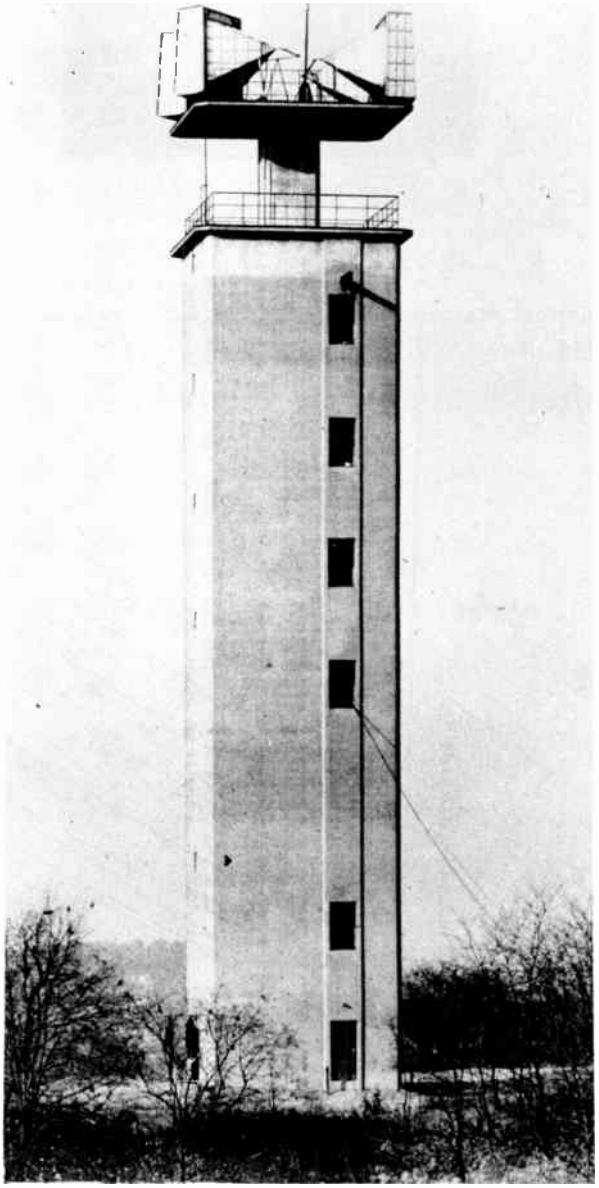


Fig. 4
Typical TD-2 repeater station.

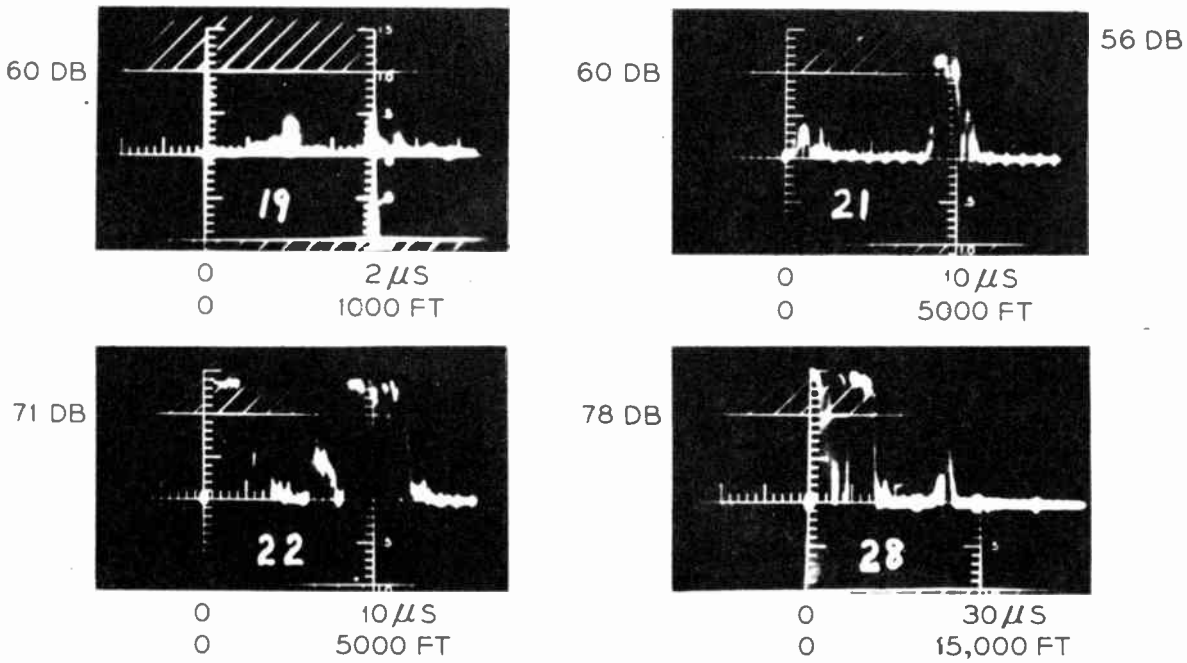


Fig. 5
Reflections observed at Buckingham, Penna.

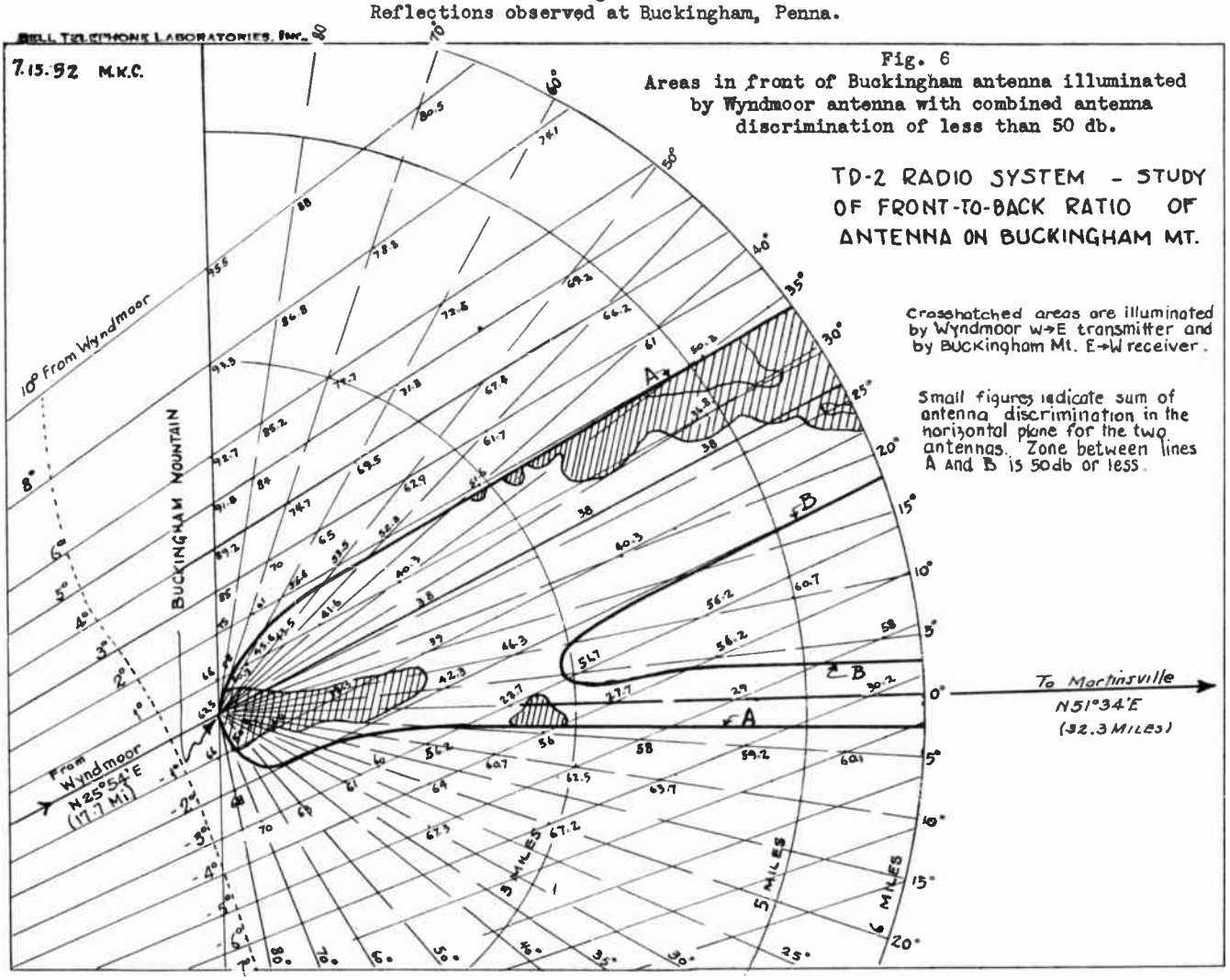


Fig. 6
Areas in front of Buckingham antenna illuminated by Wyndmoor antenna with combined antenna discrimination of less than 50 db.

TD-2 RADIO SYSTEM - STUDY OF FRONT-TO-BACK RATIO OF ANTENNA ON BUCKINGHAM MT.

Crosshatched areas are illuminated by Wyndmoor W→E transmitter and by Buckingham Mt. E→W receiver.

Small figures indicate sum of antenna discrimination in the horizontal plane for the two antennas. Zone between lines A and B is 50db or less.

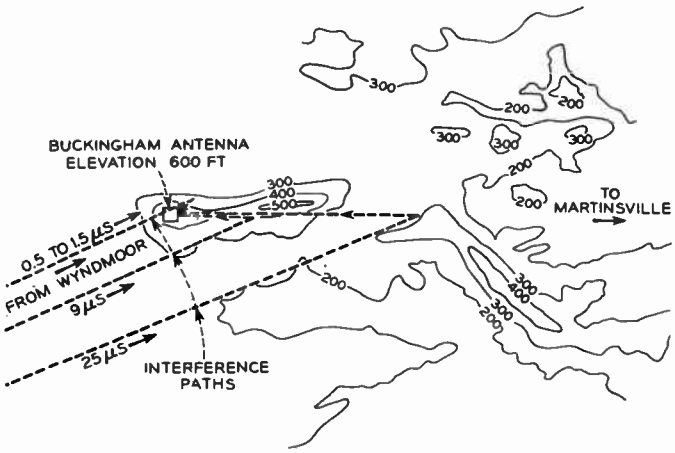


Fig. 7
Topographical detail of sources of reflection at Buckingham.

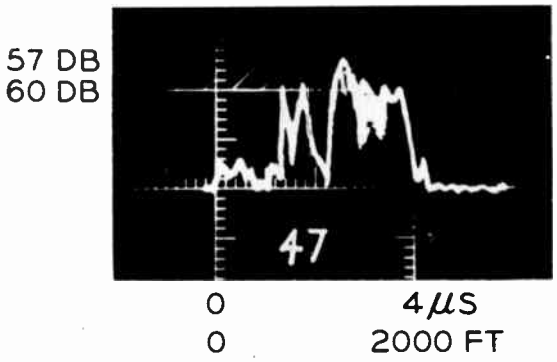


Fig. 8
Reflections observed at Wyndmoor, Penna.

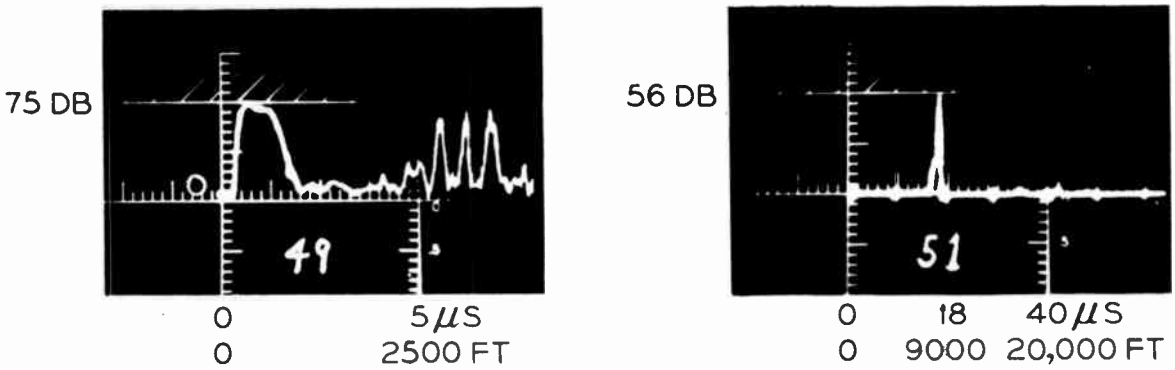


Fig. 9
Reflections observed at New Holland, Penna.

LOW SIDE LOBES IN PENCIL-BEAM ANTENNAS

by
E. M. T. Jones
Stanford Research Institute
Stanford, California

Introduction

This paper describes an investigation of the radiation properties of two perfectly focussing devices: the paraboloid reflector and the hyperboloid dielectric lens excited at their focii by a short electric dipole, a short magnetic dipole, and a plane wave source. The purpose of this investigation is to examine the radiation characteristics of these antennas which, at least in the limit of infinitesimal wavelengths, have no phase aberrations in the aperture field. The analysis is divided into two parts: (1) the aperture fields of the two antennas are first computed, and (2) the far-zone diffraction patterns are then determined. The power gain of the systems is obtained as a by-product of the diffraction pattern computation. Some experimental results are also discussed.

Paraboloid Aperture Distributions

Electric Dipole Feed

Let us first consider the paraboloid excited by an electric dipole. This problem has been previously treated, in similar types of analyses, by E. U. Condon[1] and by L. J. Chu.[2] A picture of this antenna and its feed oriented along the x-axis is shown in Fig. 1 on the left. At the right of the figure the aperture distribution arising from the presence of the electric dipole at the focus is illustrated.

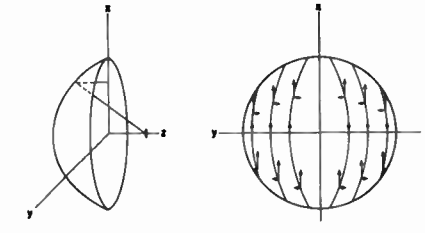


Fig. 1
Paraboloid excited by a short electric dipole feed.

Several assumptions have been made in the derivation of this aperture distribution. They are:

1. The reflector is in the far-zone of the feed, so that only fields varying as the reciprocal of the distance from the feed to the reflector are significant.

2. The feed pattern is the same with the reflector in place as when it is absent.

3. Energy traveling in the region between the reflector aperture and the feed is assumed to travel in the directions predicted by geometric optics. On the other hand, the polarization of the aperture field is determined by the plane wave boundary condition at the reflector surface, namely, that the total tangential electric field in the incident and reflected waves must be zero.

The effects of radiation from currents flowing on the reflector surface in the direction of the z-axis are neglected since they do not radiate in the direction of the z-axis. It is seen that the electric field has a unidirectional component along the x-axis, as well as a cross-polarized component parallel to the y-axis. The cross-polarized component has the interesting symmetry property that it is oppositely directed in adjacent quadrants.

Figure 2 shows the general characteristics of the patterns arising from this aperture distribution that is polarized in both the x- and y-directions. It is seen that the principal polarization patterns measured in the E- and H-planes have the customary shapes with their maxima on the polar axis. On the other hand, the cross-polarization patterns measured in planes at 45° to the principal planes, where the cross-polarization is a maximum, have minimum intensity on the polar axis and maximum intensity off axis. The position of maximum intensity corresponds approximately to the angular position of the first null in the principal-polarization patterns. Because these cross-polarized lobes often have a magnitude considerably higher than the first side lobe of the principal polarization patterns, they can be troublesome in radar applications in which the target reflects energy polarized at right angles to that incident upon it.

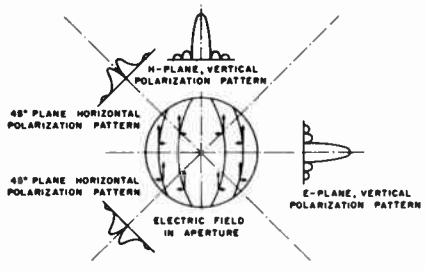


Fig. 2
Far-zone diffraction patterns of a circular aperture with principal and cross-polarized aperture fields.

Magnetic Dipole Feed

Now when the paraboloid is excited by a magnetic dipole oriented along the y-axis as shown at the left of Fig. 3, the aperture fields have the characteristics shown at the right of Fig. 3. We notice that this aperture distribution is somewhat different from that produced by the electric dipole. The electric field still has a unidirectional component along the x-axis. However, the cross-polarized component of electric field parallel to the y-axis is directed oppositely to that produced by the electric dipole.

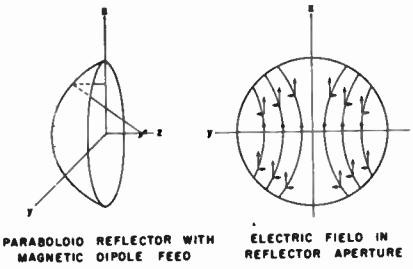


Fig. 3
Paraboloid excited by a short magnetic dipole feed.

Plane Wave Feed

If the paraboloid could be excited by a combination of electric and magnetic dipoles, oriented at right angles to one another and having the proper values of electric and magnetic current, the y-directed component of the aperture field and the cross-polarized lobes in the diffraction pattern can be made to disappear. It turns out that the proper values of electric and magnetic current are just those associated with a plane wave source that has the ratio of electric to magnetic field of 377 ohms. Figure 4 shows the paraboloid excited by a small plane wave source and the resulting aperture distribution, having only an x-directed component of electric field. For a small plane wave source the amplitude of this field is a function of radius only. For larger sources it is a function of both radius and angular position.

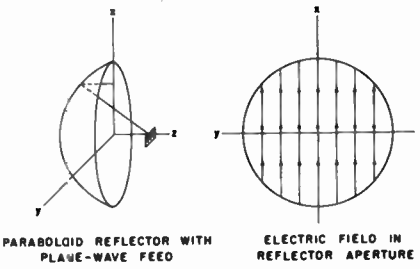


Fig. 4
Paraboloid excited by a plane wave feed.

Hyperboloid Lens

Electric Dipole Feed

The aperture field of the hyperboloid lens excited by an electric dipole is determined in a slightly different manner than the aperture fields of the paraboloid. At the left of Fig. 5 is a sketch of the lens and its electric dipole feed. At the right of this figure is the resulting aperture field which, it is noticed, has no cross-polarized components. This aperture field is computed in the following manner, subject to the conditions outlined previously. First, the field incident on the lens from the feed is resolved into a component in the plane of incidence and one perpendicular to it. The amount of each of these components transmitted through the lens is then computed from plane wave boundary conditions at the dielectric air interface, which are summarized in Fresnel's equations. Finally, these transmitted components are resolved into the x-directed component which is shown at the right of Fig. 5. The assumption is made in deriving this aperture field that the multiple transits of energy in the lens between the two surfaces are unimportant. These transits can be neglected because they give rise to small amplitude and phase variations arranged in concentric rings in the aperture, superimposed on the constant phase amplitude distribution produced by the single transit. In any lens of reasonable size they produce small amplitude diffraction patterns split off at wide angles from the polar axis.

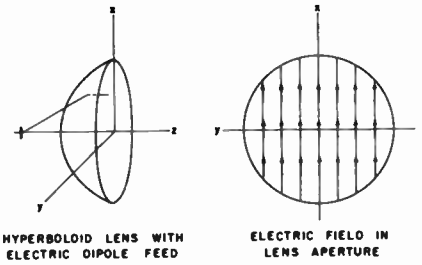


Fig. 5
Hyperboloid lens excited by an electric dipole feed.

Magnetic Dipole Feed and Plane Wave Feed

The aperture field of the hyperboloid lens excited by a magnetic dipole can be derived in the same fashion. At the left of Fig. 6 is shown the aperture field produced by the magnetic dipole feed. It is seen to be similar to that of the paraboloid excited by a magnetic dipole. At the right of Fig. 6 is the aperture field of the lens when it is excited by a plane wave source. Here again this aperture field has cross-polarized components. It turns out that for a given focal length to diameter ratio, the amplitude of the cross-polarized aperture field of the lens excited by a plane wave source is considerably lower than

the cross-polarized field of the paraboloid excited by an electric dipole.

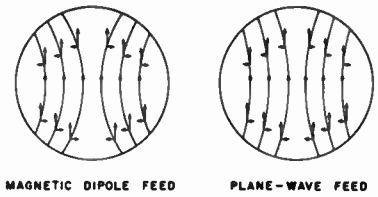
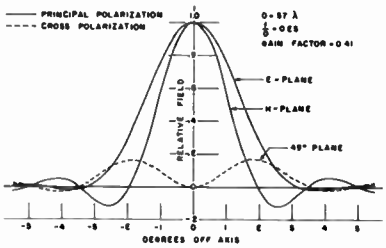


Fig. 6
Hyperboloid lens excited by a magnetic dipole feed and by a plane wave feed.

Diffraction Patterns

The diffraction patterns of aperture distributions that we have discussed have been computed from the well-known diffraction integral. The method used to evaluate this integral is to approximate the radial dependence of the aperture distributions by three or four term polynomials in the aperture radius whose coefficients are determined such that the square of the difference between the approximating polynomial and the desired function is a minimum, and then to integrate term by term. The process is straight-forward but tedious and will not be discussed further.

The diffraction pattern of a paraboloid with an aperture diameter of 37 wavelengths and a focal length to diameter ratio of 0.25 excited by an electric dipole is shown in Fig. 7. It is noted that the principal-polarization patterns in the E-plane are considerably wider than in the H-plane, and that the first side lobe in the E-plane is



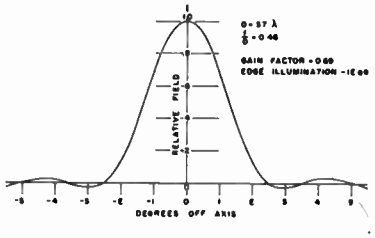
PARABOLOID DIFFRACTION PATTERNS WITH ELECTRIC DIPOLE FEED

Fig. 7

lower than in the H-plane. This behavior is consistent with the fact that the aperture distribution is much more strongly tapered in the E-plane than the H-plane. The cross-polarized lobes measured in planes at 45° to the E- and H-planes, where they attain their maximum value, are quite close to the polar axis and are higher than the first side lobe in either the E- or H-plane. The gain factor for this antenna is only 0.41. For paraboloids of the same aperture diameter but longer focal lengths, the H-plane patterns stay quite constant but the E-plane patterns have narrower half-power beamwidths and higher side lobes. The cross-polarized lobes and the gain factor both decrease as the focal length increases.

Because of the symmetry between the aperture fields of the paraboloid when excited by the electric and magnetic dipoles, it turns out that the diffraction patterns of the paraboloid excited by a magnetic dipole are related to those of the paraboloid excited by an electric dipole. The E- and H-plane patterns of the magnetic dipole excited paraboloid correspond to the H- and E-plane patterns of the electric dipole excited paraboloid respectively, while the cross-polarized patterns are the negative of one another. The gain of the two systems is the same.

Figure 8 shows the diffraction pattern of a paraboloid with a 37 wavelength diameter aperture, a focal length to diameter ratio of 0.40, and an edge illumination of -12 db. The plane wave aperture is assumed to illuminate the reflector so that this principal-polarization pattern is symmetric about the polar axis. No cross-polarized lobes are present. For this edge illumination, the computed gain factor of the antenna is a maximum at 0.69 and the first side lobe level is -29 db. Decreasing the edge illumination decreases the gain factor and widens the main beam, while the side lobe level goes through a minimum value of -45 db at about -20 db edge illumination, and then increases.

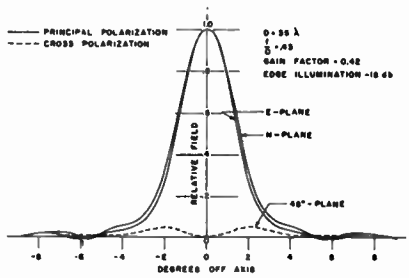


PARABOLOID DIFFRACTION PATTERN WITH PLANE-WAVE FEED

Fig. 8

Figure 9 shows the diffraction pattern of a hyperboloid lens with a 35.5 wavelength diameter aperture and a focal length to diameter ratio of 0.41, excited by a plane wave source to give an edge illumination of -15 db at the curved surface of the lens. It is noticed that although the illumination of the curved surface of the lens is symmetric about the polar axis, the E-plane pattern is narrower than the H-plane pattern. The reason is because the illumination is less tapered in this plane, since the transmission coefficient for the lens is greater in the E-plane. The computed gain factor of the lens is a maximum at 0.42. This is almost 2.2 db less than the best gain factor of 0.69 realizable with a paraboloid of

about the same f/D ratio. The cross-polarized lobes are -32 db down and lie about 2° from the polar axis. Decreasing the edge illumination decreases the amplitude of these lobes very slightly, while their position of maximum intensity remains essentially constant. As is expected, the width of the main lobe also increases as the edge illumination decreases. The side lobe level, however, shows no systematic variation with edge illumination.



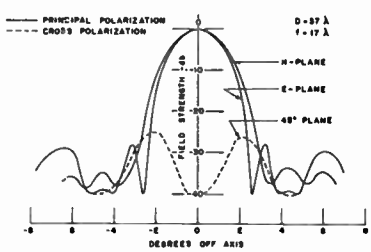
HYPERBOLOID LENS DIFFRACTION PATTERNS WITH PLANE-WAVE FEED

Fig. 9

Experimental Results

The experimentally determined radiation patterns of these antennas do not agree in detail with the computed patterns. The agreement is best in the regions near the polar axis where the radiated power density is relatively high. However, deviations between theory and experiment are to be expected in view of the many approximations made in the analysis.

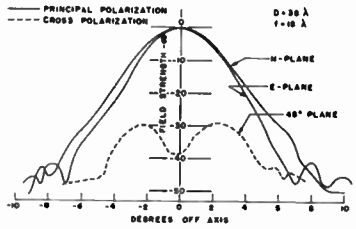
An experimental radiation pattern of a paraboloid reflector with a 37 wavelength diameter aperture excited by a pyramidal horn feed, which gives an edge illumination of -20 db in the E-plane and -21.5 db in the H-plane, is shown in Fig. 10. Here, it is noticed that the cross-polarized lobes are -25 db. It has been found that the level of these lobes decreases monotonically as the H-plane dimension of the horn is



MEASURED RADIATION PATTERNS OF PARABOLOID REFLECTOR

Fig. 10

increased. Because the wave impedance of the mode in the horn approaches 377 ohms as this dimension is increased, this behavior tends to confirm the previous theory that a paraboloid excited by a plane wave source with a wave impedance of 377 ohms would have no cross-polarized lobes at all. The level of the first side lobes in the E-plane is -28 db, while that in the H-plane is -33 db. The H-plane pattern is slightly wider than the E-plane pattern because the aperture illumination is more strongly tapered in this plane.



MEASURED RADIATION PATTERNS OF HYPERBOLOID LENS

Fig. 11

In Fig. 11 the measured radiation patterns of a hyperboloid lens excited by a horn feed are shown. The main beam is quite broad and the side lobes are very low, the highest one being -41 db. The cross-polarized lobes are -29 db, and lie close to the polar axis (compared to the width of the main lobe); hence, they should not be objectionable. The measured gain factor of this lens is 0.24.

Conclusions

The main conclusions to be drawn from this discussion are:

1. Low cross-polarized radiation lobes can be obtained from paraboloid reflectors when they are excited by small horns which approximate plane wave sources.

2. Low principal-polarization side lobes and low cross-polarized lobes can be obtained with a hyperboloid lens excited by a small horn, but the main beam is wide and the aperture efficiency is less than for paraboloids.

Acknowledgment

The work reported in this paper was sponsored by the Signal Corps Engineering Laboratory, Fort Monmouth, New Jersey.

References

1. E.U. Condon, "Theory of Radiation from Paraboloid Reflectors," Westinghouse Research Report, SR-105. (September 1941).

2. L.J. Chu, "Theory of Radiation from Paraboloidal Reflectors," Radiation Laboratory Report V-18(February 1941).

NOTES ON PROPAGATION

L. A. Byam, Jr., and J. Z. Millar
Western Union Telegraph Company
New York, N. Y.

General

These notes are intended as a brief discussion of the fading which occurred during an extended microwave propagation experiment, operated at a frequency of 4000 Mc.

Several similar tests have been conducted in recent years and have yielded considerable information concerning the behavior of microwave signals under the influence of varying atmospheric conditions. Most of this information is available in the literature, having appeared in articles published by Durkee,[1] Friis,[2] Millar and Byam,[3] Kiely and Carter,[4] Crawford and Jakes[5], and by others.

Although the test under consideration was conducted under conditions very much similar to earlier tests, there were two important differences. First, the present test included spaced diversity reception whereas in earlier tests diversity reception was not employed. Simultaneous recordings of both main and diversity signals afforded means of fade-by-fade comparison and provided data for a general analysis in determining the effectiveness of diversity action. Secondly, earlier tests were made with an r.f. power output in the neighborhood of 0.1 watt at 4000 Mc., whereas in this more recent test a c.w. type magnetron delivered approximately 10 watts at the same frequency. The hundred-fold increase in output power added 20 db to the depth of the measurable fade.

Test Conditions

A Raytheon Type QK151 magnetron, serving as the source of r.f. power, delivered approximately 10 watts of unmodulated carrier at a frequency of 4000 Mc. into the transmitting antenna situated atop the Western Union Building at New York. At Neshanic, New Jersey, separate simultaneous recordings were taken of the received signal as appearing on the main and on the diversity receiving systems. Recorders were of the motor-driven type, operated from the same 60-cycle a.c. power source, and were synchronized. The test path involved will be recognized as one used in a prior propagation test. It is situated almost entirely over land and is approximately 42 miles in length. A profile picture of this path appears in Figure 11. Since the profile was constructed on the basis of a four-thirds earth's radius, the transmission path was shown as following a straight line to represent "standard" conditions of propagation. The transmitting and receiving antennas were of the dipole type, coaxial fed and the associated reflectors were conventional parabolic surfaces, 48" in diameter. The diversity receiving antenna was situated directly below the main receiving antenna, and separated from it by a vertical distance of 31 feet, center to center. Finally, recordings were taken for a period of roughly 14 months, beginning on January 9, 1950.

Carrier-to-Noise Ratio

An exact determination of the carrier-to-noise ratio requires, among other things, accurate information as to the noise figure of the receiver used. When the absolute noise of a receiver is not known, it is convenient and usually sufficiently accurate to assume a representative noise factor such as fifty times thermal or 17 db above thermal. Modern microwave receiver design has improved noise factors to within twenty times thermal. In the present case, the factor of fifty was used as a concession to earlier design considerations. It is interesting to note the manner in which other factors influence the carrier-to-noise ratio, as indicated by the relationship shown below:

$$C/N = 2.42 \times 10^{-2} \times \frac{D^4}{L^2} \times f^2 \times \frac{P_t}{B}, \text{ where}$$

D is diameter of reflector in Feet.
L is path length in Miles.
f is frequency in Mc.
P_t is transmitter power in Watts.
B is receiver bandwidth in Mc.
C/N is carrier-to-noise ratio.

To obtain C/N in terms of db, the following form is used:

$$C/N = -16 + 40 \log D + 20 \log f + 10 \log P_t - 20 \log L - 10 \log B.$$

For the conditions of the test, as outlined, and with the further assumption of 4 Mc. as representing the receiver bandwidth, a carrier-to-noise ratio of 51 db obtains.

Analysis of Data

Although very nearly two miles, (representing 10,107 hours), of signal recording strip was examined, attention was directed primarily to those fades which dropped 15 db or more from the normal signal level. (As in the previous test, the normal signal represents the mean signal level for periods during which fading was absent or negligible, i.e., periods when transmission was stable.) For each such fade, the essential data were extracted and tabulated, including information with respect to the corresponding level of the diversity (or main) signal.

Initially, in the interests of operational considerations, these data were assembled and represented in a form, Figure 1, intended to reveal the general structure of the fading occurring during during the test. The two upper curves of Figure 1 portray the generally regular manner in which the number of fades decreases with depth. For the range of signal levels considered, there were many more fades on the main than on the diversity, the difference varying by a factor of roughly two at the -15 db level to roughly nine at the -35 db level. The two lower curves shown in Figure 1 represent the average length in minutes of all fades falling below the abscissa level considered. It will be observed that the ordinate scale associated with these two curves appears at the right-hand side of the chart. To illustrate, the average duration of fades falling below the signal level of -20 db was 4.8 minutes with respect to the main receiver and 2.6 minutes with respect to the diversity. Although some degree of similarity is suggested by the manner in which the length of fade generally decreases with increasing depth of fade on both main and diversity antennas, the marked difference between these two curves requires explanation. About midnight on February 1, 1951, the diversity signal fell to slightly more than 35 db below normal and remained at that level for well over an hour. At the same time the main signal also faded but with much less depth and for a shorter duration. Were it not for this one fade, the two curves would have exhibited a marked similarity.

Ideally, a microwave radio system is engineered to have a signal-to-noise ratio which includes, among other things, a sufficient fading-allowance margin to insure the degree of continuity dictated by service requirements. For a smaller signal-to-noise ratio, deterioration of service would obtain; a higher value would be considered as extravagant. Practical considerations therefore require reliable information to be available which will permit translating service demands in terms of fading-allowance margin. For this purpose, distribution curves of instantaneous signal strength, together with associated curves representing distribution of fade length, are helpful, particularly if such curves represent data taken over a long period of time. Curves A and B of Figure 2 represent distributions of instantaneous signal strength for the main and diversity respectively based on the data taken from the test. Similarly, the curves appearing in Figure 3 and 4 represent distributions of fade lengths. For purposes of comparison, additional curves are shown in Figure 2, as identified thereon. Curves D and E are based on data appearing in the Bell System Technical Journal of January 1952 in a paper written by Messrs. Crawford and Jakes[5]. The pronounced spread between the main and diversity curves is noteworthy and indicates that fading was less severe on the diversity than on the main. Generally, the diversity signal during periods of moderate fading was found to be three or four dbs higher than the main, a good example of which may be found in Figure 7. This effect is attributable to the geometry of the system, and where depth of fading is a determining factor, forms a basis for locating antennas somewhat lower than at the Fresnel maximum. In any event, the marked difference in fading effects as reflected by these curves, A and B in Figure 2, is remarkable in that it was occasioned entirely by the 31 foot vertical separation between the main and diversity receiving antennas. The importance attached to careful probing of proposed tower sites will thus be understood. The effect of the infrequent case of fading mentioned in discussing Figure 1 is again manifested in the shape of that portion of Curve B, Figure 2, lying below the 0.04% level. Again, were it not for this one fade, Curve B would have appeared strikingly similar to Curve A from the 0.04% level on downwards, as was found by actual (tentative) plotting of data which excluded the fade in question. Other Curves, C, D, and E, are included in Figure 2 for purposes of comparison. A comparison of Curve D with that of E would seem to indicate that the effect of path length is virtually obscured under the influence of other considerations, such as, for example, path clearance.

These distribution curves may also be considered as fading probability curves and thus

used to determine the probability of a fade having a duration greater than zero minutes below the level considered, i. e., the probability of a fade occurring below that level. For example, according to Curve B, Figure 2, the probability that a fade will occur having a depth greater than 20 db below normal is 0.0009; the probability of a fade falling more than 32 db below normal is 0.00018.

At this point, it would appear logical to consider information bearing on the relative time duration of fades. For this purpose, the charts shown as Figure 3 and 4 were prepared, based on data obtained from the experiment. The curves appearing in Figure 3 are distribution curves of fade length for fades occurring on the main receiver. Distribution curves for the diversity are shown in Figure 4. With respect to Figure 3, no explanation could be found for the coincidence of the -20 and -25 db curves. When plotted, the points for these two curves virtually coincided leaving no alternative to representation by a single curve. Data for the -40 curve were considered insufficient properly to represent a significant distribution at this level and for that reason the curve is labeled "estimated". The straight-line distributions, obtaining as shown in Figure 4, (using logarithmic probability paper), are considered noteworthy. The manner in which the distribution curves of Figure 3 and 4 supplement those of Figure 2 is best explained by example. Again selecting the diversity curves, assume the occurrence of a fade below -20 db. From the -20 db curve of Figure 4 the probability that this fade will remain below -20 db for more than 2-1/2 minutes is 0.20. But, the probability of a fade falling below -20 is 0.0009 as determined from the earlier example using Curve B of Figure 2. Hence the probability of a fade remaining below -20 db for more than 2-1/2 minutes is 0.0009 × 0.20 or 0.00018. In a like manner similar information may be obtained for various other levels and time intervals within the range of the curves. Such information is helpful in that it provides a means of determining fading-allowance margin in terms of service requirements. As a simple illustration, assuming that a particular service requires that the probability of an interruption of more than 1/2 minute shall not exceed 0.001, the fading-allowance margin is found to be 17 db.

Diversity Action

As stated earlier, attention was directed primarily to those fades with excursions of 15 db or more below normal. Usually when such a fade occurred on the main it was accompanied by a diversity fade of somewhat the same general pattern although the diversity fade was generally less severe. This characteristic of similarity is illustrated in Figure 7 and was manifest on 79.9% of the total number of occasions that the main faded below -15 db. Of the 79.9% occasions on which similarity was exhibited, the diversity fade started downward at approximately the same instant as the main on 66.4% occasions; the diversity fade preceded the main on 9.6% occasions and the diversity fade followed the main on the remaining 3.9% occasions. For 20.1% of the total number of occasions that the main faded below -15 db there was no similarity in the action of the main and diversity signals, i.e., the diversity signal faded upward or in opposition to the main, as illustrated in Figures 8 and 9, or was constant or otherwise independent of the main fade, as illustrated in Figure 10. In passing, it will be noted that the number of times a diversity fade preceded the corresponding main fade exceeded the number of times a diversity fade followed the corresponding main fade by a ratio of 2-1/2 to 1.

On 80.7% of the total number of occasions that the main faded below -15 db the diversity did not fade below -15 db; on 38.3% of the total number of occasions that the diversity faded below -15 db the main did not fade below -15 db. Stated in a different way, 19.3% of the main fades below -15 db occurred simultaneously with 61.7% of the diversity fades below -15 db. During periods of simultaneous fading below -15 db, 1 main fade mixed on the average with 1.7 diversity fades. Simultaneous fading did not occur below the level of roughly -38 db. Thus, in terms of the number of fades, diversity action may be considered as 80.7% effective with respect to the -15 db reference level and 100% effective with respect to the -38 db reference level. The manner in which the effectiveness of diversity action varied with signal level is shown by the curve appearing in Figure 5.

In the preceding paragraph consideration was given to the relative number of main and diversity fades below -15 db for both simultaneous and non-simultaneous fading conditions; nothing was said with respect to the time duration of fading under these two conditions. During non-simultaneous fading below -15 db the average length of the main fade was 4.3 minutes; the average length of the diversity fade was 2.6 minutes. During simultaneous

fading, however, the average length of the main fade was 17.7 minutes and that of the diversity 4.8 minutes.

During 77.5% of the time the main signal was below -15 db the diversity signal remained above -15 db; during 25.0% of the time the diversity signal was below -15 db the main signal remained above -15 db. Stated in another way, simultaneous fading below -15 db occurred for 22.5% of the time the main was below -15 db or 75.0% of the time the diversity was below -15 db. Simultaneous fading did not occur below roughly -38 db. Thus, in terms of time duration of non-simultaneous fading, the effectiveness of diversity action may be considered to be 77.5% at the -15 db reference level and 100% at the -38 db reference level. For other levels, the effectiveness of diversity action may be determined from the curve appearing in Figure 6.

In concluding the discussion on diversity action, consideration will be given to the mechanics of correlation in analyzing the relationship of simultaneous variations in the main and diversity signals under two general but entirely different conditions of fading. These two conditions were arbitrarily chosen.

The first conditions will be referred to as moderate fading and includes only those fades whose minima ranged roughly from -10 db to -20 db. Such fades were found generally to be of the comparatively slow type of fade and during which similarity of variations in the main and diversity signals was observed as illustrated in Figure 7. From a sample taken of such fades, the maximum depth was tabulated for each fade together with the corresponding value of the diversity (or main) signal at the same instant. Under this condition of fading a correlation coefficient of + 0.69 was obtained. Thus, a fairly close relationship is indicated and it may be concluded that the diversity was not particularly helpful under the condition defined above as moderate fading.

The second condition will be referred to as severe fading and includes only those fades whose minima exceeded -30 db. Such fades were found generally to be of the comparatively rapid type of fade, characteristic of wave interference. In this case the sample consisted of all fades falling more than -30 db and for each such fade the maximum depth was tabulated together with the corresponding value of the diversity (or main) signal at the same instant. Under this condition of fading a correlation coefficient of -0.36 was obtained. Thus, it may be concluded that the diversity was decidedly helpful under severe fading conditions as defined.

General Comments

The foregoing, although by no means an exhaustive analysis of the data obtained from the experiment, furnishes specific quantitative measurements which portray the essential aspects of the fading encountered during the test. Since the analysis included data on fading at comparatively much greater depth, the results supplement those of earlier propagation tests in which the same path was used.

Although the test extended over a long period of time, it is not intended that the results obtained may be applied to a propagation path other than the one used in the experiment. For an overland path of approximately the same length and under the same conditions of clearance, however, the results may be useful in arriving at a general prediction of fading effects.

The general similarity in form exhibited by the distribution curves would seem to be worthy of notice.

References

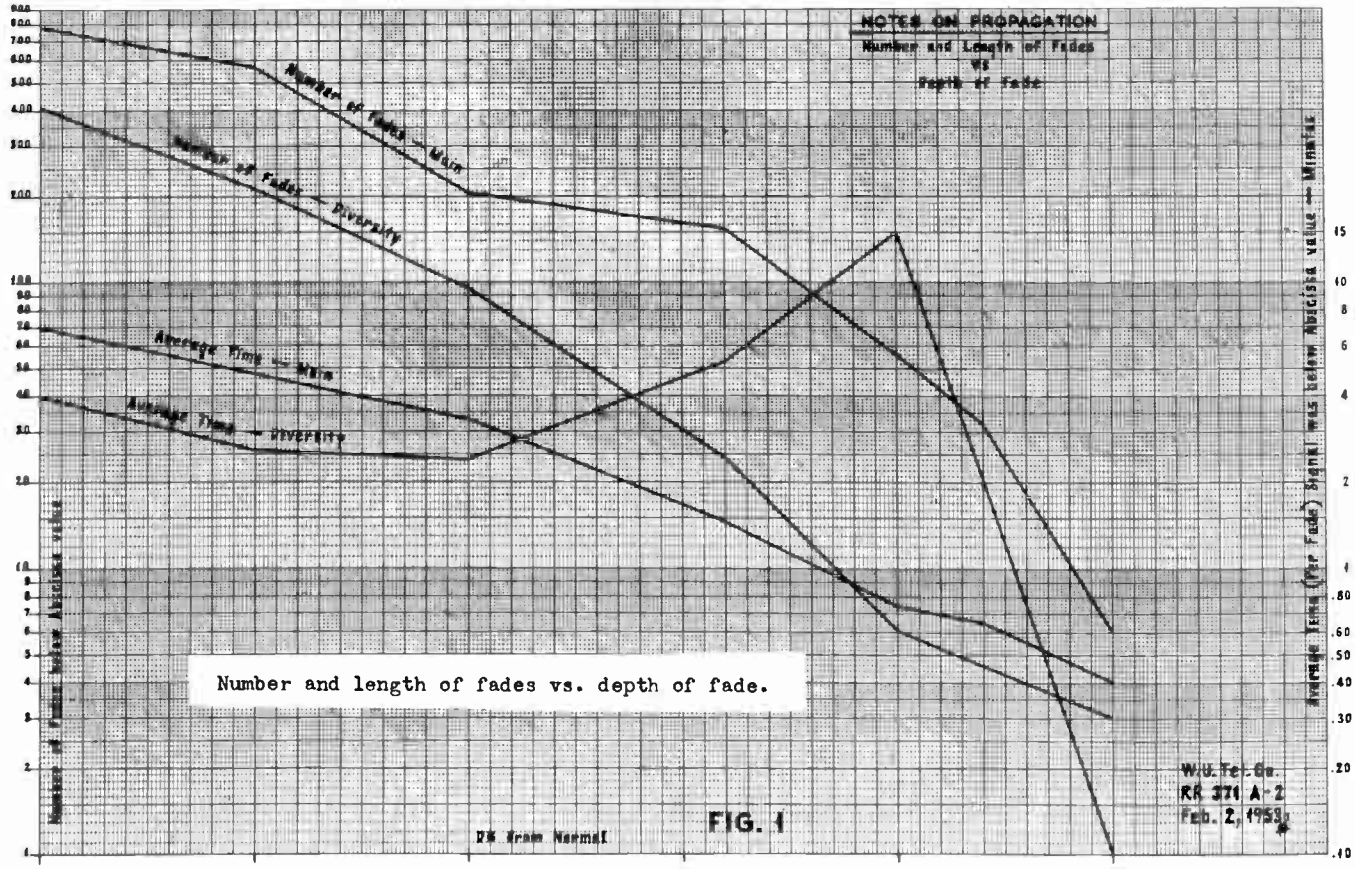
(1) Results of Microwave Propagation Tests on a 40-mile Overland Path, A. L. DURKEE, Proceedings IRE, February 1948.

(2) Microwave Repeater Research, H. T. FRIIS, Bell System Technical Journal, April 1948.

(3) A Microwave Propagation Test, MILLAR and BYAM, Proceedings, IRE, June 1950.

(4) An Experimental Study of Fading in Propagation at 3-CM Wavelength over a Sea Path, KIELY and CARTER, Proceedings of IRE, Part III, March 1952.

(5) Selective Fading of Microwaves, CRAWFORD and JAKES, Bell System Technical Journal, January 1952.



Number and length of fades vs. depth of fade.

FIG. 1

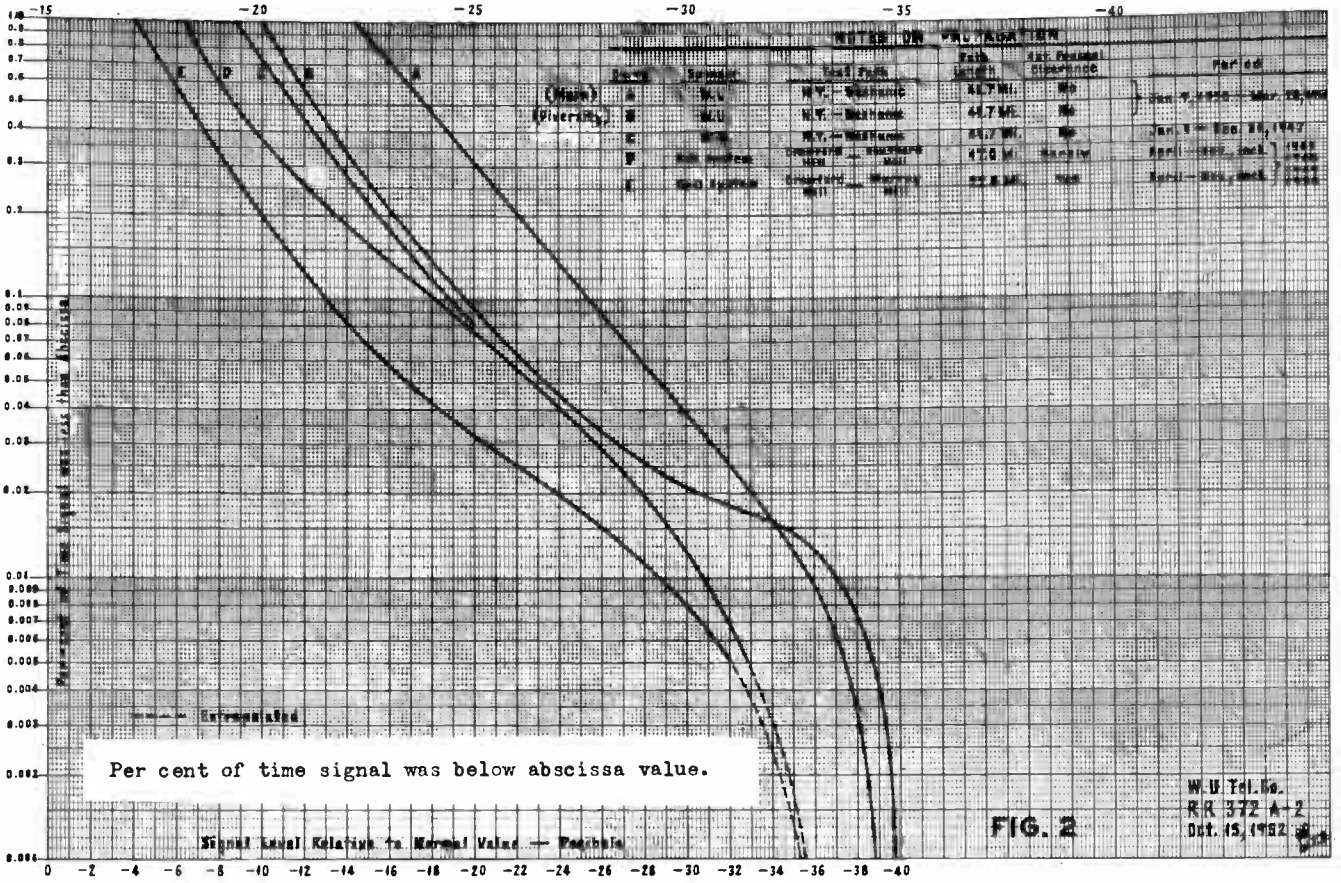


FIG. 2

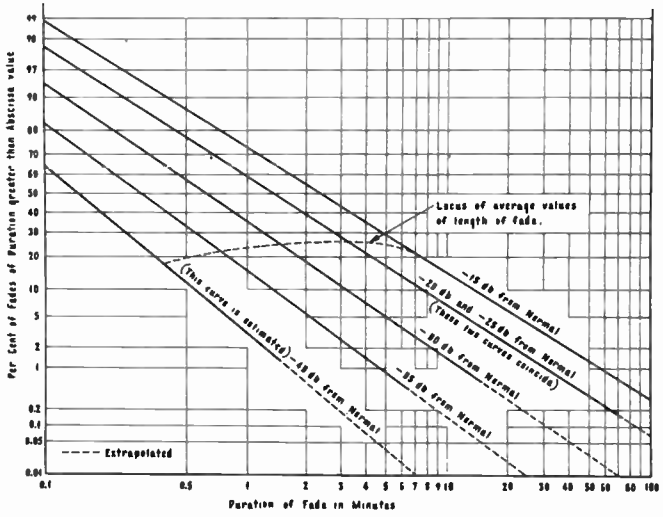


Fig. 3
Number of fades (in per cent) vs.
length of fade (main receiver).

