

Convention Record



of the I·R·E

1954 NATIONAL CONVENTION

Part 1— Antennas and Propagation

LIBRARY
JUL 7 1954
U. S. PATENT OFFICE

L
G
V
L

SESSIONS ON . . .

Antennas and Propagation I — General

Antennas and Propagation II — Microwave Antennas

Antennas and Propagation III

Antennas and Propagation IV — Symposium: UHF Television — Boom or Bust

SPONSORED BY

IRE PROFESSIONAL GROUP ON . . .

Antennas and Propagation

Presented at the IRE National Convention, New York, N.Y., March 22-25, 1954
Copyright 1954, by The Institute of Radio Engineers, Inc., 1 East 79 Street, New York 21, N.Y.

The Institute of Radio Engineers

TK
6541
I 55

Additional Copies

Additional copies of 1954 Convention Record Parts may be purchased from The Institute of Radio Engineers, 1 East 79 Street, New York 21, N. Y., at the prices listed below.

Part	Title	Sponsoring Groups	Prices for Members (M), Libraries (L), & Nonmembers (NM)		
			M	L	NM
1	Antennas & Propagation	Antennas & Propagation	1.25	3.00	3.75
2	Circuit Theory	Circuit Theory	1.25	3.00	3.75
3	Electron Devices & Component Parts	Electron Devices Component Parts	1.50	3.60	4.50
4	Electronic Computers & Information Theory	Electronic Computers Information Theory	1.50	3.60	4.50
5	Aeronautical Electronics & Telemetry	Aeronautical & Navigational Electronics	1.50	3.60	4.50
6	Audio & Ultrasonics	Radio Telemetry & Remote Control Audio Ultrasonics Engineering	1.50	3.60	4.50
7	Broadcasting & Television	Broadcast Transmission Systems Broadcast & Television Receivers	1.50	3.60	4.50
8	Communications & Microwave	Communication Systems Microwave Theory & Techniques Vehicular Communications	1.50	3.60	4.50
9	Medical & Nuclear Electronics	Medical Electronics Nuclear Science	1.50	3.60	4.50
10	Instrumentation & Industrial Electronics	Instrumentation Industrial Electronics	1.25	3.00	3.75
11	Engineering Management & Quality Control	Engineering Management Quality Control	1.00	2.40	3.00
			15.25	36.60	45.75

Responsibility for the contents of papers published in the Convention Record of the I. R. E. rests solely upon the authors, and not upon the IRE or its members.

CONVENTION RECORD OF THE I.R.E.

1954 NATIONAL CONVENTION

LIBRARY

JUL 9 1954

D. S. PAULI OFFICE

PART 1 - ANTENNAS AND PROPAGATION

TABLE OF CONTENTS

Session 30: Antennas and Propagation I - General

(Sponsored by the Professional Group on Antennas and Propagation.)

Empirical Approximations to the Current Values for Large Dolph-Tchebyscheff Arrays	Louis L. Bailin, Robert S. Wehner, and Ivan P. Kaminow	3
Gain Pattern of a Terminated-Waveguide Slot Antenna by an Equivalent Circuit Method	Leopold B. Felsen	10
A Four Slot Cylindrical Antenna for VOR Service	Andrew Alford and R.M. Sprague	12
Trapped Wave Antennas	Herman Ehrenspeck, Werner Gerbes, and Francis J. Zucker	25
Scattering of Electromagnetic Waves by Wires and Plates	J. Weber	31

Session 37: Antennas and Propagation II - Microwave Antennas

(Sponsored by the Professional Group on Antennas and Propagation.)

Reflections in Microwave Antennas and Their Harmful Effects	Peter W. Hannan	39
Surface Matching of Dielectric Lenses	E.M.T. Jones and S.B. Cohn	46
Double Parabolic Cylinder Pencil Beam Antenna	Roy C. Spencer, F. Sheppard Holt, Helen M. Beauchemin, and John L. Sampson	54
Diffuse Radiation in Pencil Beam Antennas	David Carter	60
Theoretical Gain of Flat Microwave Reflectors	D.R. Crosby	71

Session 42: Antennas and Propagation III

(Sponsored by the Professional Group on Antennas and Propagation.)

Isotropic Variable Index Media	W.O. Puro and K.S. Kelleher	76
The Characteristics of a Vertical Antenna with a Radial Conductor Ground System	James R. Wait and W.A. Pope	79
Some Information Theory Aspects of Propagation Through Time Varying Media	Joseph Feinstein	87
Comparative 100 Mc Measurements at Distances Far Beyond the Radio Horizon	Albrecht P. Barsis	98
The Measurement of the Polarization of Radio Waves Reflected from the Ionosphere at Non-Vertical Incidence	G.T. Inouye	108

Session 49: Antennas and Propagation IV - Symposium: UHF Television - Boom or Bust

(Sponsored by the Professional Group on Antennas and Propagation.)

FCC Rules and Propagation Data	E.W. Allen	116
Propagation in the UHF-TV Band (Abstract)	J.W. Herbstreit	120
Overcoming the Line-of-Sight Shibboleth with the Air and High Power	Thomas J. Carroll	121
A Comparison of Antenna Problems at UHF and VHF TV	Lloyd O. Krause	126



FRANK
JUL 7 1954
U. S. AIR FORCE

EMPIRICAL APPROXIMATIONS TO THE CURRENT VALUES FOR LARGE DOLPH-TCHEBYSCHIEFF ARRAYS

Louis L. Bailin, Robert S. Wehner, and Ivan P. Kaminow
Hughes Research and Development Laboratories
Culver City, California

Abstract

The demand for radar antennas exhibiting a narrow beam and low side-lobes has led to the development of the so called Dolph-Tchebyscheff array which optimizes the relationship between beamwidth and side-lobe level. Unfortunately, however, the calculation of the excitation coefficients for the Tchebyscheff array is quite time consuming and tedious for arrays with a large number of elements. The present paper describes a simple approximate method for calculating long Tchebyscheff arrays which cuts the calculating time by 95 per cent for a 40 element array and which does not increase in complexity with the number of elements in the array. Furthermore, sample calculations indicate errors of only 3 or 4 per cent in the excitation coefficients which is tolerable for most practical arrays. In addition, the paper contains a table of all the coefficients for Tchebyscheff arrays computed in the course of antenna development at the Hughes Aircraft Company. The table contains the exact excitation coefficients and the gain values relative to the corresponding uniform array.

Introduction

It has been shown by C. L. Dolph¹ that the far field radiation pattern for a broadside linear array which optimizes the relationship between beamwidth and side-lobe level is given by a Tchebyscheff polynomial. Such a pattern has many features which are desirable for radar applications. However, to produce this pattern, it is necessary to determine the excitation coefficient at each radiator from formulas which become more and more tedious as the number of radiators increases. One purpose of this paper is to present a table of all the current values for Dolph-Tchebyscheff arrays which have been computed by various analytical formulas in the course of antenna development at Hughes Aircraft Company (see Table I). A second purpose is to present an empirical method (for determining the approximate excitation coefficients for large arrays) which reduces the calculating time by about 95 per cent for the example described at the end of this paper. Although this method will not yield the accuracy in both the coefficients and the corresponding radiation pattern that can be obtained from the analytical formulas, the results will be given simply and the accuracy will be sufficient for most practical antennas.

The Approximate Method

Dolph¹ and others^{2,3,4} have presented exact analytical methods for determining the Tchebyscheff coefficients. Experience has shown that as the number of elements increases (say, greater

than 24), the analytical methods become impractical because of their complexity. Thus, it behooves the array designer to devise a method for quickly and easily arriving at a current distribution which will approximate the Dolph-Tchebyscheff distribution. Such a method, of course, leads to some deterioration of the optimum pattern. Nonetheless, the accuracy of the approximation can be made better than the accuracy with which the desired distribution can be effected in the actual array.

In an effort to devise a method for approximating the Dolph-Tchebyscheff currents, we consider the element excitation currents as a function of their location on the array. Thus, when all the currents are normalized to unity at the center of the array (for an even number of elements an interpolated value is used), and the radiators are equally spaced along a normalized array length, $-1 \leq x \leq 1$; the current distribution is symmetrical about $x = 0$ and decreases monotonically away from the center except for the end elements which increase discontinuously. This is shown in Figure 1 for a 20 db array of 12, 24, 48 elements; in Figure 2 for a 30 db array of 12, 24, 48 elements; in Figure 3 for a 35 db array of 66 elements; and in Figure 4 for a 40 db array of 12, 24, 48, 144 elements*. It is evident that in each case, the continuous curve which passes through the discrete excitation coefficient points does not change appreciably from 24 to 48 elements. Also, in the case of the 40 db array there is very little change from 48 to 144 elements. Therefore, we may conclude that the "envelope" or limiting curves are independent of the number of elements for arrays of more than 24 elements and given side-lobe level. From another point of view, it can be said that Figs. 1, 2, and 4 indicate that, for arrays larger than 24 elements, a discrete array can be approximated by a continuous array (or vice versa) having a similar aperture distribution (excluding end elements).

To derive an empirical expression for these envelope curves**, we consider an even function of x given by

$$f(x) = ax^4 + bx^2 + c, \quad (1)$$

and fit this function to the envelope curve at three points ($x=0, 0.5, 1.0$) as shown by the cross marks on the solid curves in Figures 1, 2, 3, and 4. The approximation is fairly good, although marked deviations occur for the lower side-lobe cases. However, if the square root of

* End element amplitudes greater than unity are not shown.

** An analytical expression, in terms of a complex Bessel function, for the envelope curves has been derived independently by G. J. van der Maas⁴ and T. T. Taylor⁵.

the envelope curve is plotted and the resulting envelope fitted by

$$g(x) = Ax^4 + Bx^2 + C \quad (2)$$

at the same three points, then excellent agreement is obtained through the range of x except for the end points. The approximation is shown in Figures 1, 2, 3, and 4 by the cross marks on the dashed curves.

Fortunately, the associated coefficients, A , B , and C , for the polynomial approximation are related simply to $\cosh^{-1}r$, where r is the main beam to side-lobe voltage ratio. It is apparent that, for a normalized envelope curve ($g(0) = 1.0$), $C = 1$ and B is always negative. A plot of $\cosh^{-1}r$ versus the value of the polynomial coefficients, A and B , is presented in Figure 5 and serves to indicate the relationship between side-lobe level and the coefficients. Empirical formulas for A and B can be written as follows:

$$\begin{aligned} A &= 0.0861 \cosh^{-1}r - 0.228; \\ -B &= 0.225 \cosh^{-1}r - 0.240 \end{aligned} \quad (3)$$

The current amplitude distribution along a 20 db, 30 db, 35 db, or 40 db Tchebyscheff array can be found directly from the curves of Figures 1, 2, 3, and 4, respectively. Similar curves for intermediate values of side-lobe level can be computed with the aid of the coefficients of Figure 5, i.e., the square root curve, $g(x) = Ax^4 + Bx^2 + 1$, can be plotted and squared to give the normalized amplitude distribution (envelope curve). It has been pointed out previously that this envelope curve is independent of the number of elements (greater than 24) in the array so that practically any long Tchebyscheff array can be computed from Figure 5 except for the excitation of the last element.

The excitation coefficient for the last element cannot be found from the envelope curve because the amplitude distribution is discontinuous at the ends of the array. Simple exact analytical expressions for the excitation of the last element in terms of the excitation of the next to last element can be derived from the equations of Dolph¹ or Stegen³. The expressions are:

$$I_N = I_{N-1} \frac{Z_0^2}{(2N-1)(Z_0^2-1)} \quad \text{for } 2N \text{ elements,} \quad (4)$$

$$I_N = I_{N-1} \frac{Z_0^2}{2N(Z_0^2-1)} \quad \text{for } 2N+1 \text{ elements;}$$

where

$$Z_0 = \frac{1}{2} \left[(r + \sqrt{r^2 - 1})^{1/M} + (r - \sqrt{r^2 - 1})^{1/M} \right]$$

$$\text{or} \quad Z_0 = \cosh \left(\frac{1}{M} \cosh^{-1}r \right), \quad (5)$$

and M is one less than the number of elements in the array.

If adequate tables are available, the second form will be easier to handle.

Example:

Suppose that we wish to find the currents for a 38 element 30 db Tchebyscheff array. The steps follow:

1. From Figure 5 or equation (3) we find the values of A and B for $r = 31.62$ (30 db) -- $A = 0.130$, $B = 0.690$ -- and plot the curve $g(x) = 0.130x^4 - 0.690x^2 + 1$, as shown in Figure 6(a).
2. Then we square the $g(x)$ curve to obtain the current distribution curve (envelope curve), $I(x) = g^2(x)$, as shown in Figure 6(b).
3. Now divide the abscissa of (b) into 37 (one less than the number of elements) equal parts with 38 marks, as in Figure 6(c). These marks indicate the positions of the elements along the normalized array length. For an odd number of elements, one of the marks will occur at the center of the array.
4. We can read directly from the curve, (c), the values of current at each element except for the end elements. The values read from the curve at $x = \pm 1.0$ are extraneous. The $I(x)$ curve need not be plotted; the excitation currents of elements at the positions x , defined in paragraph 3, can be determined by substituting for x in the equation for $I(x)$.
5. To find the excitation for the end elements, we calculate Z_0 from equation (5) and apply equation (4),

$$I_{19} = I_{18} \frac{Z_0^2}{37(Z_0^2-1)}, \quad (2N = 38),$$

where I_{18} has been found in Step 4. For this case $Z_0 = 1.0062895^*$, $I_{18} = 0.238$, and $I_{19} = 0.515$. These actual current values differ from the values given in Table I by a constant multiplier (the normalizing factor is 1/2.493).

A comparison of actual current values (from Table I), normalized to unity at the center of the array, and normalized approximate current values (calculated as indicated above) for the 38 element 30 db array is given in Table II. For this calculation the curves were not actually drawn: $g(x)$ was calculated for values of x corresponding to radiating elements as described in paragraph 4 and then $g^2(x) = I(x)$ was found. The error for the approximate method is greatest

* Note that Z_0 must be calculated to a large number of places to maintain a given accuracy since some significant figures will be lost in the $(Z_0^2 - 1)$ term.

near the ends of the array but is less than 3 per cent at the next to last and last elements*. The approximate calculating time (with an electrically driven desk calculator) for the more accurate method is 20 to 30 hours, whereas, for the approximate method, the calculating time is about one hour.

The space factor gain of this (or any discrete linear non-supergain array) can be calculated from the relations

$$G_{2N} = G_{u1} \cdot \frac{(\sum_{k=0}^{N-1} I_{k+1})^2}{N(\sum_{k=0}^{N-1} I_{k+1}^2)} \text{ for } 2N \text{ elements} \quad (6)$$

and

$$G_{2N+1} = G_{u2} \cdot \frac{(I_0 + 2 \sum_{k=1}^N I_k)^2}{2N+1 (I_0^2 + 2 \sum_{k=1}^N I_k^2)} \text{ for } 2N+1 \text{ elements,}$$

where $G_{u1} = \frac{2d \cdot 2N}{\lambda}$ and $G_{u2} = \frac{2d \cdot (2N+1)}{\lambda}$ are the gains of the corresponding uniform arrays. The gain calculated from the actual currents is .876 G_{u1} , and the gain determined from the approximate currents is .878 G_{u1} . The value G given in Tables I and II is G_{2N}/G_u or G_{2N+1}/G_u , the gain relative to the equivalent uniform array.

* Similar checks for a 33 element 25 db array, a 66 element 35 db array, and a 144 element 40 db array indicate errors of less than 5 per cent for the end elements.

References

1. C. L. Dolph; "A Current Distribution for Broadside Arrays Which Optimizes the Relationship Between Beamwidth and Side-Lobe Level", Proc. I.R.E., Vol 34, pp. 335-348, June 1946.
2. D. Barbieri; "A Method for Calculating the Current Distribution of Tchebyscheff Arrays", Proc. I.R.E., Vol 40, pp. 78-82, Jan. 1952.
3. R. J. Stegen; "Excitation Coefficients and Beamwidth of Tchebyscheff Arrays", Proc. I.R.E., Vol 41, pp. 1671-1674, November 1953.
4. G. J. van der Maas; "A Simplified Calculation for Dolph-Tchebyscheff Arrays", J. Appl. Phys., Vol 24, p 1250, September 1953.
5. T. T. Taylor; "Dolph Arrays of Many Elements", Hughes Aircraft Company Technical Memorandum No. 320, August 1953.

Acknowledgement

The authors wish to acknowledge the computing services of Mrs. Annabelle Cordova and

the Central Group of Mathematicians of Hughes Aircraft Company.

Table I

Current Values for Dolph-Tchebyscheff Arrays

<u>Contents</u>	
No. Elements	-db Side-Lobe Level
4	20
6	30
7	20
8	28, 30, 32
10	25
12	20, 30, 40
16	32, 36
18	25
24	20, 30, 40
33	25
38	30
40	36
48	20, 30, 40
66	35
144	40

4 Elements

20 db

k	I_k
1	6.34467
2	3.65532
Z_0	= 1.54043
r	= 10
G	= .93255

6 Elements

30 db

k	I_k
1	15.97290
2	10.92129
3	4.72224
Z_0	= 1.364048
r	= 31.62278
G	= .83991

7 Elements

20 db

k	I_k
0	3.76820
1	3.45054
2	2.61588
3	2.04944
Z_0	= 1.12704
r	= 10
G	= .95082

8 Elements

	28 db	30 db	32 db
<u>k</u>	<u>I_k</u>	<u>I_k</u>	<u>I_k</u>
1	9.411472	12.19580	15.76045
2	7.753466	9.90250	12.61676
3	5.116790	6.32654	7.809191
4	2.837067	3.19794	3.624265
Z ₀ =	1.160635	1.1806586	1.201956
r =	25.11887	31.62278	39.81072
G =	.862324	.841612	.822562

18 Elements

	25 db
<u>k</u>	<u>I_k</u>
1	2.74792
2	2.66535
3	2.50596
4	2.28072
5	2.00480
6	1.69616
7	1.37397
8	1.05695
9	1.45099

10 Elements

25 db

<u>k</u>	<u>I_k</u>
1	5.0500
2	4.54167
3	3.64306
4	2.55344
5	1.99461
Z ₀ =	1.07973575
r =	17.782794
G =	.904804

Z ₀ =	1.02213846
r =	17.782794
G =	.921175

24 Elements

	20 db	30 db	40 db
<u>k</u>	<u>I_k</u>	<u>I_k</u>	<u>I_k</u>
1	1.01822	3.9901	14.5769
2	1.00528	3.9080	14.1372
3	.97977	3.7478	13.3433
4	.94242	3.5173	12.2159
5	.89428	3.2274	10.8317
6	.83671	2.8918	9.2731
7	.77128	2.5256	7.62242
8	.699775	2.1444	6.07028
9	.62406	1.7643	4.58229
10	.54608	1.3991	3.26402
11	.46775	1.0614	2.15811
12	1.21436	1.4504	1.83112
Z ₀ =	1.0084801	1.016298	1.026655
r =	10	31.62278	100
G =	.94044	.87134	.77191

12 Elements

	20 db	30 db	40 db
<u>k</u>	<u>I_k</u>	<u>I_k</u>	<u>I_k</u>
1	2.09821	8.12761	29.29670
2	1.98542	7.43902	25.95992
3	1.77354	6.20007	20.21841
4	1.48752	4.64892	13.57260
5	1.16013	3.06107	7.53420
6	1.49518	2.14609	3.41817
Z ₀ =	1.03725	1.07190	1.11826
r =	10	31.62278	100
G =	.96428	.85284	.75976

16 Elements

	32 db	36 db
<u>k</u>	<u>I_k</u>	<u>I_k</u>
1	7.81	13.19830
2	7.50	12.43593
3	6.73	11.01836
4	5.70	9.13899
5	4.55	7.04060
6	3.37	4.97075
7	2.28	3.13961
8	1.86	2.15322
Z ₀ =	1.042880	1.0524605
r =	39.81072	63.09574
G =	.840	.799355

33 Elements

	25 db
<u>k</u>	<u>I_k</u>
0	1.4730106
1	1.4666424
2	1.4476607
3	1.4164483
4	1.3736270
5	1.3200464
6	1.2567533
7	1.1849740
8	1.1060732
9	1.0215286
Z ₀ =	1.0062314
r =	17.782794
G =	.919856

38 Elements

30 db			
k	I _k	k	I _k
1	2.49199	12	1.41622
2	2.47178	13	1.26664
3	2.43221	14	1.11879
4	2.37382	15	.97510
5	2.29722	16	.83613
6	2.20441	17	.70359
7	2.09746	18	.57550
8	1.97766	19	1.26063
9	1.84679		
10	1.70799	Z _o	= 1.0062895
11	1.56401	r	= 31.62278
		G	= .87665

40 Elements

36 db			
k	I _k	k	I _k
1	5.182363	13	2.436042
2	5.135927	14	2.121467
3	5.044040	15	1.818732
4	4.908665	16	1.532093
5	4.732775	17	1.265176
6	4.520148	18	1.020841
7	4.275316	19	.801212
8	4.003356	20	1.348897
9	3.709761		
10	3.400315	Z _o	= 1.00770345
11	3.080970	r	= 63.09574
12	2.757645	G	= .81739

48 Elements

20 db			
k	I _k	k	I _k
1	.49995	13	7.19977
2	.49842	14	7.14943
3	.49536	15	7.04964
4	.49080	16	6.90211
5	.48478	17	6.70939
6	.47732	18	6.47475
7	.46849	19	6.20216
8	.45834	20	5.89614
9	.44694		5.56169
10	.43437		5.20417
11	.42071		4.82916
12	.40606		4.44232
13	.39051		4.04933
14	.37416		3.65569
15	.35712		3.26668
16	.33949		2.88717
17	.32139		2.52162
18	.30293		2.17395
19	.28423		1.84749
20	.26538		1.54493
21	.24651		1.26832
22	.22772		1.01905

48 Elements (Cont'd)

20 db		30 db		40 db	
k	I _k	k	I _k	k	I _k
23	.20911	23	.43693		.79787
24	1.09993	24	1.20045		1.34718
Z _o	= 1.002029		1.003895		1.006361
r	= 10		31.62278	100	
G	= .85821		.87718		.78280

66 Elements

35 db			
k	I _k	k	I _k
1	2.73787	20	1.45000
2	2.72927	21	1.34917
3	2.71214	22	1.24926
4	2.68661	23	1.15083
5	2.65286	24	1.07115
6	2.61117	25	.96050
7	2.56184	26	.86955
8	2.50525	27	.78198
9	2.44182	28	.69815
10	2.38875	29	.61837
11	2.29635	30	.54293
12	2.21538	31	.47203
13	2.12967	32	.40585
14	2.03985	33	1.18698
15	1.94653		
16	1.85035	Z _o	= 1.00264
17	1.75197	r	= 56.23413
18	1.65202	G	= .83149
19	1.55115		

114 Elements

40 db			
k	I _k	k	I _k
1	2.3727	22	1.9852
2	2.3709	23	1.9512
3	2.3673	24	1.9161
4	2.3619	25	1.8801
5	2.3548	26	1.8432
6	2.3458	27	1.8055
7	2.3351	28	1.7669
8	2.3226	29	1.7276
9	2.3085	30	1.6877
10	2.2927	31	1.6472
11	2.2752	32	1.6062
12	2.2561	33	1.5647
13	2.2355	34	1.5229
14	2.2132	35	1.4807
15	2.1895	36	1.4383
16	2.1643	37	1.3957
17	2.1377	38	1.3531
18	2.1097	39	1.3103
19	2.0804	40	1.2676
20	2.0499	41	1.2249
21	2.0181	42	1.1824

144 Elements (Cont'd)

40 db

k	I_k	k	I_k
43	1.1400	60	.49987
44	1.0989	61	.46918
45	1.0562	62	.43943
46	1.0148	63	.41065
47	.97385	64	.38285
48	.93336	65	.35605
49	.89339	66	.33024
50	.85398	67	.30544
51	.81518	68	.28165
52	.77704	69	.25886
53	.73959	70	.23708
54	.70288	71	.21631
55	.66694	72	1.1031
56	.63180		
57	.59749	$Z_o = 1.000687$	
58	.56405	$r = 1.00$	
59	.53150	$G = .78938$	

Table II

Comparison of Actual and Approximate Current Values for 38 Element 30 db Tchebyscheff Array

k	x	$g(x)$. $130x^4 - .690x^2 + 1$	$g^2(x) = I(x)$ (Approx)	$I'(x)$ (Actual)
-	0	1.000	1.000	1.000
1	.027	1.000	1.000	.9996
2	.081	.996	.991	.9916
3	.135	.988	.975	.9755
4	.189	.976	.952	.9523
5	.243	.960	.921	.9214
6	.297	.940	.884	.8841
7	.351	.917	.841	.8411
8	.405	.890	.793	.7934
9	.459	.860	.740	.7409
10	.513	.827	.685	.6851
11	.567	.792	.627	.6274
12	.621	.753	.567	.5680
13	.675	.713	.508	.5082
14	.729	.670	.449	.4489
15	.783	.626	.392	.3911
16	.837	.580	.327	.3353
17	.891	.534	.285	.2824
18	.945	.488	.238	.2308
19	1.000	-	0.515	0.5007

$G = .878$ $G = .8764$

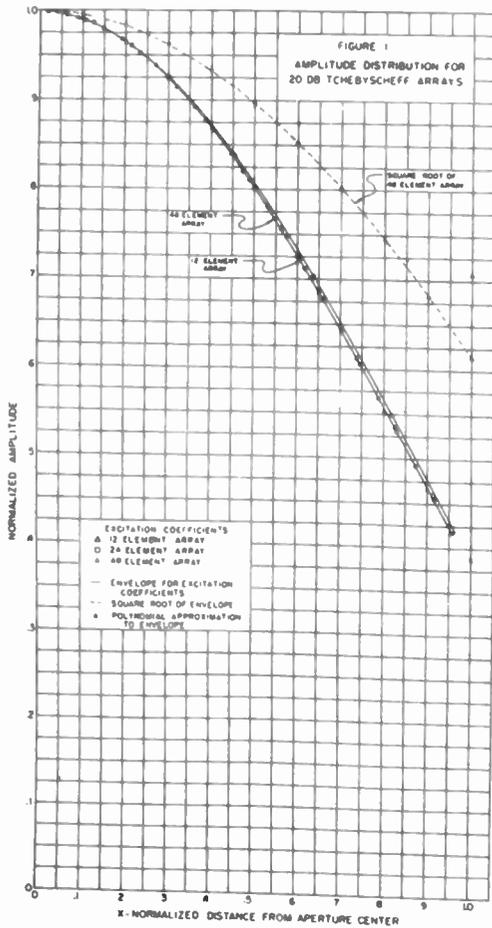


Fig. 1

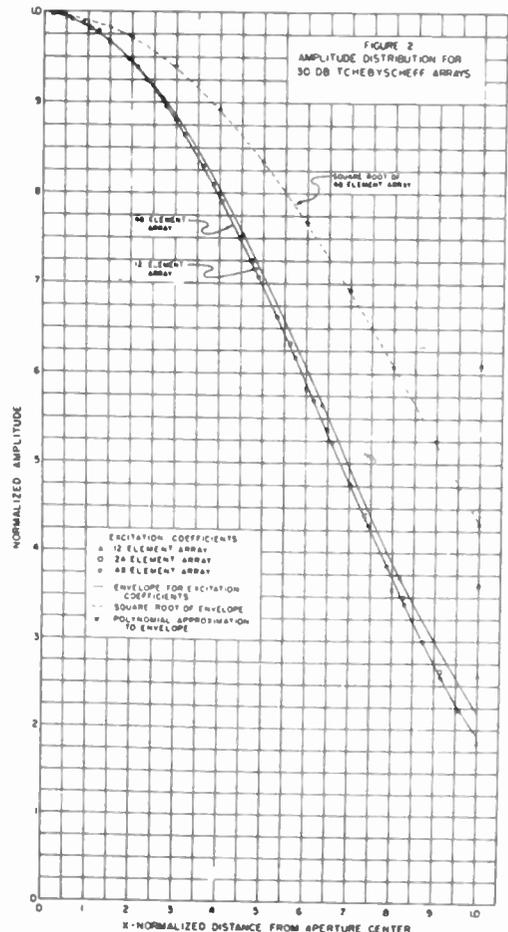


Fig. 2

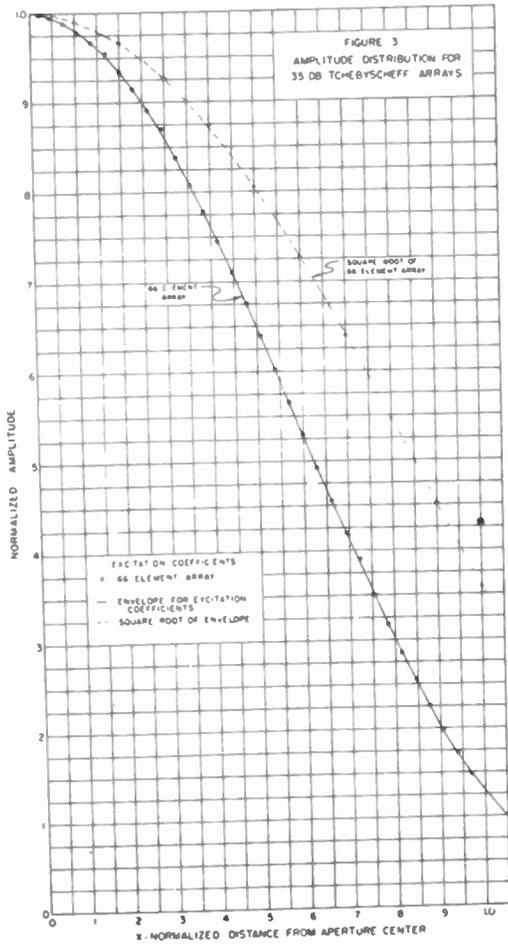


Fig. 3

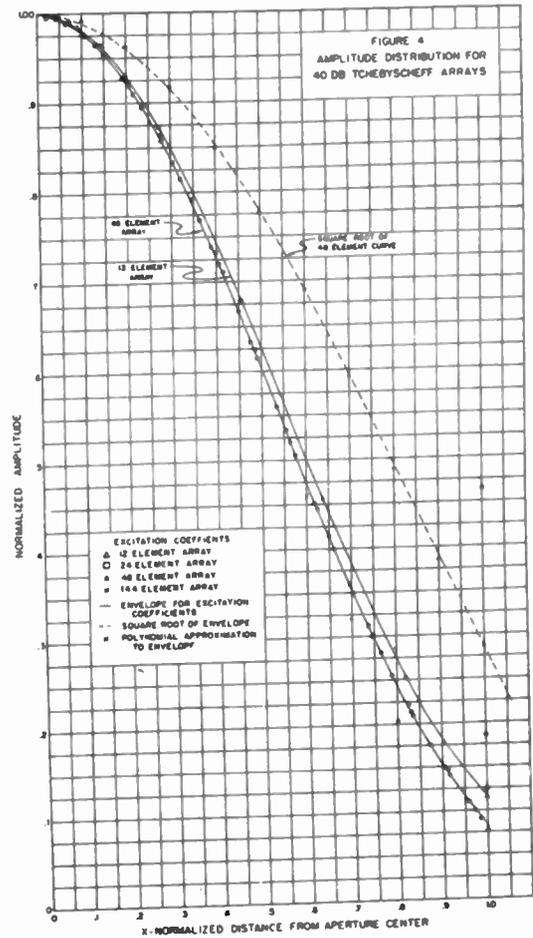


Fig. 4

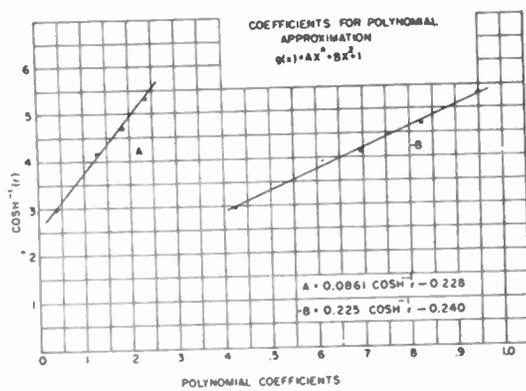


Fig. 5

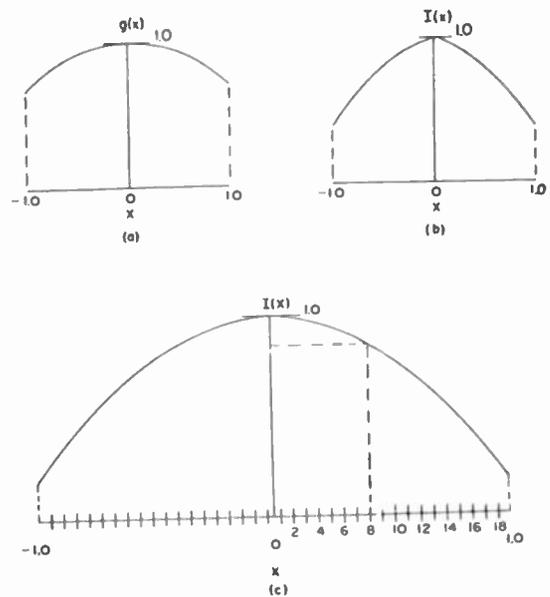


Fig. 6

GAIN PATTERN OF A TERMINATED - WAVEGUIDE SLOT ANTENNA BY AN
EQUIVALENT CIRCUIT METHOD

by

Leopold B. Felsen

Microwave Research Institute
Polytechnic Institute of Brooklyn
Brooklyn 1, N.Y.

Summary

The terminated - waveguide slot antenna consists of a slot cut in the wall of a waveguide terminated in a known load. The gain pattern of such a slot is a function of the waveguide termination and may change markedly for different load values. The slot field can be regarded as composed of two independent field types arising from symmetric and antisymmetric electric field excitations of the slot, respectively; the relative intensities of the two excitations depend on the waveguide termination. Each of these constituent fields has its own characteristic radiation pattern, and the total gain pattern is obtained by superposition. The determination of the constituent slot fields from which the radiation fields may be computed is an extremely difficult task. It is desirable, therefore, to employ a technique which enables one to find the radiated field as a function of the incident field in the waveguide without a detailed knowledge of the slot field.

If the slot radiates into a half-space (i.e., the slot is contained in an "infinite" baffle) the radiated fields are analyzed very conveniently in terms of spherical transmission line theory¹, which considers the half-space region to be a spherical waveguide representable by a number of non-uniform, spherical transmission lines. For many small slot configurations it may be shown that a single spherical transmission line suffices to represent approximately the far fields radiated by a symmetrically and by an anti-symmetrically excited slot. Thus, the half-space is represented approximately by two spherical transmission lines propagating the appropriate spherical modes, and the slot is viewed as a network coupling the feeding waveguide with the half-space region. If the equivalent circuit parameters of the slot are known, the far spherical mode voltages and currents for a given waveguide excitation and termination may be computed by simple network calculations, and the radiated fields determined directly by modal synthesis.

The specific configuration considered here is a symmetric rectangular E-plane slot radiating from the broad face of a rectangular waveguide propagating the dominant mode (Fig. 1). From field-theoretic considerations it can be shown

that the approximate four terminal-pair equivalent circuit for the slot is that shown in Fig. 2. The values of the lossless network parameters in Fig. 2 can be expressed in terms of those of the lossy two terminal-pair equivalent circuit in Fig. 3, where the slot is considered as a lossy, radiating structure in the waveguide (see Eqs. (1)). The latter circuit may be obtained either theoretically (by a variational procedure) or from measurements in the rectangular guide. The network in Fig. 2 reduces to that in Fig. 3 when a matched termination (free space characteristic impedance: $R = \lambda/\lambda_g$, relative to input guide) is connected to terminal pairs 3 and 4. With the ratio of the currents I_3/I_4 calculated for a given load, modal synthesis yields the simple expressions for the gain pattern given in Eqs. (2).

Although the above theory is valid for small slots its most useful application is for larger slots which can radiate a substantial amount of power. Measurements taken on a large slot terminated in a variable reactance have shown reasonable agreement with theory.

Computation of Network Parameters

$$\frac{1}{X_1} = \frac{X_a}{R_a^2 + X_a^2} + \frac{2}{X_b} \quad (1a)$$

$$2 X_2 = X_a - X_1 \quad (1b)$$

$$n_1^2 = \frac{2 \lambda_g/\lambda}{\frac{2}{R_b} + \frac{R_a}{R_a^2 + X_a^2}} \quad (1c)$$

$$n_2^2 = \frac{2 \lambda/\lambda_g}{R_a} \quad (1d)$$

λ = free space wavelength, λ_g = guide wavelength. All network parameters are normalized to the characteristic impedance of the rectangular waveguide.

Computation of Gain Patterns

$$\frac{|rH|^2}{A} = \cos^2 \theta + \left| \frac{I_3}{I_4} \right|^2 \sin^2 \theta, \quad y = 0 \text{ plane} \quad (2a)$$

$$= 1 + \left| \frac{I_3}{I_4} \right|^2 \sin^2 \theta + 2 \operatorname{Re} \left(\frac{I_3}{I_4} \right) \sin \theta, \quad y \geq 0, \quad x = 0 \text{ plane} \quad (2b)$$

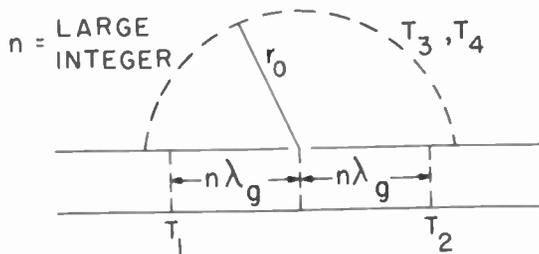
\underline{H} = magnetic field, r = radial distance from slot center ($r/\lambda \gg 1$), θ = polar angle measured from z axis (see Fig. 1b), A = constant parameter.

Acknowledgment

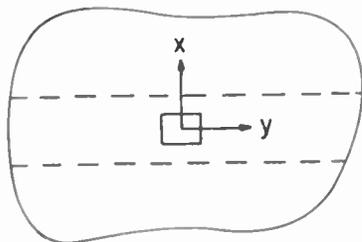
This work was sponsored by the Air Force Cambridge Research Center under Contract No. AF-19(604)-890.

References

1. See L. B. Felsen and N. Marcuvitz, "Slot Coupling of Rectangular and Spherical Waveguides", *Journal of Applied Physics*, Vol. 24, June 1953, pp. 755-770.



(a) SIDE VIEW



(b) TOP VIEW

FIG. 1 SLOT CONFIGURATION

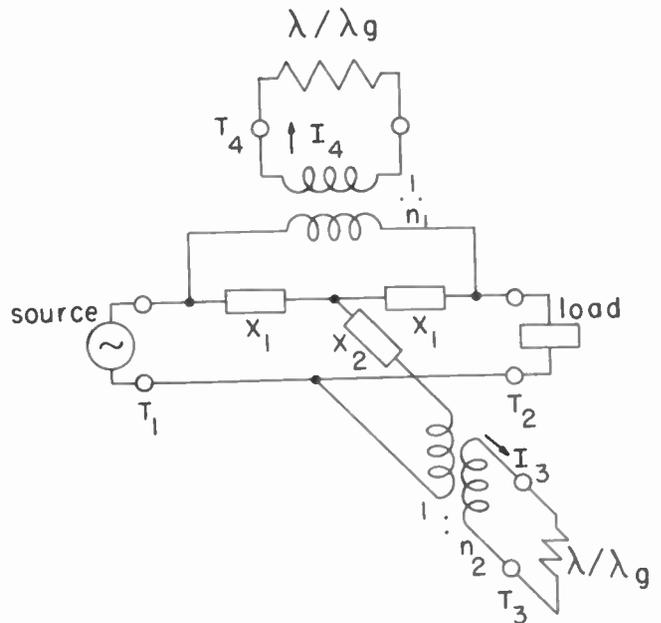


FIG. 2 FOUR TERMINAL-PAIR EQUIVALENT NETWORK

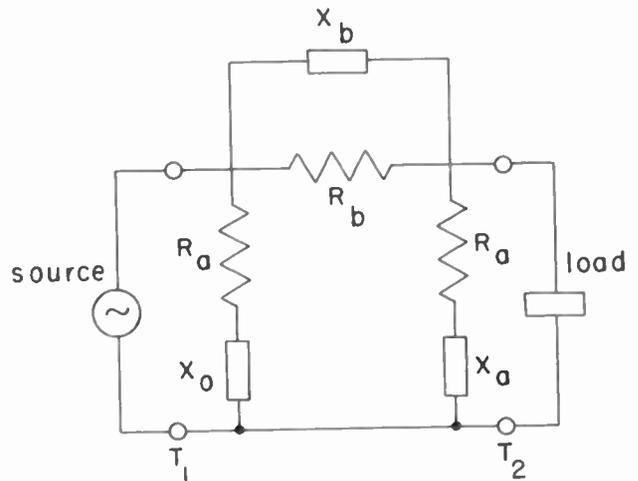


FIG. 3 TWO TERMINAL-PAIR EQUIVALENT NETWORK

A FOUR SLOT CYLINDRICAL ANTENNA FOR VOR SERVICE

Andrew Alford and R. M. Sprague
Andrew Alford Consulting Engineers
Boston, Massachusetts

Summary. This antenna was originally developed for the Air Navigation Development Board under a U. S. Navy contract. Four equally spaced longitudinal slots in the wall of a cylindrical cavity constitute the radiating elements. The four slots are energized by three independent feeders in such a way that one circular radiation pattern and two figure of eight patterns are obtained. When the antenna is used in conjunction with a goniometer, a rotating limaçon pattern results. Intercoupling among the feeders is very low. The antenna is 7 feet high, 15 inches in diameter and weighs 110 pounds. The antenna may be readily tuned with the aid of a calibration chart to any frequency between 108 mc. and 118 mc. Experience with a number of installations shows that the overall course errors are around $\angle 1.5^\circ$ although the antenna error itself can be kept down to about one half of this value. The design of the antenna, flight tests and the several sources of course errors are discussed in detail.

Introduction

A Visual Omnidirectional Range (VOR) system affords an aircraft a direct reading visual indication of the true bearing of the range station as seen from the aircraft. This is accomplished by transmitting two signals, one whose phase does not vary with azimuth, and a second whose phase varies linearly with the azimuth angle. From a comparison of the relative phase of these two signals, the true bearing can be determined.

In present day VOR stations these two signals are transmitted on the same carrier frequency which is usually between 112 mc. and 118 mc. The constant phase signal is generated by amplitude modulating the carrier by a 9960 cycle subcarrier, which in turn is frequency modulated by a thirty-cycle signal. The carrier, modulated in this manner, is radiated equally in all directions of the azimuth. This thirty-cycle signal, as received by a double detection (AM-FM) receiver, is in the same phase at all points of azimuth. The variable phase is generated by radiating a rotating figure-of-eight pattern. The RF phase in one lobe of this pattern is the same as that of the carrier. The RF phase in the second lobe of the figure-of-eight pattern is opposite to that of the carrier. If each lobe of the pattern is a true circle and the pattern is rotating at a thirty-cycle per second rate, the carrier as received in an aircraft will be effectively amplitude modulated at a thirty-cycle rate and the phase of this thirty-cycle will vary linearly with the azimuth angle.

The rotating figure-of-eight pattern can be generated in two ways:

- (a) An antenna radiating a figure-of-eight pattern can be rotated at 1800 RPM, or
- (b) Two stationary figure-of-eight patterns at right angles to each other can be modulated in

time quadrature at thirty cycles by means of a goniometer, so that a single rotating figure-of-eight pattern results. The present antenna is of the latter type.

In the development of this VOR antenna, the following points were considered to be particularly important:

- (a) The antenna should radiate only pure horizontal polarization, so that bearing errors and attitude effects introduced by even small amounts of vertical polarization would be avoided.
- (b) The antenna should be inherently accurate, possessing primary patterns which would result in small bearing errors.
- (c) The antenna should be a compact unit that can be shipped in a crate, bolted in place, and placed into operation by connecting three cables.
- (d) The antenna should preferably contain no rotating parts.

Fig. 1 shows two views of the antenna, the first, with the protective radomes in place; and the second, with the radomes removed.

General Design

Method of Feed

Metal cylinders with one or more longitudinal slots have been used in the past to provide several types of radiation patterns. If a potential is applied across a slot by means of a coaxial line whose inner and outer conductors are connected to the opposite side of the slot, currents flow around the slot. When the slot is relatively narrow in terms of the wavelength, vertically polarized radiation caused by vertical components of the currents substantially cancels, while horizontally polarized radiation results from the horizontal currents across the top and bottom of the slot.

When there are two slots on opposite sides of a cylinder and both are similarly excited but in phase opposition to each other, a figure-of-eight pattern is radiated. When there are four equally spaced slots, two figure-of-eight patterns at right angles to each other can be had by exciting alternately, now one pair of opposite slots, then the other pair of opposite slots. If all four slots are excited so that the horizontal currents associated with all four, are in the same direction a pattern omnidirectional in azimuth results.