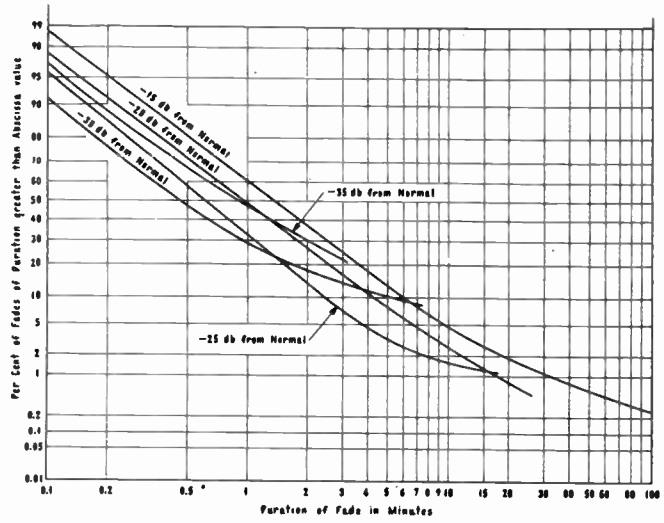


Fig. 4
Number of fades (in per cent) vs.
length of fade (diversity receiver).

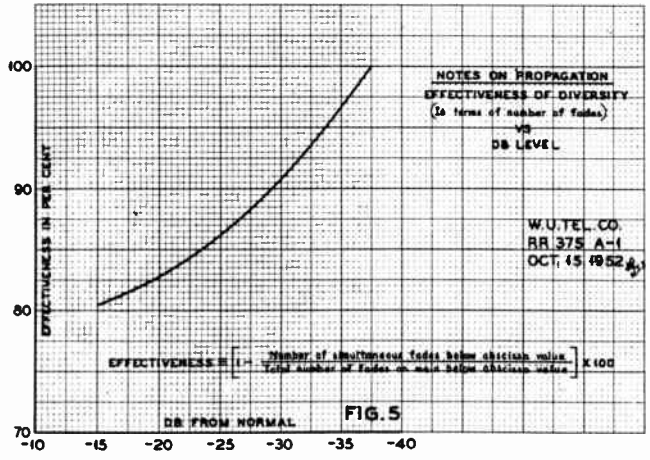


Fig. 5
Effectiveness of diversity vs. signal
level (in terms of number of fades).

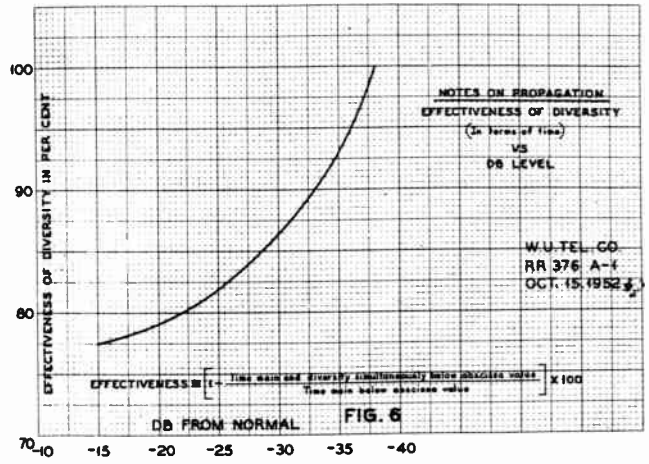


Fig. 6
Effectiveness of diversity vs. signal
level (in terms of time).

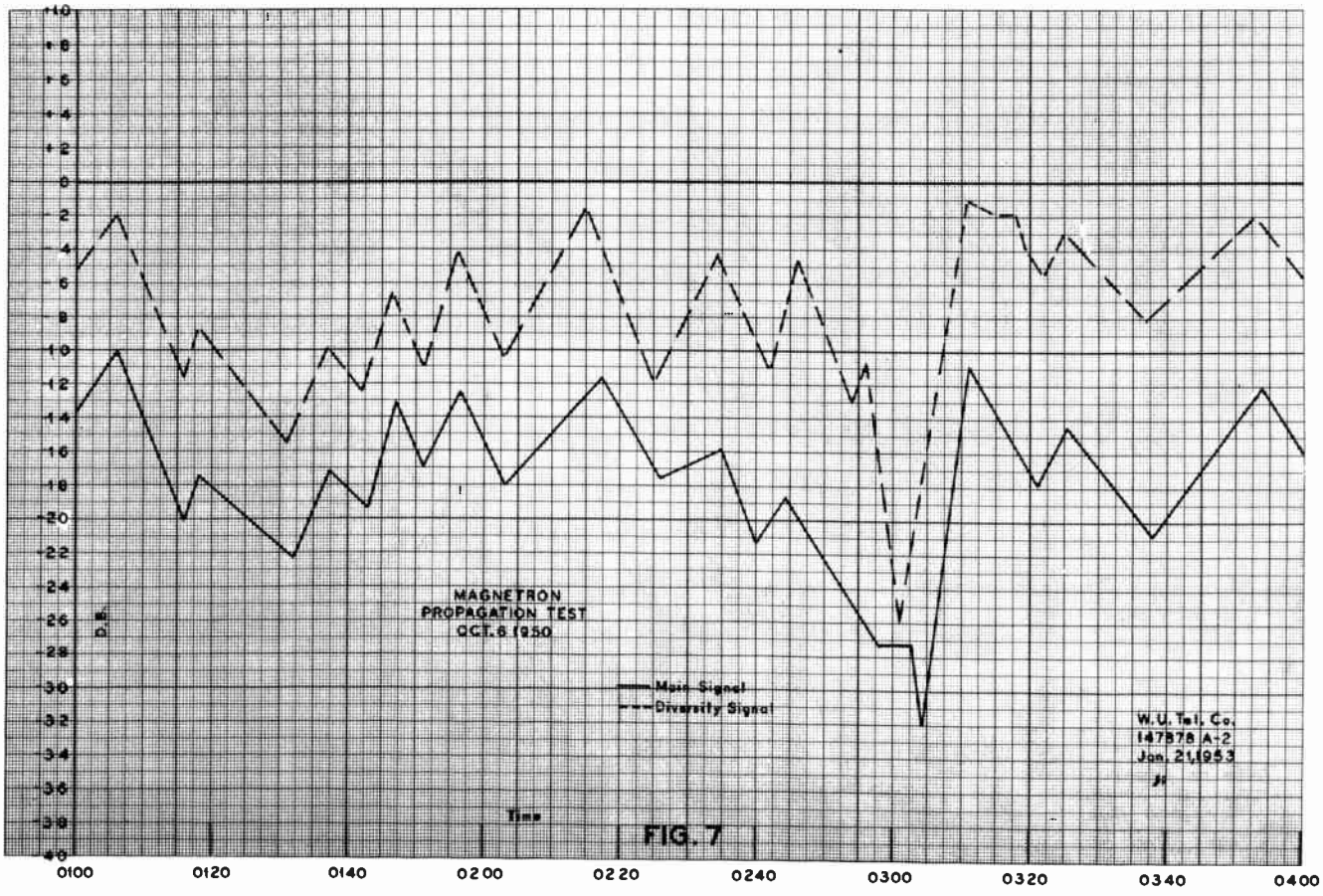


Fig. 7
Fading chart, illustrating similarity in variations
of the main and diversity received signals.

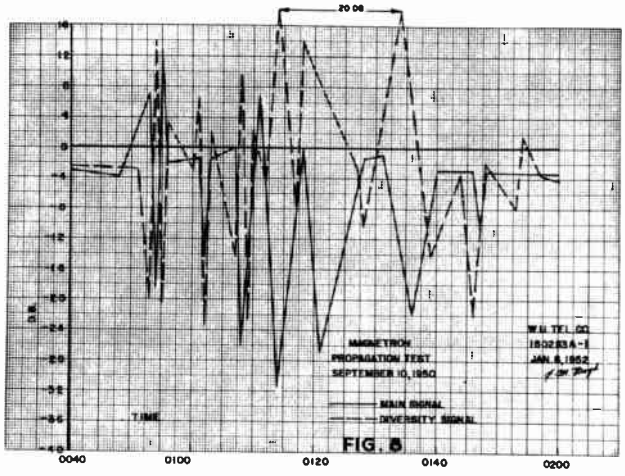


Fig. 8
Fading chart, illustrating examples of opposition
of main and diversity received signals.

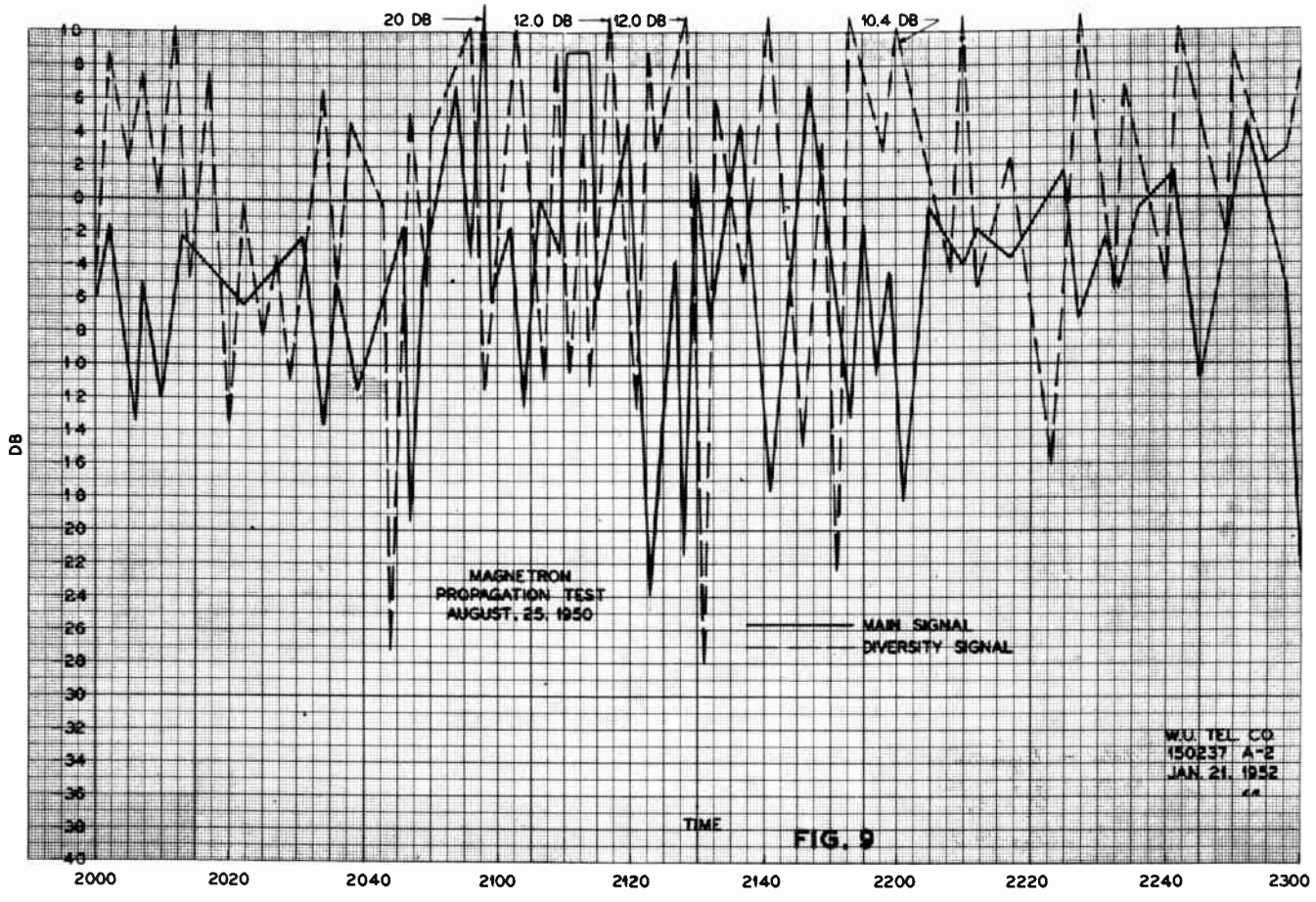


FIG. 9

Fig. 9
Fading chart, illustrating examples of opposition and also phase differences in main and diversity received signals.

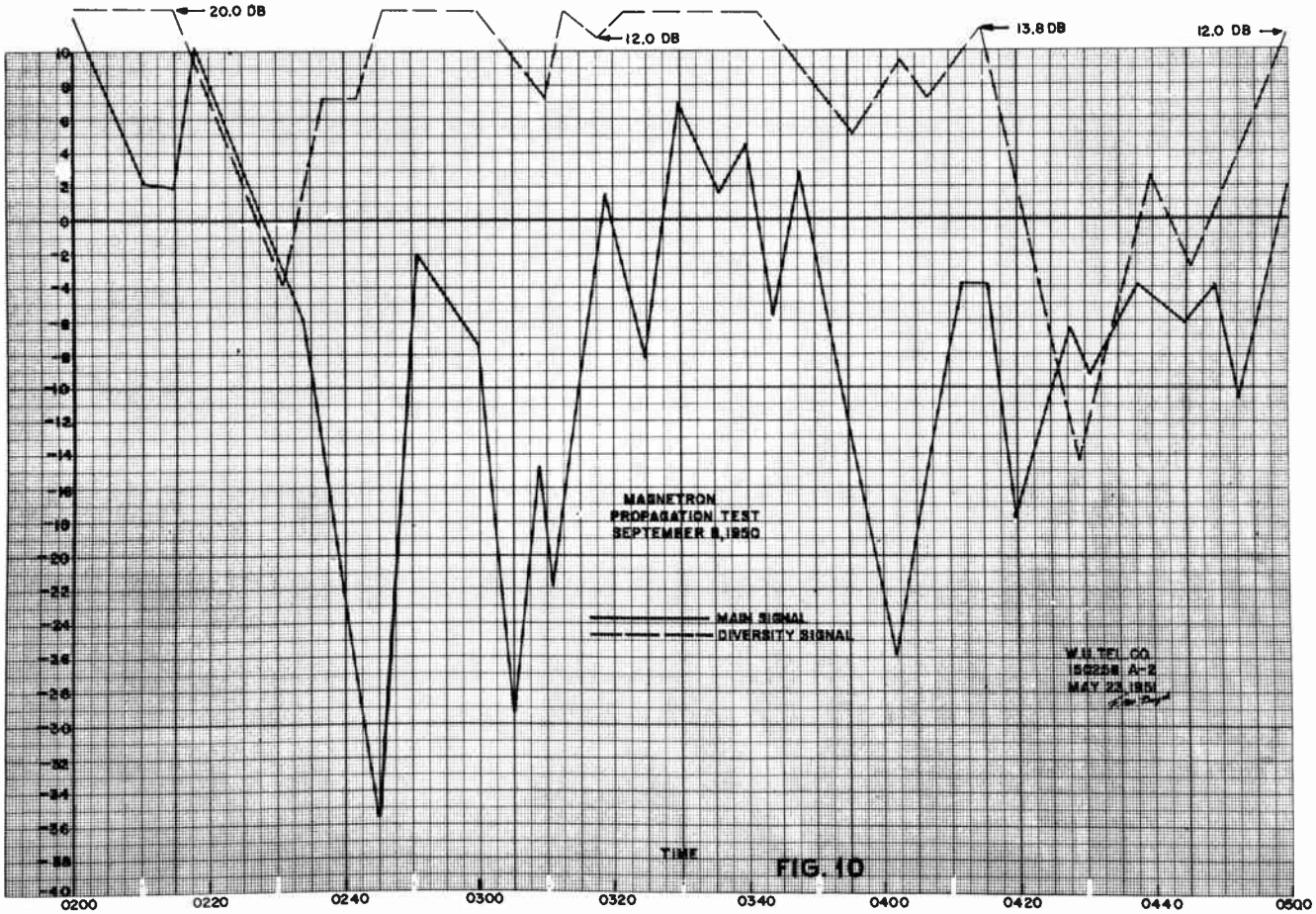


Fig. 10
Fading chart, illustrating examples of independence of diversity signal with respect to main signal.

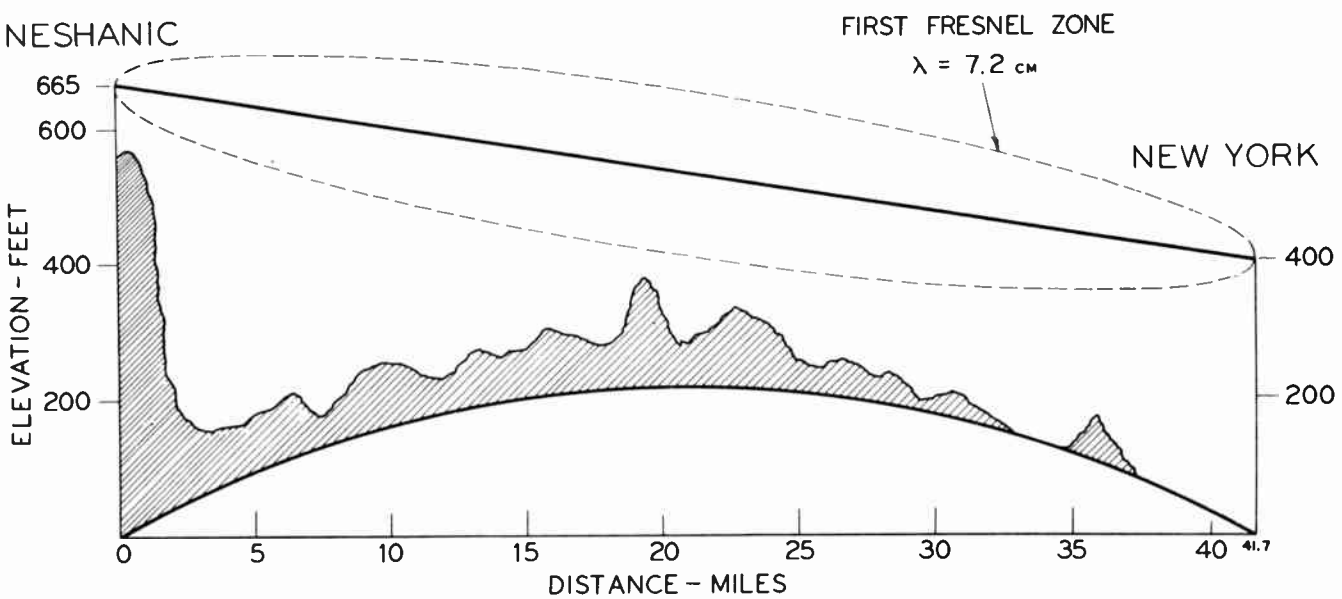


Fig. 11
Profile of New York - Neshanic propagation path.

TROPOSPHERIC PROPAGATION IN HORIZONTALLY STRATIFIED MEDIA OVER ROUGH TERRAIN

H. M. Swarm, R. N. Ghose*, and G. H. Keitel
University of Washington
Seattle, Wash.

The propagation of VHF and UHF radio waves through the troposphere is affected by the variations in the atmospheric structure as well as the physical conditions peculiar to each path. Observed field intensities far beyond the horizon are much higher than the fields predicted on the basis of the mode theory, the deviation being much greater than any possible experimental error. It is definitely known that some of the important factors in the mechanism of long-range (150 miles and more) tropospheric propagation have been omitted in the development of the mode theory. Whether these all-important factors at long distances are internal reflections, scattering due to turbulence, or a combination of both is a question that will only be decided by further experimentation.

The purpose of this paper is the study of the correlation of fields calculated on the basis of the mode theory with measured signals for a 150-mile path. Only days on which the atmospheric structure resulted in a linear index of refraction curve versus height were considered in this paper. Other atmospheric structures such as bilinear and linear exponential are being studied and will be reported in a future paper. The effect of the ignored factors is believed to be minor since, at least in the case of the internal reflection theory, the predicted signal is approximately the same as for standard refraction at this distance, according to T. J. Carroll[1]. The particular path under consideration is approximately 150 miles long, from Portland, Oregon, to Seattle, Washington. This path is characterized by a 3000-foot ridge crossing the path near the midpoint and numerous lower obstructions over most of the remaining path. The receiving site is far below line-of-sight.

One of the most important factors in determining the field strength in the diffraction zone is the effect of refraction in the troposphere. Since the refractive index is a function of the temperature, pressure, and vapor pressure of the atmosphere, the field intensity will show wide fluctuations as the meteorological conditions change. The procedure used in this analysis was to assume a horizontally stratified atmosphere in which the refractive index changes linearly with height. The assumption of horizontal stratification along this path is not unreasonable. Except in unusual instances, the entire path is subject to the same weather conditions. Normally the fronts move from West to East and are almost always parallel to the path.

In the solution of the Hertzian wave equation for a linear structure, the rate of change of the refractive index with respect to height is much more important than the magnitude of the refractive index. Therefore, the atmospheric structure enters into the calculation of the field strength through the refractive index gradient. J. E. Freehafer, W. T. Fishback and W. H. Furry[2] solved the Hertzian wave equation for the linear case and give numerical results for the standard atmosphere structure. In the following theoretical development, their method was pursued, except the refractive index gradient was retained as a variable in the final equation.

The Hertzian wave equation is

$$\nabla^2 \underline{\Pi} + \omega^2 \mu \epsilon \underline{\Pi} = 0 \qquad (1)$$

and may be reduced to the scalar wave equation since the Hertzian vector is assumed to have only a vertical component and the earth's surface becomes the x-y plane,

$$\nabla^2 \psi + k^2 N^2 \psi = 0 \qquad (2)$$

where

ψ is a function of r, the distance from the transmitter, and z, the height above the ground at the measuring point.

$k = 2\pi/\lambda$

N is the modified index refraction, used to take into account the flattening of the earth's surface and is equal to $n(1 + z/a) \approx n + z/a$.

a is the earth's radius.

Equation (2) can be solved as a boundary value problem in which the boundary conditions are

(a) ψ should be essentially zero at the ground.

(b) ψ should reduce to the form $\frac{1}{r} e^{-ikr}$ for small distances from the source.

(c) ψ should represent an outgoing wave at great distances.

ψ is assumed to be of the form of a product solution which may be put in the following form

---

* Now at the University of Illinois, Urbana, Ill.

$$\psi(r,z) = \int_0^\infty \beta\, J_0(\beta r)\, V(\beta,z)\, d\beta \qquad (3)$$

where

β^2 is the separation constant
$J_0(\beta r)$ is zero order Bessel function
$V^0(\beta,z)$ is a function which satisfies the differential equation,

$$\frac{d^2V}{dz^2} + (k^2N^2 - \beta^2)V = 0$$

Satisfying the rest of the boundary conditions, (3) may be written as

$$\psi = \sum_{m=1}^{\infty} A_m\, J_0(\beta_m r)\, U_m(\beta_m, z_t)\, U_m(\beta_m, z_r) \qquad (4)$$

where

A_m is the characteristic function
$U_m(\beta_m, z)$ are functions of the antenna heights at the receiving and transmitting locations.

Defining F as the ratio of the field strength in the presence of the earth to the free space field strength

$$F = \frac{|E|}{|E_0|} = r|\psi| \qquad (5)$$

In terms of (4), F becomes

$$F = \sum_{m=1}^{\infty} |\Phi_m(r)|\,|U_m(z_t)|\,|U_m(z_r)| \qquad (6)$$

where

$$\Phi_m(r) = 2\sqrt{\pi X}\, \exp(i A_m X)$$

$$\Phi_m(r) = 2\sqrt{0.528\,\pi r \left(\frac{dM}{dz}\right)^{2/3}\left(\frac{\pi}{\lambda}\right)^{1/3}}\; \exp\left[i\, 0.528\, A_m r \left(\frac{dM}{dz}\right)^{2/3}\left(\frac{\pi}{\lambda}\right)^{1/3}\right]$$

$$U_m(z) = i\,\frac{h_2(Z + A_m)}{h_2'(A_m)}$$

$$U_m(z) = i\,\frac{h_2\left[\frac{z}{50}\left(\frac{dM}{dz}\right)^{1/3}\left(\frac{\pi}{\lambda}\right)^{2/3} + A_m\right]}{h_2'(A_m)}$$

where

$$h_2(\xi) = \left(\frac{2}{3}\xi^{3/2}\right)^{\frac{2}{3}} H^{(2)}_{1/3}\left(\frac{2}{3}\xi^{\frac{3}{2}}\right)$$

The quantity dM/dz is the gradient of the index modulus, M, where $M = (N-1)10^6$.

In all of the discussion to follow and in all of the calculations of field strength, only the first term of the infinite series, in (6), is used. The effect of second and higher terms or "modes" becomes smaller because the imaginary part of A_m increases as m increases. For example, $A_1 = -1.169 + j\,2.02$ while $A_2 = -2.044 + j3.54$, and the attenuation goes up exponentially as the imaginary part of A_m increases. Therefore the second mode is highly attenuated and its contribution is negligible.

In (6) Φ is shown to be a function of the gradient of M, wavelength, and distance. It is usually called the attenuation factor. Fig. 1 shows Φ as a function of dM/dz for various distances from 50 to 200 miles for a for a frequency of 100 mc. In order to calculate the field strength at the receiver, it is necessary to multiply F by the free space field strength. The product of the attenuation factor and the free space field strength, ΦE_0, is plotted as a function of dM/dz in Fig. 2. This product might be interpreted as the field to be expected if the product of the rest of the factors in (6) equals unity.

Fig. 3 shows ΦE_0 as a function of distance for various dM/dz. This figure clearly shows the greater range of variation to be expected at greater distances. One curve for standard atmosphere (dM/dz = 0.036) is shown for 500 mc. This curve shows the much higher rate of attenuation with distance at 500 mc, e.g., at 100 miles distance there is 47 db more attenuation for 500 mc as compared with 100 mc.

The function $U_m(z)$, of (6), is of such a form that an increase in z results in an increase in the calculated signal, and therefore has been termed the height-gain function. Fig. 4 shows the height-gain function for various heights from 100 to 2000 feet as functions of dM/dz at 100 mc. These curves are relatively flat in the generally encountered range of dM/dz = 0.03 to 0.04. Fig. 5 gives a clearer picture of the effect of height for a given value of dM/dz. The 500 mc curve for standard atmosphere is included in this figure for comparison. An antenna height increase from 100 to 750 feet results in an increase of 30 db in the expected signal at 500 mc as compared with an increase of only 22 db at 100 mc for the same change in height.

Fig. 6 shows the calculated field in db above 1 μv/m for 1 kw on a hypothetical path

of 150 miles and antenna heights of 750 and 200 feet as a function of dM/dz. The difference in the field strengths is seen to approach zero for low values of dM/dz. However, in the observed range of dM/dz = 0.03 to 0.04, the 500 mc signal is 32 to 45 db lower than the 100 mc signal.

Experimental Work

The experimental measurements used in comparison with the theory were obtained from recording equipment operated by the University of Washington under contract to the National Bureau of Standards. The transmitter is a commercial FM station, KOIN-FM, 101.1 mc, located in Portland, Oregon, and radiates an effective power of 39.3 kw in the direction of Seattle. The antenna height is 1600 feet above sea level and 1400 feet above the 2 to 10 mile average terrain. The receiving-recording system is located at the University of Washington, 149 miles to the north of Portland. The receiving antenna is 200 feet above sea level and 100 feet above ground level. The profile of the path is shown in Fig. 7, with the region near the receiver shown on a larger scale below. To measure the effect of the hill in front of the receiver, another receiver recorded the field strength for a period of two weeks at the top of the hill some 400 feet above sea level and approximately 200 feet above the 2 to 10 mile average terrain. The profile near this receiver is shown in the middle profile of Fig. 7.

Radiosonde reports for the Seattle area, which were obtained from the U. S. Navy, included temperature, dewpoint temperature, and pressure as functions of height. These radiosondes are taken each morning and evening at 7 o'clock. Using the empirical constants suggested by E. K. Smith, Jr.,[3] the formula for the index of refraction is

$$(n - 1)\, 10^6 = \frac{77.6}{T}\left[p + \frac{4830\, e}{T}\right]$$

To simplify the numerous calculations of this quantity, a nomograph was devised which gives the quantity $(n - 1)\, 10^6$ directly from the data in the radiosonde reports. The nomograph shown in Fig. 8 is composed of three sections: multiplication to obtain 4830 e/T, addition to obtain the term within the bracket, and multiplication to obtain the final result. The values of $(n - 1)\, 10^6$ thus obtained were plotted as a function of height for each radiosonde. Two sample curves are shown in Fig. 9, displaced by 25 M units to avoid overlapping.

From approximately 178 M-curves chosen at random throughout a period of one year, 93 were selected that were reasonably linear in the height range below 10,000 feet. This height range was chosen since it is apparent from the profile shown in Fig. 7 that the signal must be propagated in the region above 3,000 feet in the middle of the path and therefore, the M gradient above this height would be most important. Fig. 10 shows a typical $(n - 1)\, 10^6$ curve and the corresponding M-curve with the 3,000 to 10,000 foot interval indicated by the double arrow. Using the M-gradient found from these curves, the signal may be calculated from (6).

The receiving site at the University of Washington is situated behind a 400-foot hill (See Fig. 7) and consequently signals arriving at very low angles must be diffracted. Simultaneous recordings atop this hill and at the University of Washington over a period of two weeks have shown the average difference between the hourly medians at these two locations is about 10 db. A plot of the difference versus the measured field strength atop the hill indicates a possibility that the difference increases with increasing signal. However, not enough data is available to show this conclusively. The average difference between the hourly medians agrees very well with the diffraction theory for a signal arriving at approximately 1 degree above the horizon.

Thus the field strength was calculated for the top of the hill using the heights above average terrain in the height-gain expression of (6). The average diffraction loss was then subtracted to obtain the field strength at the University of Washington site. This calculated field strength based on a linear atmosphere was then compared with the measured median field averaged for the two hours preceding the time of the radiosonde observation. A comparison of the calculated versus the measured field is shown in Fig. 11.

The error distribution is shown plotted on normal probability paper in Fig. 12. This figure shows the distribution is essentially normal, i.e., a straight line with a median error of approximately 11.5 db. Also the deviations were such that only 10% of the time did the error exceed 15 db while 90% of the time the error was greater than 8 db.

Conclusions

For the path considered with the assumption of linear index of refraction in the atmosphere between 3,000 and 10,000 foot heights, the field strength may be predicted within 15 db for 90% of the time. The predicted field strength is based on first mode propagation in which the gradient of the refractive index is the primary variable. The atmosphere is assumed to be

stratified vertically but homogeneous horizontally. The roughness of the terrain affects the heights at which the refraction is important at the middle of the path and is of particular importance near the receiving site. The calculated field strength was always lower than the measured field strength by a median value of 11.5 db. Thus apparently other mechanisms of propagation contribute to the increased field strength by this median value.

Acknowledgements

The authors wish to thank Mr. W. J. Siddons and Mr. Carlos Chong who performed much of the data reduction. This work was done through fellowships of the Engineering Experiment Station of the University of Washington and under contract CST-481 of the National Bureau of Standards.

Bibliography

1. T. J. Carroll: Tropospheric Propagation Well Beyond the Horizon, Trans. I. R. E. Prof. Group on Antennas and Propagation, 3, p. 98, August, 1952.

2. D. E. Kerr: Propagation of Short Radio Waves, MIT Radiation Laboratory Series, Vol. 13, pp 58-140; 1951

3. E. K. Smith, Jr.: The Constants of the Equation for the Refractive Index of Air, Trans. I. R. E. Prof. Group on Antennas and Propagation, 3, p. 179, August, 1952.

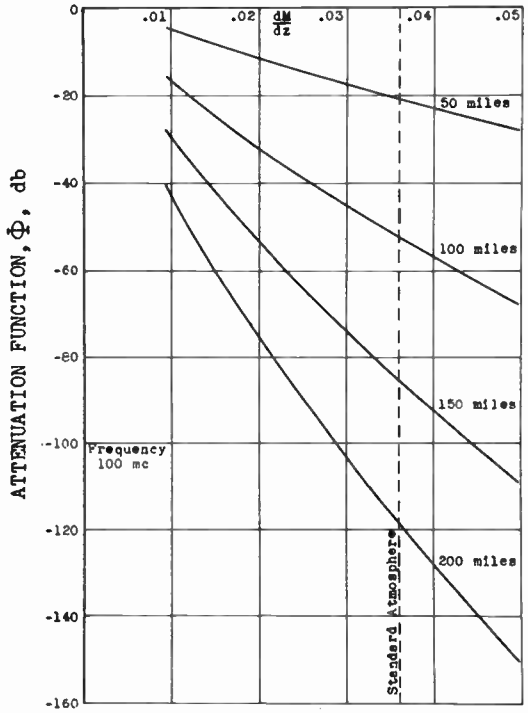


Fig. 1
Attenuation function vs. $\frac{dM}{dz}$ for linear $\frac{dM}{dz}$.

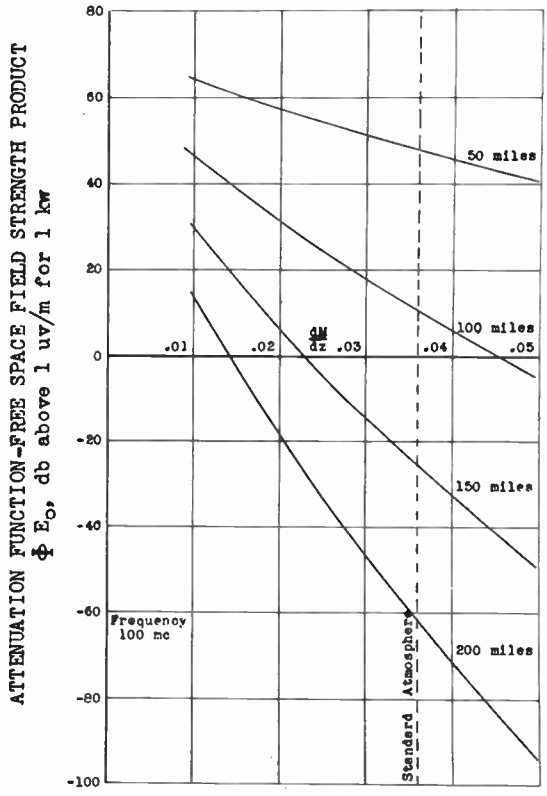


Fig. 2
Attenuation function-free space field strength product vs. $\frac{dM}{dz}$ for linear $\frac{dM}{dz}$.

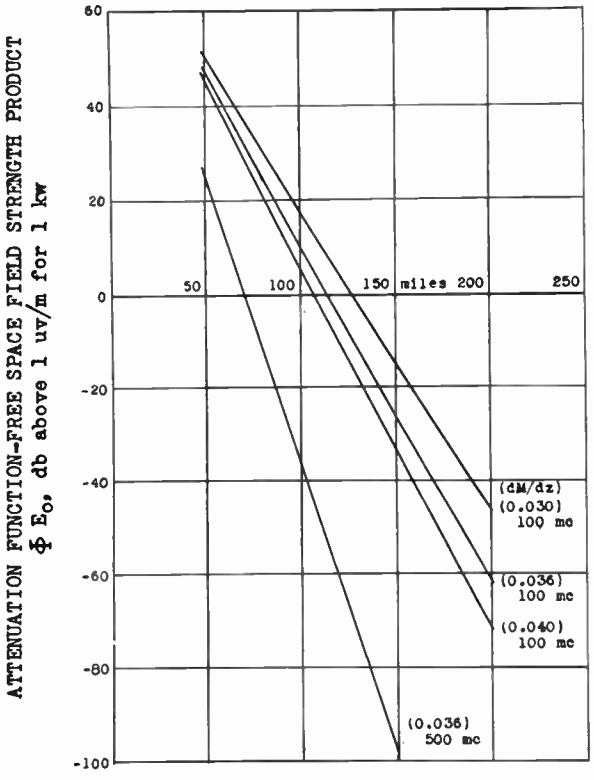


Fig. 3
Attenuation function-free space field strength product vs. distance for linear $\frac{dM}{dz}$.

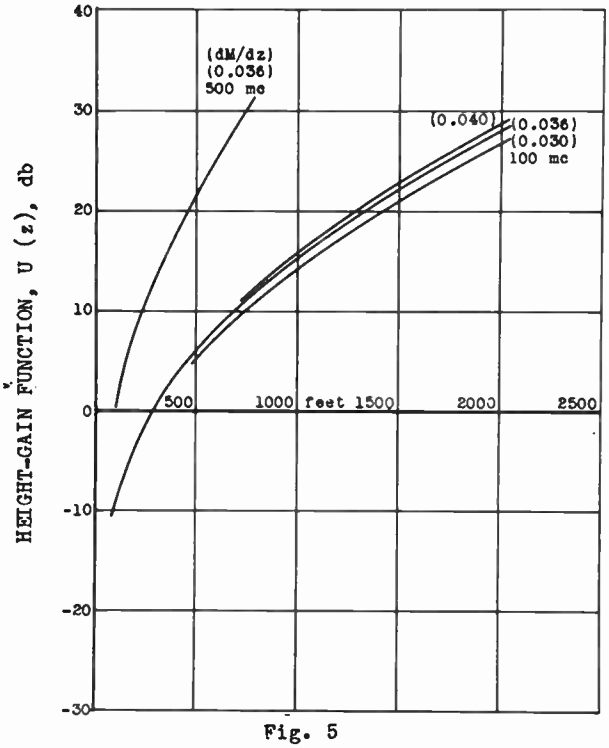


Fig. 5
Height-gain function vs. height for linear $\frac{dM}{dz}$.

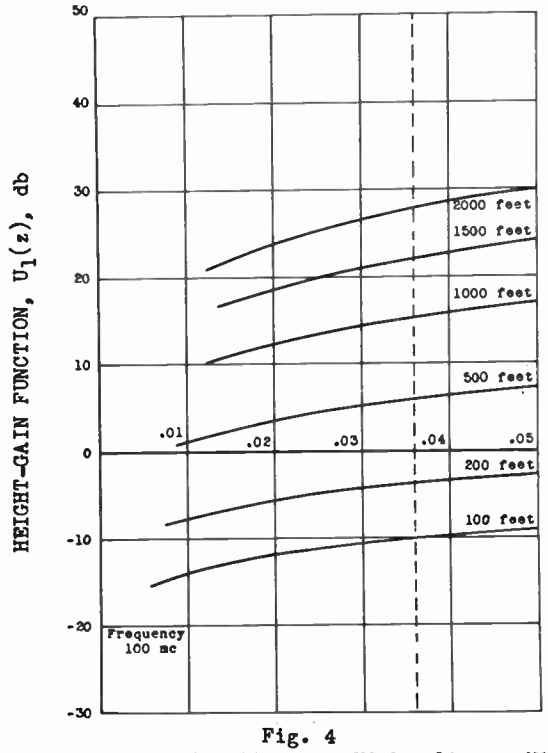


Fig. 4
Height-gain function vs. $\frac{dM}{dz}$ for linear $\frac{dM}{dz}$.

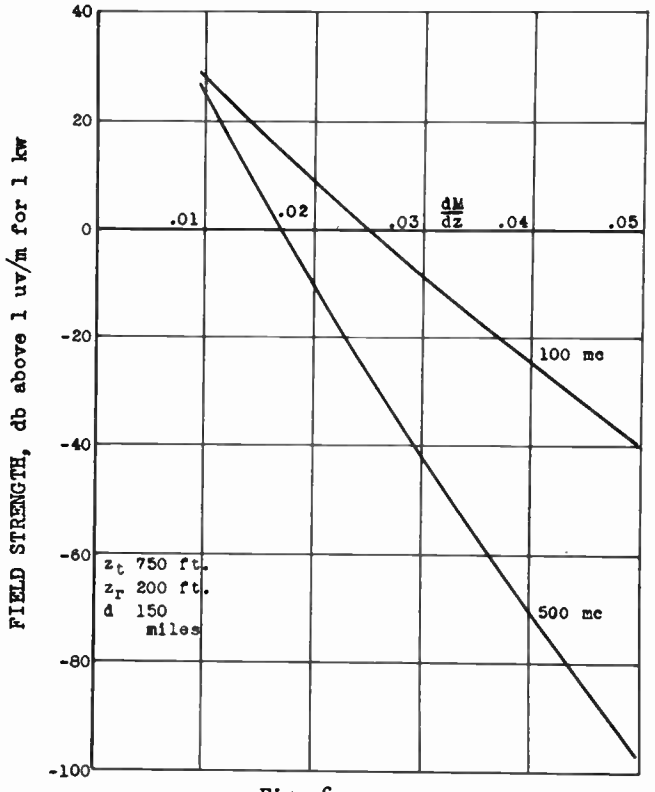


Fig. 6
Comparison of 100-mc and 500-mc propagation.

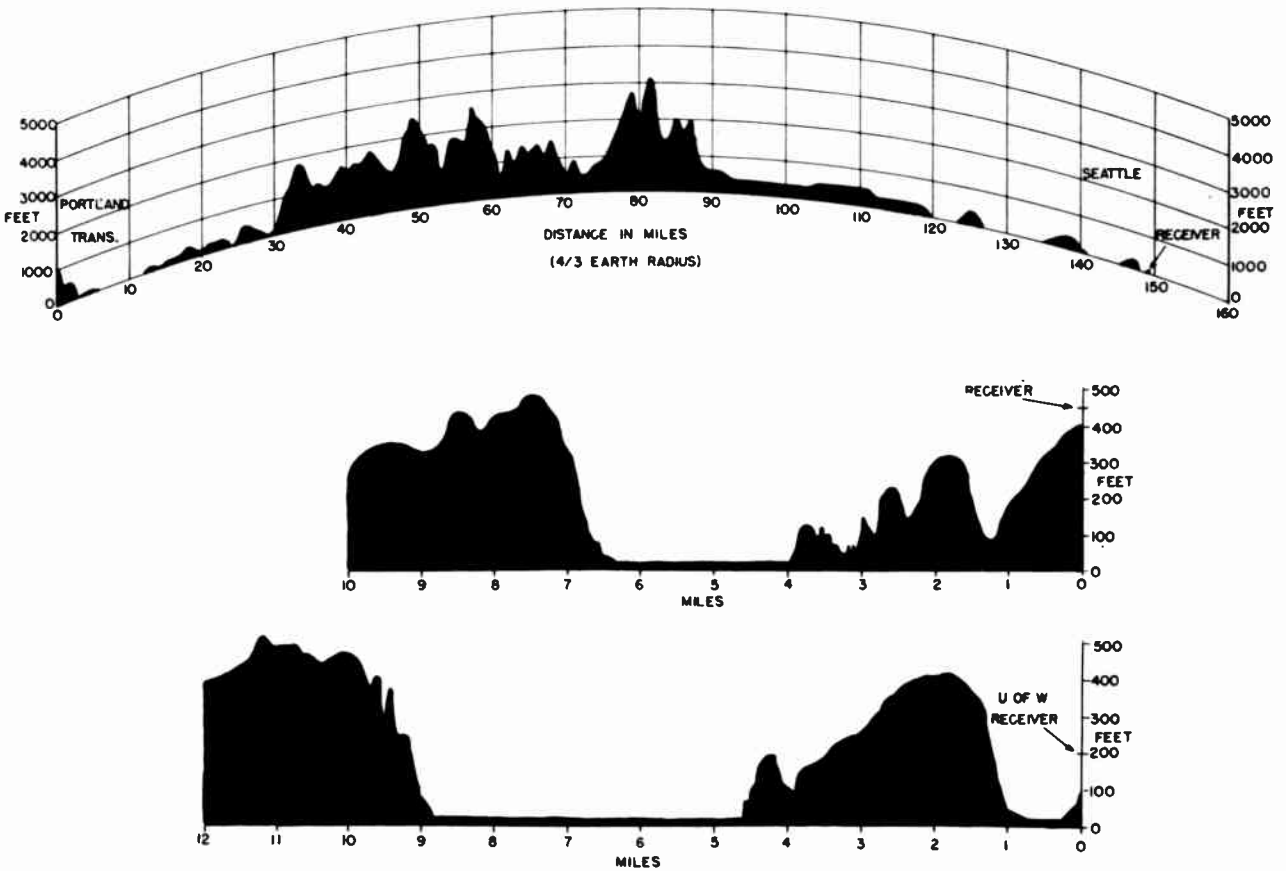


Fig. 7

Fig. 8

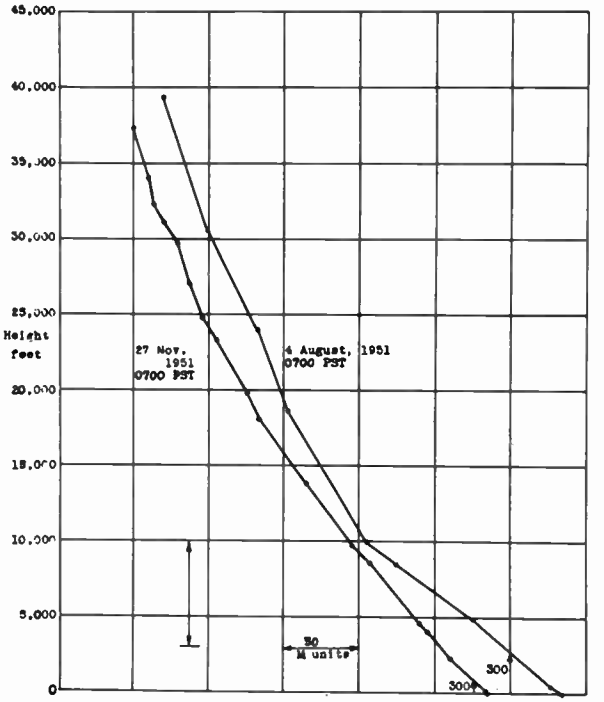


Fig. 9
$(n-1)10^6$ curves.

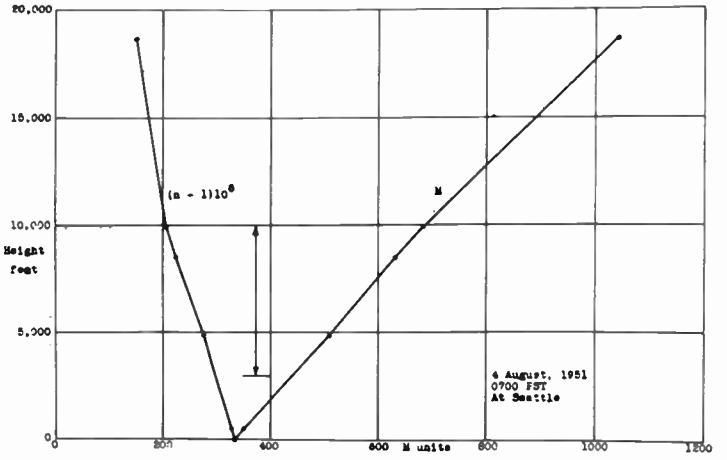


Fig. 10
Typical $(n-1)10^6$ and M curves.

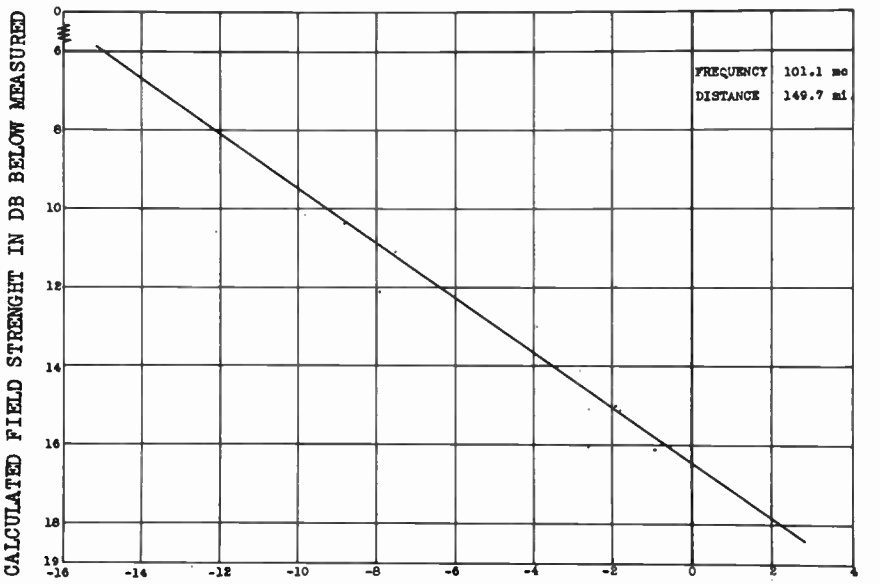


Fig. 11
Comparison of measured and calculated field strengths.

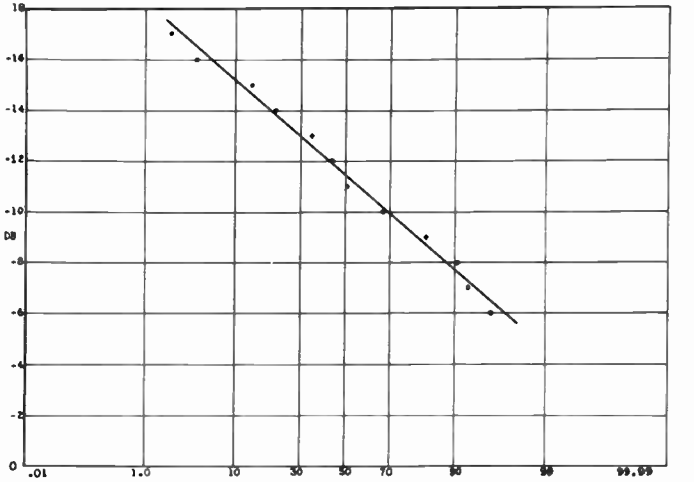


Fig. 12
Percent of time for which the error is greater than ordinate.

RADIO WAVE SCATTERING IN TROPOSPHERIC PROPAGATION

J. W. Herbstreit, K. A. Norton, P. L. Rice, and G. E. Schafer
National Bureau of Standards
Boulder, Colorado

Summary

The scattering theory of Booker and Gordon has been developed, assuming the correlation function $C(r) = C(0)\exp(-r/\ell)$, so as to be suitable for easy numerical calculation of the transmission loss expected with this mode of transmission; $C(0)$ denotes the variance with time of the refractive index of the atmosphere, and ℓ denotes the scale of turbulence. In this development the parameter, $[C(0)/\ell]$, emerges as a direct measure of the radio wave power transmitted by this mode of propagation. The theory is used to calculate the transmission loss to be expected on the transmission paths involved in the National Bureau of Standards Cheyenne Mountain experiment which cover the transmitting antenna height range from 30 to 7800 feet, distance range 223 to 628 miles, and frequency range 100-1046 Mc. The dependence of $[C(0)/\ell]$ on height above the earth's surface which must be assumed in these predictions to obtain agreement with the experimental transmission loss data is in qualitative agreement with the meager meteorological data which are now available for estimating this parameter.