A sketch of a four-slot antenna is shown in Fig. 2, View A. Four slots 1, 2, 3, and 4, are cut in the cylinder, equally spaced around the circumference. The diameter of the cylinder is chosen to be approximately 0.15 wavelengths, as

the best compromise between two factors. A cylinder of too large a diameter results in a deviation of the lobes of the figure of eight patterns from true circles. A cylinder of too small a diameter would reduce the radiation resistance of the slots, making impedance matching difficult.

The feeding method is depicted in view B of Fig. 2, in which a developed or spread open view of the interior of the cylinder is shown. Four 200 ohm open wire lines 11, 12, 13, 14, using the inner surfaces of the cylinder as ground return, are terminated across slots 1, 2, 3, and 4, each in the same direction. Approximately a quarter wave from the slots, these four lines are connected to a single 50 ohm coaxial line 21. Power fed to line 21 will then excite each of the four slots equally, producing a continuous field around the cylinder, resulting in an omnidirectional radiation pattern. Substantially no reflection is introduced at the junction of line 21 with lines 11, 12, 13 and 14.

Two 100 ohm coaxial lines 15 and 17 are terminated in opposite directions across slots 1 and 3. Approximately a quarter wave from the slots, these two lines are connected to a single 50 ohm coaxial line 19. Power fed to line 19 excites slots 1 and 3 in phase opposition, resulting in a figure-of-eight radiation pattern with nulls in the directions of slots 2 and 4. In a similar manner, slots 2 and 4 are fed by lines 16 and 18, producing a second figure-of-eight pattern displaced ninety degrees from the first. Essentially no reflection is introduced at the junction of line 19 with lines 15 and 17, or line 20 with lines 16 and 18.

The feeding arrangement is such that no bridges or other isolating means are required to eliminate cross talk among the three sets of feeders. A potential applied across line 19 results in equal but opposite potentials across lines 11 and 13. The waves started by these potentials along lines 11 and 13 cancel at the junction of lines 11, 13 on the line 21. Therefore, assuming perfect symmetry no signal can be transmitted from line 19 into line 21. Similarly, no signal is transmitted from line 20 into line 21. By the law of reciprocity no signal is transmitted from 21 into 19 or into 20.

The two sideband feeders 19 and 20 are also isolated from each other. When slots 1 and 3 are energized by a potential applied to line 19, equal potentials of the same sign appear on both sides of slot 2. Similarly, the potentials appearing across slot 4 are also equal and of the same sign. Therefore, no potentials are produced across lines 16 and 17, so that no signal is transmitted from line 19 into line 20 or vice versa.

The length of each of the branch feeders 15, 16, 17, and 18 is approximately a quarter wavelength at the mean frequency in order that the lines will present a high impedance across the slots to potentials from feeder 21. Similarly, each of the lines 11, 12, 13, and 14 is a quarter

wave length long at the mean frequency in order that all lines will present a high impedance across the slots to potentials from lines 19 and 20.

The above described method of slot excitation, while well suited to an antenna of this type, is not the only one that could be used. For instance, in Fig. 3, slots 1, 2, 3, and 4 are fed by only one line per slot, respectively by lines 11, 12, 13 and 14. Lines 11 and 13 are connected to terminals 21 and 23 of a coaxial bridge 5. Lines 12 and 14 are connected to terminals 22 and 24 of a similar coaxial bridge 6.

When a potential is applied to terminal 25 of coaxial bridge 5 two equal but opposite potentials are produced at terminals 21 and 23. Slots 1 and 3 are then excited in phase opposition and a figure-of-eight pattern is produced. When a potential is applied to terminal 27 of Bridge 5 equal potentials in the same relative phase are produced at terminals 21 and 23. Slots 1 and 3 are then excited the same relative phase. Bridge 6 has similar properties. Terminals 27 and 28 are connected to the reference feeder. When a potential is applied to this feeder all four slots are excited in the same relative phase resulting in a circular radiation pattern. Terminals 25 and 26 are connected to the sideband output terminals of the goniometer.

The Two Modes in the Cavity

Because of radiation through the slots, the inside of the cylinder behaves like a lossy cavity in which two different types of modes can be excited. The mode of, say, Type I is excited by either of the two sideband feeders. The mode of Type II is excited by the reference feeders. When the reference feeder is energized the lines of electric field are distributed as shown in view A of Fig. 4. The cylinder is then in effect divided into four equal parts indicated by the imaginary partitions shown in view A of Fig. 4 by the dotted lines. The cross sectional area of the effective cavity back of each slot is therefore one quarter of the cross sectional area of the cylinder. When one of the sideband feeders is energized the electric lines of force are distributed as shown in view B of Fig. 4. In this case, the effective cavity behind each slot is much larger than in the previous case. Therefore, the "cut off frequency" for the Type II mode shown in view A of Fig. 4 is substantially higher than the "cut off frequency" for the Type I mode shown in view B. If a lossy cavity is operated near or below its "cut off frequency" the distribution of voltage v along a slot is of type $v = v_0 e^{\alpha x} \sin(\beta x + \gamma)$ where x is the distance along a slot and α, β, γ are constants. A voltage distribution of this kind is illustrated by view C in Fig. 4. When a cavity of this type is operated well above its cut off frequency the voltage distribution is approximately sinusoidal as shown in view D of Fig. 4. The dimensions of the antenna are such that the cut off frequency of the Type II mode must be artificially lowered to avoid exponential distribution of the reference

signal along the slots. This latter condition would be undesirable for the following reasons:

(a) With an exponential or a semi-exponential distribution of voltage along the slots the radiation resistance is low making the impedance too frequency sensitive.

(b) Unless both the reference and the sideband feeders are connected at the exact centers of the slots, the center of radiation for the sideband signal is not at the same height as the center of radiation for the reference signal, because the sideband distribution along the slot is approximately sinusoidal with the voltage maximum at the center of the slot, whereas the reference voltage distribution is unsymmetrical.

Antenna Tuning

If four short circuiting bars were symmetrically placed across each of the four slots at some distance from say the bottoms of the slots, the electrical length of each of the slots would be shortened by equal amounts, and the slots would be "tuned" to a higher frequency. If, instead of short circuiting bars, equal finite inductances were placed across the slots, the tuning frequency would also be raised, to an extent depending on the values of the inductances. If, in lieu of inductances, equal finite capacities were placed across the slots, the electrical slot length would be increased and the tuning frequency would be lowered. Thus several means are available to tune the antenna over an appreciable frequency range.

The tuning of each slot separately would involve a lengthy operation, and would result in pattern dissymmetry unless relatively complex measurements were used to control the operation. A method of tuning all slots simultaneously and symmetrically, for both reference feeders and sideband feeders, is shown in the schematic diagram of Fig. 6. By cross connecting opposite pairs of slots, the impedance placed between the opposite pairs is effectively placed across each slot. Cross connections of this type are made both at the top as well as at the bottom of the loading fins. The two cross connecting means will be referred to as "tuning bridges" or simply "bridges". Both views in Fig. 6 are intended to show the same set of fins cross connected by the same tuning bridge. View A indicates the polarities of the potentials which exist when only the sideband feeders are energized. View B shows the distribution of potentials when only the reference feeder is energized. In both cases the plus and the minus signs refer to the polarities of the potentials. The numbers are used to identify the fins.

When the sideband feeders are energized, the junctions of the cross connecting conductors, shown as circles in the diagram, are at zero potential. The lengths of the cross connecting conductors determines the value of the inductance shunted across the slots of opposite polarities. The length of these cross connecting conductors can be varied by sliding the cross connecting assembly

(bridge) up or down. Slide bar extensions to the fins are provided upon which the bridge rides.

When the reference feeder is energized, the lengths of the cross connecting conductors is immaterial because they cross connect fins at equal potentials. The capacitance between the two sets of the cross connecting conductors (capacitance between the two concentric circles in the diagram) is the parameter that can be used to control the impedance presented to the reference feeders. Thus, there are two independent adjustments, one which has an effect only on the impedance presented to the reference feeder. One adjustment is made by translating the tuning bridge. The other adjustment is accomplished by moving a plunger which varies the capacitance between the two sets of cross connecting conductors in the bridge. The nature of these impedance controls is such they allow one to tune the antenna to any frequency within its operating range without disturbing the symmetry of the radiation pattern.

The lower and the upper bridges are provided with graduate scales which enable one to clamp the bridges at various positions which correspond to different frequencies. A calibration chart provided for each individual antenna gives the proper settings of the bridges for any frequency between 108 mc. and 118 mc. The variable condenser consists of a stationary cylinder and a movable concentric metal plunger. The capacitance can be varied by moving the plunger. The upper as well as the lower bridge is provided with a variable capacitor of this type. Both plungers are graduated. The plunger in the upper bridge is adjusted only when it is desired to change from the 108 mc. - 112 mc. band to 112 mc. - 118 mc. band or vice versa. The calibration chart gives proper settings of the plungers for the frequencies in operating range of the antenna. Experience shows that the retuning of the antenna from one frequency to another can be done in something less than an hour. This operation may be performed in the field by relatively unskilled personnel. It is not necessary to make either impedance or pattern measurements.

When the slots are tuned by means of the condenser in the lower bridge, the impedance presented to the feeder at the junction of the four open wire lines is around $45 \angle j 12.5$ ohms. Therefore, a small shunt capacity of the proper value can be used to match the reference feeder. The required value of this capacity remains sufficiently constant over each half of the operating band. Standing wave along the four individual branch lines feeding the slots does not exceed 1.25 to 1.

The four coaxial lines feeding the slots with the sideband power are not connected directly across the slots. On the contrary, the inner conductors are terminated in fixed series inductances. The value of these inductances is such, that when the slots are correctly tuned by the lower bridge, the impedance at junction of each pair of slot feeders is around $75 \angle j 0$ ohms.

A small shunt capacity of the proper value located three-eighths of a wavelength from this junction toward the generator, is used to match the feeder to the 52.5 ohm line. Five values of this capacity are required to cover the frequency band of 108 to 118 megacycles. It is noted that the matching problem is relatively minor because the feeding arrangement is such that without any matching the SWR even along the branch feeders is under 1.5.

Details of Construction

In Figures 7, 8 and 9 are shown out away drawings of the antenna. Fig. 7 shows the lower third, Fig. 8 the center third, and Fig. 9 the upper third. In the lower section, the supporting structure can be seen. The mast A-7215 passes through the hubs of two heavy wheel like castings separated by the outer shell A-7214. The two sideband cables W-7201 and W-7203 from the goniometer, enter the antenna through the supporting mast and terminate in Tees P-7201 and P-7202. To these Tees are connected matching stubs W-7209 and W-7218. To the same Tees are connected cables W-7212 and W-7221. Cable W-7212 feeds coaxial lines W-7214 and W-7215 which are connected to one pair of opposite slots. Cable W-7221 feeds coaxial lines W-7223 and W-7224 which are connected to the other pair of slots. Cable W-7206 is the reference signal cable. It is normally connected to the modulation eliminator. This cable is carried up inside the cylinder between the slots and is connected to the four individual feeders supplying the reference signal to the four slots. For mechanical strength, four 1/8 inch thick fiber glass strips, similar to A-7209, are bolted across the slots.

In Fig. 8 can be seen the sideband feeders W-7214, W-7215, W-7223 and W-7224. The inner conductors of these lines terminate in the series inductance sections W-7216, W-7217, W-7225 and W-7226. The four sets of loading fins are visible in this drawing. Extending downward from the bottom of the loading fins are the eight calibrated slide bars E-7207 through E-7214 on which the tuning bridge E-7204 rides. When this bridge is moved downward the inductance across the slots is increased. The calibrated coaxial condenser C-7203 is used to adjust the impedance presented to the reference signal feeder.

Four heaters such as HR-7202 and HR-7203 are arranged symmetrically close to the inside wall of the cylinder. They provide up to 1250 watts dissipation for sleet melting. Trimmer condensers C-7204, C-7205, C-7206 and C-7207 across the center of each set of fins are used to compensate for manufacturing tolerances in the fin spacings. They are adjusted at the factory and locked in place.

In Fig. 9 can be seen the main reference cable W-7206, the stub W-7227, and the four "open wire feeders" W-7228, W-7229, W-7230 and W-7231, each terminated across one of the slots. The upper bridge, C-7207, used for changing from the (108 mc. - 112 mc.) band to the (112 mc. - 118 mc.) band is also shown. The top casting provides a support for a DME.

The antenna is designed for 40 G vertical shock, and winds up to 150 miles per hour. The coaxial lines are hermetically sealed. All joints in the upper 2/3 of the antenna structure are also sealed. Gasketed radomes are provided over each of the four slot openings.

The seams in the shells of the early models of the antenna were not sealed and drain holes were provided to dispose of the small amount of water that might be driven through the seams by wind. Experience showed that when water was driven in, some of it tended to cling in the form of droplets to the inside surfaces of the loading fins, causing an increase in capacitance between pairs of fins. This change in capacitance resulted in bearing errors sometimes as high as $\pm 3^\circ$. By sealing the seams in the upper 2/3 of the antenna this trouble was eliminated.

While heaters are provided for sleet removal, experiments show that about 1/2 inch of ice, even when deposited on only one side of the antenna, produces a bearing error of about 0.4° . A 1/4 inch coating accumulated during a sleet storm resulted in something less than 0.2° .

Approximately 125 watts of heater power is used continuously in Liberia in order to retard growth of fungus. This is believed to be in accordance with the now more or less standard practice followed in tropical climates.

Antenna Adjustments

Upon assembly, the antenna is accurately balanced by means of the trimmer condensers installed in the loading fins. These trimmers are adjusted so that the nulls of the two figure-of-eight patterns are separated by 180 degrees. The input impedances looking into the sideband feeders are then measured at several frequencies in the frequency range in order to check the fundamental symmetry of the antenna. Radiation patterns are also taken at these frequencies.

The second operation is the tuning calibration. The antenna is checked over the entire 108 to 118 megacycle band, being tuned to unity standing wave ratio at each of the test frequencies. From this data is drawn a calibration curve. Between the calibration points the standing wave ratio does not exceed 1.2 and is usually less than 1.1.

After the calibration has been completed, the antenna is ready to be installed in the field with no further operations than bolting in place, connecting three cables, and phasing of the carrier feeder.

In Fig. 10 is reproduced a pattern as recorded on a polar recorder. Each lobe is just ninety degrees wide at the three db points, and one hundred twenty degrees wide at the 6 db points. This is as it should be for true circles. The power levels of the four lobes are equal within $\pm .1$ db. The intersections of the figure-of-eight

patterns are spaced by ninety degrees $\pm 1/2^\circ$. The reference pattern deviates from a circle by $\pm .2$ db.

In Fig. 11 is a family of figure-of-eight patterns. These patterns were obtained by connecting both sideband feeders to a goniometer, and recording patterns with the goniometer settings of 0, 22.5, 45, 67.5 and 90 degrees. These patterns were obtained with the antenna tuned to the frequency at which the patterns were recorded. A similar family of figure-of-eight patterns shown in Fig. 12 were recorded with the antenna deliberately tuned to a wrong frequency so that a standing wave ratio of 4.7 existed on the sideband feeders. No appreciable increase in bearing error is apparent in spite of the 4.7 SWR along the sideband feeders. This corresponds to a detuning of 1.65 megacycles. Thus, it is apparent that the antenna may be operated over a considerable band without retuning, although such operation is not recommended.

Figure 13 shows the vertically polarized component of radiation. In this figure the larger figure-of-eight is a recording of the normal horizontally polarized radiation. The smaller figure-of-eight pattern shows a record of the vertically polarized component. This record was obtained by turning the receiving antenna through 90° and after increasing the power by 30 db (1000 to 1). The vertical component is 36 db below the horizontal component.

Counterpoise Effects

A study of the effect of the counterpoise was made with the aid of microwave models. The data obtained from this study serves to show that a counterpoise of too small a diameter has the effect of increasing the size of the "cone of confusion" above an antenna. In Fig. 14 is recorded the vertical pattern of the reference radiation of a model antenna mounted so that its center was 0.4 wavelengths above a counterpoise one wavelength in diameter. A deep well occurs above the antenna, as expected. In Fig. 15 is recorded the vertical pattern of the sideband radiation under the same conditions as in Fig. 14. Here, radiation at 90° above the horizon is only 10 db below the maximum of the lobe. The counterpoise seems to be radiating under these conditions, the amount depending on the proximity of the counterpoise to the antenna. When the counterpoise is placed even closer to the center of the antenna. The sideband radiation directly upwards may be only several decibels below the main lobe.

Currents are induced in the counterpoise by the reference radiation are circular. If the counterpoise is also circular, no current paths are interrupted, regardless of the size of the counterpoise. The radiation pattern of circular currents has a well along the axis of symmetry.

Tests were conducted to determine the counterpoise size for substantially no sideband radiation at 90° above the horizon. Fig. 16 shows the radiation pattern in the vertical plane for the sideband field with the counterpoise increased

to two wavelengths in diameter. There is no field directly above the antenna. With a counterpoise 1.5 wavelengths in diameter, radiation straight up is still very low. When the antenna is to be used in conjunction with a counterpoise, it is recommended that it be placed 0.4 wavelengths (from its center) above a counterpoise at least 1.5 and preferable two wavelengths in diameter. Under such conditions, the "cone of confusion" should be very small.

Flight tests made at Friendship Airport, at Baltimore, Maryland showed that the antenna mounted above a round counterpoise 9 feet in diameter (approximately one wavelength) produced a cone of confusion which was only a fraction of the width of the cone produced by a standard CAA 5 loop VOR near Baltimore.

Some installations in which a counterpoise 35 feet in diameter was used resulted in what seemed to be a complete absence of the cone of confusion at least when flown in some aircraft. There is also some evidence that a VOR installation which has no cone of confusion may appear to have one when flown in an aeroplane equipped with some types of receiving antenna installations.

The field from the antenna is very small in comparison with the field in the directions along the ground. It is, therefore, likely that objects or irregularities of ground in the vicinity of the antenna may reflect sufficient field in the upward direction to override the very small primary signal at high angles above the horizon. The result of such reflected signals would be to produce a cone of confusion that could not be reproduced if the VOR installation were moved to another spot at the same airport or to some other site.

Flight Tests

One of the earliest models of this antenna was lent by the Air Navigation Board to the Air Transport Association for flight tests at Friendship Airport, Baltimore, Maryland. A demonstration of experimental TVOR equipment, using this antenna, was held under the auspices of the Air Transport Association on March 5, 6 and 7, 1952. Fig. 17 is a photograph of the experimental installation being prepared for the tests.

Atlantic Division representatives of the CAA made theodolite controlled flights, with Mr. George Luecker as the observer. Five flights were made around the station, all at a constant distance of five miles, at various altitudes corresponding to vertical angles above the horizontal ranging from 3.26 degrees to 12.79 degrees. On Fig. 18 are recorded three representative flights. The three different flight altitudes correspond to vertical angles of 3.26, 8.10, and 12.79 degrees. A portion of the last flight record was interrupted in the vicinity of 100 degrees because of poor visibility. The error curves for all three plots track quite well, and each shows errors around plus or minus one and one-half degrees.

The predominant error is of the single cycle type which, as explained in the appendix to this paper, is suspected to be due to 30 cps amplitude modulation of the reference signal caused by small non-linearity in the experimental transmitting equipment.

Flight checks of later installations made mostly at airports confirm the early data obtained at Baltimore. Although later antennas were substantially more accurate than the early models, the overall bearing errors observed in flight check remained about the same (except in a few cases where $\pm 1^\circ$ bearing errors were observed). It is likely that a substantial portion of the overall error measured in a flight test is caused by factors independent of the antenna. For example site errors or single cycle errors due to 30 cps modulation of the reference signal, goniometer errors and others may well mask the bearing errors due to the antenna itself.

In all flight checks the aircraft attitude effects were found to be very small (approximately the thickness of the needle as observed on the standard cross pointer instrument) showing that the vertical component of the field radiated by the antenna is sufficiently low in comparison with the horizontal component.

APPENDIX

Some Sources of Bearing Errors

The following discussion is believed to be applicable to VOR Stations in general, not only to those using the antenna described in this paper. It is noted that most of this discussion is equally applicable to antennas using rotating dipoles.

(1) Deviation of the primary pattern from a figure of eight having perfectly circular lobes.

The magnitude of the bearing error introduced by a given deviation of the lobes of the figure-of-eight patterns from true circles can be calculated as follows:

Let the E-W figure-of-eight pattern be described by function $\phi(\theta)$. Then the N-S figure-of-eight will be given by $\phi(90-\theta)$.

The signal at time t in direction θ is

$$S = \phi(\theta) \cos \omega t + \phi(90^\circ - \theta) \sin \omega t$$

$$= \sqrt{\phi^2(\theta) + \phi^2(90^\circ - \theta)} \cos(\omega t - \beta)$$

where β the bearing error.

$$\tan \beta = \frac{\phi(90^\circ - \theta)}{\phi(\theta)}$$

In the ideal case when the lobes of the figure-of-eight patterns are perfect circles

$$\begin{aligned} \phi(\theta) &= \cos \theta \\ \phi(90^\circ - \theta) &= \sin \theta \end{aligned}$$

so that the bearing error is zero.

In the general case when the lobes are not perfect circles the error vs. azimuth curve has four complete cycles.

(2) Non-Linearity in Transmitting Equipment.

One-cycle errors can be produced by the presence of thirty-cycle amplitude modulation of the carrier signal. Such modulation changes the phase of the space modulation produced by the rotating figure-of-eight in such a manner as to cause one-cycle error. The original source of this thirty-cycle modulation of the carrier is probably the tone wheel or its equivalent. The 9960-cycle frequency modulated subcarrier has been observed to be also amplitude modulated by as much as five to ten percent at a thirty-cycle rate. If a slight non-linearity exists in the transmitting equipment, partial rectification takes place resulting in thirty-cycle amplitude modulation of the reference field.

Let the reference field be

$$S_R = \cos \omega t [1 + \sigma \cos(\Delta t + \psi)]$$

where σ = modulation factor due to the spurious 30 cycle signal

$$\Delta = \text{modulation frequency} \times 2\pi$$

$$\psi = \text{modulation phase (could have almost any value)}$$

The sideband field from one figure-of-eight at azimuth θ is

$$S_1 = m \cos \omega t \cos \theta \cos \Delta t$$

where m = modulation factor (usually 0.3)

$$\theta = \text{azimuth angle}$$

The sideband field from the second figure-of-eight is

$$S_2 = m \cos \omega t \cdot \cos \theta \sin \Delta t$$

The total signal T at azimuth angle θ is then

$$\begin{aligned} T &= S_R + S_1 + S_2 \\ &= \cos \omega t [1 + \sigma \cos(\Delta t + \psi) + m \cos(\theta - \Delta t)] \end{aligned}$$

The envelope E of T is then

$$E = 1 + \sigma \cos(\Delta t + \psi) + m \cos(\theta - \Delta t)$$

This expression may be transformed into the following form

$$E = 1 + \sqrt{a^2 + b^2} \cos(\theta - \Delta t - \beta)$$

where $a = \sigma \cos(\theta + \psi) + m$

$$b = \sigma \sin(\theta + \psi)$$

and the bearing error β is given by

$$\tan \beta = \frac{\sin(\theta + \psi)}{\cos(\theta + \psi) + \frac{m}{\sigma}}$$

The maximum value of error β is at the azimuth angles which results in

$$\cos(\theta + \psi) = -\frac{\sigma}{m}$$

The maximum value of the bearing error is given by

$$\tan \beta_{max} = \frac{\sqrt{1 - \left(\frac{\sigma}{m}\right)^2}}{-\frac{\sigma}{m} + \frac{m}{\sigma}} \approx \frac{\sigma}{m}$$

Example: Let $m = 0.3$ which is the usual value. Assume that the undesirable amplitude modulation of the carrier by the spurious 30 cps is 1.5%. Then

$$\sigma = .015$$

A value of 6 degrees is the bearing error obtained under such conditions. It is noted that a bearing error of this type would be obtained with an absolutely perfect antenna. Single cycle errors of this origin are often assumed to be caused by antennas.

(3) Inequality of lobes in one of the two figure-of-eight patterns.

In Figure 19 are shown two figure-of-eight patterns. The maxima of the two lobes in, say, the North-South pattern are equal to each other and each is equal to unity. The lobes in East-West pattern are assumed to be unequal; the maximum of one lobe is $1 - \alpha$, the maximum of the other lobe is $1 + \alpha$. The nulls defined in the East-West pattern are at azimuth angles $\angle(90^\circ + \gamma)$ with respect to the line joining the maxima. When α is a small fraction, angle α is also small. This state of affairs is observed when a small fraction of the power supplied to one of the two sideband feeders is radiated in the form of omnidirectional reference pattern because of some lack of symmetry.

Under such conditions at azimuth θ the total side band field F (from both figure-of-eight patterns) is given by

$$F = [(a \cos \theta + b) \cos \Delta t + (a \sin \theta + b) \sin \Delta t] \times \sin \omega t$$

This expression may be transformed into the following form

$$F = F_0 \cos(\theta - \Delta t - \beta)$$

in which β is the bearing error. When this is done it is found that

$$\tan \beta = \frac{\alpha \sin \theta}{1 + \alpha \cos \theta}$$

The maximum value of $\tan \beta$ is

$$\tan \beta_{max} = \frac{\alpha}{\sqrt{1 - \alpha^2}}$$

The polar equation of the figure-of-eight pattern with unequal lobes is

$$h = \alpha + \cos \theta$$

The nulls $h = 0$ occur when $\cos \theta = -\alpha$. If we let $\theta = 90^\circ + \gamma$, it is found that

$$\alpha = \sin \gamma$$

If this value of α in terms of γ is substituted into the expression for the maximum value of β it is found that $\tan \beta = \tan \gamma$ which means that $\beta_{max} = \gamma$. Moreover the maximum value of error occurs at the same azimuths as the nulls of the figure-of-eight pattern with unequal lobes. Error of this kind, when plotted against azimuth results in a single cycle curve.

(4) Inequality of the Two Figure-of-Eight Patterns.

Let it be assumed, as shown in Figure 20, that the amplitude of the North-South figure-of-eight pattern is "a" while the amplitude of the East-West figure-of-eight pattern is "b". Then the sideband field at azimuth θ is

$$F = \cos \omega t (a \cos \theta \cos \Delta t + b \sin \theta \sin \Delta t)$$

This expression may be transformed into the following form

$$F = \beta \cos(\theta - \beta - \Delta t)$$

where β is the bearing error. It may be shown that

$$\tan \beta = \frac{(1 - q) \tan \theta}{(1 + q) \tan^2 \theta}$$

$$\text{where } q = \frac{b}{a}$$

The maximum value of error occurs when

$$\tan \theta = \frac{1}{q}$$

The maximum error is given by

$$\tan \beta_{max} = \frac{1-q}{2\sqrt{q}}$$

(5) Course Error Caused by Mis-phasing of Sidebands.

When the two figure-of-eight patterns are not radiated in the same RF phase, a small bearing error results. Consider the case where the RF phase of one figure-of-eight is delayed ϕ electrical degrees behind the carrier, while the RF phase of the second figure-of-eight is advanced the same amount. The total field T is then

$$T = \sin \omega t + m \sin(\omega t - \phi) \cos \Delta t \cos \theta + m \sin(\omega t + \phi) \sin \Delta t \sin \theta$$

The envelope of T

$$E = \left\{ [1 + \cos \phi \cos(\Delta t - \theta)]^2 + m^2 \sin^2 \phi \cos^2(\Delta t + \theta) \right\}^{\frac{1}{2}}$$

By expanding E into a series it is found that the bearing error β is given by the following approximate expression.

$$\tan \beta \approx \frac{m^2}{\theta} \sin^2 \phi \sin 4\theta$$

It is noted that this is a four cycle error.

Example: assume that the modulation factor $m = 0.3$, and that the total phase error is $2\phi = 18^\circ$. The maximum course error is found to be only 0.016 degrees.

(6) Course Error Caused by Mis-phasing of the Carrier.

If the two figure-of-eight patterns are in the same RF phase, but carrier pattern phase is in error by ϕ electrical degrees, the total field is

$$T = \sin(\omega t + \phi) + m \sin \omega t \cos \Delta t \cos \theta + m \sin \omega t \sin \Delta t \sin \theta$$

The envelope reduces to the form

$$E = [1 + 2m \cos \phi \cos(\theta - \Delta t) + m^2 \cos^2(\theta - \Delta t)]^{\frac{1}{2}}$$

$$E \approx 1 + m \cos \phi \cos(\theta - \Delta t)$$

Terms involving $(1 - \cos \phi)^2$ and higher powers of $(1 - \cos \phi)$ have been neglected. Thus, as long as $(1 - \cos \phi)$ is a very small quantity, which it usually is, there is practically no bearing error. The only effect is a change in the percentage of modulation.

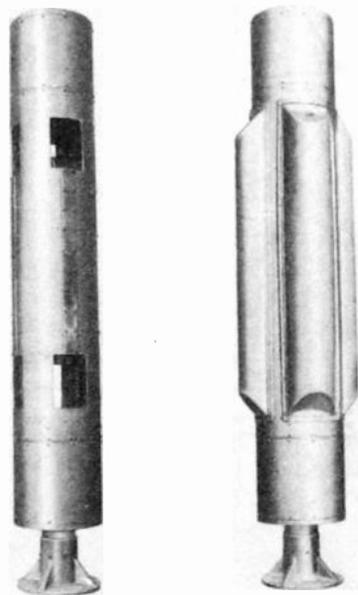


Fig. 1

The four slot VOR antenna with (a) protective radomes removed, and (b) protective radomes in place.

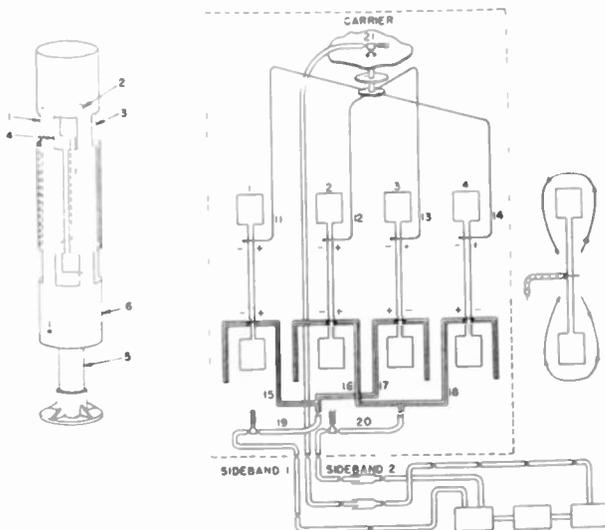


Fig. 2

A spread open view of the four slots, showing the method of exciting the slots for both the reference patterns and the two figure-of-eight patterns.

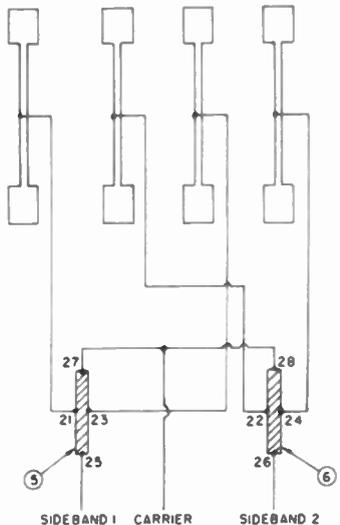


Fig. 3
A spread open view of the four slots,
showing an alternate method of slot
excitation employing bridges.

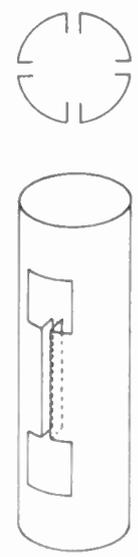


Fig. 5 - A slot loaded with fins.

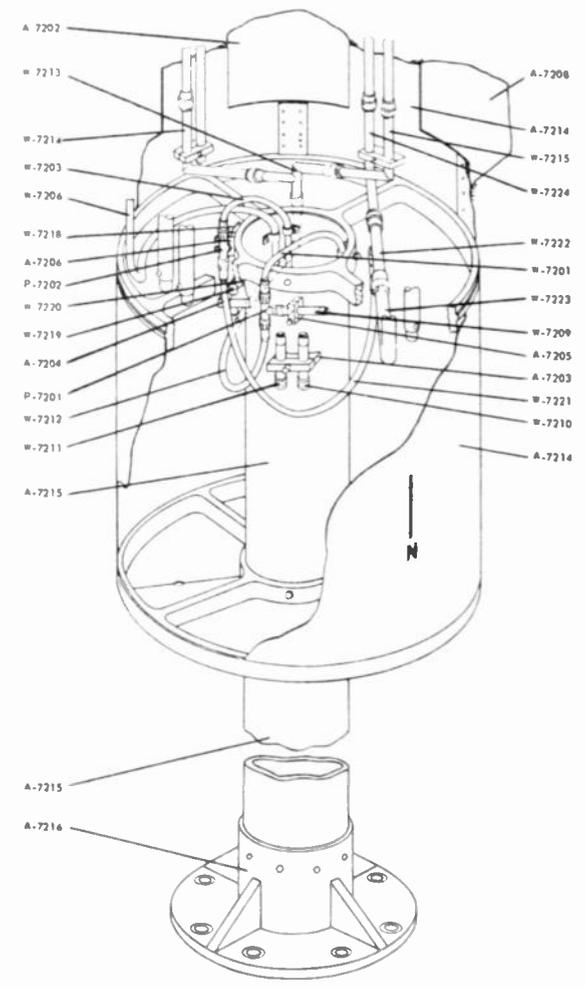


Fig. 7
Cutaway drawing of the lower third
of the antenna.

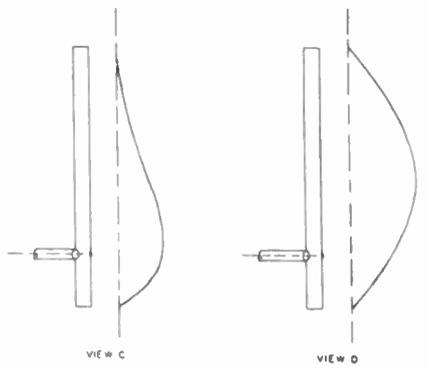
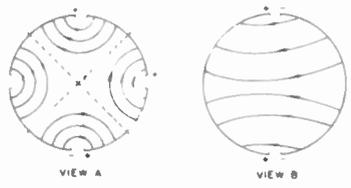


Fig. 4
Cavity modes of type I and type II
and voltage distributions along
a slot.

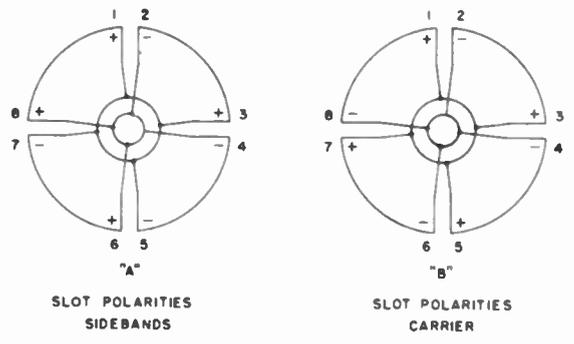


Fig. 6 - Cross connections between fins.

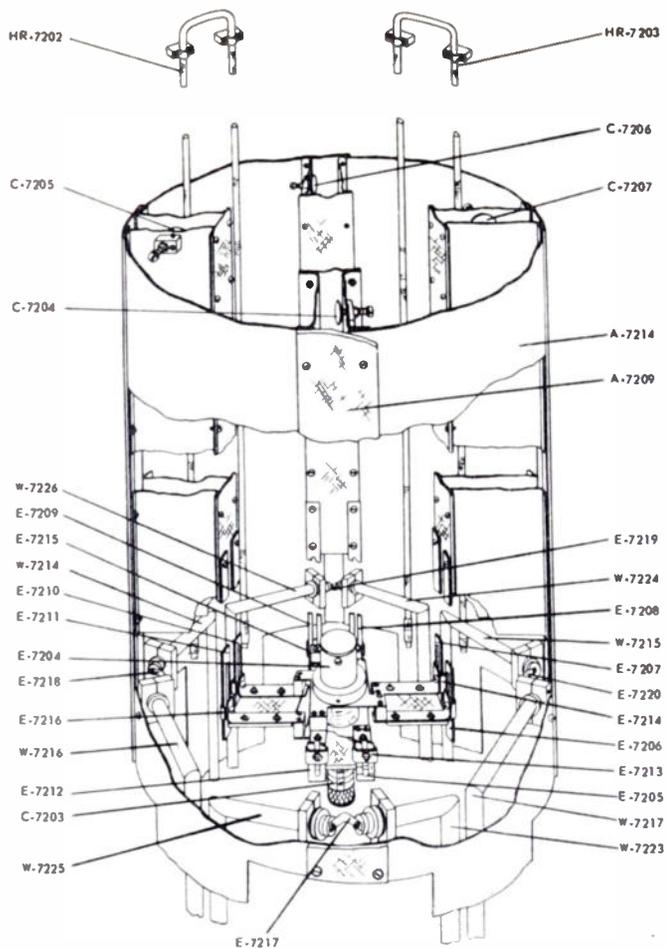


Fig. 8

Cutaway drawing of the center third of the antenna.

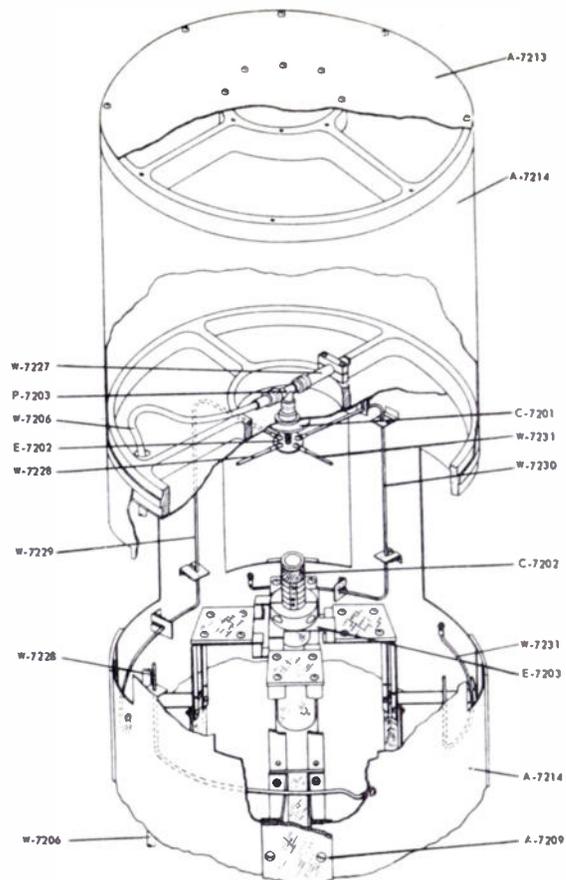


Fig. 9

Cutaway drawing of the upper third of the antenna.

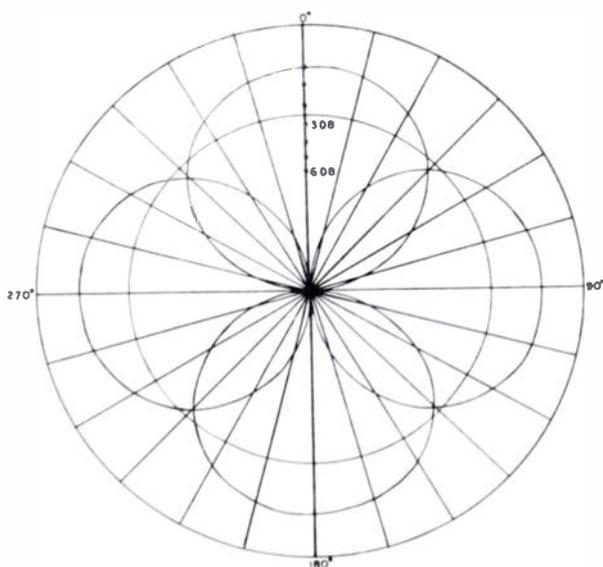


Fig. 10

Primary patterns of the antenna, showing the reference pattern and the two figure-of-eight patterns.

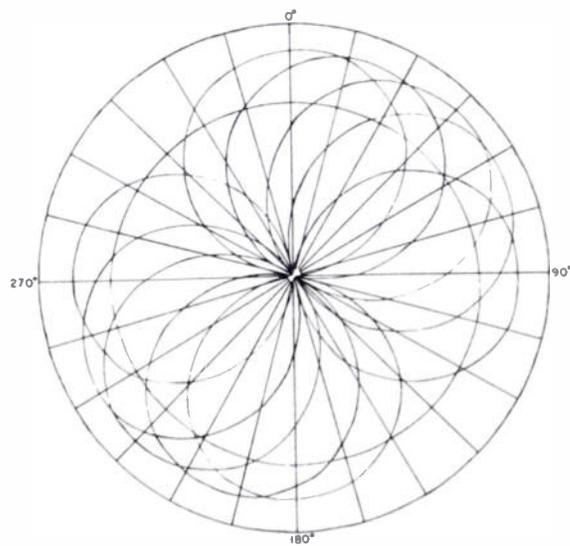


Fig. 11

Antenna patterns obtained with the side band feeders connected to a goniometer set at 22 1/2 degree intervals. Feeders matched.

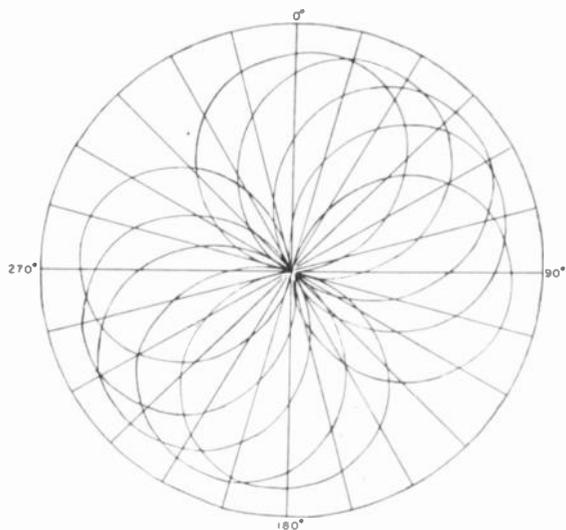


Fig. 12
Antenna patterns obtained with the side band
feeders connected to a goniometer set at 22 1/2
degree intervals. Antenna detuned 1.7 megacycles.

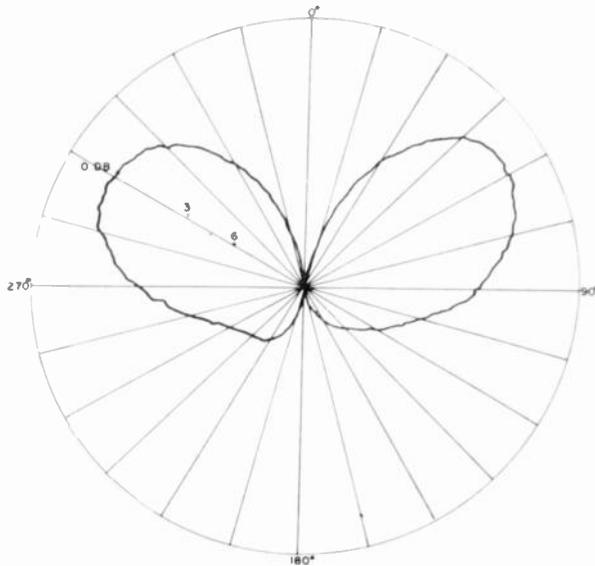


Fig. 14
Vertical plane pattern of the reference field with
the antenna center placed 0.4 wavelengths above a
counterpoise one wavelength in diameter.

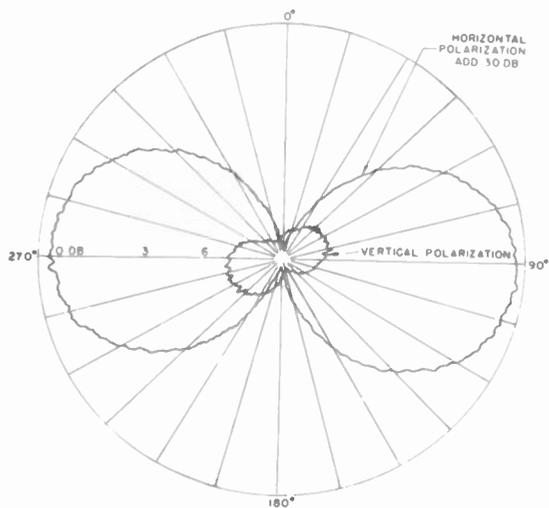


Fig. 13
Vertically polarized component of antenna
radiation.

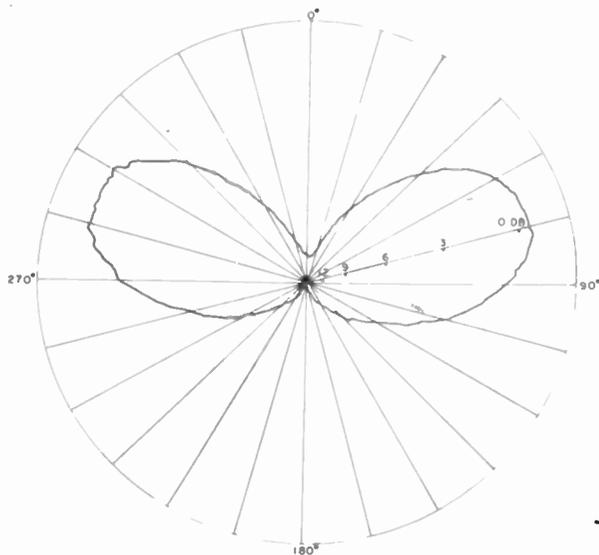


Fig. 15
Vertical plane pattern of the sideband field with
the antenna center placed 0.4 wavelengths above a
counterpoise one wavelength in diameter.