The purposes of this paper are to further develop the scattering theory, recently outlined in papers by Pekeris,[1] Booker and Gordon,[2] Staras,[3] and LaGrone[4] and to check this theory for consistency with the tropospheric propagation data obtained by the National Bureau of Standards in connection with the Cheyenne Mountain experiment. A detailed description of most of the facilities used in this experiment was given in a recent paper by Chambers, Herbstreit and Norton.[5]

The concept of transmission loss developed in a recent paper by Norton[6] will be used in the following presentation; the transmission loss of a radio system is defined to be the ratio of the power radiated from the transmission antenna to the resulting signal power available from a loss-free receiving antenna.

Using the variables of integration, α, β, and ϕ defined in figures 1 and 2, it may be shown that the average transmission loss, in decibels, may be expressed:

$$L_a = -10 \log_{10}\left[\frac{\lambda^2}{2d_T(4\pi)^3}\int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} d\phi \int_{\beta_m}^{\pi-\beta} d\beta \int_{\alpha_m}^{\pi-\alpha_m} \frac{[C(0)/\ell]\, g_t g_r P_1 P_2 d\alpha}{\left[\sin^2\left(\frac{\alpha+\beta}{2}\right) + (\lambda/4\pi\ell)^2\right]^2}\right] \qquad (1)$$

In the above ℓ, d_T, and the free space wavelength, λ, are to be expressed in the same units.

The lower limits of integration, α_m and β_m in (1) may be expressed:

$$\alpha_m = \tan^{-1}(\tan \alpha_o / \cos \phi) \qquad (2)$$

$$\beta_m = \tan^{-1}(\tan \beta_o / \cos \phi) \qquad (3)$$

The transmitting and receiving antenna gains, g_t and g_r, may be expressed in terms

of four new angles ψ_t, γ_t, ψ_r, and γ_r. Thus dv can be located by orthogonal coordinates (ψ_t, γ_t, R_o) with origin at the transmitting antenna and another set of orthogonal coordinates (ψ_r, γ_r, R) with origin at the receiving antenna, the angles ψ_t and ψ_r being angles of elevation above the horizontal planes at the transmitter and receiver, respectively, and γ_t and γ_r measuring the angular displacement from the great circle plane. Thus:

$$\sin \gamma_t = \sin \alpha \sin \phi / \cos \psi_t \qquad (4)$$

$$\sin \gamma_r = \sin \beta \sin \phi / \cos \psi_r \qquad (5)$$

$$\sin \psi_t = \cos \alpha_o \sin \alpha \cos \phi - \sin \alpha_o \cos \alpha \qquad (6)$$

$$\sin \psi_r = \cos \beta_o \sin \beta \cos \phi - \sin \beta_o \cos \beta \qquad (7)$$

The above described geometry is directly applicable only for transmitting and receiving antennas on the surface of the earth, and the modifications and approximations involved in its use with elevated antennas will be presented elsewhere. The following approximate expressions are suitable for making allowance for the antenna gains:

$$g_t = g_t(\psi_t, \gamma_t)\ 4 \sin^2\left[\frac{2\pi}{\lambda} h_t' \sin \psi_t\right] \qquad (8)$$

$$g_r = g_r(\psi_r, \gamma_r)\ 4 \sin^2\left[\frac{2\pi}{\lambda} h_r' \sin \psi_r\right] \qquad (9)$$

In the above $g_t(\psi_t,\gamma_t)$ and $g_r(\psi_r,\gamma_r)$ represent the free space gains of the antennas and the remaining factors represent the lobe structure, arising from ground reflections. The heights h_t' and h_r' represent the heights of the antennas above the plane tangent at the earth at the point of ground reflection.

P_1 is a polarization factor to allow for the fact that a receiving antenna with a particular polarization will not respond equally well to the radiation received from all portions of the scattering volume. With horizontally polarized transmitting and receiving antennas, as were used in obtaining the results presented here, a negligible loss in accuracy is involved by setting this factor $P_1 = 1$.

The factor

$$P_2 = \left[1 + \frac{3\ell \sin(\alpha+\beta)}{4d \sin\left(\frac{\alpha+\beta}{2}\right)} \left(\frac{1}{\sin \alpha} + \frac{1}{\sin \beta}\right)\right] \qquad (10)$$

allows for the fact that the incident wave front in the scatterers is not quite plane; this phase factor becomes appreciable only at short distances and has been set equal to one in our present calculations.

In order to obtain an appreciation of the relative importance of the various parameters entering into (1), it will be evaluated first for several limiting cases under the assumption that $[C(0)/\ell]$ may be removed from under the sign of integration and replaced by its value $[C(0)/\ell]$ at the center of gravity of the resulting integral; the center of gravity lies in the great circle plane at the elevation angles ψ_t', ψ_r'. In the following approximate evaluations, we have set $P_1 = P_2 = 1$, and isotropic transmitting and receiving antennas are assumed so that $g_t(\psi_t,\gamma_t) = g_r(\psi_r,\gamma_r) = 1$.

Case I

$(h_t/\lambda).(\alpha_o+\beta_o) > 1$,
$(h_r/\lambda).(\alpha_o+\beta_o) > 1$ and
$(\ell/\lambda).(\alpha_o+\beta_o) > 1$

For this case (h_t/λ) and (h_r/λ) are both taken to be sufficiently large so that the oscillations of the sine squared factors in (8) and (9) occur within such small intervals of ψ_t and ψ_r that each of these factors may be replaced by its average value (1/2).* The product $g_t g_r$ in (1) thus becomes equal to 4; this factor arises due to the assumption, implicit in (8) and (9), that the earth is perfectly reflecting and thus doubles the scattered power available to the receiver twice, i.e. the power incident on the scatterers is doubled since it is confined to the half plane above the earth, and the power re-radiated from the scatterers is again doubled at the receiving antenna for the same reason.

* This rapid oscillation apparently violates the basic assumption,[2,3] made in deriving (1), that the incident field strength is constant over a macroscopic volume element of dimensions of the order of the scale of turbulence; the justification for this apparent violation will be presented elsewhere.

Thus (1) becomes:

$$L_a\left[(h_t/\lambda).(\alpha_o+\beta_o) > 1 \text{ and } (h_r/\lambda).(\alpha_o+\beta_o) > 1\right] = -10 \log_{10}\left[\frac{2\lambda^2\left[\overline{C(0)/\ell}\right] I(\ell/\lambda)}{d_T(4\pi)^3}\right] \quad (11)$$

$$\text{where } I(\ell/\lambda) = \int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} d\phi \int_{\beta_m}^{\pi-\beta} d\beta \int_{\alpha_m}^{\pi-\alpha_m} \frac{d\alpha}{\left[\sin^2\left(\frac{\alpha+\beta}{2}\right) + (\lambda/4\pi\ell)^2\right]^2} \quad (12)$$

In this case ℓ is assumed to be so large that $(\lambda/4\pi\ell) << \sin\left(\frac{\alpha_o+\beta_o}{2}\right)$ and we will write $I\left[(\ell/\lambda) = \infty\right] \equiv I_o$; the integrations with respect to α and β may then be made:

$$I_o = \int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} I(\alpha_m+\beta_m)\,d\phi \equiv \phi_1(\alpha_o+\beta_o) I(\alpha_o+\beta_o) \quad (13)$$

where

$$I(\alpha_m+\beta_m) \equiv \frac{2}{3}\cot^2\left(\frac{\alpha_m+\beta_m}{2}\right) - \frac{8}{3}\log \sin\left(\frac{\alpha_m+\beta_m}{2}\right) \quad (14)$$

The factor $\phi_1(\alpha_o+\beta_o)$ has been evaluated by graphical methods and its values for the cases $\alpha_o=\beta_o$, $\alpha_o=9\beta_o$ and $\beta_o=9\alpha_o$, together with the resulting values of I_o, are given on figure 3.

In this case $I(\ell/\lambda) \cong I_o$ and is thus independent of the scale of turbulence or the frequency; thus we see by (11) that the scale of turbulence will enter into our calculations only through the factor $\left[\overline{C(0)/\ell}\right]$. Furthermore, it may be shown for this case that the center of gravity of (12) is at the elevation angles:

$$\psi_t'' = \psi_r'' = \alpha_o+\beta_o \quad (15)$$

From the geometry of figure 1 we obtain:

$$\alpha_o = \frac{d}{2a} - \frac{d_{Lt}}{a} + \frac{(h_t-h_r)}{d} \quad (16)$$

$$\beta_o = \frac{d}{2a} - \frac{d_{Lr}}{a} - \frac{(h_t-h_r)}{d} \quad (17)$$

$$\alpha_o+\beta_o = \frac{d-d_{Lt}-d_{Lr}}{a} \quad (18)$$

In the above $d_{Lt} \cong \sqrt{2ah_t}$ and $d_{Lr} \cong \sqrt{2ah_r}$ and represent the distances to the radio horizons of the transmitting and receiving antennas, respectively. We see by (18) that the important parameter $(\alpha_o+\beta_o)$ is directly proportional to the path distance beyond the radio horizon. An effective radius of the earth, a, larger than the actual radius, a', by the factor

$$\left\{1/\left[1 + \frac{a'}{n}\cdot\frac{\Delta n}{\Delta h}\right]\right\}$$

is employed in order to allow for the systematic effects of air refraction. In the following calculations this factor is set equal to 4/3 and then a becomes 5280 miles.

Note that L_a is, in this limiting case (i.e. h_t, h_r and ℓ large or λ very small), a function of the frequency only through the factor $(\lambda^2/4\pi)$ in (11) which is the absorbing area of an isotropic receiving antenna; thus the field strength or the attenuation relative to the free space value will be independent of the frequency in this particular limiting case.

Case II

$(h_t/\lambda).(\alpha_o+\beta_o) > 1$ and
$(h_r/\lambda).(\alpha_o+\beta_o) > 1$ but
$(\ell/\lambda).(\alpha_o+\beta_o) < 1$

If we let $c = (\lambda/4\pi\ell)^2$ and make the substitution $t = \cot\left(\frac{\alpha+\beta}{2}\right)$ in (12), it may be integrated with respect to β:

$$I(\ell/\lambda) = \int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} d\phi \int_{\alpha_m}^{\pi-\beta_m} f\left[\alpha,\beta_m,(\ell/\lambda)\right] d\alpha \equiv \phi_2\left[(\ell/\lambda),\alpha_o+\beta_o\right] \int_{\alpha_o}^{\pi-\beta_o} f\left[\alpha,\beta_o,(\ell/\lambda)\right] d\alpha \qquad (19)$$

where
$$f\left[\alpha,\beta_m,(\ell/\lambda)\right] \equiv \left[c(1+c)\right]^{-3/2} \left[(1+2c)\tan^{-1} X - \frac{X}{1+X^2}\right] \qquad (20)$$

and
$$X = \sqrt{\frac{c}{1+c}} \cot\left(\frac{\alpha+\beta_m}{2}\right) \qquad (21)$$

Graphical integration of (19) first with respect to α and then with respect to ϕ, assuming $\alpha_o=\beta_o$, provided the curves of $\phi_2\left[(\ell/\lambda),\alpha_o+\beta_o\right]$ and $F\left[(\ell/\lambda),\alpha_o+\beta_o\right]$ shown on figure 4. Note that we may write:

$$10 \log_{10} I(\ell/\lambda) = 10 \log_{10} I_o + F\left[(\ell/\lambda),\alpha_o+\beta_o\right] \qquad (22)$$

and we see by figure 4 that $F\left[(\ell/\lambda),\alpha_o+\beta_o\right]$ is practically negligible (i.e. nearly equal to zero) for scales of turbulence, frequencies, and distances such that $(\ell/\lambda).(\alpha_o+\beta_o) > 1$; fortunately this relation was amply satisfied for all of the experimental data considered in this paper with the result that the scale of turbulence enters the calculations only through the factor $\left[\bar{C}(0)/\ell\right]$.

Case III

$(\ell/\lambda).(\alpha_o+\beta_o) > 1$ and either
$(h_t/\lambda).(\alpha_o+\beta_o) > 1$ or
$(h_r/\lambda).(\alpha_o+\beta_o) > 1$

When one (but not both) of the antennas is low (we will evaluate the particular case of a low transmitting antenna but reciprocity makes the results applicable to the case of a low receiving antenna), we may express the median transmission loss as follows:

The first term in (23) arises from the fact that the median transmission loss for Rayleigh distributed losses is larger by 1.5915 decibels than the average transmission loss.[6] The term $H_o\left[(h_t/\lambda),\alpha_o+\beta_o\right]$ has been determined by graphical integration of (1) and is shown on figure 5 as a function of $(h_t/\lambda).(\alpha_o+\beta_o)$. We see by this figure that h_t may be considered to be large provided $(h_t/\lambda).(\alpha_o+\beta_o) > 1$. For lower antennas the lobes in the ground reflection pattern do not illuminate the scattering volume uniformly, and there is an addition to the transmission loss. The elevation angle of the center of gravity is raised in this case to the value ψ_t'' which is also shown graphically on figure 5.

Case IV

Isotropic antennas and $\left[\bar{C}(0)/\ell\right]$ replaced by its value $\left[\bar{C}(0)/\ell\right]''$ at the center of gravity ψ_t'', ψ_r''.

When both antennas are low we may, to a first approximation, neglect the interaction of the transmitting and receiving antennas (this interaction principally affects the center of gravity), neglect the difference between $\phi_1(\alpha_o+\beta_o)$ and $\phi_2\left[(\ell/\lambda),\alpha_o+\beta_o\right]$ and obtain our final approximate formula for the expected median transmission loss:

$$L_m\left[(h_r/\lambda).(\alpha_o+\beta_o)>1 \text{ and } (\ell/\lambda).(\alpha_o+\beta_o)>1\right] = 1.5915 - 10 \log_{10}\left[\frac{2\lambda\left[\bar{C}(0)/\ell\right]I_o}{d_T(4\pi)^3}\right] - H_o(h_t/\lambda) \qquad (23)$$

$$L_m \cong 1.5915 - 10 \log_{10}\left[\frac{2\lambda^2 \overline{[C(0)/\ell]} I_o}{d_T (4\pi)^3}\right] - H_o(h_t/\lambda) - H_o(h_r/\lambda) - F(\ell/\lambda) \qquad (24)$$

By substituting the measured median transmission loss[7] on the Cheyenne Mountain paths for L_m in (24), it is possible to estimate the values of $\overline{[C(0)/\ell]}$ at the centers of gravity of the integrals. Table I gives the values of the various parameters required for these calculations together with the resulting values of $\overline{[C(0)/\ell]}$.

The height, h, above the surface in the great circle plane corresponding to particular values of α and β may be determined from the following approximate expression:

$$h \cong 0.5(h_t+h_r) + \frac{\alpha\beta d}{\alpha+\beta}\left[1 - \frac{d}{2a(\alpha+\beta)}\right] \qquad (25)$$

In Table I, h_o represents the lowest height involved in the integration, i.e. the height obtained by setting $\alpha=\alpha_o$ and $\beta=\beta_o$ while h_c represents the height of the center of gravity obtained by setting $\alpha=\alpha_o+\psi_t''$ and $\beta=\beta_o+\psi_r''$.

In these approximate evaluations of $\overline{[C(0)/\ell]}$ it has been assumed that the transmitting and receiving antenna free space gains are realized (i.e. $g_t(\psi_t,\gamma_t)$ and $g_r(\psi_r,\gamma_r)$ are replaced by the fixed values $g_t(o,o)$ and $g_r(o,o)$, and this constitutes one of the errors in these estimates.

TABLE I

DATA FROM THE CHEYENNE MOUNTAIN EXPERIMENT

No.	d (miles)	f_{mc} (Mc)	h_t (feet)	h_r (feet)	$\alpha_o+\beta_o$ (radians)	$(h_r/\lambda)\cdot(\alpha_o+\beta_o)$	$(h_t/\lambda)\cdot(\alpha_o+\beta_o)$	$(\ell/\lambda)\cdot(\alpha_o+\beta_o)$ (For ℓ = 300 meters) (db)	$H_o[(h_t/\lambda),(\alpha_o+\beta_o)]$ (db)	$H_o[(h_r/\lambda),(\alpha_o+\beta_o)]$ (db)
1	237.1	100	6920	20.1	0.02079	.0425	14.6271	2.08	0	-7.35
2	226.5	100	2685	19.7	0.02828	.0566	7.7187	2.83	0	-6.00
3	226.5	1046	2575	30.6	0.02675	.8705	73.249	28.0	0	-0.25
4	223.6	100	30	19.75	0.03787	.0760	0.11551	3.79	-3.22	-4.70
5	404.1	100	7600	40.9	0.05091	.2117	39.340	5.10	0	-1.63
6	393.5	100	2890	41.0	0.05845	.2436	17.1737	5.85	0	-1.41
7	393.5	1046	2840	7.5	0.05972	.4763	180.384	62.5	0	-0.59
8	390.7	100	30	40.6	0.06808	.2810	0.20765	6.81	-1.92	-1.19
9	628.1	100	7880	38	0.09157	.3538	73.361	9.16	0	-0.83
10	617.7	100	2865	38	0.09840	.3802	28.663	9.84	0	-0.80
11	614.0	100	30	38	0.1076	.4157	0.3282	10.8	-0.98	-0.66

TABLE I (continued)

DATA FROM THE CHEYENNE MOUNTAIN EXPERIMENT

No.	Ψ''_t (radians)	Ψ''_r (radians)	$g_t(o,o)$ (db)	$g_r(o,o)$ (db)	10 $\log_{10} I_o$ (db)	L_m (db)	$\overline{[C(0)/\ell]}$ (per meter)	h_o (feet)	h_c (feet)	σ (db)
1	0.02079	0.07904	8.55	2.15	39.87	184.6	6.00×10^{-16}	5,271	25,791	6.63
2	0.02828	0.08840	9.98	2.15	37.08	185.1	8.65×10^{-16}	3,959	30,458	6.05
3	0.02675	0.02675	26.0	2.15	37.72	178.0	2.74×10^{-15}	3,574	19,331	5.03
4	0.07385	0.09712	18.2	2.15	34.53	183.7	4.90×10^{-16}	4,867	55,130	6.12
5	0.05091	0.06852	8.46	16.40	32.06	180.4	2.82×10^{-16}	11,293	72,700	5.11
6	0.05845	0.07359	9.98	16.40	30.87	179.4	3.02×10^{-16}	12,918	80,228	5.38
7	0.05972	0.05984	26.0	25.65	30.71	174.3	2.78×10^{-16}	13,452	75,565	5.42
8	0.09293	0.08020	17.9	16.2	30.56	178.7	9.42×10^{-17}	16,800	105,790	3.51
9	0.09157	0.09899	8.46	19.40	27.17	192.8	3.23×10^{-17}	31,708	189,246	4.41
10	0.09840	0.10352	9.98	19.55	26.59	194.8	1.55×10^{-17}	35,350	199,670	3.59
11	0.1193	0.1117	17.6	19.95	25.80	191.2	8.05×10^{-18}	41,614	228,737	3.28

The determinations of α_o, β_o, h_t, h_r, h_o, and h_c all present special problems over the irregular terrain involved in the Cheyenne Mountain experiment. Space does not permit a discussion here of the particular methods used in solving these special problems, and they will be described elsewhere.

The estimates of $\overline{[C(0)/\ell]}$ given in Table I and obtained by comparing the measured median transmission loss with that calculated using the approximate formula (24) are shown in figure 6 as a function of the height of the approximate centers of gravity, also given in Table I. These values, together with meteorological data available at still lower heights,[8,9] suggest that we may express the average value of $\overline{[C(0)/\ell]}$ as an exponential function of the height:

$$\overline{[C(0)/\ell]} = Ae^{-Bh} \qquad (26)$$

Using this form of variation of $\overline{[C(0)/\ell]}$ with height and assuming $B = 1.62 \times 10^{-4}$ per meter, we have re-calculated L_m for transmission path No. 8 of Table I. These calculations have been made by numerical methods and allow not only for the variations of $\overline{[C(0)/\ell]}$ with height but also allow for the loss in gain due to leaving the free space gain factors under the sign of integration. Finally, by comparing this calculated value of L_m with the value of L_m measured on this path, the value of A in (26) has been determined to be 8.24×10^{-15} per meter. The curve of $\overline{[C(0)/\ell]}$ versus height on figure 6 was determined in this way. The comparatively small scattering of the points on figure 6 relative to the line representing our tentative estimate of $\overline{[C(0)/\ell]}$, constitutes very favorable evidence as to the ability of the scattering theory to accurately predict the influence on transmission loss of distance, frequency and antenna height at points far beyond the horizon. The center of gravity of the integrand of (1), as calculated by the more precise methods described above for path No. 8, was equal to 19.6 kilometers instead of the value 32.2 kilometers estimated by means of the values Ψ''_t and Ψ''_r determined from figure 5. Using this value, we see by figure 5 the error (over 5 decibels) involved in the use of (24) for calculating $\overline{[C(0)/\ell]}$ for a given value of L_m.

The above described more accurate method for evaluating the integral (1) for path No. 8 is now being employed for all of our Cheyenne Mountain transmission paths and, when completed, should provide a still better estimate of the probable variation of $[C(0)/\ell]$ with height and of the validity of the scattering theory.

By calculating for path No. 8 the difference in transmission losses (a) with the free space gain factors under the sign of integration and (b) replaced by the constant values $g_t(o,o)$ and $g_r(o,o)$ we find that the loss in antenna gain relative to that expected in free space is equal to 2.58 decibels for this particular transmission path. In addition to the above evidence in favor of the Booker-Gordon theory that scattering can be the principal mode of transmission to large distances in this frequency range, two other experimental results also provide confirmation. The first is the fact that the instantaneous transmission loss is Rayleigh distributed over the short periods of time (of the order of an hour) during which the average transmission loss remains substantially constant; the evidence for this is presented in a forthcoming paper by Barsis, Janes, and Roubique.[10] Finally the hour-to-hour and day-to-day variations in hourly median transmission loss also behave in a manner consistent with the scattering theory. It has been found that these variations are normally distributed when expressed in decibels,[5] and the estimates of the standard deviation of these approximately normal distributions given in Table I were obtained from the following relation:

$$\sigma = 0.39\,[F_m(90 \text{ percent}) - F_m(10 \text{ percent})] \quad (27)$$

where $F_m(90 \text{ percent})$ and $F_m(10 \text{ percent})$ denote the hourly median transmission losses exceeded for 90 percent and 10 percent of the hours, respectively. Reference to (1) or (24) indicates that the principal variations with time in the median transmission loss would be expected to arise from variations in $[C(0)/\ell]$. Thus σ may, in accordance with our theory, be assumed to measure the variance in $[C(0)/\ell]$. We see by Table I that σ decreases more or less monotonically with the height, h_c, of the centers of gravity of the integrals, and we conclude that the atmosphere is not only less turbulent at the higher heights but also that this turbulence is less variable from hour to hour and from day to day.

References

1 C. L. Pekeris, "Note on Scattering in an Inhomogeneous Medium," Physical Review, Vol. 71, p. 268, February 1947.

2 H. G. Booker and W. E. Gordon, "A Theory of Radio Scattering in the Troposphere," Proc. IRE, Vol. 38, pp. 401-412, April 1950.

3 Harold Staras, "Scattering of Electromagnetic Energy in a Randomly Inhomogeneous Atmosphere," National Bureau of Standards Report 1662, May 12, 1952; also Journal of Applied Physics, Vol. 23, pp. 1152-1156, October 1952.

4 A. H. LaGrone, "Volume Integration of Scattered Radio Waves," Proc. IRE, Vol. 40, p. 54, January 1952.

5 G. R. Chambers, J. W. Herbstreit, and K. A. Norton, "Preliminary Report on Propagation Measurements from 92-1046 Mc at Cheyenne Mountain, Colorado," National Bureau of Standards Report 1826, July 23, 1952.

6 K. A. Norton, "Transmission Loss in Radio Propagation," Proc. IRE, Vol. 41, pp. 146-152, January 1953.

7 A. P. Barsis, "Report on Comparative 100-Mc Measurements for Three Transmitting Antenna Heights," in preparation.

8 George Birnbaum, H. E. Bussey, and R. R. Larson, "The Microwave Measurement of Variations in Atmospheric Refractive Index," Trans. of the IRE Professional Group on Antennas and Propagation, PGAP-3, pp. 74-78, August 1952. Also, National Bureau of Standards Report 1874.

9 C. E. Von Rosenberg, C. M. Crain, A.W. Straiton, "Atmospheric Refractive-Index Fluctuations as Recorded by an Airborne Microwave Refractometer," Electrical Engineering Research Laboratory of the University of Texas, Report No. 6-01, February 6, 1953. Data shown on figure 6 were measured in Ohio in July and August 1952.

10 A.P. Barsis, H.B. Janes, C.J.Roubique, "An Analysis of Short-Term Fading in 100-1000-Mc Propagation Measurements beyond the Radio Horizon," in preparation.

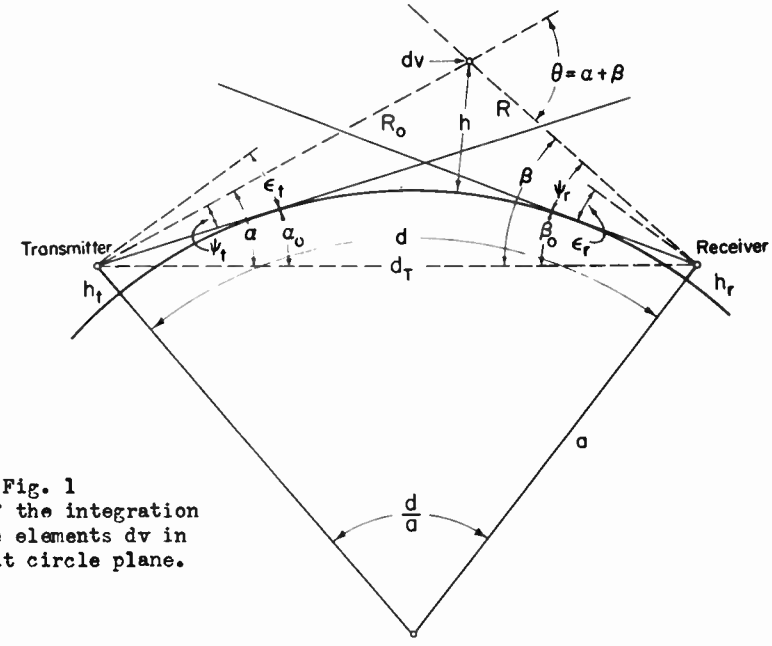


Fig. 1
Geometry of the integration for volume elements dv in the great circle plane.

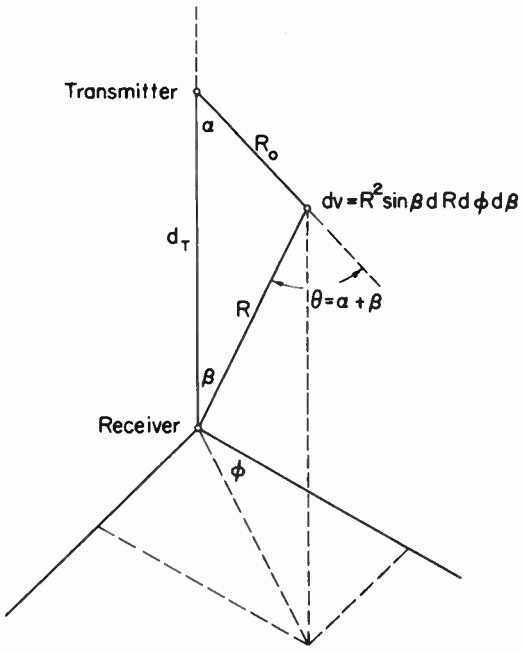


Fig. 2

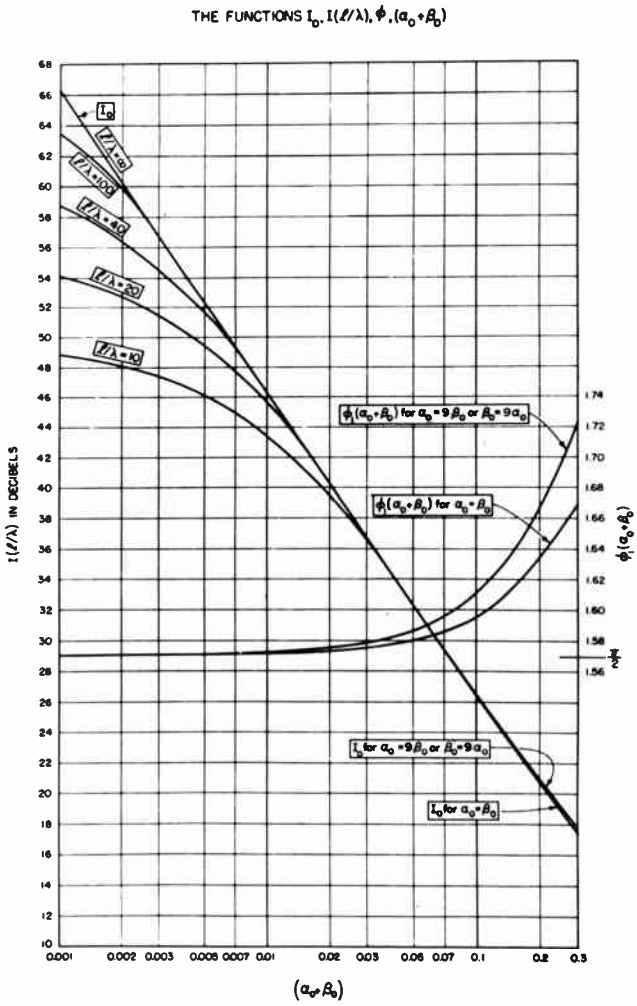


Fig. 3

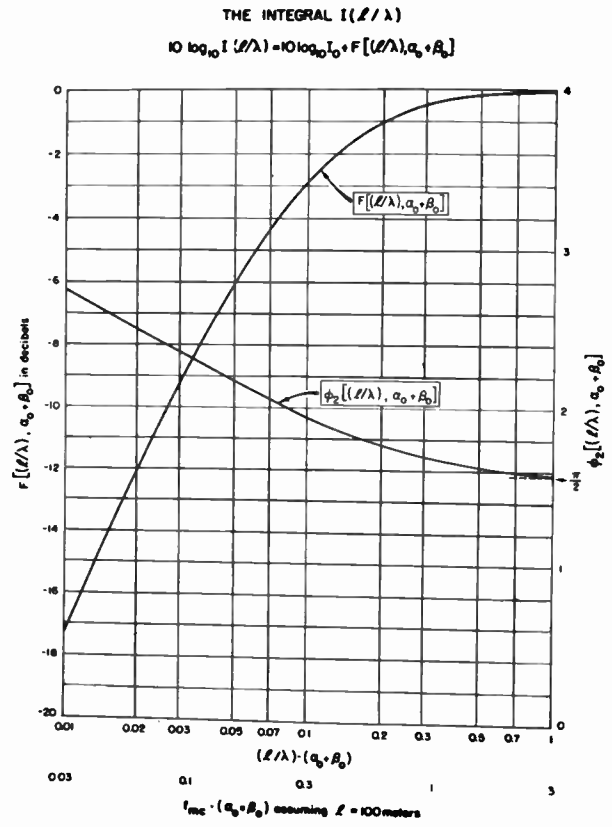


Fig. 4

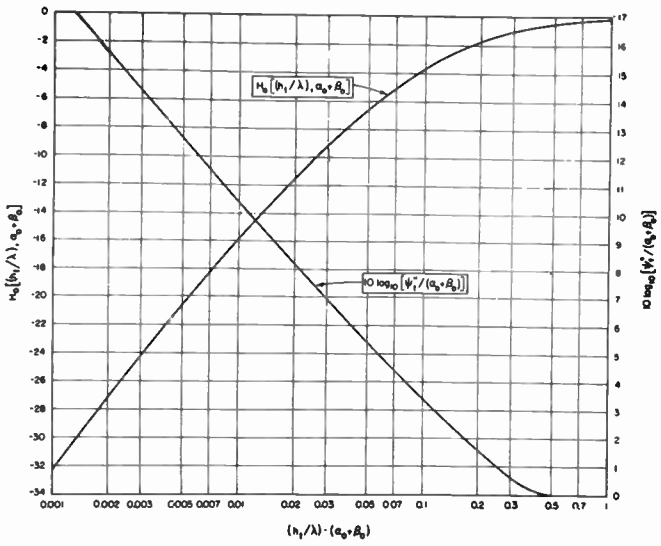


Fig. 5 - The frequency-gain function.

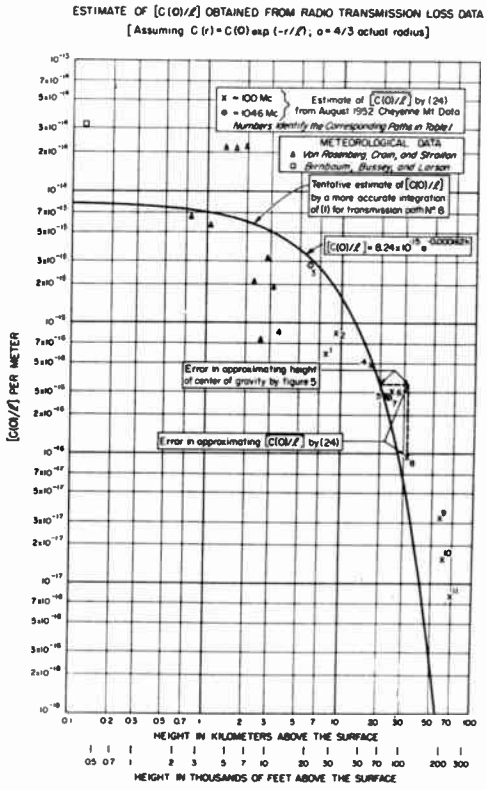


Fig. 6

EXTENDED-RANGE RADIO TRANSMISSION BY OBLIQUE REFLECTION FROM METEORIC IONIZATION

O. G. Villard, Jr., A. M. Peterson, L. A. Manning and V. R. Eshleman
Radio Propagation Laboratory
Stanford University
Stanford, California

This work was in part supported by the Navy Department (Office of Naval Research), the U. S. Army (Signal Corps) and the U. S. Air Force under ONR Contract N6-ONR-251 Consolidated Task No. 7.

ABSTRACT

It has been found that radio communication between relatively low-power stations operating at 14 megacycles and separated by distances of roughly 1200 km may be maintained at times when no layer transmission to any point on the earth's surface can be demonstrated to be present. The signal obtained is subject to considerable fading, but some signal is nearly always detectable. The contribution of overlapping oblique-incidence meteor reflections to the observed signal is considered in the light of some preliminary theoretical and experimental findings. It is clearly important to assess the meteoric contribution with care, since the possibility that meteoric reflections alone could account for the signal does not seem unreasonable.

INTRODUCTION

During tests of a technique for determining those areas on the earth's surface to which strong radio transmission is possible at any given time and frequency,[1,2] it was found that radio communication with a station 1200 km away was still possible at a time when no sizeable reflecting region - such as the F, E, or sporadic-E layer - could be shown to be present. The frequency employed was 14.3 megacycles. Both stations had power outputs of the order of 500 watts, and were located at sites free of obstructions to low-vertical-angle radiation. Rotatable beam antennas mounted at heights of roughly 40 feet, and having gains of 6-8 db, were used. The receivers were of the standard communications type having fair sensitivity and half-power bandwidths of the order of 4 kilocycles.

Characteristics of the signal obtained under these conditions may be summarized as follows. The strength varied over wide limits, dipping occasionally into the noise level, but frequently rising as much as twenty or thirty decibels above. The 5-minute average value is perhaps 10 decibels above the noise. The most significant feature of this signal is the fact that it is always present, except for the fades. It has been observed on a large enough number of occasions to rule out the possibility that it might be a transient or sporadic effect. With suitable limiting, the signal is entirely satisfactory for hand-keyed cw transmission.

The sudden, strong increases in signal strength are clearly caused by reflections from individual meteor ionization trails. The relatively continuous background, however, requires further explanation. It is the purpose of this paper to investigate the contribution to this signal provided by overlapping reflections from the relatively large number of meteor trails present at any given time.

ABSENCE OF LAYER REFLECTIONS

In order to establish that no layer-type transmission was contributing to the measured field strength, the scatter-sounding apparatus described in reference 1 was used. This consisted of a one-kilowatt 14-megacycle transmitter connected to a rotatable directional antenna and adapted so as to excite and display back-scatter echoes from the ground. These echoes appear in detectable strength only when a radio-reflecting region of reasonable size (of the order of several hundred square kilometers or larger) is present. Tests have shown the indication to be quite sensitive; as soon as one-way layer-propagated signals of medium strength are heard, back-scattered echoes appear on the indicator.

Scatter-soundings and transmission tests were performed at Stanford University, California between the hours of 11 p.m. and 4 a.m., Pacific Standard Time, during the months of December, 1951 and January, 1952. The remote station providing the opposite terminus of the transmission path was located at Spokane, Washington, at a bearing of 16 degrees true from Stanford, California.

At those hours and at that time of year, F-layer transmission is almost invariably absent. Sporadic-E transmission, however, is always a possibility. Although the presence of a sporadic E cloud of the usual type along the transmission path would manifest itself as a tremendous increase in signal strength over that being considered here, together with a large reduction in the average fading period, the possibility always exists that

such a cloud might form to one side of the path and contribute to the total observed signal at each terminus by oblique back-scattering from the ground. The strength and fading characteristics of signals propagated by such oblique back-scattering make them very difficult to distinguish from signals propagated by the type of transmission being considered herewith. Hence it is essential in tests of this sort to rule out the possibility of a back-scattered contribution to the total signal.

To demonstrate the absence of a back-scattered component, repeated scatter-soundings were made during the course of the transmission tests. On about half the nights during the period under discussion, sporadic E patches of varying intensity, and in varying locations, were found with the aid of the scatter-sounder. (A vertical-incidence model C-3 ionosphere recorder, operated at Stanfore University, showed far less sporadic-E activity, as would be expected in view of its limited area of surveillance.) Since tests on these occasions showed clearly that a suitably-located off-path sporadic-E patch did contribute to the total signal, forward-transmission test results were considered reliable only when all detectable sporadic-E was absent.

Use of rotatable unidirectional antennas at both ends of the forward-transmission circuit provided another means for assessing the contribution of off-path scatter to the total signal. In the absence of sporadic-E, it was found that if either antenna was directed away from the other station, the steady background component of the signal faded down and was soon lost. In fact, the signal was readily detectable at the aforementioned power levels only when the extra gain provided by the beam antennas was present. When an off-path sporadic-E patch was present, a relatively strong signal could usually be obtained when each antenna was pointed not at the opposite station, but rather in the general direction of the scatter patch and the ground scattering area which lies beyond it.

FURTHER TRANSMISSION CHARACTERISTICS

It was found that an increase in transmission frequency caused the background signal to decrease very rapidly. Tests with Spokane using moderately high-power pulsed transmissions at 23 megacycles showed that the steady background had disappeared; a signal was received only during the strongest meteor bursts. The same was true for continuous-wave signals over a similar path when the operating frequency was increased from 14.2 to 28.6 megacycles, with all other circumstances of the test essentially equivalent.

In an effort to determine the maximum range at which this type of transmission could be maintained, a 14.2 megacycle continuous-wave test was made between Stanford, California and a station in Aberdeen, South Dakota, which is some 2280 kilometers distant. This range is close to the maximum for one-hop transmission via E-region reflection. Once again, the antenna arrangement, power, and site characteristics at both ends of the circuit were substantially equivalent. It was found that signals from the opposite station could be received only at relatively infrequent intervals by reflection from individual meteor columns. There was no trace of a steady background.

COMPARISON OF FORWARD AND BACK-REFLECTION RECORDINGS

Field strength recordings of the 1200-kilometer 14-megacycle transmissions were made with the aid of a direct-inking magnetic "penmotor" recorder, capable of a response up to 100 cps. These records were found to have a great many features in common with recordings previously made with the same equipment at Stanford University for the purpose of meteor ionization studies.[3,4] Continuous-wave signals at approximately the same power levels and operating frequencies had been used for the meteor investigations, which had been concerned with backward reflection from ionization trails at perpendicular incidence.

The most striking difference between the backward and oblique recordings is the presence in the latter of a steady background component. In both cases, individual meteor reflections can be clearly discerned insofar as their time of beginning is concerned. Although the duration of each echo is well-defined at perpendicular incidence, since the individual echoes are well separated along the trace, at oblique incidence this duration cannot readily be determined. It is clear that the individual echoes have, on the average, considerably increased durations, although the exact time of ending of each one tends to be obscured since it fades into and merges with the continuously-fluctuating background.

The question at once arises, can the background signal itself be accounted for as a consequence of the overlapping of many individual meteor echoes? Clearly, if overlapping occurs on a consistent-enough basis, a mechanism is at hand whereby continuous transmission could be supported. In partial answer to this question, it is first necessary to consider the expected increase in echo duration when reflection from a meteor column occurs at oblique rather than perpendicular incidence.

THEORETICAL CONSIDERATIONS

On the basis of experience a model electron column has been postulated[5] whose radio echo characteristics agree fairly closely with those of the majority of meteor echoes observed in practice.

Using generally-accepted recombination and diffusion coefficients for the "E" region of the ionosphere, in which the majority of ion columns occur, it may be shown that if the number of electrons per meter of trail length is less than 10^{14}, and the initial radius of the column is of the order of a molecular mean free path (a

reasonable assumption), there is virtually no loss of electrons by recombination. Back-reflection meteor studies made at Stanford University[3] at a frequency of 23 megacycles have indicated that the majority of meteor trails - i.e. those whose characteristics are best matched by the model - have a line density of less than 10^{14} electrons per meter. On the basis of this and other evidence, recombination is not believed to be an important factor in meteor ionization.

The model column, accordingly, has an initial size as given above, and expands outward in accordance with the classical diffusion equation as applied to cylindrical geometry. Solution of this equation leads to a Gaussian distribution of electron density as a function of radius measured from the axis of the column. Although it is highly probable that actual meteor trails have a more complicated distribution of ionization than this, the Gaussian distribution represents a useful working model.

Accepting these premises, we have a column at first small compared with a wavelength. When a radio wave strikes this column perpendicularly, the contributions of each individual electron to the back-scattered energy in the vicinity of the perpendicular reflection point add up substantially in phase, and a strong back-reflected or back-scattered signal results. Then, as the column expands and its size becomes comparable to a wavelength, the contribution from the outer-most edge begins to cancel that from the inner edge, owing to the path-length difference. For a given rate of diffusion, this will happen relatively quickly during the life of a back-scattering ionization column. However when the reflection or scattering is oblique, a longer time must elapse before the meteor trail can expand to the point where an equivalent path-length difference exists.

For both forward and back-scattering, the amplitude of the radio signal propagated via the column will decay with time in an exponential manner. An example of an echo showing this behavior will be found in Fig. 1 of reference 3.

ECHO DURATION AND AMPLITUDE

Evaluation of the ratio of the time of echo duration for forward scattering τ_F, to the time of echo duration for back-scattering τ_B, may proceed with the aid of the geometry illustrated in Fig. 1. Both of these times are defined as the time interval between initial and 37% amplitude. Let 2ϕ be the forward-scattering angle.

For the geometry of either (A) or (B) of Fig. 1, it may be shown on the basis of the preceding assumptions that the echo duration to 37% of the initial amplitude τ_F is given by:

$$\tau_F = \frac{\sec^2\phi}{4k^2 D} \text{ seconds} \qquad (1)$$

where $k = \frac{2\pi}{\lambda}$

D = the diffusion coefficient, approximately 3 meter^2/sec in the E-region.

From equation (1), the ratio of durations for the forward and back-scattering cases is

$$\frac{\tau_F}{\tau_B} = \sec^2\phi \qquad (2)$$

In assessing the average duration increase produced by this effect for a particular forward-scattering path, account must be taken of the location of the individual meteors with respect to the end points, since this will sharply affect the value of ϕ and thus the echo duration. The situation is illustrated in Fig. 2 (A) and (B).

From the standpoint of the amplitude of the reflected signal, on the other hand, the geometry of Fig. 1 (B) proves to be the more favorable, owing to the way in which the Fresnel zone contributions add up in comparison with part (A) of the figure. Whereas in (A) the forward-scattered signal will have the same strength as a back-reflected signal traveling over a path of equivalent length, in (B) the amplitude of the forward-scattered signal exceeds that of an equivalent back-scattered signal by the factor $\sec\phi$.

A further factor affecting reflected-signal amplitude can be understood by reference to Fig. 2. As the separation distance d increases, reflections of the sort shown in part (B) become weaker owing to the very low angle of takeoff or arrival required for signals at one end of the circuit. Antenna systems in practice rapidly lose sensitivity at very low angles.

EXPERIMENT INVOLVING THREE PATHS OF DIFFERING LENGTHS

In an effort to obtain at least a qualitative experimental verification of these considerations, a test was undertaken whereby a simultaneous recording was made at Stanford of three 44-megacycle signals on closely adjacent frequencies, one originating from a location 15 kilometers away, another 475 kilometers away, and a third 1175 kilometers away. All three stations had powers of the order of 500 watts. The closest-in station employed a quarter-wave dipole antenna, one quarter-wave above ground, but both the more distant ones had Yagi beams giving 6-8 db gain, and located about 40 feet above ground. Absence of layer transmission of any sort was verified by scatter-sounding. Ground-wave transmission from the nearest station was suppressed by suitable choice of location and antenna orientation.

An effort was made to determine the total duration of the discernable meteor echoes for each of the three paths by examination of the records. The <u>total</u> duration of a given meteor

was selected as a criterion, because the non-linearity and limited dynamic range of the recorder made it impractical to determine the time interval between a given value and 37% of that value. Total duration can be determined with accuracy only for the shortest path, where the individual echoes are well separated. It could only be estimated for the longer paths because of the difficulty of distinguishing the time of cessation of a given echo in the presence of the continuously-fluctuating background.

Fig. 3 shows a sample of the records obtained. The "diffuse" echo shown on the 15-km recording is believed on the basis of considerable experience to be a non-perpendicular reflection from a meteor column grown so large (or so distorted by winds) as to have become amorphous. These echoes are relatively rare compared with the "simple" type echoes whose behavior is well predicted by the assumed model.

At a frequency of 44.2 megacycles, and assuming a diffusion coefficient of 3 meters^2 / second, the calculated forward-scattering times to 37% amplitude, τ_F, and the ratios of τ_F to τ_B for the two oblique paths, are summarized in the following table:

For Reflection at Midpoint (Fig. 2A)

	τ_F	τ_F / τ_B
475 km path	6.2 secs	6.7
1175 km path	21.4 secs	23.1

For Reflection over One Station (Fig. 2B)

	τ_F	τ_F / τ_B
475 km path	1.5 secs	1.6
1175 km path	1.6 secs	1.7

Data was taken for the three paths simultaneously for a total time of roughly three hours. The rate of detection of meteor reflections distinguishable as such, was 2.7 per minute for the 15-km path, 3.2 per minute for the 475-km path, and 1.0 per minute for the 1175-km path. The lower rate for the longer path is unquestionably due to the great difficulty of distinguishing individual echoes in the fluctuating background.

For those meteor reflections which could be distinguished as such, it was found that the average duration for the 475 km path was 3.1 times as great as that for the local path. The average duration for the 1175 km path was 4.5 times as great as for the local path.

An attempt was made to add up the total time during which distinguishable meteor echoes were present on the three records. Such echoes were present for 7.5% of the time over the local path; 27% of the time over the 475-km path, and 12% of the time for the 1175-km path. The last figure is again indicative of the difficulty of distinguishing individual echoes in the fluctuating background.

CONCLUSION

These experimental results, rough as they are, are interpreted as providing qualitative support for the theory. It is evident that durations are considerably increased over the oblique path, as predicted. If echoes are present for 7% of the time at vertical incidence, and if a duration increase of up to 23 times for favorably-situated columns is to be expected over the oblique path, it seems clear that the contribution of overlapping meteor echoes to the total forward-scattered signal must be assessed with care, since the possibility that these echoes themselves may account for the total observed signal is not unreasonable.

No attempt is made at this point to offer a final explanation for the signal actually observed over the 1200-km oblique path. It should be kept in mind that the scatter-sounding apparatus could only demonstrate the absence of strong layer transmission. Very weak and patchy sporadic-E ionization along the path between the stations could conceivably account for the observed signal, and might not be detectable on the back-scatter equipment.

Although observations similar in many details have been made at the higher frequencies,[3] it is not yet demonstrated that the same explanation would be applicable to the two cases.

ACKNOWLEDGEMENT

The authors wish to express their thanks to Messrs. Rodney O. Beaudette, Cameron G. Pierce and Alvin C. Haugen for their assistance in the tests involving Spokane, Washington; Pasadena, California; and Aberdeen, South Dakota, respectively.

REFERENCES

1. O. G. Villard, Jr., and A. M. Peterson, "Instantaneous Prediction of Radio Transmission Paths", QST, Vol. XXXVI, No. 3, pp. 11-20, March 1952.

2. O. G. Villard, Jr., A. M. Peterson and L. A. Manning, "A Method For Studying Sporadic-E Clouds at a Distance", Proc. I.R.E., Vol. 40, No. 3, pp. 992-994, August 1952.

3. L. A. Manning, O. G. Villard, Jr., and A. M. Peterson, "Radio Doppler Investigation of Meteoric Heights and Velocities", Journal of Applied Physics, Vol. 20, No. 5, pp. 475-479, May 1949.

4. L. A. Manning, O. G. Villard, Jr., and A. M. Peterson, "Meteoric Echo Study of Upper Atmosphere Winds", Proc. I.R.E., Vol. 38, No. 8, pp. 877-883, August 1950.

5. Von R. Eshleman, "The Mechanism of Radio Reflections from Meteoric Ionization", Technical Report No. 49, Electronics Research Laboratories, Stanford University, July 15, 1952.

6. D. K. Bailey et al, "A New Kind of Radio Propagation at Very High Frequencies Observable over Long Distances", Physical Review, Vol. 86, pp. 141-145, April 15, 1952.

7. D. W. R. McKinley and P. M. Millman, "A Phenomenological Theory of Radar Echoes from Meteors", Proc. I. R. E., Vol. 37, pp. 364-375, April 1949.

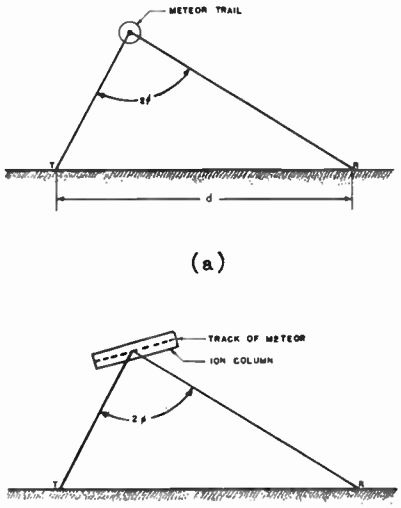


Fig. 1
Reflection geometry.
(a) trail in a plane perpendicular to line joining "T" and "R";
(b) trail in a plane containing the line joining "T" and "R".