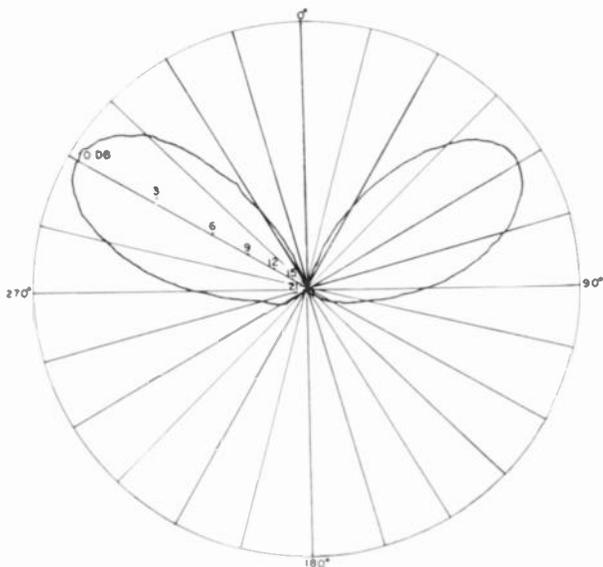


Fig. 16
Vertical plane pattern of the sideband field with the antenna center placed 0.4 wavelengths above a counterpoise two wavelengths in diameter.

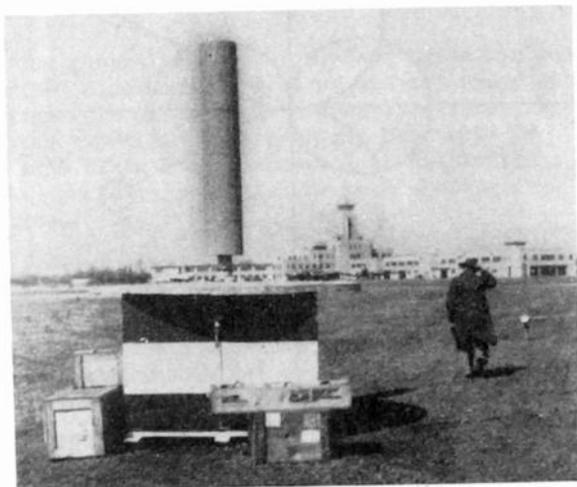


Fig. 17
The experimental TVOR installation at Friendship Airport, Baltimore, Md., March 5, 1952.

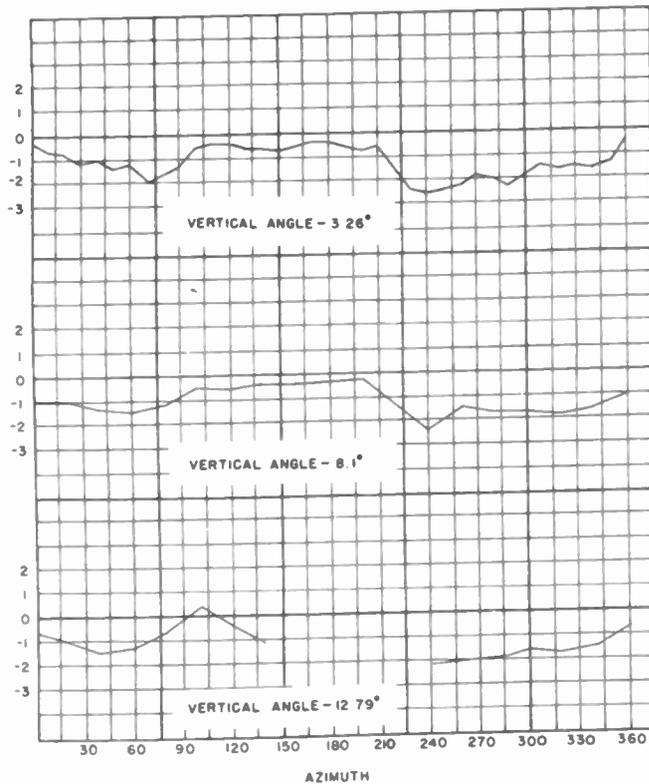


Fig. 18
Theodolite flight checks of the Friendship Airport experimental TVOR installation at three representative vertical angles.

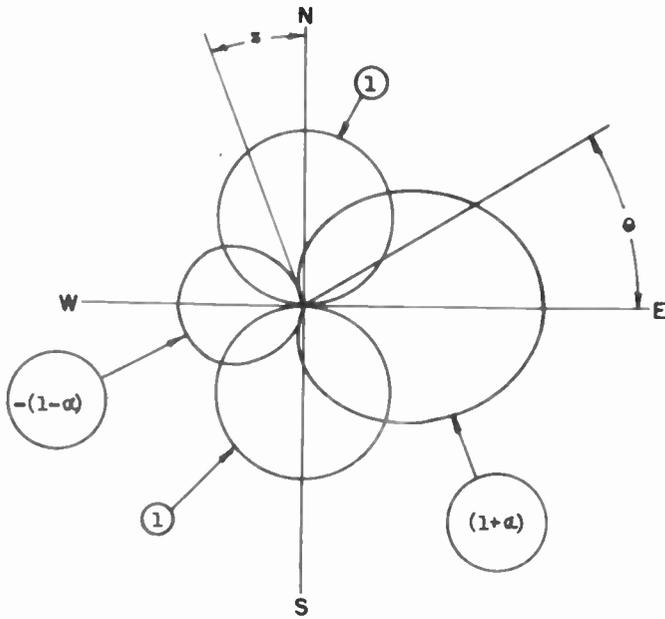


Fig. 19
Course error caused by inequality of the lobes
of one of the two figure-of-eight patterns.

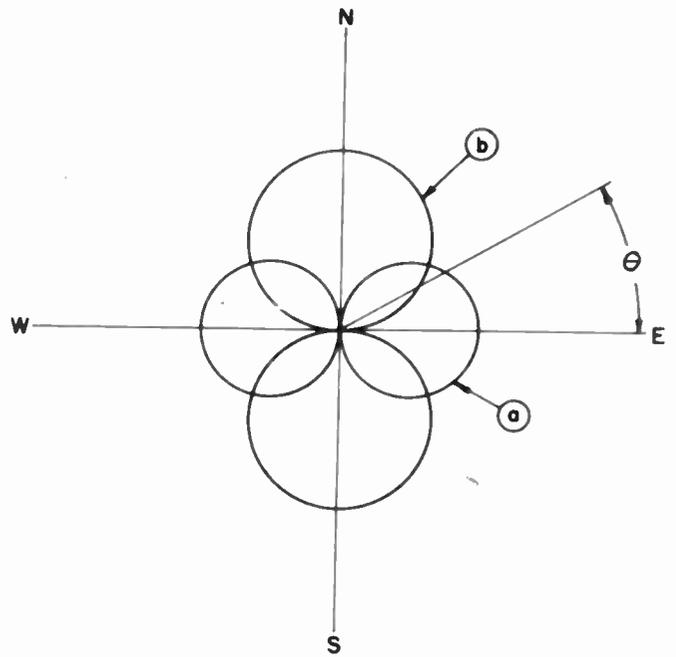


Fig. 20
Course error caused by inequality of the two
figure-of-eight patterns.

TRAPPED WAVE ANTENNAS

Herman Ehrenspeck, Werner Gerbes and Francis J. Zucker

Air Force Cambridge Research Center
Cambridge, Mass.

SUMMARY

The question is first raised as to how a 'trapped' wave radiates, and what beam shapes it can produce. The results are compared with the radiation fields produced by conventional antenna apertures.

A flush-mounted type of antenna is discussed which utilizes waves trapped in single and multiple dielectric layers. The calculated mode characteristics are in very good agreement with experimental results. By means of phase and amplitude control, it should be possible to design such antennas with a great variety of beam shapes in azimuth and elevation.

HOW DOES A TRAPPED WAVE RADIATE?

A surface wave is one which propagates along an interface between two media. If the surface wave is slower than light, it carries most of its energy within a small distance from the interface, and will not radiate unless a discontinuity impedes its progress. We can therefore speak of it as a guided, or 'trapped' wave. Its phase fronts are perpendicular to the interface (see Fig. 1), while the amplitude decays exponentially upward.

The general theory of surface waves has been described elsewhere.¹ Here we are concerned with the manner in which a surface wave radiates, and with the types of pattern it can produce. We will also discuss phase control on dielectric sheets, as a prerequisite for beam shaping in practice.

At first sight the expression 'trapped wave antennas' appears like a contradiction in terms. How can a wave be 'trapped' and radiate at the same time? To explain this, we consider an infinitely long unshielded waveguide (dielectric slab or corrugated surface.) One or more trapped modes will propagate along it. They do not radiate as long as the guide is uniform. If the guide is cut short, however, to form a finite slab, then a radiation pattern is set up. This means that the existence of an electromagnetic far-field is somehow connected with the missing parts of the infinite surface. As usual, the far-field pattern is obtained by integrating over the tangential E or H field along the slab surface. Choosing the origin of coordinates in the center of the slab, and paying attention only to the limits of integration, we note first that

$$-\infty \int^{\infty} = 0$$

(no radiation). This implies, however, that

$$\int_{-\infty}^{-d/2} + \int_{-d/2}^{d/2} + \int_{d/2}^{\infty} = 0, \text{ or } \int_{-d/2}^{d/2} = - \int_{-\infty}^{-d/2} - \int_{d/2}^{\infty}$$

In words: the radiation pattern of the finite slab is equal in magnitude and opposite in phase to the pattern of the two semi-infinite slabs which have been cut away.

We can shuffle the limits of integration in another way, more useful than the first:

$$\int_{-d/2}^{d/2} = \int_{-\infty}^{d/2} - \int_{-\infty}^{-d/2}$$

(see Fig. 2). This shows that the finite-panel pattern can be regarded as a superposition of the patterns of two semi-infinite panels, with origin at $-d/2$ and $d/2$, respectively. The two panels are fed out of phase by the time it takes the trapped wave to travel from $-d/2$ to $d/2$, plus a delay of π occasioned by the phase-reversal of the shorter semi-infinite slab (minus sign in front of the last integral). To obtain maximum radiation in the end-fire direction, this phase delay must be compensated for by choosing d appropriately. We are merely restating here in physical terms the well-known Hansen - Woodyard condition, for which a number of pseudo-explanations have appeared in the literature from time to time.

It also appears from Fig. 2 that the far-field pattern of the longer semi-infinite slab could be obtained alternatively by integrating over a vertical plane through $d/2$ (vertical plane through $-d/2$ for the shorter semi-infinite slab). In this plane, the phase is constant and the amplitude decreases exponentially away from the slab (it is trigonometric within the slab). The radiation of the finite slab is therefore due to its two transverse planes of discontinuity. As before, the radiation from the plane through $d/2$ is delayed by the slow-travelling surface wave. In addition, we have the phase reversal for the aperture distribution in the plane through $-d/2$, because, just as before, we are supposed to subtract the $(-d/2)$ -pattern from the $(d/2)$ -pattern. Physically speaking, we might say that the phase reversal is due to the different nature of the field discontinuities at $-d/2$ and at $d/2$: in the first aperture the trapped wave is being set up, while in the second it is being terminated. At any rate, it is the front and the back end of the trapped wave which radiates, and there is no continuous leakage of energy along the slab surface itself.

We cannot go much further in our physical interpretation of the manner in which a trapped wave radiates. There seems to be little point in drawing lines of power flow from the finite slab to its far-field, since the impossibility of localizing energy in a Maxwellian field leads to an infinite number of power flow pictures, all equally correct.

WHAT BEAM SHAPES ARE OBTAINABLE FROM TRAPPED WAVE ANTENNAS?

Dunbar has shown² how to utilize travelling wave antennas for beam shaping. By suitable amplitude and phase control, it is possible to obtain shaped beams (such as csc^2) as well as patterns with deep nulls. Dunbar's design equations, as he himself noted, break down when the phase velocity of the travelling wave is less than that of light. This means that our trapped wave antennas are excluded from Dunbar's analysis. The question therefore arises as to the beam shapes obtainable from trapped wave antennas, and as to the variety of these beam shapes when compared to the faster-than-light travelling wave antennas.

Integration³ over the semi-infinite slab of Fig. 3 reveals that the radiation pattern due to a trapped wave is elliptic, with front-to-back ratio given by $(c + v)/(c - v)$. This assumes that the transition from slab to free space is made gradual enough to give rise to a negligible reflected wave. If an appreciable reflected wave does exist, its pattern will also be elliptic, but with reversed front-to-back ratio and a maximum amplitude equal to the incident wave amplitude multiplied by the reflection coefficient. In what follows, we shall omit this reflected wave pattern. Combining the two semi-infinite slab patterns according to the prescription in Section 1, we find by simple algebra that the finite slab pattern is of the conventional $\sin x/x$ form, with the added observation that the envelope of its polar plot is still the ellipse of the semi-infinite panel, doubled in amplitude but with the same front-to-back ratio.

Comparison with faster-than-light travelling waves shows (see Fig. 4) that the pattern of the semi-infinite antenna is now hyperbolic, with one branch of the hyperbola mirror-imaged about a vertical axis through the phase center. This results in a pattern infinity in the direction of the emerging main beam. The infinity disappears as soon as two hyperbolae are added (with suitable phase difference) to constitute a finite-aperture pattern. The conventional $\sin x/x$ polar plot emerges again, this time with hyperbolic envelope, and with maxima in the direction of the envelope infinities. The limiting case of $v = c$ results in a parabolic envelope, and applies to broadside arrays.

The stationary phase points which play an important part in Dunbar's analysis correspond to real angles of emergence for the main beam, and therefore do not appear in the far-field integral

for trapped waves. This suggests that a far greater variety of beam shapes is obtainable from fast-wave apertures than from trapped wave antennas. To achieve a prescribed radiation pattern it is necessary to vary the phase in the aperture so that the main beam emerges at a calculable real angle from each point on the antenna. This cannot be done for slow waves, since their angle of emergence is imaginary.

The situation is not quite as dark as it seems, for three reasons. To begin with, the pattern of a trapped wave antenna is always accompanied by the direct radiation of its feed. If the transition from waveguide to dielectric slab is sudden, as in Fig. 5, this feed radiation can be suppressed some 15 db or more below the energy transferred to the slab, and the total pattern is therefore essentially $\sin x/x$. If, on the other hand, a waveguide horn is used with gentle flare and large aperture to effect as gradual a transition to the open surface as possible, then the trapped wave appears to originate at minus infinity and the resulting pattern is elliptic. Putting it another way, the direct feed radiation is increased so as to fill in the deep nulls of the finite slab pattern, until in the limit the feed radiation exactly cancels the field due to the rear surface of discontinuity of the trapped wave, leaving us with the pattern of a semi-infinite slab. This shows that shaped patterns are in principle obtainable from trapped wave antennas.

The second reason is that the extent of the ground plane in which the trapped wave antenna is embedded can be utilized to influence the total pattern.⁴ Utilizing this method in conjunction with control over the direct feed radiation, Dr. M. Ehrlich⁵ succeeded in building a trapped wave antenna with excellent csc^2 pattern.

The third, and most important reason for optimism is based on the idea of phase and amplitude modulation along a trapped wave antenna. The first man to experiment with such an antenna was G.E. Mueller,⁶ although he did not analyze it from our present point of view. Fig. 6(a) illustrates his dielectric rod loaded with periodic disks of high dielectric constant. The disks bind their trapped wave so closely that its amplitude is negligible compared to the amplitude of the trapped wave on the rod itself; the disks are then spaced in such a way that the in-between sections of dielectric rod radiate in phase to yield a broadside pattern. Generalizing Mueller's idea, we may examine the far-field integral of any periodic phase and amplitude modulation, and soon discover the presence of one or more stationary phase points of the type utilized by Dunbar. In a recent conversation, Dr. J.C. Simon⁷ mentioned to one of us that he approached this problem via the Fourier spectrum of the modulated phase function which, when properly chosen, contains one or more lines corresponding to phase velocities faster than light. The decisive point is that Dunbar's beam shaping methods are now seen to be applicable to modulated trapped wave antennas, which therefore exhibit a wealth of pattern

potentialities equal to that of the fast-wave apertures and the unmodulated trapped wave antennas combined.

The antenna of Fig. 6(a) contains sharp discontinuities which cause reflections and high side lobes. To avoid these difficulties, the gentler modulation shown in Fig. 6(b) is proposed. This figure shows a dielectric slab above a metal sheet, first in longitudinal cross-section, and then from above. The periodicity in any given cross-section is so low that the resulting pattern would be multi-lobed. It will therefore be necessary to stagger the cross-sections in the manner shown and thus cancel the extraneous lobes. We hope to test such a "washboard" antenna in the near future.

PHASE CONTROL ON DIELECTRIC SHEETS

It is clear from the foregoing that success in beam shaping with trapped wave antennas rests on the possibilities of phase and amplitude control. This requires, first of all, a knowledge of the TE and TM modes in slabs and multiple layers. The modal properties are usually determined by solving a transverse eigenvalue problem, which can alternatively be expressed as a transverse resonance condition. We decided, however, to carry out the modal analysis by a new method, due to one of us (W. Gerbes). It introduces the Laplace transform to permit application to arbitrary pulse shapes propagating along the slab, and uses matrix notation to allow for an arbitrary number of layers. Only plane sheets have been considered thus far.

Figures 7 to 10 deal with the properties of the lowest TE and TM mode in a slab of thickness τ_2 spaced at a height τ_1 above a perfectly conducting ground plane. A given phase velocity can be realized in an infinite number of ways. If we start with a slab directly lying on metal, we can keep the phase velocity of a TM mode constant by raising the slab off the surface and increasing its thickness at the same time, until at very large distance from the ground plane it is twice what it was to begin with. To keep the phase velocity of a TE wave constant, the slab would have to lose in thickness as it is raised upward.

The dotted line in Figs. 7 and 8 gives those values of slab and air-gap thickness for which the phase velocity is the same in the TM and the TE mode. These are therefore the geometric configurations for which a dielectric trapped wave antenna can be elliptically polarized. If both modes are excited with equal strength, the resulting polarization is circular. Fig. 9 shows the conditions for elliptic polarization once more, and illustrates furthermore that the mode velocities are always restricted within a band terminated by the velocity of light on one side and by the free propagation velocity in an infinitely large dielectric slab on the other.

A large number of data were taken by one of us (H. Ehrenspeck) to verify the theoretical predictions of mode characteristics. The excellent agreement shown in the right half of Fig. 10 is typical of the results. The worst deviations obtained are those shown in the left half of the figure. They occur when the slab lies directly on the metal surface, and are due to the minute but non-negligible air gaps still between them. A simple error calculation was made, with the result that a good fit was obtained for an average air gap thickness as indicated on the figure.

When more than one slab is used, the number of different geometrical structures corresponding to a single phase velocity increases further. Fig. 11 shows three possible configurations (drawn to scale), all yielding the same phase velocity. The general formula for multiple layers is as follows:

$$th_1 = \frac{\alpha_{r+1} - \sum_{\nu} \alpha_{\nu} th_{\nu} + \alpha_{r+1} \sum_{\mu} \sum_{\nu} \frac{\alpha_{\nu}}{\alpha_{\mu}} th_{\mu} th_{\nu} - \dots}{1 - \alpha_{r+1} \sum_{\nu} \frac{1}{\alpha_{\nu}} th_{\nu} + \sum_{\mu} \sum_{\nu} \frac{\alpha_{\mu}}{\alpha_{\nu}} th_{\mu} th_{\nu} - \dots}$$

$$- \sum_{\phi} \sum_{\mu} \sum_{\nu} \frac{\alpha_{\nu} \alpha_{\phi}}{\alpha_{\mu}} th_{\phi} th_{\mu} th_{\nu} + \dots$$

$$- \alpha_{r+1} \sum_{\mu} \sum_{\nu} \sum_{\phi} \frac{\alpha_{\mu}}{\alpha_{\phi}} th_{\phi} th_{\mu} th_{\nu} + \dots$$

where

$$\phi > \mu > \nu > 1$$

and

$$th_{\nu} = \tanh ik_{\nu} \tau_{\nu}$$

(k is the transverse wave number),

$$\alpha_{\mu} = \epsilon k_{\mu} / \epsilon_{\mu} k_1,$$

and where the layers are numbered in ascending order. The formula for TE modes looks much the same, except that the tanh on the left-hand side is replaced by a coth, and that all dielectric constants are replaced by permeabilities.

Fig. 12 illustrates a number of the points which we have been making. It is the phase and amplitude plot of a dielectric slab placed on a large ground plane. The slab is indicated in black on the bottom of the picture, with the feed horn being just beyond the right-hand margin. The phase fronts are perpendicular to the slab near its surface, while the amplitude contours are essentially parallel and very densely packed due to the exponential decay (5 db per line). The slight

wiggle in the amplitude lines comes from a small reflected wave. The strong discontinuity on the left is due to the slab end. As a result of the very smooth transition from waveguide mode to trapped wave, the only disturbance in the vicinity of the feed is a group of amplitude islands on the extreme right which are some 35 db below the field intensity at the slab surface. The length of the antenna is such as to satisfy very nearly the Hansen - Woodyard condition for maximum gain: this can be seen by following the radiation from the region of the source, across the dielectric slab (where its phase is seen to interfere with the phase of the trapped wave), and into the main lobe region which begins on the extreme left and in which the original phase fronts align themselves quite well with the phase fronts arising at the terminal discontinuity.

To sum up, then, we might say that the process of radiation from trapped wave antennas seems well understood at the present time, that some of the basic information required for phase and amplitude control is now at hand, and that the introduction of phase and amplitude modulation along a trapped wave antenna should provide

us with beam shapes suitable for a great variety of applications.

REFERENCES

1. F.J. Zucker, "Theory and Applications of Surface Waves," Microwave Optics Symposium Issue, Supplemento No. 3, Nuovo Cimento 2 (1952). See also bibliography with this paper.
2. A.S. Dunbar, "On the Theory of Antenna Beam Shaping," J. Appl. Phys. 23, 847 (1952).
3. A.F. Kay and J. Kotik, paper to be published in the Proceedings of the McGill Symposium on Microwave Optics, to be issued by Air Force Cambridge Research Center in Summer, 1954.
4. R.S. Elliott, "On the Theory of Corrugated Plane Surfaces", TM 317, Contract AF 19(604)-262, Hughes Aircraft Company (1953).
5. Formerly at Hughes Aircraft Company, now with Microwave Radiation Company, Venice, California.
6. G.E. Mueller, "A Broadside Dielectric Antenna," Proc. I.R.E., 40, 71 (1952).
7. With the Compagnie de Telegraphie Sans Fils, Paris, France.

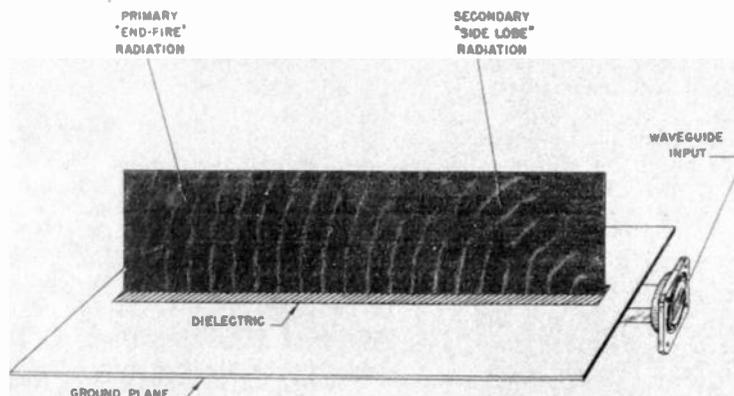


Fig. 1
Phase front contours of a metal-clad dielectric antenna.

The far field pattern of



is equivalent to



minus



Fig. 2



Minus

$$F/\theta = \frac{c+v}{c-v}$$



Equals



$v < c$

Fig. 3

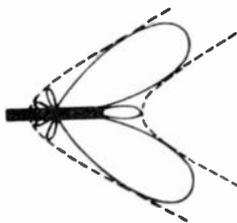


Minus

$$F/\theta = \frac{v+c}{v-c}$$



Equals



$v > c$

Fig. 4

Spectrum in region I: discrete; II: continuous; III: discrete + continuous; IV: continuous.

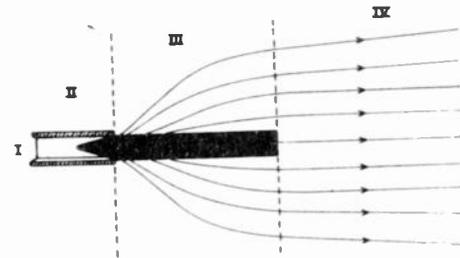


Fig. 5 - End-fire antennas.

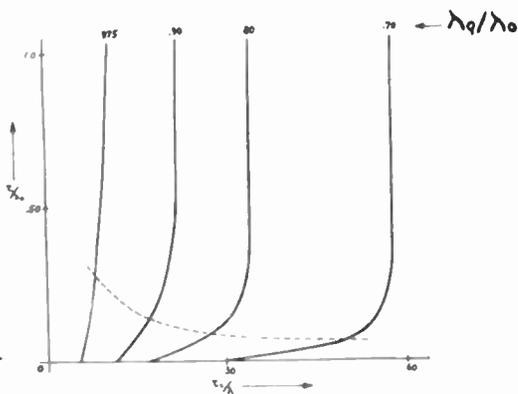


Fig. 6 - TM-mode.

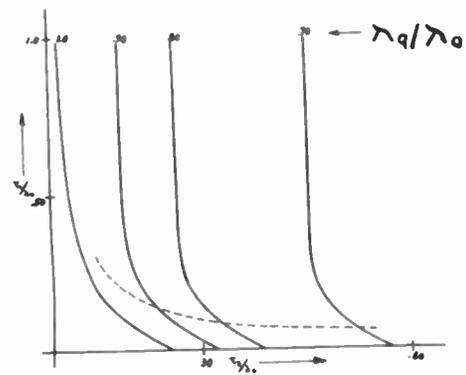


Fig. 7 - TE-mode.

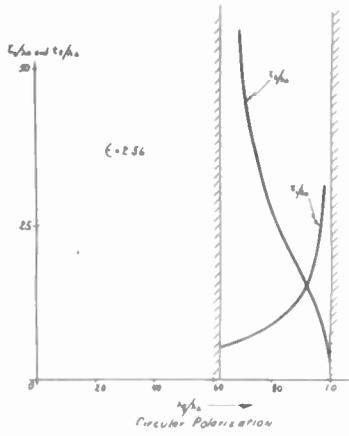
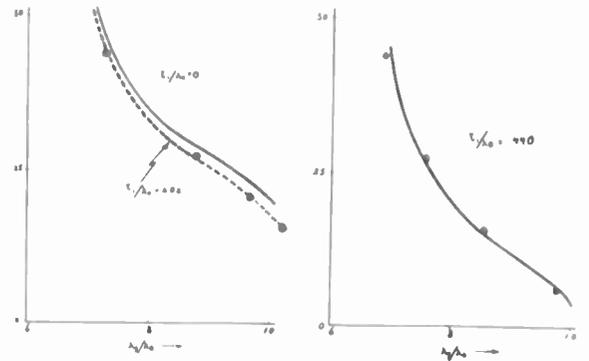


Fig. 8



TE-MODE

Fig. 9

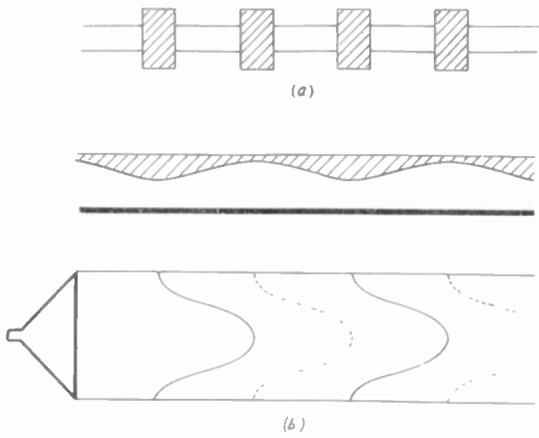


Fig. 10

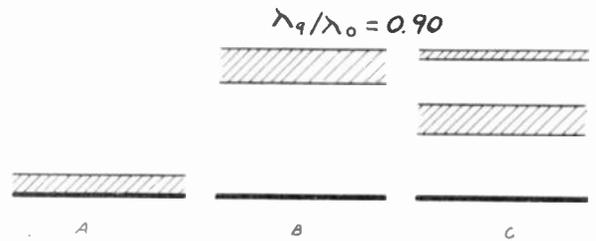


Fig. 11

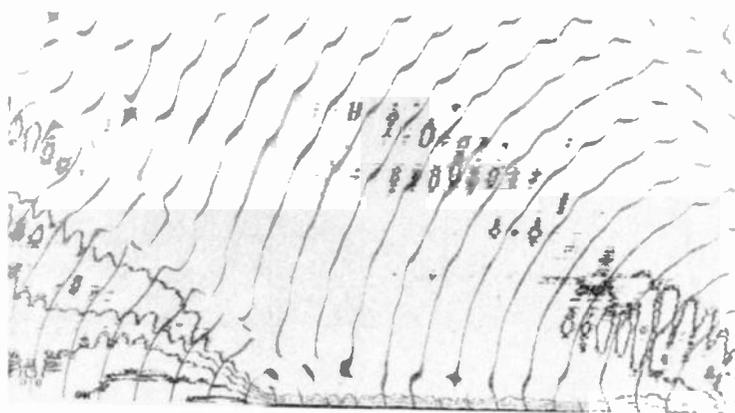


Fig. 12

SCATTERING OF ELECTROMAGNETIC WAVES BY WIRES AND PLATES

J. Weber

U. S. Naval Ordnance Laboratory, White Oak, Maryland

and

University of Maryland, College Park, Maryland

Abstract

The scattering of electromagnetic waves by wires and plates is discussed, particularly with reference to new polarization components which appear. The problem of scattering of radiation by a rectangular plate, for arbitrary angle of incidence, and arbitrary direction of polarization, is solved. The problem of scattering by wires is formulated in terms of the current distributions on the wire when center driven, and when driven as a receiving antenna. A relation between the magnitude of the input impedance of a center driven wire and the scattering properties of the wire is given. The use of simple sinusoidal current distributions is shown to give good results for the scattering of wires of length less than a wavelength.

The scattering properties of a wire are shown to provide a method for precisely measuring the magnitude of the input impedance of a center driven wire. This avoids the unknown impedance usually associated with measurements on a center driven wire.

Introduction

Experiments have shown that certain receiving antenna sites have the property of changing the polarization of incident electromagnetic waves. In order to study these effects, we have calculated the scattering of electromagnetic waves by plates, and wires.

Consider a wire ab upon which linearly polarized electromagnetic waves are incident, with the direction of polarization given by the vector \vec{E} , the component of the electric field vector \vec{E} parallel to ab will excite ab and the scattered fields will have components perpendicular to \vec{E} .

The cross polarization scattering of plates can be understood from the following model. Imagine that a conducting paraboloid of revolution is excited by a dipole, as shown in Figure 2.



Figure 2

The direction of the currents induced in the paraboloid is shown by the arrows. In the forward direction the polarization of the radiation is the same as that of the dipole, because the cross polarization field from element A is cancelled by that of element A'. Suppose we remove the entire paraboloid except element A. Then we still have currents induced in A with direction shown in Figure 3.



Figure 3

If the element A is of the order of a wavelength on a side or larger and remains oriented in a direction tangent to the paraboloid, the radiation scattered in the direction of the paraboloidal axis will be large. This is because the maximum of the diffraction pattern can be expected to be in the direction of the specular reflection angle. It will be shown later that this is so.

Scattering of Electromagnetic Waves by Plates

Consider a rectangular conducting plate oriented as in Figure 4.

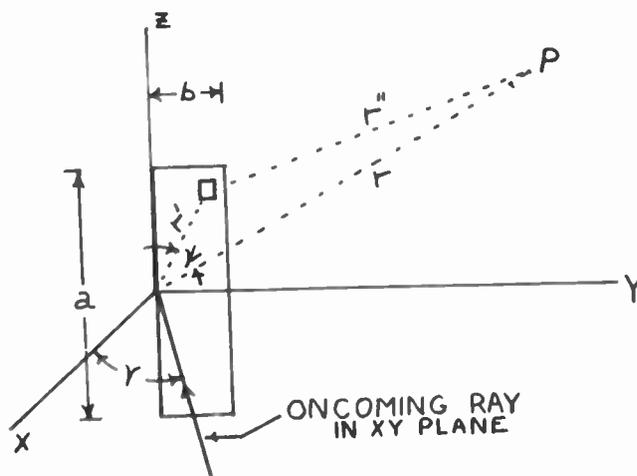


Figure 4

To simplify the calculation it will be assumed that the plate is rectangular in shape and lying in the YZ plane, with the origin at one edge of the plate (Figure 4). The oncoming

electromagnetic wave will excite currents in the plate. The radiation due to these currents will now be calculated.

If we are interested only in field components which vary as $\frac{1}{r}$, where r is the distance from the plate, then the electric field components E_θ and E_ϕ at point P can be shown to be given¹ by

$$E_\theta = -\frac{j\omega\mu}{4\pi r} N_\theta \quad (1)$$

$$E_\phi = -\frac{j\omega\mu}{4\pi r} N_\phi \quad (2)$$

where

$$N_\theta = N_x \cos\theta \cos\phi + N_y \cos\theta \sin\phi - N_z \sin\theta \quad (3)$$

$$N_\phi = -N_x \sin\phi + N_y \cos\phi$$

where θ and ϕ are the coordinates of point P and

$$\bar{N} = \iint \bar{J} e^{-jk r''} dA \quad (4)$$

In (4) \bar{J} is the surface linear current density on element dA of the plate, r'' is the distance between dA and point P. \bar{J} is a complex vector whose phase accounts for the difference in phase of the currents in the plate as a result of excitation by an obliquely incident wave.

If, for instance the wave normal of the incident wave is in the xy plane and makes an angle γ with the x axis, then the wavefront arrives at a point whose coordinate is $+y$ sooner than at the origin. The difference in distance (Figure 4) is $y \sin\gamma$. The difference in phase is $(y \sin\gamma) \left(\frac{2\pi}{\lambda}\right) = k y \sin\gamma$ where $k = \frac{2\pi}{\lambda}$

$$\bar{N} = \bar{J}_0 \int_{-\frac{a}{2}}^{\frac{a}{2}} \int_0^b e^{jk(y \sin\gamma - r'')} dA \quad (5)$$

Let us now consider the field scattered from the plate, at large distances.

In this case it is known that a good approximation is the relation $r'' \approx r - r' \cos\psi$ where r' is the radius vector from the origin to dA and ψ is the angle between r' and r . It can be readily shown² that

$$\cos\psi = \cos\theta \cos\theta' + \sin\theta \sin\theta' \cos(\phi - \phi') \quad (6)$$

where θ and ϕ are the angle coordinates of P

θ' and ϕ' are the angle coordinates of dA . In this case $\phi' = \frac{\pi}{2}$ for all points on the plate.

From the boundary conditions at the surface of a perfect conductor, we know that the value of \bar{J} is given by $\bar{J} = \bar{n} \times 2H_T$ where H_T is the

tangential component of the incident magnetic field H at the plate and \bar{n} is a vector of unit length normal to the plate. The tangential electric field at the plate is zero. In order to calculate the scattered field at point P it is only necessary to consider the tangential components at the plate. Then, combining expressions 5 and 6 we obtain

$$\bar{N} = 2 \int_0^b \int_{-\frac{a}{2}}^{\frac{a}{2}} \bar{n} \times \bar{H}_T e^{jk(y \sin\gamma + r' \cos\theta \cos\theta' + r' \sin\theta \sin\theta' \cos(\phi - \frac{\pi}{2}))} dy dz \quad (7)$$

$$r' \cos\theta' = z \quad r' \sin\theta' = y \quad \cos(\phi - \frac{\pi}{2}) = \sin\phi$$

With these substitutions the integral becomes

$$N = 2 \bar{n} \times \bar{H}_T \int_0^b \int_{-\frac{a}{2}}^{\frac{a}{2}} e^{jk(y \sin\gamma + y \sin\theta \sin\phi + z \cos\theta)} dy dz \quad (8)$$

Carrying out the integration, we obtain

$$N = \frac{2[\bar{n} \times \bar{H}_T] \sin\left(\frac{k a}{2} \cos\theta\right) \sin\left[\frac{k b}{2} (\sin\theta \sin\phi + \sin\gamma)\right] (ab)}{\left(\frac{k a}{2} \cos\theta\right) \left(\frac{k b}{2} [\sin\theta \sin\phi + \sin\gamma]\right)} \quad (9)$$

$$\text{LET } \alpha = \frac{k a}{2} \cos\theta, \quad \beta = \frac{k b}{2} (\sin\theta \sin\phi + \sin\gamma)$$

$$A = ab$$

$$N = 2A \bar{n} \times \bar{H}_T \left(\frac{\sin\alpha}{\alpha}\right) \left(\frac{\sin\beta}{\beta}\right)$$

If the field vector H makes an arbitrary angle δ with the z axis the component of H parallel to the z axis is $H_z = H \cos\delta$. The component of H parallel to the y axis is

$$H_y = H \sin\delta \cos\gamma$$

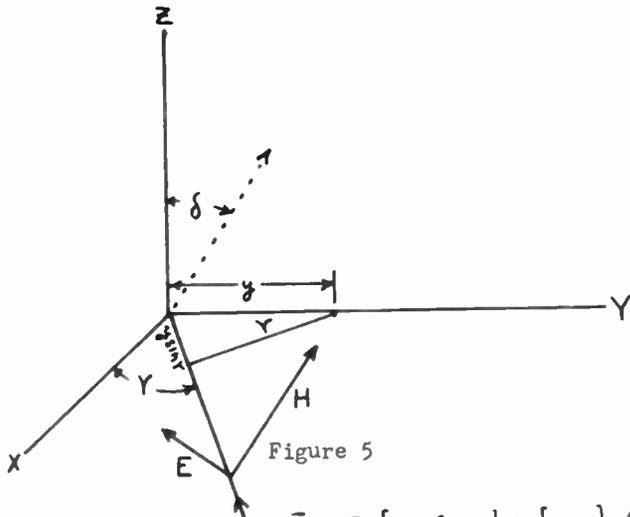


Figure 5

$$\vec{H}_T = \vec{a}_y [H \sin \delta \cos \alpha r] + \vec{a}_z [H \cos \delta] \quad (10)$$

where H_T is the tangential component of H at the plate. Now a unit normal to the plate is given by $\vec{n} = \vec{a}_x$, since the plate is perpendicular to the x axis.

$$\vec{n} \times \vec{H}_T = \vec{a}_z [H \sin \delta \cos \alpha r] + \vec{a}_y [-H \cos \delta]$$

Now

$$\vec{N} = 2AH \left(\frac{\sin \alpha}{\alpha} \right) \left(\frac{\sin \beta}{\beta} \right) \left[\vec{a}_y (-\cos \delta) + \vec{a}_z (\sin \delta \cos \alpha r) \right] \quad (11)$$

where we have set $A = ab$, A is the area of the plate.

$$N_\theta = (N_x \cos \psi + N_y \sin \psi) \cos \theta - N_z \sin \theta$$

$$N_\psi = -N_x \sin \psi + N_y \cos \psi$$

Making use of (11) we obtain

$$N_\theta = 2AH \left(\frac{\sin \alpha}{\alpha} \right) \left(\frac{\sin \beta}{\beta} \right) \left[-\cos \delta \cos \theta \sin \psi - \sin \delta \cos \alpha r \sin \theta \right]$$

$$N_\psi = 2AH \left(\frac{\sin \alpha}{\alpha} \right) \left(\frac{\sin \beta}{\beta} \right) \left[-\cos \psi \cos \delta \right]$$

The scattered field components E_θ and E_ψ are given¹ by:

$$E_\theta = -\frac{j\omega\mu}{4\pi r} N_\theta \quad E_\psi = -\frac{j\omega\mu}{4\pi r} N_\psi$$

Inserting the values of N_θ and N_ψ , simplifying and making use of the relation

$$\frac{E}{H} = 120\pi \quad \text{we obtain}$$

$$E_\theta = \frac{AE}{\lambda r} \left(\frac{\sin \alpha}{\alpha} \right) \left(\frac{\sin \beta}{\beta} \right) \left[\cos \theta \sin \psi \cos \delta + \sin \theta \sin \delta \cos \alpha r \right] \quad (12)$$

$$E_\psi = \frac{AE}{\lambda r} \left(\frac{\sin \alpha}{\alpha} \right) \left(\frac{\sin \beta}{\beta} \right) \left[\cos \psi \cos \delta \right]$$

Where:

E_θ and E_ψ are the θ and ψ components of the scattered fields.

E is the incident field

λ is the wavelength

A is the area of the plate

r is the distance from the plate to the point at which the scattered fields are required.

a is the length of the plate

b is the width of the plate

$$k = \frac{2\pi}{\lambda}$$

$$\alpha = \frac{ka}{2} \cos \theta$$

$$\beta = \frac{kb}{2} (\sin \theta \sin \psi + \sin \tau)$$

θ and ψ are the spherical coordinates of the point at which the scattered fields are required. The incident ray is in the XY plane.

γ is the angle between the direction of the incident ray and the normal to the plate, δ is the angle between direction of the incident magnetic field vector and the z axis.

The relations 12 are the general solution for the distant scattered fields on a small rectangular plate oriented as shown, for any angle of incidence and any direction of polarization. It is apparent that although the formulas were derived for a rectangular area they will be valid for small areas of any shape provided that the shape is such that one can reasonably fit a similarly oriented rectangle to it. Only the factor $\left(\frac{\sin \alpha}{\alpha} \right) \left(\frac{\sin \beta}{\beta} \right)$ needs to be modified for other shapes. For example if the shape is circular the factor $\left(\frac{\sin \alpha}{\alpha} \right) \left(\frac{\sin \beta}{\beta} \right)$ would probably be a product of two Bessel functions.

If a and b are larger than a wavelength

$$\alpha > \pi \cos \theta$$

$$\beta > \pi (\sin \theta \sin \psi + \sin \tau)$$

Under these conditions the factor $\left(\frac{\sin \alpha}{\alpha}\right) \times \left(\frac{\sin \beta}{\beta}\right)$ will be small unless θ , γ , and ψ are such that α and β both approach zero. The quantity $\frac{\sin \alpha}{\alpha} \frac{\sin \beta}{\beta}$ therefore is the dominant

term in the expression 12. The scattered fields are a maximum if $\alpha \rightarrow 0$, $\beta \rightarrow 0$ i.e. if $\theta \rightarrow \frac{\pi}{2}$ and $\gamma \rightarrow -\psi$. These conditions will be recognized as the ordinary specular reflection condition for infinite planes. This constitutes a proof that if a scattering plate is of dimensions of the order of a wavelength by a wavelength, or larger, the maximum of the scattering occurs for the angle of incidence equal to the angle of reflection. If the plate is very small or very narrow it behaves like a wire and the specular reflection condition has no special significance.

To obtain the cross polarization and parallel polarization components we require the components of the total field in directions perpendicular and parallel respectively to the original electric field.

A unit vector perpendicular to the incident electric field vector is, from the geometry of Figures 4 and 5, given by

$$U_{\perp} = \bar{a}_x (-\sin \delta \sin \gamma) + \bar{a}_y (\sin \delta \cos \gamma) + \bar{a}_z (\cos \delta)$$

A unit vector parallel to the incident electric field is similarly given by

$$U_{\parallel} = \bar{a}_x (\cos \delta \sin \gamma) + \bar{a}_y (-\cos \delta \cos \gamma) + \bar{a}_z (\sin \delta)$$

The total scattered electric field can be written as

$$E_{\text{Total Scattered}} = \bar{a}_x (E_{\theta} \cos \theta \cos \psi - E_{\psi} \sin \psi) + \bar{a}_y (E_{\theta} \cos \theta \sin \psi + E_{\psi} \cos \psi) + \bar{a}_z (-E_{\theta} \sin \theta) \quad (12A)$$

where E_{θ} and E_{ψ} are given by (12)

To obtain the parallel and cross polarization components we form the scalar products $\bar{U}_{\perp} \cdot \bar{E}_{\text{total}}$ and $\bar{U}_{\parallel} \cdot \bar{E}_{\text{total}}$. This is a tedious calculation, the results are

$$E_{\text{scattered parallel polarized}} = \frac{AE}{\lambda r} \frac{\sin \alpha}{\alpha} \frac{\sin \beta}{\beta} \left[\sin \delta \cos \delta \sin \theta \cos \theta \cos \psi \right. \\ \left. (\sin \gamma \cos \gamma \cos \psi - \sin \psi - \cos^2 \gamma \sin \psi) + \sin \psi \cos \psi \right. \\ \left. (\cos^2 \theta \cos^2 \delta \sin \gamma - \cos^2 \delta \sin \gamma) - \cos^2 \theta \sin^2 \psi \cos^2 \delta \cos \gamma \right. \\ \left. - \sin^2 \theta \sin^2 \delta \cos \gamma - \cos^2 \delta \cos^2 \psi \cos \gamma \right] \quad (13A)$$

$$E_{\text{scattered cross polarized}} = \frac{AE}{\lambda r} \frac{\sin \alpha}{\alpha} \frac{\sin \beta}{\beta} \left[\sin \delta \cos \delta \cos \psi \right. \\ \left. \sin \psi \cos \psi \sin \gamma \sin^2 \theta + \sin \delta \cos \delta \cos \gamma (\sin^2 \psi \cos^2 \theta + \cos^2 \psi - \sin^2 \theta) + \sin \theta \cos \theta \sin \psi (\sin^2 \delta \cos^2 \gamma - \cos^2 \delta) \right. \\ \left. - (\sin^2 \delta \sin \gamma \cos \gamma \sin \theta \cos \theta \cos \psi) \right] \quad (13B)$$

Expressions (13) are the general solution for the distant parallel and cross polarized scattered fields in terms of the quantities which have already been defined under equations (12), for any angle of incidence and any direction of polarization of the incident wave.