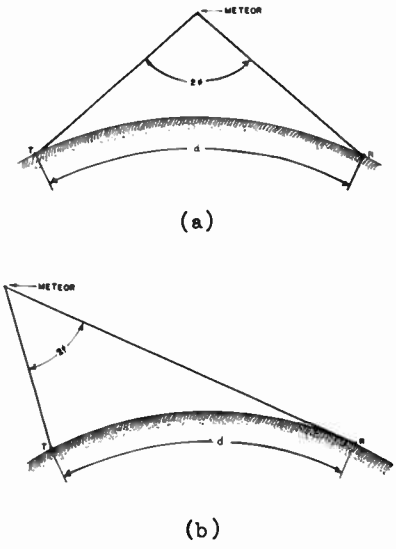


Fig. 2
Difference between mid-point and end-point reflection.
(a) mid-point; (b) end-point.

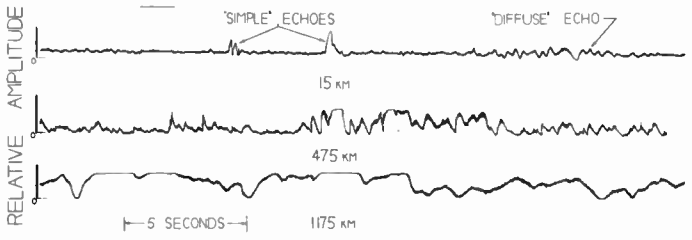


Fig. 3
Sample of 14-mc field strength recordings made simultaneously over paths of differing length, at a time when no layer-type reflection to any point on the earth's surface could be shown to be present.

AN INTERPRETATION OF VERTICAL INCIDENCE EQUIVALENT HEIGHT VERSUS TIME RECORDINGS ON 150 KC/S*

Rune Lindquist**

Ionosphere Research Laboratory
The Pennsylvania State College
State College, Pennsylvania

Abstract

Results of virtual height versus time vertical incidence pulse recordings, obtained on 150 Kc/s, are presented and discussed. Monthly median values of the reflection heights are shown in a series of graphs. The coupling echo, predicted by current wave theory, is definitely shown to exist. The results of measurements during undisturbed and disturbed days are discussed. It is concluded that one form of echo regularly noticed during magnetically disturbed nights must be due to one type of sporadic E. Recorded group and phase heights are compared and the differences checked against those predicted theoretically. Finally, the results are given of a preliminary investigation of the effects of solar flares.

Introduction

Low frequency vertical incidence pulse soundings of the ionosphere have been underway at this Laboratory since 1949. The part of the work dealt with in this paper is the recording and interpretation of virtual height versus time. The experimental work considered covers the period January 1950 - April 1952.

Due to various reasons, we have not been able to obtain a complete coverage of data for this period. As far as the normal behavior of the lower E-layer is concerned, this paper, therefore, will discuss results obtained during the following months: January 1950; February, March, August-December 1951; and January-March 1952. Most of the recordings obtained during the entire period can, however, be used for investigations of special phenomena; for example, the behavior of the E-region during disturbed conditions, solar flare correlations, sunrise phenomena, etc. The paper also gives a survey of the information hitherto obtained on these questions.

Results of low frequency virtual height soundings using the pulse technique have been reported earlier by R. A. Helliwell[1,2] and J. C. Blair, J. N. Brown and J. M. Watts[3,4,5]. Some of their results will be compared with those obtained in this paper.

General Discussion Concerning the Records and Their Interpretation

For this recording program we have used the conventional method of transmitting a short pulse of radio frequency. The direct or ground pulse and the pulses reflected from the ionosphere are picked up at the receiving site, detected, applied as intensity modulation on a cathode ray tube, and photographed on moving film in a conventional manner.

This system has certain limitations when applied to a low frequency carrier frequency; namely, the problem of obtaining a sufficiently steep leading edge and short duration of the transmitted pulse. By using some compromises a satisfactory solution has been found, the resulting transmitted pulse, approximately Gaussian, having a rise-time corresponding to between five and ten kilometers virtual height and a total duration of about 350 μsec.

Our main interest in the present work is to study the fine structure of the lower region of the ionosphere; that is, we should be able to measure the recorded virtual heights of reflection within a few kilometers. When recording a transmitted pulse of the above-mentioned shape, we may choose between one of two methods. One method is to preserve the shape of the ground pulse and the echoes throughout the receiver. The virtual heights can then be found very accurately by scaling, for example, from the peak of the ground pulse to

*The research reported in this paper has been supported by the Geophysics Research Division of the Air Force Cambridge Research Center under Contract AF19(122)-44 and by the Department of State under Contract SCC-2680.

**On leave from the Chalmers University of Technology, Gothenburg, Sweden.

the corresponding peak of the echo. This method has been discussed in a recent paper by H. A. Whale[6]. The other method of accomplishing the same degree of precision is to utilize a limiting amplifier in the video stages of the receiver. The records can then be scaled from the leading edge of the ground pulse to the leading edge of the echo. For the sake of simplicity we have chosen to use the latter system. The few disadvantages, for example the tendency of giving too high readings for very weak echoes, can easily be overcome by careful scaling.

It is also necessary to include a differentiating circuit in the video amplifier. This makes it possible to, within certain limits, record the closely spaced split echoes which appear quite frequently on the nighttime records, as discussed later in this paper.

A series of sample records illustrating most of the phenomena observed appear in Figures 1-3. These are direct recordings obtained by the movement of a suitable photographic paper past an image of the recording oscilloscope trace. Time increases from left to right for each record and is in local mean solar hours as indicated under each interruption of the record which is used as a time marker. Equivalent height increases vertically, the bottom solid trace being the direct or ground pulse while solid traces above this are ionospheric echoes. The thin, uninterrupted, horizontal lines are height marker traces. These are spaced 10 kilometers virtual height apart. In scaling these records it is necessary to add 4 Km to account for the transmitter-receiver separation of this amount. The interpretation of these records is discussed in a following section.

As far as routine scaling is concerned, the same principles as those internationally accepted for scaling equivalent height versus frequency records have been used. These are summarized in, for example,[7].

In order to obtain average monthly values for the heights of the E-layer reflections, we have used the system of computing median values. We have considered the part of the E-layer we are interested in, namely, the bottom, as regular in its diurnal behavior as the higher parts of the layer. However, as will be discussed later in this paper, the nighttime E-layer may be much more irregular in its behavior than the daytime layer, especially during magnetically disturbed conditions.

The Diurnal Variations of the E-Layer Reflections Under Undisturbed Conditions

Before discussing the experimental results regarding the diurnal variations of the E-layer reflections at 150 Kc/s, let us briefly summarize certain theoretical work which has appeared in earlier publications from this Laboratory[8,9]. These consider the wave equations for the two magneto-ionic components, π_1 and π_2,

$$\pi_1'' + (k_o^2\varepsilon_1 + M^2)\pi_1 = -\pi_2 M' - 2\pi_2' M$$
$$\pi_2'' + (k_o^2\varepsilon_2 + M^2)\pi_2 = -\pi_1 M' - 2\pi_1' M \qquad (1)$$

where $M = \frac{u'}{1+u^2}$ is the coupling factor and u is the polarization of the wave as discussed in detail in the references. A set of certain models of the electron distribution and the collisional frequency in the height range of interest, in the E-layer, are assumed. It is shown that, for the frequency of interest, these should result in two returned echoes, under suitable conditions, originating in the vicinity of the 300 and 3000 electrons/cm^3 points in the E-layer. The signal returned from the 300 el/cm^3 region is shown to be mainly due to the coupling factor terms in Eq. (1). We will call this region the coupling region. The high collisional frequency at this level prevents the existence of reflection conditions during daytime hours. During the night, the height of the 300 electrons/cm^3 level increases, taking this region into a lower collision level. As shown in the references, this increases the coupling effect with consequent reflection from this level.

The other electron density of interest is the region where N = 3000 electrons/cm^3. Here the conditions for reflection of one of the magneto-ionic components is fulfilled. This region will, therefore, be referred to as the main reflection region. We will, in general, refer to the signals from these two levels as coupling echo and main reflection echo.

It should be emphasized that this theoretical work, since it is based on the use of continuous waves, does not account for the dispersive properties of the medium, an extension of a portion of this work to include the dispersion of the medium appears elsewhere[10]. This will be referred to later in this paper.

Returning now to Figures 1-3, an inspection of the group height records shows that the occurrence of split echoes is very common during certain hours of the day. See, for example, Figure 2B. From the beginning of the record up to about 2330 LMST, we can only distinguish one single first and second hop echo. From 2345 - 0215 we can very definitely observe two first hop echoes, the center traces. The separation between the echoes and their respective amplitudes, estimated from the width of the traces, are varying somewhat with time. At 0215, the upper of the split hop echoes decreases in height by about ten kilometers and the lower echo disappears. This condition lasts up to about 0340, when the echoes once more separate and remain separated until sunrise is reached at about 0640. This latter part of the record, an hour or so before sunrise up till sunrise, is very typical for most nights. The two types of echoes just described are believed to be what we earlier referred to as main reflection and coupling echo; the latter being, of course, at the lower height under conditions of splitting.

In Figure 1A is depicted another type of nighttime record. This record characterizes a, magnetically, very quiet night. We observe, mainly, one single, continuous first hop echo throughout the whole night. There is very little indication of height changes in the layer and split echoes are difficult to detect. However, there are some split echoes present during the later part of the night. The reason why these cannot be recorded as two separate echoes is, as mentioned earlier in this paper, the width of the ground pulse. When the main reflection and coupling levels are separated by less than about ten kilometers, it is very difficult to record them as two separate echoes. The coupling echo will only manifest itself as a downward extension of the main reflection echo. But as, in general, the coupling echo shows more variations in amplitude than the main reflection, this provides a means for detecting both components during the early morning hours.

Finally, in Figure 1B is illustrated a daytime record. This record shows a single reflection throughout the day with no sign whatsoever of a coupling echo. The interruption of the echo during the interval 1049 - 1155 LMST is, as explained in a later section of this paper, caused by a solar flare. Normally, at this time of the year, we would expect to observe an echo continuously throughout the whole day.

In the routine scaling of the records, we have been mainly interested in establishing a definite proof for the existence of the two above-mentioned reflections from the E-layer and also to find their diurnal and seasonal variations. The monthly median heights of reflection for both echoes have been computed and are shown in Figure 4. These graphs definitely establish the existence of the two reflecting regions during the early morning hours. The theory also predicts the greatest separation of the echoes, and the strongest amplitude of the coupling echo, at low critical frequencies and increased layer heights, a condition which is most likely to exist in the early morning hours.

A closer inspection of the diurnal heights of the coupling level reveals that, on some months, the median virtual heights of the coupling echoes are greater than the corresponding median heights for the main reflection echoes. This condition indicates that our method of determining median heights must, on occasion, be in error. As a method of tracking down this error, all magnetically disturbed days during these months were selected and excluded from the computation. The computed median values of the reflection heights using only undisturbed days are also plotted on Figure 4. However, as the Figure shows, we still have occasional cases where the median height of the coupling echo exceeds the height of the main reflection echo. The reason for this must, therefore, be sought in the number of the individual values used for computing a median value. In this work we have considered five or more values as satisfactory for computing a median value. This is definitely justified for the daytime echoes. But the abnormal behavior of some of the nighttime median plots indicates that the E-layer at that time is very irregular. It would, therefore, be preferable to consider a median value based on less than five values as unacceptable, as also has been done in this paper. But a value based on five-nine values should be considered as questionable, instead of certain as we have done. However, the error introduced by accepting less than ten values as giving a good median value is, in general, probably not too great.

In Table I below we have listed some typical heights for different quantities. It is very interesting to note, for example, that the median virtual height of the coupling region is

Table I. Various Heights as Deduced from the Monthly Median Plots. All Heights are given in Kilometers. Numbers in Brackets Are Considered Doubtful Because of Lack of Sufficient Data.

Height (Km.)	Jan.1950 Jan.1952	Feb.1951 Feb.1952	Mar.1951 Mar.1952	Aug. 1951	Sept. 1951	Oct. 1951	Nov. 1951	Dec. 1951
A. Smoothed Noon Value	90 90	92 91	(95) 92	(94)	(94)	93	92	88
B. Night Peak Main Reflection	104 108	111 110	107 110	107	109	106	109	106
C. Smoothed Night Coupling	93 94	96 93	- 96	-	95	93	96	95
D. Smoothed Night Main Reflection (Max. Value)	102 106	108 107	106 110	(106)	105	101	107	106
E. Maximum Value Noon ±1 Hour	93 93	96 95	- 96	-	94	97	97	91
B - A	14 18	19 19	(12) 18	(13)	(15)	13	17	18
C - A	3 4	4 2	- 4	-	(1)	0	4	7
D - A	12 16	16 17	(11) 18	(12)	(11)	8	15	18
E - A	3 3	4 4	- 4	-	0	4	5	3

almost constant throughout the year, ranging between 93 and 96 kilometers. This height will very closely correspond to the bottom of the nighttime E-layer. A comparison with the smoothed noon values of reflection heights gives about the same heights; the noon values being slightly lower, the difference varying from zero to seven kilometers.

We have considered two different height values for noon (A and E). The reason for this is quite obvious if, especially, the latter, and more reliable, data of Figure 4 are examined. Inspection of the various monthly plots shows that the main reflection first decreases in height from sunrise up to an hour or so before noon. Thereafter, centered near noon, we can see, on most of the monthly plots, a very definite tendency of increasing height. This increase reaches a maximum somewhat after noon, after which the reflection height starts to decrease. However, the decreasing part of the curve is not so pronounced; it merely merges into the expected afternoon increase of the reflection height. Since the possible errors involved in the determination of the reflection height, as discussed earlier in the general discussion, have positively been removed in the scaling procedure, we are forced to accept this increase in noon height as real. The most likely explanation is to attribute the increase to an extra retardation in the D-region. We have, therefore, listed in Table I both the maximum median height recorded near noon, E, and the smoothed median height at noon, A. The latter height was determined by plotting the best smooth curve through the daytime data, excluding the part near noon. This curve should, therefore, correspond very closely to the changes in the E-layer. We can see from Table I, E-a, that the extra group

retardation, possibly introduced by the D-layer, is of the order of three to five kilometers. It is indeed very unfortunate that we are unable to obtain values for the period April - July due to the lack of system gain. This information should be very valuable, especially for establishing seasonal trends.

The very abrupt decrease in reflection height near sunrise is characteristic for all months. This problem has been studied in detail by the author elsewhere[11]. No further discussion of this phenomenon should, therefore, be necessary at this time.

A comparison of the median curves obtained for the same months in different years shows a very close correspondence. The median values for January 1950 and January 1952, for example, are identical within a few kilometers. This, perhaps, can be taken as an indication that solar activity has less influence on the ionization near the bottom than at the maximum of the E-layer.

As mentioned earlier, it has been shown[10], that the effect of the dispersion in the medium for our pulse signals would be, under certain conditions, to increase the time delay of the main reflection echo. That is, increase the group height compared to the phase height. As indicated in the reference, tha maximum time delay is to be expected at low E-layer critical frequencies. Efforts have been made to verify this theory by comparing group and phase height records. The phase height records have been obtained by R. E. Jones of this Laboratory. His recording program is to be described elsewhere at a later date.

In Figure 5 is shown on of the representative diurnal phase height records together with the group heights for the same day, December 23, 1951. As the phase height recordings only give changes in phase height, and not the absolute phase height, we must match the two records at one point. This was done at the point during daytime where both records showed the lowest height. We, consequently, assume negligible difference in phase and group height at that time, a procedure which appears to be valid in view of reference[10]. It can be seen from this Figure how changes in group and phase height correspond throughout the day, the relative difference in the two heights being of the order of three to four kilometers immediately after sunset when the corresponding E-layer critical frequency probably is fairly low. This difference decreases at a fairly fast rate during the hours immediately following sunrise, the difference in height being of the order of one kilometer near noon. This behavior checks very well with the predicted values given in[10]. Also, the nighttime behavior gives differences of about the same order of magnitude as those predicted.

The only part of the phase-height records that has to be considered as doubtful in this example is the part from 0245 to 0715. During this time an increase in group-height corresponds to a decrease in phase-height. However, as the group-height record shows, coupling reflections were present immediately preceding this interval. The recorded phase-heights are, therefore, possibly contaminated by the coupling reflections. It is, consequently, impossible to estimate the maximum difference in phase and group height in this interval. It would appear to be of the order of eight kilometers.

The Reflections on 150 Kc/s During Magnetically Disturbed Conditions:

It has been found during this recording program of virtual reflection heights versus time on 150 Kc/s that there is a considerable difference between the behavior of the E-layer during a disturbed and an undisturbed night. This fact is in contrast to the behavior of the daytime layer for the two conditions. Here the only difference during a magnetic storm is an increase in daytime absorption with no significant change in the shape of the h'-t curves when compared with a preceding undisturbed day.

Records obtained during disturbed nights are characterized by a very irregular behavior of the reflections. Sudden jumps from one reflecting level to another appear and the virtual heights of reflection show comparatively rapid fluctuations. Traces showing the typical behavior of changing retardation are obtained quite frequently.

A selection of commonly observed records obtained during magnetically disturbed conditions appear in Figures 2 and 3. The most predominant effect to be observed during these conditions is the appearance of a new reflecting stratum somewhat above the normal E-layer. Figure 2A depicts a typical example of this new reflecting medium, the reflections lasting from 2044 to 2345 LMST. This Figure also clearly demonstrates the behavior of the normal E-reflections during the period of the disturbance.

These reflections seem to nearly disappear. We will return to this question somewhat later in the paper.

As a first attempt to obtain information on the sporadic reflections above the E-layer, their frequency of appearance was studied. The day was divided into three-hour intervals corresponding to the three-hour divisions of the Grennwich Day. The total duration in hours of the presence of sporadic reflections within every three-hour interval was compiled and tabulated and is herein titled the index of intensity of sporadic reflections. This is the basic information for comparing these echoes with certain other manifestations of magnetic activity. It may be mentioned in this connection that during certain of the months of interest, it has not been possible to obtain this information. The reason is that the height range of the recording system had been adjusted to a restricted height range, too low to record these reflections. These conditions have, especially, prevailed during November, December 1950 and May-August 1951.

In Figure 6 is shown a typical example of the three-hour range index of intensity of the sporadic reflections. The index is plotted for every three-hour period during March and April 1952. There is also plotted the three-hour range magnetic K-index as measured at Cheltenham, Maryland[12] for these months. Besides the vertical lines indicating the intensity of the sporadic reflections, there appear certain letter designations on the Figure. These are defined as follows:

A = Record unsatisfactory for some reason that particular night.

X = Record satisfactory but positively no sign of sporadic reflections during that particular night.

Diagrams similar to Figure 6 have been plotted for the period January 1950 - April 1952 and all show the same general tendency as Figure 6. This Figure clearly demonstrates the close correlation between the intensity of the sporadic reflections and the magnetic activity both in detailed and in "gross" behavior.

In order to have a closer check on the dependence of these reflections on the magnetic conditions every individual reflection was rechecked. The total duration in hours of every single continuous reflection was plotted versus the average K-index during the corresponding time. These plots were made on a monthly basis. The general tendency of the individual points on these plots seemed to be to group on or above a straight line with a slope of about 45° and going through the origin of the coordinate system. In Figure 7, all these individual monthly plots are replotted in the same graph. This Figure very clearly demonstrates the grouping on or above the above-mentioned straight line of the majority of the points. This result appears to indicate that, in order to produce a sporadic echo of a certain duration, the magnitude of the disturbance has to exceed a certain minimum value of the K-index, this minimum value being greater the longer the produced echo lasts.

Finally, the total duration in hours of the sporadic reflections during each month was computed. The result is plotted in Figure 8. The mean values for the months have been connected to give a picture of the seasonal variation of the sporadic echoes. The general shape of this curve corresponds, for example, to the well, known seasonal distribution of auroral activity, with peaks in March and September.

This plot has also been compared with the indices of auroral activity in North America as published by Gartlein et al[13]. However, his published data did not extend beyond June 1950. A satisfactory correlation was obtained for the period January-May 1950. This is also true when his average curve for the year and our average curve for that and later years are compared. It is especially interesting to note the third maximum in auroral activity in June, obtained by Gartlein for the years near the sunspot maximum. The seasonal distribution of the sporadic reflections shows the same maximum in June. However, this point is based on only one year's data, 1950. On the other hand, as this year is closest to the sunspot maximum of the three years covered by this investigation, it might well be significant.

Before discussing the possible origin of the sporadic reflections, a few comments will be made regarding the virtual heights of reflection. As can be seen from the sample records in Figures 2 and 3, these virtual heights usually show typical retardation characteristics. However, in most cases, the echo appears at a fairly great height, decreases in height until a minimum value has been reached, remains for a certain

period of time at that height and, finally, disappears with increasing height. We, therefore, have recorded both the hourly heights and the minimum height of the reflection. Of these two quantities, the minimum height is of special interest. The minimum heights recorded during each month in the period January 1951 - April 1952 are given in Table II below.

Table II. Minimum Virtual Heights in Kilometers of the Sporadic Reflections Recorded During the Period January 1950 - April 1952

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1951	124	124	124	124	143	--	--	133	123	123	137	131
1952	131	143	127	123								

From this Table we can see that the minimum heights appear to have a tendency to fall within the height range 123-131 kilometers. This is also clearly demonstrated in Figure 9, where all the individual points during this period have been plotted. There is no pronounced seasonal variation to be noted.

When discussing the possible height of reflection for these sporadic reflections it would be very valuable to have some knowledge of their height versus frequency characteristic. Some tests were, therefore, performed using a manual sweep frequency recorder. The usable frequency range of this recorder is 560 Kc/s - 3600 Kc/s. On the number of occasions we have had this manual recorder in operation, the existence of an echo trace up to about 900 Kc/s has been noted. These reflections have been noted at the same time as we have observed sporadic echoes on 150 Kc/s. The virtual heights of the high frequency reflections have fallen in the range 135 - 155 kilometers, with the heights in the individual cases corresponding closely to those noted on 150 Kc/s.

Up to this time it has only been possible on a few occasions to obtain polarization records of the sporadic echoes in view of the height range to which this equipment is normally set. In the cases recorded, mostly manual observations, the polarization has been elliptic (almost circular) with a left handed sense of rotation. Whether this holds for all sporadic reflections is, in view of the scarcity of data, impossible to say, but it is hoped that this question will be settled by a continued check of such records.

What now will all this information on the sporadic reflections tell us? The first conclusion will, of course, be that they manifest themselves only under disturbed conditions. In view of the fact that they appear above the regular E-layer we have a choice of two possibilities. Either their origin actually is at this height or they are located below the E-layer and the reflections are of oblique origin. Tests performed on antennas with different directive characteristics have indicated that the reflections arrive from the vertical. Another strong argument favoring the statement that the reflections arrive vertically is the agreement between measurements on 150 Kc/s and on 500-1000 Kc/s. We, therefore, assume that these reflections are at vertical incidence.

The reflections found in the same height range at higher frequencies would definitely be classified as a form of sporadic E. In view of the close correspondence in height and time between the echoes in these two frequency ranges on all the occasions simultaneous recordings have been made, it seems justified to assume that the sporadic reflections we see on 150 Kc/s also are to be classified as sporadic E. In this connection it should be mentioned that sporadic reflections of apparently the same type have been reported earlier by Brown and Watts [3,4,5], and Helliwell[1,2] on low frequencies. Another interesting paper also mentioning reflections of this type is that of Gerber and Werthmuller[14].

The last paper discusses reflections on several fixed frequencies in the frequency band 556-1167 Kc/s. The authors did not quote any height measurements, but their published records show heights of the same order of magnitude as those we have obtained. This is a very interesting further check on the theory that our echoes are from one form of sporadic E. Helliwell[2] has also pointed out correlations between certain types of low frequency echoes and sporadic E on higher frequencies.

As was shown in a preceding paragraph, the maximum nighttime E-layer ionization usually does not go below 3000 el/cm^3, one of the reflection points for the 150 Kc/s waves. This is the case on quiet nights. On the other hand, during disturbed nights, we have many reasons to believe this is no longer true. See, for example, Figure 3A. At 2300, the layer appears quite normal with no detectable sign of split echoes indicating we are approaching the critical frequency. However, only ten minutes later the virtual height starts increasing, and at 2315 we have a well-developed split. At about 2348 the upper echo, corresponding to N = 3000 el/cm^3, completely disappears. The last part of the trace shows the retardation cusp so typical for an echo just penetrating a layer. At about the same time we note a reflection from a stratum above the E-layer. During the first hour this echo shows a decreasing height, while the coupling echo shows increasing height and increasing absorption. At about 0100 the conditions gradually change, the sporadic echo increasing in height and the coupling echo decreasing in height and absorption. Finally, at 0215, the sporadic echo suddenly disappears and, at the same time, both echoes from the E-layer show up again.

Our interpretation of these phenomena is that the maximum electron density of the E-layer started to decrease slightly before 2300, passed through a value of 3000 el/cm^3 at 2348, continued decreasing and reached its minimum value at 0100. From that time it began increasing again, passing N_{max} = 3000 el/cm^3 at 0215 and then, probably, continued to increase somewhat more. This Figure is typical of many of the examples we have noted of sporadic E.

The other main group of sporadic E records is represented by Figure 3C. In this record it is possible to see reflections from both the 3000-el/cm^3 level and from the sporadic E for a fair amount of the time that the sporadic E was visible. It should be noted that, on this type of record, the group height of the sporadic E is greater than, for example, on Figure 3A. This would indicate that N_{max} is not below but probably very close to 3000 el/cm^3, causing an appreciable amount of retardation. It would also indicate that the E-layer is fairly thin, part of the energy leaking through and part reflected. Somewhat later on, Figure 3C at about 0225, we see a typical example of the waves completely penetrating the E-layer. The corresponding decrease of extra retardation in the E-layer is seen to amount to about 18 kilometers.

The above explanation should mean that we are only able to detect sporadic E on 150 Kc/s under the condition that the maximum electron density of the E-layer drops close to, or below, 3000 el/cm^3. This, on the other hand, means that it is not necessary that the presence of a reflecting stratum slightly above the E-layer be limited only to disturbed conditions. It could well be there all the time, but normally the E-layer would prevent us from detecting its existence. Measurements on other frequencies at latitudes appreciably South of the Auroral Zone have, in General, shown that sporadic E is not confined to disturbed conditions. On the other hand, as State College is situated but little South of the Auroral Zone, geomagnetic latitude 52°N, it could well be possible that this type of sporadic E is related to some of the forms so frequently noted in, say, Southern Canada. Distant Auroral displays have, for example, been visible here on several occasions when sporadic E has been recorded.

However, it still remains to explain why the normal electron density in the E-layer should show such a marked decrease during disturbed conditions. The generally accepted value of the effective recombination coefficient, α is, in the E-layer, $\alpha_{day} = 1.1 \times 10^{-8}$ [15] and $\alpha_{night} = 2 \times 10^{-9}$ cm^{3}sec^{-1} [16]. However, as Bates and Massey have pointed out[17], the nocturnal value is probably too low because of the neglect of sources of ionization. They suggest that the true night value is the same as the day value.

We have made an approximate determination of α_{night} during disturbed conditions. Records showing the typical

cusp indicating that N = 3000 el/cm^3 was reached and, immediately thereafter, showing sporadic E were selected for the investigation. Utilizing the point where N = 3000 el/cm^3 and published E-layer data[12], it was possible to determine the critical frequency of the E-layer at sunset. The effective recombination coefficient, assumed to be independent of height, could then be computed by solving the well-known equation

$$\frac{dN}{dt} = - \alpha N^2 \qquad (2)$$

where

$$\alpha = \alpha_e + \lambda \alpha_i \qquad (3)$$

α_e = true recombination coeff. between electrons and positive ions.

α_i = true recombination coeff. between negative and positive ions.

λ = ratio between negative ions and electrons.

N = the electron density.

It was possible to find 25 records where there was no doubt that the maximum electron density decreased below 3000 el/cm^3. The median value of α_{night} obtained from these determinations is $\alpha = 3.9 \times 10^{-8} cm^3 sec^{-1}$. The minimum value is 1.6×10^{-8} and the maximum value $7.4 \times 10^{-8} cm^3 sec^{-1}$. The order of magnitude of the effective recombination coefficient, therefore, seems to be at least a factor of ten greater during disturbed conditions than under undisturbed conditions.

It is well known that during a magnetic disturbance a redistribution usually takes place in the F2-layer; maximum electron density decreases and the thickness increases. A similar behavior of the E-layer could explain this apparent increase in the recombination coefficient. However, an inspection of the individual records shows this effect to be unlikely. We would expect an appreciable increase in the group retardation if a redistribution occurred. As this does not happen, we conclude that the increase in the effective recombination coefficient probably is real. The frequent records showing reflections from both the normal E and the sporadic E with a separation in virtual height of only some 15 kilometers also indicate that the shape of the layer is probably unaltered.

As long as we do not have a more complete picture on what is going on during a magnetic disturbance, it is difficult to draw a conclusion as to what actually causes this apparent change in the recombination coefficient. A possible explanation would be to assume, as already suggested by Appleton and Naismith[18], an impact of high energy corpuscles, able to cause the ionization observed as sporadic E. Due to a similar mechanism, we could expect reactions in the normal E-layer, possibly increasing the value of λ in Eq. (3). However, it is apparent that before anything can be concluded in this direction, we should have some knowledge of, for example, the night sky spectrum during these conditions.

Finally, it should be mentioned that it has not yet been possible to detect any sign of reflections from the F2-layer on 150 Kc/s. In the above-mentioned cases, when the maximum electron density of the E-layer decreased below 3000 el/cm^3, we would expect F2-reflections on the records. However, this has never been the case. Every time they seem to have been obscured by the sporadic E-layer developed above the normal E-layer.

The Effect of Solar Flares

It is a well known fact that some solar flares give rise to a phenomenon called SID, sudden ionospheric disturbances. The SID's manifest themselves by causing a complete disappearance of high frequency signals received via the ionosphere and also a complete disappearance of echoes received on an ionospheric high frequency recorder. The only possible explanation from this point of view would be the assumption that the radiation from some of the flares would be of such wave-length and such an intensity as to be able to build up a highly absorbing stratum of ionization below the E-layer. Prevailing suggestions place this ionization within the D-layer.

Recently published data from the British workers[19] claims a very good correlation between practically all solar flares and recordings of 16 Kc/s phase heights. The main effect is a change in phase height of the order of 30 to over 500 degrees, corresponding to a change in effective height of reflection from 0.9 to 15 kilometers.

These results, namely, that both 16 Kc/s waves and waves of a frequency higher than 1 Mc/s, are affected by solar flares immediately suggest that there should be some effect noted at 150 Kc/s. And, as the British results definitely indicate that the ionization appearing in connection with the flare is located below the E-layer, it is most likely that the effect would be one of increased absorption.

As an approach to determine if such an effect exists, the h't-records were checked against all reported high frequency SID's during the period February 1950 - March 1952. However, partly due to lack of records and partly due to the considerable daytime absorption of the signal during the summer months, when the SID's appear most frequently, this survey cannot give us more than an indication of the conditions. The numerical results are listed in Table III.

Table III. Result of Comparison of SID's and the 150 Kc/s Virtual Height Records During February 1950-March 1952

Total No. of Reported SID's	Total No. of SID's When 150 Kc/s Records Are Avail.	Number of Cases Showing Total Absorp. During SID	Number of Cases Showing Partial Absorp. During SID	Number of Cases When The Over-All Absorp. Was Too High	Number of Cases Where No Effect Was Detectable
198	82=100 o/o	19=23.0 o/o	23=28.0 o/o	37=45.1 o/o	3=3.7 o/o

From this Table it can be seen that the existing records only cover about 40 o/o, or 82 cases, only 54.9 o/o, or 45 cases, can be used for the investigation. In the 19 cases showing total absorption, the effect was very positive. These records showed continuous reflections before and after the SID. Within ±2 minutes of the time reported for the start of the SID, the 150 Kc/s reflections disappeared suddenly and completely. No change at all in group height could be detected. The time of the return of the echoes corresponded, in most cases, to within a few minutes of the reported end of the SID. On the other hand, the 23 cases reported as showing partial absorption during the SID, started with the same precision as the last group, but had considerable fluctuations in the times of return of the echoes. In some cases the echoes returned within ten minutes after the SID, as observed at short wavelengths, was over, but in several cases the no-echo condition could last as long as several hours. As long as the time of onset shows such a remarkable coincidence with the start of an SID, it seems most reasonable to assume that this long period of complete absorption was due to an SID. We do not need much increased electron density below the E-layer to account for complete absorption of the E-reflections at 150 Kc/s during daytime.

The remaining three cases, when no effect at all was detectable, are difficult to explain. A determination of the reflection heights shows a height of about 85 kilometers in two cases and 94 in the third. The time of day was near noon in all three cases, a time when the 150 Kc/s reflections would be most sensitive to an increase in absorption. It seems, therefore, most likely, in these three cases, that the increased ionization causing the SID occurred at a height above the 3000 electrons/cm^3 level in the E-layer.

A direct comparison between reported solar flares and the virtual height records has also been made. The results obtained during the period September 1951 - February 1952 are presented in Table IV.

It is interesting to note that the percentage of cases showing a total absorption of the 150 Kc/s reflections was 23.2 o/o both by comparison with solar flares and SID's. This remarkable agreement must be coincidental, but as far as the order of magnitude is concerned, it could be significant;

Table IV. Result of a Comparison of Solar Flares and the 150 Kc/s Virtual Height Records During September 1951 - February 1952

Total No. of Reported Flares	Total No. of Flares When 150 Kc/s Records Are Avail.	Number of Cases Showing Total Absorp. During Flare	Number of Cases Showing Partial Absorp. During Flare	Number of Cases When The Overall Absorp. Was Too High	Number of Cases Where No Effect Was Detectable
121	69-100 o/o	16-23.2 o/o	8-11.6 o/o	35-50.7 o/o	10-14.5 o/o

especially as a check of the individual cases showed that almost all of the reported solar flares causing total absorption on 150 Kc/s did not cause fade-outs at higher frequencies. The fact that solar flares can affect low frequency reflections but not high frequency reflections is also in accordance with the results of the British investigations on 16 Kc/s[19]. However, the number obtained for cases when no effect was noticeable at 150 Kc/s, 14.5 o/o, shows that the 150 Kc/s reflections are a much less sensitive indicator of solar flares than 16 Kc/s reflections.

Sample records showing total absorption during solar flares are depicted in Figures 1B and 1C. The fade-out depicted in Figure 1B was recorded on December 15, 1951. The corresponding solar flare is reported during the period 1049-1154 LMST and was of importance 1. No corresponding high frequency fade-out was reported. It can be seen from the record that the time of beginning of the fade-out on 150 Kc/s coincides, within one minute, with the onset of the flare. The recovery of the reflections is somewhat delayed; however, a weak reflection is detectable immediately after the end of the flare.

In Figure 1C another fade-out is depicted, this time recorded on September 17, 1951. Two different flares were reported near the time of the fade-out. One flare, of importance 1, was reported as active during 1539-1639 LMST with its maximum at 1552 LMST. The other flare was reported at 1549 LMST and of importance 2. A corresponding high frequency fade-out is reported at 1547-1604 LMST. A comparison with the 150 Kc/s record shows that the low frequency fade-out started at exactly the same time and lasted one minute longer than the high frequency fade-out. Figure 1C also indicates that the solar flare of intensity 2 is more likely to produce a fade-out than the one of intensity 1. These two flares were located very close together on the sun's disk. The recovery of the fade-out in Figure 1C is very typical of most of the fade-outs we have observed. The shape of the recovery curve suggests that the reflection is exposed to an extra group retardation which gradually decreases. However, as explained earlier in the paper, a very high degree of absorption would also produce this effect. As this is very likely to be the case, it is safer to discard this piece of information. It is believed that a careful investigation of phase height records obtained on 150 Kc/s could settle this question.

Finally, it must be stressed that, owing to the lack of sufficient mid-summer daytime data during the interval of interest, the numbers quoted for the correlations with SID's and solar flares should be regarded as preliminary.

Acknowledgments

The author wishes to express his sincere thanks and appreciation to R. J. Nertney, N. Davids, R. E. Jones, A. H. Waynick and other members of the Ionosphere Research Laboratory for their many helpful suggestions and valuable criticism. The author also wishes to thank A. J. Haley especially for his unfailing interest in supervising the recording equipment.

References

1. R. A. Helliwell, A. J. Mallinchrodt, F. W. Kruse, Jr. and B. A. Wanbsganss, "Pulse Studies of the Ionosphere at Low Frequencies", Electr. Res. Lab. Stanford Univ., Calif., March 15, 1950.

2. R. A. Helliwell, "Sporadic E Stratification and Correlation with Low-Frequency Soundings", Paper presented at the URSI Meeting in Washington, D.D., April 1952.

3. J. N. Brown and J. M. Watts, "Ionospheric Observations at 50 Kc", Jour. of Geophys. Res., Vol. 55, No. 2, June 1950.

4. J. M. Watts and J. N. Brown, "Effects of Ionosphere Disturbances on Low Frequency Propagation", Jour. of Geophys. Res., Vol. 56, No. 3, September 1951.

5. J. C. Blair, J. N. Brown and J. M. Watts, "A Sweep Frequency Ionospheric Recorder for the Low Frequencies", Paper presented at the URSI Meeting in Washington, D.C., April 1952.

6. H. A. Whale, "Fine Structure of the Ionospheric Region E", Jour. Atm. and Terr. Physics, Vol. I, No. 4, 1951.

7. "Ionospheric Data, Series F", issued by CRPL, Nat. Bur. of Standards, Washington, D.C., January and February 1952.

8. J. J. Gibbons and R. J. Nertney, "A Method for Obtaining the Wave Solutions of Ionospherically Reflected Long Waves, Including all Variables and Their Height Variation", Journ. Geoph. Res., Vol. 56, No. 3, September 1952.

9. J. J. Gibbons and R. J. Nertney, "Wave Solutions, Including Coupling of Ionospherically Reflected Long Radio Waves for a Particular E-Region Model", to be published in Journ. of Geophysical Research, September 1952.

10. N. Davids, "Dispersion Effects at 150 Kc/s for a Particular E-layer Model", Submitted for Publication.

11. R. Lindquist, "An Investigation of the Ionizing Effect in the E-layer Near Sunrise", Submitted for publication.

12. Published for example in "Ionospheric Data, Series F", issued by CRPL, Nat. Bur. of Standards, Washington, D.C.

13. C. W. Gartlein and R. K. Moore, "Southern Extent of Aurora Borealis in North America", Journ. of Geophys. Res., Vol. 56, No.1, March 1951.

14. W. Gerber and A. Werthmuller, "Nahschwundtatigkeit der Rundspruchsender", Technische Mitteilungen der Schveizerischen P.T.T., Vol. XXV, No. 1, 1947.

15. O. E. H. Rydbeck, "Chalmers Solar Eclipse Ionospheric Expedition 1945", Transactions Chalmers Univ. of Tech., Gothenburg, No. 53, 1946.

16. E. V. Appleton, "Regularities and Irregularities in the Ionosphere", Proc. Roy. Soc. A, 162, pp 451-479, October 1937.

17. D. R. Bates and H. S. W. Massey, "The Basic Reactions in the Upper Atmosphere. I.", Proc. Roy. Soc., Vol. 187A, pp 261-296, 1946.

18. E. V. Appleton and R. Naismith, "Normal and Abnormal Region-E Ionization", Proc. Phys. Soc., Vol. 52, pp 402-415, 1940.

19. R. N. Bracewell and T. W. Straker, "The Study of Solar Flares by Means of Very Long Radio Waves", Monthly Not. of the Roy. Astr Soc., Vol. 109, No. 1, 1949.

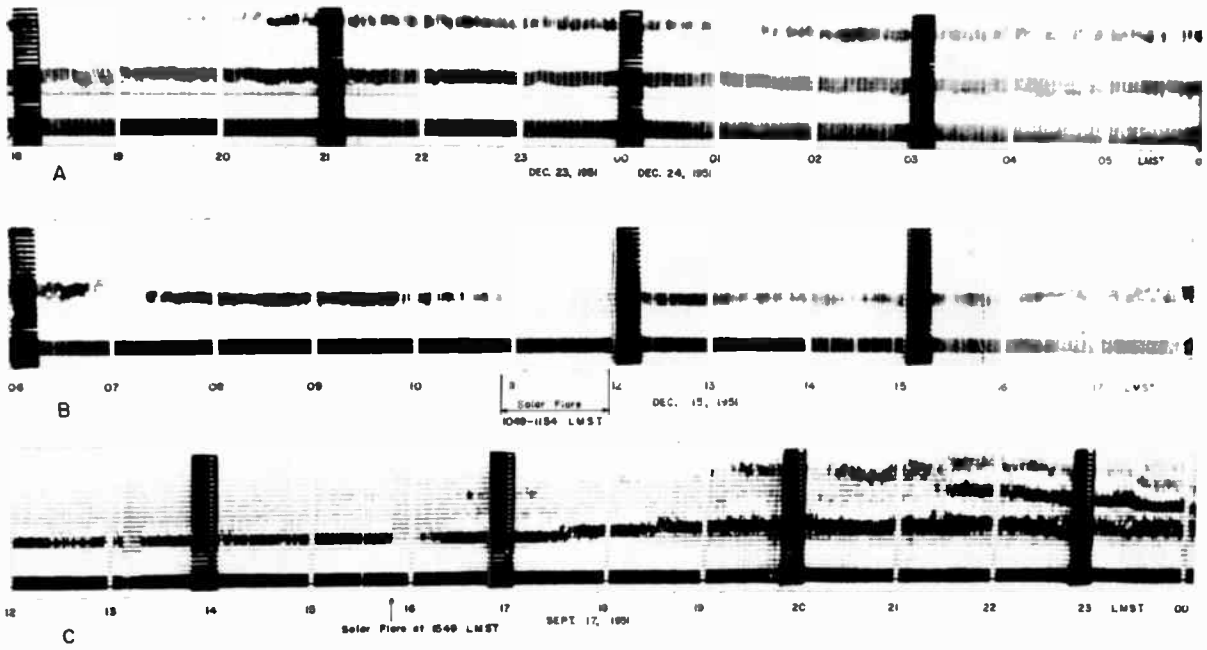


Fig. 1
150 kc vertical incidence equivalent height versus time records.

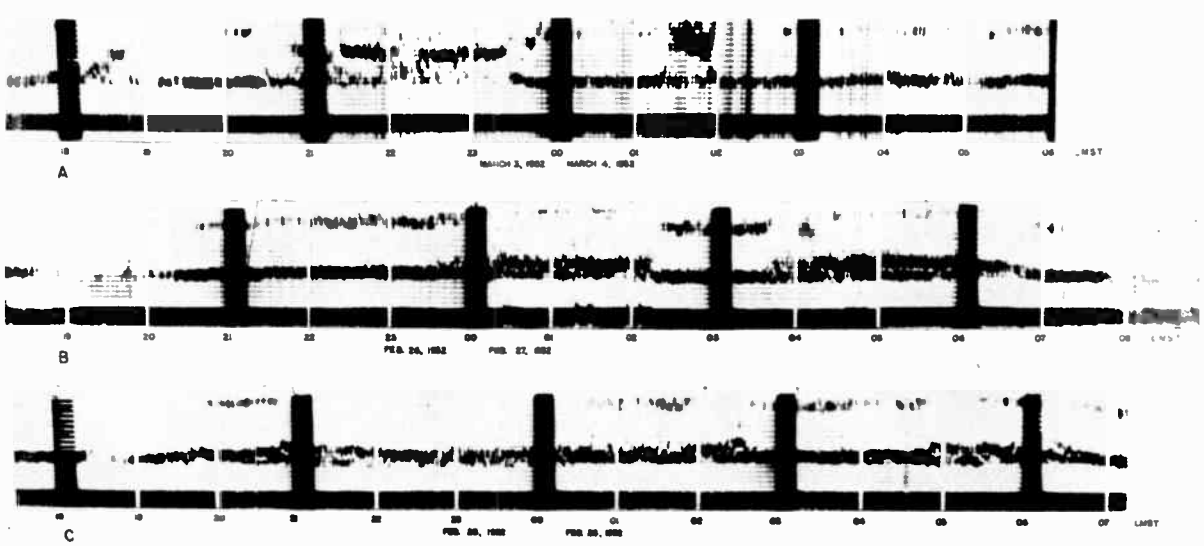


Fig. 2
150 kc vertical incidence equivalent height versus time records.

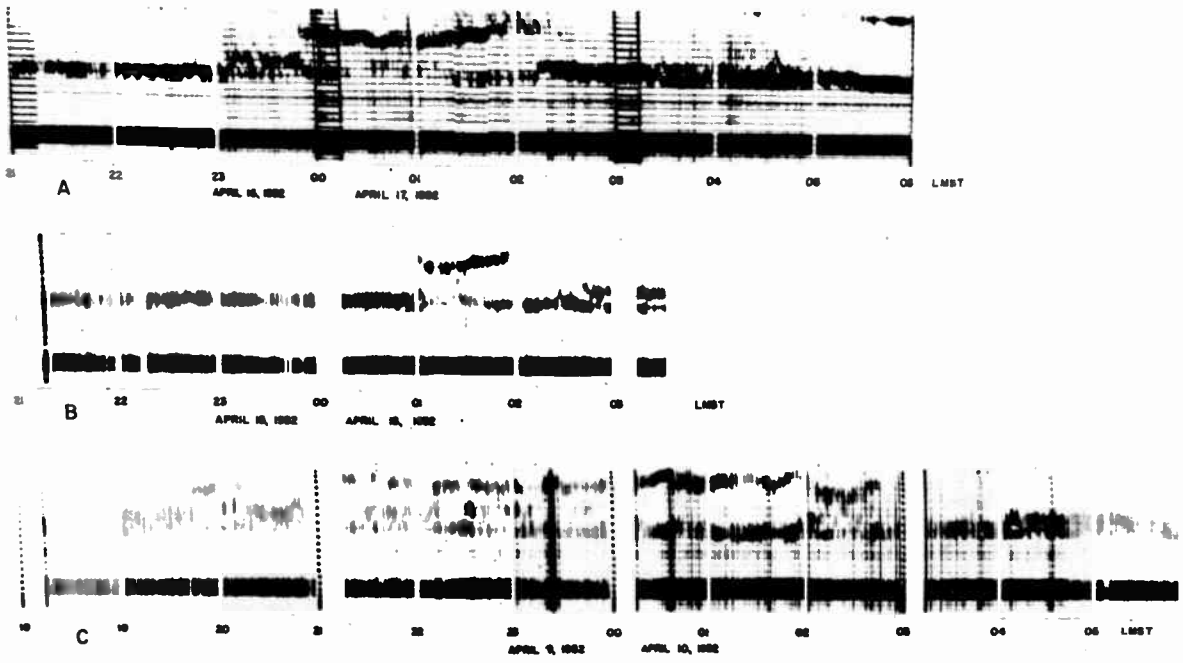


Fig. 3
150 kc vertical incidence equivalent height versus time records.

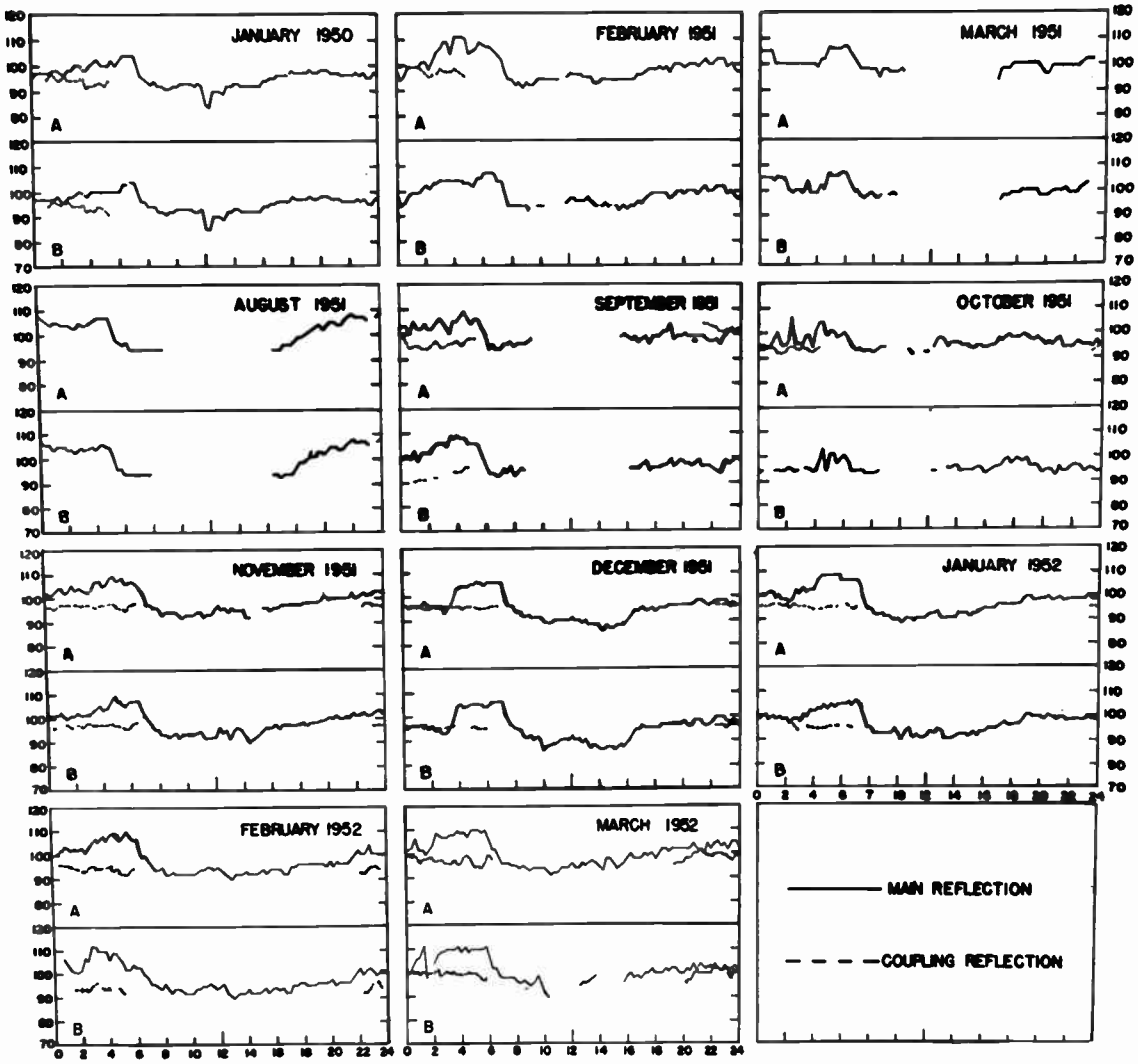


Fig. 4

Monthly median 150 kc vertical incidence equivalent heights versus diurnal time.