If the dimensions of the plate are of the order of a wavelength by a wavelength or larger then the maximum values of the scattered fields will occur, as discussed previously, for $\gamma = -\psi$ and $\theta = \frac{\pi}{2}$. In this case expressions (13) reduce to:

$$E_{\text{scattered cross polarized}} = \frac{AE}{\lambda r} \sin \delta \cos \delta \cos \psi [\cos^2 \psi - 1] \quad (14A)$$

and

$$E_{\text{scattered parallel polarized}} = \frac{AE}{\lambda r} \cos \psi [\cos^2 \delta \cos^2 \psi + \sin^2 \delta] \quad (14B)$$

Expressions (14) give the scattered fields for the direction in which the angle of reflection is equal to the angle of incidence. Expression 14A can be shown to have a maximum when $\delta = \frac{\pi}{2}$ and $\psi = \cos^{-1} \frac{1}{\sqrt{3}}$. In this case 14A becomes

$$E_{\text{cross polarized (maximum)}} = \frac{2AE}{3\sqrt{3}\lambda r} \quad (15)$$

where A is the area, E is the incident field, λ is the wavelength, and r is the distance from the plate. Expression (15) gives us the maximum value of the cross polarized field, under the condition that the angle of incidence is equal to the angle of reflection, the electric field intensity vector makes an angle of 45° with the plane of incidence*, and the angle of incidence is $\cos^{-1} \frac{1}{\sqrt{3}} = 54.8^\circ$.

Scattering by Wires

In order to calculate the scattering from a wire it is necessary to solve the problem of the diffraction of electromagnetic waves by a cylinder. It is one of the purposes of this

* The plane of incidence is a plane containing the incident ray, and normal to the scattering plane.

paper to point out a close connection between this diffraction problem and the antenna boundary value problems which have been considered by radio engineers.

We imagine first that our scattering wire is broken at the center and an impedance Z_L is inserted. Then the scattering wire is equivalent to a receiving antenna. An equivalent circuit for the antenna is given by Figure 6.



Figure 6

Z_{IN} is the input impedance of the antenna. E is the electric field and h is the effective height of the antenna. If we remove Z_L and again join the two halves of the antenna at the center the equivalent circuit is represented by Figure 7, and this represents the scattering wire.

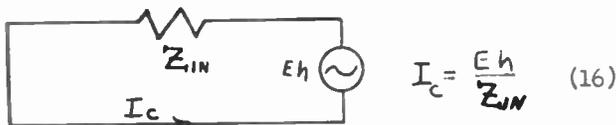


Figure 7 h is the effective height.

A formula for h can be obtained using the methods of Schelkunoff⁶. The formula is

$$h = \omega \int_{-l}^l e^{jkz' \cos \theta'} f(z') dz' \quad (17)$$

where $f(z')$ is the current distribution function of the antenna when driven at the center, as a transmitter, l is the half length of the antenna, θ' is the angle between the incident ray and the axis of the wire, ψ' is the angle between the incident electric field intensity vector and the axis of the antenna, $k = \frac{2\pi}{\lambda}$. When the antenna is employed as a receiving antenna it is driven by an oncoming electromagnetic wave and the driving forces are now distributed along the entire antenna rather than at the center. We cannot assume that the current distribution function is the same as for the center driven antenna. Let the new current distribution function be $\varphi(z)$. From a knowledge of the current at the center of the antenna and the new current distribution function $\varphi(z)$ we can calculate the fields. The radiation field³ is given by

$$E_\theta = \frac{j\omega \mu}{4\pi r} \int_{-l}^l I(z') \sin \theta' e^{-jkr} dz' \quad (18)$$

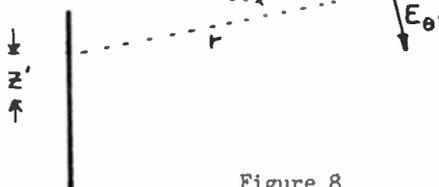


Figure 8

where l is the half length of the antenna (see Figure 8). I is the current on the receiving antenna as a function of z . Now $I = I_c \varphi(z)$ where I_c is the current at the center of the antenna. Taking the value of I_c from expression 16, we obtain

$$I = \frac{Eh \varphi(z)}{Z_{in}} = \frac{E \varphi(z) \omega \int_{-l}^l e^{jkz' \cos \theta'} f(z') dz'}{Z_{in}} \quad (19)$$

We have used both z and z' to represent the coordinates of a point on the antenna to avoid confusion in evaluating the definite integrals in 20.

Utilizing expressions (17) and (18) we obtain

$$E_\theta = \frac{j\omega \mu E \sin \theta \omega \int_{-l}^l e^{jkz' \cos \theta'} f(z') dz' \int_{-l}^l \frac{\varphi(z') e^{-jkr}}{r} dz'}{4\pi Z_{in}} \quad (20)$$

In (20) E is the incident electric field intensity, ψ' is the angle between the incident field direction and the wire axis, θ' is the angle between the incident ray and the wire axis, z' is the coordinate of a point on the wire, measured from the center, $f(z')$ is the current distribution function of the wire when broken at the center and driven as a transmitting antenna, $\varphi(z)$ is the current distribution function of the wire as a receiving antenna when excited by a plane wave with direction of propagation making an angle of θ' with the wire axis and electric field intensity vector making an angle ψ' with the wire axis, Z_{in} is the input impedance of the scattering wire when center driven. It is worth noting that the magnitude of the input impedance for a cylindrical wire is well defined by (19) and (20) in terms of the scattering properties of the wire, and the current distributions. Measurements of the scattering properties could be used to obtain $|Z_{in}|$ without the difficulties usually encountered in measuring Z_{in} by center driving the wire. This is because the unknown impedance associated with the antenna feed does not enter into (19) and (20).

If the problem of the receiving antenna is solved rigorously, as a boundary value problem, an expression for the current $I(z)$, is obtained. If this is inserted into expression (18), the scattered fields can be obtained, provided the integrations can be carried out. The integrations are tedious and have to be done numerically.

A simpler procedure which gives good results in certain cases is to utilize (20), and make simplifying assumptions concerning the current distribution functions $f(z')$ and $\varphi(z)$

based on the approximate analogy between an antenna and a transmission line. For example we could assume

$$f(z') = \frac{\sin[k(l-z')]}{\sin kl}, \quad z' > 0$$

$$f(z') = \frac{\sin[k(l+z')]}{\sin kl}, \quad z' < 0 \quad (21)$$

Expressions (21) are well known, to obtain an approximate formula for $\Phi(z)$ we can proceed as follows, see Figure 9.

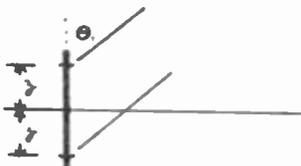


Figure 9

We regard the two antenna elements dy located at $+z$ and $-z$ as a section of a transmission line driven by the oncoming electromagnetic wave and we have, from the transmission line equations

$$\frac{dV}{dz} = -I Z' + E \cos \psi' \left[e^{-jk(r_0 - j\omega z')} + e^{-jk(r_0 + j\omega z')} \right] \quad (22)$$

$$\frac{dI}{dz} = -Y V \quad (23)$$

where I is the current along the line
 V is the potential along the line
 Z' is the series impedance per unit length
 Y is the shunt admittance per unit length
 ψ' is the angle between the electric field vector and the wire axis
 θ' is the angle between the incident wave-normal and the wire axis
 $k = \frac{2\pi}{\lambda}$, $\lambda = \text{wavelength}$
 z is the distance along the transmission line

If equation (23) is partially differentiated with respect to z and equation (22) substituted in it we obtain

$$\frac{d^2 I}{dz^2} = -k_1^2 I - 2Y E \cos \psi' \cos(k_1 z \cos \theta') e^{-jk_1 r_0} \quad (24)$$

$$\text{where } k_1^2 = Z' Y$$

The solution of the differential equation (24) satisfying the boundary conditions of zero current at the ends is

$$I = I_c \frac{\cos k_1 z \cos(k_1 l \cos \theta') - \cos k_1 l \cos(k_1 z \cos \theta')}{\cos(k_1 l \cos \theta') - \cos k_1 l} \quad (25)$$

where I_c is the current at the center of the antenna.

The current distribution function $\Phi(z)$ is therefore given by:

$$\Phi(z) = \frac{\cos k_1 z \cos(k_1 l \cos \theta') - \cos k_1 l \cos(k_1 z \cos \theta')}{\cos(k_1 l \cos \theta') - \cos k_1 l} \quad (26)$$

If we employ this expression and the expressions 21 for $f(z')$, the integration required by expression 20 can be carried out, the result is:

$$E_\theta = \frac{j\omega \mu E_0 \cos \psi'}{4\pi r^2 Z_{in}} \left[\frac{2k_1 \sin \theta}{(\sin k_1 l)(k_1^2 - k^2 \cos^2 \theta')} \right] X$$

$$\left[\cos(k_1 l \cos \theta') \left[\frac{2k_1 \sin \theta \cos(k_1 l \cos \theta')}{k_1^2 - k^2 \cos^2 \theta'} - \frac{2k_1 \cos \theta \cos k_1 l \sin(k_1 l \cos \theta')}{k_1^2 - k^2 \cos^2 \theta'} \right] \right. \\ \left. - \cos k_1 l \left(\frac{\sin[k_1 l (\cos \theta + \cos \theta')]}{k_1 (\cos \theta + \cos \theta')} + \frac{\sin[k_1 l (\cos \theta - \cos \theta')]}{k_1 (\cos \theta - \cos \theta')} \right) \right] \quad (27)$$

For most purposes we can take $k_1 = k$, since the phase velocity for waves along the antenna is very close to the free space velocity. In this case (27) becomes

$$E_\theta = \frac{j\omega \lambda E_0 \cos \psi'}{\pi r^2 Z_{in}} \left[\frac{\sin \theta}{\sin^2 \theta' \sin k_1 l} \right] \left[\cos(k_1 l \cos \theta') \right] X$$

$$\left[\frac{\sin k_1 l \cos(k_1 l \cos \theta) - \cos \theta \cos k_1 l \sin(k_1 l \cos \theta')}{\sin^2 \theta} \right. \\ \left. - \frac{\cos k_1 l}{\sin^2 \theta} \left(\frac{\sin[k_1 l (\cos \theta + \cos \theta')]}{\cos \theta + \cos \theta'} + \frac{\sin[k_1 l (\cos \theta - \cos \theta')]}{\cos \theta - \cos \theta'} \right) \right] \quad (28)$$

where:

E_{θ} is the distant scattered electric field

E is the incident field

ψ' is the angle between incident field and wire axis

λ is the free space wavelength

r is the distance between the wire and the point at which the scattered fields are desired

Z_{in} is the input impedance of the wire when broken at the center and driven as an antenna. The radius of the wire will partially determine the value of Z_{in}

$$k_1 \cong k = \frac{2\pi}{\lambda}$$

λ is the free space wavelength

l is the half length of the wire

θ' = angle between axis of the wire and the incident wavenormals

θ = angle between axis of the wire and the radius vector to the point at which the scattered fields are desired

The author is well aware of the fact that the sinusoidal current distributions are not correct for an antenna. The use of the correct input impedance Z_{in} and the sinusoidal current distributions appears to be a good procedure for antenna lengths less than a full wavelength. To show this we have calculated the back scattering cross section σ . For a plane wave incident normally on the wire, with electric field intensity vector parallel to the wire, the back scattering cross section can be calculated using expression 28, and is given by

$$\sigma = 4\pi\lambda^2 \left[\frac{60}{\pi Z_{in}} \sin k l \left(\sin k l - k l \cos k l \right) \right]^2 \quad (29)$$

σ is 4π times the power scattered back towards the source per unit solid angle divided by the incident power per unit area

λ is the wavelength

Z_{in} is the input impedance of the wire when it is center driven

$$k = \frac{2\pi}{\lambda}$$

l is the half length of the wire

The very simple formula (29) has been used to calculate the back scattering cross section for the .01" diameter wire, at 3000 mcs, which was used by Dike and King⁴. The results are plotted on Figure 10. It is apparent that the rough approximations used here agree considerably better with experiments than do existing first order solutions of the problem using the Hallén integral equation. For Z_{in} we have used the input impedances of cylindrical antennas calculated by Schelkunoff.

The expressions given here for the scattering from wires are not valid for wire lengths near a multiple of a full wavelength. King and Harrison⁵ have given a simple approximate formula for $f(\mathbf{z}')$, the current distribution function for the center driven antenna. They also describe a procedure for introducing a fictitious length for antenna lengths near a multiple of a full wavelength. The use of their relation for $f(\mathbf{z}')$ and their procedure, makes it possible to use the formulations of this paper for wire lengths up to the vicinity of two wavelengths.

Conclusion

We have obtained expressions for the scattering of electromagnetic waves by a rectangular plate. The simple formula (15) gives the maximum value of the cross polarized field. A relation between the current distributions, input impedance, and scattering properties of a wire has been given. This suggests a method for measuring the magnitude of the input impedance by measuring scattering cross sections. This avoids the unknown terminal impedance associated with most center feed methods of impedance measurement. Relations have been obtained for the scattered fields and back scattering cross section of a wire, based on assumed sinusoidal current distributions. These expressions are simple and reasonably accurate up to antenna lengths approaching a full wavelength.

Bibliography

1. Romo and Whinnery: Fields and Waves in Modern Radio, 1st edition page 455
2. Romo and Whinnery: Fields and Waves in Modern Radio, 1st edition page 454
3. Romo and Whinnery: Fields and Waves in Modern Radio, 1st edition page 431
4. Dike and King, Proc. I.R.E., 40, 853, 1952
5. Harrison and King, Proc. I.R.E., 31, 548, 1943
6. S. A. Schelkunoff, Electromagnetic Waves, page 479

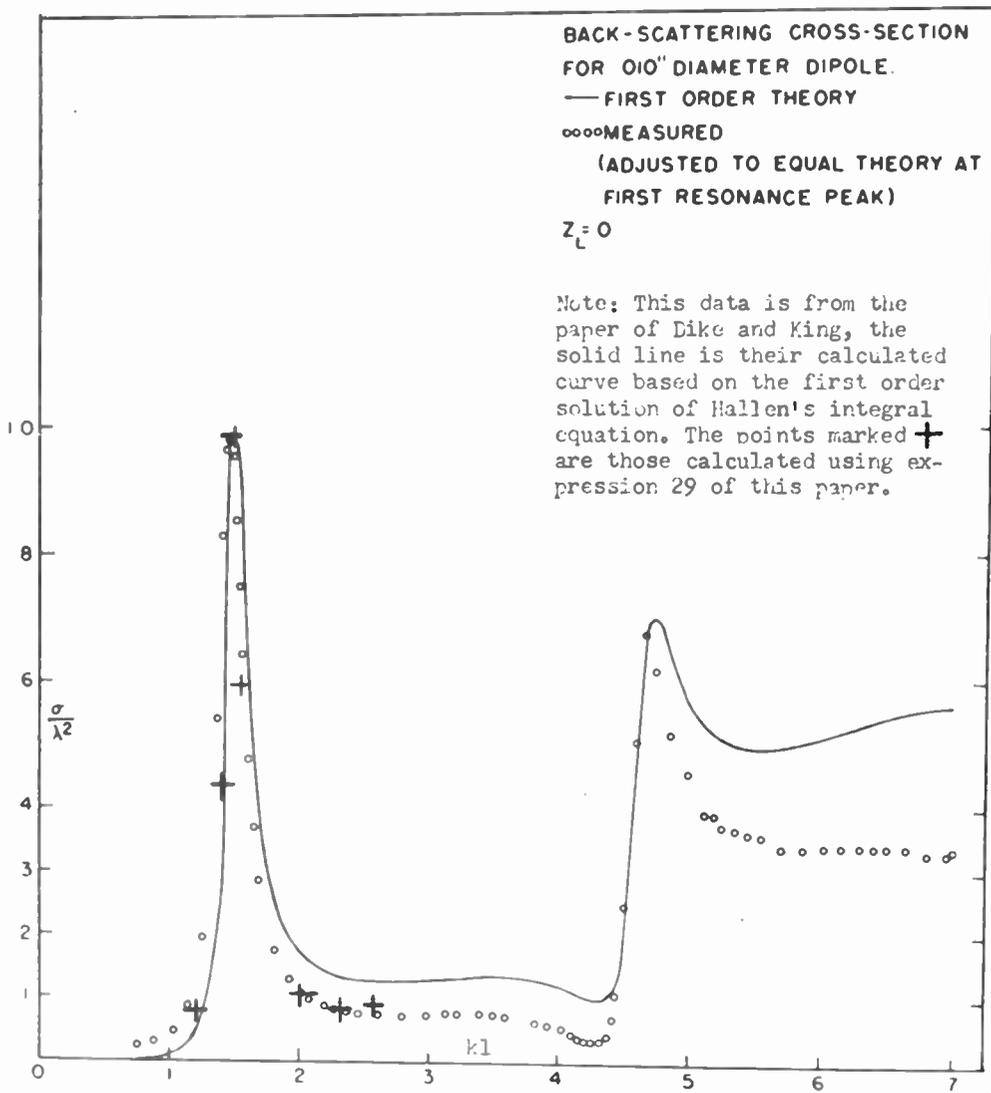


Fig. 10 — Broadside back-scattering cross section of unloaded 0.010-inch diameter dipole.

REFLECTIONS IN MICROWAVE ANTENNAS AND THEIR HARMFUL EFFECTS

Peter W. Hannan
Wheeler Laboratories
Great Neck, N. Y.

Abstract

In a microwave antenna of the focusing type, the reflecting or refracting element is illuminated by power from the waveguide assembly. At any boundary surface of the focusing element, some of this power may be reflected and return to the waveguide assembly. Here, this reflected power may cause troubles such as instability of the transmitting oscillator. Formulas for computing this reflection are presented, and techniques for its reduction are reviewed.

There is also reflection from the waveguide assembly when it receives power by way of the focusing element. This reflection is composed of two components: that caused by incompletely matched plumbing elements, and that caused by scattering from the feed. The feed reflection is determined by the aspect presented by the feed to the incoming focused wave.

The combination of the reflections from the focusing element and from the waveguide assembly causes a sinusoidal variation of antenna gain with frequency, at a rate dependent on the distance between the two reflections. In a lobing antenna these reflections, if asymmetrical, cause an error in its direction-finding information, which likewise varies sinusoidally with frequency. Formulas for computing these effects are presented, and techniques for their reduction are reviewed.

Introduction

As the requirements on microwave antennas become ever more demanding, one aspect of their design is receiving increased attention; namely, the harmful effects of reflections within the antenna. In particular, these reflections are likely to degrade certain properties of the antenna, when these properties are supposed to be maintained over a wide frequency band. This paper discusses some of these effects, and some possible techniques for minimizing them.

The general class of antenna which is considered consists of a focusing element and a feed, as shown in Fig. 1. Attached to the feed is the plumbing; the feed and plumbing together are given the name waveguide assembly. Most microwave antennas are highly directive; in these antennas, both the aperture size of the focusing element, and the distance between it and the feed, are many wavelengths. In some systems, the feed and plumbing are close together, while in others, they are connected by a waveguide which is many wavelengths long.

Ordinarily, all three elements reflect in some manner and degree. Individually, the reflections may degrade the performance of the system in some respects; when taken in combination, still further harmful effects may ensue. The following three categories are considered:

1. Reflection from the focusing element.
2. Reflection from the waveguide assembly.
3. Combination of reflections from the focusing element and the waveguide assembly.

Reflection from the Focusing Element

When the focusing element is illuminated by power from its waveguide assembly, some power is reflected from the focusing element and returns to the waveguide assembly. Approximate formulas for the reflection coefficient seen by the feed are given in Figs. 2 and 3 for two common antenna types.

The dish-type antenna, shown in Fig. 2, has complete reflection at the dish, but only part of this reflection re-enters the feed (Ref. 1). The magnitude of this part is determined by assuming that the power at the paraboloidal apex travels back toward the feed as would that of a uniform plane wave. The voltage reflection

coefficient seen by the feed may then be computed from the feed gain, focal length, and wavelength (Ref. 2,3,4); or alternately, by assuming a relation between feed gain and antenna dimensions, the reflection coefficient may be computed from the dish area, focal length, and wavelength, as shown. It may be seen that an increase of antenna size will, if the angular width of the feed radiation is held constant, decrease the reflection into the feed of the dish-type antenna.

The lens-type antenna, shown in Fig. 3, has partial reflection at the lens surfaces; but in the case of a flat outer surface, all the reflection from this surface is focused back to the feed. This occurs because the antenna is focused for an incoming wavefront that also is flat, and the reflection at every part of the outer surface adds in-phase at the location of the feed. Thus, in the absence of intermediate losses, the reflection coefficient seen by the feed is that of the flat outer surface. When the antenna has losses, such as those caused by phase distortion or spillover, the reflection into the feed is decreased. It is assumed in Fig. 3 that the curved inner surface is sufficiently defocused so that it contributes only a comparatively small effect.

The reflection coefficient of the lens surface itself depends on the particular lens material used. Some materials change with frequency; also, in the case of a curved or tilted lens surface, the dependency upon the relative orientation of the incident wave and the lens surface must be considered (Ref. 3,5,6,7).

The reflection from the focusing element affects the load presented to the transmitting oscillator, as shown in Fig. 4 for the case of a typical wide-band system. When this reflection is added to that which is already present in the waveguide assembly during transmission, the resulting reflection varies with frequency at a rate which depends on the distance between the focusing element and the waveguide assembly. At certain frequencies, the two reflections interfere to yield a minimum; the waveguide-assembly reflection may even be adjusted so that this minimum is zero. However, at other frequencies the reflections reinforce, and the maximum reflection seen by the transmitting

oscillator is greater than that of the waveguide assembly alone. This increase of reflection may bring the oscillator close to an unstable operating point, or "sink" (Ref. 8). Perhaps even more important, the reflection from the focusing element may cause frequency jumping because of the large distance in wavelengths from the oscillator to the focusing element (Ref. 9,10). In addition, in high-power-pulse radar systems the reflection may contribute to voltage breakdown, and in microwave relay systems it may increase the distortion of information (Ref. 11).

The dish and lens antennas analyzed above are special cases of a more general analysis which handles a reflecting surface having any degree of defocusing. For simplicity, it is assumed that the surface is spherical and that the aperture of the focusing element is circular and is small compared to the spherical radii of both the reflecting surface and incident wavefront. The result, given for the case of a totally reflecting surface, is shown in Fig. 5. Two types of analyses are given. The ray analysis, previously used for the dish antenna, modifies the result given in Fig. 2 (second formula) to account for divergence or convergence of the energy reflected back toward the feed. The wave analysis, previously used for the lens antenna, modifies the result given in Fig. 3 to account for out-of-phase contributions to the reflection. For simplicity, the wave analysis assumes that the reflecting surface is uniformly illuminated by the feed, and it is given for the case of zero antenna losses.

The ray analysis is a good approximation only when the illumination of the focusing element is tapered down at the edge and when the reflection is defocused. The wave analysis is exact, but it provides the indicated simple result only when the illumination is uniform or when the reflection is perfectly focused. Since most antennas provide tapered illumination of the focusing element, the regions in which the two analyses usually yield a good approximation are shown by the solid curves in Fig. 5, and the other regions are shown by broken curves. It may be seen from the curves that when the reflection is defocused, the ray analysis yields approximately the average of the wave analysis. If the wave analysis were carried out for

the case of tapered illumination, the curve of Fig. 5 would have its nulls filled in and its minor maximums reduced, and would approach that obtained by the ray analysis when the reflection is defocused.

There are a number of possible methods for reducing the reflection from the focusing element into the feed; some of these are shown in Fig. 6. Shortening the focal length of a dish (Fig. 6a), or curving the front surface of a lens (Fig. 6b) spreads or defocuses the reflection. When this is allowable by the mechanical and electrical design, it offers one solution. Tilting a paraboloidal dish will direct the reflection away from the feed, but a large tilt (Fig. 6c) is usually required to get a substantial improvement. A large tilt may cause phase aberrations (Ref. 3, 4); to prevent this, an offset section of the paraboloid is often used (Ref. 1,3,4). Tilting of lenses (Fig. 6d) is also a popular technique (Ref. 12,13), and in the case of a focused reflection a small tilt will effect a large improvement. A similar procedure involves stepping half of the lens a quarter wave back from the other half (Ref. 13). Apex plates (Fig. 6e) are often used to cancel the reflection from a dish into the feed (Ref. 1,2, 3,4); this is done at the expense of diverting a small fraction of the power from the main lobe into sidelobes. Matching techniques (Fig. 6f) have been applied to reduce the reflection at the surface of a lens (Ref. 14). When this is done to a non-refracting surface (such as the flat outer surface), there is no change of the focusing action of the lens; however, when performed on a refracting surface, the lens contour may have to be modified to maintain proper focusing. Another method for reducing the reflection at a lens surface would be, of course, to employ a lens material having the same impedance as that of free space (Ref. 6,15). The possibility for achievement of non-reflecting surfaces, either by design of the lens material or by design of matching plates, is an advantage of the lens antenna over the dish antenna.

Cancellation of the focusing-element reflection with another distant one introduced in the plumbing is a possibility (Fig. 6g and 6h); a small loss is introduced by this technique (Ref. 16), and the design must be such that the magnitude and

phase of the cancelling reflection accurately tracks that from the focusing element. Polarization crossing may also be utilized: in a dish antenna, a polarization twister placed on the dish (Fig. 6i) will cross-polarize the reflected wave so that none re-enters the feed (Ref. 3). In an antenna having a lens construction which is alike in the E and H planes, a quarter-wave plate between the feed and lens (Fig. 6j) will also cross-polarize the reflection; another quarter-wave plate on the other side of the lens will restore the linear polarization. Another technique, not shown in Fig. 6, which can eliminate some effects of the reflection, utilizes a non-reciprocal element (Ref. 17) in sections of the plumbing where only one-way energy flow is required.

Reflection from the Waveguide Assembly

When the waveguide assembly receives power by way of the focusing element, some of this power is reflected back toward the focusing element, as shown in Fig. 7. The reflection is the sum of two components: the power scattered from the feed, and that reflected from within the attached plumbing. The plumbing reflection is caused by incompletely matched plumbing elements and is measurable by standard transmission-line techniques; it returns to the focusing element with a power distribution given by the ordinary radiation pattern of the feed. On the other hand, determination of the feed reflection during reception must be based on the aspect presented by the feed to the wave incident on it from the focusing element. Its magnitude cannot be determined by measuring the impedance match of the feed to its waveguide during transmission (Ref. 3,18, 19), nor, in general, is its distribution that of the ordinary radiation pattern of the feed.

Not only must the feed reflection in question be determined in reception, rather than transmission, but the incident wave which must be considered is the focused one. For example, the horn-type feed shown in Fig. 8 has been matched as seen from its waveguide, but a uniform wave being received by the horn will be partially reflected by the edges parallel to the electric field. However, if the received wave is focused, the reflection

is critically dependent on the nature of the focused field. For instance, if the focused field has amplitude minimums at the horn edges, the feed reflection will be small; conversely, if it should have maximums, the reflection will be large.

The plumbing reflection is usually made as small as possible; this may be done by the usual waveguide matching techniques. Reduction of the feed reflection during reception may be accomplished by minimizing those metallic edges which are parallel to the focused electric field and are located in regions of high field intensity.

When combined with a reflection from the focusing element, the reflection from the waveguide assembly causes some effects which are described in the next section.

Combination of Reflections from the Focusing Element and the Waveguide Assembly

When a wave is incident on the complete antenna, the power entering the waveguide assembly is the sum of direct power and spurious power. The spurious power arises from a multiple reflection, as shown in Fig. 9: this power is first reflected from the waveguide assembly and then reflected from the focusing element, and finally enters the waveguide assembly. Here, the spurious signal and the direct signal add and, depending on their relative phase, the total signal is increased or decreased. Thus the apparent antenna gain may be increased or decreased by the spurious signal.

Since the path lengths of the direct and spurious signals differ, their relative phase will change with frequency. The antenna gain will therefore vary sinusoidally with frequency as shown in Fig. 9, at a rate which depends on the distance between the focusing element and the waveguide assembly. Formulas for the magnitude and period of the gain variation are given; these are based on the following simplifying assumptions: small and constant reflections from the waveguide assembly and focusing element, and a distribution of energy reflected from the feed equal to that given by the normal radiation pattern of the feed.

Since the spurious signal is proportional to the product of the focusing-element and waveguide-assembly reflections, reduction of either one will reduce the variation of antenna gain. Reduction of the first may be accomplished by any of the methods described in the first section, except that those methods which operate within the plumbing may not be designed to eliminate the effect of the multiple reflection between the feed and the focusing element. Reduction of the second may be accomplished by reducing the reflection from either the plumbing or the feed, or both. When the plumbing and feed are separated by a large distance, the greatest average improvement over a wide frequency band is obtained by reducing the larger of the two reflections which comprise that of the waveguide assembly.

In a lobing antenna, the doubly-reflected spurious power may cause an error of direction-finding information. This occurs when the focusing element is tilted with respect to the feed, as shown in Fig. 10. The spurious power, reflected from the waveguide assembly and then from the focusing element, arrives at the feed off-center, and appears to come from a distant target located off the antenna axis by twice the tilt angle of the focusing element. This spurious target, or ghost, combines with the real target to yield an apparent target whose direction is different from that of the real target by some fraction of the tilt angle. As before, this direction error varies sinusoidally with frequency, at a rate which depends on the distance between the focusing element and the waveguide assembly.

In Fig. 11, curves are shown of the spread of the direction error vs. tilt of the focusing element. Ideally, the spread goes to zero when the tilt is zero; however, if either the focusing element or the feed is asymmetrical about its own axis, the spread does not go to zero.

Reduction of the direction-error spread may be accomplished by either improving the symmetry of the antenna, or by reducing the magnitude of the doubly-reflected signal, or both. As illustrated in Fig. 11, complete antenna symmetry requires symmetrical orientation of the antenna components relative to each other

and also symmetry of each individual component. These symmetries may be obtained by first limiting the antenna design to one which is nominally symmetrical, and then by providing close manufacturing tolerances, or perhaps adjustments, to assure that this symmetry is actually obtained.

The magnitude of the doubly-reflected signal may be reduced by any of the methods previously discussed, except that those methods which operate within the plumb line may not reduce the direction error. When attempting to reduce the doubly-reflected signal, care must be taken not to increase the asymmetry. For instance, tilt of the focusing element will reduce the reflection but will degrade symmetry, and the direction error may actually be increased.

Conclusion

When a microwave antenna of the focusing type is to operate over a wide frequency band, reflections within the antenna are likely to cause a degradation of performance. Reflection from a boundary surface of the focusing element back to the waveguide assembly may cause troubles such as instability of the transmitting oscillator. A multiple reflection between the waveguide assembly and the focusing element may cause a sinusoidal variation of antenna gain and direction error with frequency.

References

- (1) C. C. Cutler, "Parabolic-Antenna Design for Microwaves," Proc. I.R.E., vol. 35, Nov. 1947, pp. 1284-1294. (Mentions reflection from the dish into the feed, and reduction of the reflection by means of either an apex plate or an offset section of the dish, pp. 1288-1289.)
- (2) A. B. Hippard, H. Elson, "The Elimination of Standing Waves in Aerials Employing Paraboloidal Reflectors," Jnl. I.E.E., vol. 93, III A, 1946, pp. 1531-1535. (Calculation of reflection from dish; design of apex plate; calculation of spurious radiation from apex plate.)

- (3) S. Silver, "Microwave Antenna Theory and Design," Radiation Laboratory Series, vol. 12, McGraw-Hill, 1949. (Both a general and a practical calculation of reflection from a dish, pp. 155-158 and 439-442; reflection of lens surfaces, pp. 401 and 410-412; offset section of dish, pp. 453-457; aberration caused by tilt of dish, pp. 487-491; design of apex plate, pp. 443-447; polarization crossing of reflection from dish, pp. 447-448; antenna scattering and the resulting sinusoidal error of a particular antenna-gain measurement, pp. 587-592.)

- (4) D. W. Dry, F. K. Goward, "Aerials for Centimetre Wave-Lengths," Cambridge University Press, 1950. (Calculation of reflection from dish, p. 59; design of apex plate, pp. 59-61; aberration caused by tilt of dish, pp. 62-64; use of hoghorn to obtain tilt, p. 91.)

- (5) N. Marcuvitz, "Waveguide Handbook," Radiation Laboratory Series, vol. 10, McGraw-Hill, 1951. (Formulas for reflection from metal-strip media, pp. 280-293.)

- (6) S. B. Cohn, "Analysis of the Metal-Strip Delay Structure for Microwave Lenses," Jnl. Appl. Phys., vol. 20, March 1949, pp. 257-262.

- (7) B. A. Lengyel, "Reflection and Transmission at the Surface of Metal-Plate Media," Jnl. Appl. Phys., vol. 22, March 1951, pp. 265-276. (Phase-advancing media.)

- (8) D. R. Hamilton, J. K. Knipp, J. B. H. Kuper, "Klystrons and Microwave Triodes," Radiation Laboratory Series, vol. 12, McGraw-Hill, 1948. (The "sink" region on the Rieke diagram of a klystron, p. 418.)

- (9) G. B. Collins, "Microwave Magnetrons," Radiation Laboratory Series, vol. 6, McGraw-Hill, 1948, pp. 320-329.

- (10) J. F. Hull, G. Novick, R. Cordray, "How Long-Line Effect Impairs Tunable Radar," Electronics, vol. 27, no. 2, Feb. 1954, pp. 168-173.

- (11) H. T. Friis, "Microwave Repeater Research," Bell Syst. Tech. Jnl., vol. 27, April 1948, pp. 183-246. (Horn reflector to obtain tilt, pp. 202-204; distortion of information caused by reflections, pp. 242-245.)

(12) W. E. Kock, "Metal-Lens Antennas," Proc. I.R.E., Nov. 1946, pp. 828-836. (Phase-advancing lens types. Mentions lens tilt, p. 833.)

(13) W. E. Kock, "Metallic Delay Lenses," Bell Syst. Tech. Jnl., vol. 27, Jan. 1948, pp. 58-82. (Mentions lens tilt and stepping, pp. 72-74, 77.)

(14) W. K. Kahn, "Impedance Matching of an X-Band Lens Face by a Dielectric Cover," Wheeler Laboratories Report 486 to Bell Telephone Laboratories, December 1951. (Matching the flat outer surface of a phase-advancing type lens.)

(15) W. E. Kock, "Path-Length Microwave Lenses," Proc. I.R.E., August 1949, pp. 852-855.

(16) W. C. Jakes, "Broad-Band Matching with a Directional Coupler," Proc. I.R.E.,

vol. 40, October 1952, pp. 1216-1218.

(17) C. L. Hogan, "The Microwave Gyrotator," Bell Syst. Tech. Jnl., vol. 31, Jan. 1952, pp. 1-31. (Use of ferrite to obtain only one-way propagation, p. 25.)

(18) A. B. Pippard, O. J. Burrell, E. E. Cromie, "The Influence of Re-radiation on Measurements of the Power Gain of an Aerial," Jnl. I.E.E. vol. 93, III A, 1946, pp. 720-722. (Describes reflection from some antennas during reception and the resulting sinusoidal error of a particular antenna-gain measurement.)

(19) R. M. Redheffer, "Microwave Antennas and Dielectric Surfaces," Jnl. Appl. Phys., vol. 20, April 1949, pp. 397-411. (Reflection from an antenna during reception and transmission; describes a situation which involves a sinusoidal variation of antenna gain.)

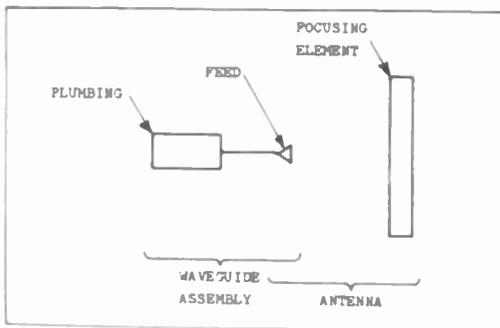


Fig. 1 - General antenna and plumbing.

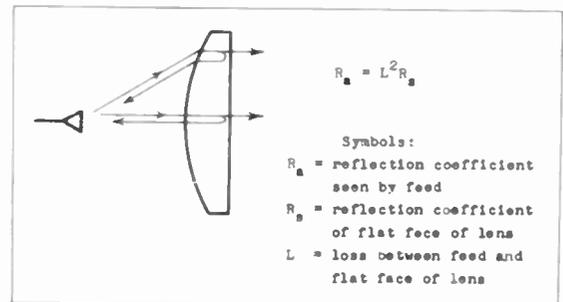


Fig. 3
Reflection into feed from lens having flat outer face.

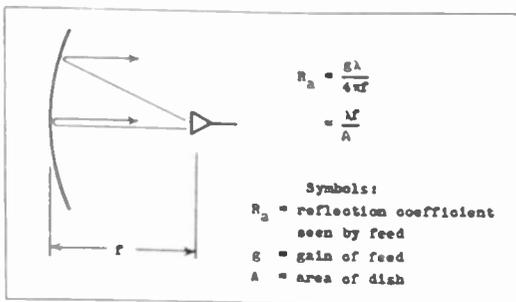


Fig. 2
Reflection into feed from paraboloidal dish.

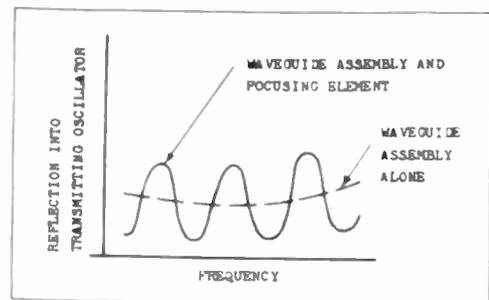


Fig. 4 - Typical reflection into oscillator.

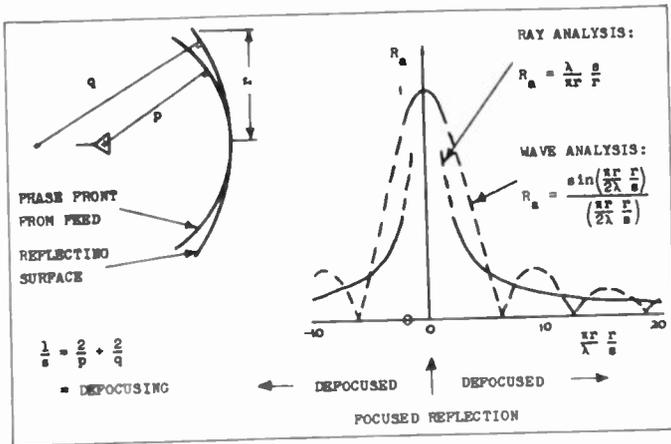


Fig. 5
 General formulas for reflection into feed.

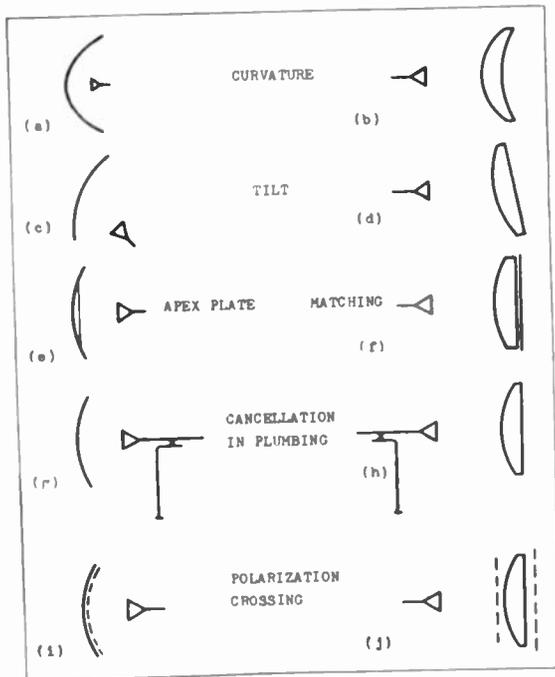


Fig. 6 - Techniques for reducing reflection.

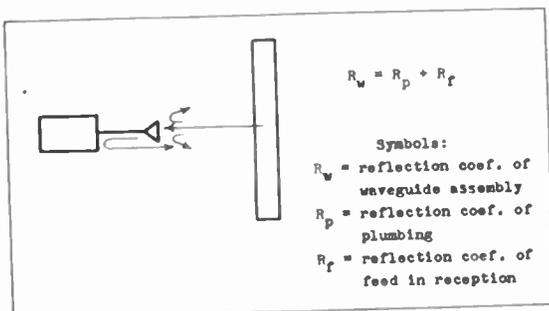


Fig. 7
 Reflection from waveguide assembly.

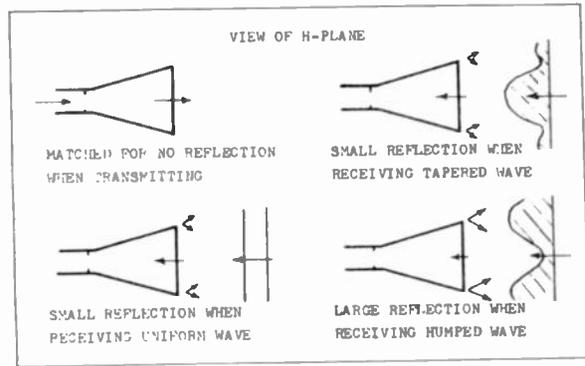


Fig. 8 - Reflection from a feed horn.

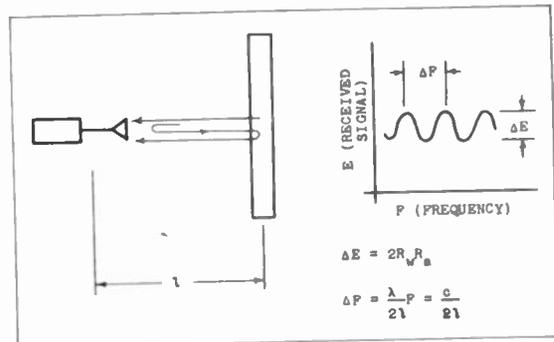


Fig. 9
 Variation of antenna gain with frequency.

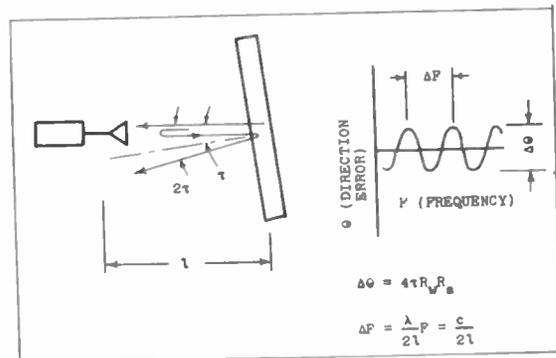


Fig. 10
 Variation of direction-error with frequency.

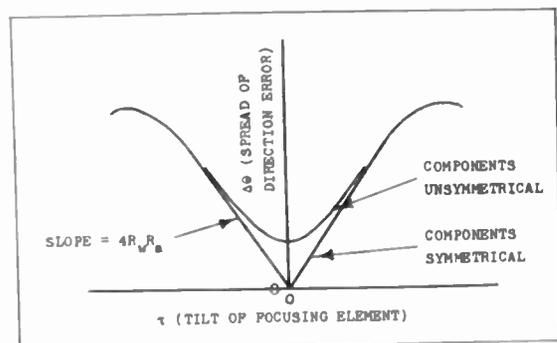


Fig. 11 - Direction-error spread vs. tilt.