A. Monthly median values for all available data.

B. Monthly median values for days with no significant magnetic disturbance (half-day K-index of Cheltenham, Md., 4).

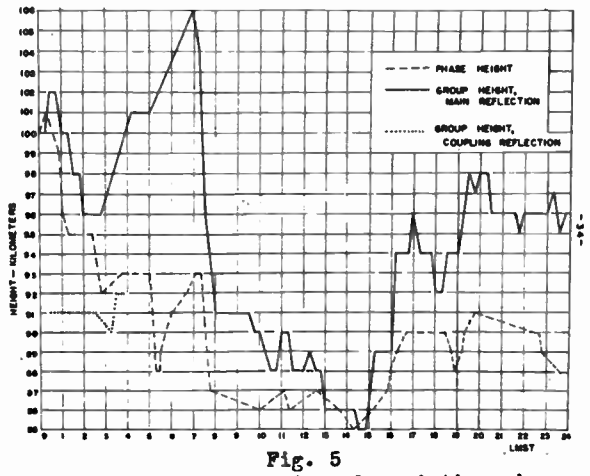


Fig. 5

Comparison between the diurnal variations in group and phase heights on 150 kc for a typical day.

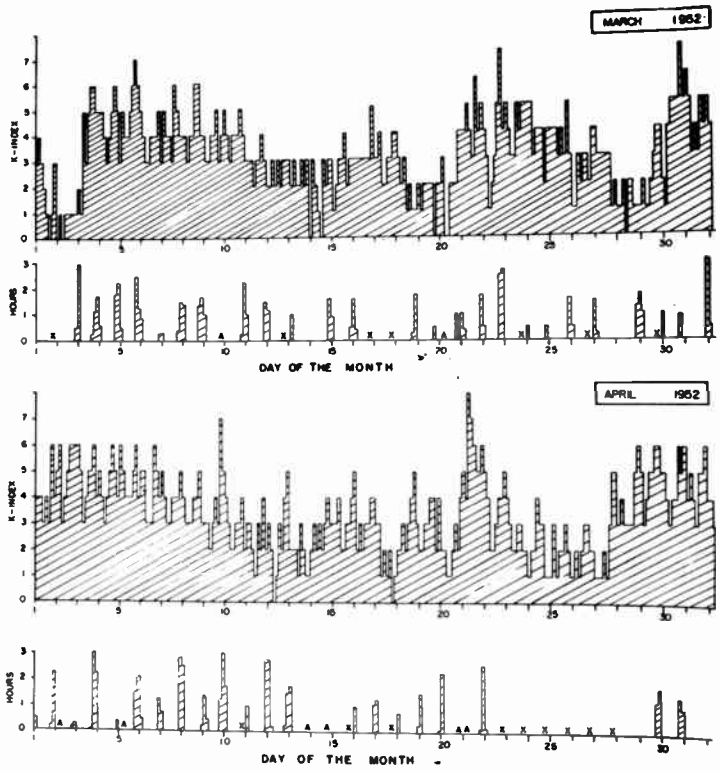


Fig. 6
Correlation between the magnetic K-index and index of intensity of sporadic reflections above the E-layer for two typical months.

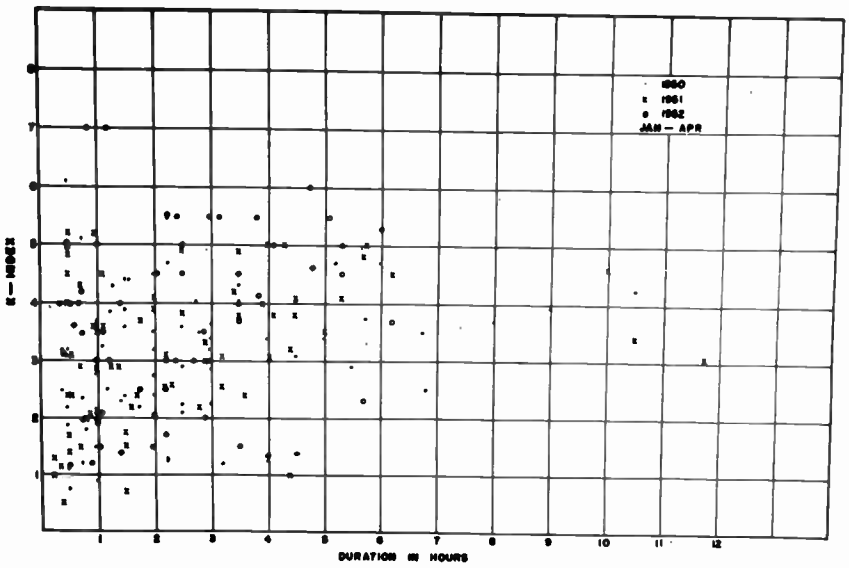


Fig. 7
Mass plot of all data correlating the magnetic K-index with the index of intensity of sporadic reflection.

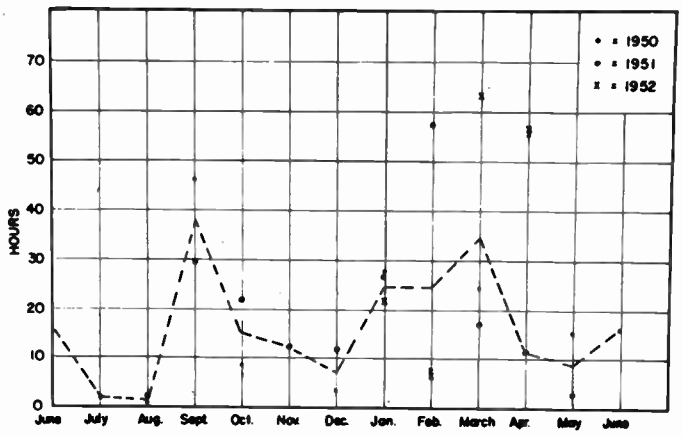


Fig. 8
Seasonal distribution of the appearance
of sporadic reflections on 150 kc.

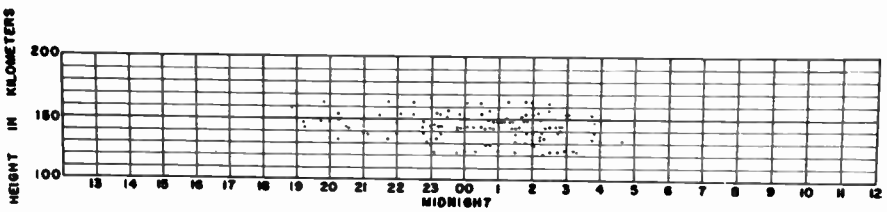


Fig. 9
Mass plot of all data showing the diurnal variation of the
minimum heights of the sporadic reflections on 150 kc.

THE EFFECTS OF SELECTIVITY, SENSITIVITY, AND LINEARITY IN RADIO CIRCUITS ON COMMUNICATIONS RELIABILITY AND COVERAGE

John G. Schermerhorn
Rome Air Development Center
Rome, N. Y.

1. Introduction

With the growth in use of the communication facilities of any frequency band, a major factor eventually arises that limits the number of assignable channels that can be usefully employed. This major factor is the interference between radio systems, which may become greatly aggravated in potential possibilities of interference by the need for close geographical disposition of multiple channel transmitters and receivers. This is particularly so if the systems involved are mobile and contain air-ground links. When the frequency assignment and disposition within a given geographical area are made specific, some of these possibilities become severe casualties[1]. By analysis and co-ordination of such specifications for a given application, considerable relief from interference may be obtained [2,3] at the expense of useful bandwidth.

Isolation from such interference, other than space attenuation, is inherent in the circuitry design and components used in the R-F sections of receivers and transmitters[4]. Since the highest level of interference at the greatest distance from a certain given space distribution of transmitters arises from distortion and intermodulation effects in the R-F section of a receiver, this paper will deal in particular with limits imposed on the communication reliability and coverage of a typical UHF mobile super-heterodyne receiver by the selectivity, sensitivity, and linearity of its R-F circuits. That such a study has not resulted in specific design criteria is to be regretted, but the statistical nature of the solution and its variability with equipment and application will become apparent in the following discussion, and may serve to illuminate some of the "gremlins" that are bound to be always present in this business of radio.

2. Nomenclature

Ps - Audio signal power in watts
Vs - Audio signal volts
Af - Audio frequency amplifier
Pa - Power amplifier
Pt - Transmitter power in watts
Vt - Transmitter R-F volts
Lt - Transmitter loss factor
Wt - Transmitter antenna watts
Gt - Transmitter antenna gain
Eo - Free space field strength volts/mtr
La - Transmission loss factor above free space
Er - Received field strength volts/mtr
Gr - Receiver antenna gain
Wr - Receiver antenna watts
Lr - Receiver loss factor
Pr - Received power watts
Vr - Received R-F volts
Rf - Radio frequency circuitry
If - Intermediate frequency circuitry
D - Distance in meters
Ae - Antenna effective area sq. meters
λ - Wave length meters
Zo - Free space impedance ohms
Ra - Antenna resistance ohms
A - Radian frequency
Δ - Frequency difference
Vu - Undesired R-F voltage
Ys - Threshold level of sensitivity
Sr - R-F selectivity factor
Si - I-F selectivity factor
Ga - R-F amplifier gain
Ca - R-F amplifier tube parameter
Cm - R-F mixer tube parameter
Br - BYRNE'S figure of merit, DB
DBI- Isolation to interference, DB

Volts are relative to one micro-volt noise level.

Watts are relative to one watt.

3. A Typical System

If a radio communication system accomplished its transmission through a medium with fixed characteristics, the

statement of the problem of reliable communication and the specification of design parameters for its solution would be considerably simplified. However, because of propagation variables, the signal that is received is a statistical quantity, the true mathematical representation does not present a closed solution, and an air of uncertainty, (or the presence of "gremlins are indicated, if the expression may be allowed) creeps into any engineering statement of the problem.

For the engineer to state the basic physics of such a complex system, he must simplify such variables to basic essentials by the device of statistical averages so as to better concentrate attention to the problems of equipment design and manufacture. The system diagram and associated power and voltage relation equations shown in Figure 1 represent such a simplification, and the numbers that an engineer would substitute in such equations would be those averages representative of the particular situation encountered where such is known. Many years of controversy and tests have centered around how to treat propagation and its attendant uncertainties in a mathematical closed form on a theoretical basis, and it is still a subject of intense investigation. For our purpose, however, the physical concepts of the situation and engineering "know how" is best served by such a simplified representation, and is more fully discussed in current literature[5,6,7,8].

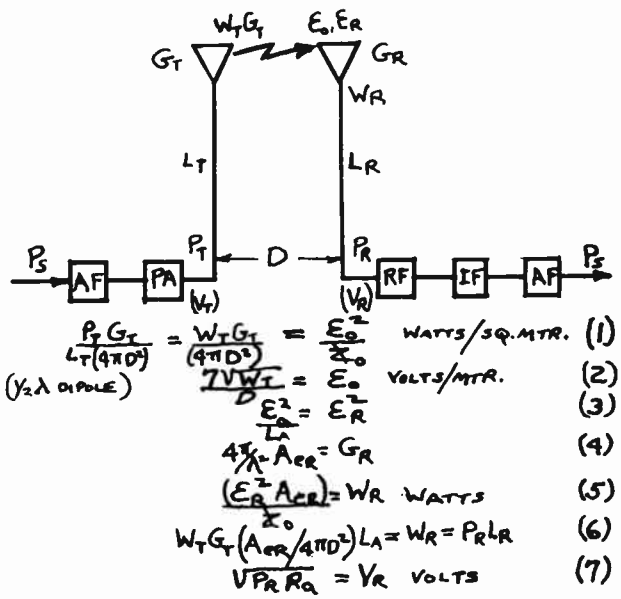


Fig. 1 - Typical system. Power-volt relations.

4. Propagation

In spite of the many factors that determine the radio wave signal intensity at the receiver, line of sight transmission, that is, short distances ground to ground and hundreds of miles air to ground and air to air, may be simply related for average values by the recognizable watt-voltage relation of equation (1). The departure from the usual engineering E^2/R term is because our transmission line is a sphere in space, and the amount of energy intercepted at the receiver in a relatively straight line path from the transmitter is a matter of space geometry. The voltage is applied to this line through the antenna which may have a gain in the direction of transmission due to the space distribution of power excited by the antenna. Therefore, the resultant volts/meter field intensity depends on the size of this sphere and a constant known as the impedance of free space. With a half-wave dipole for a transmitter antenna, a simple relation to the transmitted power is shown in equation (2). Since other than line of sight paths may be treated by insertion of a path loss attenuation factor [5,6] the relation for a given path below line of sight may be expressed by using equation (3). The watts in the antenna at the receiving location depends fundamentally upon the effective area of interception of the receiving antenna, and this is related to the transmitted power and the power in the receiver RF stage after losses by equations (5) and (6). In the well-known manner, receiver RF volts may be determined from the received power and antenna resistance by equation (7).

Depending on the relative locations of receiver and transmitter, the true instantaneous values of receiver RF volts may vary from twice as much to as little as one-tenth the average values of the equations shown but do not always represent total interruption of communication, and may be treated on a statistical basis where reliable field data is obtainable [8]. Thus, this may still provide a realistic basis for the estimation of interference and maximum field ranges in a given case where experience or spot field checks indicate that the variables of terrain and location represent a known condition. This view is also confirmed by the recent experiences of mobile units both ground and airborne, that the deep fades expected by theory are not experienced in practice in the UHF range and are attributed to a "fill-in" by multiple reflections from irregular terrain and structures. Antenna location and design,

with emphasis on gain antennas, aid in the reduction of variations of signal strength [7,8,9], and is mentioned here to point out that the variations due to space direction from transmitter to receiving antenna are involved in the definition of the effective area (Ae) or gain (G) in that direction. Since the direction to a mobile receiver changes with time, there is involved another probability that is associated with the service area in addition to instantaneous changes of signal strength experienced at a fixed location due to propagation variables. That is the probability of maintaining communication throughout an area within which there are located "dead spots" and "interference strips" that the mobile vehicle may encounter. This can be handled statistically after analysis with average values.

5. Signal Levels and Sensitivity

Thus with a fair amount of assurance that our average values realistically represent transmission conditions, we may construct a signal transmission chart such as Figure 2. Starting with audio signal level at the transmitter, the level of the signal is shown throughout its transmission and reception to the audio stage of the receiver. The receiver RF - IF selectivity, RF amplifier and mixer, and IF - AF amplification are shown separately to illustrate the effects of receiver selectivity, linearity and sensitivity in combatting interference produced by unwanted signals. The DB scale represents signal power relative to one watt, and the noise level indicated represents one micro-volt in a fifty ohm antenna. The signal threshold for reliable reception is assumed as 12 DB above one micro-volt. The three propagation conditions for free space, air to ground, and ground to ground show typical relative effects of distance and terminals. Such a chart thus provides an overall picture of the excursions of the average of the signal during transmission through a typical system.

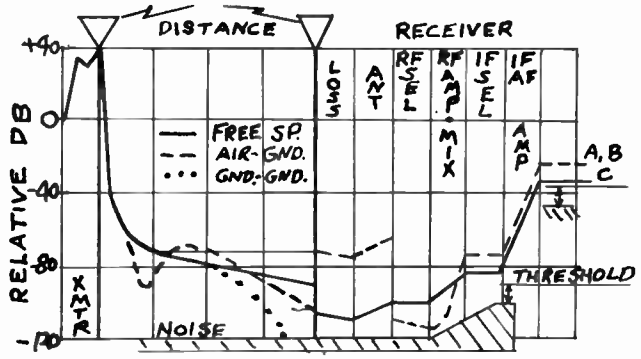


Fig. 2 - Signal transmission.

6. Vehicular Combinations

The number of possible transmitter and receiver combinations that could result in potential interference is illustrated in Figure 3. Hypothetical vehicles are indicated for generality since they may be ground or airborne. Two fixed sites and three mobile vehicles are grouped to show possible relative positions and the jagged arrows indicate which are transmitting. One mobile and one fixed receiver are indicated at (C) and (A'). About 16 total possible combinations result where two undesired transmitters are operating in addition to the desired transmitter with both alternatives fixed and mobile receivers. Of these, two cases with the receiver (C) airborne mobile and two cases with the receiver (A') fixed could be most severe and involve all three propagation conditions of Figure 2. These are diagrammed in Figure 4 where the crosses represent fixed locations and the elipses represent mobile vehicles. The circled quantity represents the receiver that is potentially in trouble due to interference from other transmitters than its own home station.

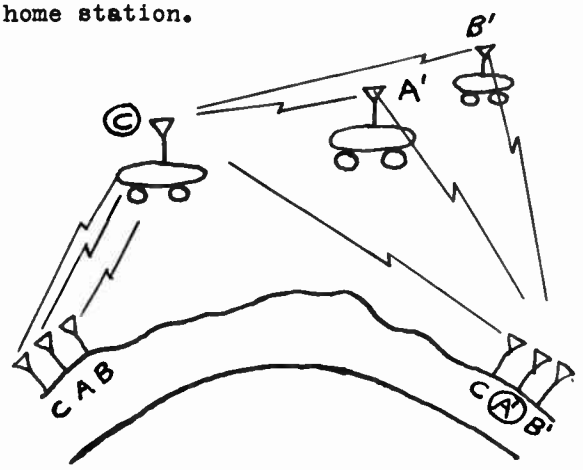


Fig. 3 - Vehicular combinations.

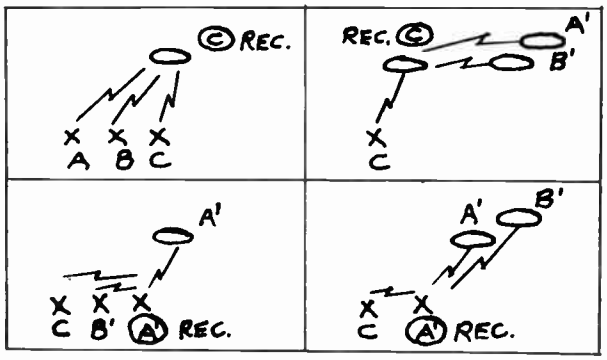


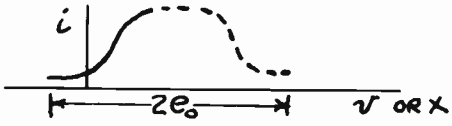
Fig. 4 - Interference combinations.

7. R-F Intermodulation

We have been discussing propagation and interference without specific reference to the physical basis for another frequency interfering with a desired communication channel frequency. The effect bears a name that, depending on the writer, ends in modulation with prefixes of inter - or cross - or often termed common frequency, co-channel, or similar waveform interference. Of the considerable literature pertaining to this subject [10, 11, 12, 14], some refer to the effects of particular modulation or demodulation processes, and a few [13, 15], specifically deal with the matter as the result of non-linear circuit action with generalized results.

Two such non-linear culprits exist in the R-F section of our typical system. They are the R-F amplifier and the R-F mixer. Surgical removal is not the answer since they must perform their given functions. Perfectly linear vacuum tubes are not highly probable of attainment in the immediate future, so that the R-F circuitry with its selectivity before amplification and after mixing, the degree of amplification for sensitivity requirements and amplifier gain controls are elements that need to be judiciously balanced to result in the lowest level intermodulation products at the I-F amplifier [4].

For clarity we will review the physics of the formation of these interfering products. With a tube characteristic as shown in Figure 5, it is possible to express the output current as the function of the input R-F voltage as indicated.



$$i = \alpha_0 + \alpha_1 V + \alpha_2 V^2 + \alpha_3 V^3 \cdots \quad (8)$$

$$\text{OR } i = \beta_0 + \beta_1 \cos\left(\frac{\pi x}{e_0} + \theta_1\right) + \beta_2 \cos\left(\frac{2\pi x}{e_0} + \theta_2\right) \cdots \quad (9)$$

WHERE

$$V = a\cos A + b\cos B + c\cos C \quad (10)$$

$$V_A \sim \frac{3}{4}\frac{\alpha_3}{\alpha_1} V_u^3 \qquad \text{RF AMP.} \quad (11)$$

$$V_M \sim \frac{3}{2}\frac{\alpha_4}{\alpha_2} V_u^3 \qquad \text{RF MIX.} \quad (12)$$

Fig. 5 - Vacuum tube nonlinearity.

The product combinations of the frequencies A, B, C are called intermodulation products, and the term second, third, etc., order refers to the powers of (V) in the non linear device. These products must not be confueed with harmonics. As may be seen on Figure 6, only the circled quantities represent harmonics, and are far less in number, and of a lower level than combinations formed by more than one frequency. The main interference to a message on (C) frequency would be those combinations of (A) and (B) frequencies that produce a (C) frequency component. The product spectrum of Figure 7 indicates a third order (2B-A) product that is a direct interference to (C) frequency. The magnitude of this effect may be computed [4,13,15] from the tube coefficients (α_n) as given by equations (11) and (12), but is determined with more certainty at the present time by the test procedure outlined in Figure 8 that gives directly the level of unwanted signal that will produce an interfering output from the receiver.

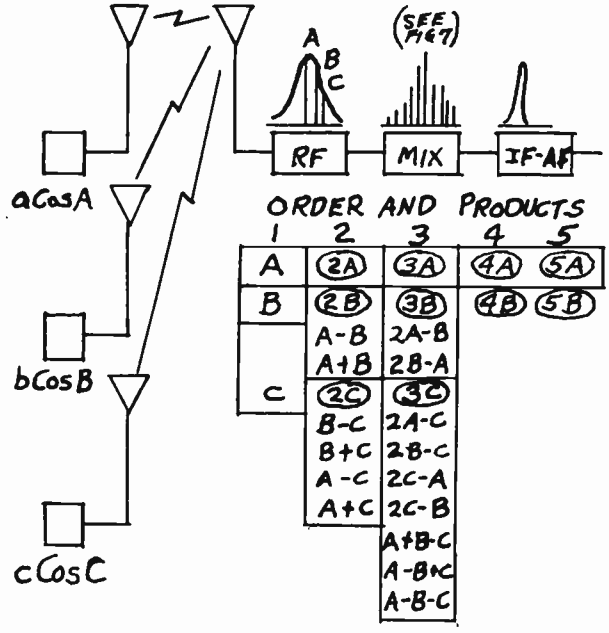


Fig. 6 - Intermodulation products.

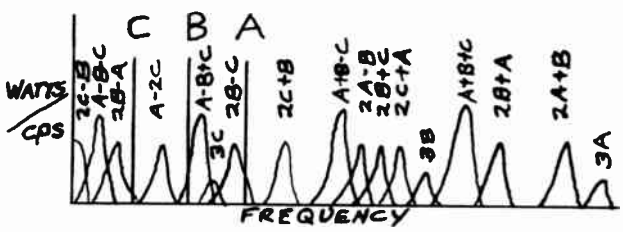
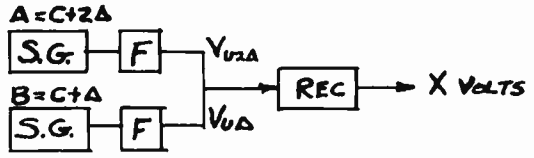


Fig. 7 - Intermodulation spectrum.

8. Selectivity And Interference Isolation

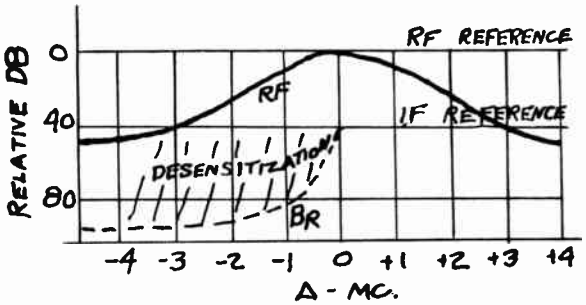
When one thinks of selectivity as applied to a radio receiver, it is quite natural to suppose that a major contribution is from the tuned R-F circuit. This may be quite true in single channel equipment, and by careful design, may materially reduce or even avoid the problem of adjacent channel interference. However, in broad-band multi-channel equipment, the number of tuned elements to accomplish this are excessive to be practical, and less RF amplification and more IF selectivity has to be employed [4]. Thus, a measure of overall selectivity is something that combines all these factors. Such a measure has been proposed by John F. Byrne [4], and has been referred to in the industry as "Byrne's figure of merit". It is measured in the fashion described in figure 8. It is the level of undesired signal above one micro-volt that will give a receiver response equivalent to a one micro-volt on-channel signal.



$2B - A = C$ $\quad V_S = 1\mu V.$ → C → REC → X VOLTS

BYRNE'S FIGURE OF MERIT $\quad B_R = V_{U_{MIN}} = \frac{2V_{U\Delta} + V_{U2\Delta}}{3}$ (13)

Fig. 8 - Intermodulation tests.



$$B_R \sim \frac{C_M}{S_I G_A}\left[\frac{C_A V_{U_{MIN}} G_A}{S_R}\right]^3 \quad (14)$$

$$DBI_{MAX.} = B_R + Y_S \quad (15)$$

$$\text{RF NON-LINEARITY GAIN} \sim \frac{C_M}{G_A}\left[\frac{C_A V_U G_A}{S_R}\right]^3 \quad (16)$$

$$DBI \sim B_R + Y_S - 20\,Log_{10}\left[\frac{C_M}{G_A}\left(\frac{C_A G_A}{S_R}\right)^3 V_U^2 V_{U2\Delta}\right] \quad (17)$$

Fig. 9 - Selectivity.

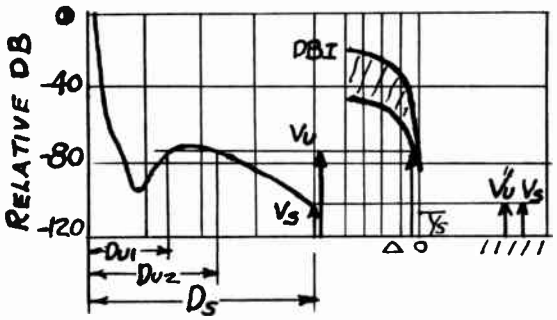


Fig. 10 - Interference distance.

In Figure 9 is plotted a typical RF selectivity curve, and the results of a measurement of typical (Br) values. An area marked 'desensitization' indicates the effect of receiver overloading that may occur where AGC action is not employed to limit strong off-channel signals. The net results are to disable a certain random number of channels that increase with the strength of the off-channel signals. This represents a condition on the ability of (Br) values to represent intermodulation protection of the receiver, but this uncertainty can be met for most signal levels encountered by inserting a factor in the probability of channel use as explained later on.

If the sensitivity (Ys) is included, the basic receiver protection is given by equation (15) below Figure 9. This is in DB's, and is termed DBI, meaning decibels isolation. The non-linearity of the RF amplifier and mixer is proportional to equation (16), and demonstrates how the RF gain nullifies the RF selectivity[4]. This effect is represented on the signal level chart of Figure 2, where a non-linear gain is shown. Equation (17) provides an approximate expression for the interference isolation as a function of the undesired signal level (Vm).

Fig. 11 - Interference areas.

9. Signal And Interference Areas

The existence of an interference distance is best shown by the graphical construction of Figure 10. If the (Br) curve of Figure 9 is inverted with its peak placed at the sensitivity level (Ys) adjacent to the signal level-distance curve of Figure 2, then a horizontal reference gives the solution at what distance from the transmitter interference may be expected at the receiver. When these distances are plotted as radii from some typical combination of transmitters such as diagrammed in Figure 4, interference and signal areas result as in Figure 11. Here an airborne mobile vehicle (C) is out of contact with his home transmitter because of (A) and (B) interference. The airborne transmitters (A') and (B') are another potential source of interference that may engulf him when he moves out of the interference "strip" that, at present, troubles him. For a while, signal transmission will be clear depending on the direction of flight taken. This presentation furnishes a basis for estimating probabilities of interference and illustrates the variables involved.

10. Probabilities and Multi-channels

From the picture evolved, it should be possible to outline the field tests required to determine overall communications reliability for given areas and locations of fixed and mobile transmitters and receivers. This is, in essence, a statistical evaluation for each given operational situation, and would involve an expression of the results in probabilities of X% operational reliability.

Besides the percentage time useable signal and desensitization disabling probability previously referred to, and the interference area conditions just mentioned, the multi-channel operation imposes an additional probability that concerns channel use time[17], and the pyramiding effect of the increased third order intermodulation products where perhaps 24 channels are in operation rather than 12. Figure 1 of reference 3 indicates that the probability of interference increases from 0.04 to 0.40 when the increase from 12 to 24 is a random selection of 100 channels (50 mc). The use of co-ordinated area location and frequency allocation [2,3], techniques [16] can reduce this probability to zero at the expense of band width. Figure 3 of reference 3 gives 12 interference free channels out of 100 (50mc) as an example of this expense.

11. Recommendations

It is hoped that this discussion will aid in the focusing effort on communication systems analyses of the type outlined with the objectives of engendering design data and awakening engineering ingenuity to further increase communications reliability in civil and military applications.

It is particularly recommended that attention be given to R-F circuitry to improve rejection of interference produced by neighboring facilities.

12. Acknowledgements

The author wishes to express his sincere thanks to Mr. A. A. Kunze and others at RADC, and those members of industry that have aided in the formulation of the ideas presented in this paper.

References

1. W. Rae Young, Jr.,"Interference beween VHF radio communication circuits," Proc. I.R.E., Vol. 36, July 1948.

2. K. Bullington, "Frequency economy in mobile radio bands", Bell Sys. Tech. Journal, Vol.XXXII, Jan. 1953.

3. W.C.Babcock, "Intermodulation interference in radio systems", Bell Sys. Tech. Journal, Vol. XXXII, Jan. 1953.

4. J.F.Byrne, "The selectivity and intermodulation problem in UHF communications equipment", a paper presented to the National Conference on Airborne Electronics, Dayton, Ohio, May 12, 1952.

5. K. Bullington, "Radio propagation at frequencies above 30 megacycles", Pro. I.R.E., Vol. 35, Oct. 1947.

6. K. Bullington, "Radio propagation variations at VHF and UHF", Bell Sys. Tech. Mono. #1802, Mobile Radio Propagation, Feb. 1951.

7. K. Bullington, "Radio transmission beyond the horizon in the 40 to 4,000 mc band", Proc. I.R.E., Vol. 41, Jan. 1953.

8. W.Rae Young, Jr., "Comparison of mobile radio transmission at 150, 450, 900, and 3700 mc", Bell Sys. Tech. Jour., Vol. XXXI, Nov. 1952.

9. R.S.Kirby, J.W.Herbstreit, K.A.Norton, "Service range for air-to-ground and air-to-air communications at frequencies above 50 mc", Proc. I.R.E. Vol. 40 May 1952

10. W.R.Bennett, "Cross modulation requirements on multichannel amplifiers below overload", Bell Sys. Tech. Jour., Vol. XIX, Oct. 1940.

11. C.B.Aiken "The detection of two modulated waves which differ slightly in carrier frequency", Proc. I.R.E. Vol. 19, Jan. 1931.

12. J.C.Schelleng, "Some problems in short-wave telephone transmission", Proc. I.R.E., Vol. 18, June 1930.

13. W.L.Barrow, "A contribution to the theory of non-linear circuits", M.I.T. Elec. Eng. Dept. Publication #97, Oct. 1934.

14. C.B.Aiken, "Theory of the detection of two modulated waves by a linear rectifier", Proc. I.R.E. Vol. 21, April 1933.

15. W.J.Warren, W.R.Howlett, "An analysis of the intermodulation method of distortion measurement", Proc. I.R.E., Vol. 36, April 1, 1948.

16. A.J.Beauchamp, "A technique of intermodulation interference determination", paper #22.5 delivered at the I.R.E. National Convention, March 25, 1953.

17. B.D.Holbrook, J.T.Dixon, "Load-rating theory for multi-channel amplifiers", Bell Sys. Tech. Jour. Vol. XVIII, Oct. 1939.

SINGLE SIDEBAND FOR MOBILE COMMUNICATION

Adamant Brown and Robert H. Levine
Signal Corps Engineering Laboratories
Fort Monmouth, N. J.

Introduction

There is a dual purpose in presenting a paper on single sideband as applied to mobile communications. First, to acquaint other engineers with the progress made by the Signal Corps in the single sideband field with a brief description of some of the systems embodying single sideband techniques under consideration by the Signal Corps Engineering Laboratories. Second, to stimulate interest and provide a challenge to others to look into the problems involved and perhaps evolve a superior system.

It is not intended that this paper be a rigorous discussion of some new found technique in the realm of single sideband communications. Neither is it directed to those among us who may be considered experts in the single sideband field. Rather, it is intended to point out the advantages of employing this method of communication and indicate the relative simplicity with which it may be accomplished.

Advantages of Single Sideband

It would be well, at this point, to review the advantages of single sideband over conventional double sideband communication. While these points have been recounted in numerous articles and publications, it is our experience that by and large only those directly engaged in this field ever read such articles. A few minutes will, therefore, be spent in refreshing our memories.

The following points are fundamental to any consideration of single sideband systems:

Single sideband offers the equivalency of higher transmitted power when compared to double sideband systems. Various ways may be used in expressing the equivalent power gain, dependent upon which transmitter parameter is considered to be fixed. The most commonly used figure is 9 db, which is derived as shown in Figure 1, by considering the peak power capability of the transmitter to be fixed, and then allowing 6 db for the elimination of the carrier and an additional 3 db for the reduction in receiver accepted noise due to halving of the accepted bandwidth.

There is also an improvement in transmission characteristics through the ionosphere which is not included in the 9 db figure. This improvement may be explained in several ways, such as elimination of harmonic distortion caused by severe carrier fading, and the elimination of the two sidebands phasing one another to produce frequency distortion. In other words, a single sideband system can be considered as much as ten times as powerful as a conventional double sideband system with the same transmitter power output.

Not to be overlooked in mobile applications are the reduction of antenna and antenna feeder voltages and the reduction of power source requirements which may be obtained with no loss of effectiveness of communications.

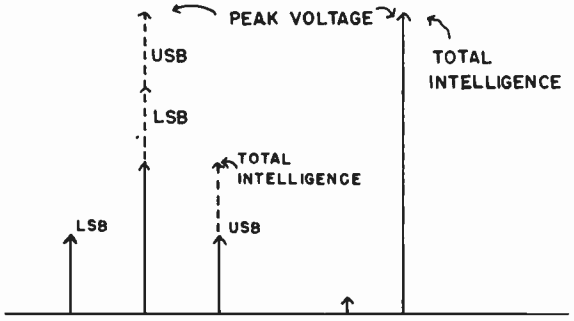


Fig. 1 - Improvement due to Single Sideband Transmission.

Spectrum economy is automatically obtained by the use of single sideband. Because of the fundamental limits of the HF spectrum, reductions in bandwidth are almost a necessity to avoid hopeless overcrowding. This is an immediate need, inasmuch as the HF band is very nearly approaching the hopeless classification today, as a cursory spin of a receiver dial through these frequencies will quickly prove.

Present Applications

The application of single sideband to mobile communications of the army would have seemed extremely dubious and even visionary only a few short years ago. The military requirements for mobile equipment of extreme ruggedness, dependability and simplicity of operation were not characteristic of single sideband radio equipment developed up to that time. With the possible exception of dependability, the reverse was true. Although this type of field use of single sideband seemed, at that time, unobtainable in the foreseeable future, single sideband had more than demonstrated its worth in fixed plant systems.

Single sideband has advanced from an initial commercial high frequency application to a position of undisputed superiority for providing reliable long range high frequency communications within a period of less than 20 years. Today,

single sideband circuits provide the backbone for the principal military global radio circuits, and are widely used for inter-continental commercial telephone service. Its natural characteristics of spectrum and power conservation, together with relative freedom from multi-path propagation distortion, insure that its future employment will be widespread.

Methods of Obtaining Single Sideband

There are two general methods by which a single sideband signal may be derived, namely, the filter method and the phase shift method. Each system has its own respective advantages and disadvantages, and a selection of the particular method of deriving the single sideband signal is greatly dependent upon the particular application. While either system may be used at some fixed low frequency to derive the single sideband signal, only the phase shift method is adapted to generate the single sideband signal at the final operating frequency in the HF range. Both systems use balanced modulators in order to suppress or greatly reduce the carrier output.

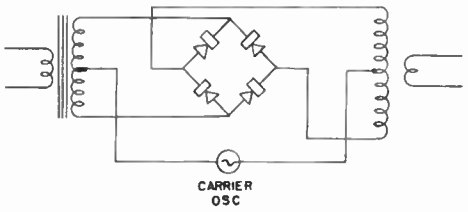


Fig. 2 - Simple Balanced Modulator.

Figure 2 shows a very simple type of balanced modulator, such as is used with the filter derived single sideband generator. The filter would be connected directly after the balanced modulator. It is noteworthy that only diodes are employed in this modulator. These can be semi-conductors rather than vacuum tubes. Hence, there is no power supply drain. Also, diodes can be selected for this application which will last almost as long as the life of the equipment. The filter derived single sideband provides greater stability with time, since modulator balance affects only the reduction of carrier, while the elements controlling the rejection of the unwanted sideband are purely passive and remain constant.

Illustrated in Figure 3 is one type of double balanced modulator such as is employed in obtaining phase derived single sideband. In order to operate this modulator over the wide frequency range desired in this system, it is necessary to provide for wide band RF phase shift control and to provide a means of equalizing the outputs of the various tubes as they change due to aging, basic differences in dynamic characteristics of the same tube types, or when replacing tubes.

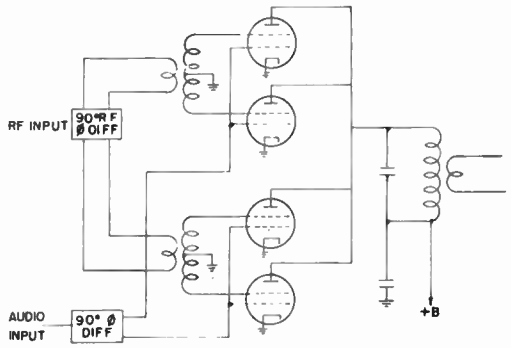


Fig. 3 - Double Balanced Modulator

In the double balanced modulator required for the phase derived system unbalance in either modulator will cause an increase in carrier output, as well as an increase in the level of the undesired sideband. Errors in the radio frequency phase shift adjustment between the pairs of balanced modulators will result in increases in the level of the undesired sideband. This basically means that unless very accurate automatic balancing systems are used with phase derived single sideband, that frequent adjustments will have to be made such as naturally result from changing tube parameters. In addition, adjustments with changes in frequency are required with phase derived systems. This is offset somewhat by the advantage of being able to generate the single sideband at the final operating frequency.

While some work has been done on the phase derived approach to single sideband, these demanding requirements of balanced quadrature drive voltages modified by varying tube parameters present a challenging problem in applied engineering when considered for continuous tuning wide frequency coverage transmitters, such as those needed by the Signal Corps. These problems have not been solved to the high degree of reliability required of a device to be turned over to a relatively unskilled military operator in a mobile radio unit.

New Components

There have been recent developments of new transmitting tubes with high power gains, requiring virtually no grid driving power. Thus, the need for heavy grid swamping of linear amplifiers to reduce distortion due to the non-linear nature of the grid drive load as presented to the preceding stage, is obviated. The resultant large increase in actual power gain attainable in one stage makes possible the application of one or two linear amplifier stages to raise the power of the low level single sideband signal to a value of several hundred watts. Thus, the filter method discussed briefly in the preceding paragraphs deserves renewed attention. One of the principal

objections to a low level single sideband generation is largely eliminated, that is, a long string of difficult to tune cascade Class B linear amplifiers is no longer a necessity.

The very recent development of suitable mechanical filters furthers the attractiveness of such a system. Mechanical filters occupy a small fraction of the space required by conventional crystal filter units and, happily, are substantially lower in price.

The Transmitter Exciter

It has always been a problem to provide continuous tuning in a single sideband transmitter due to the relationship of the fixed frequencies used in the low level stages of the transmitter to heterodyne, usually in two steps, the single sideband signal to the final operating frequency. These frequencies did not bear a simple harmonic relationship with one another and, therefore, were not adaptable to ganged tuning.

A scheme has recently been derived which may be likened to the action of a superheterodyne receiver with a variable frequency intermediate frequency amplifier. The main difference is that in the superheterodyne receiver the objective is to remove a signal from somewhere in the high frequency range and heterodyne it down to a lower frequency in order to obtain selectivity and detection. The object in the transmitting case, however, is to take a fixed low frequency signal and conveniently heterodyne it to some frequency that may lie anywhere in the high frequency spectrum.

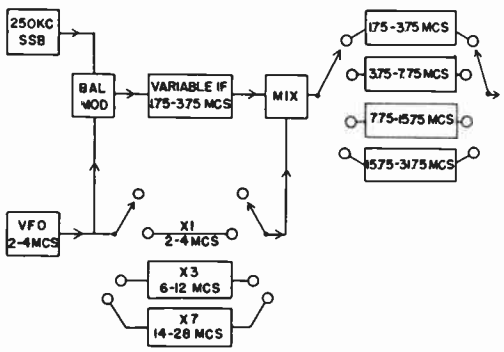


Fig. 4 - Exciter Scheme.

Illustrated in Figure 4 is one method of accomplishing this. A single sideband signal at 250 kc, which may have been derived by either the phase shift method or by means of mechanical filters, is applied to one of the inputs of a balanced modulator. To the other input of the balanced modulator is applied the output of a variable frequency oscillator operating in the frequency range of 2 to 4 mc. Neglecting for a moment the frequency multiplier stages at the lower portion of the Figure, we see that the output of the balanced modulator will contain the sum and the difference frequencies of the variable frequency oscillator and the 250 kc single sideband signal. In this particular case we will choose the difference frequency. The difference frequency will lie in the range between 1.75 and 3.75 mcs, depending on the frequency of the variable frequency oscillator.

If this frequency is fed through the variable frequency IF stages, which are ganged tuned with the variable frequency oscillator, and are fed through the amplifier stages of the transmitter, we can cover the output frequency range of 1.75 to 3.75 mcs per second.

Let us now examine the operation of the second mixer and the frequency multiplier stages. We may also apply the variable oscillator frequency of 2 to 4 mcs to one of the inputs of the second mixer, and at the same time apply the already generated 1.75 to 3.75 mc single sideband signal to the other input of the second mixer. This time we will choose the sum frequencies in the output of the second mixer so, therefore, the ganged tuned stages at the output of the second mixer will contain the sum of 1.75 and 2 mcs, or 3.75 mcs, which is the beginning of the next band. When the VFO is tuned to its highest frequency, or to 4 mcs, the single sideband signal passing through the variable IF will be at 3.75 mcs. The sum of 4 mcs and 3.75 mcs is 7.75 mcs, which is the upper end of our second band.

We can do this same trick with a multiple of the VFO frequency. Let us multiply the VFO frequency by three, so that the one input to the second mixer will contain the frequency 6 to 12 mcs, while maintaining the same single sideband input at the other input to the second mixer. Let us see what band of frequency this will give us. Starting once again with the VFO at its lowest frequency, we will have 6 mcs plus 1.75 mcs, which gives us 7.75 mcs, the beginning of band three. Raising the VFO up to 4 mcs will give us the other end of the band, which is 12 mcs plus the 3.75 mcs, or an upper band limit for band three of 15.75 mcs. The entire HF range may be adequately covered using this scheme. You will note that in all cases that the output frequency is always an integral multiple of the variable frequency oscillator minus 250 kc. This scheme therefore appears particularly adaptable to ganged tuning where continuous frequency coverage is required. Of course, where it is not necessary to provide continuous coverage somewhat simpler schemes may be applied.

Need for Automatic Frequency Control

Some additional consideration must be given to a peculiarity of suppressed carrier single sideband systems. That is the basic problem of supplying a locally generated carrier to use in detecting the single sideband signal. This can be readily understood by referring to Figure 5.

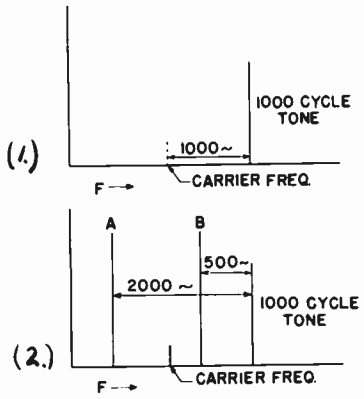


Fig. 5 - Dependency of the Detected Output on the Proper Local Carrier Frequency.

The single RF signal in (1) is a 1000 cycle tone transmitted as the upper sideband on a suppressed carrier. It thereby loses its audio sense, and the detected audio frequency is dependent upon the frequency of the locally generated carrier of the receiver. Any error as shown in (2) in the local oscillator results in the wrong audio output frequency. For speech communications it has been found that an error more than approximately 50 cycles results in almost complete loss of intelligibility.

In view of the above, the need for an AFC system can be clearly seen, inasmuch as no mobile equipment with such a degree of frequency stability appears feasible at this time. Even if such stability were attainable, unapproached settability of the receiver frequency would be called for, if it were desired to tune to a new frequency in advance of any transmitted signal so that no coordinated line-up of the system would be needed. For mobile applications with push-to-talk operation this is obviously impossible.

Alternatively, the suppressed carrier can be separated from the sideband intelligence by suitable filtering in the receiver IF, and then "reconditioned" by means of limiters so that fades and noise are removed, and it is suitable for use in detection. In fixed plant installations both means are available in the same equipment, although reconditioned carrier is seldom used. While the reconditioned carrier provides exact audio frequency restoration, severe fading conditions render it inferior to a locally generated carrier for normal use. This system offers no advantages over the locally generated carrier other than slightly more precise restoration of the audio frequency, inasmuch as an extremely narrow filter must be used to separate the carrier from the sidebands and, hence, AFC of the same order as that required for the locally resupplied carrier is necessary.

From the preceding comments, it is obvious that some type of fast acting wide range AFC is an essential for any practical push-to-talk radio system.

One of the basic difficulties with any scheme for automatic frequency control having a wide and rapid pull-in range is the vulnerability of the receiving equipment to capture by strong adjacent signals or even strong components within the sidebands of the desired signal.

In the single sideband receiving equipment in use at present for fixed plant application in more or less fixed frequency service with highly stable transmitters and receivers, the use of a narrow band filter approximately 20 cycles wide in the carrier channel prevents capture by strong adjacent signals, or by other signals and tones within the sideband. Capture is still possible, however, if an extraneous signal drifts slowly through the carrier filter. Of course, the major difficulty with this scheme for mobile systems of communication is the necessity of requiring a somewhat skilled operator to tune the receiver with sufficient accuracy so as to place the receiver carrier within the 20 cycle band pass of the carrier filter. Because of drift, it would be necessary to repeat this tuning process each time the carrier was turned off for more than a few seconds.

AFC Scheme

In the fixed plant application of single sideband it had been suggested that some form of carrier coding be used in order to automatically tune the receiver to the precise frequency, the operator only being required to provide approximate tuning adjustments within several kilocycles. One of these schemes seems particularly suited to providing the required fast pull-in automatic frequency control system essential to mobile single sideband communication.

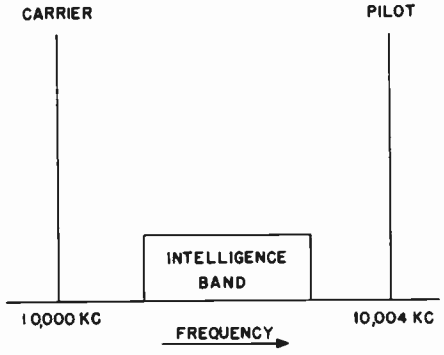


Fig. 6 - Relation of Carrier and Pilot Tone to Intelligence Band.

As shown in Figure 6, this scheme utilizes an additional carrier or pilot tone displaced somewhat from the carrier frequency by, say, three or four hundred cycles more than the maximum intelligence bandwidth desired to be transmitted. For example, if the intelligence bandwidth was to be 300 cycles to 3500 cycles, we might place this auxiliary pilot tone at about 4 kcs from the carrier. Now, what we have done

is to provide a transmitted beating oscillator frequency which will be utilized much in the same manner as the picture carrier of a television signal is used to demodulate, or rather to heterodyne the sound carrier of a television signal in intercarrier sound systems. This should be familiar to most of us.

It is not intended that this transmitted carrier and pilot tone be sent continuously at the full power output of the transmitter. This would defeat the purpose of single sideband insofar as power gain is concerned. Fortunately, the characteristics of an AFC circuit in the receiver are such that frequency information is not continuously required, but is necessary only during the very beginning of a transmission, and occasionally in between to provide minor adjustments for slight drift. The carrier and pilot tone, therefore, can be transmitted at full power output between pauses in speech transmission. This is easily accomplished by controlling the carrier and pilot insertion at the transmitter in inverse proportion to the speech input to the transmitter.

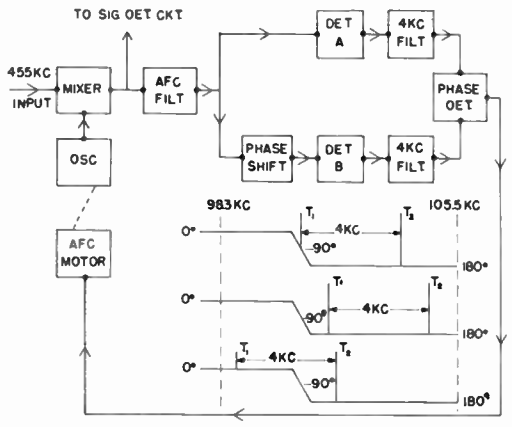


Fig. 7 - AFC Scheme to Utilize Carrier and Pilot Tone.

As illustrated in Figure 7, you will note that the carrier and pilot tone pass through a mechanical filter of such bandwidth as to limit the maximum pull-in range of the automatic frequency control system. The AFC information now goes to two paths, path A and path B.

The signal that follows path A, consisting of both the carrier and the pilot tone, reaches detector A, where the difference frequency of approximately 4 kcs is detected. Of course other signal products may also be detected, so the detector is followed by a 4 kc filter. This 4 kcs is present as long as both the carrier and pilot are within the band pass of the AFC mechanical filter. We will now leave the signal in path A.

The signal following path B receives the same treatment as the signal following path A, with the exception that interposed before the detector, that is, between the mechanical filter for AFC and the detector B, is a phase shifting network with the phase characteristic as shown on the slide. The 4 kc outputs from detector A and detector B are fed to the phase detector. The phase detector compares the phase of the 4 kc output from detector A with that from detector B. Now it so happens that the phase relationships between the two detector outputs bear the same phase relationship as that between the carrier and the pilot tone in the path containing the phase shift network.