SURFACE MATCHING OF DIELECTRIC LENSES

by

E. M. T. JONES

AND

S. B. COHN

of

STANFORD RESEARCH INSTITUTE
STANFORD, CALIFORNIA

Summary

Two methods of cancelling the surface reflections of dielectric lenses are described in this paper. The first utilizes a simulated quarter-wave matching layer, and the second a reactive wall embedded within the dielectric. The reactive wall may take a variety of physical forms, such as arrays of thin conducting discs, which have a capacitive reactance, or arrays of thin wires, which have an inductive reactance. Surface matching is obtained when the discs are placed approximately $3/8$ wavelength inside the lens, or the wires $1/8$ wavelength. Curves are presented that show how the reflections at the air and dielectric boundary are reduced for various angles of incidence and polarization when quarter-wave layer and reactive-wall matching are employed.

The reactance of the array of discs for waves incident at various angles and polarizations is computed by means of Bethe's small aperture theory, and Babinet's principle. Measurements in waveguide of the reactance of an array of circular discs for various angles of incidence and for both E - and H -plane polarization show close agreement with the theory.

Introduction

The biggest disadvantage in using natural dielectric lenses as focusing elements for microwave antenna systems is that a portion of the energy incident on the air-dielectric interfaces is not transmitted. This decrease in transmission results in a net decrease in the power gain of the system, while the reflected energy raises the input standing wave ratio and can cause various anomalies in the far-zone diffraction pattern. If the surface of the lens is matched in some fashion, these difficulties disappear and the lens becomes a much more useful microwave antenna. Two basic methods of achieving a surface match are considered in this paper. One entails the use of quarter-wave sheets on the lens surface, while the other involves the placing of reactive walls inside the surface of the lens. These two methods will be discussed in detail in the following paragraphs.

Matching Methods

The first method requires a quarter-wave layer having a dielectric constant intermediate to that of

air and the main body of the lens. This layer could consist of a foamed dielectric, but such a material is believed to be impractical for this purpose due to physical limitations. The quarter-wave layer may also be simulated by perturbing the surface of the dielectric lens, as illustrated in Fig. 1. Dr. Tetsu Morita of this laboratory has determined by theoretical and experimental means that these surface perturbations behave similarly to a homogeneous dielectric. A possible production technique would be to die cast the entire lens complete with this surface.

In the second method the reactive wall can assume a variety of forms. A rectangular grid of thin wires or arrays of thin conducting discs are typical physical elements. The wires have an inductive reactance and therefore must be imbedded about $1/8$ wavelength within the dielectric to match the surface. The discs on the other hand, have a capacitive reactance and, therefore, must be imbedded about $3/8$ of a wavelength in the dielectric to cancel surface reflections. A practical method of construction for the reactive wall is to print conducting obstacles or inductive strips on the lens surface, and then to cement over this the necessary additional layer of dielectric material.

Theoretical Performance of Various Matching Methods

The relative performance of various matching techniques can be evaluated most easily in terms of the power reflection coefficient of the air-dielectric interface. For quarter-wave matching sheets let the external region be region 1, the quarter-wave sheet region 2 and the lens region 3. Then R_{12} and R_{23} the reflection coefficients⁽¹⁾ measured at the interfaces of regions 1 and 2 and regions 2 and 3 are both real. The power reflection coefficient R^2 becomes

$$R^2 = \frac{(R_{12} + R_{23})^2 - 4R_{12}R_{23} \sin^2\phi}{(1 + R_{12}R_{23})^2 - 4R_{12}R_{23} \sin^2\phi} \quad (1)$$

while the electrical length ϕ of the quarter wave sheet is

$$\phi = \frac{2\pi L}{\lambda} \sqrt{n_2^2 - \sin^2\theta_1} \quad (2)$$

where

- λ = free-space wavelength
- θ_1 = angle of incidence in region 1
- n_2 = index of refraction of the quarter-wave sheet
- L = thickness of the quarter-wave sheet

When reactive wall matching is used, let the region outside the lens be region 1, the region between the lens surface and the matching wall be region 2 and the region within the matching wall be region 3. Then R_{12} is real and $R_{23} = a_{23} + jb_{23}$. The power reflection coefficient in this case becomes

$$R^2 = \frac{R_{12}^2 + 2R_{12}(a_{23} \cos 2\phi + b_{23} \sin 2\phi) + a_{23}^2 + b_{23}^2}{1 + 2R_{12}(a_{23} \cos 2\phi + b_{23} \sin 2\phi) + R_{12}^2(a_{23}^2 + b_{23}^2)} \quad (3)$$

while the electrical spacing ϕ of the reactive wall from the lens surface is

$$\phi = \frac{2\pi L}{\lambda} \sqrt{n^2 - \sin^2 \theta_1} \quad (4)$$

where

$$a_{23} + jb_{23} = \frac{-\left(\frac{B}{Y_2}\right)^2 - j2\frac{B}{Y_2}}{4 + \left(\frac{B}{Y_2}\right)^2}$$

Y_2 = the characteristic admittance of the dielectric in region 2 (i.e. the ratio of the components of magnetic to electric field that are parallel to the reactive wall).

- B = susceptance of the reactive wall
- n = refractive index of the lens
- L = separation of the reactive wall from the lens surface.

The conditions for matching a lens with a quarterwave layer on its surface are seen from (1) to be that $R_{12} = R_{23}$ and $\phi = 90^\circ$. When a reactive wall is used for surface matching, it can be determined from (3) that the spacing of the wall from the dielectric surface is determined by

$$\frac{Y_2}{Y_1} = \tan^2 \phi \quad (5)$$

while the normalized susceptance of the wall is given by

$$\frac{B}{Y_2} = \cot 2\phi \quad (6)$$

where

Y_1 = the characteristic admittance of free space (i.e. the ratio of the components of magnetic to electric field that are

parallel to the lens surface)

The single boundary power reflection coefficient for a dielectric lens of refractive index 1.57 matched by a quarter-wave sheet at normal incidence is plotted in Fig. 2 by means of (1) and (2). For comparison the power reflection coefficient for an unmatched surface is also shown. For perpendicular polarization the quarter-wave matching sheet reduces the reflected power for all angles of incidence while for parallel polarization it reduces the reflected power up to angles of 51 degrees and increases it above 51 degrees.

A rectangular array of wires, as shown in Fig. 3, will appear inductive to plane waves incident upon it because the inductive susceptance of the wires running in the x direction is orders of magnitude greater than that of the capacitive susceptance of the wires running in the y direction. Therefore, this capacitive susceptance will be neglected in the following discussion. For the values of susceptance needed to match the surface of ordinary dielectric lenses the required wire diameter D will be so small that in the analysis (and also in practice) it will be permissible to replace the wires of diameter D by flat strips of width $2D$.

The reactance of an array of thin strips of width $2D$ (oriented in the x direction) to a wave polarized in the plane of incidence, can be obtained by use of Babinet's transformation from the solution for a capacitive slit in a wave-guide operating in a TE_{10} mode. The solution for the capacitive slit is given by Marcuvitz.⁽²⁾ Applying the transformation the inductive reactance of the wires becomes

$$\frac{X}{Z_2} = \frac{s}{\lambda_2} \cos \theta_2 \left[\ln \left(\csc \frac{\pi D}{s} \right) + F \left(\theta_2, \frac{s}{\lambda_2}, \frac{D}{s} \right) \right] \quad (7)$$

where

- $Z_2 = 1/Y_2$
- s = center to center spacing of the wires
- λ_2 = wavelength in the dielectric medium surrounding the wires.

The correction term F has a small value for normal incidence, and its value decreases further as the angle of incidence increases.

The reactance of this array of wires (or thin strips) for an incident wave polarized perpendicular to the plane of incidence has been computed by MacFarlane.⁽³⁾ He obtains

$$\frac{X}{Z_2} = \frac{s}{\lambda_2} \cos \theta_2 \left[\ln \frac{s}{\pi D} + F' \left(\theta_2, \frac{s}{\lambda_2} \right) \right] \quad (8)$$

The correction term F' has the same value as F for normal incidence but has much greater values for angles off the normal.

In Fig. 4, the single boundary power reflection coefficient is plotted from (3), (4), (7), and (8) for a wave incident on a dielectric lens of refractive index 1.57 matched at normal incidence by a grid of wires spaced 0.45 wavelengths between

centers. The power reflection coefficient of an unmatched lens is also included in Fig. 4. Here it is seen that with matching the power reflection is always greater for parallel polarization than for perpendicular polarization. This grid does not give as small a power reflection at angles off the normal as does the quarter-wave section.

The susceptance of a reactive wall composed of thin circular discs arranged in hexagonal array as shown in Fig. 5 will be analyzed later in this paper. The single boundary power reflection coefficient of a dielectric lens of refractive index of 1.57 that has been matched by such an array of discs spaced on half-wavelength centers, together with the power reflection coefficient for the unmatched lens is shown in Fig. 6. Here it is seen that for perpendicular polarization, matching reduces the power reflection for all angles of incidence, while for parallel polarization the power reflection is reduced for angles up to 42° , and is increased for angles greater than 42° .

Theoretical Calculation of the Susceptance of an Infinite Array of Thin Circular Discs

The susceptance of an array of small, thin metal obstacles can be computed for various angles of incidence with the help of Bethe's small-aperture theory, and Babinet's principle as follows.

Consider an array of obstacles immersed in a uniform dielectric medium, of relative permeability μ_2 and relative permittivity ϵ_2 , lying in the plane $z = 0$ (as shown in Fig. 5) with two plane waves incident on it at angles θ_2 and $-\theta_2$, each wave being polarized perpendicular to the plane of incidence. The field components of the resultant wave in Gaussian units are

$$\begin{aligned} E_x &= \sqrt{\frac{\mu_2}{\epsilon_2}} E_0 \cos(k_2 y \sin \theta_2) \epsilon^{j(\omega t - k_2 z \cos \theta_2)} \\ H_y &= E_0 \cos \theta_2 \cos(k_2 y \sin \theta_2) \epsilon^{j(\omega t - k_2 z \cos \theta_2)} \\ H_z &= j E_0 \sin \theta_2 \sin(k_2 y \sin \theta_2) \epsilon^{j(\omega t - k_2 z \cos \theta_2)} \end{aligned} \quad (9)$$

where

$$k_2 = \frac{2\pi}{\lambda} \sqrt{\mu_2 \epsilon_2}$$

Imagine for the moment another wave incident on the array identical to that described in (9) but propagating in the negative z direction. It then becomes obvious that a magnetic wall can be inserted in the plane of the obstacles since the sum of the tangential H field of the two waves in this plane is zero. With the magnetic wall in place, the array of metal obstacles appears as a pure susceptance to the wave traveling in the positive z direction. The value of this susceptance is equal to one-half the value of the susceptance of the metal obstacles be-

fore the magnetic wall was inserted.

Babinet's principle states that if a solution $E(x, y, z)$ and $H(x, y, z)$, to Maxwell's equation in a medium is known, then another solution may be obtained by replacing $E(x, y, z)$ by $\sqrt{\mu_2/\epsilon_2} H'(x, y, z)$ and $H(x, y, z)$ by $-\sqrt{\epsilon_2/\mu_2} E'(x, y, z)$. Obviously for this solution to hold at boundaries, it is also necessary to replace all electric walls by magnetic walls and conversely. As a corollary, it is seen that all susceptances are replaced by reactances.

Applying Babinet's principle to the case of a wave terminated by the reactive wall, one sees that the magnetic wall is replaced by the electric wall and the electric-wall obstacles are replaced by magnetic-wall obstacles. The reactance $2X'/Z_2'$ of this composite wall is equal to the susceptance $B/2Y_2$ of the previous wall.

The field components of the incident wave after the Babinet transformation are

$$\begin{aligned} H'_x &= E_0 \cos(k_2 y \sin \theta_2) \epsilon^{j(\omega t - k_2 z \cos \theta_2)} \\ E'_y &= -\sqrt{\frac{\mu_2}{\epsilon_2}} E_0 \cos \theta_2 \cos(k_2 y \sin \theta_2) \epsilon^{j(\omega t - k_2 z \cos \theta_2)} \\ E'_z &= -j \sqrt{\frac{\mu_2}{\epsilon_2}} E_0 \sin \theta_2 \sin(k_2 y \sin \theta_2) \epsilon^{j(\omega t - k_2 z \cos \theta_2)} \end{aligned} \quad (10)$$

Now the magnetic wall can be removed by the reverse procedure used to insert it, and there remains an electric wall with small apertures. The reactance X' of the apertures is related to the susceptance B of the obstacles according to the relation

$$\begin{aligned} \frac{2X'}{Z_2'} &= \frac{1}{2} \frac{B}{Y_2} \\ \text{or} \quad \frac{-Y_2'}{H'} &= \frac{1}{4} \frac{B}{Y_2} \end{aligned} \quad (11)$$

In order to have the array of apertures excite a transmitted wave traveling only in the direction θ_2 it is necessary that the center-to-center spacing s be related to the angle of incidence θ_2 by the relation

$$s \leq \frac{\lambda_2}{1 + |\sin \theta_2|} \quad (12)$$

If s were to exceed this limit one or more undesired waves in other directions would emanate from the array of apertures.

Because there is no variation of H'_x in the x direction, magnetic walls can be inserted at the planes $11'$ without affecting the incident wave. It can also be seen from (10) that there are planes parallel to the xz plane a distance $A = \lambda_2/2 \sin \theta_2$ apart where H'_z is zero. Therefore, magnetic walls

also can be inserted here. The behavior of the cell bounded by these walls will be the same as that of the infinite array.

Bethe⁽⁴⁾ has shown that the normalized susceptance B'/Y_2' of a single small aperture in a thin metal wall of such a cell can be computed in Gaussian units from the relation

$$\frac{Y_2'}{B'} = \frac{-2\pi}{\lambda_2 S} [M_x H_x'^2 + M_y H_y'^2 + P_z E_z'^2] \quad (13)$$

His formula can be extended to the present case of m identical small apertures by summing up the contributions from the m apertures. One obtains

$$\frac{Y_2'}{B'} = \frac{-2\pi}{\lambda_2 S} \sum [M_x H_{x,m}'^2 + M_y H_{y,m}'^2 + P_z E_{z,m}'^2] \quad (13a)$$

Where

M_x and M_y = magnetic polarizabilities of an aperture the same size as the obstacle

P_z = the electric polarizability of an aperture the same size as the obstacle

$H_{x,m}'$ and $H_{y,m}'$ = tangential components of magnetic field at the m th aperture

$E_{z,m}'$ = normal component of electric field at the m th aperture

$$S = \int_0^B \int_0^A \bar{n} \cdot \bar{E}' \times \bar{H}' \cdot dx dy.$$

For an integral number of apertures within the cell substitution of (10) in (13) gives

$$\frac{Y_2'}{B'} = \frac{-4\pi}{\sqrt{3} \lambda_2 S^2 \cos \theta_2} [M_x - P_z \sin^2 \theta_2] \quad (14)$$

which is the normalized susceptance of an array of apertures for a wave incident on them polarized in the plane of incidence. Although (14) has been derived for the specific case of an integral number of apertures within a unit cell it holds for non-integral numbers of apertures between the planes 22' as long as (12) is satisfied. To find the susceptance of the array of discs with an incident wave polarized perpendicular to the plane of incidence the use of (11) gives

$$\frac{B}{Y_2} = \frac{16\pi}{\sqrt{3} \lambda_2 S^2 \cos \theta_2} [M_x - P_z \sin^2 \theta_2] \quad (15)$$

Proceeding in the same fashion it can be shown that the susceptance of the array of obstacles for an incident wave polarized in the plane of incidence is

$$\frac{B}{Y_2} = \frac{16\pi \cos \theta_2}{\sqrt{3} \lambda_2 S^2} M_y \quad (16)$$

The value of the electric and magnetic polarizabilities of a circular aperture of diameter d have been computed by Bethe.⁽⁴⁾ He finds

$$M_x = M_y = 2P_z = \frac{d^3}{6}$$

The values for other shapes have been measured by electrolytic tank means.^(5,6) A correction formula for the polarizability of wave guide irises whose diameter is an appreciable fraction of the wave length has been found to give good agreement with theory.⁽⁷⁾ This correction method can also be applied to an array of obstacles, with the aid of Babinet's transformation. When a wave is incident on the array with the electric field perpendicular to the plane of incidence, the normalized susceptance becomes, using this correction method

$$\frac{B}{Y_2} = \frac{16\pi}{\sqrt{3} \lambda_2 S^2 \cos \theta_2} \left[\frac{M_x}{1 - \left(\frac{\lambda_{c1}}{\lambda_2}\right)^2} - \frac{P_z \sin^2 \theta_2}{1 - \left(\frac{\lambda_{c2}}{\lambda_2}\right)^2} \right] \quad (17)$$

When a wave is incident on the array with the magnetic vector perpendicular to the plane of incidence the normalized susceptance is

$$\frac{B}{Y_2} = \frac{16\pi \cos \theta_2}{\sqrt{3} \lambda_2 S^2} \cdot \frac{M_y}{1 - \left(\frac{\lambda_{c1}}{\lambda_2}\right)^2} \quad (18)$$

here λ_{c1} is the cutoff wavelength of the TE_{11} mode in a circular wave guide having the same radius as the obstacle and λ_{c2} is the cutoff wavelength of the TM_{01} mode in a circular waveguide having the same radius as the obstacle.

Phase Shift Through Surfaces Matched at One Incident Angle

When quarter-wave sheets or reactive walls are used to cancel the surface reflection from dielectric lenses they introduce a certain amount of phase shift to a wave passing through the lens surface that would not be present in the absence of matching devices. For quarter-wave sheets this excess phase shift results from a combination of increased path length due to the addition of the sheet to the lens, as well as multiple reflections in the quarter-wave sheet. For a normal incidence match the phase shift ψ_1 due to the first effect can be easily shown to be

$$\psi_1 = \frac{2\pi L}{\lambda \cos \theta_2} [\sqrt{n} - \cos(\theta_1 - \theta_2)] \quad (19)$$

while that due to the second effect is

$$\psi_2 = -\tan^{-1} \frac{R_{12} R_{23} \sin 2\phi}{1 + R_{12} R_{23} \cos 2\phi} \quad (20)$$

Figure 7 shows this total excess phase shift as a function of incident angle and polarization for the same conditions as Fig. 2. It is seen to

be not excessive even for grazing angles of incidence.

When reactive walls are used to achieve a match, the excess phase shift results only from multiple reflections between the wall and the air dielectric interface. This phase shift can be computed from the formula

$$\psi = -\tan^{-1}\left(\frac{b_2}{1+n_2}\right) - \tan^{-1}\left[\frac{P_{12}(a_{23} \sin 2t - b_{23} \cos 2t)}{1+R_{12}(a_{23} \cos 2t + b_{23} \sin 2t)}\right] \quad (21)$$

Equation 21 has been plotted in Fig. 8 and 9 for a wire grid inductive wall and a circular-disc capacitive wall respectively. The parameters for these curves are the same as those for Figs. 4 and 6. In each of these cases the excess phase shift is smaller than $\lambda/16$.

Design Information for Circular Disc Wall

In order to achieve a perfect match at various incident angles and polarizations it is necessary to vary the spacing of the reactive wall from the lens surface as well as the reactance of the wall. Figure 10 shows the spacing of a capacitive wall necessary to achieve a match as computed from (5) as well as the disc diameters necessary to achieve a match as computed from (6) and the experimental data presented below.

Measurement of Disc Susceptance in Waveguide

When waves that have only a vertical component of electric field are incident upon the array in Fig. 5, electric walls can be inserted along the planes 11' without affecting the performance of the array. If these waves are incident at angles θ_2 and $-\theta_2$ with respect to the array normal, there are planes perpendicular to 11' such as 22' spaced a distance $A = \lambda_2/2 \sin \theta_2$ apart where electric walls can be inserted that also image all the other obstacles in the array. The susceptance of the obstacles within the confines of these walls is exactly the same as the susceptance of the infinite array of obstacles to a wave incident at an angle θ_2 polarized perpendicular to the plane of incidence

When waves that have only a vertical component of magnetic field are incident on the array of obstacles at angles θ_2 and $-\theta_2$, it is possible to insert magnetic walls at the same positions occupied by the electric walls in Fig. 5. The susceptance of the obstacles within the confines of this magnetic-wall waveguide is exactly the same as the susceptance of the infinite array of obstacles to a plane wave incident at an angle θ_2 polarized parallel to the plane of incidence. Since magnetic-wall waveguide is not available, it is necessary to apply the Babinet transformation and replace the magnetic-wall waveguide by an electric-wall waveguide and the array of obstacles in space by an array of apertures in a metal screen. Then the normalized reactance X'/Z_2'

of these holes is related to the normalized susceptance of the obstacles by (11).

Figure 11 shows the non-standard waveguide cross sections, with obstacles and apertures in place, used in determining the susceptance of a hexagonal array of circular discs for various angles of incidence and polarization. In each case the center-to-center spacing of the discs and apertures is 1.010 inches.

The test sections were machined from aluminum, and the inside dimensions are accurate to within ± 0.001 inch. The obstacle and aperture patterns were punched from copper sheets, and shadowgraph measurements have shown that the average radii of the circular obstacles and apertures are all within 0.63% of the desired radii. For each waveguide test section, discs and apertures of radii 0.202 in., 0.268 in., and 0.333 in., are available. For each radius, pattern thicknesses of 0.002 in., 0.005 in., 0.010 in., and 0.020 in. are available, making a total of 12 obstacle and 12 aperture patterns for each test section.

Figure 12 shows a block diagram of the test set up used in measuring the susceptance of the hexagonal array of circular apertures. Note that there is a tuner on each side of the test section as well as one preceding the slotted line. By properly adjusting the tuners at each end of the test section it is possible to tune out the mismatch caused by the abrupt change in waveguide cross section at the junctions of the test sections and the standard size waveguide on either side. The tuner preceding the slotted line transforms the generator impedance to the waveguide characteristic impedance. Although the aperture susceptance can be determined in terms of the input standing wave ratio the most accurate value of susceptance of these apertures can be determined by measuring the ratio t of the power transmitted with the apertures in place to that transmitted when they are removed. The normalized susceptance of the apertures B'/Y_2' is given in terms of t by

$$-\frac{B'}{Y_2'} = 2 \sqrt{\frac{1-t}{t}} \quad (22)$$

A block diagram of the test set up used to measure the susceptance of the obstacles corresponding to waves incident on the array with perpendicular polarization is shown in Fig. 12. The detailed measuring procedure used is as follows: An obstacle pattern is placed in the mid-plane of the test section, and a shorting plunger is placed one quarter guide wave length behind it. A position of minimum voltage is then determined in the slotted line. Next the obstacle pattern is removed from the test section and the shorting plunger is displaced a distance x until the voltage minimum is observed at the same place in the slotted line. The normalized susceptance of the obstacles is then computed from the relation

$$\frac{B}{Y_2} = \tan \frac{2\pi x}{\lambda g} \quad (23)$$

To determine the susceptance of the obstacles and apertures for zero thickness the values of susceptance for each obstacle and aperture pattern is plotted as a function of thickness and the curve extrapolated back to zero thickness. This procedure is necessary since the Babinet transformation used to establish the theoretical basis of these measurements applies only to infinitesimally thin obstacles and apertures.

Experimental Values of Disc Susceptance

It is estimated that any individual measurement of obstacle susceptance by these methods is within 2% of the true value. A curve drawn through the measured points for various angles of incidence should be within 1% of the true value.

In Fig. 13 are shown typical measured values at one frequency of disc susceptance for various disc radii and polarizations. It is seen that for the two smaller disc radii the measured points lie very close to the theoretical curves. For the largest disc radius the experimental points are considerably below the theoretical values, indicating that the high frequency corrections proposed in (17) and (18) are not valid for a disc radius of 0.165λ .

Measurements of the susceptance of these discs have been made at a number of frequencies. As shown previously, each frequency corresponds to a different angle of incidence in each waveguide cross section. It has been found that the error between experiment and theory does not depend on the center-to-center spacing of the discs in terms of wavelength or on the angle of incidence, but only on the disc radius in terms of wavelength. In Fig. 14 is shown the average error between the theoretical curves and experimental points as a function of disc diameter for each polarization.

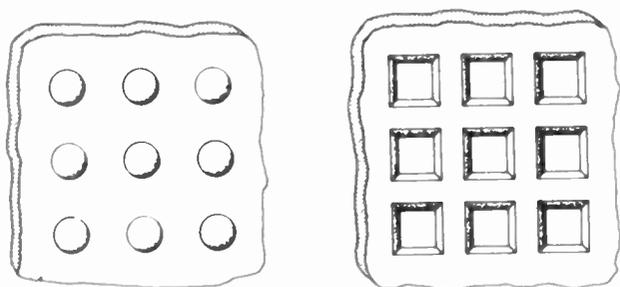


Fig. 1 - Simulated Quarter-Wave Sheets

Conclusion

Two methods of reducing the surface reflections from dielectric lenses have been discussed. One employs simulated quarter wave plates on the lens surface, while the other uses a reactive wall embedded within the dielectric. The theoretical calculations of the performance of lenses matched by these methods indicate that substantially improved performance can be obtained by their use.

Acknowledgment

The work reported in this paper was supported by the Signal Corps under Contract No. DA-36-039-sc-52595

References

1. J. A. Stratton, "Electromagnetic Theory," McGraw-Hill Book Co., p. 492; 1941.
2. N. A. Marcuvitz, "Wave-Guide Handbook," McGraw-Hill Book Co., p. 218, Eq. (2A); 1951.
3. G. G. MacFarlane, "Surface Impedance of an Infinite Parallel-Wire Grid at Oblique Angles of Incidence"; Jour. I.E.E., pp. 1523-1526; 1947.
4. H. A. Bethe, "Lumped Constants for Small Irises," Radiation Laboratory Report 43-22; March 24, 1943.
5. S. B. Cohn, "Determination of Aperture Parameters by Electrolytic Tank Measurement," I.R.E. Vol. 39, pp. 1416-1421; 1951.
6. S. B. Cohn, "Electric Polarizability of Apertures of Arbitrary Shape," I.R.E. Vol. 40, pp. 1069-1071; Sept. 1952.
7. S. B. Cohn, "Microwave Coupling by Large Apertures," I.R.E. Vol. 40, pp. 696-699; 1952.

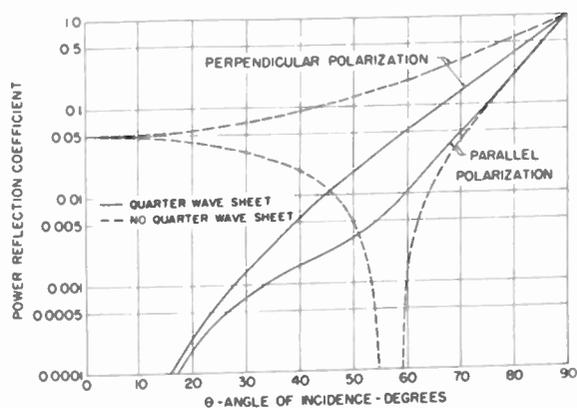


Fig. 2

Single Boundary Power Reflection Coefficient for a Dielectric Lens of Refractive Index 1.57 With and Without a Quarter-Wave Matching Sheet.

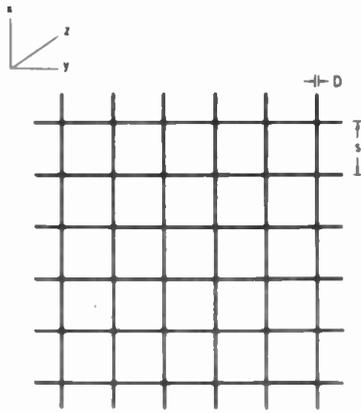


Fig. 3- Inductive Wall Composed of a Wire Grid.

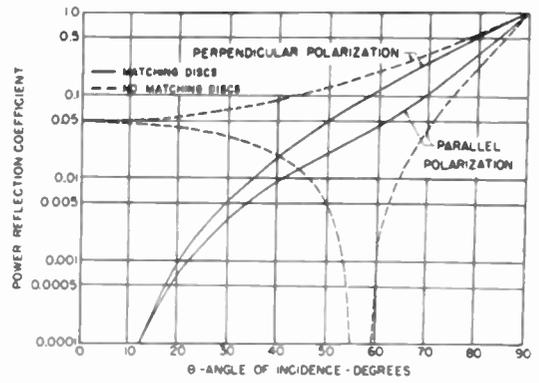


Fig. 6
Single Boundary Power Reflection Coefficient for a Dielectric Lens of Refractive Index 1.57 With and Without a Circular Disc Matching Wall.

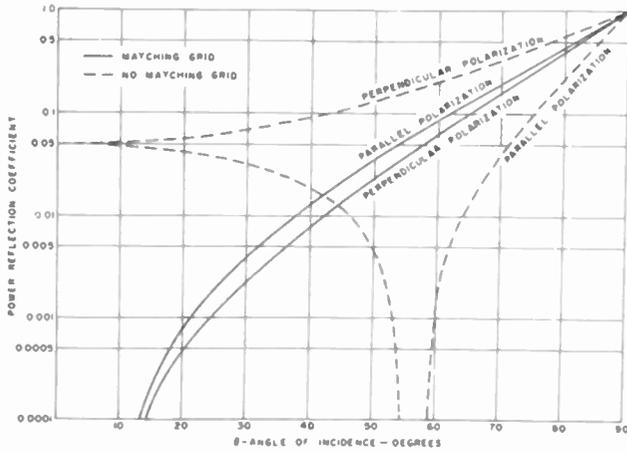


Fig. 4
Single Boundary Power Reflection Coefficient for a Dielectric Lens of Refractive Index 1.57 With and Without a Wire Grid Matching Wall.

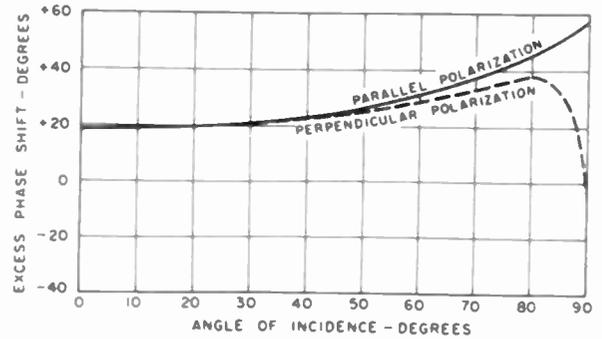


Fig. 7
Single Boundary Excess Phase Shift for a Lens Matched by a Quarter Wave Sheet at Normal Incidence ($n = 1.57$).

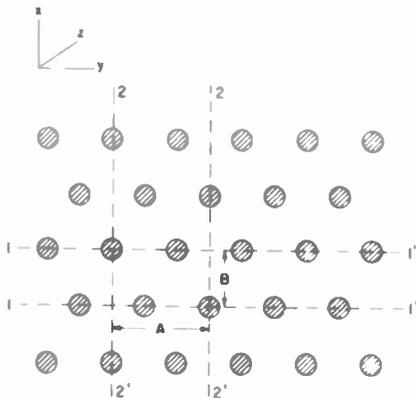


Fig. 5
Capacitive Wall Composed of Thin Metallic Discs.

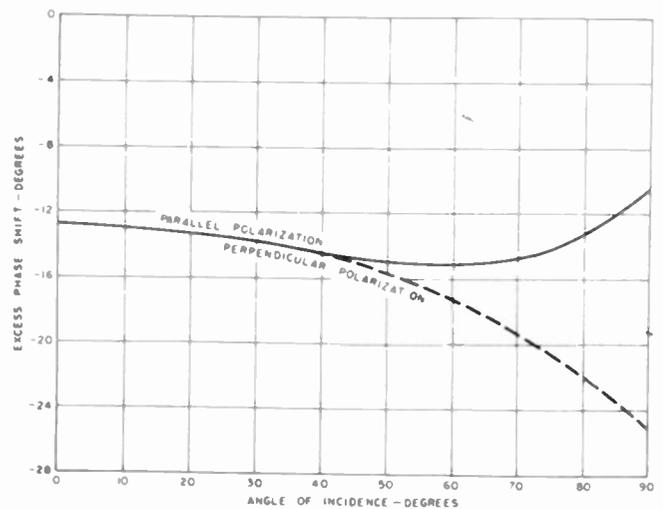


Fig. 8
Single Boundary Excess Phase Shift for a Lens Matched by a Wire Grid at Normal Incidence ($n = 1.57, s = \lambda/1.42$).

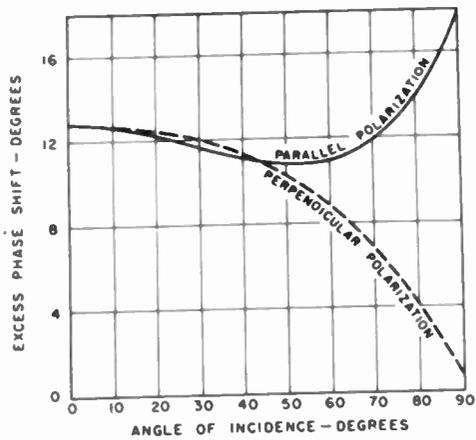


Fig. 9

Single Boundary Excess Phase Shift for a Lens Matched at Normal Incidence by a Circular Disc Wall ($n = 1.57$, $s = \lambda/1.28$).

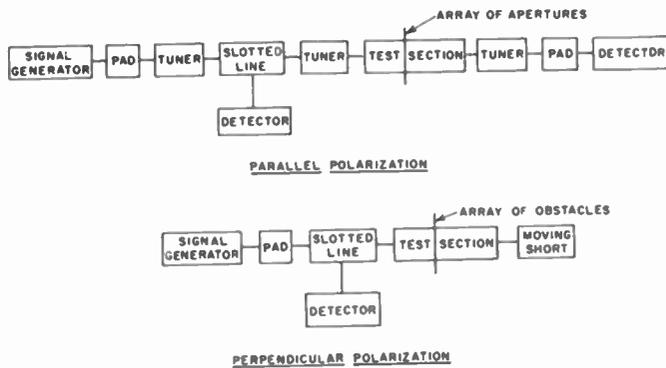


Fig. 12

Block Diagrams of Test Set-Ups Used in Measuring the Susceptance of the Hexagonal Array of Circular Discs.

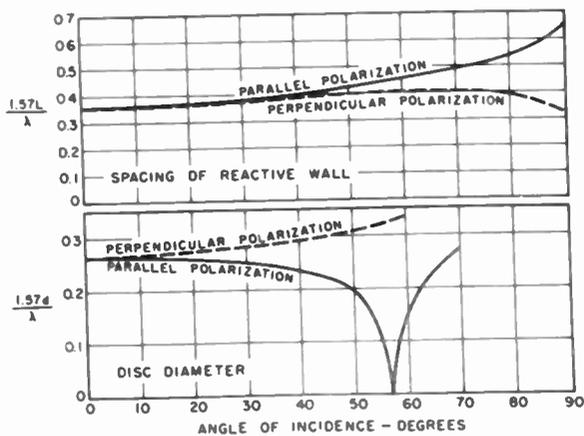


Fig. 10

Disc Diameters and Spacing of Disc Array from Lens Surface for a Match ($n = 1.57$)
 $b = \lambda/3.14$.

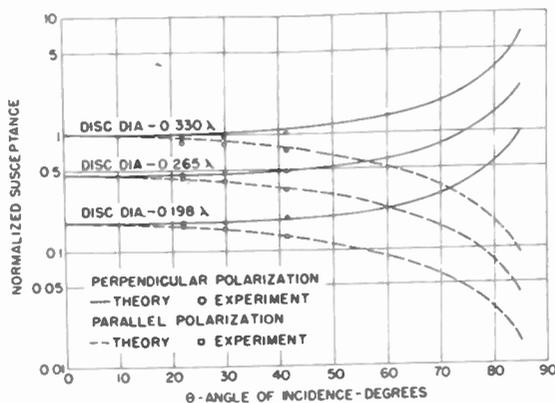


Fig. 13

Normalized Susceptance of a Hexagonal Array Thin Circular Discs on Half Wavelength Centers.

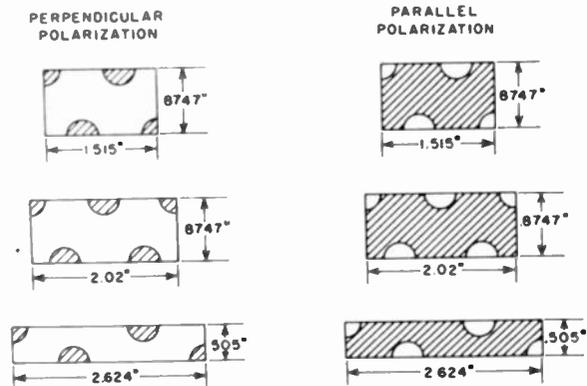


Fig. 11

Cross Section of Wave Guide Test Sections With Obstacles and Apertures in Place Used in Determining the Susceptance of Discs for Various Angles of Incidence and Polarizations.

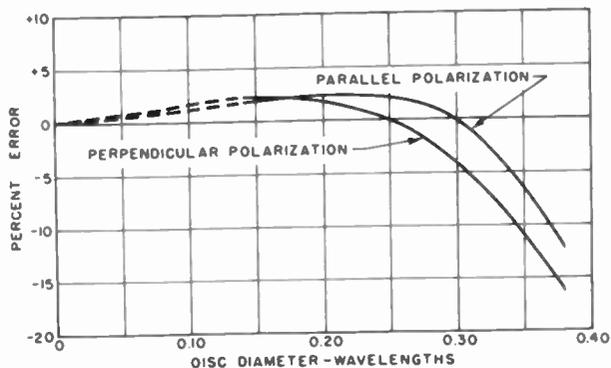


Fig. 14

Percent Error from Theory of the Measured Susceptance of an Array of Circular Discs.

DOUBLE PARABOLIC CYLINDER PENCIL BEAM ANTENNA*

Roy C. Spencer, F. Sheppard Holt, Helen M. Beauchemin and John L. Sampson†

Air Force Cambridge Research Center
Cambridge, Mass.

SUMMARY

Radiation from a point source placed on the focal line of a parabolic cylinder is reflected in succession from this cylinder and from a second parabolic cylinder crossed so that its focal line coincides with the directrix of the first cylinder. The two reflections result in a parallel beam. The theory is applicable to both microwaves and light. The advantages of shipment of the cylinders in the form of flat sheets and the possibilities of independent control of horizontal and vertical beamwidths and shapes are pointed out. Experimental models have been built and tested.

INTRODUCTION

In contrast to the usual microwave pencil beam antenna formed by allowing energy from a point source to reflect once from a paraboloid of revolution, the device to be discussed makes use of successive reflections from two parabolic cylinders with elements at right angles, each cylinder collimating the beam in one plane.

Functionally, the point source and first parabolic cylinder could be replaced by a parabolic pillbox¹ or any other equiphase line source.

The idea of using crossed cylindrical lenses has long been known in optics², but the application to pencil beams appears to be new. More recently, the use of crossed cylindrical reflectors has been introduced in connection with x-ray microscopes^{3,4} where the phenomenon of total reflection of x-rays at grazing incidence lends itself to the use of curved metal reflectors.

P. Kirkpatrick and A.V. Baez³ discuss the use of two crossed elliptic cylinder mirrors to form a real image of a point and the use of two parallel parabolic reflectors to form a real line image of a point. In both reference 3 and reference 4, the reflectors are approximated by sections of circular cylinders. In reference 4, the elliptic cylinder is approximated by a bent optical flat.

GEOMETRY

Figure 1 shows a perspective drawing of the two parabolic cylinders. The first cylindrical surface S_1 is given by the equation

$$y^2 = 4fx = 2lx, \quad (1)$$

where f is the focal length and $l = 2f$ is the semilatus rectum. The point F with coordinates $(f, 0, 0)$ lies on the focal line of S_1 . The line DD' is the image of the point F in S_1 ; that is, any ray emitted from F and incident on S_1 at P_1 is reflected as though it originated on DD' .

The formula for the second cylindrical surface S_2 is given by the equation

$$(x + f)^2 = 4f(z + f). \quad (2)$$

This surface has been so oriented that: (a) its focal line coincides with DD' , (b) it contains the point F , and (c) any ray emitted radially from DD' is reflected parallel to the positive z axis. Under these conditions, S_2 has the same focal length f as S_1 , and the point V with coordinates $(-f, 0, -f)$ lies on the vertex line of S_2 .

It is evident, therefore, that any ray FP_1P_2 emitted from a point source located at F will, after successive reflections from S_1 and S_2 , be directed parallel to the positive z axis. If either condition (b) or condition (c) on S_2 were relaxed, then other combinations of relative focal lengths and orientations would be possible.

CALCULATION OF THE COORDINATES OF COMMON INTERSECTION

The x , y , and z coordinates of the intersection of the two parabolic cylinders with $2f = l = 1$ were calculated using Eqs. (1) and (2). The curve of intersection may be written in the following parametric form:

$$\left. \begin{aligned} x &= t^2/2 \\ y &= t \\ z &= \frac{(t^2 + 1)^2}{8} - \frac{1}{2} \end{aligned} \right\} \quad (3)$$

Table 1 contains numerical values of x , y , and z as computed from Eq. (3). The range on t (that is, y) is from 0 to 2 in increments of 0.1.

*The geometric design of the antenna was described in a report by the same title by Roy C. Spencer, Air Force Cambridge Research Center, Report No. E5084, (April, 1952).

†Now in the U.S. Navy.

The curve of intersection of the two surfaces S_1 and S_2 is a three-dimensional space curve that we shall denote as C (see Fig. 2). If the surfaces S_1 and S_2 are separately developed into planes, the intersection curve in each case becomes a plane curve; however, these two plane curves are not identical. In the construction of the double cylindrical reflector, it is desirable to know the coordinates of these plane curves in order that they may be laid out on flat sheets before the sheets are bent into the parabolic cylindrical shape.

Curves of $z = \text{constant}$ in the surface S_1 are all identical parabolas and s_1 is defined to be the arc length along any of these parabolas as measured from the z axis (see Fig. 2). Similarly, the curves of $y = \text{constant}$ in the surface S_2 are all identical parabolas and s_2 is defined to be the arc length along any of these parabolas as measured from the vertex line HH' . In Table 1 we already have the x, y , and z coordinates of the intersection curve C. In order to obtain the s_1, z coordinates of C as measured in the surface S_1 , it is necessary to determine s_1 as a function of either z or y , that is, the arc length of a parabola as a function of its coordinates. Similarly, in order to obtain the s_2, y coordinates of C as measured in the surface S_2 , it is necessary to determine s_2 as a function of either x or z . Tables of parabolic arc length vs coordinates were computed by the Center of Analysis, Massachusetts Institute of Technology.⁵ The last two columns of Table 1 were obtained from these parabolic arc length tables.

The straight-line elements of the cylinder S_1 along which the z coordinate is measured, together with the parabolas along which s_1 is measured, form an orthogonal curvilinear coordinate system on the surface S_1 . When the cylinder is developed into a plane, this coordinate system becomes a rectangular system in the plane. The s_1, z coordinates of the intersection curve C are given in Table 1 and a plot of this curve on the developed surface S_1 is shown in Fig. 3(a). Similarly, the cylinder S_2 contains an s_2, y orthogonal curvilinear coordinate system that becomes a rectangular system when S_2 is developed into a plane. The s_2, y coordinates of the curve of intersection are given in Table 1 and a plot of this curve on the developed surface S_2 is shown in Fig. 3(b). Note that certain points along the intersection curves, as shown in Figs. 3(a) and 3(b), are numbered according to the first column of Table 1. When the developed surfaces S_1 and S_2 are returned to their parabolic cylindrical form and fitted together, identically numbered points will coincide.

Assume that the feed is polarized in the direction

$$\hat{u} = \hat{j} \cos \alpha + \hat{k} \sin \alpha, \quad (4)$$

where \hat{i} , \hat{j} , and \hat{k} are unit vectors in the x , y , and z directions respectively (see Fig. 4), and consider the ray that leaves the feed in the direction determined by θ and ϕ . Assuming that the radiation flows along ray paths, then the satisfaction of the boundary condition that the tangential component of the electric field vanish on each reflecting surface results in the following expression for the direction of polarization of the electric field on the aperture plane of the double parabolic cylinder antenna:

$$\bar{E} = \frac{1}{\sin \gamma} \left[\hat{i} (\cos \phi \sin \alpha - \sin \phi \sin \theta \cos \alpha) + \hat{j} (\cos \alpha \cos \theta) \right]. \quad (5)$$

Since θ and ϕ are independent variables it is evident that the x -component of \bar{E} can never be identically zero for any choice of α . The y -component is, however, identically zero when $\alpha = \pi/2$ --that is, when the original polarization is parallel to the elements of the first reflecting cylinder--and under these conditions the electric field on the aperture plane has only an x -component.

ILLUMINATION OVER THE APERTURE

An exact ray-tracing analysis of the antenna leads to a geometric optics approximation of the aperture illumination. Assuming an isotropic radiating source at the feed position, the illumination over the aperture is found to be proportional to a function separable in the normalized coordinates x/f and y/f as follows:

$$P = kF(x/f)G(y/f), \quad (6)$$

$$F(x/f) = 24.63 \frac{1 + x/f}{[4 + (1 + x/f)^2]^2}, \quad (7)$$

$$G(y/f) = \frac{4}{4 + (y/f)^2}. \quad (8)$$

Plots of these normalized functions are shown in Figs. 5(a) and (b).

EXPERIMENTAL MODEL DESIGN

An experimental model of the double parabolic cylinder antenna with aperture 1 ft by 1 ft,

f = 6 in. was designed and built. Full use was made of reference 5 in calculating the shapes of the flat sheets that were later bent to form the parabolic cylinder surfaces. In Fig. 6(a) the reflector surfaces and the component parts of the supporting framework are shown laid out prior to assembly. The assembled antenna is shown in Fig. 6(b). The mechanical design in this particular model features compact packaging in the disassembled state as well as simple rapid assembly and take-down procedures through use of dowels and slide snap fasteners.

The amplitude-tapering effects of the system as shown in Figs. 5(a) and 5(b) were taken into account in the design of a K-band ($\lambda = 1.25$ cm) horn feed that theoretically produced at least a 10 db illumination taper over the final aperture in both the x and y directions. It is evident from Fig. 5(a) that if the primary pattern of the feed were symmetric in ϕ (see Fig. 4), then the final aperture illumination would be asymmetrical in the x direction. This asymmetry in illumination introduced purely by the geometry of the system was counteracted to some extent by orienting the feed horn so that its primary pattern was asymmetrical in ϕ .

Experimental patterns taken with the K-band model are shown in Figs. 7(a) and 7(b). Asymmetry of the side lobe levels in the E-plane pattern indicate the presence of asymmetrical edge effects.

SHAPED BEAM DESIGN (COSECANT SQUARED)

After reflection from the first parabolic cylinder S_1 all radiation striking the second parabolic cylinder S_2 appears to originate from a line source located along the line DD' (see Fig. 1). Hence, it is necessary to redesign only the cylinder S_2 in order to produce a shaped fan beam.

Using the method suggested by L.J. Chu,⁶ the design of the second cylindrical surface S_2 of the original experiment model was modified to produce a $\text{csc}^2\theta$ fan beam over the range 2° to 20° . Experimental elevation and azimuth patterns are shown in Fig. 8. (The azimuth patterns should be shifted horizontally until their peaks match the ordinate of the elevation patterns.)

SUMMARY OF ADVANTAGES

The double parabolic cylinder antenna has several advantages over the conventional paraboloid. Each of its surfaces, being cylindrical, can be stamped from flat sheets. The system itself can be designed so that it can be easily dismantled and stored or transported in a compact flat package. There is an absence of cross polarization. Since the first reflector effectively focuses in one plane and the second focuses in the orthogonal plane, a certain degree of independent control can be obtained over the beam shapes in the two planes by modifying the appropriate reflecting surfaces. In particular, the second surface S_2 can be modified to obtain $\text{csc}^2\theta$ patterns.

ACKNOWLEDGMENTS

The authors wish to acknowledge the aid of Norris Hansen and the Antenna Laboratory Shop in the design and fabrication of the experimental models, and the aid of the Ipswich Antenna Pattern Test Unit in obtaining the patterns.

REFERENCES

1. S. Silver, *Microwave Antenna Theory and Design*, Mass. Inst. Tech. Rad. Lab. Series, vol. 12, McGraw-Hill, New York (1949); pp. 151, 494.
2. J.P.C. Southall, *Mirrors, Prisms and Lenses*, 3rd ed., MacMillan, New York (1933); Chapter IX, "Astigmatic Lenses."
3. P. Kirkpatrick and A.V. Baez, "Formation of Optical Images by X-rays," *J. Opt. Soc. Amer.* **38**, 766-774 (Sep. 1948).
4. W. Ehrenberg, "The Production of Converging Beams by Total Reflection," *J. Opt. Soc. Amer.* **39**, 741-746 (Sep. 1949).
5. R.C. Spencer and G.E. Reynolds, "Tables of Normalized Parabolic Coordinates and Arc Lengths," Report No. E4083, AF Cambridge Research Center, Cambridge, Mass. (July, 1951).
6. R.C. Spencer, "Synthesis of Microwave Diffraction Patterns with Application to $\text{Csc}^2\theta$ Patterns," M.I.T. Rad. Lab. Report No. 272, (54-24), (23 June, 1943); pp. 5-7.

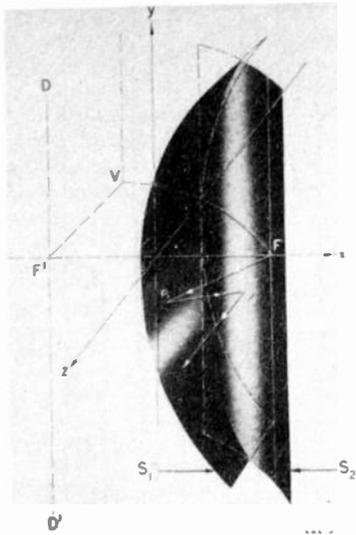
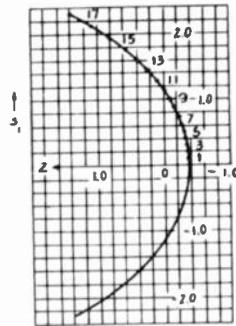
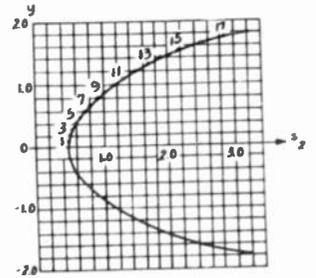


Fig. 1
Double reflection from orthogonal
parabolic cylinder reflectors.



(a)



(b)

Fig. 3
Curves of intersection developed
on surfaces S_1 and S_2 .

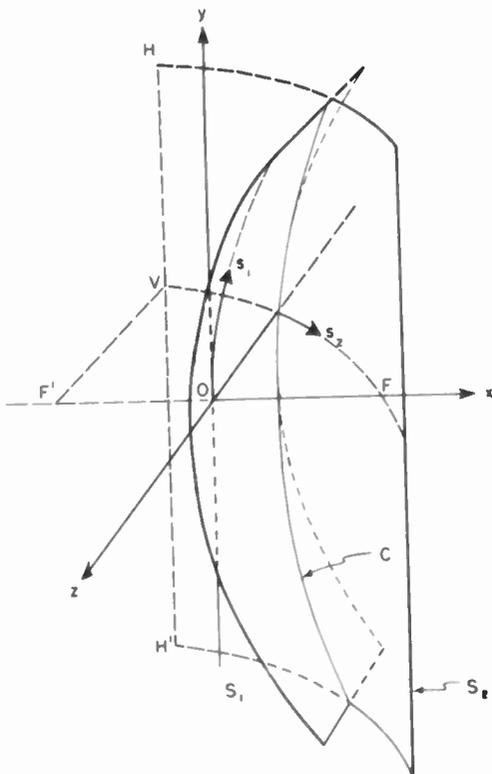
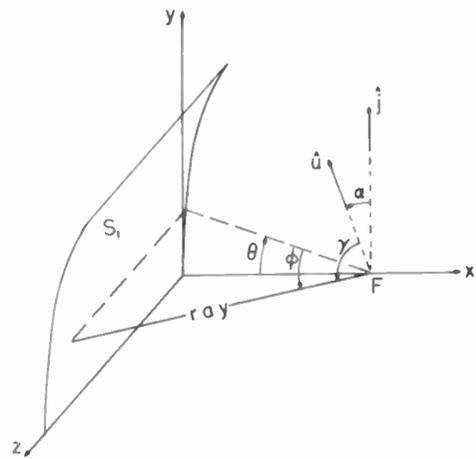


Fig. 2
Orthogonal curvilinear coordinate
systems on surfaces S_1 and S_2 .



$$E = \frac{1}{\sin \gamma} [\hat{i}(\cos \phi \sin \alpha - \sin \phi \sin \theta \cos \alpha) + \hat{j}(\cos \alpha \cos \theta)]$$

Fig. 4
Feed orientation and polarization.

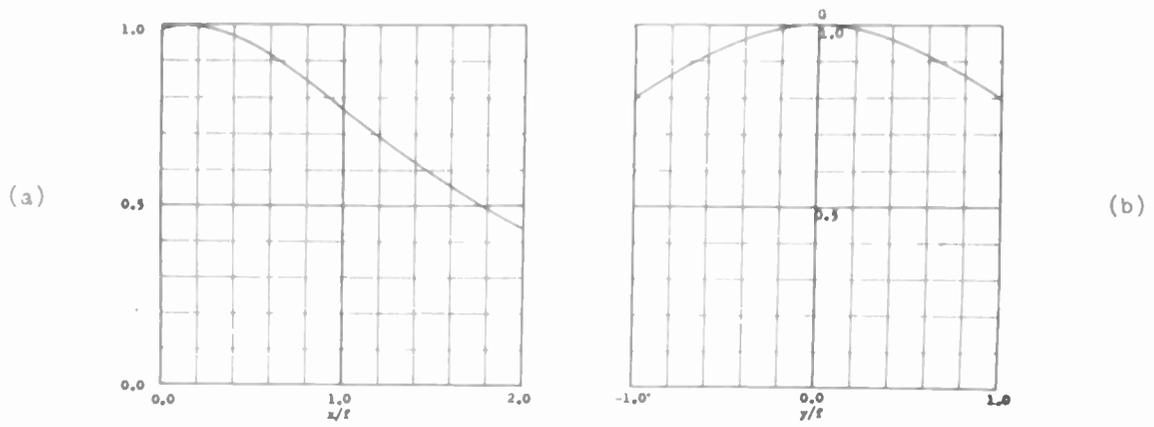


Fig. 5 - Aperture illumination. (a) x-direction; (b) y-direction.

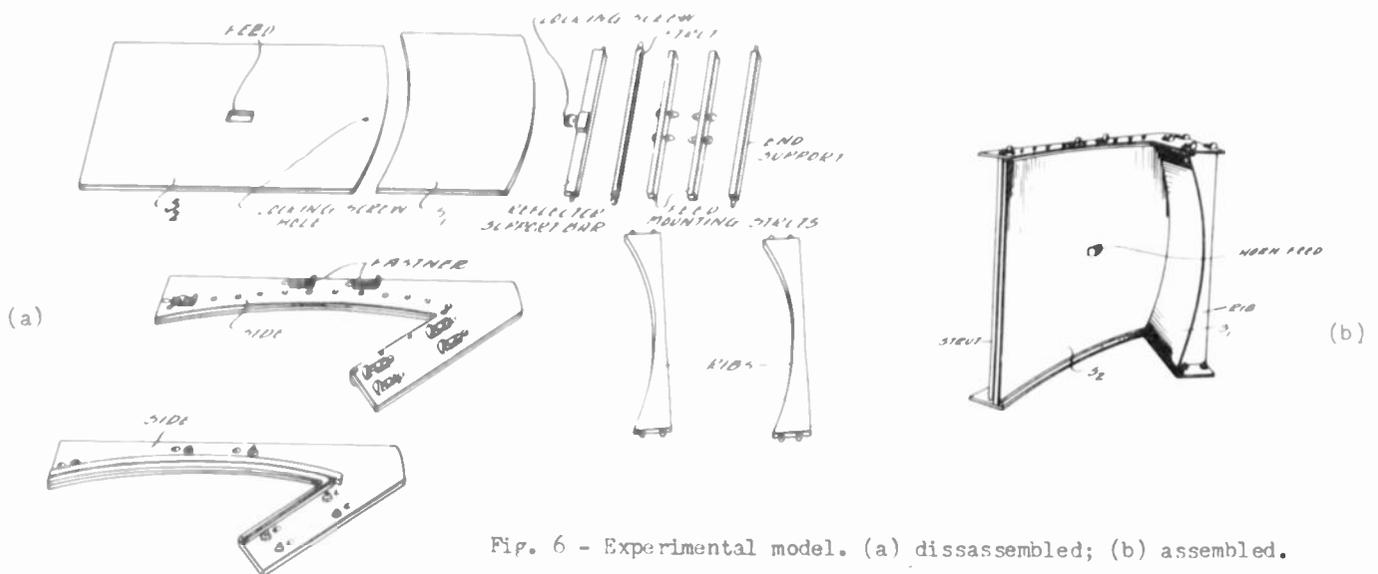


Fig. 6 - Experimental model. (a) dissassembled; (b) assembled.

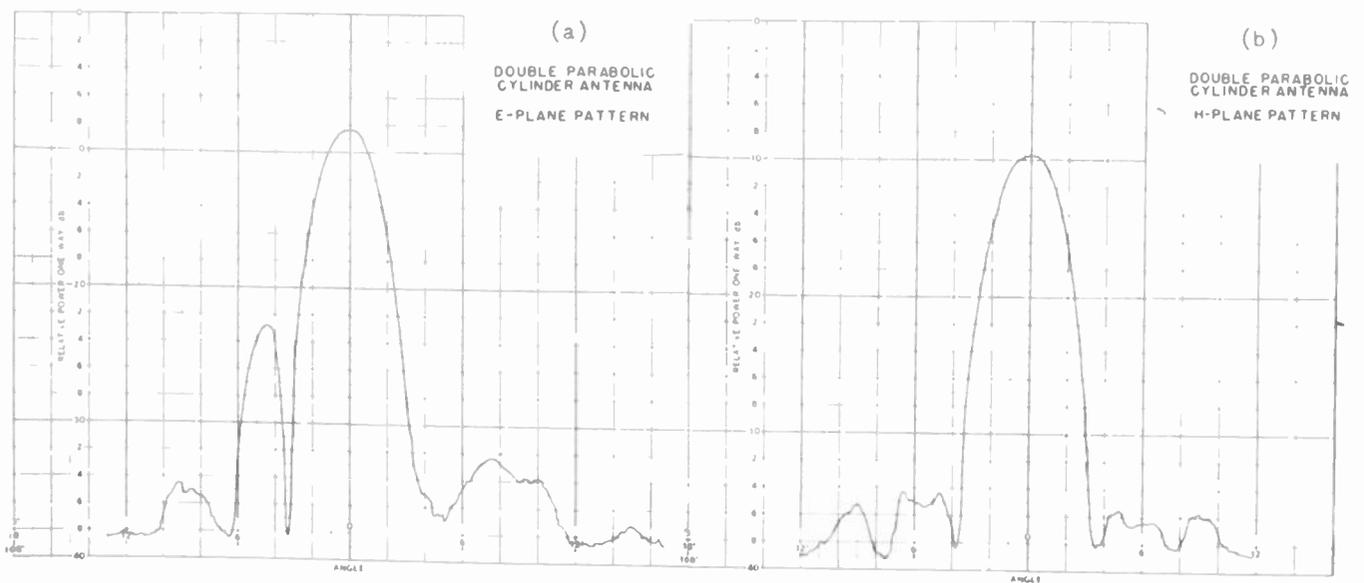


Fig. 7 - Experimental patterns. (a) E-plane; (b) H-plane.

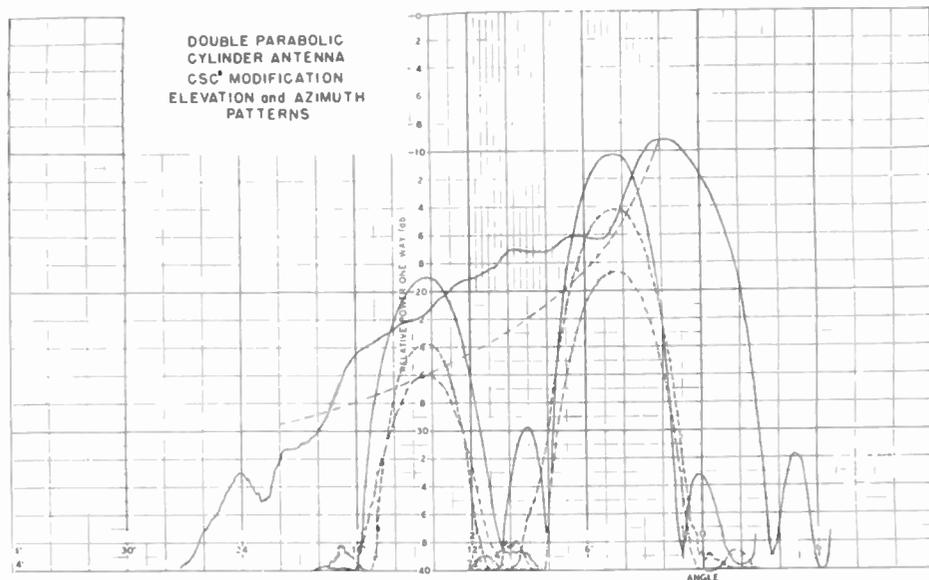


Fig. 8 - Experimental patterns of csc^2 model.

Point No.	t	x	y	z	s_1	s_2
0	0.0	0.000	0.0	-0.3750	0.0000	0.5201
1	.1	.005	.1	-.3725	.1002	.5257
2	.2	.020	.2	-.3648	.2013	.5426
3	.3	.045	.3	-.3515	.3044	.5709
4	.4	.080	.4	-.3318	.4104	.6110
5	0.5	0.125	0.5	-0.3047	0.5201	0.6636
6	.6	.180	.6	-.2688	.6343	.7293
7	.7	.245	.7	-.2225	.7536	.8091
8	.8	.320	.8	-.1638	.8786	.9043
9	.9	.405	.9	-.0905	1.0098	1.0166
10	1.0	0.500	1.0	0.0000	1.1478	1.1478
11	1.1	.605	1.1	.1105	1.2928	1.3003
12	1.2	.720	1.2	.2442	1.4452	1.4766
13	1.3	.845	1.3	.4045	1.6053	1.6799
14	1.4	.980	1.4	.5952	1.7733	1.9136
15	1.5	1.125	1.5	0.8203	1.9495	2.1814
16	1.6	1.280	1.6	1.0842	2.1339	2.4874
17	1.7	1.445	1.7	1.3915	2.3269	2.8363
18	1.8	1.620	1.8	1.7472	2.5284	3.2327
19	1.9	1.805	1.9	2.1565	2.7387	3.6819
20	2.0	2.000	2.0	2.6250	2.9579	4.1893

Table I
Coordinates of the curve of intersection
of the two parabolic cylinders.

DIFFUSE RADIATION IN PENCIL BEAM ANTENNAS

David Carter
Consolidated Vultee Aircraft Corporation
San Diego, California

Summary

This paper is concerned with theoretical estimates of wide angle radiation in pencil beam antennas. In reflector type antennas this energy consists of direct radiation from the feed and scattered energy from the reflector.

Approximate methods for the evaluation of these contributions are discussed together with their simplifying assumptions.

To get some numerical indication, calculations were made for paraboloidal reflectors of different f/D ratios and a class of primary patterns which provide an approximate representation of a great many common feeds. The results are presented in graphical form.

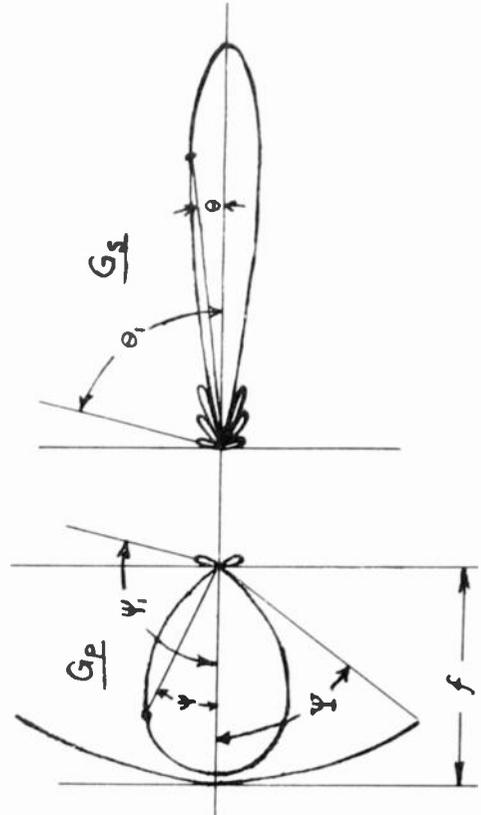


Fig. 1

In many applications of pencil beam antennas it is important to know the amount of energy directed at large angles away from the axis of the main beam. While this energy is usually very small, some operations may require all wide angle radiation to be down 50 db or more. Physical proximity of antennas may impose an extremely small upper limit on the absolute power in certain directions in order to reduce cross-talk or prevent crystal burn-out. This may in turn, for certain geometries, add a very stringent specification of side lobe discrimination at large angles to the usual gain and small angle side lobe specifications. Then feed and reflector design will require a theoretical estimate of the antenna pattern at large angles off axis.

In reflector type antennas, the energy radiated at large angles off the beam axis consists of direct radiation from the feed and scattered energy from the reflector. In any direction where these two contributions are of the same order of magnitude a knowledge of their relative phases is necessary to obtain the power from the resultant of the two field intensities. In the usual case, however, the contribution from one far exceeds that of the other, and the smaller one can then be neglected.

It is a simple matter to approximate the contribution of the direct radiation from the feed to the secondary pattern. From the definitions of primary feed gain function, $G_p(\psi, f)$, and secondary pattern gain function, $G_s(\theta, \phi)$, the power per unit solid angle in any direction, for a system having rotational symmetry, is given by

$$P_p(\psi, \xi) = P_p(\psi) = G_p(\psi) \frac{P_T}{4\pi}$$

$$P_s(\theta, \phi) = P_s(\theta) = G_s(\theta) \frac{P_T}{4\pi}$$

where P_T is the total radiated power and the angles are defined in Figure 1. Then

$$P_s(0) \approx \frac{G_s(0)}{G_p(0)} P_p(0)$$

and for the case of negligible diffracted energy at θ_1 ,

$$P_s(\theta_1) \approx P_p(\psi_1)$$

or

$$\frac{P_s(\theta_1)}{P_s(0)} \approx \frac{G_p(0)}{G_s(0)} \frac{P_p(\psi_1)}{P_p(0)}$$

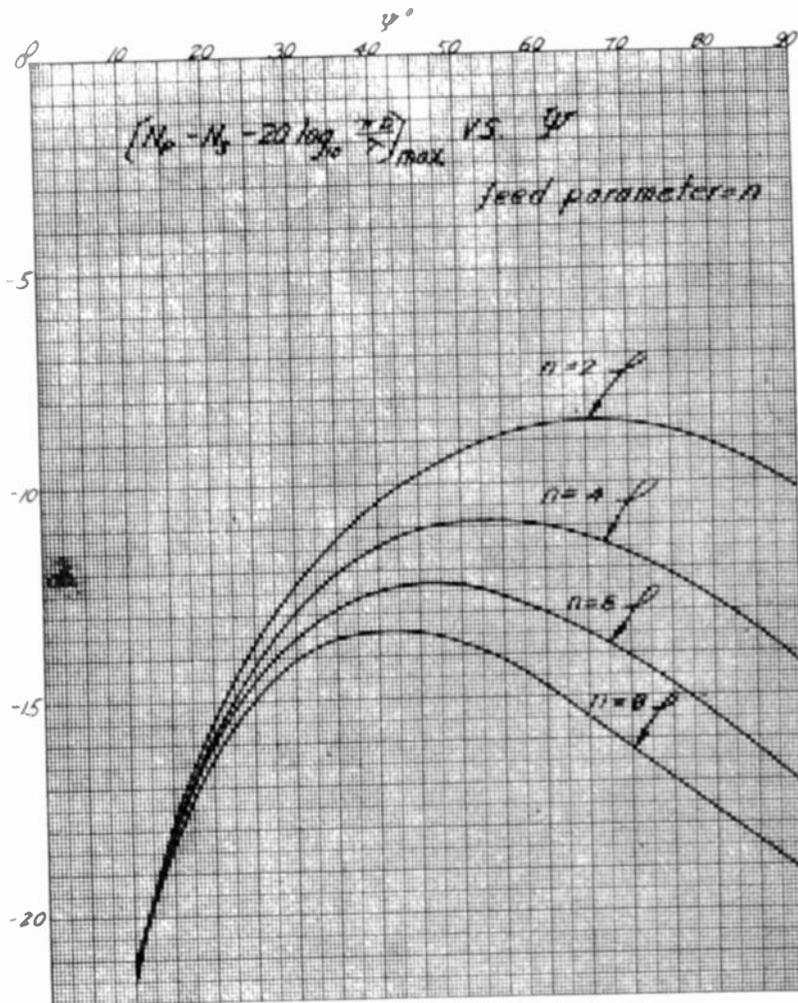


Fig. 2

Writing this in a slightly more practical form, - if N_s is the maximum allowable db level of a side lobe in the secondary pattern, produced solely by direct radiation from the primary feed as shown in Figure 1., then the db level of the corresponding radiation in the primary feed pattern, $N_p \equiv 10 \log_{10} \left[\frac{P_p(\psi_1)}{P_p(0)} \right]$,

must be given by

$$N_p \leq N_s + 10 \log_{10} \frac{G_s(0)}{G_p(0)}$$

This indicates, what was intuitively apparent, that the larger the secondary gain and the smaller the primary feed gain, the less stringent will be the requirement on side lobe discrimination in the primary feed in order to meet a specified maximum wide angle side lobe level in the secondary pattern.

To get some numerical indication, this relationship was evaluated for paraboloidal reflectors of different f/D ratios and a class of primary patterns which provide an approximate representation of a great many of the common feeds.

Assuming

$$G_p(\psi) = \begin{cases} G_p(0) \cos^n \psi & 0 \leq \psi \leq \frac{\pi}{2} \\ \epsilon \delta(\psi - \psi_1) & \text{for } \frac{\pi}{2} \leq \psi \leq \pi \end{cases}$$

where $\epsilon \ll 2$ and δ is the Dirac delta function, it can be shown¹ that

$$G_s(0) = \left(\frac{\pi D}{\lambda} \right)^2 G_p(0) \left[\cot \frac{\Psi}{2} \int_0^{\Psi} \cos^{n/2} \psi \tan \frac{\psi}{2} d\psi \right]^2$$

where D is the aperture diameter, λ the operating wavelength and Ψ is the angular aperture. For such antennas then

$$N_p - N_s - 20 \log_{10} \frac{\pi D}{\lambda} \leq 10 \log_{10} \left[\cot \frac{\Psi}{2} \int_0^{\Psi} \cos^{n/2} \psi \tan \frac{\psi}{2} d\psi \right]^2$$

This relationship has been plotted in Figure 2. for the different feed patterns characterized by the values of the parameter $n = 2, 4, 6, \text{ and } 8$.

It controls the choice of a reflector for a given primary feed and side lobe spec when the side lobe in the secondary pattern is due to direct radiation from the feed alone, or speci-

fies the maximum allowable level of this radiation for a fixed reflector.

The other contributing source of radiation in the secondary pattern is the energy scattered by the reflector. Usually this is calculated as a diffraction pattern from the field distribution over the aperture by means of the scalar integral $U(P) \approx \text{const} \int_{\mathcal{A}} u e^{jk\vec{p}\cdot\vec{R}} dS$, where \vec{p} is the vector from the origin to the surface element dS , \vec{R} , is the unit vector pointing from the origin to the point P at which the field is being calculated, $k = 2\pi/\lambda$, $j = \sqrt{-1}$ and u denotes the principal polarization component of the aperture

field.² Unfortunately this aperture distribution method involves approximations which, among other things, limits the applicability of the results to angles near the axis of the main lobe.

Another approach which removes this restriction to small angles and reduces to the aperture distribution formula for small angles off the main beam axis is the method of calculating the scattered pattern from the current distribution over the reflector. This formulation differs from the former only by replacing the scalar integration over the aperture by a vector integration of the equivalent surface current density over the entire surface of the reflector. However, the calculation of the far field leads to significant differences at large angles off the main lobe axis.

Using the latter approach then to obtain an estimate of the scattered field, it can be shown that the far field is given by

$$E_{\theta\phi} \approx \text{const} \frac{e^{jkR}}{R} i_{\theta\phi} \cdot I$$

$$\text{and } H_p \approx \left(\frac{\epsilon}{\mu}\right)^{1/2} R_1 \times E_p$$

$$\text{where } I = \int_S N \times H e^{jk\vec{p}\cdot\vec{R}} dS$$

i_{θ} and i_{ϕ} are unit vectors at the field point P in the polar and azimuthal directions respectively, N is the unit normal vector at dS , R is the distance from the origin to the field point, ϵ and μ are the electric and magnetic inductive capacities, respectively, and S denotes the reflector surface.

This may be obtained from an integration of Maxwell's equations, expressing the field in terms of the sources.³ The lowest order terms in R for the case of a source free region bounded by one infinitely conducting surface then leads to the far field approximation above.

Many simplifications and assumptions are made in order to obtain workable formulas. To anticipate discrepancies between theory and experiment, it seems pertinent to review some of the assumptions and approximations that are made

in the application of the current distribution formulas.

The scattered field in the far zone may be calculated once the H distribution over the entire surface of the reflector is known. To obtain an approximation in terms of the primary feed pattern, it is assumed that the primary radiation is geometrically reflected and that multiple scattering between the feed and reflector may be neglected because of their geometry and large separation. These assumptions imply that the reflector is in the far zone of the feed, and that the curvatures of both the reflector and primary feed wave front are so small that the transformation from incident to reflected wave front may be calculated, at each point of the reflector, as though the tangent plane at that point reflects a uniform plane wave having the same intensity and polarization as the incident wave front at that point.

The tangential component of the total H at any point on the reflector is then given in terms of the reflected primary field at the same point by

$$N \times H = 2N \times H_r = 2 \left(\frac{\epsilon}{\mu}\right)^{1/2} N \times (S_1 \times E_r)$$

where S_1 is a unit vector in the direction of Poyntings vector on the reflected wave front and the subscript r refers to reflected quantities. The far field relations for a point source feed gives E_r (which, from the geometrical optics approximation, differs from the incident E at the reflector only in polarization), in terms of the primary gain function as

$$E_r = \text{const} \frac{e^{-jk\rho}}{\rho} G_p^{1/2}(\psi, \xi) e_r(\psi, \xi)$$

where e_r is a unit vector defining the polarization in the reflected wave front.

Substitution of these approximations then allows the vector integral I, which determines the scattered field, to be expressed as an integral of the primary feed pattern over the surface of the reflector. Thus

$$I = \text{const} \int_S \frac{G_p^{1/2}(\psi, \xi)}{\rho} e^{-jk\rho} \left[N \cdot e_r S_1 - (N \cdot S_1) e_r \right] dS$$

where \vec{p}_1 is the unit vector pointing from the origin to the element dS .

The field on the shadow side of the reflector is assumed to vanish, which is a consequence of using the approximations of the geometrical-optics method of calculating the field over the reflector. Normally there will be small currents at the edge and on the shadow side of the reflector which will contribute to the diffraction field, especially in the shadow region. In some cases this might be the worst assumption. However, if the edge of the reflector is very near a deep null of the primary feed pattern, these currents will be negligibly small, and this will be assumed in what follows.

To obtain numerical estimates of this second contribution to wide angle radiation, the scattered pattern was calculated under the same conditions as in the previous calculations of the direct radiation from the feed, viz., paraboloidal reflectors of different f/D ratios and the previously defined class of primary patterns, which provide an approximate representation of a great many common feeds. (Generally G_p is a function of both ψ and ξ . However, for primary feed patterns having approximately the same half power beam widths in the principal E and H planes, it is assumed that the feed pattern can be represented by a pattern, having rotational symmetry about the paraboloid axis, which is the average of the patterns in the two principal planes.)

To calculate the diffraction pattern, the integral of I is expanded in terms of the coordinates shown in Figure 3.

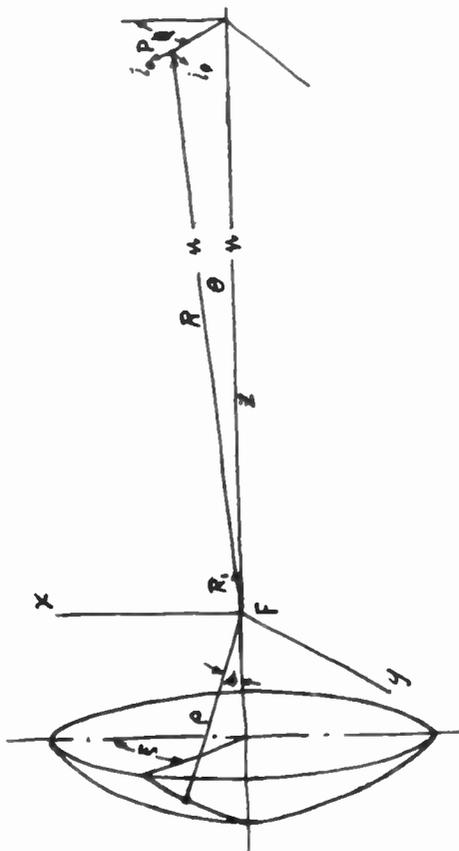


Fig. 3

Thus, for a paraboloidal reflector ($\rho = f \sec^2 \psi$) it can be shown that

$$I = \text{const.} \int_0^{2\pi} \int_0^{\psi} G_p(\psi) e^{-jkf \sec^2 \frac{\psi}{2}} [1 + \cos \theta \cos \psi - \sin \theta \sin \psi \cos(\xi - \theta)] [(\mathbf{N} \cdot \mathbf{e}_r)_s - (\mathbf{N} \cdot \mathbf{s}) \cdot \mathbf{e}_r] \frac{\tan \frac{\psi}{2}}{\cos \frac{\psi}{2}} d\psi d\xi.$$

There are several other simplifying assumptions which limit the accuracy of the results. One of these is that $S_1 = i_z$ over the entire surface of the reflector. This implies that the far field of the primary feed is more than just quasi-point source. It follows from the previous assumptions applied to a paraboloid only if the feed is a true point source with a single center of phase. Furthermore, since the geometry of the paraboloid makes all reflected rays parallel to the axis, \mathbf{e}_r can have no axial component and it is set equal to i_x at each point on the reflector. This assumes that the cross polarization component of \mathbf{e}_r may be neglected in calculating the principle polarization diffraction pattern because the cross polarization energy is a very small fraction of the total energy in most feeds of interest.

With these simplifications, the integral defining the scattered pattern becomes

$$I = \text{const.} \int_0^{2\pi} \int_0^{\psi} G_p(\psi) e^{-jkf \sec^2 \frac{\psi}{2}} [1 + \cos \theta \cos \psi - \sin \theta \sin \psi \cos(\xi - \theta)] [i_x + i_z \tan \frac{\psi}{2} \cos \xi] \tan \frac{\psi}{2} d\psi d\xi,$$

and using the relations

$$\int_{\alpha}^{2\pi + \alpha} e^{ja \cos \xi} d\xi = 2\pi J_0(a),$$

$$\int_{\alpha}^{2\pi + \alpha} \cos \xi e^{ja \cos \xi} d\xi = 2\pi J_1(a),$$

$$\int_{\alpha}^{2\pi + \alpha} \sin \xi e^{ja \cos \xi} d\xi = 0$$

to perform the integration over ξ , it follows that

$$I = \text{const.} \int_0^{2\pi} \int_0^{\psi} [i_x J_0(2kf \sin \theta \tan \frac{\psi}{2}) + i_z \cos \theta \tan \frac{\psi}{2} J_1(2kf \sin \theta \tan \frac{\psi}{2})] G_p(\psi) \tan \frac{\psi}{2} e^{-jkf \sec^2 \frac{\psi}{2}} (1 + \cos \theta \cos \psi) d\psi$$

It can be shown that in the region near the axis ($\theta \rightarrow 0$), I reduces approximately to the familiar integral of the aperture distribution method of calculating the diffraction pattern. However, for wide angles, the contribution of the axial component of I , which has no counterpart in the aperture distribution formulation, cannot be neglected, and the two formulations give significantly different answers.

For feeds having approximately rotationally symmetric patterns then, evaluation of this last form of I leads to the scattered field of the paraboloid in any direction (θ, ϕ). Restricting attention to the principal E and H plane patterns and changing integration variable to

$x = \tan \frac{\psi}{2}$, it follows that, in the principle

E plane ($\phi = 0$)

$$I = \text{const.} e^{-j2kfc \cos^2 \frac{\theta}{2}} \int_0^x G_p(x) \frac{x}{1+x^2} e^{-j2kfx^2 \sin^2 \frac{\theta}{2}} \left[i_1 J_0(2kfx \sin \theta) + i_2 j x J_1(2kfx \sin \theta) \right] dx,$$

and in the principle H plane ($\phi = \frac{\pi}{2}$)

$$I = \text{const.} i_x e^{-j2kfc \cos^2 \frac{\theta}{2}} \int_0^x G_p(x) \frac{x}{1+x^2} e^{-j2kfx^2 \sin^2 \frac{\theta}{2}} J_0(2kfx \sin \theta) dx.$$

Since $i_\theta \cdot i_z = 0$ and $i_\theta \cdot i_x = -\sin \phi$, it can be seen that the z field in the E plane has θ polarization, and, since $i_\theta \cdot i_x = \cos \theta \cos \phi$, the field in the H plane has ϕ polarization.

For the actual calculation, the previously defined feed patterns were used, viz.,

$$G_p(\psi) = G(0) \cos^n \psi = G(0) \left(\frac{1-x^2}{1+x^2} \right)^n,$$

and the working formulae for the scattered field amplitude became

$$E(\theta) = A e^{-j2kfc \cos^2 \frac{\theta}{2}} \begin{cases} i_\theta \left[I_{z2} \sin \theta - I_{x1} \cos \theta + j(I_{x1} \sin \theta + I_{x2} \cos \theta) \right] & \text{in the E plane} \\ \pm i_\phi (I_{x1} - j I_{x2}) & \text{in the H plane} \end{cases}$$

$$I_{x1} = \int_0^x \left(\frac{1-x^2}{1+x^2} \right)^n \frac{x}{1+x^2} \cos(2kfx^2 \sin^2 \frac{\theta}{2}) J_0(2kfx \sin \theta) dx,$$

$$I_{x2} = \int_0^x \left(\frac{1-x^2}{1+x^2} \right)^n \frac{x}{1+x^2} \sin(2kfx^2 \sin^2 \frac{\theta}{2}) J_0(2kfx \sin \theta) dx,$$

$$I_{z1} = \int_0^x \left(\frac{1-x^2}{1+x^2} \right)^n \frac{x^2}{1+x^2} \cos(2kfx^2 \sin^2 \frac{\theta}{2}) J_1(2kfx \sin \theta) dx,$$

$$I_{z2} = \int_0^x \left(\frac{1-x^2}{1+x^2} \right)^n \frac{x^2}{1+x^2} \sin(2kfx^2 \sin^2 \frac{\theta}{2}) J_1(2kfx \sin \theta) dx,$$

and A is a function of R alone.

To normalize the pattern, the maximum value of E,

$$E_{\max} = \left| A I_{x1}(0) \right| = \left| A \int_0^x \left(\frac{1-x^2}{1+x^2} \right)^n \frac{xdx}{1+x^2} \right|,$$

is easily evaluated for any value of n. Substituting in the power pattern equation

$$r = 10 \log_{10} \left| \frac{E}{E_{\max}} \right|^2$$

then gives the diffraction pattern. Thus, in the E plane,

$$r_E = 10 \log_{10} \frac{1}{I_{x1}^2(0)} \left[(I_{x1}^2 + I_{x2}^2) \cos^2 \theta \right.$$

$$\left. + (I_{z1}^2 + I_{z2}^2) \sin^2 \theta + (I_{z1} I_{x2} - I_{x1} I_{z2}) \sin 2\theta \right]$$

and in the H plane,

$$r_H = 10 \log_{10} \frac{1}{I_{x1}^2(0)} (I_{x1}^2 + I_{x2}^2)$$

These patterns have been numerically calculated on IBM computers for a number of values of each of the parameters N, f/D and D, corresponding to primary gain, aperture efficiency, and uniformly illuminated aperture gain, respectively. These have been plotted below (Figs. 4-14).

Firstly it should be said that these patterns are not complete since they were evaluated at discrete values of θ , and some lobe maxima and nulls were undoubtedly missed. However, in the regions of most interest the intervals were made small in order to get the general level.

It can be seen that for small θ , all the patterns conform to the most significant results of aperture theory:

- As primary gain increases (n increases), secondary gain decreases (broader main beams)
- As primary gain increases (n increases), side lobe level decreases.
- As D and f/D increase (altho f/D variation is not shown here) so does secondary gain.

A significant departure from aperture theory is that minor lobe intensity does vary with D. This is also true for the wide angle energy.

In the region of greatest interest for this investigation the level at 90° was taken as an indication of the level of the diffracted energy at wide angles, and here the most important conclusions are as follows:

- a. The sharpest variation by far was with primary gain, going in the same direction as the small angle side lobes; for example, there was a drop of approximately 5 db for unit increase in n in all the very large D cases calculated. Individual variations are shown in the curves below.
- b. The second significant factor was $\frac{D}{\lambda}$, the wide angle radiation level decreasing with increasing $\frac{D}{\lambda}$. Thus, for a tapered illumination corresponding to $n = 6$, the 90° level decreased approximately 2 db for each db increase in the gain of the uniformly illuminated dish. The variation was smaller for smaller primary gains.
- c. The last factor, f/D , produced very small variations. The largest calculated was (for large n and large D in the H plane) 3.7 db change in 85 db, as f/D went from .3 to .25. Variations were even smaller for smaller n .
- d. The H plane and E plane patterns differed very little.

It should be noted that the requirements on primary gain are in opposite directions for the two contributions to wide angle energy, - reduction of the level due to directly radiated feed energy calling for a low gain feed whereas a high gain feed is required to reduce the level of the scattered energy.

In conclusion, one might say in general, that design for low side lobe at any angle in the scattered pattern calls for large secondary and primary gain.

References

1. S. Silver, "Microwave Antenna Theory and Design", McGraw-Hill Book Co., 1949, Ch. 12.
2. S. Silver, loc. cit., pg. 167.
3. J. A. Stratton, "Electromagnetic Theory," McGraw-Hill Book Co., 1941, pg. 467.
S. Silver, op. cit., Ch. 3.