The DC output of the phase detector is fed around the AFC loop in order to control the oscillator that heterodynes the signal to the last intermediate frequency.

You will note that as long as both signals are to the right of the phase reversal of the AFC phase shift network, that the DC output of the phase detector will always drive the system in such direction as to place the carrier at the 90° phase shift point, so that the signal is at the proper frequency for demodulation. If one of the tones is to the left of the phase reversal and the other to the right of the phase reversal, the DC output of the phase detector is so connected as to drive the composite signal to the right or higher in frequency until once again the carrier falls at the 90 degree point. Remember that the output of the phase detector following the 4 kc filters is zero at the 90 degree point. If both carriers should appear to the left of the phase reversal, erroneous information would be presented by the phase detector since it could not tell the difference between this condition and the condition when both carriers are to the right. However, this is prevented by the bandwidth limitations of the AFC mechanical filter. So much for the AFC system. You will note the block diagram, Figure 8, shows a signal mechanical filter followed by a detector. The detector, of course, must have a local carrier generator in order to heterodyne the single sideband signal down to audio frequency.

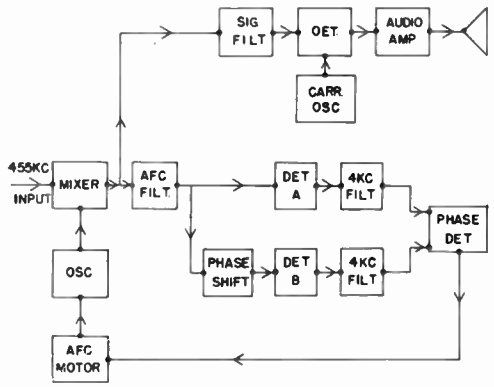


Fig. 8 - Receiving Circuit.

It should be noted that the zero or in-tune condition of the AFC scheme, determined by the 90 degree point of the AFC phase shift network, has to fall with considerable precision at the frequency of the local carrier oscillator.

It can easily be seen that the maximum pull-in range of the system is dependent upon the pilot tone displacement from the carrier and, hence, the transmitted intelligence bandwidth. This is not overly narrow for speech service with the frequency stabilities and settabilities obtainable by application of techniques currently employed in many new Signal Corps equipments.

Summary

To sum up, single sideband communications offers:

1. Increased efficiency and higher effective power than conventional double sideband systems of the same power handling capabilities.

2. Resultant decreased power supply drain in mobile use.

3. Spectrum economy.

4. Reduced antenna voltages.

While single sideband has not been practical for mobile use in the past, new components and techniques can overcome the objections formerly encountered. It deserves reconsideration in view of these developments.

We have described only one of several possible systems of mobile single sideband communications. While we feel that it is a workable system of high merit, we do not wish to close the door on completely different ideas. On the contrary, as stated in the opening words of this paper, we hope that the result of our talk is a fresh outlook devoted to the problems involved, with enough encouragement provided by our example of one workable system to attract further attention to mobile single sideband communications.

Acknowledgements

We wish to thank Mr. Edwin D. Evans, Mr. Bernard Goldberg, and many other co-workers too numerous to mention whose efforts have contributed to the material presented in this paper.

MAJOR FACTORS IN MOBILE EQUIPMENT DESIGN WITH EMPHASIS ON 460 MC. MOBILE EQUIPMENT CHARACTERISTICS

By John F. Byrne and Angus A. Macdonald
Motorola, Inc., Communications & Electronics Div., Chicago, Illinois

The rapid expansion of mobile radio in the 30 to 50 and the 152 to 162 megacycle bands has left most of us who are associated with the field somewhat desperate, and a little short of breath. The feeling of desperation comes from the knowledge that the available spectrum is being chewed up at a terrific rate. The shortness of breath is caused by trying to keep up with the rapid advances in the art, which have permitted greater and greater utilization of the available spectrum. Most of you here will recall that only a few years ago, channel allocations of 120 kilocycles were considered closely spaced. Now 40 kilocycles in the low band and 60 kilocycles in the high band are the rule. At the present time, some 20 kilocycle channel allocations are being used in the low band. A subcommittee of the Joint Technical Advisory Committee has been charged with the task of determining whether or not the high band channel separation can be reduced to 30 or even 20 kilocycles! While this struggle for channel space in the low and high land mobile communication bands is at its height, adjoining services, whose requirement for a large quantity of stations is not as acute, are calmly squandering their assigned spectrum with channel spacings of 200 and, in a few cases, 100 kilocycles.

After worrying with the struggle for additional channels in the 30 to 50 and the 152 to 162 megacycle bands, it was a great relief in 1949 when the Federal Communications Commission provided 100 additional channels for the land mobile services by opening the Ultra High Frequency band from 450 to 460 megacycles for developmental licenses.

Most of the radio users who are now pioneering this new band have done so out of desperation. They had to have radio, and the new channels were the only ones available. Their initial fear and hesitancy has been supplanted in most cases, by admiration for the new band. The two "bugaboos" which were; New and untried equipment, and propagation problems were found, as is so often the case, to fade away when confronted.

This paper will outline the special characteristics of the receiver and the transmitter which are peculiar to the 450 to 460 megacycle band. Finally, the overall system performance will be discussed with reference to both propagation data and field experience.

The equipment to be discussed was designed to cover the complete range from 450 to 470 megacycles. This was done so that it could also be used in the 40 Class A channels allocated at the edges of the so-called Citizen's band, which extends from 460 to 470 megacycles. The Citizen's band is defined as: A fixed and mobile service intended for use for private or personal radio communication, radio signalling, control of objects or devices by radio, and other purposes not specifically prohibitted. Any citizen of the United States, eighteen years of age or over, is elegible for a station license in this service.

Allocations in the U.H.F. band have been made on the basis of 100 Kc. channel separation. Other than this, no specific standards have been set down by the FCC. However, good engineering practice would lead to performance specifications similar to those required in the 160 Mc. band. A transmitter frequency tolerance of $\pm$ 0.001% or $\pm$ 4.6 Kc. at 460 Mc. might seem reasonable. Frequency deviation should be limited to $\pm$ 15 Kc., and some form of automatic sideband radiation control should be employed to prevent modulation splatter into adjacent channels.

Receiver:

In order to operate effectively in a 460 megacycle system, the receiver must meet several very stringent requirements: First of all, for reasons that will be shown later, it should remain within $\pm$ 2.5 Kc. or $\pm$ 0.00055% of the transmitter frequency. Second, it should have a sensitivity of 1 microvolt or better. Third, it must reject intermodulation products to the greatest possible extent. Finally it must have sufficient selectivity to provide a high rejection to the adjacent channel signals and all spurious responses.

The relative frequency stability between the receiver and transmitter at 460 megacycles turns out to be quite a problem. Exhaustive laboratory and field tests have lead to the "rule of thumb" that the relative drift between the receiver and transmitter should not exceed 1/12th of the receiver bandwidth if good system performance and ignition suppression are to be maintained. For a bandwidth of 30 Kc., then, the relative drift should not exceed 2.5 Kc. If the allowable drift is divided evenly between the receiver and transmitter, we can allow only $\pm$ 1.25 Kc. for each, or $\pm$ 0.00027%. This high degree of stability is a little beyond that obtainable with reasonably priced crystals and ovens.

The problem has been solved in this equipment by applying automatic frequency control to the receiver high frequency crystal oscillator. The transmitter frequency is controlled by means of an oven mounted crystal, within $\pm$ 2.3 Kc. The receiver crystal is not oven controlled and the receiver is allowed to drift approximately $\pm$ 13.7 Kc. Therefore, the maximum relative drift is about 16 Kc. The correction ratio of the automatic frequency control circuit is 8 to 1. Under the worst conditions, therefore, the receiver detuning will not exceed 2 Kc., which is well within the allowable limit of our 2.5 Kc. "rule of thumb". The automatic frequency control is designed to correct the tuning of any signal within 30 kilocycles of the receiver frequency. Precautions are taken in the design to prevent tuning to the next channel.

The second basic problem is that of receiver sensitivity. From field experience and from data on solar and man-made noise levels in this band, we find that a sensitivity of one microvolt would be useful. Sensitivities of 0.45 microvolt have

been obtained in this frequency range; however, a practical figure for a production receiver with a minimum of tube selection is 0.6 to 0.8 microvolts for 20 db. of quieting. This corresponds to a noise figure of about 8.5 to 11 db. In order to provide sufficient safety factors to allow for tube aging and test equipment variations, the receiver sensitivity is advertised to be 1 microvolt. All the sensitivity figures mentioned are "closed circuit" microvolts, or, the signal voltage across the receiver terminals.

The "closed circuit" microvolt mentioned above is better known as the "advertising microvolt" since it is usually half the I.R.E. or R.T.M.A. standard "open circuit" microvolt. We resort to it as a protective measure since the "advertising microvolt" is commonly used in competetive sales literature and sales talk. If I have managed to confuse the issue it is only because it is a confusing issue. Since receivers are power operated devices, it would seem very sensible to rate their sensitivities in decibels below 1 watt. This rating would eliminate the confusion in everyone's mind concerning the big and little microvolt. On a power basis, the receiver in question would be rated at -137 Dbw. Because of the frequency advantage, a good 160 Mc. receiver would have a sensitivity of about -143 Dbw.

The receiver sensitivity depends upon several factors. First is the choice of the R.F. amplifier tube, or, tubes and circuit. The use of disc seal types of tubes in the R.F. amplifier would lead to a good noise figure, but the tube cost would be high. It was felt that the use of more conventional miniature tubes would provide adequate sensitivity and reasonable tube cost. Many tube types were carefully tested for noise figure and gain. Of the miniature tubes tested, the 6J4 provided the best noise figure and uniformity, and was selected for use in a grounded grid circuit for production.

The second important factor in receiver sensitivity is radio frequency gain. If a good noise figure is to be obtained, sufficient radio frequency gain must be provided to override the noise generated in the first mixer stage. In order to achieve this result, it was necessary to cascade two grounded grid 6J4 R.F. stages.

Corralary to the problem of achieving good receiver sensitivity is the problem of designing high Q, compact, tunable resonators for use in the grounded grid circuits. After rather extensive theoretical and experimental work, during which time everything that would tune the frequency range from gold tie clips to coaxial cavity tuners was tried, an extremely satisfactory coaxial cavity tuner was finally developed.

Figure 1 shows a cross section view of the tuner, which is about 4-1/2 inches long and 7/8 inch in diameter. The unit is tuned by unscrewing the end cap and inserting a standard metal screwdriver into the tuning slot. Contact is made between the movable center conductor and the stationary center conductor by silver plated contacts. Smooth tuning adjustment is accomplished by carefully dimensioning the capacitive end slug. Temperature compensation is built into the construction of the center conductor assembly by the use of two types of metal with different thermal coeficients of expansion.

Input and output coupling loop leads are supported and brought out of the tuner through glass seals. Soldering the loops into the correct position is done by the aid of special assembly fixtures which permit accurate control of input and output impedances.

The removable end cap compresses a circular neoprene ring to provide an hermetic seal in the cavity. Contact fingers are further protected against corrosion by a small application of DC-4 Silicone.

Three of these tuners are used in conjunction with the two 6J4 grounded grid amplifiers, and a fourth is used to filter the injection voltage fed to the first mixer tube.

Figure 2 shows a block diagram of the receiver. It should be noted that the receiver employs triple conversion. The first I.F. tunes from about 71 to 75 megacycles. The second I.F. tunes from 8 to 8.9 megacycles. It should be noted that the injection voltages for both the first and second conversion are derived from the same crystal oscillator which operates from about 31.5 to 33 megacycles. Automatic frequency control is applied to this oscillator to maintain the receiver frequency within 2 kilocycles of the transmitter frequency regardless of relative drifts.

The use of grounded grid R.F. amplifiers makes the control of intermodulation more difficult than it is in a low or high band set because biasing a grounded grid R.F. amplifier increases intermodulation. Therefore, intermodulation control must be attained by other means. In this receiver, control of intermodulation reception interference is achieved by reducing the injection voltage to the first mixer, and hence the receiver gain, as the desired signal strength increases above the sensitivity threshold.

In the final conversion the desired signal is converted to 455 Kc. Again the injection voltage is crystal controlled by an oscillator whose frequency ranges from 8.5 to 9.4 megacycles.

The use of triple conversion permits attaining the same selectivity at 460 megacycles as is obtained in the low or high bands. At the same time, all spurious responses including all images are attenuated by more than 85 db.

The receiver selectivity is achieved at 455 Kc. by means of the same type of packaged filter used in high and low band receivers. The Motorola Permakay Filter is permanently protected against tampering, humidity, and shock by potting. There are now over 100,000 of these filters in use that

have proven themselves in other bands. The filter is guaranteed for the life of the equipment.

Low standby drain is achieved in the receiver by biasing the audio stages to cutoff when the receiver is squelched. Low drain is further aided by the omission of the receiver crystal oven.

Figure 3 gives a summary of the receiver specificatcns which are being met by this Ultra High Frequency mobile receiver.

Figure 4 shows a top view of the complete receiver. The coaxial cavity tuners can be seen on the far side of the chassis at the right. Remember that the receiver chassis is only about 4-7/8 by 16-1/2 inches and weighs about 7-1/2 lbs.

Transmitter:

Now let us turn our attention to the special problems associated with the 450 to 470 megacycle mobile transmitter which is to be a companion to the mobile receiver just described.

The mobile transmitter should provide a power output of about 20 watts. This should be accomplished using tubes which will provide long life, for low maintenance; and high efficiency, for minimum battery drain. A high drain transmitter will appreciably shorten the life of the automotive battery. In order to operate successfully with the receiver automatic frequency control, we have seen that the transmitter frequency stability should be $\pm$ 0.0005% or better.

The universal problem confronting all transmitter engineers is the problem of attaining a desired power output at a reasonable efficiency and tube replacement cost. This problem becomes particularly hard to handle in the frequency range of 450 to 470 megacycles. In addition, the fact that we are dealing with a mobile transmitter, imposes the requirement that the tube type used be rugged enough to withstand vibration and hard knocks; and it stresses the requirement of high efficiency, for reduced battery drain.

After consideration and test of the various tube types available which would operate in the 450-470 Mc. range, it was apparent that 2C39A disc seal triode would best meet the above requirements.

As used in the new Motorola transmitter, the 2C39A yields a plate efficiency of about 65%. An understanding of the importance of this efficiency can be gained by a comparison between it and the plate efficiency to be obtained from a more conventional tube at these frequencies. A good, conventional, tube of a type suitable for use at 160 Mc. provides a plate efficiency of only about 25% when operated in the 460 Mc. band. This difference in plate efficiency in a 20 watt transmitter represents a difference in drain from the battery in a 6 volt system of approximately 12 amperes. Of course, the saving in power supply size and cost where plate efficiencies are high goes without saying.

The 2C39A easily meets the requirements of being rugged and being operated well within its ratings. The 2C39A has a maximum plate dissipation rating of 100 watts and in the mobile transmitter its plate dissipation is only about 10 watts.

The initial cost of a 2C39A is higher than the cost of conventional tubes designed for use at lower frequencies, but since it is used at a small percentage of its maximum ratings it will have a life expectancy many times that of a cheaper tube used up to the limit of, or in excess of, its ratings and hence the cost of tube replacement in dollars per hour for the expensive tube will be less than the cost per hour of the more conventional tube.

Along with the problems of power output and selection of a suitable tube goes the corallary problem of the design of a stable, high efficiency tank circuit to work with the tube. Figure 5 shows a drawing of the plate cavity circuit, designed for use in the 2C39A driver and grounded grid power amplifier in the 20 watt transmitter strip.

In the power amplifier the grid is bypassed to ground by means of a metal plate to which the grid connector is soldered. This plate is insulated from ground by means of a sheet of silvered mica. The center conductor of the cavity consists of a piece of rectangular pipe to which the plate connector is soldered. The end of this pipe opposite the tube is flared out into a rectangular flange which forms one plate of a bypass capacitor whereby the center conductor is connected "RF wise" to the outer cavity. Dielectric insulation in this condenser consists of two sheets of silvered mica. The rectangular center conductor serves a dual role, both as a portion of the RF circuit and also, in the case of the 40 watt unit, to be discussed later, as an air duct through which cooling air from a blower is directed at the anode fins on the tube. The heater and cathode connections underneath are provided with heavy heat radiating connectors to draw heat away from the heater seal.

The tripler-driver stage is operated as a conventional grounded cathode frequency multiplier. This arrangement is employed so that a good thermal path to the chassis can be provided thereby reducing cathode seal temperatures in the absence of a blower. No blower is required for the 20 watt mobile unit.

R.F. drive for the final amplifier is injected directly into the cathode circuit which is fixed tuned. The cathode tuned circuit consists of a heavy aluminum strap which provides the necessary thermal conductivity to the chassis to reduce seal temperatures of this cathode. The plate cavity circuit is tuned capacitively by means of a single disc capacitor located directly over the end of the line. Power is coupled out by means of a rotatable loop coupled into the cavity near the current maximum.

The transmitter block diagram is shown in Fig-

ure 6. The crystal oscillator employs a harmonic mode crystal oscillating at about 19 megacycles. The oscillator frequency is multiplied 24 times to obtain the final frequency. The techniques employed are similar to those used in the 152-162 megacycle band except for the 2C39A tripler-driver and 2C39A final stages. It should be noted that the Motorola standard instantaneous deviation circuits are used in the audio amplifier.

Figure 7 outlines the basic transmitter specifications which have been met in this design. Note the small size and weight of the completed unit.

The pattern of the transmitter spurious radiations is shown on Figure 8. It is interesting to note that there are no significant radiations within the 450 to 470 megacycle band. All spurious radiations are at least 60 db. below the carrier level. The only spurious radiations which are significant are at one third the final frequency, and its second harmonic, and the second harmonic of the final frequency.

Figure 9 is a photograph of the top of the complete transmitter. Note that the tripler and final amplifier cavities are placed end to end. This is done so that in base station applications this same basic unit can be provided with a blower to cool the 2C39A's and can be rated at 40 watts output.

Figure 10 shows the complete mobile receiver, transmitter, and 6 volt power supply, removed from the mobile housing. The complete assembly is 16-1/4 inches wide by 17-3/4 inches long by 6 inches high and weighs 58 lbs. The battery drain is 14 amperes standby and 39.5 amperes transmit.

System Performance:

The major design considerations and performance characteristics of this relatively new 450 Mc. equipment have been outlined briefly. It will now be worth while to spend a few moments on some important aspects of overall system performance.

First, it must be said that the performance of the equipment described has lived up to expectations. Field experience has shown that the design aims of ruggedness, dependability, long life, and low maintenance have been successfully achieved. Transmitter tube life has proven to be particularly satisfactory. A mobile unit on life test has run over a million transmissions at a duty cycle of 33%, with the same set of tubes. This is gratifying at a frequency where tube life is hard to get. U.H.F. systems in operation in many sections of this country have demonstrated that in general the maintenance problem is no more severe for this equipment than the maintenance of 160 Mc. equipment.

The most pleasant surprise of all is the propagation situation. It was thought for some time that users of the 450 Mc. band would suffer an appreciable loss in system performance relative to a 160 Mc. system. This unhappy thought has turned out to be false. In a straightaway range test on smooth earth with equal powers and antenna gains, the 160 Mc. system would appear to be about 6 db. or about 1.5 to one in range better than the 450 Mc. system in rural areas. However, the level of man-made noise even in suburban areas is about 9 db. higher in the 160 megacycle band than in the 460 megacycle band. Therefore, for the majority of actual systems, the 460 megacycle band will turn out to be about 3 db. better than the 160 megacycle band.

Figure 11 presents a summary of propagation data from 100 megacycles to 3700 megacycles. The abscissa is frequency and the ordinate is relative system merit in db. for suburban coverage. The fixed station antenna at each frequency was a half wave dipole, and the mobile antenna was a quarter wave whip. It can be seen from the curve that the 460 megacycle system has a slight edge over other bands even before high gain antennas are used. Of course, a high gain antenna for the base station is much smaller and cheaper at 460 megacycles than it is at 160 megacycles. The data for this curve was obtained from a recent article published in the Bell System Technical Journal by W. Rae Young of the Bell Telephone Laboratories entitled; "Comparison of Mobile Radio Transmission at 150, 450, 900 and 3700 Mc."

In urban areas also, the 450 Mc. band has proved itself to be superior to any other mobile frequency available. This is due primarily to the fill-in of holes which is obtained at the higher frequencies. Coverage can be obtained in tunnels and cuts which cannot be covered at the lower frequencies. The majority of urban coverage is not line of sight, and hence is dependent upon multipath reflection, which is increased at the higher frequencies. In addition to this advantage, man-made noise is greatly reduced in urban areas at 450 Mc. W. Rae Young found the noise level to be about 12 db. less at 450 megacycles than 150 Mc.

The results of these exhaustive field tests made by W. Rae Young, Jr. and others at Bell Telephone Laboratories, serve to corroborate field experience and other comparative field tests made in the 160 and 460 megacycle bands.

We have discussed some of the characteristics of the transmitter and receiver peculiar to the 460 megacycle band and have shown how some of the special problems have been solved. We have also shown that propagation phenomenon favor the use of the new 460 megacycle band for the land mobile services.

Field experience too has shown the 460 megacycle band to be equal or superior to the 160 Mc. band. Therefore, we should expect to see a rapid expansion of the land mobile services into the U.H.F. channels. We believe that the future of the 460 megacycle band is bright, and we at Motorola are proud of taking a leading part in its development.

In closing we would like to give credit to James Clark and Frederick Hilton, who were responsible for the receiver and transmitter developments respectively.

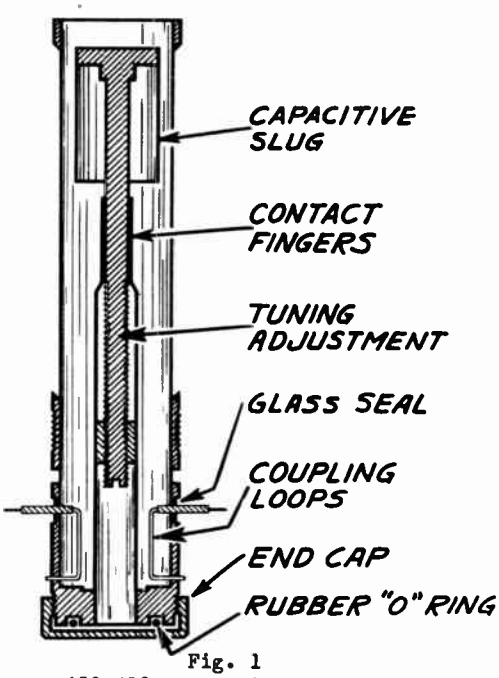


Fig. 1
450-470 mc receiver cavity.

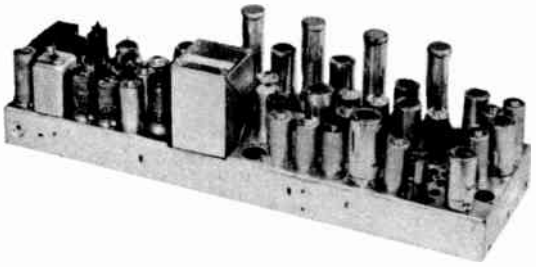


Fig. 4
450-470 mc receiver top view.

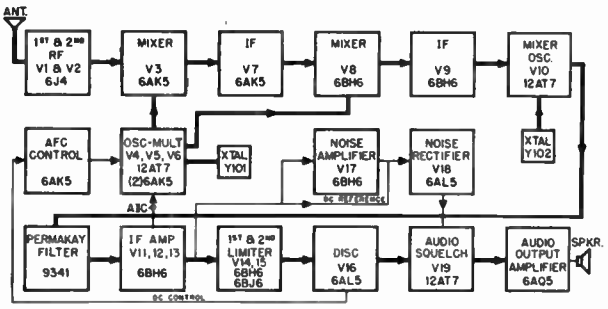


Fig. 2
Motorola 450-470 mc Sensicon receiver chassis.

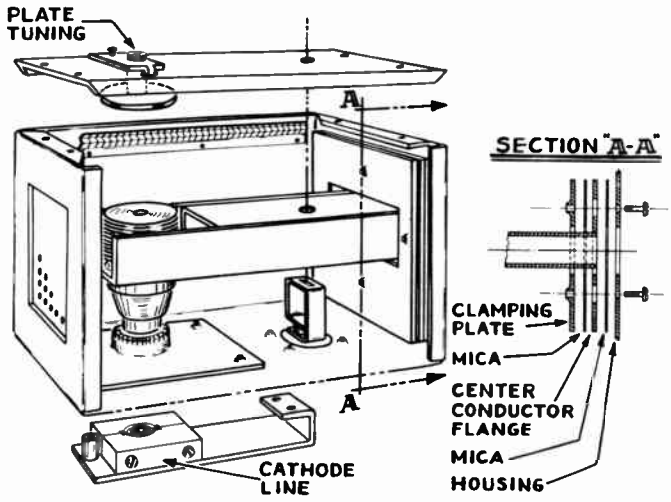


Fig. 5
450-470 mc transmitter cavity.

SENSITIVITY	1 microvolt for 20-db quieting
SELECTIVITY	± 15 kc at 6 db, ± 60 kc at 85 db
SPURIOUS	all below 85 db
STABILITY	with AFC, over-all stability better than ± 0.0005% from -30 to + 60° C
SQUELCH SENSITIVITY ..	better than 0.3 microvolts
AUDIO OUTPUT	1 watt less than 10% distortion
WEIGHT	7½ lbs.
SIZE	16½ long x 4 7/8 wide x 4 9/16 high

Fig. 3
Motorola 450-470 mc receiver characteristics.

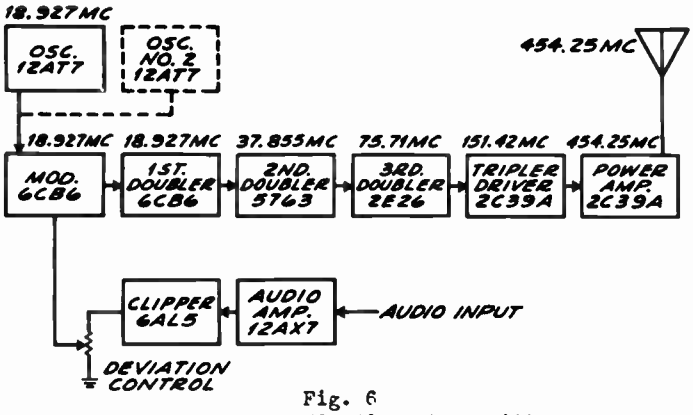


Fig. 6
Block diagram 450-470 mc transmitter.

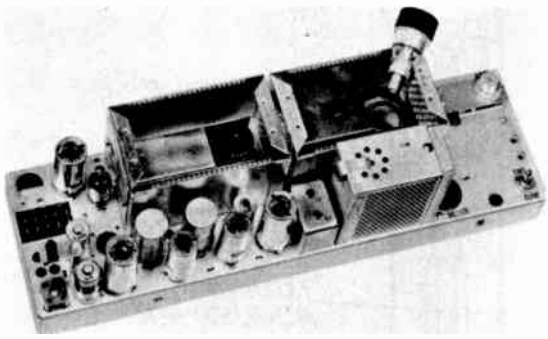


Fig. 9
Transmitter top view.

RF POWER OUTPUT	20 or 40 watts
SPURIOUS EMISSION	all in band below 85 db; all others, 60 db
FREQUENCY STABILITY ...	± 0.0005% (± 2.25 kc) from -30 to +60° C
MODULATION	± 15 kc for 100 %
AUDIO INPUT	about 0.2 v for 100 %
WEIGHT	7½ lbs.
SIZE	16 1/2 long x 4 7/8 wide x 5 1/4 high

Fig. 7
Motorola 450-470 mc transmitter characteristics.

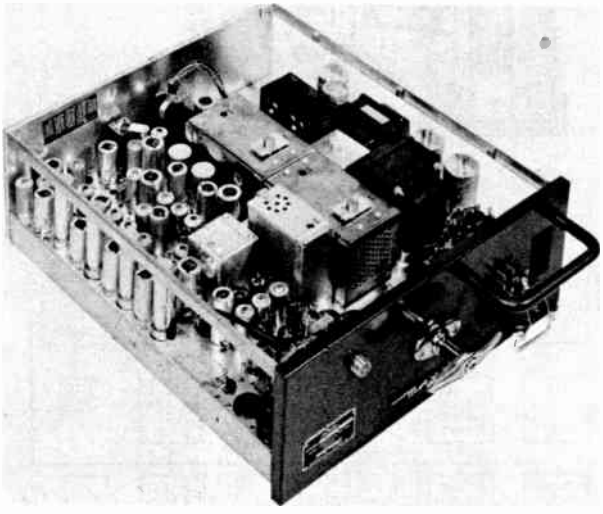


Fig. 10
450-470 mc transceiver.

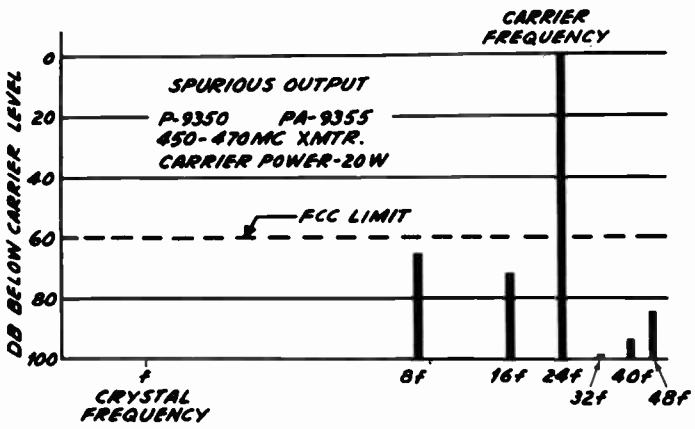


Fig. 8
450-470 mc transmitter spurious radiations.

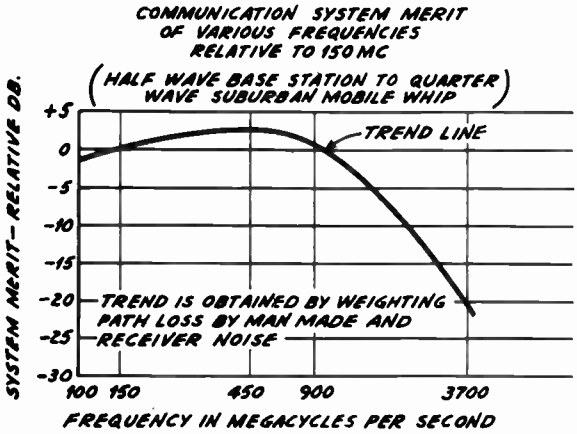


Fig. 11
Summary of propagation data.

FIELD EXPERIENCE WITH 450 MC MOBILE SYSTEMS

James Q. Montress
Bell-Mont Communications Service Corp.
Englewood, New Jersey

It has become increasingly apparent that the number of channel frequencies allocated to the mobile radio services are rapidly becoming saturated to the point where users of these services such as public safety, taxicabs, public utilities and news services are reluctant to use the present frequencies extending through 174 mc. It is now obvious that the only recourse is to expand the use of the new 450 mc band of frequencies.

We have conducted a series of very successful 450 mc surveys in a number of the larger cities throughout the country and have also engineered several large 450 mc systems. For the past year we have been actively engaged in the maintenance of these systems.

Initial Surveys

Our first tests using 450 mc equipment for comparison purposes were conducted in New York City from the roofs of the Daily Mirror and the Daily News buildings, where there were existing 150 mc systems in operation. From these locations excellent fill-in of the downtown walled canyons was achieved with the 450 mc signals. Actual coverage from both locations was approximately 90 to 95 percent of the 150 mc coverage with the emphasis on a complete lack of random noise and man-made interference.

The equipment used for these tests was a 100 watt base station and 15 watt mobile units. Receiver sensitivity at the time was 2 microvolts for 20 DB quieting and has since been improved to 1 microvolt or better. The antennas used were a Workshop 6HW, coaxial, coliniar array and standard one quarter wave rooftop whips on the mobile units.

On the basis of these tests both the Daily News and the Broadway Maintenance Corporation purchased and have had in operation for the past year 25-car systems using 450 mc equipment with antenna sites atop the News Building.

We conducted further comparison surveys in Philadelphia and Baltimore for the Raymond Rosen Company, engineering and maintenance contractors for the Yellow Cab Companies in both cities. From the existing 150 mc locations the 450 mc coverage was equal to the 150 mc coverage or surpassed it -- always with the same freedom from noise. These advantages are borne out by the fact that 450 mc equipment is currently being installed by both cab companies.

Typical 450 mc Operation

One of the foremost systems in the country today of the 20th Century Taxicab network in Newark, New Jersey. This system is comprised of two 100 watt base stations on 452.05 mc using one Workshop 6HW antenna and one Andrews 6 bay antenna. At the present time both units are operating single channel; however, there are plans to expand to two channel operation in the near future. The mobile fleet consists of 250 cabs using 15 watt transmitters on 452.55 mc. This system represents a typical large taxicab operation.

In our initial tests at Newark we experimented with many sites that would undoubtedly prove satisfactory for 150 mc operation, but they did not give the desired complete coverage of the City of Newark and its suburbs at 450 mc. The chosen location of the antenna is atop one of the two tallest buildings in downtown Newark which is also the site of three 150 mc systems. From this point, using 450 mc, it is definitely possible to talk as far as 25 miles with a solid coverage radius of 10 miles.

Before the inception of this system, radio was definitely out of the question for this user because of the heavy channel occupancy of the 150 mc band in the Newark area. These crowded conditions exist throughout the country, and where the factors are favorable to its use, 450 mc equipment is the answer and should be recommended. The additional cost of this type of equipment can easily be justified by its increased efficiency and freedom from competitive interference.

Conditions Observed at 450 mc

I would like to relate the experiences we encountered which are peculiar to these new frequencies.

As previously stated an antenna site giving adequate 150 mc coverage will not necessarily give good 450 mc operation. On the contrary it may fall considerably short of the 150 mc coverage. However, although it has been found that excellent or above average 150 mc locations will usually make good 450 mc sites, it it still very difficult to predict the field coverage from a proposed location.

High gain base station antennas are far superior to simple coaxial antennas and have definitely established their necessity in order to obtain solid coverage and fill-in of low signal areas. One of the most important factors is the proper location of the antenna. It should be as high as practical above all surrounding objects, including particularly large roof surfaces. If this precaution is not observed, there will be serious shading of certain areas due to the low angle radiation of the high gain antennas. Solid dielectric transmission line of the RG-8/U type should be used only for runs under fifty feet and RG-17/U for runs under one hundred feet. For all runs exceeding one hundred feet, air or teflon insulated line should be used.

It is interesting to note that ice and sleet formations on the base station antennas severely effect the signal propogation more so than at the lower frequencies, even though the final amplifier has been retuned to correct for the ice loading.

One condition encountered at 450 mc which is similar to that noticed at 150 mc is a rapid flutter when the mobile unit is in motion through what would be considered a low signal or fringe area. It is interesting to note that as the car is slowed down and moved in multiples of quarter wave lengths, the signal can fall from R-5 to R-1 in a distance of approximately six inches. Unlike 150 mc the nulls thus encountered appear to constitute a more complete absence of signal. On the other hand at a given normal car speed the rate of flutter is three times as high as at 150 mc and therefore does not chop out as much of the speech intelligence. This is a condition which may be improved at some later date, possibly with a mobile antenna having a large surface area. Another noticeable factor is that there seems to be no appreciable fringe area as encountered at the lower frequencies. The signal is either solid copy or it completely disappears within a relatively short distance.

In the approximately 30 square miles of the City of Newark there is only one small area of about one square mile where the signal strength is down in the region of 3 to 4 microvolts. This area is about 2 miles airline from the transmitter and is being shaded by a building of equal height, located about 150 ft. from the 450 mc antenna and directly between it and the shaded area.

By utilizing one or the other of the two antennas which are located on diagonally opposite corners of the building, it is possible to shift the shadow angle area by several city blocks. The lack of signal in this area can be attributed to the fact that beyond the shaded area there are no large reflecting surfaces either natural or man-made to bounce a signal back in. On the other hand, several miles from this area there are many points well below line of sight which receive excellent signals due to reflection from the mountain ranges beyond them.

Maintenance

Maintenance presents no special problems that the communications serviceman has not already encountered at the lower frequencies. Additional tubes, grounded grid RF amplifiers and cavity-tuned front ends are items of particular interest to the average serviceman, but after some indoctrination he will find the RF amplifiers extremely simple and the cavities very stable and easy to tune.

Of interest to the potential user of 450 mc equipment and the serviceman are the additional tripler and final amplifier tubes which are higher in cost than the average transmitting tubes used in the past. Contrary to first impressions the tube life is exceeding all expectations.

It must be remembered that with the additional tubes and current drain it is more difficult to maintain excitation to the tripler and amplifier. Care should therefore be exercised in maintaining the recommended drive to the higher frequency stages.or their tube life will be reduced considerably.

From a service standpoint some means of automatically controlling the frequency appearing at the IF system and the discriminator of the receiver is highly desirable. Because of the increased amount of frequency multiplication in the receiver oscillator chains and the frequency multipliers of the transmitter, there is room for more percentage of error in maintaining tolerances and it becomes more difficult to hold the receiver on frequency with the transmitter.

We have found by experimentation that as long as the system receivers are carefully maintained with respect to frequency and sensitivity, it is possible to reduce transmitter power output by a ratio of 10 to 1 with no apparent deterioration of the primary system coverage.

With the exception of the receiver 450 mc front end, the serviceman will find he is working with a common 40 mc receiver with the usual sensitivity and maintenance problems. He will be introduced to crystal diode mixers and longer oscillator chains to provide the high frequency injection, but metering of all stages simplifies the alignment or adjustment of the individual stages.

A requisite in servicing this type of equipment is a good on-channel signal source, such as a stable signal generator or preferably a crystal controlled frequency multiplier capable of being accurately attenuated.

The foregoing discussion has not been intended as a means of selling 450 mc equipment. Our intention has only been to present our experiences as we have interpreted them, and perhaps enlighten or assist those who will be stepping into what we feel will be a most interesting challenge to both the system engineer and the communications serviceman.

At this time the author would like to acknowledge the assistance given to me by the Link Radio Corporation and Mr. C. E. Denton of the New York Daily News.

AN AUTOMATIC TUNING COMMUNICATION TRANSMITTER

M. C. Dettman
Federal Telecommunication Laboratories, Inc.
Nutley, New Jersey

After the experience of World War II with communication transmitters in the low and medium frequency range, and similar experience with much later design VHF equipment, it became evident that a modernization program was highly desirable for communication transmitters operating in the medium and high frequency bands. Accordingly, a development program was instituted to build a modern communications transmitter which could effect communication under the rigorous conditions encountered in the Services with precision and reliability and without the necessity of preliminary calling and with absence of interference with other units of the communications system. It was desirable that this transmitter be rather versatile so that it might be used on a variety of surface and under-surface vessels.

A development program was instituted at Federal Telephone and Radio Corporation in 1947 which was subsequently transferred to Federal Telecommunication Laboratories. This project was to develop a transmitter meeting the needs indicated above. The first development model of this transmitter was presented by the contractor in the latter part of 1948. A development model substantially the same as the final production model was presented in August, 1950. Extensive laboratory tests were made on this unit. After the laboratory tests were completed, the equipment was installed on vessels and seven months of operational evaluation followed.

This new transmitter meets the following broad specifications. The transmitter may be automatically tuned and matched to specified antennas at any frequency between 300 kc and 26 mc. The operator may either select any one of ten preset channels or manually set up any frequency in the operating range by manipulating a group of knobs which directly indicate the frequency. The remainder of the tuning is accomplished automatically, including the antenna matching. Tune up time for most frequencies with proper antennas in use is approximately 30 seconds. Maximum time is 90 seconds. In the event that the operator wishes to tune the equipment under manual control, he may do so. The nominal power output of the transmitter is either 100 or 500 watts depending on the type of equipment and the radio frequency in use. Power output is limited to 100 watts nominal over the range of 300 kc to 2 mc. From 2 mc to 26 mc, power may be either 100 watts or 500 watts. The transmitter may be keyed at speeds up to 600 words per minute. Either polar or neutral keying signals are acceptable. Radio phone operation using a standard carbon or dynamic microphone is available. Facsimile operation with a 600 ohm input circuit and utilizing frequency shift for transmittal of intelligence is incorporated. Frequency shift keying operation with or without phase modulation is also available. All the aforementioned features are integral with the equipment.

In addition to these features, compactness of size and flexibility of installation were required. To this end, the entire equipment is broken down into groups as follows.

1. The transmitter group which itself is broken down into a basic 100 watt equipment and additional frames and chassis capable of raising the nominal power level to 500 watts. (Figure 1)

2. The antenna control group which may be mounted at any convenient spot but preferably adjacent to the transmitter bay. (Figure 2)

3. The antenna tuner unit and capacitor assembly which is always mounted at the base of the antenna to be used with the transmitter. (Figure 3)

4. The remote channel selector consisting of a waterproof case containing a telephone dial for selecting preset channels and an indicator showing the channel in use. (Figure 4)

These basic units may be assembled into three standard equipments known as AN/URT-2, AN/URT-3 or AN/URT-4.

The AN/URT-2 consists of the basic 100 watt transmitter bay shown in Figure 5, the antenna control group, antenna tuner, capacitor assembly and remote channel selector. The AN/URT-3 consists of an AN/URT-2 100 watt system plus a booster section composed of a high level modulator and a high voltage rectifier with associated frames, capable of raising the operating power output to 500 watt nominal. Figure 1 shows the transmitter bay. The AN/URT-4 consists of two AN/URT-2 systems and a booster section. The frames of the transmitter group are bolted together forming an integral assembly as shown in Figure 6. The AN/URT-4 makes available two complete systems capable of simultaneous operation on two channels at 100 watt level. At the option of the operator, power may be boosted to 500 watts on either one of these two channels provided the operating frequency is above 2 mc. Automatic interlocking is incorporated to ensure that power is limited to the 100 watt level below 2 mc even though the booster is selected for

500 watt operation.

Now that we have viewed and discussed the overall equipment, let us see what the individual chassis look like. Starting at the top of the main transmitter frame, we have the RF Amplifier unit. (Figure 7). This unit is basically designed for 500 watt operation. The vacuum tube lineup includes a 6AG7 feeding a 5933 (807W). These two stages are gang-tuned, driven by a tuning motor which in turn is controlled by a servo system when automatic tuning is used. The power amplifier tube is a 4-400A whose output circuit is designed to feed a 50 ohm line. This power amplifier has its own tuning motor and servo for automatic tuning. Included on the chassis are the servo system tubes and keying tubes. Most of the relays and control circuitry used for automatic tuning sequencing are also located on this chassis.

The next drawer down is the Low Level Radio Modulator and is shown in Figure 8. This unit includes the audio amplifier tubes, squelch circuits, clipper tubes and AVC circuits which are used for radio phone operation. The audio circuits have sufficient power output to modulate the transmitter in 100 watt condition. A group of tubes are utilized to take either hand keying or teletypewriter keying voltages and transform them into suitable signals for either A1 emission or frequency shift operation.

Another section of this chassis contains the equivalent of a standard two inch test oscilloscope with circuit sampling switches for checking the operation of various sections of the equipment. Additionally, two power supplies are included on the chassis. One is a regulated 250 volt power supply and the other, a 12 volt d-c supply for standard radiophone control unit remote box operation. A large number of the operating controls are centralized on the front panel of this unit.

Figure 9 shows the RF Oscillator. Its function is to provide the fundamental radio frequency required at a 2 volt signal level sufficient to drive the r-f amplifier unit. It must do this with a high degree of frequency precision, in the order of plus or minus 1½ parts per million ±20 cycles under all operating conditions. Included on the front panel are the preset switch banks which allow presetting to any desired 10 channels. Selection of any of these channels is by means of a telephone dial and an associated channel indicator meter over a two wire line. A direct reading dial indicating amount of frequency shift is on the front panel. This is used on frequency shift and facsimile to set the amount of deviation.

The chassis below this is the Low Voltage Power Supply (Figure 10) and includes 24 volt d-c supplies for control circuit relays and tuning motor operation, bias supplies, 250 volt and 300 volt supplies.

The bottom chassis in the basic 100 watt frame is the Medium Voltage Power Supply (Figure 11) It supplies 500 volts for the modulator tubes and r-f driver tubes and either 1300 volts or 1100 volts as required for the r-f amplifier tube for phone or CW operation.

The five individual chassis just discussed are mounted in the 100 watt frame shown in Figure 12. This frame contains all interconnecting wires and necessary connectors for plug-in operation of the chassis. Terminal boards are provided at the bottom for the tieing in the external cables to make an installation. For the AN/URT-2 100 watt installation, this frame is mounted on a standard base mount (Figure 13). This mount contains the main blowers which supply the necessary filtered air for cooling the units of the 100 watt frame. Four shock mounts at the bottom of the base mount support the frame. Two additional mounts at the top of the 100 watt frame prevent excessive sway.

In order to provide operation at a 500 watt power level at frequencies above 2 mc, the plate voltage of the final stage of the r-f amplifier is supplied from a three phase rectifier. The output of this unit is 2400 volts d-c for phone operation and 3000 volts d-c for other types of service. In order to plate modulate the final r-f amplifier at this level, a high level radio modulator is used. A pair of 4-125's are used as modulators. Part of the high voltage rectifier filter is in this chassis. The power selector switch for 100 or 500 watt operation is on the front panel. The high voltage rectifier and the high level radio modulator are located in two frames, making up the booster section (Figure 14). These may be mounted one above the other on top of the standard base mount and when used in conjunction with a single 100 watt frame the AN/URT-3 is formed.

The r-f power developed in the transmitter is fed via a 50 ohm coaxial cable to the antenna tuner unit, mounted at the base of the antenna. An exploded view of this unit is shown in Figure 15. The function of this unit is to automatically match the antenna to the 50 ohm transmission line with a normal standing wave ratio performance of better than 2:1 and a specification limit of 4:1. The antenna tuner is effectively a coaxial stub with a lumped center conductor. Its electrical length is about 18° at 300 kc. Two special broadband iron core transformers are used in this unit; one steps up the impedance of the 50 ohm circuit to 160 ohms and the other is used in the sensing bridges. These bridges supply information to the antenna control group for automatically tuning the antenna to the frequency in use.

In order to tune some specified antennas over the wide frequency range, an additional unit, the capacitor assembly is utilized. An internal view is shown in Figure 16.

This assembly varies the electrical length of antenna as presented to the antenna tuner when the

tuner fails to resonate the load with the capacitor assembly out of the circuit.

The operation of the antenna tuning unit and the capacitor assembly is determined by the antenna control group. Automatic tuning sequence information from the transmitter bay, capacitor assembly step position information, sensing bridge and tuning element position information from the antenna tuning unit are collected in the antenna control group. The input information thus collected is used to operate two two-phase motors in the antenna tuning unit in such a manner as to accomplish the necessary antenna match. In the event that a full scan of the tuner fails to resonate the antenna, the antenna control group changes the impedance of the antenna by controlling the capacitor assembly step position. The antenna tuner unit then cycles again. When the antenna circuit is resonated and matched, information is sent to the transmitter bay that proper tuning has been accomplished and communications may proceed. In the event that the antenna can not be resonated, the antenna control group sends this information to the transmitter bay and the operator may then take appropriate action.

The antenna control group is composed of a Control and Indicator unit shown in Figure 17. The upper two meters indicate the positions of the tuning elements. The special dual movement meter in the center indicates the output of the sensing bridges during the tuning process. When tuning is completed the meter indicates SWR on the 50 ohm line. A power supply for operating control relays is included on this chassis. There are two identical preamplifier units (Figure 18). After the sensing circuit bridge information has passed through the Control and Indicator unit it is fed to the proper preamplifier unit where it passes through a chopper converting it to 60 cycles a-c and amplified. These signals are then passed to the proper Power Amplifier (Figure 19) where the signal is further amplified to a level sufficient to drive the appropriate servo motor in the Antenna Tuning Unit. One amplifier chain controls the tuning servo motor and the other controls the coupling servo motor of the Antenna Tuning Unit. Figure 20 shows the dual regulated 300 volt power supply chassis which feeds the vacuum tubes of the Antenna Control Group.

The automatic tuning communications transmitter just described has met the requirements of precision frequency control, rapidity of tuning and provision for all usual types of emission required for a modern compact medium power communications transmitter.

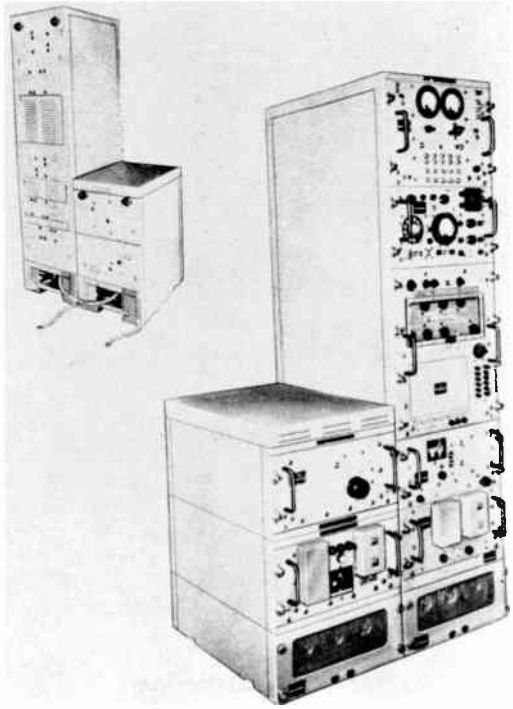


Fig. 1
AN/URT-3 transmitter bay.

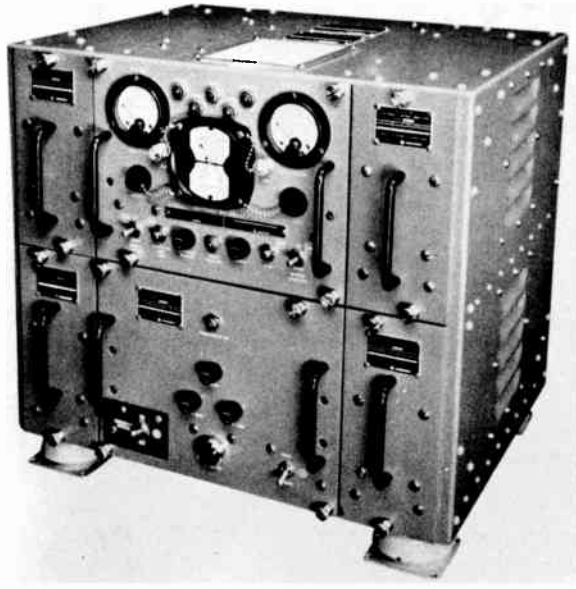


Fig. 2
Antenna control group.