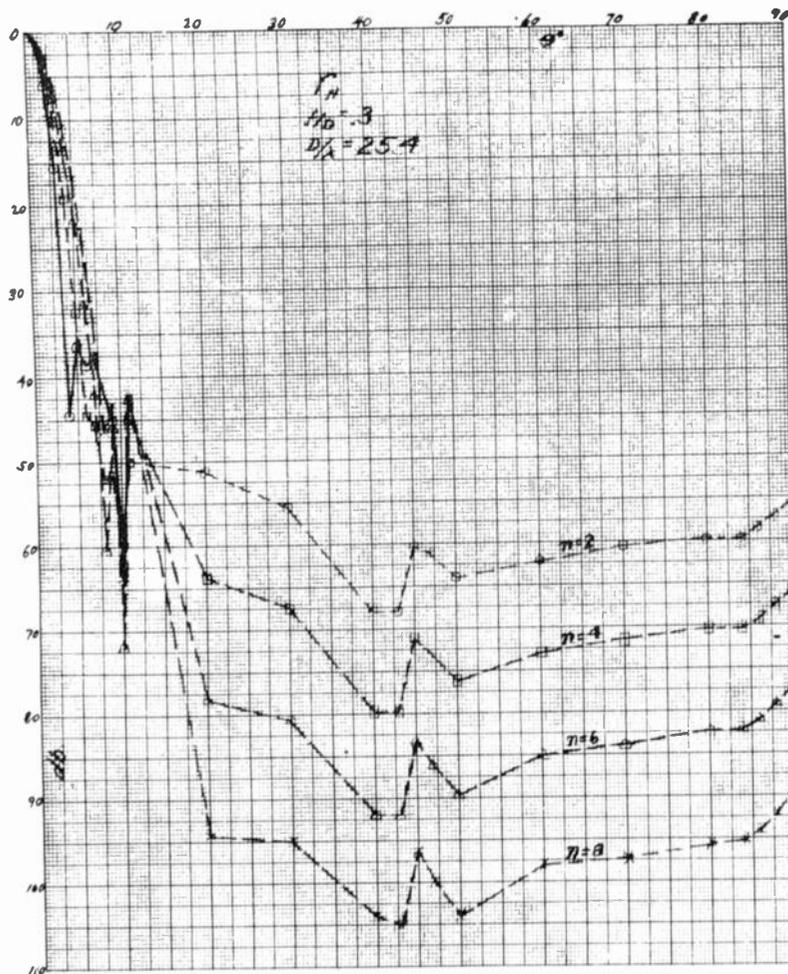


Fig. 4

Fig. 6

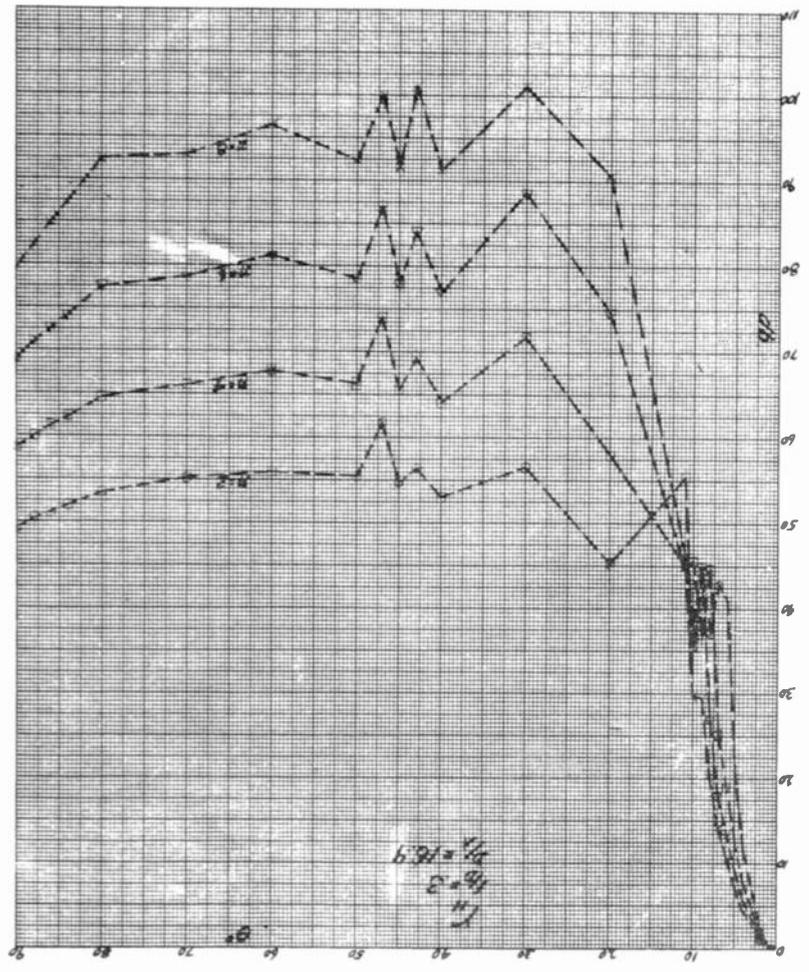
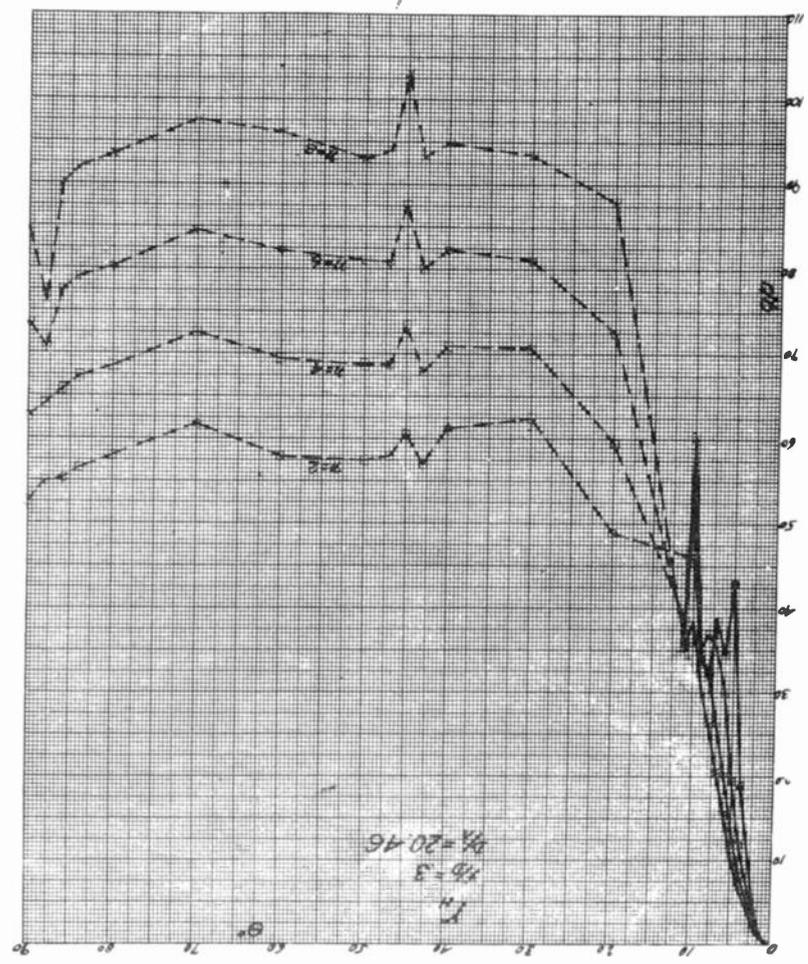


Fig. 5



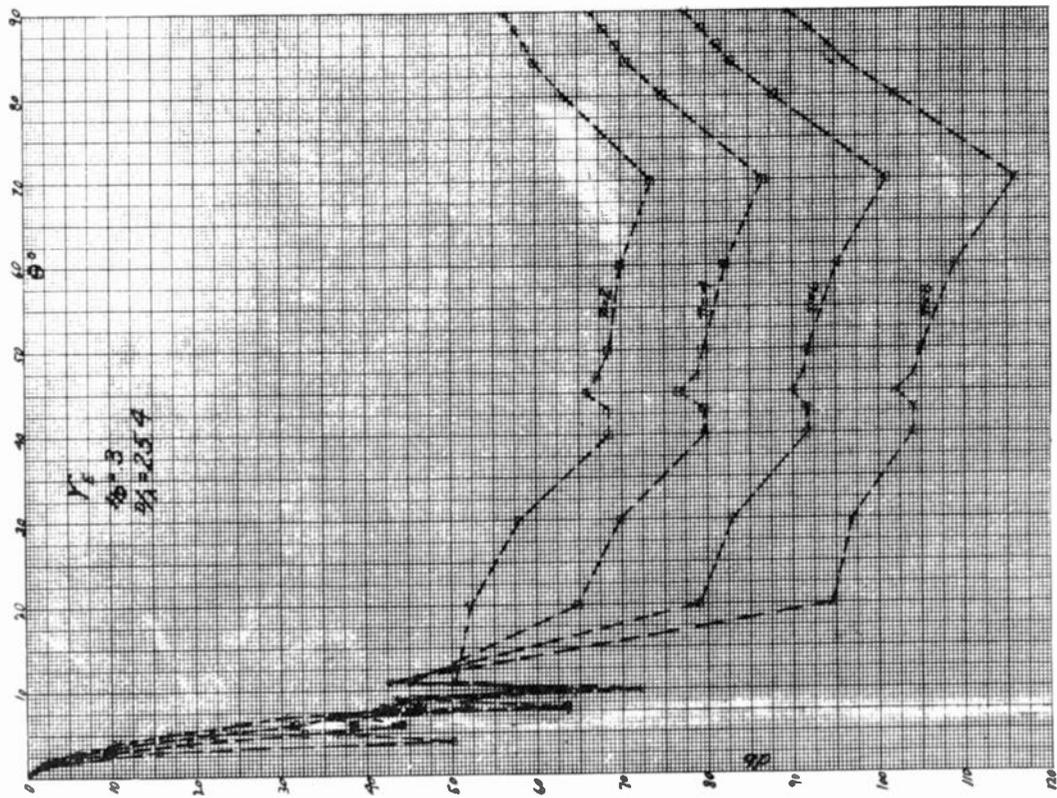


FIG. 8

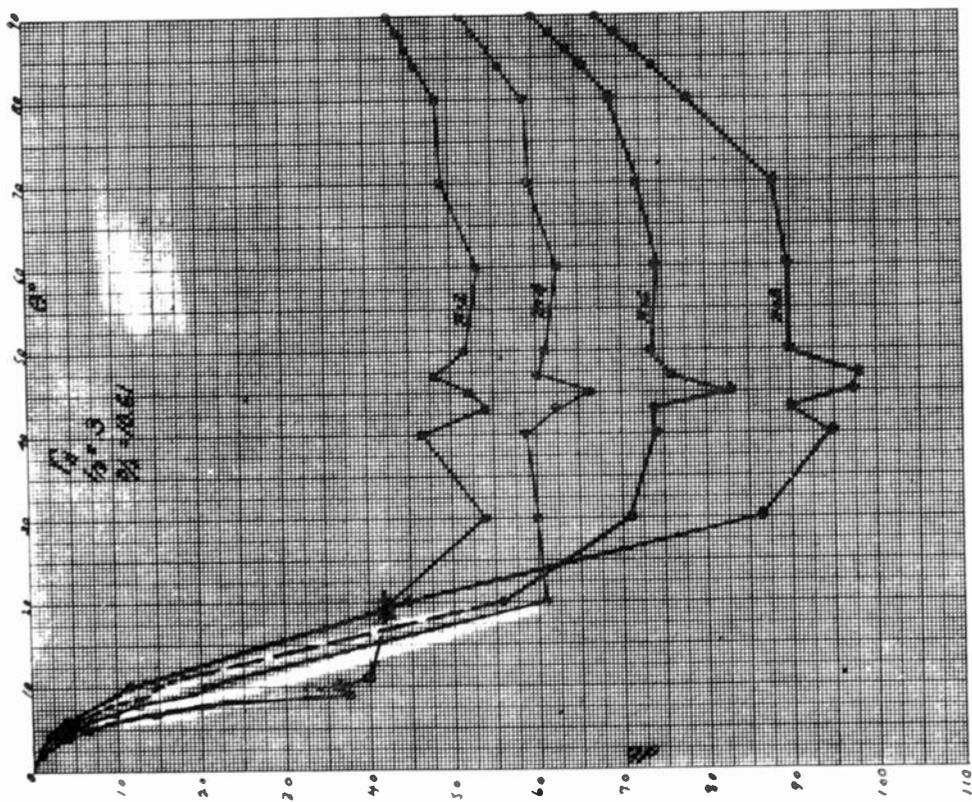


FIG. 7

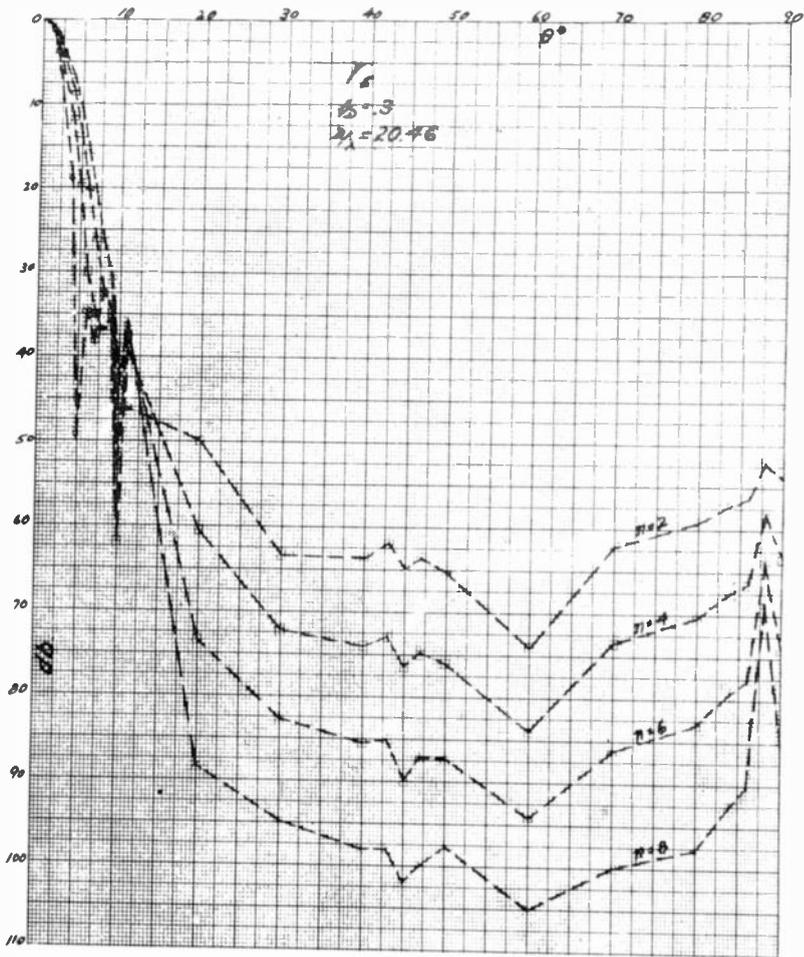


Fig. 9

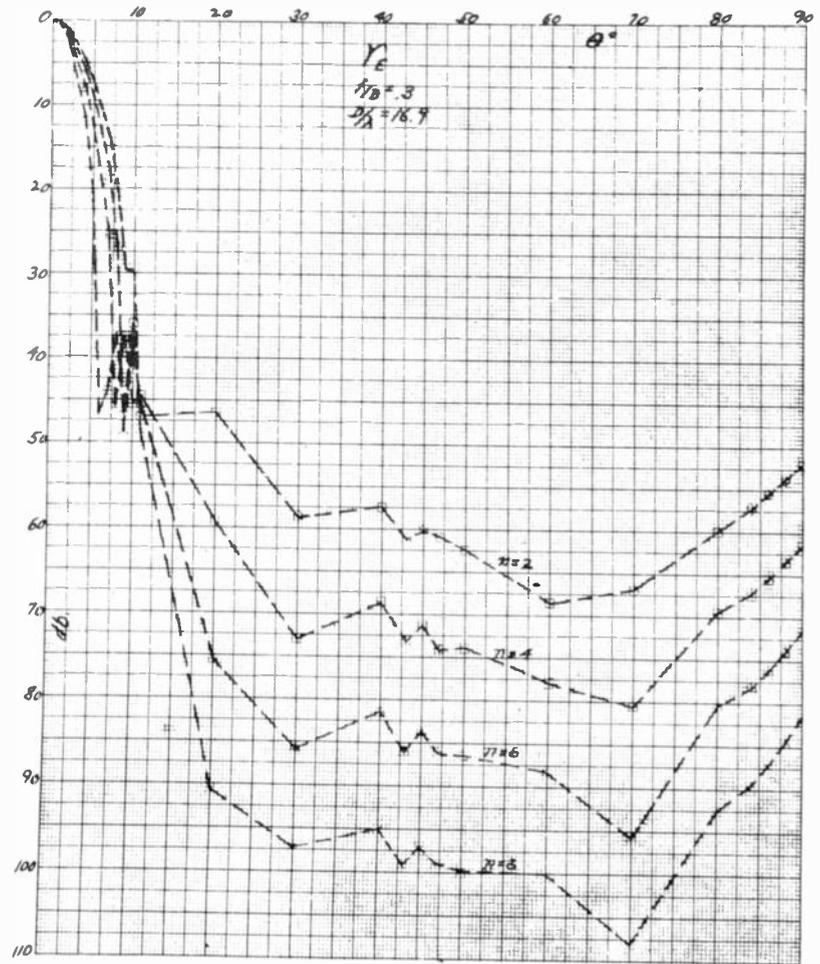


Fig. 10

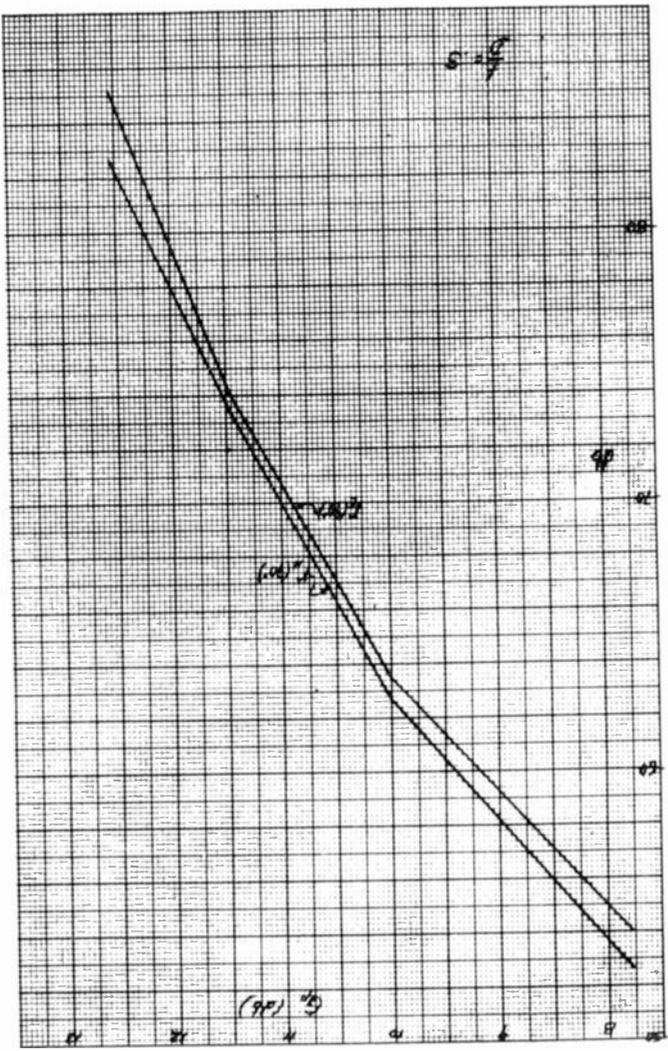


Fig. 12

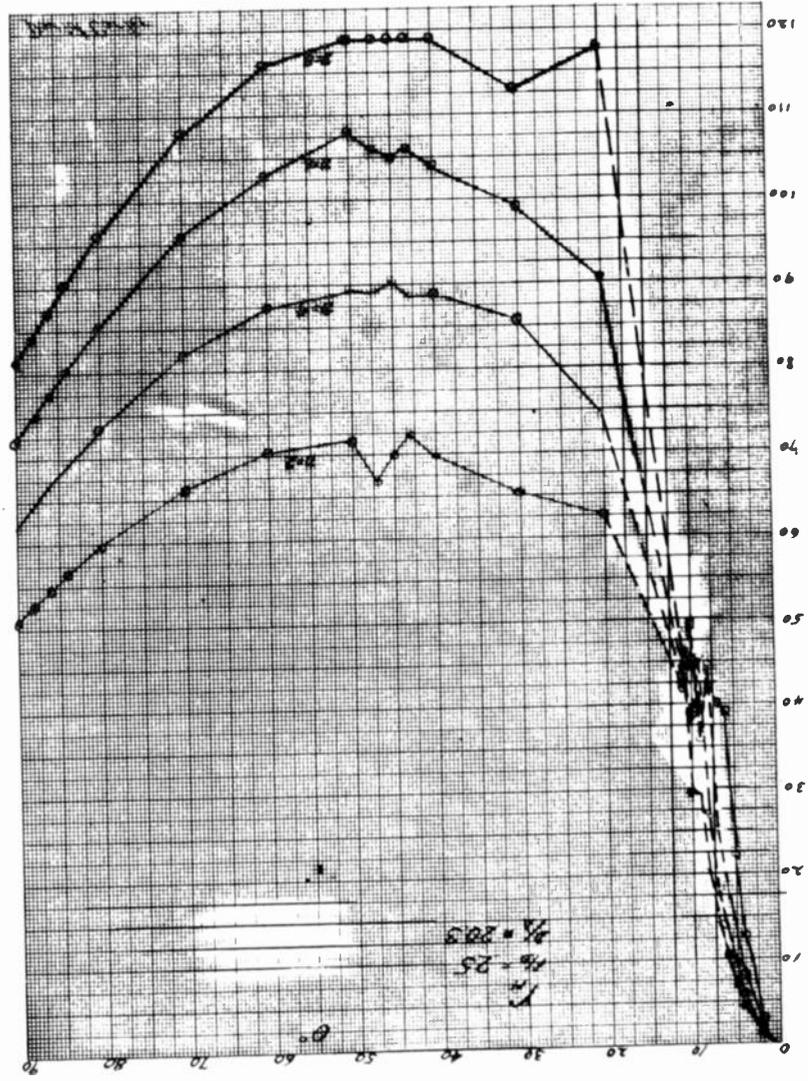


Fig. 11

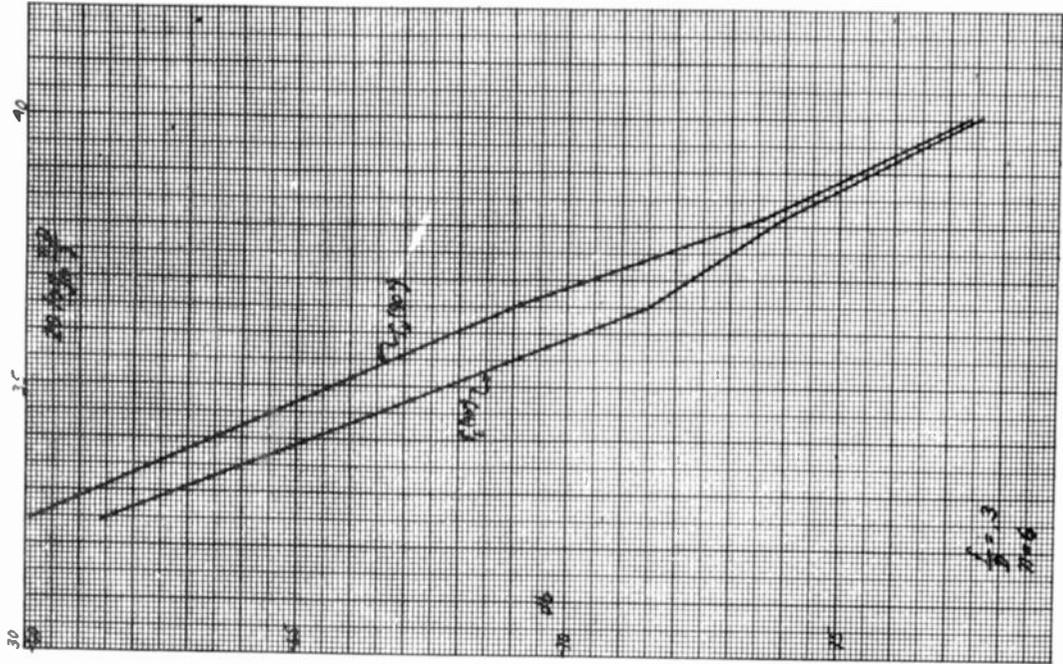


Fig. 14

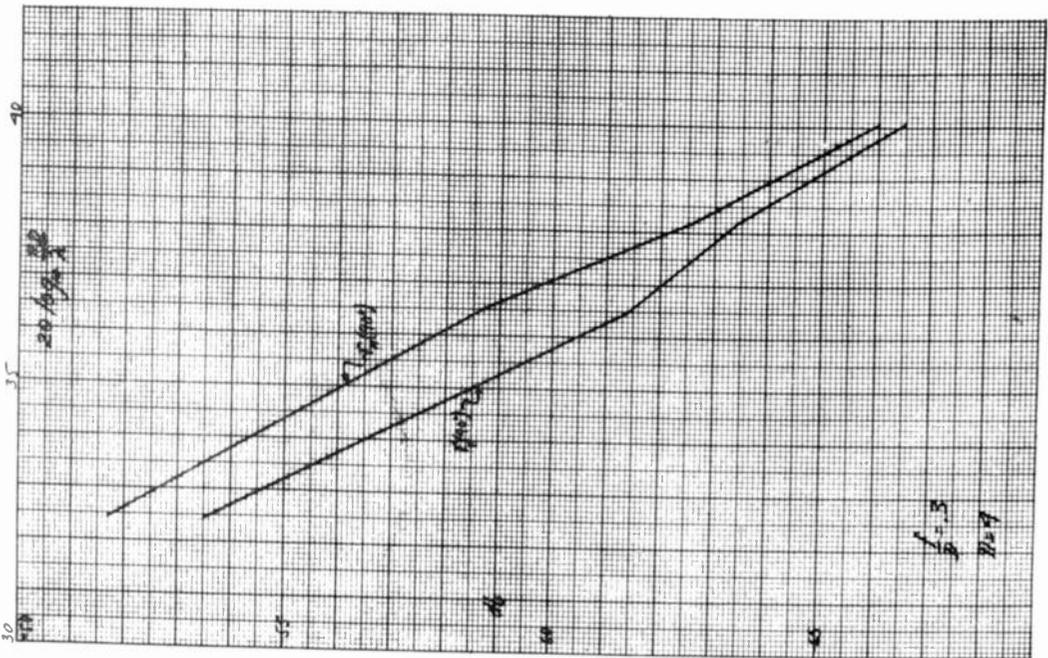


Fig. 13

THEORETICAL GAIN OF FLAT MICROWAVE REFLECTORS

D. R. Crosby
R.C.A. Victor Division
Camden, New Jersey

Summary

Flat sheets are sometimes used as microwave reflectors at the top of towers. Calculations are presented showing how these reflectors may produce a system gain rather than a loss, however this gain exists in a qualified sense, as the antenna at the ground level is taken as the gain standard. The reflectors exhibit no gain when the gain standard is an antenna having the same area as the reflector. Reflectors exhibit several characteristics which complicate their use in practice.

The Two Possible Approaches

The system to be analyzed consists of a paraboloid antenna at the base of the tower, and a flat reflector tilted at a 45° angle at the top of the tower. This flat reflector is elliptical so that its projection on a horizontal plane is a circle. When we speak of the diameter of the reflector we shall mean its smaller or projected diameter.

Any antenna system may be analyzed either as a transmitting system or as a receiving system. By an extension of the reciprocity theorem, it may be shown that the same system gain results from either approach, however the physical interpretation is quite different in the two approaches.

Using the receiving point of view, we have a plane wave incident on the reflector. This wave sets up currents on the surface of the reflector which in turn cause energy to be radiated in the direction of the ground level antenna. A solution of the problem involves analyzing the near field radiation pattern of the reflector, and the response of the ground level antenna to this radiation pattern.

The following analysis uses the transmitting viewpoint. We consider a particular type of transmitting antenna at the ground level, and study its radiation field to determine the currents produced on the surface of the reflector. Using these currents we compute the intensity of the far field radiation from the reflector.

Physical Behavior

The radiation of the ground level antenna coming from the antenna aperture occupies at first a cylindrical region, but a gradual expansion occurs so that at several hundred feet the pattern becomes conical.

For example a 4 ft. diameter antenna may have a tapered illumination so that the intensity at the edge of the cylindrical region is 10 db below that at the center. At a distance of 100 ft. from the antenna, assuming 6000 mc operation, the diameter of the -10 db line has increased from its original 4 ft. to 8 ft. The amplitude of the reference field on the major axis will be less at the 100 ft. level than that at ground level.

If the radiation in the cylindrical region is intercepted by a large flat reflector, a current pattern is set up on the reflector very similar to the equivalent current sheet in the aperture of the antenna. The effect of the reflector is then to merely change the direction of the central axis of the original radiation pattern. It follows that if a reflector is to have some particular gain greater than zero, there will be a minimum tower height necessary to obtain this gain.

Consider a reflector located where the radiation from the ground level antenna occupies an appreciably expanded cross section. The reflector may intercept a substantial fraction of the total radiated energy. When the reflector is larger than the ground level antenna, its far field pattern is usually more directive or "sharper". One effect on the far field of this pattern sharpening is to increase the field intensity, and this effect may more than offset the loss of energy in the portion of the beam that was not intercepted. Thus using the ground level antenna as the gain standard, the reflector may be said to have gain.

Consider a particular height, where the beam is sufficiently diverged so there is a possibility of a particular gain. There will exist a minimum size reflector for zero gain. As the reflector is increased beyond this size, the gain will increase, reach a maximum, and then decrease. In practical cases this maximum gain may be 2 or 3 db or less. Thus as Fig. 1 shows, the maximum gain at 6000 mc with a 100 ft. tower and a 4 ft. diameter ground level antenna, is less than 3 db. However, if the tower were 300 ft. high, there is a theoretical maximum gain at 6000 mc of over 4 db. The practicality of achieving such gains is not assured due to the large size of the required reflector, in this case with a diameter of over 13 feet.

The higher gains are only realized by reflectors that intercept a small portion of the

radiated energy. From Fig. 1 we see that the optimum reflector for the 900 ft. tower is less than twice the diameter of the optimum reflector for the 300 ft. tower. The reflector for the 900 ft. tower intercepts an appreciably smaller solid angle and a smaller fraction of the radiated energy, then does the reflector on the 300 ft. tower.

A key factor in reflector performance is the phase of the illumination produced by the low level antenna on the reflector. The illumination may be divided into phase zones. The central part of the reflector comprises the first zone in which the phase varies from zero degrees at the center out to 180 degrees at the edge of the zone. The second zone contains the illumination with phase in the range 180 to 360 degrees. The reflector having maximum gain approximately coincides with the first phase zone. The diameter of this zone is approximately proportional to the square root of the tower height. Thus as the tower height increases, the intercepted angle of the first phase zone decreases and the fraction of the total intercepted energy decreases. Thus whether the reflector is mounted on a low tower or a high tower, there exist fundamental restrictions on its gain possibilities. Gain is obtained only by a considerable sharpening of the radiation pattern. This sharpening of the pattern may also occur when the reflector is so small that it introduces a loss in the system.

Results of the Gain Computations

The results of this study are shown in two families of curves; the first in Fig. 1 where the variable is the reflector diameter and the gain standard is the low level antenna. The second family is shown in Fig. 2, where the variable is the diameter of the low level antenna and the gain standard is an antenna having the diameter of the reflector.

In Fig. 1 we consider a four foot diameter low level antenna, as this is a typical value for such applications, and in Fig. 2 we consider an eight foot diameter reflector, as this is also a typical value.

Each curve is a universal curve in that it applies for a particular ratio of frequency to tower heights. For example, the first curve of Fig. 1 applies to a 100 ft. tower at 6000 mc, and also to a 150 ft. tower at 9000 mc.

The middle curve of Fig. 1 has been computed over enough range of reflector diameters to show the nature of the first minimum, and the beginning of the damped oscillatory approach of the gain curve to the line of zero gain. From physical reasoning, the reflector would be expected to have substantially zero gain when it is large enough to intercept almost all of the radiated energy.

A non-realizable limit curve is shown dotted in Fig. 1, evaluated for the 300 ft. tower and 6000 mc. This limit curve assumes that the phase

of the illumination on the reflector is due only to the path length differences from the center of the low level antenna. It is an accurate approximation for sufficiently high towers and small low level antennas. If the gain is of the order of -6 db or less, the limit curve formula holds for all three of the curves shown in Fig. 1. For the regions of maximum gain, the limit curve formula is a good approximation only for very high towers, thus it holds fairly well throughout the range of reflector diameters shown in Fig. 1 for the 900 ft. tower at 6000 mc.

Fig. 2 is based on the same calculations as is Fig. 1. When converted to the parameters of Fig. 2 these calculations were not extensive enough to carry the gain curves up to their maximum values, however they cover most practical situations. That these gain curves do not reach the zero db line can be understood from physical considerations. Since the gain standard is now the same size as the final radiating surface, the pattern from the reflector is much like that from the gain standard. Thus any lost energy due to the finite size of the reflector results in a gain less than zero. To be made strictly quantitative, the above argument must account for the difference in gain between a uniformly illuminated flat sheet and an antenna of the same area but with tapered illumination. These corrections do not effect the above conclusions.

We note from Fig. 2 that for a 4 ft. diameter low level antenna, at 6000 mc, the reflector loss for towers in the range of 100 to 300 ft. is from 3.5 to almost 7 db.

Comparison With Other Work

The reflector problem has also been studied by W.C. Jakes¹, who used the receiving system point of view. His conclusions are presented in the form of curves plotted to a different set of parameters. Replotting his curves to the parameters of Fig. 1 shows an agreement within 1/2 db. This check can be made only in the region covered by both sets of curves, which is the region of maximum gain and down to about -2 db. His curves confirm the conclusion that the curves of Fig. 2 do not reach the zero gain line.

Complicating Practical Factors

In considering the application of reflectors to microwave relay service, we note several factors tend to complicate their use:

1. The increased cost of tower to take care of the added wind load and to have the torsional rigidity demanded by the narrow beams.

2. The added difficulty of aligning the antenna system due to its having twice the number of components to align, since these components must be aligned with increased precision, and since the reflector is so unwieldy.

3. The added subsequent cost of realignment

due to drift in tower set.

4. The poor side lobe patterns that may interfere with other microwave systems.

5. The high cross talk between the R.F. circuits on the same tower, sometimes necessitating the use of 4 frequencies at a relay station instead of 2 frequencies as is practical when using antennas with superior shaped patterns.

6. The difficulty of adding reflectors on an installed system to provide expansion to include a side leg circuit.

7. The location of the ground level antenna being restricted by the azimuth of the microwave path.

Calculations

From the definition of the gain of the reflector, we have:

$$(DB) = 20 \log \frac{|F_T|}{|F_T'|} \quad (1)$$

where

F_T = volts per meter, field strength at the reference far field point on the central axis, remote from the reflector.

F_T' = volts per meter, at the reference point when the ground level antenna is in the normal position of the reflector.

We get $|F_T|$ by integrating in equation (2) over the surface of the reflector a quantity proportional to the surface current on the reflector.

$$|F_T| = \frac{\omega\mu}{4\pi} \left| \int_0^{A_2} \frac{\sigma_2 dA_2}{h_2} \right| \quad (2)$$

where

A_2 = reflector area, square meters

h_2 = distance from an element on the reflector surface to the far field reference point, meters

σ_2 = amperes per meter, current density on the reflector

ω = radian frequency

μ = space permeability

The distance factor h_2 appearing in the de-

nominator of the integrand may be chosen arbitrarily large so that its variations over the range of integration are small. Thus we may take it outside the integral sign, and use for it the center-line distance h_2 from reflector to the far field reference point. Both F_T and σ_2 are sinusoidal in time and directed in space. We represent the phase and amplitude of their time dependence by expressing them as complex numbers, the modulus being equal to the rms amplitude of the sinusoid. Specification of their space direction is not required, as each vector is parallel to other vectors in the regions of interest.

We get σ_2 by integrating in equation (3) over the aperture of the low level antenna a quantity proportional to the surface current on the aperture. The complex factor in the integrand provides for the phase of the illumination on the reflector. The distance factor h_1 can be removed from the denominator of the integrand as before. The values of h_1 are sufficiently large and the areas over which we integrate are sufficiently small to permit this approximation. Note that equation (2) and (3) differ by a constant of $2\sqrt{\frac{\epsilon}{\mu}}$ which is the ratio of the current density on the conductor surface to the electric field intensity in a traveling wave when normally incident on a perfect, plane conductor.

$$\sigma_2 = 2\sqrt{\frac{\epsilon}{\mu}} \frac{\omega\mu}{4\pi} \int_0^{A_1} \frac{E e^{-j \frac{2\pi\delta}{\lambda}}}{h_1} \sigma_1 dA_1 \quad (3)$$

where

A_1 = aperture area of low level antenna, square meters.

h_1 = distance from a point on reflector to a point on aperture of low level antenna, meters

σ_1 = amperes per meter, current density over aperture of low level antenna

δ = path length variation from the central axis distance of an element in A_1 to an element in A_2 , meters

λ = free space wavelength, meters

We get δ from geometry, with suitable approximations.

$$\delta = \frac{R_1^2 + R_2^2 - 2R_1R_2 \cos(\theta_1 - \theta_2)}{2H_1} \quad (4)$$

where

R_1 = meters, radius of aperture of low level antenna, considered as a coordinate variable.

R_2 = meters, radius of reflector, con-

sidered as a coordinate variable.

θ_1 = radians, azimuth angle of aperture of low level antenna, considered as a coordinate variable.

θ_2 = radians, azimuth angle of reflector, considered as a coordinate variable

In equation (4), we consider the case where the reflector is horizontal, and so parallel to the low level antenna. We thus assume in these calculations that, providing the projected area is held constant, the only significant effect of tilting the reflector to its final 45° position, is to change the direction of the principal axis of the radiation pattern.

We consider that the antenna can be represented by a current sheet across its aperture, of constant phase over the aperture, but with a quadratic 10 db taper, thus giving

$$\sigma_r = \sigma_0 \left[1 - \left(\frac{2\rho}{D_1} \right)^2 \times .684 \right] \quad (5)$$

σ_0 = amperes per meter, current density at center of low level antenna

D_1 = meters, diameter of low level antenna

The remainder of the analysis consists of substituting equations (2), (3), (4), (5), in (1), simplifying and completing the integrations. We can evaluate $|E'|$ by putting (5) into an equation of the form of (2). Completing this integration and making the indicated substitutions we get:

$$(DB) = 20 \log \frac{1}{\lambda H_1 A_1^2 \cdot 658} \left| \int_0^{A_2} \int_0^{A_1} E^{-j \frac{2\pi}{\lambda} \left(\frac{\rho_1^2 + \rho_2^2 - 2\rho_1 \rho_2 \cos(\theta_1 - \theta_2)}{2H_1} \right)} \left[1 - \left(\frac{2\rho}{D_1} \right)^2 \cdot 684 \right] dA_1 dA_2 \right| \quad (6)$$

After making changes in variable and suitable approximations, we get from (6) the final gain equation (7). The numerical computations were a direct evaluation of equation (7). The integrands were calculated using a slide rule, and then plotted on 8 x 11 paper, from which the integrals were evaluated using a planimeter. For a particular set of values of D_1 , λ , and H_1 , the gain was evaluated as a function of D_2 . This evaluation took about two-man-weeks of work.

$$(DB) = 10 \log \frac{1}{4} \left\{ \left[\int_0^x \sum_{n=1}^4 \cos(x^2 + y_n^2) J_0(2xy_n) x dx \right]^2 + \left[\int_0^x \sum_{n=1}^4 \sin(x^2 + y_n^2) J_0(2xy_n) x dx \right]^2 \right\} \quad (7)$$

where

J_0 = Bessel Function of zero order

$$x = \frac{D_2 \sqrt{\pi}}{2\sqrt{\lambda H_1}}$$

$$y_1 = \frac{D_1 \sqrt{\pi}}{2\sqrt{\lambda H_1}} \times .210$$

$$y_2 = \frac{D_1 \sqrt{\pi}}{2\sqrt{\lambda H_1}} \times .516$$

$$y_3 = \frac{D_1 \sqrt{\pi}}{2\sqrt{\lambda H_1}} \times .703$$

$$y_4 = \frac{D_1 \sqrt{\pi}}{2\sqrt{\lambda H_1}} \times .897$$

The summation in (7) results from considering the aperture of the low level antenna as composed of four concentric zones. Each zone is then approximated by a narrow circular band of current, whose field can be integrated in terms of Bessel Functions.

For sufficiently high towers and small low level antennas, (7) reduces to the limit formula (8). Note that this formula is independent of the illumination taper or size of the ground level antennas.

$$(DB) = 20 \log \left[2 \sin \frac{D_2^2 \pi}{8\lambda H_1} \right] \quad (8)$$

if $D_1, D_2 < \frac{\lambda H_1}{2}$ & $D_1 < \sqrt{\lambda H_1}$

For reflectors less than the optimum size and having gains less than about -4 db, equation (8) reduces to (9):

$$(DB) = 20 \log \frac{D_2^2 \pi}{4 \lambda H_1} \quad (9)$$

The gain for other diameters of the low level antenna may be obtained by replotting the data of Fig. 1 in terms of gain versus X, with y as a parameter:

$$y = \frac{D_1 \sqrt{\pi}}{2 \sqrt{\lambda H_1}} \quad (10)$$

Each of the 3 curves in Fig. 1 will become a curve with a constant value of y. This plot may

then be entered for a range of arbitrarily chosen values of D_1 , D_2 , λ and H_1 .

The gain used in Fig. 2 is that from Fig. 1 after correction for the reflector size, by equation (11):

$$(DB)_{FIG 2} = (DB)_{FIG 1} - 20 \log \frac{D_2}{D_1} \quad (11)$$

As incidental information, the divergence of the beam from the low level antenna can be obtained from the integrands of equation (7).

References

1. "A Theoretical Study of An Antenna-Reflector Problem", W.C. Jakes, Jr., Proc. I.R.E., February 1953, page 272.

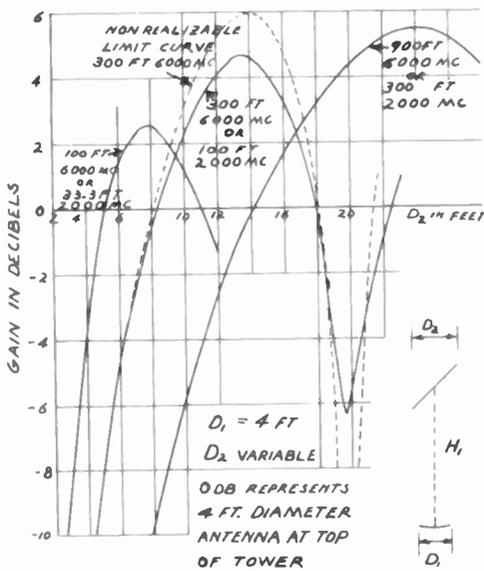


Fig. 1 - Gain of a flat reflector.

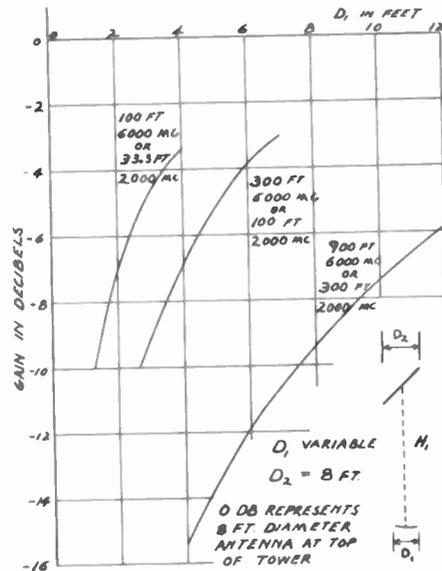


Fig. 2 - Gain of a flat reflector.

ISOTROPIC VARIABLE INDEX MEDIA
W.O. Puro and K.S. Kelleher
Melpar, Inc.
Alexandria, Va.

Introduction

For some time it has been realized that an artificial dielectric could be formed by embedding obstacles in a base dielectric. In all of the cases considered up to the present time the base dielectric has been air or some light weight plastic whose dielectric constant is very close to that of air. Such artificial dielectrics are structurally no better than their base material--which is inherently poor. The most successful isotropic media was studied by Corkum¹ at the Air Force Cambridge Research Center. He considered spherical obstacles of glass or metal embedded in Styrofoam. No tests were made on the mechanical properties of this media, but it can be assumed that its application would be limited to devices not subject to shock or vibration.

This paper presents a study of an artificial dielectric material which has better mechanical properties than the materials previously considered. This media is obtained by utilizing spherical voids in a relatively high dielectric base material. The base material may be polystyrene or a similar plastic which has the necessary strength characteristic. Through use of several such base materials, it is possible to obtain a variation in dielectric constant from 1.1 to 2.6, a range which has been of interest in antenna applications.

Theoretical Analysis

A formula can be developed which yields the value of the dielectric constant in terms of the radii of the spherical voids, the number of voids per unit volume, and the dielectric constant of the base material.

The mechanism employed in a theoretical analysis of this problem consists of considering that a uniform field exists in the base medium and then imagining that some of the base material is removed to leave the spherical voids. The addition of these voids causes an alteration in the field which can be related to an effective change in dielectric constant. The change in the field is usually evaluated by introduction of the polarization vector P which has two equivalent definitions: $P = D - \epsilon_0 E$ and $P = \alpha E' N/V$. Here D and E are the displacement vector and the original electric field vector, while E' is the electric field vector at one of the spherical voids. N is the number of voids in the volume V , α is the polarizability of a single void and ϵ_0 is the permittivity of free space. It should first be noted that since E is the field in the base medium before addition of the voids, the first definition must be written $P = D - \epsilon_1 E$ where ϵ_1 is the permittivity of the base medium without voids. The vector P then vanishes, as required in the case where no voids are present ($D = \epsilon_1 E$).

In order to determine the permittivity ϵ of the medium with spherical voids and so obtain the dielectric constant we need only equate the two expressions for P and substitute the appropriate values of D , E' , and α . By definition, $D = \epsilon E$. The

expressions for E' and α can be found by standard techniques.

The field E' at the center of a spherical void is determined in a manner similar to that used by Slater and Frank.³ The spherical hole in that analysis is replaced by a sphere of material with permittivity ϵ_1 , and the charge on the spherical surface is given by $-|P| \cos \theta$ rather than $|P| \cos \theta$. It can then be shown that $E' = (|E| - |P|/3\epsilon_1)e$ where e is a unit vector in the direction of E , ($E = |E|e$).

The polarizability α of a spherical void in a base medium of permittivity ϵ_1 is given by Slater and Frank⁴ as:

$$\alpha = \frac{4\pi\epsilon_1 R^3}{2\epsilon_1 + \epsilon_0} \left(\frac{\epsilon_0 - \epsilon_1}{2\epsilon_1 + \epsilon_0} \right). \quad (1)$$

Since it is known that $P = \alpha E' N/V$, we can write

$$P = (4\pi\epsilon_1 R^3 CN/V) (|E| - |P|/3\epsilon_1)e, \quad (2)$$

where $C = (\epsilon_0 - \epsilon_1)/(2\epsilon_1 + \epsilon_0) = (1 - K_1)/(1 + 2K_1)$ and K_1 , the dielectric constant of the base medium, is by definition ϵ_1/ϵ_0 . Equation (2) can be simplified somewhat if we introduce the concept of fractional volume used by Corkum.⁵ For an isotropic array of spheres, the fractional volume $F = 4/3(\pi R^3 N/V)$. If we use this expression and the fact that $P = D - \epsilon_1 E = (\epsilon - \epsilon_1)E$ in Equation (2), there results: $\epsilon - \epsilon_1 = FC(2\epsilon_1 + \epsilon)$. This can be rewritten and solved for $\epsilon/\epsilon_1 = K/K_1$, or

$$K/K_1 = (1 + 2FC)/(1 - FC) \quad (3)$$

It might be noted that this result is related to that obtained by Corkum.