Fig. 3
Antenna tuner unit (top)
and capacitor assembly (bottom).

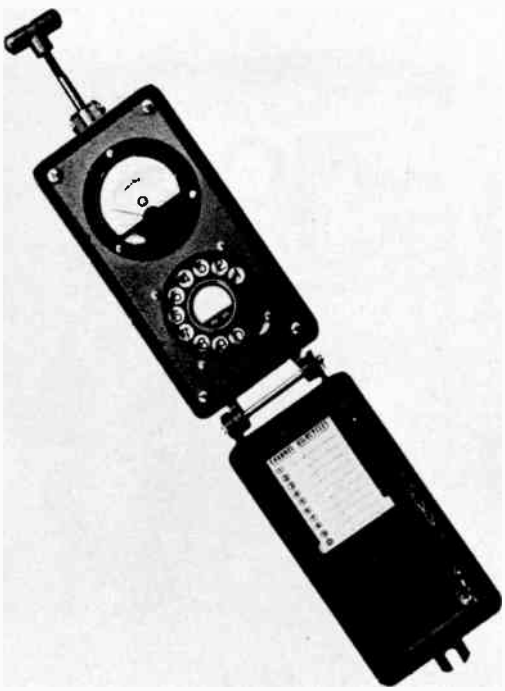


Fig. 4
Remote channel selector.

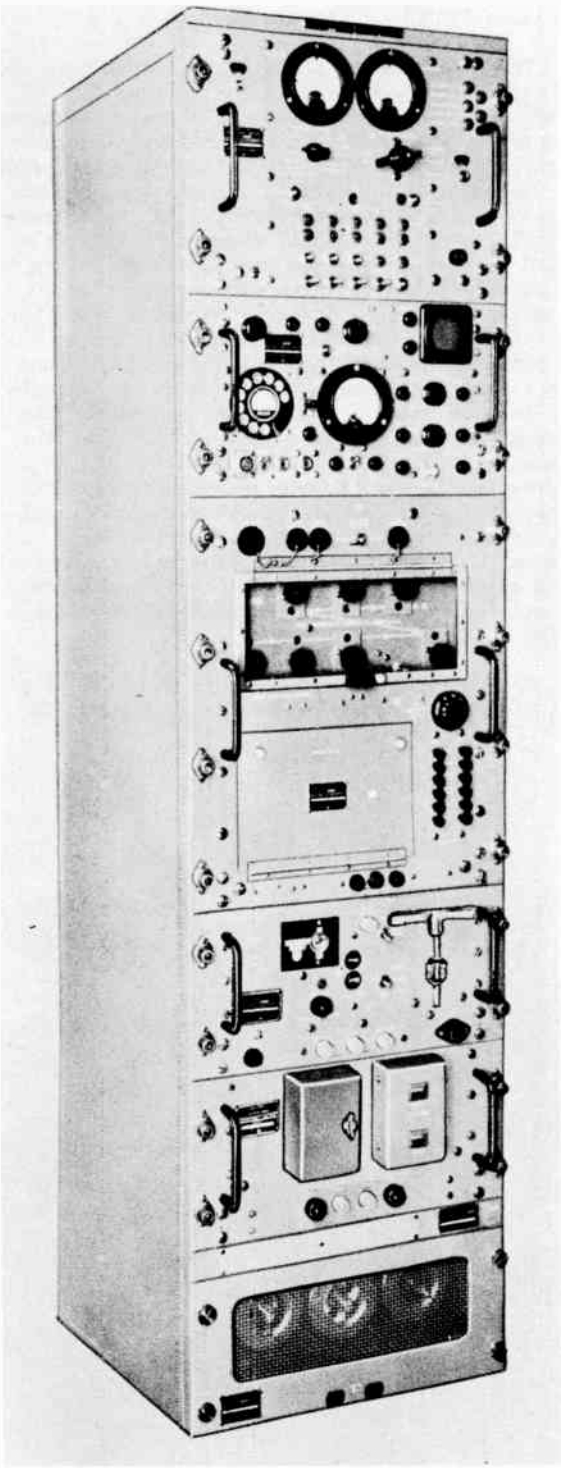


Fig. 5
AN/URT-2 transmitter bay.

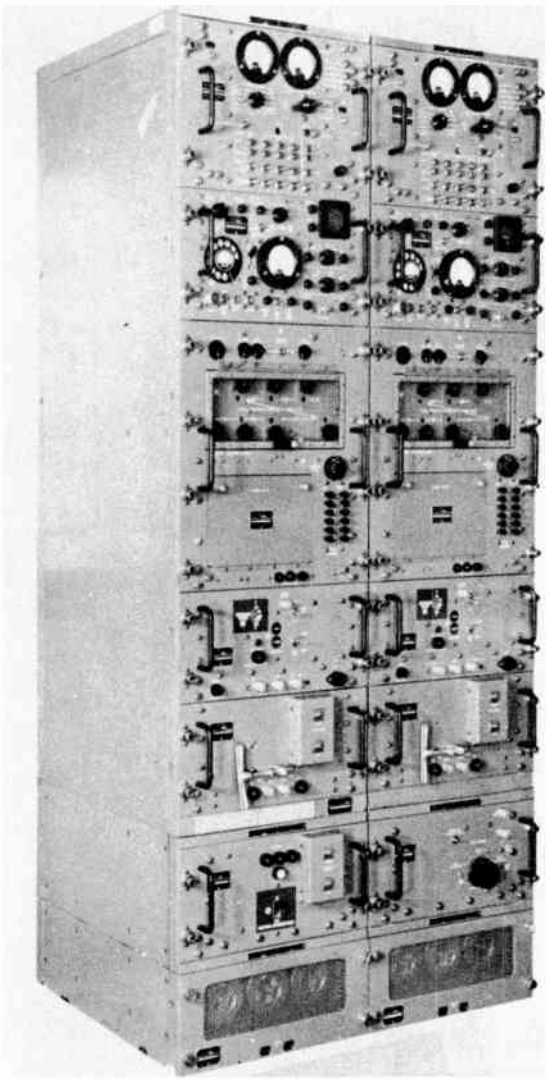


Fig. 6
AN/URT-4 transmitter bay.

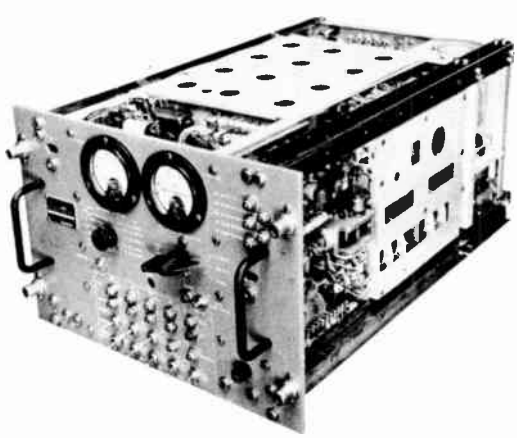


Fig. 7
Radio frequenoy amplifier.

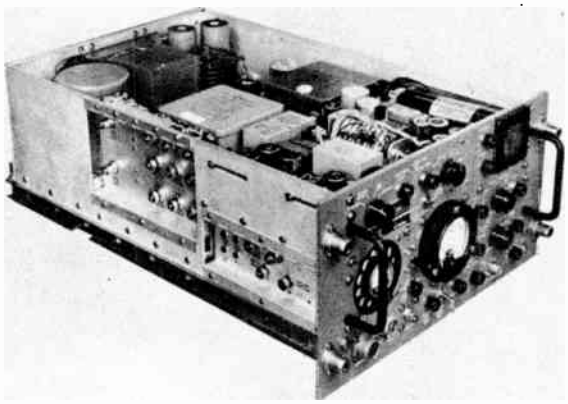


Fig. 8
Low level radio modulator.

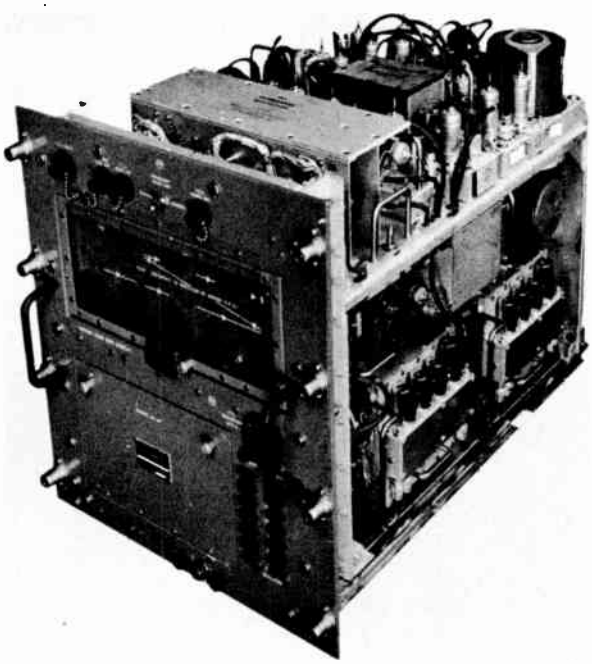


Fig. 9
Radio frequency osoillator.

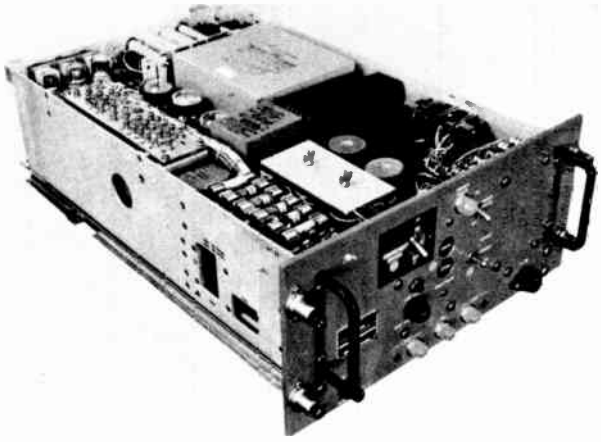


Fig. 10
Low voltage power supply.

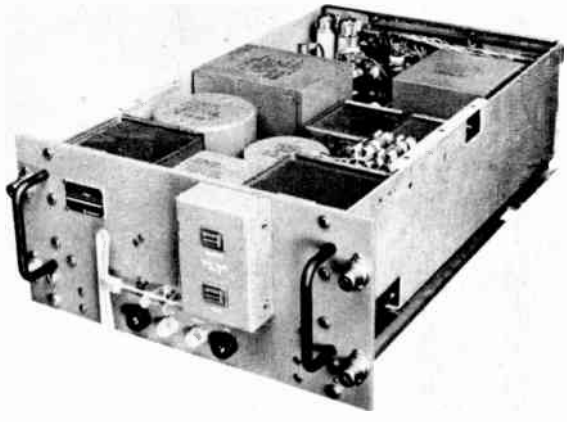


Fig. 11
Medium voltage power supply.

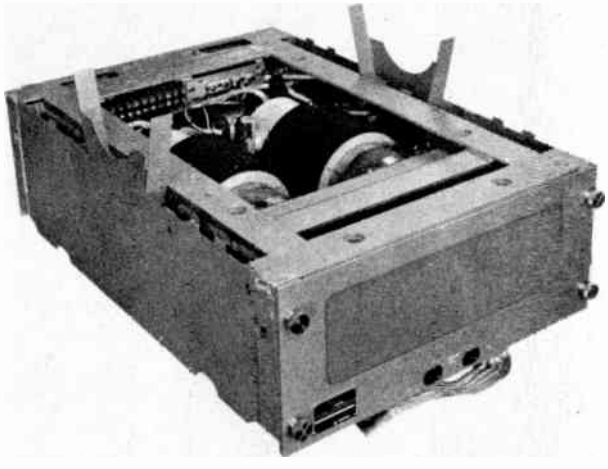


Fig. 13
Base mount.

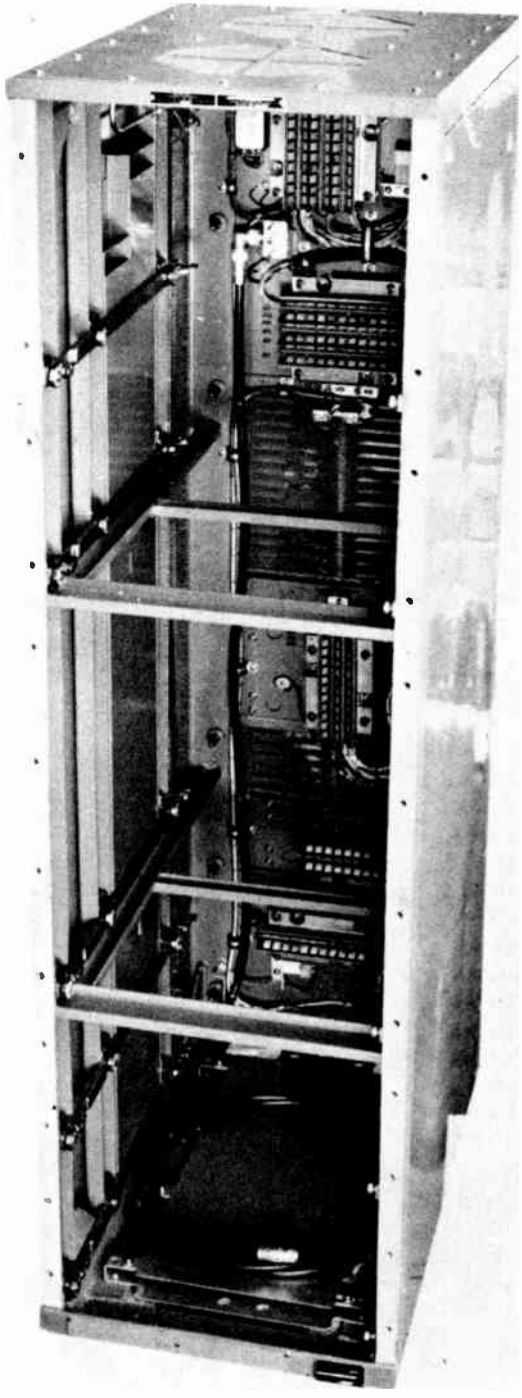


Fig. 12
100 watt transmitter frame.

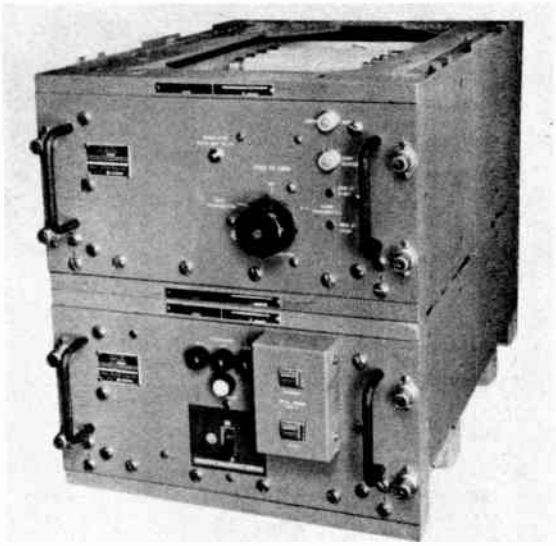


Fig. 14
Booster assembly.

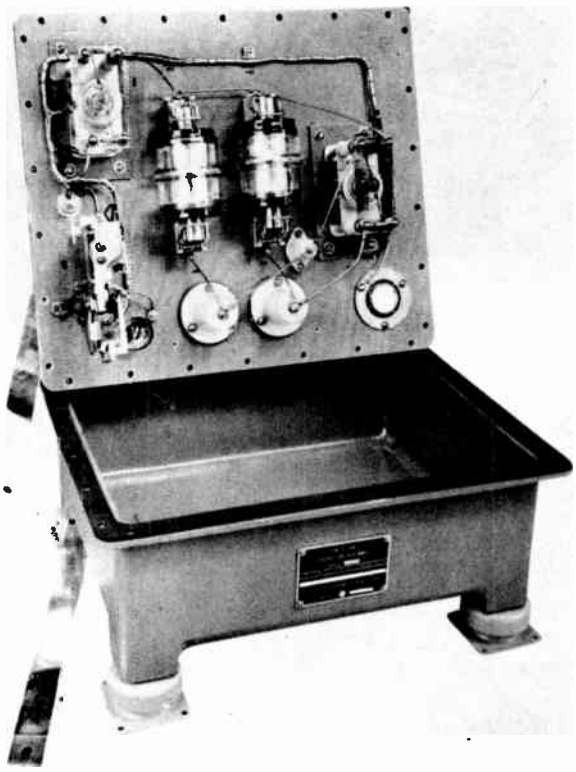


Fig. 16
Capacitor assembly (interior).

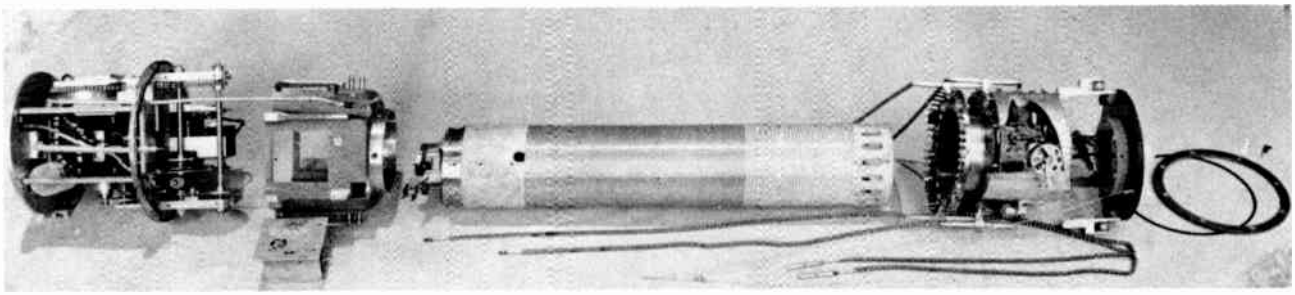


Fig. 15
Antenna tuner unit (exploded view).

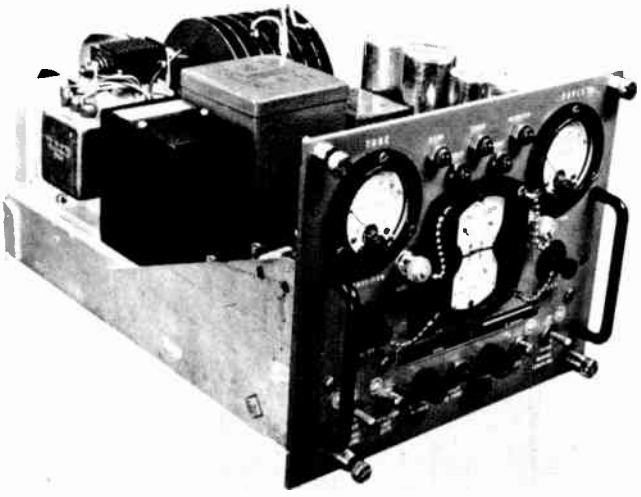


Fig. 17
ACG control and indicator unit.

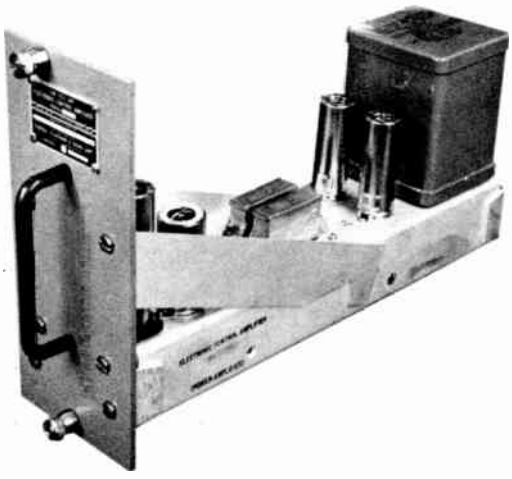


Fig. 19
ACG power amplifier unit.

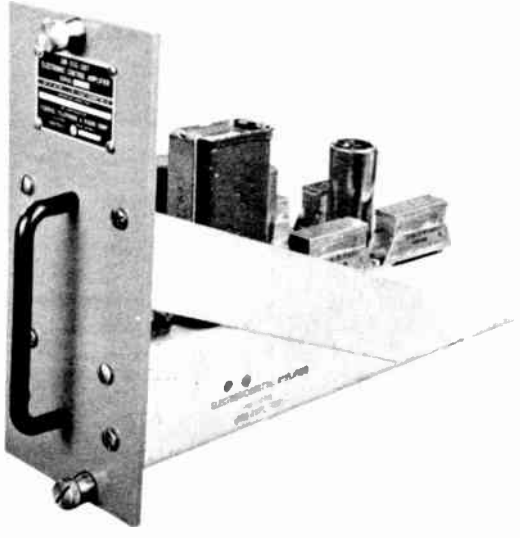


Fig. 18
ACG pre-amplifier unit.

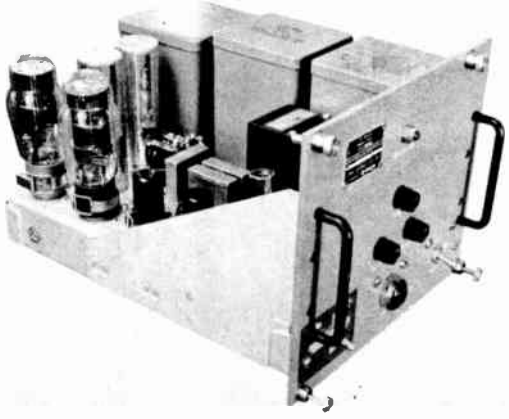


Fig. 20
ACG power supply unit.

Clifford D. May, Jr.
Communication Specialist
Office of the Chief Signal Officer
Washington 25, D. C.

The purpose of this paper is to discuss the practical application of known multiplexing techniques to derive additional communications channels on long range single sideband radio circuits.

For a number of years the United States Army Signal Corps has employed a variety of techniques to make maximum use of the high frequency assignments allocated for fixed military point to point communications circuits. The brief history below will trace the progress in utilizing single sideband high frequency radio circuits to their maximum capabilities.

The Signal Corps operates a world wide communications network connecting major overseas commands with the Department of the Army Headquarters in Washington, D. C. At the beginning of World War II the primary means of HF radio communications used on this network was C.W. telegraphy, both manual and Boehme. In the early stages of the war, it became obvious that a system employing C.W. telegraphy would not be capable of handling the tremendous volume of wartime message traffic due primarily to the requirement for highly trained operators. It was decided to investigate semi-automatic systems of telegraph transmission. After an investigation of existing systems, radio teletypewriter was accepted as the system best suited to meet Army requirements.

Two types of radio teletypewriter systems are now in general use on high frequency point to point radio circuits, (1) frequency shift, using single channel terminal equipment and time division multiplex terminal equipment and, (2) single sideband, using frequency division multiplex telegraph terminal equipment.

The single sideband equipment employed is of the type manufactured by the Western Electric Company and used by the American Telephone and Telegraph Company on their overseas radio telephone circuits. This equipment is designed to provide two (2) telephone channels each being approximately six (6) kilocycles wide. In order to provide the necessary telegraph multiplexing equipment for this single sideband system, the Bell Telephone Laboratories designed a six channel frequency division multiplex terminal which was later procured for military use.

A block diagram of this terminal is shown in Figure 1. The twelve (12) oscillators are arranged in pairs so that adjacent frequencies spaced 170 cycles apart provide the marking (current condition on the teletypewriter loop) and spacing (no current condition on the teletypewriter loop) tones for any given channel. These tones are then keyed in accordance with the DC telegraph intelligence on the teletypewriter loop (Ch 1), and, after passing through filters, are combined on a common bus for transmission over the single sideband system. The band occupied by the combined group of tones is 425-2295 cycles. This group of tones is transmitted over the sideband of the SSB system which has been designated in Figure 1 as SIDEBAND A. Less than one half of the capability of SIDEBAND A is utilized for the transmission of these tones. After receiving the group of telegraph tones from the single sideband receiver, they are separated by filters into appropriate channels and then passed through limiters, filters and rectifiers which restore the keyed tones to their original D. C. telegraph form, less, of course, the degradation resulting from transmission. It can be seen from Figure 1 that SIDEBAND B is used to provide a 2500 cycle telephone channel required to maintain continuity on the six telegraph channels.

The radio telegraph channels provided by the arrangement shown in Figure 1 did not prove to be too reliable during unstable radio conditions. Loss of individual channel frequencies caused by selective fading would cause many errors on the telegraph circuits.

A system was then devised to operate channels in parallel so that the same intelligence could be transmitted simultaneously on two different channel frequencies. This tone diversity arrangement provided a tremendous increase in reliability, although it reduced the capacity from six to three teletypewriter channels.

In order to provide tone diversity with its increased reliability without the loss of three channels, the system shown in Figure 2 was developed by Bell Telephone Laboratories. With this arrangement, the original six channel group of tones is shifted in frequency by the use of a 5270 cycle oscillator and modulator, thus creating a group of frequency diversity tones. This shifted group of tones is then combined with the original set and transmitted over SIDEBAND A. A band of 425-4845 cycles is occupied by the combined group of tones. At the receiving end, the 425-4845 cycle group is passed through filters which separate the normal tones from the diversity tones. The diversity tones are then demodulated to a 765-2635 cycle group. As can be seen, the diversity tones for any given channel are 340 cycles higher in frequency than the normal tones. This frequency separation is required to permit the use of a common channel limiter. After being separated into normal and diversity groups, the tones are separated into channels by filters, then are passed through limiters, filters, and rectifiers in the same manner as Figure 1. The tone diversity system shown in Figure 2 provides reasonable protection from selective fading, while still retaining a six channel capability in the terminal equipment.

As noted previously a telephone channel was operated over the B SIDEBAND of the single sideband system. The use of this channel was limited to that as an engineering order wire, as no requirement existed in the Department of the Army to operate radio telephone circuits over these single sideband facilities. In the latter part of World War II, requirements for facsimile transmissions over the world wide Signal Corps

communication system increased. When the single sideband circuits were used for these transmissions, it meant the sacrifice of the engineering order wire. Once again, frequency division multiplexing techniques were employed to provide the facsimile channel without the loss of the engineering order wire. As can be seen in Figure 3, a modulator associated with a 5300 cycle converting oscillator is employed to shift the order wire channel to a band above the facsimile channel. At the receiving end, a demodulator associated with a 5300 cycle oscillator restores the order wire channel to its normal frequency band.

In Figure 3 it can be seen that an air channel approximately ten kilocycles wide is required for the transmission of a single sideband signal carrying six telegraph channels, a facsimile channel and a voice order wire. Although it appears that the capability of the single sideband system is not fully used, no use can be made of the additional capacity due to the limitation of a ten kilocycle channel on the frequency assignments allocated for Army single sideband circuits. By the end of World War II, the single sideband circuits operated by the Department of the Army had apparently reached the saturation point as far as deriving any additional communication channels.

During the past few years the world situation has once again caused increased military communications requirements on overseas radio circuits. On some existing circuits the traffic requirements have overtaxed the capabilities of the system. An attempt to establish additional circuits to handle this increased traffic load would be extremely difficult due to the scarcity of frequencies in the high frequency spectrum. With this in mind, efforts were made to devise a system which would permit additional traffic to be handled on existing circuits.

By again noting Figure 3, it can be seen that 50% of the Sideband A is occupied by the diversity tones which carry the same intelligence as the normal tones. It was reasoned that if a space diversity system could be utilized, it would not be necessary to transmit the group of tones required for frequency diversity, thus releasing this band for transmission of additional channels. The problem was further complicated by the decision that it would be necessary to utilize the available existing type terminal equipment with a minimum of modification to accomplish the task.

First, the feasibility of space diversity with this type of terminal equipment was tested with favorable results. Only the normal tones were transmitted over the single sideband system. At the receiving end an additional single sideband receiver was installed and operated with an antenna properly spaced from the normal antenna for effective diversity. A double conversion process was then used to shift the group received on the diversity receiver so that the tones would pass through the diversity filters. This arrangement proved to be superior to the tone diversity system formerly employed.

After having eliminated the requirement to transmit the frequency diversity tones on SIDEBAND A, the tones from a second set of six channel terminal equipment could be transmitted in the band formerly occupied by the diversity tones. This is accomplished with minor modifications to the terminal equipment as shown in Figure 4. The modulator and 5270 cycle oscillator, formerly used to derive the diversity tones, are used to shift the tones from a second six channel terminal into the band formerly occupied by the diversity tones. At the receiving end, the normal tones from one terminal are separate from the normal tones from the second terminal by filters. The tones from the second terminal are restored to their original frequencies by a demodulator and 5270 cycle oscillator. These two groups are then passed through the normal filters of the two six channel terminals. The two groups of tones from the diversity receiver are separated by filters and are then shifted in frequency to permit them to pass through the diversity filters of the two six channel terminals. Thus twelve telegraph channels are transmitted in the band formerly required for transmission of six telegraph channels.

It can be seen that by the application of space diversity and frequency multiplexing techniques to single sideband circuits, the Signal Corps is able to utilize the maximum potential traffic handling capabilities of these circuits. Circuits of this type are in operation today and are contributing reliable communications to our armed forces overseas.

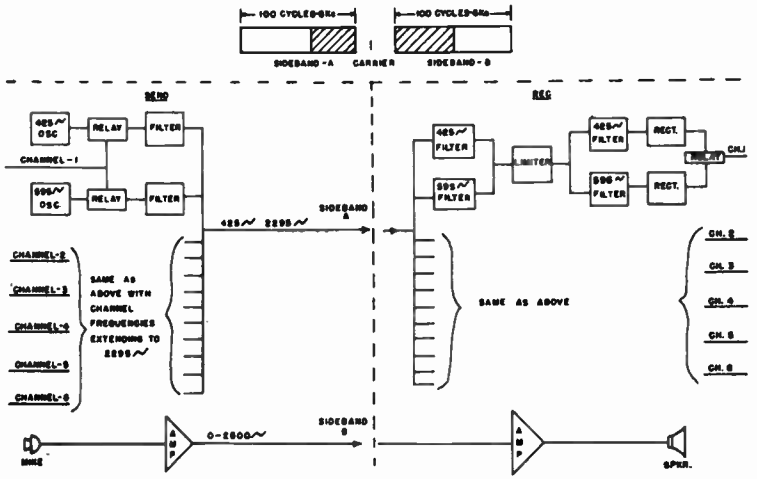


Fig. 1

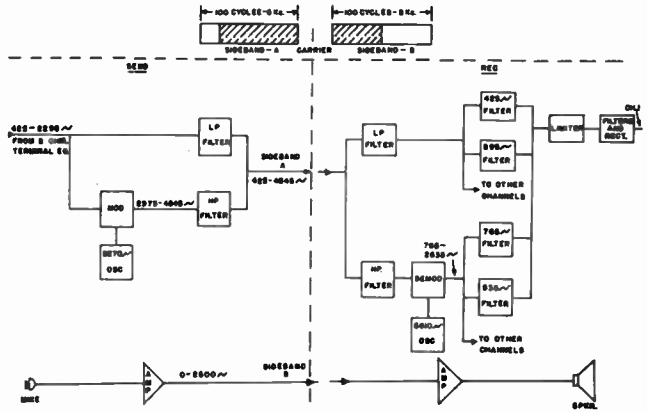


Fig. 2

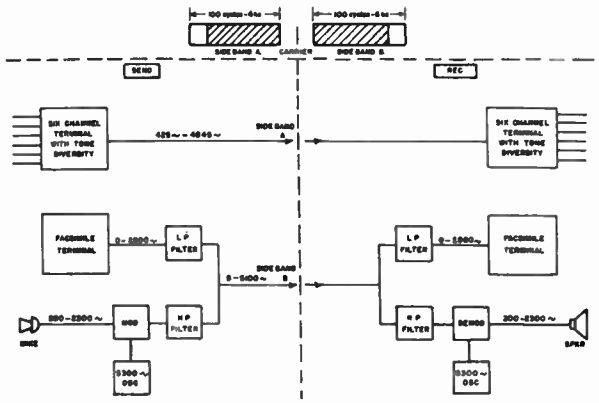


Fig. 3

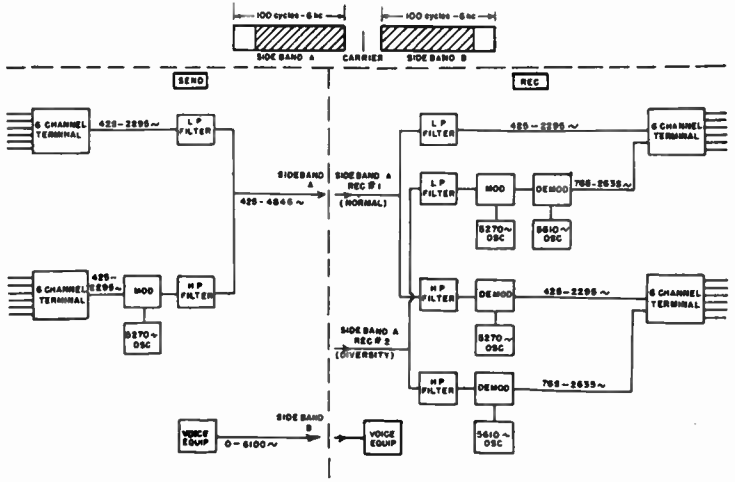


Fig. 4

PERFORMANCE OF SPACE AND FREQUENCY DIVERSITY RECEIVING SYSTEMS

by

R. E. Lacy and M. Acker
Signal Corps Engineering Laboratories
Red Bank, New Jersey

and

J. L. Glaser
Formerly with Washington University, St. Louis, Mo.
Now with Bell Telephone Laboratories

INTRODUCTION

Diversity systems represent a method for improving radio communications under fading conditions. Communications by way of ionospheric paths frequently become degraded because signal components at the location of the receiving antenna combine in an out-of-phase manner. The composite signal can differ instantaneously at different locations. Space diversity is a means whereby two or more antennas are displaced from each other in an effort to obtain the strongest signal at any instant from the available elements. Similarly, the signal combining effect can also differ instantaneously at different frequencies because of frequency selective fading. Another form of diversity is therefore frequency diversity, wherein a displacement in frequency is utilized by operating on two or more radio carrier frequencies to obtain the strongest signal from these available carriers.

While it has been known that signal improvement results from the use of diversity techniques, little quantitative data have been available. Some quantitative information was published in the Proceedings of the IRE in 1951[1] based on preliminary space diversity data obtained over one circuit early in this investigation.

This report presents additional quantitative information on the performance of space and frequency diversity systems. The data were obtained during an investigation of diversity systems conducted jointly by the Signal Corps Engineering Laboratories and Washington University, St. Louis, Mo., for a period of over four years. Three experimental circuits and six different radio frequencies were used. The circuits were from Red Bank, N. J. to Weldon Spring, Mo., a distance of 850 miles, on frequencies of 6.985, 11.66 and 15.72 megacycles; from Minneapolis, Minn., to Weldon Spring, Mo., 450 miles distance on 2.63, 4.25, and 6.985 megacycles; and from the Panama Canal Zone, to Weldon Spring, Mo., 2200 miles in distance on 11.66, 15.72 and 22.0 megacycles.

DEFINITIONS

At this point, it is also desirable to define several terms that are used throughout this report:

First, median level is that level which the received radio signal exceeds for one-half the operating time.

Second, least usable level is the lowest signal level that can be received intelligibly by the equipment, and is determined by the signal to noise ratio requirement of the system. For a certain expected received noise, a least usable level is therefore established.

Third, percent unusable time is that portion of the operating time during which the received radio signal falls below the least usable level. The percent unusable time can be expressed as a function of the ratio of the median signal amplitude to the least usable level. This is a conveniently used relationship[2] and can be obtained by determining the percent of time that the signal exists below discrete amplitude levels.

The advantages of diversity reception arise from the favorable statistical behavior of short-period fading of signals. It is this short-period fading that is treated here. Such fading can be thought of as being superimposed on long period fading of a diurnal or seasonal nature. The latter type of fading can be treated in an established manner[2] as will be mentioned later.

GENERAL RESULTS

Certain definite patterns have evolved from this investigation. It has been determined that the most practical spacing between individual antennas of space diversity receiving systems is approximately 600 feet. Little additional improvement resulted from increasing the spacing beyond this distance, although spacings up to 2400 feet were checked. Considering the usual unavailability and value of real estate as well as the maintenance of antennas and antenna ground areas, it is economical to keep the space to a minimum consistent with desired communication performance.

Antenna separations of less than 600 feet exhibited reduced diversity improvement. When the spacing between individual antennas was made 400 feet or less, antennas spaced in a direction parallel to the direction of propagation provided

greater diversity improvement than perpendicular orientation. At 600 feet, this difference in performance became small, though parallel orientation was still preferable.

PERFORMANCE OF SPACE DIVERSITY SYSTEMS

Methods for expressing diversity performance quantitatively were discussed in some detail in the 1951 papers[1] and will be reviewed briefly here.

An example of the method used to express diversity performance is shown in Figure I. The diversity curves show the percentage of the time that neither antenna provided a signal above the least usable level for various ratios of the non-diversity median-signal-amplitude to the least usable level. The reference curve in this figure shows similar information for a single non-diversity receiving channel during the time of the diversity measurements. At the frequency of 6.985 megacycles, the spaced-antenna arrangement represented here was found to give considerably less unusable time than the non-diversity reference antenna. As an example, for a ratio of non-diversity median signal amplitude to least usable level of 15 db, the unusable time was reduced from 2.2% to 0.045% by means of this diversity arrangement.

For a particular percentage of unusable time, the diversity system allows a reduction in transmitter power relative to the power required without diversity. For example, for one percent unusable time, this "diversity gain" amounted to about 8 db at 6.985 megacycles. It must be noted, however, that "diversity gain" has no meaning unless the percentage of unusable time or the ratio of median non-diversity signal amplitude to least usable level is specified.

Also of interest is the good diversity performance obtained at the lowest frequency tested. At 6.985 megacycles, the separation of 600 feet corresponds to about 4 wave-lengths. At 15.72 megacycles, the same separation amounts to about $9\frac{1}{2}$ wave-lengths, more than twice the former wavelength value. In spite of this difference, very little diversity improvement was obtained at 15.72 megacycles.

The data shown in Figure I were obtained with signals transmitted from Red Bank, N. J. and received at Weldon Spring, Mo. The frequencies of 11.66 and 15.72 megacycles were considered good usable daytime frequencies and 6.985 was a good usable nighttime frequency for the operating periods. The dual diversity receiving antennas were separated 600 feet in a direction parallel to the great circle route between Red Bank and Weldon Spring. All antennas used were half-wave horizontal doublets.

A similar variation in diversity performance with frequency is seen in Figure II, which shows the results of measurements at Weldon Spring, Mo., of signals transmitted from Minneapolis, Minn. In these tests, poor diversity performance was obtained at 6.985 megacycles. This frequency represents a relatively high one for the circuit. Better performance was obtained at 4.25 and 2.63 megacycles.

Similar results were obtained on the 2200 mile path from Panama to Weldon Spring, as shown in Figure III. Diversity performance at 22.0 megacycles was quite poor, with performance improving at 15.72 and 11.66 megacycles respectively. It should be noted that the diversity antennas were spaced perpendicularly during the Minneapolis and Panama tests. This occurred because the superiority of parallel orientation had not yet been established and it was not possible to rerun the tests thereafter.

Experiments with triple-spaced antenna-diversity were also included in the test program. Some of the results of these measurements are shown in Figure IV, for the Red Bank-Weldon Spring circuit. The three antennas were in a line parallel to the direction of propagation, with adjacent antennas separated 300 feet.

A noticeable improvement over dual diversity on the same circuit resulted from the installation of the third antenna. For example, at 6.985 megacycles, with the ratio of median non-diversity signal amplitude to least usable level approximately 8 db, the unusable time was reduced from 10% to about 0.01%. This value of unusable time would have been reduced more if the spacing between individual antennas was 600 instead of 300 feet. Diversity performance increases here as the frequency decreases from 15.72 to 6.985 megacycles.

Results of triple diversity tests made with signals transmitted from Minneapolis are shown in Figure V. The same inverse relationship between diversity performance improvement and frequency exists for the frequencies checked.

Results of triple diversity measurements with signals originating in Panama are shown in Figure VI. The same general trends are exhibited here.

Assuming independent fading on each of the diversity antennas, it is possible statistically to predict diversity performance. It is interesting to note that actual diversity improvement obtained at the lowest frequencies used in the Red Bank-Weldon Spring and Minneapolis-Weldon Spring tests was approximately equal to the values predicted statistically. These frequencies were 6.985 and 2.63 megacycles, respectively. Performance at the higher frequencies was below the statistically predicted values for independent fading.

PERFORMANCE OF FREQUENCY DIVERSITY SYSTEMS

In addition to the space diversity phase of the investigation, the performance of frequency diversity systems was also determined. This was accomplished by transmitting two signals with frequency separations ranging from 170 to 1870 cycles per second. At the receiving site they were analyzed after being separately detected and combined. The frequency separations corresponded to channel separations commonly used in multi-channel radio telegraph equipment. Presumably, the fading of the two signals would be

statistically independent with sufficient separation. However, the performance obtained with separations up to 1870 cycles per second was considerably worse than that obtainable with independent fading.

Results of frequency diversity tests are shown in Figures VII and VIII. Figure VII shows the results obtained at a nominal operating frequency of 6.985 megacycles on the Red Bank to Weldon Spring path. The behavior is shown for frequency separations of 510, 850, 1190 and 1870 cycles per second. Although the performance improved slightly as the frequency separations were increased, the overall performance was considerably inferior to that obtained with a 600 foot spaced dual diversity system at the same frequency.

Figure VIII shows the frequency diversity performance measured at 11.66 megacycles over the same path. Here again the difference in performance with respect to frequency separations was small. The overall improvement was approximately the same as at 6.985 megacycles.

GENERAL INFORMATION

A typical problem in the engineering of a communication link is that of determining the transmitter power required to maintain communication in the presence of atmospheric noise. Methods have been developed by the Bureau of Standards for estimating the received median signal amplitude for certain predictable propagation conditions[2]. This same source provides means for estimating required median signal amplitude, based on the world-wide distribution of noise at various frequencies and periods. Thus, a specified system reliability can be predicted, given various combinations of transmitter power, transmitting antenna gain, discrimination gain of receiving antenna, and characteristics of the diversity system.

SUMMARY

In summarizing, the following are the major points of this report, based upon experimental data obtained in the investigation:

a. The performance improvement of space diversity systems increased with separation between antennas up to about 600 feet. Little or no additional improvement resulted from increasing the spacing beyond this distance, though spacings up to 2400 feet were checked.

b. Spacing of diversity antennas parallel to the direction of propagation resulted in greater performance improvement than with perpendicular orientation when separations between individual antennas were less than 400 feet. No appreciable differences were observed with greater spacings, although parallel orientation was preferable.

c. The performance improvement of space diversity systems exhibited an inverse relationship with frequency for any particular circuit. Diversity performance at lower frequencies was found to be approximately equal to values predicted statistically based on the assumption of independent fading for each of the diversity antennas. Improvement at the higher frequencies was substantially less than this.

d. For a fixed configuration, the degree of performance improvement exhibited by diversity systems is a function of the ratio of median signal level to the least usable level at the receiving location.

e. The performance improvement of a dual frequency diversity system does not vary much in the range of frequency spreads from 170 to 1870 cycles per second. The lower frequency spreads were slightly less effective than the higher.

f. The performance improvement of the frequency diversity system remained fairly constant over the range of the two nominal frequencies (6.985 and 11.66 megacycles) tested.

g. The best performance improvement obtained with frequency diversity systems at 6.985 megacycles was definitely inferior to that obtained on the same frequency with a dual space diversity system having a 600 foot separation between antennas. Frequency diversity improvement at 11.66 megacycles was only slightly less than that obtained with a 600 foot spaced dual diversity system at the same frequency.

ACKNOWLEDGMENT

In closing, acknowledgment must be made to all who contributed in making this investigation successful. The number of contributors was so numerous that credit is due entire groups of testing, operating and computing personnel, both at Coles Signal Laboratory, Fort Monmouth, N. J. under the direction of Mr. Harold F. Meyer and at Washington University, St. Louis, Missouri, under the direction of Professor S. H. Van Wambeck.

REFERENCES

1. Glaser & Van Wambeck; Van Wambeck & Ross, Proc. IRE March 1951.

2. National Bureau of Standards Circular #462 "Ionospheric Radio Propagation" Dept. of the Army Technical Manual TM 11-499 and others.

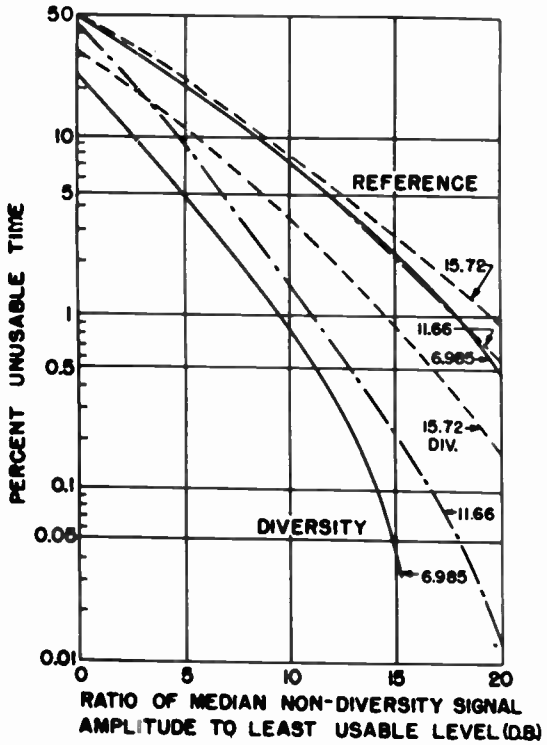


Fig. 1
Performance of dual diversity system, 850 mi. E-W circuit, Red Bank, N.J. to Weldon Spring, Mo. Antennas spaced 600 feet; parallel orientation.

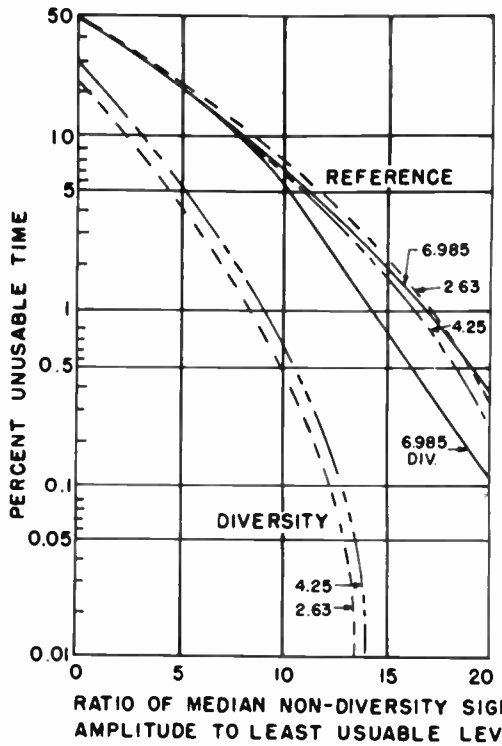


Fig. 2
Performance of dual diversity system, 450 mi. N-S circuit, Minneapolis, Minn. to Weldon Spring, Mo. Antennas spaced 600 feet; perpendicular orientation.

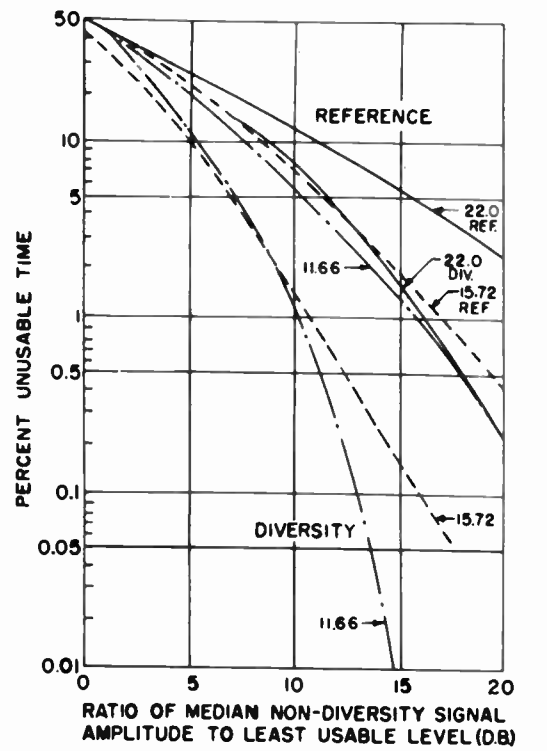


Fig. 3
Performance of dual diversity system, 2200 mi. S-N circuit, Panama to Weldon Spring, Mo. Antennas spaced 600 feet; perpendicular orientation.

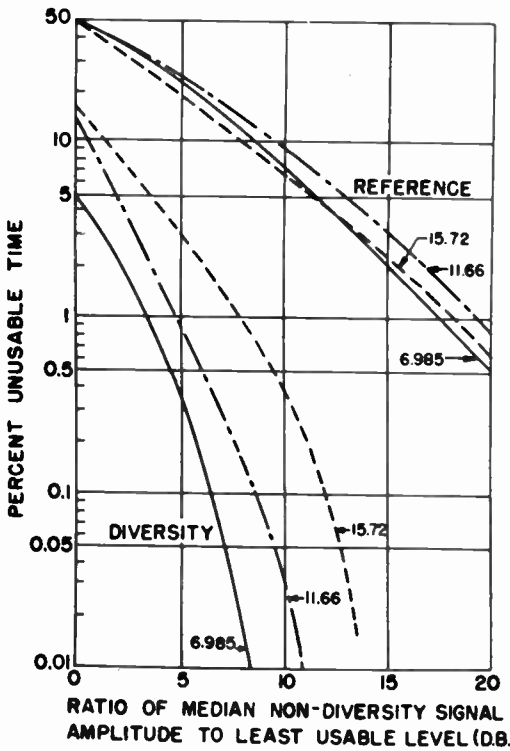


Fig. 4
Performance of triple diversity system, 850 mi. E-W circuit, Red Bank, N.J. to Weldon Spring, Mo. Adjacent antennas spaced 300 feet; perpendicular orientation.

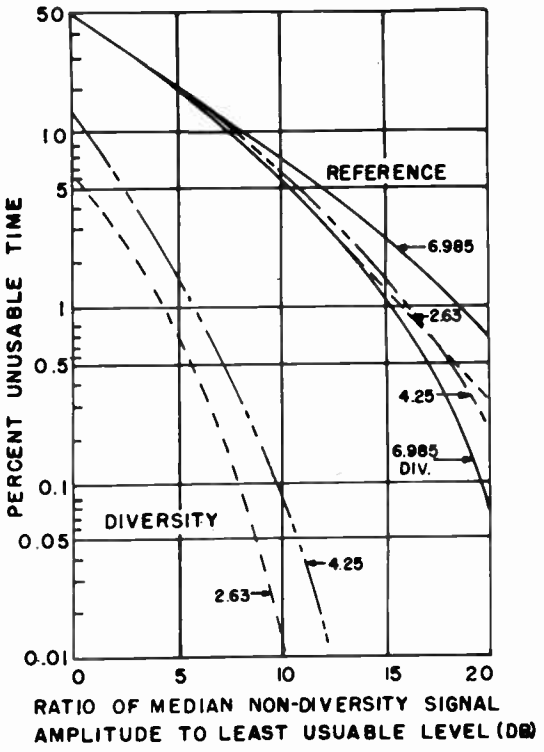


Fig. 5
Performance of triple diversity system, 450 mi. N-S circuit, Minneapolis, Minn. to Weldon Spring, Mo. Adjacent antennas spaced 300 feet; perpendicular orientation.

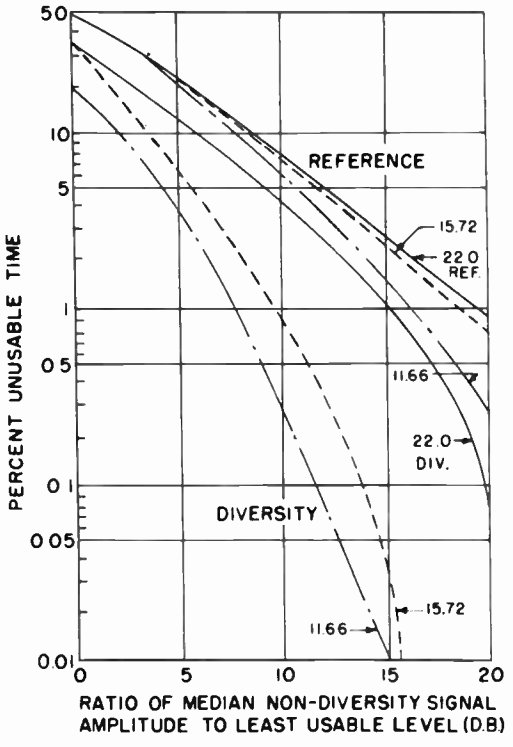


Fig. 6
Performance of triple diversity system, 2200 mi. S-N circuit, Panama to Weldon Spring, Mo. Adjacent antennas spaced 300 feet; perpendicular orientation.

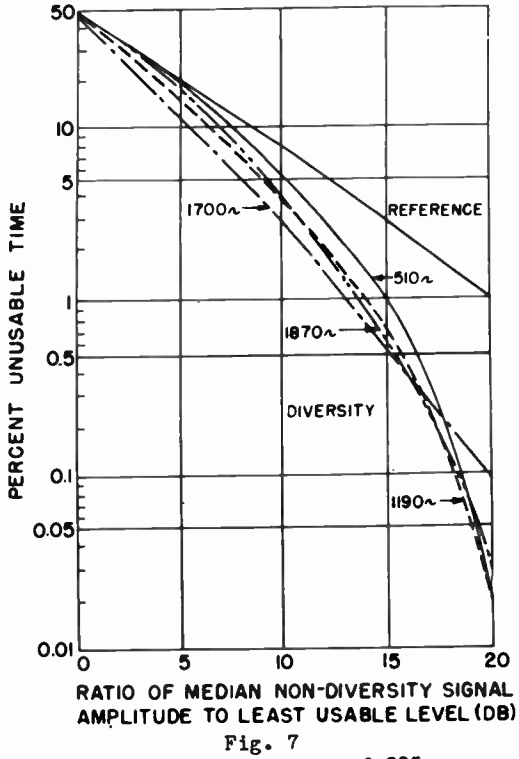


Fig. 7
Frequency diversity performance 6.985 mc over 850 mi. E-W path, Red Bank, N.J. to Weldon Spring, Mo.

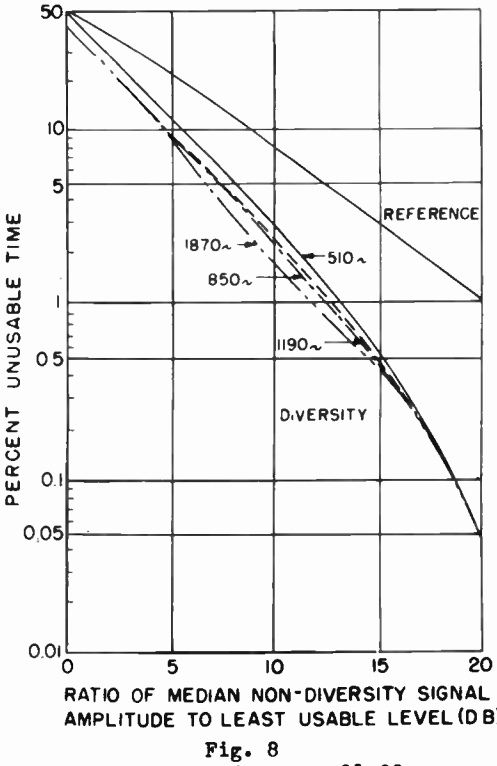


Fig. 8
Frequency diversity performance 11.66 mc over 850 mi. E-W path, Red Bank, N.J. to Weldon Spring, Mo.

EFFECT OF HITS IN TELEPHOTOGRAPHY

P. Mertz and K. W. Pfleger
Bell Telephone Laboratories, Inc.
New York, New York

1. Introduction

On many long communication links there are brief and sudden attenuation variations, which may come from a large variety of causes. These disturbances, known as "hits", are frequently barely perceptible and quickly forgotten if they occur in telephony. In telephotography they may be permanently recorded as objectionable marks in received pictures unless sufficiently low in hit intensity and hit duration. The present work represents a tentative evaluation on how objectionable are hits of various intensities and durations on individual pictures in telephotograph transmission. The technique employed is adapted from one used in television. It consists in having observers rate hit-affected pictures according to pre-worded comments. The ratings are then correlated with the measured intensities and durations of the hits.

2. Results and Conclusions

A tentative boundary between acceptable and unacceptable hits, on individual pictures, is presented below as curve (C) of Fig. 12. This boundary is plotted as a tentative design tolerance on the hit intensity over a gamut of hit durations. The intensity, expressed in db, is the increase in circuit loss during the hit over normal loss. The hit duration is measured in picture elements, or half cycles of the maximum effective modulation frequency (or "keying frequency"). For the particular machine used, in which the picture element was 1/100 inch, the hit duration is translated into inches at the top of the plot. Again for the particular machine, the picture element was 1/2 millisecond. Thus, the hit duration in milliseconds is half the figure indicated in picture elements at the bottom of the plot.

For a variety of obvious reasons there is scattering of the data. This is sufficient to justify not presenting the boundary as anything more complicated than the simple straight line shown on the logarithmic plot. Some analysis is made of this scattering at the close of the paper.

The details of some 20 per cent of the hits which constituted the extreme cases of scattering were examined individually. Half of these had received what appeared as unusually lenient ratings compared with the general group, and half unusually critical ratings. It was found that in general:

(a) The hits rated rather leniently are in parts of the picture of little interest or where the background is mottled or dark so that hits are inconspicuous or mistaken for a part of the picture, or where the picture could be retouched easily.

(b) The hits rated rather severely are in parts of the picture of considerable interest or against light or plain background where they show conspicuously or are near other hits so that the effects of all were considered unconsciously.

(c) There existed differences among observers in the method of rating a hit. This accounts for some scattering of the data. For example, one technically inclined judge consciously took the picture content into account in rating a particular hit, because most people who are not technically inclined would do this unconsciously. Another technically-minded observer tried not to let his rating of a hit be influenced by the fact that the surrounding picture detail did or did not tend to obscure the hit.

Since the test pictures all contained a number of hits, the judges were also asked to give ratings for each picture as a whole based on the simultaneous effects of all the hits therein. It is interesting to note that in nearly every case, the worst rating assigned to the picture was about the same as for the worst single hit in the picture.

3. Experimental Procedure

Sixteen positive prints of received telephotographs containing hits were available, including half tones as well as black and white subject matter. The hits which were noticeable appeared as dark streaks in white or gray portions of the pictures; for example, see Fig. 1. It is first to be noted that the scanning lines in this picture run vertically (unlike television pictures) and therefore hits appear as vertical lines. There are two faint hits near each other at about the center of the sky, the right hit being slightly worse than the left one. Just above and to the left of the middle of the picture as a whole, there is a short hit about a half dozen picture elements long of maximum intensity corresponding to a momentary break in the circuit. Between the horizon and the top of the archway at the left edge of the picture is another such hit.

About two hundred hits having substantially uniform intensities throughout their durations were selected. Ten judges gave independent ratings of each hit according to the following comments:

1. Not perceptible.
2. Just perceptible.
3. Definitely perceptible but only slight impairment to picture.
4. Impairment to picture but not objectionable.
5. Somewhat objectionable.
6. Definitely objectionable.
7. Not usable.

The judges were selected from technical staff normally engaged in engineering or mathematical work. Somewhat over half had previous experience in rating telephotograph pictures. Two had substantial experience in discussing picture defects with customers of Bell System telephotograph networks.

The lengths of the hits were measured with a ruler. The intensity of each hit was measured by evaluating the photographic density in its area and that of the surrounding area with a calibrated density wedge. This was converted to decibels by the use of the typical signal to positive density characteristic curve shown in Fig. 2.

4. Plotting and Smoothing the Data

There were differences in comment numbers assigned to each hit due to variations in the abilities, experiences, and viewpoints of the observers. Hits of a given intensity and duration were rated differently depending on location in the picture. Further, the simple process used often made it difficult to measure the hit intensity closer than ± 2 db. Consequently, there resulted scattering of the data. This scattering is considered in more detail in part 5 below.

The median rating for each hit was found. Sometimes this number turned out fractional. This was inconvenient because of the method to be described for plotting the results. In such cases, an adjustment was made in the hit intensity to that which it was estimated would have received the nearest integer rating. The adjustment was determined by the relation between the hit magnitude and observer rating. The method for making this adjustment is next illustrated.

Every individual observation was plotted, separate sheets of paper being used for each rating number. The intensity of each hit in db was the ordinate and its duration or length was the abscissa, using logarithmic scales. A provisional summarizing straight line was drawn approximately through the midst of the data on each sheet. These straight lines were also re-drawn all on an eighth sheet which is reproduced in Fig. 3. These give a provisional value of the change in db on going from one rating number to another at a particular abscissa. Now suppose, for example, that a median rating number for a particular hit is 4 1/4 and that from the family of curves just drawn it is seen that hits of the same duration have the intensity 10 db for rating number 4 and 14 db for rating number 5. The adjustment to be made to the observed intensity of the hit in order to change its median rating number to the nearest integer, 4, is to lower its observed intensity by one db, which is 1/4 of the difference between 14 and 10 db.

After fractional median ratings had been eliminated by this method of adjusting the intensity, the data were re-plotted. Only the median observations were used. A separate sheet was

employed for each median rating number. These showed one point for each hit given that original or adjusted rating, and exhibited less scattering than the provisional plots. The abscissa indicated the duration or length of the hit, and the ordinate the original or adjusted intensity in db. Examples of some of these sheets are shown in the next three figures.

Fig. 4 shows such data for median comment No.2. Small circles indicate hits which occurred on a light background of density ranging from white to 5 db below white. Crosses show hits which occurred on a gray background ranging from 5 to 10 db below white. Dots represent hits against a darker background than 10 db below white. Close examination of the figure might show some slight systematic variation among the dots, crosses, and circles. There is, however, no decided trend indicated.

The scattering of the data on Fig.4 and on other Figures is due in part to the errors in estimating the intensity but mainly to the wide differences in rating by the observers. A summarizing straight line is drawn on this sheet through the midst of the data, but its exact course was determined after final smoothing of the data following examination of similar sheets dealing with the other comment numbers.

Fig. 5 was similarly obtained for comment No. 4, and Fig. 6 for comment No. 6. The final smoothing of the data was accomplished on Fig. 7 after making some cross-plots. The straight lines each indicate a smoothed relation between hit intensity and duration corresponding to a given median rating.

In addition to the plots of median observations, similar plots were prepared for the most severe rating of the ten observers (most critical 10 per cent observations) in each case.

Examples of these are shown in the following four figures. Fig. 8 gives the data for comment No. 2 whereon a straight line appears drawn nearly through the midst of the data. Its exact course was determined after final smoothing of the data following examination of similar sheets dealing with other comment numbers.

Fig. 9 was similarly obtained for comment No. 4, and Fig. 10 for comment No. 6. The final smoothing of the data was accomplished on Fig. 11 after making some cross-plots. The straight lines each indicate a smoothed relation between hit intensity and duration for the most critical 10 per cent observations. There is no curve on Fig. 11 for comment No. 1 ("not perceptible"), because the most critical 10 per cent of the observations did not include this rating.

In the previous use of this rating technique it has been the experience that comment No. 2 ("just perceptible") for the median observations comes out not far different from comment No. 4 ("impairment to picture but not objectionable") for the most critical 10 per cent observations.

Accordingly, it is of interest to compare the corresponding curves from Figs. 7 and 11 by replotting them on Fig. 12. Curve (A) shows the smoothed intensity vs. duration for the comment "just perceptible" for the median observations and curve (B) shows the same for the comment "impairment to picture but not objectionable" for the most critical 10 per cent observations. Due to the scattering of the basic data the agreement between curves (A) and (B) is only fair. Curve (C), substantially the average of curves (A) and (B), is based on judgement, and may be assumed as a suitable compromise to denote a possible demarcation objective between acceptable and unacceptable amounts of the impairment.

It appears from Fig. 12 that for short hits lasting 1 picture element (1/2 millisecond) the smoothed boundary between acceptable and non-acceptable intensity runs between 4 and 8 db net loss change, and is probably close to 6 db. For long hits lasting 1000 picture elements (1/2 second) the boundary runs between 2 and 4 db net loss change and is probably close to 3 db. When the hit intensity (in db) and hit duration are both plotted to logarithmic scales, the smoothed boundary (C) is assumed as a straight line between the extremely short and extremely long hit durations.

5. Analysis of Scattering

In the evaluation of such an indefinite phenomenon as a hit, it is to be expected that there will be a scattering of results. As already noted Figs. 4, 5, and 6 illustrate this. It is desirable to examine the scattering more carefully to obtain a more definite picture of the various contributions to it. There are four important sources:

1. Each of the individual observers has a systematic bias, as compared with the group, in the severity with which he rates the hits.

2. Each of the individual observers has a random component of inconsistency, from judgment to judgment.

3. The specific hit being examined in each case may come on the particular picture in which it occurs at a more or less vulnerable spot as mentioned previously.

4. In the process of smoothing the final data, great weight is attached to representing the general trend, and detailed deviations from that trend have been ignored.

The departures of each observer from the median judgment of the group form a set of distributions. Two of these are illustrated in Fig. 13. At A is that for the observer most consistent with the group, at B is that for the observer least consistent. The displacements are about the median judgment for the group.

The distribution of the bias values among the observers, about the median judgment, is shown in Fig. 14. Their standard deviation is 0.88 of a comment number, as illustrated by the dotted line in the figure. This evaluates the effect of the first source of scattering mentioned above.

Each distribution in Fig. 13, and for the other observers (not shown), has a standard deviation about its median, and a range between its extreme negative extent and its extreme positive. The distribution of each of these is shown respectively in Figs. 15 and 16. The median of the standard deviations is 0.79 of a comment number, and of the ranges is 4.75 comment numbers. The latter is six times the former. Each is shown by a dotted line in the figures. These represent an evaluation of the effect of the second source of scattering.

The result of the combination of the effects of sources one and two taken together comes out as the scattering in the data reported by all the observers. Each hit rated showed a distribution in this rating, with a standard deviation about the median rating. The distribution of these standard deviations is plotted in Fig. 17. Its median is indicated at 1.05 comment number.

The influence of sources three and four together may be ascertained from a study of the departures of the median judgments in each case, from the smoothed curves of Figs. 4, 5 and 6. It is difficult to separate out the effects of the two respective sources, because the entire sample of observations is not large. Thus not all types of incidence of the hit on the picture can be covered in any systematic fashion for a given hit intensity or hit duration.

The departures of the data from the final smoothed curves were evaluated by noting, for each point, the fractional comment number on which it stood in the array of smoothed curves of Fig. 7. The departure was taken as the difference between this figure and the median rating actually assigned to that hit.

Problems are, of course, posed by points which lie outside the range of all the smoothed curves of Fig. 7. These were evaluated as best they could be, by assuming extrapolations beyond both ends of the range of comment numbers. It is recognized that in some of the extreme cases, the evaluation is rather arbitrary.

The distributions of these departures were evaluated for the median ratings of comment numbers 2, 4 and 6. The distributions are plotted in Fig. 18. The standard deviations as indicated are, respectively, 1.67, 1.81, and 2.06 comment numbers.

These are larger figures than would of course be desirable. Considering, however, the indefiniteness of the phenomenon being studied they are not unreasonable. Further, the contributions to the irregularities from the various principal sources seem to be of the same orders of magnitude.

With the data available it is not possible to tell too precisely the course of what might be considered a "typical" smooth curve, for any one comment. It does not appear for example, as has already been stated, that one can justify any greater complexity in shape than a mere straight line, for the smoothed curves of Figs. 3 to 12. Straight lines at an angle, however, intersect. This would mean that on one side of the intersection a more intense hit would be interpreted as more acceptable than a less intense hit of the same duration. This would represent a contradiction, and therefore the straight lines have been so chosen that the point of intersection would lie at the end of or beyond the range of hit durations used in the measurements. This was one guiding principle, therefore, in selecting the array of smoothed curves of Figs. 7 and 11. It is because of this choice that some residual systematic bias exists in the distributions of Fig. 18. For example, it is next shown why there is appreciable negative bias for comment 4 in this figure:

Smoothed curves drawn through the middle of the data plotted on Figs. 4, 5 and 6 might have eliminated the systematic bias in Fig. 18, but such curves when combined as in Fig. 7 would have intersected within the range of hit durations which might occur. In Fig. 5 the data appear to decrease more rapidly in intensity with increasing duration than the smoothed curve shown. By drawing this smoothed curve somewhat less steep than the general trend of the data and in the particular location shown, and by making minor changes in the smoothed curves which were first drawn through the data on Figs. 4 and 6 until they became as shown, it was possible to bring the point of intersection, when they were combined in Fig. 7, outside the range of hit durations. Unfortunately the smoothed curve in Fig. 5 lies above a majority of the plotted data. This accounts for the negative bias when the points are measured with respect to the smoothed curve. For comments 2 and 6, as shown in Fig. 18, this bias becomes positive.

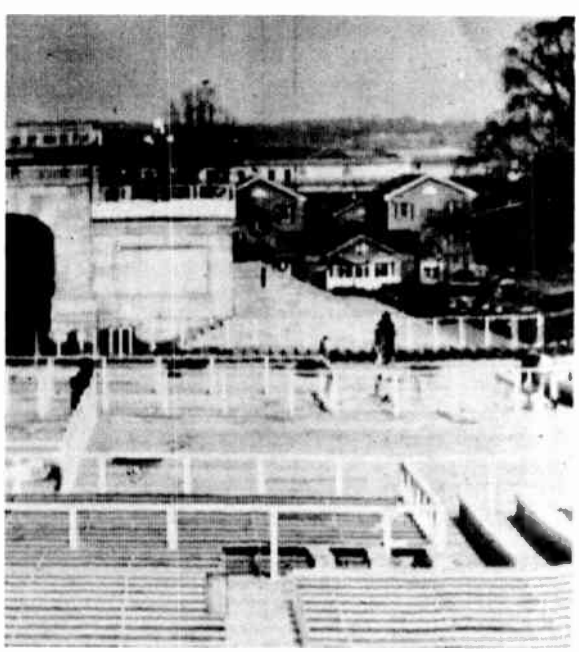


Fig. 1
Portion of telephotograph containing hits. (Original telephotograph contains 100 scanning lines per inch.)

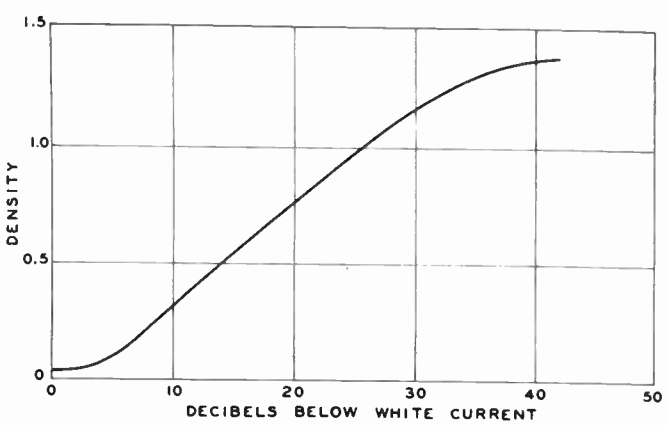


Fig. 2
Calibration curve of density wedge showing photographic density on positive print of telephotograph versus received current in decibels below maximum (white) current.

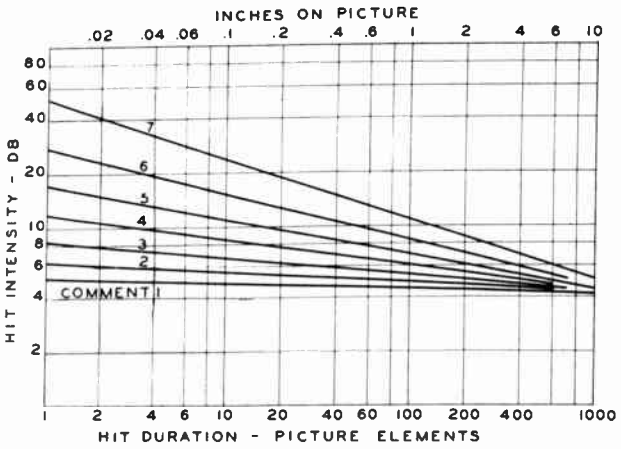


Fig. 3
Preliminary smoothed curves of original data relating intensity with duration of hit.

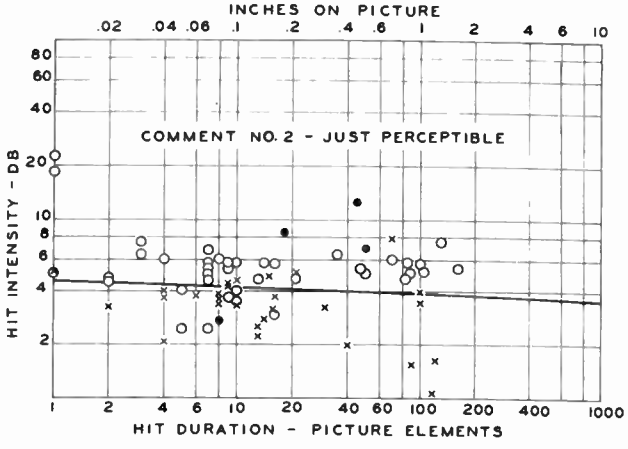


Fig. 4
Median observations relating intensity with duration of just perceptible hits.

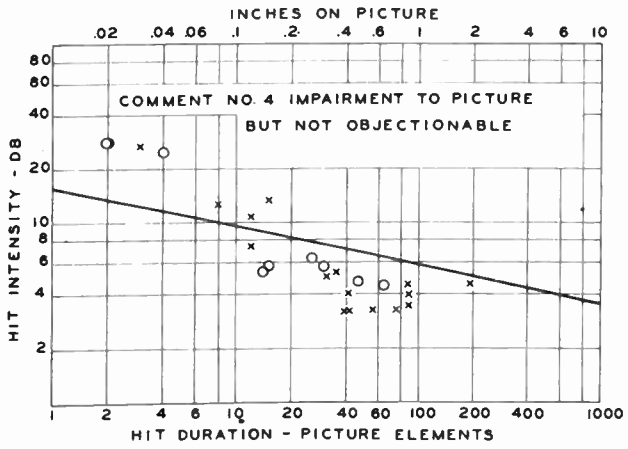


Fig. 5
Median observations relating intensity with duration of hits which are an impairment to the picture but not objectionable.

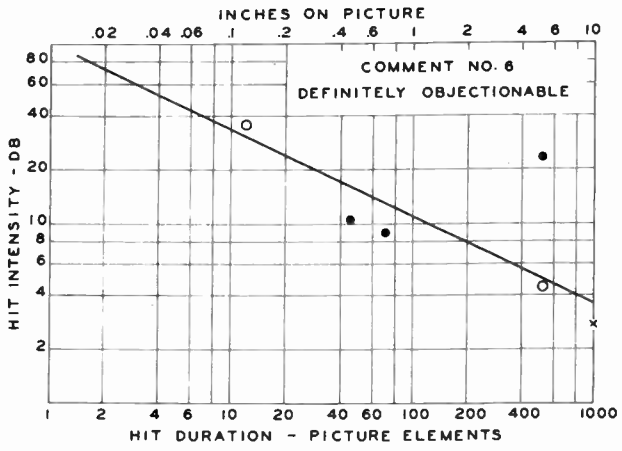


Fig. 6
Median observations relating intensity with duration of definitely objectionable hits.

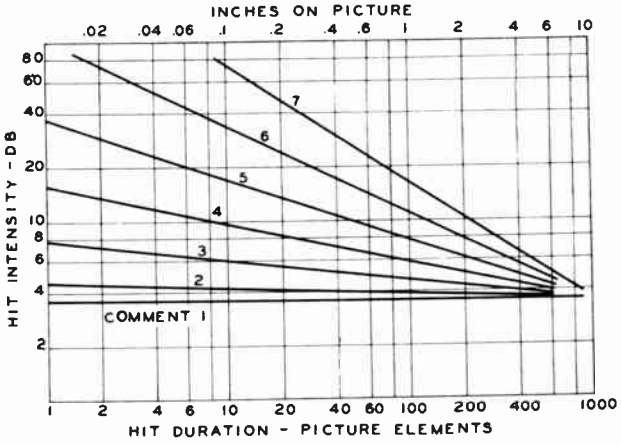


Fig. 7
Smoothed curves of median observations relating intensity with duration of hit for various comment numbers.

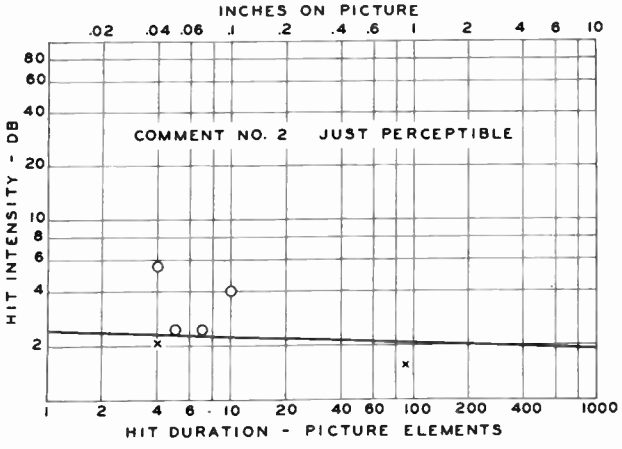


Fig. 8
Most critical observations relating intensity with duration of just perceptible hits.

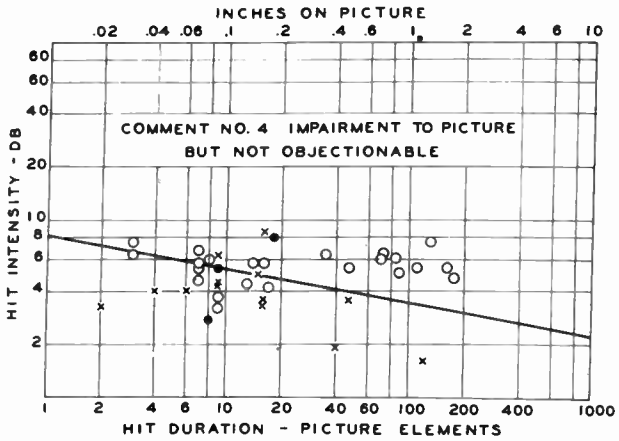


Fig. 9
Most critical observations relating intensity with duration of hits which are an impairment to the picture but not objectionable.

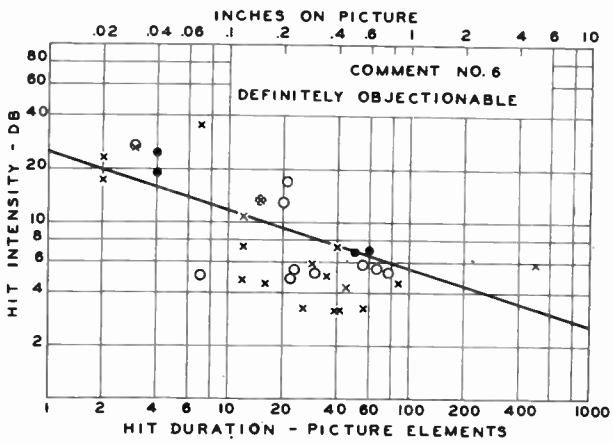


Fig. 10
Most critical observations relating intensity with duration of definitely objectionable hits.

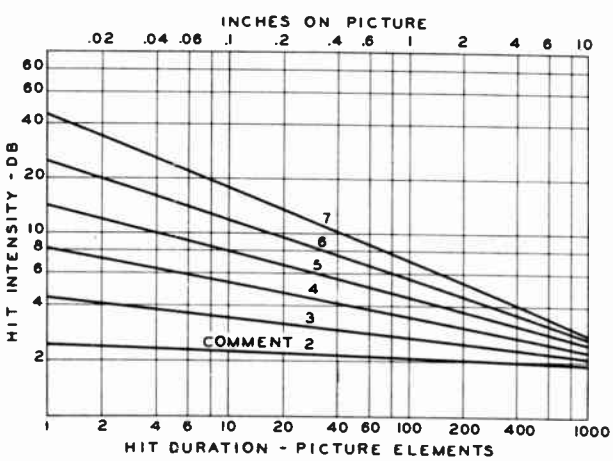


Fig. 11
Smoothed curves of most critical observations relating intensity with duration of hit for various comment numbers.

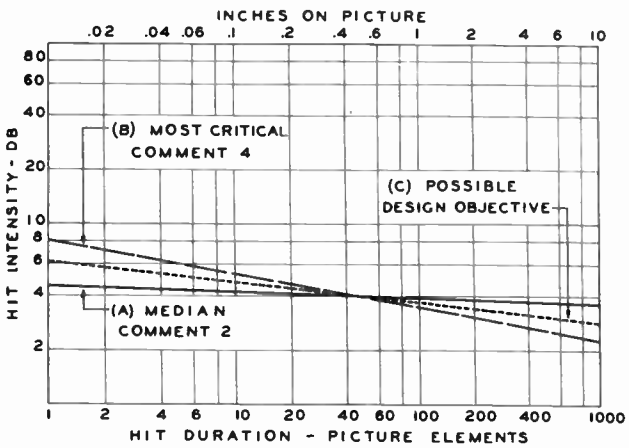


Fig. 12
Summary of results, intensity vs. duration of hits.

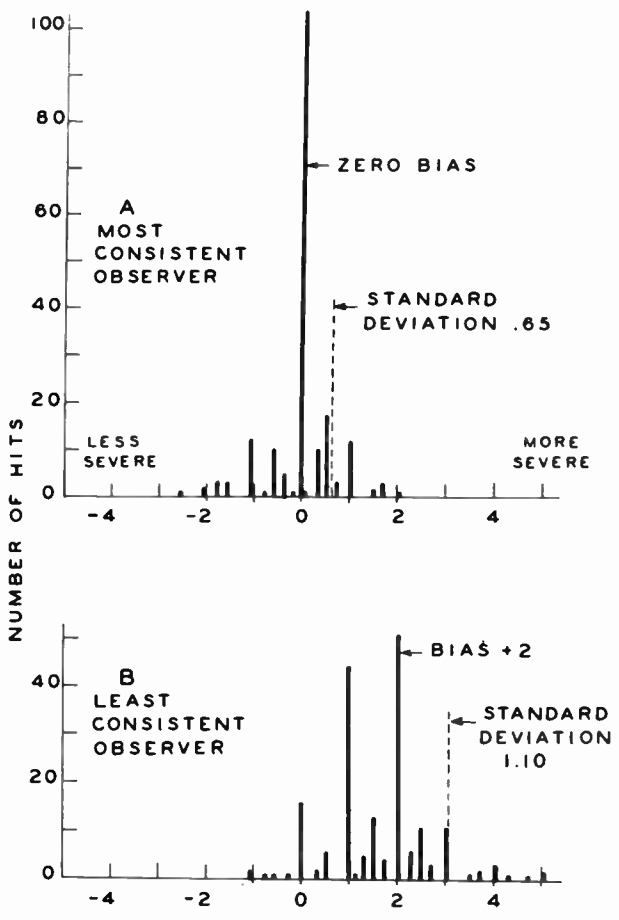


Fig. 13
Departures of individual observers from median judgment of the group.

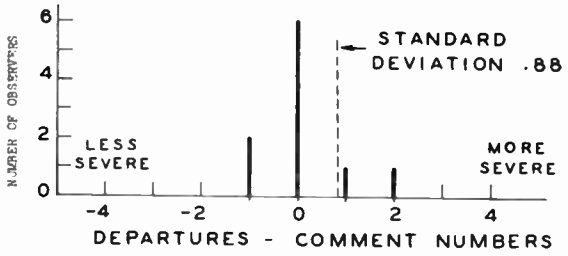


Fig. 14
Distribution of bias values among ten observers.

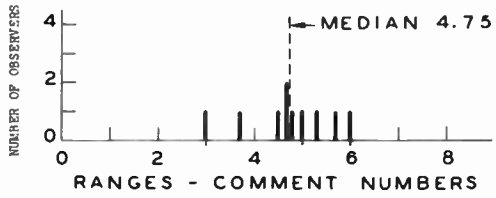


Fig. 16
Distribution of ranges of individual observers.

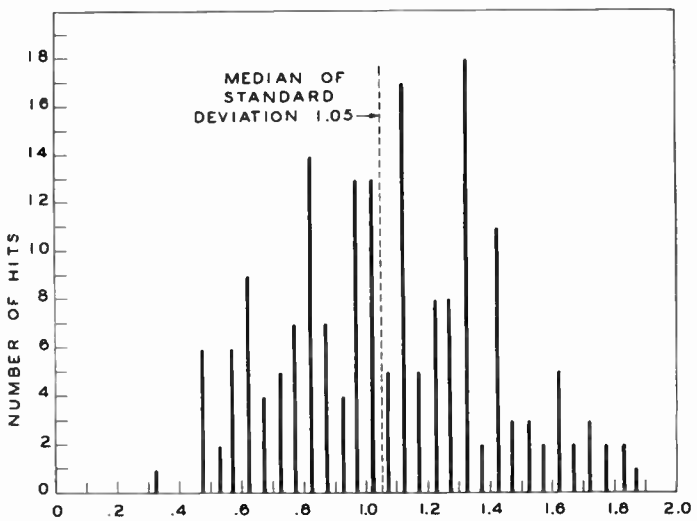


Fig. 17
Distribution of standard deviation of comment numbers from median comment for each hit.

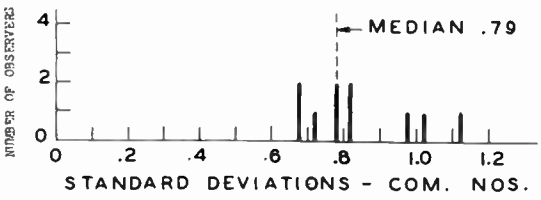


Fig. 15
Distribution of standard deviations of individual observers.

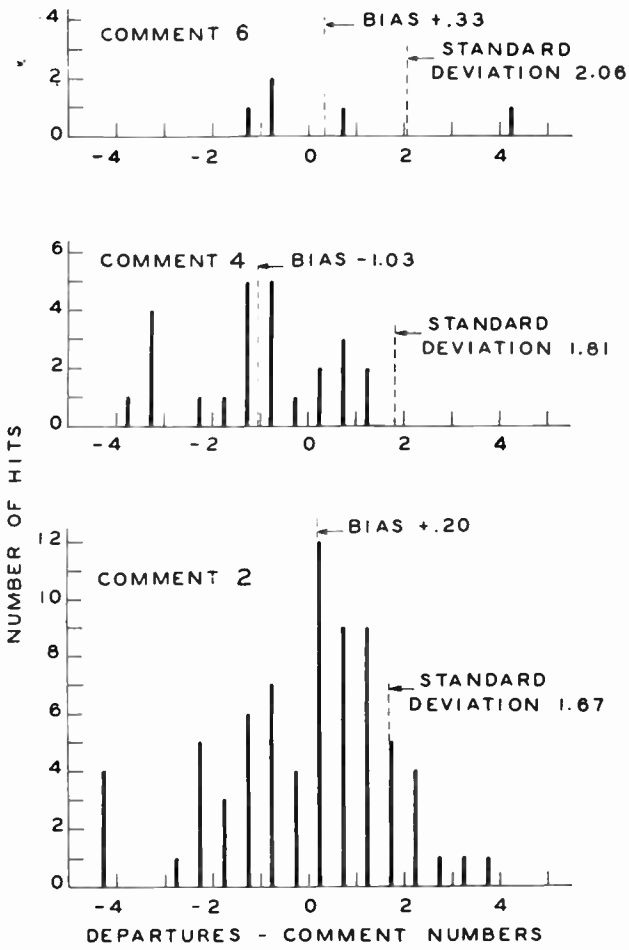


Fig. 18
Distribution of departures from median curve for various comment numbers.

RELIABILITY OF MILITARY ELECTRONIC EQUIPMENT AND OUR ABILITY TO MAINTAIN IT FOR WAR

by

Archibald S. Brown
Stanford Research Institute
Stanford, California

That military electronic equipment has been too unreliable, too complicated and too difficult to maintain is a conclusion reached by an increasing number of engineers during the past two or three years. Within the recent 12 months or so it has been interesting to note the ever growing published sentiment, one being a severe criticism which I quote as follows, "Electronic engineers have been engaged in producing wasteful, witless complexity side by side with indispensible complication."

And we have the important publication *Aviation Newsweek* coming forth with an editorial on September 18, 1952 entitled "Halting the March of Complexity" which I quote in part, "Some top Air Force officials are concerned enough with the problem of aircraft and equipment complexity to order a thorough study.....The time has passed when extra gadgetry can be permitted to substitute for simplicity."

However one sentence contained in this editorial tends to bear out my own fear of lack of sufficiently rapid progress. This is, "Obviously, any noticeable slowdown in this fantastic race toward complexity is out of the question immediately." But I sincerely hope, as stated, that "Increasing attention is being given in upper echelons of the need to apply the brakes."

Hardly a week goes by without some publication having comments on this subject, which is encouraging from the standpoint of pressure to improve the situation but a discouragement in awakening us to realize how far we still have to go.

It is of course unfortunate that during the five years from VJ day to 1950 we did nothing to improve these features, vitally affecting usefulness of many weapons for national defense during both peace and war. Actually we aggravated the problem by designing equipment more and more complicated, being convinced that the main job of scientists was to "expand the parameters of science." During these five years I know of no project which said, "Simplify, make it reliable and easier to maintain."

In 1945-46 we considered that the next war was a long ways off and most of World War II equipment would be useless for World War III, so research was the only thing worthwhile. I do not discount the value of research, much of it being essential for the solution of the problems now under discussion.

It took the shock of the Korean emergency to rouse us somewhat to face the fact that the next war needing electronic equipment might not be so far distant after all. About this time some of the high ranking commanders of our military forces were decidedly annoyed over failure of missions due to faulty electronic equipment. As a result, in meetings, a number of engineers agreed that something should be done to improve reliability. Projects were then authorized with funds to improve the life and reliability of tubes and components, while increased importance was placed on the development of transistors.

But we must not forget that as a basic rule of effectiveness for war, only the equipment actually in the hands of combat units, fighting airplanes, warships, tanks and other machines necessary for military operations can be counted as effective completion of projects. No matter how much we have in future budgets, present research and development projects or even production contracts, in shipment, in supply, going into new construction, only that in the hands of the ultimate users in completely sufficient quantities in usable condition can be considered a weapon in commission. Finally only that which can be kept in operating condition provides an efficient and satisfactory weapon.

For those who remain unconvinced as to unreliability or think because their television set is no worse than mine totaling in failures over two years - one 7N7 tube at 3 months and another 18 months later on the vertical sync. circuit, the military situation is something that proper organization can and should remedy, I would like to give a few examples out of a tremendous number.

During the war I was part of the Navy electronic organization in the New York Area, whose responsibility, among other things, was the correction of any difficulty with electronic equipment on any naval vessel entering, leaving or within that district. This was composed of the Navy Yard, Brooklyn, the Annex at Bayonne, the 100 commercial shipyards located largely on New York Harbor, the Hudson and East Rivers. There was everything from patrol craft and motor torpedo boats to battleships and carriers. We had in the organization civilian civil service engineers, technicians and mechanics, civilian field service engineers from the contractors, and Navy engineer officers. Night and day they had to be available to go down on the ships and make things work that previously did not.

Hardly anyone connected with any similar military activity could fail to be permanently impressed with the lack of reliability and real serviceable quality. I don't know how many times we were routed out of bed because a destroyer scheduled to sail at dawn as an Atlantic convoy escort had a foot or more of water in the sound room. To put it mildly it emphasized the unsatisfactory mechanical design which had guide rods that could slip out through the bottom of the hull, leaving 1 in. holes open to the sea. This was especially imprinted on our minds if the ship was delayed to miss its rendevouz with the convoy and it became necessary to explain why to the Admiral so he could inform the Commander in Chief, Atlantic Fleet.

Then to pursue the subject of underwater sound a little further, consider the failure of projector head windings that occurred with great frequency in the early part of the war. This meant placing the ship in dry dock, which not only was a big job, time-consuming and expensive, but sufficient dry dock space was not available. A big job meant nothing with all the other bigger ones, and expense was no object as we only consider the poor taxpayer at our own income tax time. To replace a $1000.00 electronic coil, however, costs upward of $100,000.00 for dry docking, not counting wastage of workmen's time while docking and undocking because they could not get on or off the ship and generally couldn't work. We bear with these things during the war, but delay in getting into a dock because of scarcity and length of docking time meant a valuable ship out of action when needs were urgent. And last but not least, failure of the projector head at sea meant loss of submarine detecting ability with possible torpedoing and even loss of the ship itself.

A few hours on a ship in the East River, New York, in mid-winter with sleet icing the masts was sufficient for us to condemn pressurized coaxial lines which could not be kept tight. There was also the thing called a "rotary-joint" for coaxial lines, passing into a rotating structure such as a gun turret or a fire control director, which rarely worked properly. Trying to keep in repair the radar antenna rotating mounts which repeatedly failed was another chore causing us to wonder if there were any good mechanical engineers in the U. S. who could properly design the gears. At other intervals pulse transformers burned out or vacuum tubes blew up leaving few dull moments in the life of an electronic engineer, technician, officer or enlisted man responsible for keeping military equipment in operating condition. A long list could be compiled including Loran receivers for aircraft as well as marine vessels, and on it would be other airborne equipment of great importance such as radio compasses, command and liaison sets, equipment for instrument landing, gunnery radar, etc. All these have histories pointing to lack of simplicity, reliability and ease of maintenance.

During wartime there is bound to be more or less trouble with apparatus that is new, and lack of time has not permitted previous service shakedown trials. Our scientists and engineers are then striving hard to invent new weapons, thus we may excuse some, but not all, of the difficulties. So we come to the question of maintenance, and the service organization to take care of these troubles.

I previously stated that we did not have enough technical electronic personnel in the New York area during the war, and I am sure this was true of every military organization. We can only conclude that World War II indicated that the Armed Forces did not have a sufficiently trained and comprehensive organization for maintenance of electronic equipment. Just how far short of the optimum goal is open to question, but anyhow our main job now is to estimate the future potentialities.

Like many engineers who grew up working largely in design, development, research and manufacturing, I believed it was fundamental that good maintenance was largely a matter of good organization. Through long experience the slogan "The Navy takes care of its own" applied to electronics also. There were always oldtime warrant officers, experienced chief petty officers, and well trained enlisted personnel to take care of the equipment after the engineers were finished with the installation.

Because of this long existant situation, despite the examples of poor reliability which were encountered, I continued to wonder when the Armed Forces would set up better organization to cope with the ever increasing complication of military electronics. Noting that conditions actually became worse for several years following the war and that since the beginning of Korea improvements have been made with great difficulty, the only logical conclusion remaining is to seek some other solution, if possible.

At least one reason for the troubles in the Armed Forces is the extreme shortage of technical manpower available to them due to the constantly increasing complication of equipment, the length of time required for training and the rapid attrition because of men leaving the service. This shortage extends in less degree or different manner also to civilian organization as you have possibly read in many publications.

So I have come to doubt entirely the practicability of throwing a bigger and bigger job of maintenance during wartime onto the folks in uniform. The real answer lies in the hands and brains of the engineers and manufacturers who must make the equipment of utmost reliability such as we provide in a service rifle. So far I have never heard of a piece of electronic equipment capable of being taken apart and reassembled blindfolded!

You are all somewhat familiar with the commercial pattern of organization for servicing television, radio and electrical appliances whereby all kinds of troubles are more or less routine, and service men of considerable training and experience are available for hire to correct the difficulties. Organizations of similar numbers,

breadth and depths of capability as part of the Armed Forces would undoubtedly be a fine thing but not possible to accomplish even for total war, the age and physical condition of these skilled service men being largely unsuitable for military duty if no other bars existed.

On a much smaller scale for maintenance of present day electronic equipment are the Airlines organizations of repair and overhaul shops and line maintenance. Here will be noted servicemen of great skill and many years of experience efficiently repairing and maintaining any and all types of equipment. Obviously, however, military requirements are many, many times the size of even the largest airline.

I trust that the impractibility if not impossibility of equalling these organizations on the tremendous scale required for World War III by the Armed Forces is evident. Therefore we must depart from the present philosophy of maintenance and insist that military electronic equipment supplied in the future be more rugged and reliable. It is imperative that in the earliest future years the Armed Forces be provided with equipment of such reliability that tubes and parts will rarely fail.

The military concept of maintenance should be able to depart radically from the commercial one where skilled service is always available because there will not be enough technical manpower for World War III. This should be especially true of line, field, and aboard ship maintenance where methods of minimization need to be applied. One method on radar equipment was demonstrated by an Office of Naval Research Contract to Stanford Research Institute and described by Wm. D. McGuigan at the National Convention of the IRE, 6 March, 1952.

The military need for increased technical manpower to work on new important projects is faced with the definite limitations shown by estimates of decreasing numbers of engineering graduates in the next few years. Money may also be a problem because the Armed Forces are being called upon for more and more justification of expenditures. Therefore I have a proposal to present to you having a bearing on technical manpower, funds and maintenance of electronic equipment.

Let us use the familiar commercial television receiver as an example with a lot of different designs produced and sold by many manufacturers, such as General Electric Co., Raytheon, Westinghouse, Philco, Sylvania, Hofman, etc. Suppose that in place of these multiple designs it was decided to produce only one, for instance, G. E. All the factories would then be turning out the same set and all the engineering manhours spend in the design of the others could be diverted to something else important. The cost of design of the General Electric could be spread over many more thousands of equipments. The new technique of automatic fabrication and assembly that is gaining emphasis among radio and television manufacturers might be more advantageously utilized.

Service organizations would need to carry only one kind of parts and servicemen would need to be trained for only one make of set. There would be greatly reduced numbers of instruction books, service manuals and parts lists.

Of course with our free enterprise system, which we would not want otherwise, this is entirely impractical and impossible commercially, but the Armed Forces are not prohibited from utilizing the advantages of less number of designs to save manpower and money. Likewise the assembly people might turn out more perfect military equipment with increased repetition and mechanized production may produce more uniform reliability. Printed circuit methods, automatic assembly of components and dip soldering of the chassis already show this.

There are many laws governing purchasing by government activities such as those requiring contract awards to lowest bidders and limitations on orders to sole sources. Nothing prevents government engineers from freezing one design except the human desire to obtain always the ultimate in perfection. Nothing prevents procurement officers, in need of multiple sources, from insisting upon identical designs. This does not limit multiple research and development to bring out major improvements but merely proposes to limit the final product reaching operating military forces to simpler numbers of designs. It does discourage shopping around even when the "grass on the other side of the fence looks greener," and it does require more intestinal fortitude to resist salesman "X" and his tales of superior performance on his company's commercial design. Sometimes it will be like continuing to be satisfied with your fine running '49 model car instead of turning it in for the shiny '53 which you find has a few new "bugs" also. It may be really tough, like holding onto your '51 television in the face of advertised super-duper performance on the '53 models for sale by madman "Dunce."

An example of one research development and production program is indicated in Figure 1. In 1945 the Joint Chiefs of Staff of Army, Navy and Air Force, upon competent technical advice, approved a new requirement. In '46 joint Army-Navy Specifications were agreed upon, followed by dual research-development contracts A & B to obtain test models. In 1948 these were evaluated, but in '49 a third contractor C received the production contract and B a limited interim order only, indicated by sub L. Contractor A dropped out, thus the engineering time and dollars which went into his model were lost. In 1950 B_L and C continued while a contractor D worked on another development for the same purpose which was evaluated in '51. By this time despite further orders to B_L, the output combined with C was insufficient to meet military demands. So additional sources were sought in '52 and by now, 1953, production by B_L, C, E, F, plus D may be underway.

If E and F are producing sets completely identical with C it is excellent but if not the Armed Forces end up with at least five types, which I hope is not the situation.

'45	JOINT CHIEFS
'46	A/N SPEC.
'47	CONTR. A B
'48	EVAL. A B
'49	CONTR B_L C
'50	CONTR B_L C D
'51	EVAL. D PROD. B_L C
'52	PROD. B_L C D
'53	PROD. B_L C D E_C F_C

FIG. 1

To summarize, if reduction of numbers of designs could be accomplished in the Armed Forces engineering-procurement program for electronic equipment, the beneficial results might be as follows:

(a) Reduced engineering manhours now spent in developing competing designs for production.

(b) Reduced production costs due to larger quantities of a single design.

(c) Easier maintenance, less training for maintenance and probably for operation plus improved reliability because of less numbers of designs.

(d) Reduced costs of stocks of parts.

(e) Saving in costs involved in installation after delivery of airplane, ship or tank instead of on the production line of airplane, ship or tank manufacturer because of earlier deliveries of electronic equipment.

(f) More improved equipment in the hands of combat forces at earlier dates due to higher production.

Recommendation:

I have only one, and that is the appointment of a single competent engineer, special assistant to Mr. Wilson, Secretary of National Defense, to examine the programs of the Army, Navy and Air Force to indicate where improvements should be made. Six months' time would accomplish much as only the large quantity programs apply.