Experimental Program

In any practical application the size of the voids and the spacing between voids is an appreciable fraction of a wavelength. Since the analysis does not explicitly consider such a medium, some extended experimental work was necessary in order to determine the validity of the final formula for dielectric constant. All of this work was based on the shorted waveguide technique⁶ at a frequency of 5000 mc.

Description of Apparatus

The shorted waveguide technique for measuring the properties of a dielectric offers the convenience of small samples, freedom from external influences, and adaptability to ordinary laboratory facilities. Choice of a frequency of 5000 mc furthermore permits samples to be of a size which can easily be machined to reasonable tolerances.

A photograph of the waveguide apparatus is shown in Figure 1. A 2K43 reflex klystron with square wave modulation energizes the waveguide assembly, comprising a frequency meter, 25 db attenuator, slotted line, and shorted cavity. A dial indicator coupled to the movable probe of the slotted line permits reading the position of the standing wave minimum to 0.001 cm. Sufficient padding is provided to prevent the frequency of the

oscillator from changing when a dielectric is inserted into the shorted cavity.

The shorted cavity, Figure 2, consists of a length of waveguide with cross section 0.937 inches x 1.874 inches I.D. to accept the test samples and a $\lambda_g/2$ long tapered section to connect the main body to standard RG-49/U waveguide (0.874 inches x 1.874 inches I.D.). A shorting plate terminates the cavity and facilitates removal of the test pieces while the two screws which hold it insure a constant position of the electrical short circuit. A choke flange on the cavity permits proper coupling to the slotted line.

Description of Samples

The dielectric sample is made up of sheets of base dielectric which have hemispherical depressions milled in one face. When two sheets are placed together spherical voids are created, forming a cubical lattice.

Two base dielectric materials were used,-- Teflon ($K_1 = 2.02$) and a Sponge Rubber Company expanded dielectric material ($K_1 = 1.62$). Two types of lattice were considered--the first had one row and two rows of voids across the narrow and wide dimensions of the waveguide respectively, while the second had two rows and four rows of voids. In each case the spacing between the center of the outer voids and the waveguide wall is one half the center to center spacing. In the case of the cubical lattice consisting of two and four rows of voids, the two diameters of spheres used, 1/4 inch and 3/8 inch, give fractional volumes of 7.96 per cent and 26.85 per cent respectively. The lattice with one and two rows of voids of 1/2 inch diameter gave a fractional volume of 7.96 per cent. The maximum number of transverse rows in the one by two lattice was five while in the two by four lattice it was seven.

Experimental Procedure

Determination of the dielectric constant of materials having a low loss factor by the shorted line method requires the measurement of the guide wavelength and the distance from the face of the dielectric sample to the first minimum of the standing wave pattern.

The dielectric constants of solid samples of the two dielectrics were first determined in order to form a basis for predicting values to be expected when the spherical voids were added. In each case several different lengths were measured in order to estimate the accuracy of the measurement technique.

The following relations give the dielectric constant after the separation of the first standing wave minimum from the face of the dielectric sample has been established:

$$\tan \beta d / \beta d = -(\lambda_c / 2 \pi d) \tan (2\pi x_0 / \lambda_g),$$

$$K_1 = [1 + (\lambda_c \beta d / 2 \pi d)^2] / [1 + (\lambda_c / \lambda_g)^2],$$

where λ_g = guide wavelength outside the sample
 λ_c = cutoff wavelength of the waveguide
 d = length of the sample
 x_0 = distance from sample to the first standing wave minimum.

Data for four samples of the dielectric media are shown in Figures 3, 4, and 5. The samples were progressively shortened by cutting off one transverse row of voids between successive measurements. As can be seen, the experimental values are quite close to the predicted theoretical values. It is estimated that the measurements can be reproduced with an accuracy of approximately 1 per cent. Because the theoretical values are based on the measured values for the base dielectric, they include the errors involved in the measurement of the properties of the base material. For the case of 1/4 inch spheres in the dielectric, Figures 3 and 4 and the upper curve in Figure 5, the measured value for the dielectric constant, did not vary more than 2 per cent from the theoretical value.

In this case, the diameter of the spheres is approximately 0.15 of the wavelength in the dielectric, and the spacing between centers of adjacent spheres is 0.28λ .

The measured values of dielectric constant shown in the lower curve of Figure 5 represent the case of 1/2 inch voids spaced 0.937 inches on centers. In this instance the diameter of the void is 0.30 of the wavelength in the dielectric, and the spacing is 0.56λ . Even in this sample, when the diameter of the void is no longer small compared to the wavelength, the maximum variation in the measured value is not more than 4 per cent.

As a further check of the applicability of the theoretical expressions to the dielectric media under study, the samples were inserted into the shorted cavity in such a manner that the waveguide walls passed through the centers of the spherical voids. The measured values did not differ from the previous results. All of the results indicated that Equation (3) is valid even for small sample lengths and for voids relatively large with respect to a wavelength.

Summary

A new type of dielectric medium, consisting of spherical voids in a base dielectric, has been investigated. A simple expression for the dielectric constant has been obtained in terms of the constant of the base medium and the fractional volume occupied by the spherical voids. The experimental measurements showed that this expression was valid even when the diameter of the voids was 0.3λ and the spacing between voids was 0.56λ .

References

1. R.W. Corkum, "Isotropic Artificial Dielectric," Proc IRE, May 1952, pp. 574-587.
2. J.C. Slater and N.H. Frank, "Electromagnetism," p. 41, McGraw Hill, New York, 1947.
3. Ibid., p. 109.
4. Ibid., p. 50.
5. Loc. cit., p. 574.
6. T.W. Dakin and C.N. works, "Microwave Dielectric Measurements," Journal of Applied Physics, Vol. 18, No. 9, September 1947.

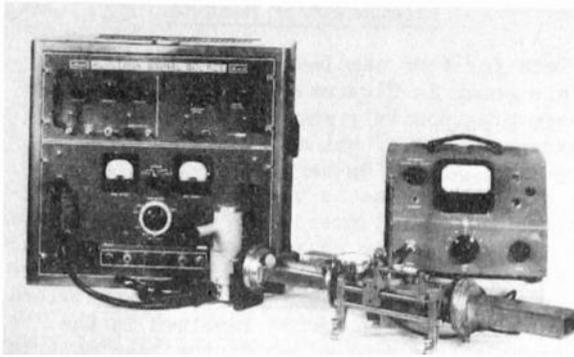


Fig. 1
Experimental arrangement for determining dielectric constant.

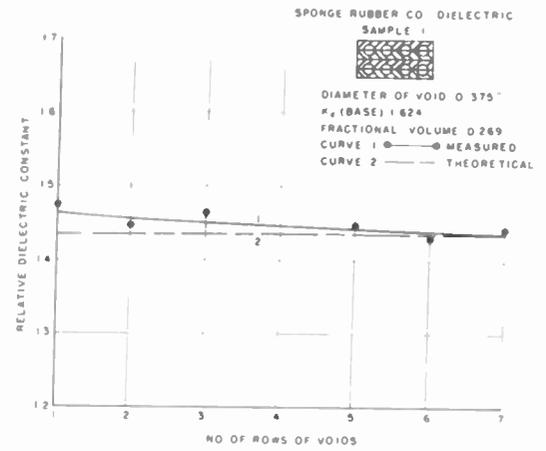


Fig. 3
Dielectric constant of Sponge Rubber Products Company material with voids.

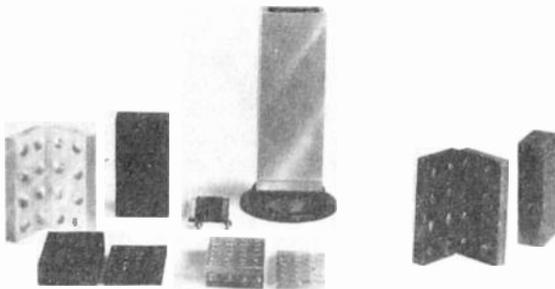


Fig. 2
Test cavity with dielectric samples containing spherical voids.

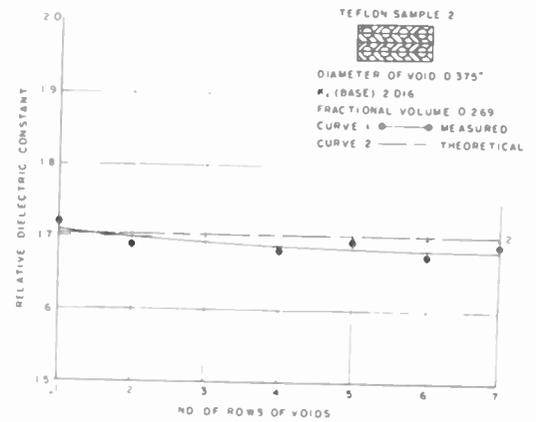


Fig. 4
Dielectric constant of Teflon with voids.

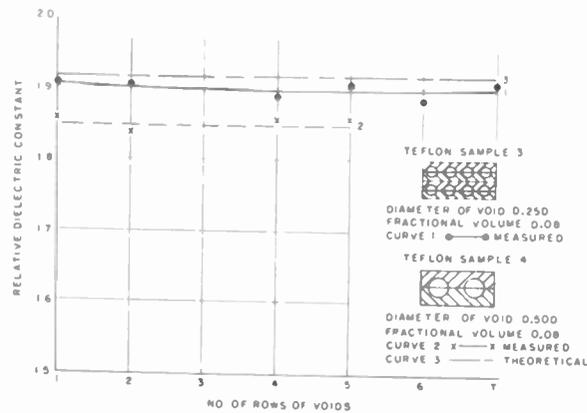


Fig. 5
Dielectric constant with same fractional volume but with different size voids.

THE CHARACTERISTICS OF A VERTICAL ANTENNA WITH A RADIAL CONDUCTOR GROUND SYSTEM *

James R. Wait and W.A. Pope
Radio Physics Laboratory
Defence Research Board
Ottawa, Ont. Canada

Summary

Employing an approximate method the input impedance of a ground based vertical radiator is calculated. The ground system consists of a number of radial conductors buried just below the surface of the soil. The integrals involved in the solution are evaluated, in part, by graphical methods. The final results are plotted in a convenient form to illustrate the dependence of the impedance on number and length of radial conductors for a specified frequency, antenna height, and ground conductivity. It is finally shown that under usual conditions the radiated fields are modified by only a few percent due to the presence of the ground system.

Introduction

Antenna systems for low radio frequency are designed, usually, to work in conjunction with a radial wire ground system buried just below the surface of the earth. The purpose of this wire grid is to provide a low-loss return path for the antenna base current and consequently to improve the efficiency of the transmission.

The rules for ground system design are usually empirical and based on the results of experiments on existing installations. The first systematic study of this problem was carried out by Brown^{1,2} and his associates who were mainly concerned with the operation of half-wave antennas for the broadcast band. Sometime later Abbott³ developed a procedure to select the optimum number of radial conductors, given the values of the electrical constants of the ground. An important related problem is the actual change of input impedance of the antenna for different sizes and types of ground systems. This analysis has been carried out by Leitner and Spence⁴ and more recently by Storer⁵, for a vertical antenna situated over a perfectly conducting disc. However, they only considered the case where the surrounding medium was free space.

The purpose of this paper is to consider, in some detail, the characteristics of a vertical antenna of any length situated over a circular ground screen composed of N radial wires of equal spacing situated at the interface between the air and a semi-infinite homogeneous ground. An approximate method to calculate the input impedance will be employed similar to that described by Monteath⁶ who developed an extension to the compensation theorem of electric-circuit theory. The current distribution on the antenna

is considered to be sinusoidal.

With reference to a cylindrical polar coordinate system (ρ, ϕ, z) the antenna of height h is coincident with the positive z axis as indicated in Fig. 1. The ground screen is of radius a and lies in the plane $z = 0$ which is also the surface of the ground. The conductivity and dielectric constant of the ground are denoted by σ and ϵ respectively and the dielectric constant of the air by ϵ_0 . The permeability of the whole space is taken as μ which is taken to be that of free space. The intrinsic propagation constant γ and characteristic impedance η of the earth medium are defined by

$$\gamma = [i\mu\omega(\sigma + i\omega\epsilon)]^{1/2}$$

and

$$\eta = [i\mu\omega / (\sigma + i\omega\epsilon)]^{1/2}$$

where ω is the angular frequency. The propagation constant γ_0 and characteristic impedance η_0 of the air are then defined by

$$\gamma_0 = i(\epsilon_0\mu)^{1/2}\omega = i2\pi/\lambda_0 = i\beta$$

and

$$\eta_0 = (\mu/\epsilon_0)^{1/2} \approx 377 \Omega$$

where λ_0 is the wavelength in air.

The self impedance at the terminals of the antenna is now denoted by Z_T and can be broken into two parts by setting, $Z_T = Z_0 + \Delta Z_T$ where Z_0 is the self-impedance of the same antenna if the ground plane were perfectly conducting and infinite in extent. On the other hand ΔZ_T is the difference between the self-impedance of the antenna over the imperfect and the perfect ground plane. It is called the self-impedance increment and can be written in terms of a real and imaginary part as follows:

$$\Delta Z_T = \Delta R_T + i\Delta X_T \quad (1)$$

where ΔR_T and ΔX_T represent the resistance and reactance increment. If the current at the terminals of the antenna is I_0 , the power required to maintain this current is $I_0^2 R_T$ watts. If the ground were perfectly conducting the input power would be $I_0^2 R_0$ where R_0 is the real part of Z_0 . The additional power required to maintain the same current I_0 at the terminals is $I_0^2 \Delta R_T$.

* Work carried out under Project No. DL8-95-55-07.

It is seen, therefore, that the quantity ΔR_T represents an important parameter of a radio frequency antenna.

The Impedance Calculation

It is shown in appendix I that the impedance increment is given by

$$\Delta Z_T = -\frac{1}{I_0^2} \int_0^{\infty} H_{\phi}^{\infty}(\rho, 0) E_{\rho}(\rho, 0) 2\pi\rho d\rho \quad (2)$$

where $H_{\phi}^{\infty}(\rho, 0)$ is the magnetic field of the antenna tangential to a perfectly conducting ground plane and $E_{\rho}(\rho, 0)$ is the tangential electric field on the imperfect ground. If the current on the antenna is $I(z)$ amps it follows that

$$H_{\phi}^{\infty} = -\frac{1}{2\pi} \frac{\partial}{\partial \rho} \int_0^h \frac{e^{-i\rho(\gamma^2 + \rho^2)^{1/2}}}{(\gamma^2 + \rho^2)^{1/2}} I(z) dz \quad (3)$$

For thin antennas it can be usually assumed that the current distribution is sinusoidal, that is

$$I(z) = I_0 \sin(\alpha - \beta z) / \sin \alpha \quad (4)$$

if the antenna is fed at the base (i.e. a monopole). The quantity α is determined by the height of the antenna and the top-loading, that is

$$\alpha = \beta(h + h')$$

The quantity h' specifies the amount of top loading and is usually obtained from experiment. For a thin antenna without top-loading (i.e. unterminated case), $\alpha = \beta h$. The integrals of the type indicated in equation (3) can be expressed in closed form for a sinusoidal current distribution. The result is given by,

$$H_{\phi}^{\infty}(\rho, 0) = -\frac{iI_0}{2\pi \sin \alpha} \left[\frac{e^{-i\rho r}}{\rho} \cos(\rho h - \alpha) - \frac{e^{-i\rho \rho} \cos \alpha - \frac{h}{\rho} e^{-i\rho r} \sin(\rho h - \alpha)}{\rho} \right] \quad (5)$$

where $r = (\rho^2 + h^2)^{1/2}$

Since the electric field $E_{\rho}(\rho, 0)$ is an unknown quantity it is necessary to make several simplifications at this stage. Since $|\gamma| \gg \beta$ an approximate boundary condition (see appendix II) is employed, expressed by

$$E_{\rho}(\rho, 0) \approx -\eta_c H_{\phi}(\rho, 0) \quad (6)$$

where $H_{\phi}(\rho, 0)$ is the tangential magnetic field for the imperfect ground system and η_c is the surface impedance of the air-ground interface. If ρ is greater than a , the radius of the ground screen, η_c can be replaced by η . If ρ is less than a , η_c is the intrinsic impedance η_a of the ground screen which is in parallel with the impedance η of the ground. In a previous investigation^{7,8} of this problem it was assumed that η_a was zero so the ground system was equivalent to a perfectly conducting disc of radius a . The limits of the integration in equation (2) are then from $\rho = a$ to $\rho = \infty$. A more general case is when η_a is comparable in magnitude to η , in which it follows that

$$\eta_c = \frac{\eta \eta_a}{\eta + \eta_a} \quad \text{for } 0 \leq \rho \leq a \quad (7)$$

where

$$\eta_a = \frac{c \eta_0 d}{\lambda} \log_e \frac{d}{2\pi c} \quad \text{and}$$

where d is the spacing between the radial conductors and c is the radius of the wire. The expression for η_a has been derived⁹ for a wire grid in free space where it was necessary to assume that $|\gamma_0 d| \ll 1$. Since the grid is lying on the ground plane, this restriction must be replaced by $|\gamma_e d| \ll 1$ where γ_e is the effective propagation constant for propagation along a thin wire in the interface and is given by¹⁰

$$\gamma_e = \left(\frac{\gamma_0^2 + \gamma^2}{2} \right)^{1/2}$$

If there are N radial conductors it can be seen that d can be replaced by $2\pi\rho/N$ since N is usually of the order of 100.

It is assumed also that $H_{\phi}(\rho, 0)$ is not very different from $H_{\phi}^{\infty}(\rho, 0)$ in the region of the ground plane where the losses are significant. This approximation has also been discussed previously^{7,8} and it certainly appears to be valid if $|\gamma| \gg \beta$.

The impedance increment ΔZ_T is then written in the following form

$$\Delta Z_T = \Delta Z + \Delta Z_a \quad (8)$$

where

$$\Delta Z \approx \eta \int_a^{\infty} [H_{\phi}^{\infty}(\rho, 0)]^2 2\pi\rho d\rho \quad (9)$$

and

$$\Delta Z_a \approx \int_0^a \frac{\eta_a \eta}{\eta + \eta_a} [H_{\phi}^{\infty}(\rho, 0)]^2 2\pi\rho d\rho \quad (10)$$

The first expression ΔZ corresponds to the self-impedance of the monopole over a perfectly conducting discoid, whereas the second expression ΔZ_a accounts for the finite surface

impedance of the radial conductor system.

It is instructive to consider ΔZ , first in some detail, since the integrations can be carried out and the result expressed in terms of the exponential integral defined by

$$Ei(-i\beta a) = - \int_a^{\infty} \frac{e^{-i\beta\rho}}{\rho} d\rho \quad (11)$$

The procedure was outlined previously^{7,8} where an expression for ΔZ was given for any value of the height h and top loading h' . The special case where the antenna is unterminated ($h' = 0$) is given by

$$\begin{aligned} \Delta Z = & \frac{\eta}{4\pi \sin^2 \beta h} \left\{ Ei[-i\beta(r_0+h)] e^{i\beta h} \right. \\ & + Ei[-i\beta(r_0-h)] e^{-i\beta h} + 2 \cos^2 \beta h Ei(-2i\beta a) \\ & + 4 \cos \beta h \left\{ Ei(-i\beta h q) - Ei[-i\beta h(q+i)] e^{i\beta h} \right. \\ & \left. \left. - Ei[-i\beta h(q-i)] e^{-i\beta h} \right\} \right\} \quad (12) \end{aligned}$$

where

$$q = \frac{1}{h} [a + (a^2 + h^2)^{1/2}] \quad \text{and} \quad r_0 = (a^2 + h^2)^{1/2}$$

This equation may be put in a suitable form for computation by employing the relation

$$Ei(-i\beta a) = Ci(\beta a) + i \left[\frac{\pi}{2} - Si(\beta a) \right] \quad (13)$$

where $Ci(\beta a)$ and $Si(\beta a)$ are the cosine and sine integrals respectively as defined and tabulated by Jahnke and Emde¹¹. The results of the calculations are presented in a most general form by plotting $4\pi\Delta Z/\eta$ as a function of a/h for various values of h/λ as shown in Figs. 2a and 2b. It can be seen that the magnitude of ΔZ increases without limit as a approaches zero. This formulation is not actually valid in this limiting case since one terminal of the generator would then be connected directly to the earth medium (or to a disc of vanishing radius) and would lead to an infinitely resistive path for the antenna base current.

A slightly more convenient way to illustrate these calculations is to plot $(\sigma/\epsilon)^{1/2} \Delta R$, where ΔR is the real part of ΔZ , as a function of a/λ as shown in Fig. 3 where f is the frequency in c/s. This is only permissible if displacement currents in the ground are negligible (note if $\epsilon\omega \ll \sigma$, $\eta \approx (i\eta\omega/\sigma)^{1/2}$). It is interesting to note that ΔR actually assumes negative values under certain conditions. In this case, the input power to obtain a given current I_0 at the antenna terminals is actually less for an imperfect ground than for a perfect

ground. This fact can be reconciled by showing that the radiated power is actually reduced if $\Delta R < 0$. A physical explanation for the oscillating nature of the curve is that a wave is reflected from the discontinuity in the surface impedance at $\rho = a$. As the radius a increases the phase lag of the reflected wave will continually increase. This viewpoint is substantiated when it is noted that the period of the oscillations is nearly equal to twice the diameter of the ground screen.

When the ground screen of finite surface impedance is considered the integrations indicated by equation (10) must be evaluated. This appears to be a formidable task for the case when the antenna is of arbitrary length. However, if the antenna is a quarter-wave monopole, without top-loading, the integrations can be carried out fairly readily by graphical means. In this case

$$H_{\phi}^{\infty}(\rho, 0) = - \frac{i I_0}{2\pi\rho} e^{-i\beta(\rho^2 + \lambda^2/4)^{1/2}} \quad (14)$$

which is a special case of equation (5) with $h' = 0$ and $h = \lambda/4$, and hence the integral in equation (10) can then be expressed in the following form:

$$\Delta R_a = \int_0^A \frac{p q \cos \left[\pi(1-4R) + \frac{3\pi}{4} - \tan^{-1} \frac{p+q}{p} \right] dp}{\sqrt{2\pi} [p^2 + (p+q)^2]^{1/2} p} \quad (15)$$

and, similarly, ΔX_a with $\cos(\quad)$ replaced by $\sin(\quad)$. The real dimensionless quantities p , q , A and R are defined by

$$p = 120\pi \delta / \sqrt{2} \quad \text{with} \quad \delta = (\epsilon_0 \omega / \sigma)^{1/2},$$

$$q = \frac{240\pi^2 P}{N} \log_e \frac{P}{NC} \quad \text{with} \quad C = c/\lambda_0,$$

$$A = a/\lambda \quad \text{and} \quad R = \sqrt{p^2 + (p+q)^2}$$

It has been assumed in writing equation (15) in this form that displacement currents in the ground are negligible, that is $(\epsilon\omega/\sigma) \ll 1$. Using the results of the numerical integration for ΔR_a , values of $\Delta R_T (= \Delta R + \Delta R_a)$ are plotted as a function of A in Figs. 4, 5 and 6 for various values of δ and N with $C = 0.1 \times 10^{-5}$, and in Fig. 7 for various values of C for $N = 100$ and $\delta = 0.1$. The value of δ can be readily obtained from Fig. 8 when the ground conductivity σ and the frequency in kilocycles are specified.

It is immediately evident that the oscillations in the curves for ΔR_T have been damped if N is finite. This can be expected

It is seen, therefore, that the quantity ΔZ_r represents an important parameter of a radio frequency antenna.

The Impedance Calculation

It is shown in appendix I that the impedance increment is given by

$$\Delta Z_r = -\frac{1}{I_0} \int_0^{\infty} H_{\phi}^{\infty}(\rho, 0) E_{\rho}(\rho, 0) 2\pi\rho d\rho \quad (2)$$

where $H_{\phi}^{\infty}(\rho, 0)$ is the magnetic field of the antenna tangential to a perfectly conducting ground plane and $E_{\rho}(\rho, 0)$ is the tangential electric field on the imperfect ground. If the current on the antenna is $I(z)$ amps it follows that

$$H_{\phi}^{\infty} = -\frac{1}{2\pi} \frac{\partial}{\partial \rho} \int_0^h \frac{e^{-i\rho(\gamma^2 + \rho^2)^{1/2}}}{(\gamma^2 + \rho^2)^{1/2}} I(z) dz \quad (3)$$

For thin antennas it can be usually assumed that the current distribution is sinusoidal, that is

$$I(z) = I_0 \sin(\alpha - \beta z) / \sin \alpha \quad (4)$$

if the antenna is fed at the base (i.e. a monopole). The quantity α is determined by the height of the antenna and the top-loading, that is

$$\alpha = \beta(h + h')$$

The quantity h' specifies the amount of top loading and is usually obtained from experiment. For a thin antenna without top-loading (i.e. unterminated case), $\alpha = \beta h$. The integrals of the type indicated in equation (3) can be expressed in closed form for a sinusoidal current distribution. The result is given by,

$$H_{\phi}^{\infty}(\rho, 0) = -\frac{iI_0}{2\pi \sin \alpha} \left[\frac{e^{-i\rho r}}{\rho} \cos(\rho h - \alpha) - \frac{e^{-i\rho \rho} \cos \alpha - \frac{h}{\rho} e^{-i\rho r} \sin(\rho h - \alpha)}{\rho} \right] \quad (5)$$

where $r = (\rho^2 + h^2)^{1/2}$

Since the electric field $E_{\rho}(\rho, 0)$ is an unknown quantity it is necessary to make several simplifications at this stage. Since $|\gamma| \gg \beta$ an approximate boundary condition (see appendix II) is employed, expressed by

$$E_{\rho}(\rho, 0) \approx -\eta_s H_{\phi}(\rho, 0) \quad (6)$$

where $H_{\phi}(\rho, 0)$ is the tangential magnetic field for the imperfect ground system and η_s is the surface impedance of the air-ground interface. If ρ is greater than a , the radius of the ground screen, η_s can be replaced by η . If ρ is less than a , η_s is the intrinsic impedance η_a of the ground screen which is in parallel with the impedance η of the ground. In a previous investigation^{7,8} of this problem it was assumed that η_a was zero so the ground system was equivalent to a perfectly conducting disc of radius a . The limits of the integration in equation (2) are then from $\rho = a$ to $\rho = \infty$. A more general case is when η_a is comparable in magnitude to η , in which it follows that

$$\eta_s = \frac{\eta \eta_a}{\eta + \eta_a} \quad \text{for } 0 \leq \rho \leq a \quad (7)$$

where

$$\eta_a = \frac{c \eta_0 d}{\lambda} \log_e \frac{d}{2\pi c} \quad \text{and}$$

where d is the spacing between the radial conductors and c is the radius of the wire. The expression for η_a has been derived⁹ for a wire grid in free space where it was necessary to assume that $|\gamma_0 d| \ll 1$. Since the grid is lying on the ground plane, this restriction must be replaced by $|\gamma_e d| \ll 1$ where γ_e is the effective propagation constant for propagation along a thin wire in the interface and is given by¹⁰

$$\gamma_e = \left(\frac{\gamma_0^2 + \gamma^2}{2} \right)^{1/2}$$

If there are N radial conductors it can be seen that d can be replaced by $2\pi\rho/N$ since N is usually of the order of 100.

It is assumed also that $H_{\phi}(\rho, 0)$ is not very different from $H_{\phi}^{\infty}(\rho, 0)$ in the region of the ground plane where the losses are significant. This approximation has also been discussed previously^{7,8} and it certainly appears to be valid if $|\gamma| \gg \beta$.

The impedance increment ΔZ_r is then written in the following form

$$\Delta Z_r = \Delta Z + \Delta Z_a \quad (8)$$

where

$$\Delta Z \approx \eta \int_a^{\infty} [H_{\phi}^{\infty}(\rho, 0)]^2 2\pi\rho d\rho \quad (9)$$

and

$$\Delta Z_a \approx \int_0^a \frac{\eta_a \eta}{\eta + \eta_a} [H_{\phi}^{\infty}(\rho, 0)]^2 2\pi\rho d\rho \quad (10)$$

The first expression ΔZ corresponds to the self-impedance of the monopole over a perfectly conducting discoid, whereas the second expression ΔZ_a accounts for the finite surface

impedance of the radial conductor system.

It is instructive to consider ΔZ , first in some detail, since the integrations can be carried out and the result expressed in terms of the exponential integral defined by

$$Ei(-i\beta a) = - \int_a^{\infty} \frac{e^{-i\beta\rho}}{\rho} d\rho \quad (11)$$

The procedure was outlined previously^{7,8} where an expression for ΔZ was given for any value of the height h and top loading h' . The special case where the antenna is unterminated ($h' = 0$) is given by

$$\begin{aligned} \Delta Z = & \frac{\eta}{4\pi a \sin^2 \beta h} \left\{ Ei[-2i\beta(r_0+h)] e^{i2\beta h} \right. \\ & + Ei[-2i\beta(r_0-h)] e^{-2i\beta h} + 2 \cos^2 \beta h Ei(-2i\beta a) \\ & + 4 \cos \beta h \left\{ Ei(-i\beta h q_2) - Ei[-i\beta h(q_2+1)] e^{i\beta h} \right. \\ & \left. \left. - Ei[-i\beta h(q_2-1)] e^{-i\beta h} \right\} \right\} \quad (12) \end{aligned}$$

where

$$q_2 = \frac{1}{h} [a + (a^2 + h^2)^{1/2}] \quad \text{and} \quad r_0 = (a^2 + h^2)^{1/2}$$

This equation may be put in a suitable form for computation by employing the relation

$$Ei(-i\beta a) = Ci(\beta a) + i \left[\frac{\pi}{2} - Si(\beta a) \right] \quad (13)$$

where $Ci(\beta a)$ and $Si(\beta a)$ are the cosine and sine integrals respectively as defined and tabulated by Jahnke and Emde¹¹. The results of the calculations are presented in a most general form by plotting $4\pi\Delta Z/\eta$ as a function of a/h for various values of h/λ as shown in Figs. 2a and 2b. It can be seen that the magnitude of ΔZ increases without limit as a approaches zero. This formulation is not actually valid in this limiting case since one terminal of the generator would then be connected directly to the earth medium (or to a disc of vanishing radius) and would lead to an infinitely resistive path for the antenna base current.

A slightly more convenient way to illustrate these calculations is to plot $(\sigma/\epsilon)^{1/2} \Delta R$, where ΔR is the real part of ΔZ , as a function of a/h as shown in Fig. 3 where f is the frequency in c/s. This is only permissible if displacement currents in the ground are negligible (note if $\epsilon\omega \ll \sigma$, $\eta \approx (i\eta\omega/\sigma)^{1/2}$). It is interesting to note that ΔR actually assumes negative values under certain conditions. In this case, the input power to obtain a given current I_0 at the antenna terminals is actually less for an imperfect ground than for a perfect

ground. This fact can be reconciled by showing that the radiated power is actually reduced if $\Delta R < 0$. A physical explanation for the oscillating nature of the curve is that a wave is reflected from the discontinuity in the surface impedance at $\rho = a$. As the radius a increases the phase lag of the reflected wave will continually increase. This viewpoint is substantiated when it is noted that the period of the oscillations is nearly equal to twice the diameter of the ground screen.

When the ground screen of finite surface impedance is considered the integrations indicated by equation (10) must be evaluated. This appears to be a formidable task for the case when the antenna is of arbitrary length. However, if the antenna is a quarter-wave monopole, without top-loading, the integrations can be carried out fairly readily by graphical means. In this case

$$H_{\phi}^{\infty}(\rho, 0) = - \frac{i I_0}{2\pi\rho} e^{-i\beta(\rho^2 + \lambda^2/4)^{1/2}} \quad (14)$$

which is a special case of equation (5) with $h' = 0$ and $h = \lambda/4$, and hence the integral in equation (10) can then be expressed in the following form:

$$\Delta R_A = \int_0^A \frac{\beta q_2 \cos \left[\pi(1-4R) + \frac{3\pi}{4} - \tan^{-1} \frac{\beta q_2}{P} \right] dP}{\sqrt{2} \pi [P^2 + (\beta + q_2)^2]^{1/2} P} \quad (15)$$

and, similarly, ΔX_A with $\cos(\quad)$ replaced by $\sin(\quad)$. The real dimensionless quantities β , q_2 , A and R are defined by

$$\begin{aligned} \beta &= 120 \pi \delta / \sqrt{2} \quad \text{with} \quad \delta = (\epsilon_0 \omega / \sigma)^{1/2}, \\ q_2 &= \frac{240 \pi^2 P}{N} \log_e \frac{P}{NC} \quad \text{with} \quad C = c / \lambda_0, \\ A &= a / \lambda \quad \text{and} \quad R = \sqrt{P^2 + (1/4)^2} \end{aligned}$$

It has been assumed in writing equation (15) in this form that displacement currents in the ground are negligible, that is $(\epsilon\omega/\sigma) \ll 1$. Using the results of the numerical integration for ΔR_A , values of $\Delta R_T (= \Delta R + \Delta R_A)$ are plotted as a function of A in Figs. 4, 5 and 6 for various values of δ and N with $C = 0.1 \times 10^{-5}$, and in Fig. 7 for various values of C for $N = 100$ and $\delta = 0.1$. The value of δ can be readily obtained from Fig. 8 when the ground conductivity σ and the frequency in kilocycles are specified.

It is immediately evident that the oscillations in the curves for ΔR_T have been damped if N is finite. This can be expected

since there is a smaller change of the surface impedance at $\rho = a$ if the ground conductivity is reasonably high ($\delta < 0.1$). The limiting case where $N = \infty$ corresponds to the perfectly conducting disc discussed previously. It is quite apparent from these curves that a ground screen radius greater than about $1/3$ of a wavelength is wasteful. On the other hand it would be quite feasible to choose a large number of radials to reduce the resistance increment to a low value. Although, in practice, it is usual to employ No. 8 wire ($C = 0.5 \times 10^{-5}$ at 1 Mc) some improvement could be obtained by using larger wire diameters. From a theoretical standpoint, however, it would appear that, for a given total weight of wire, it is preferable to use a conductor of smaller diameter, say No. 22 wire ($C = 0.1 \times 10^{-5}$ at 1 Mc.) and to employ 150 or more radial conductors.

The Earth Currents

It is interesting to examine how the antenna base current is shared by the radial conductors and the ground itself. If the current flowing in the ground is denoted by I_e and the total current in the radial wires by I_w then the ratio I_e/I_w is equal to the ratio of the surface impedance of the grating of the wires composing the earth system to the surface impedance of the ground. Therefore it follows that

$$\frac{I_e}{I_w} = \frac{\eta_a}{\eta} \approx \frac{i q}{(1+i) p} \quad (16)$$

where p and q have been defined previously. Since the total current is given by, $I_t = I_e + I_w$ it follows that

$$\left| \frac{I_w}{I_t} \right| = \left| \frac{1}{1 + \eta_a/\eta} \right| = \left[\frac{1}{p^2 + (p+q)^2} \right]^{1/2} \quad (17)$$

It has been assumed here that displacement currents in the ground are negligible, (i.e. $\frac{\omega}{\sigma} \ll 1$)

Equation (17) is not a function of the height of the antenna and therefore it would apply also to top-loaded antennas as long as the circular symmetry is essentially retained. Employing this equation, curves are plotted in Figs. 9 and 10 to show the dependence of the current in the radial wires on the various parameters. The abscissae are the lengths of the radial wires in wavelengths measured from the base of the antenna. It is noted that if the radial wires are increased beyond a certain length nearly all the current flows in the ground. When the ratio $|I_w/I_t|$ is equal to $1/2$ the current in the ground is equal to the total current carried by the radial wires.

The Radiated Power

While the main subject of this paper has been to evaluate the input impedance of the antenna, it is also of some interest to know if the presence of the ground screen appreciably changes the radiated power. A simple and approximate analysis is now carried out which indicates

that this change is small.

The power dissipated in an elemental area of the ground is equal to the real part of $\eta |H_\phi|^2$. It is then evident that the change of power lost ΔP_L in the ground due to the presence of the ground screen is given by

$$\Delta P_L = \text{Re} \int_0^a \Delta \eta |H_\phi|^2 2\pi \rho d\rho \quad (18)$$

where $\Delta \eta$ is the difference between the surface impedance η_c of the radial conductor earth system and the surface impedance η of the ground. However, the change of input power at the antenna terminals is given by

$$\Delta P = \text{Re} \int_0^a \Delta \eta H_\phi^2 2\pi \rho d\rho \quad (19)$$

Now, since there is conservation of power, the change of the total radiated power ΔP_r beyond the edge of the earth system is equal to $\Delta P - \Delta P_L$ or

$$\Delta P_r = \text{Re} \int_0^a \Delta \eta [H_\phi^2 - |H_\phi|^2] 2\pi \rho d\rho \quad (20)$$

For a good ground screen, $\Delta \eta \approx -\eta$ and, if the antenna is a quarter wave monopole, ($h = \lambda/4$) the integration can be carried out in closed form to yield

$$\begin{aligned} \frac{\Delta P_r}{I_0^2} = \text{Real part of } \frac{\eta}{4\pi} \left\{ \left[2 \log \frac{4a}{\lambda} + 0.5773 \right. \right. \\ \left. \left. + \log \frac{\pi}{2} - \text{Ci}(\sqrt{\pi^2 + (2\rho a)^2} - \pi) + \text{Ci}(2\pi) \right. \right. \\ \left. \left. - \text{Ci}(\sqrt{\pi^2 + (2\rho a)^2} + \pi) \right] + i \left[\text{Si}(\sqrt{\pi^2 + (2\rho a)^2} - \pi) \right. \right. \\ \left. \left. - \text{Si}(2\pi) + \text{Si}(\sqrt{\pi^2 + (2\rho a)^2} + \pi) \right] \right\} \quad (21) \end{aligned}$$

If the ground screen radius, a , is small compared with a wavelength the change of power radiated is given approximately by

$$\frac{\Delta P_r}{I_0^2} \approx \text{Re} \cdot i \eta \frac{\beta^2 a^2}{2} \approx -\eta \frac{1}{2\sqrt{2}} (\beta a)^2 \quad (22)$$

This is usually a small quantity with the total radiated power so it is of minor significance at low radio frequencies.

The Radiation Pattern

The affect of the ground screen on the radiation pattern is also of some interest. For convenience, it is assumed here that the ground screen is equivalent to a thin, perfectly conducting, circular disc laid on a homogeneous ground. The magnetic field $H_\phi(\rho, z)$ in the air can be written as the sum of the field $H_\phi^\infty(\rho, z)$ for an infinite screen and a secondary field $H_\phi^a(\rho, z)$. In appendix I it is shown that

$$H_{\phi}^{\Delta} = \frac{\gamma_0}{\eta_0} \int_{\rho=a}^{\infty} \int_{\lambda=0}^{\infty} J_1(\lambda \rho) J_1(\lambda \rho') e^{-4.0 \gamma} u_0^{-1} E_{\rho}(\rho', \theta) \rho' d\rho' \lambda d\lambda \quad (23)$$

where $u_0 = (\lambda^2 + \gamma_0^2)^{1/2}$.

The change of the magnetic field ΔH_{ϕ} due to the presence of the screen is now given by

$$\Delta H_{\phi} = H_{\phi}^{\Delta} - (H_{\phi}^{\Delta})_{a=0}$$

$$= -\frac{\gamma_0}{\eta_0} \int_{\rho=0}^a \int_{\lambda=0}^{\infty} E_{\rho}(\rho', \theta) \rho' d\rho' J_1(\lambda \rho') J_1(\lambda \rho) e^{-4.0 \gamma} u_0^{-1} \lambda d\lambda \quad (24)$$

The integration with respect to λ can be carried out by the saddle point method of integration since $\beta \rho \gg 1$ in the radiation zone. The result is

$$\Delta H_{\phi} = \beta \frac{e^{-i\beta \bar{R}}}{\eta_0 \bar{R}} \int_0^a E_{\rho}(\rho', \theta) J_1(\beta \rho' \cos \theta) \rho' d\rho' \quad (25)$$

where $\bar{R} = \sqrt{\rho^2 + z^2}$, $\beta = 2\pi/\lambda_0 = -i\gamma_0$

and $\theta = \tan^{-1} z/\rho$

The approximate boundary condition, $E_{\rho}(\rho', 0) \approx -\eta H_{\phi}(\rho', 0)$ can now be employed so that

$$\frac{\Delta H_{\phi}}{H_{\phi}} = \frac{-\beta \eta}{\eta_0 \bar{R}} \int_0^a \frac{H_{\phi}(\rho', 0)}{H_{\phi}(\rho, z)} J_1(\beta \rho' \cos \theta) \rho' d\rho' \quad (26)$$

An approximate expression for ΔH_{ϕ} is now obtaining by replacing $H_{\phi}(\rho, z)$ on the right hand side of equation (26) by $H_{\phi}^{\infty}(\rho, z)$. If the antenna is a quarter wave monopole

$$H_{\phi}^{\infty}(\rho', 0) = -\frac{iI}{2\pi \rho'} e^{-i\beta \sqrt{\rho'^2 + (\lambda_0/4)^2}} \quad (27)$$

and for $\beta \rho \gg 1$

$$H_{\phi}^{\infty}(\rho, z) \approx -\frac{iI}{2\pi \rho} e^{-i\beta R} \cos\left(\frac{\pi}{2} \sin \theta\right) \quad (28)$$

so that

$$\frac{\Delta H_{\phi}}{H_{\phi}} = -\frac{\cos \theta}{\cos\left(\frac{\pi}{2} \sin \theta\right)} \frac{\eta}{\eta_0} \int_0^{1/4 \lambda_0} e^{-i\beta \sqrt{\rho^2 + \frac{\pi^2}{4}}} J_1(\rho \cos \theta) d\rho \quad (29)$$

The right hand of this equation is of the order of $|\eta/\eta_0|$ which is small compared with unity. For small screens where $a \ll \lambda_0$ the relation simplifies to

$$\frac{\Delta H_{\phi}}{H_{\phi}} \approx -\frac{\eta}{\eta_0} e^{-i\frac{\pi}{4}} \left(\frac{\pi a \cos \theta}{\lambda_0}\right)^2 \frac{1}{\cos\left(\frac{\pi}{2} \sin \theta\right)} \quad (30)$$

which is of second order magnitude.

Conclusion

The results of this analysis, while not exhaustive, are sufficiently developed to be useful in the design of vertical antennae with radial conductor ground systems. The work has shown that the input impedance of the type of antenna discussed is dependent mainly on the number and the length of the radial ground conductors and on the conductivity of the ground in which the wires are buried. The dependence of antenna impedance on ground wire size is shown to be very slight. It may be concluded, from this discussion, that a sensible design criterion for an optimum ground system is attained by a suitable choice of number and length of ground wire radials so that they will always carry an appreciable fraction of the total earth current.

Appendix I

Formulation of the Input Impedance

An expression is here formulated for the input impedance at the terminals of an antenna situated over a circular screen. The total flux F of the vector $\vec{E} \times \vec{H}$ over a surface surrounding the antenna is given by

$$F = \int_S \vec{E} \times \vec{H} \cdot \vec{n} dS \quad (31)$$

where \vec{n} is the unit outward vector normal to S . As is customary in other problems of this type, S is chosen to be a slender cylindrical surface of vanishing radius ρ concentric with the antenna so that

$$F = -\lim_{\rho \rightarrow 0} 2\pi \rho \int_0^h E_z H_{\phi} dz \quad (32)$$

It then follows that

$$Z = \lim_{\rho \rightarrow 0} \left[-\frac{1}{I_0^2} \int_0^h E_z I(z) dz \right] \quad (33)$$

It is now convenient to let, $E_z = E_z^{\infty} + E_z^{\Delta}$, where E_z^{∞} is the corresponding value of the electric field for a perfectly conducting ground plane and E_z^{Δ} is the change of the field to account for the finite conductivity in the soil and the ground system. The impedance increment ΔZ_T is then given by

$$\Delta Z_T = \left[-\frac{1}{I_0^2} \int_0^h E_z^A I(z) dz \right] \lim_{\rho \rightarrow \infty}$$

$$= \left[-\frac{\gamma_0}{\gamma_0 I_0^2} \int_0^h \left(\frac{\partial}{\partial \rho} \rho H_\phi^A \right) I(z) dz \right] \lim_{\rho \rightarrow \infty} \quad (34)$$

Since H_ϕ^A is a solution of the wave equation

$$H_\phi^A(\rho, z) = \int_0^\infty J_1(\lambda \rho) e^{-\gamma_0 z} f(\lambda) d\lambda \quad (35)$$

for $z \geq 0$, where $u_0 = (\lambda^2 + \gamma_0^2)^{1/2}$.
From Maxwell's equations it is seen that

$$E_\rho(\rho, z) = \frac{\gamma_0}{\gamma_0} \int_0^\infty J_1(\lambda \rho) f(\lambda) u_0 \lambda d\lambda \quad (36)$$

and by applying the Fourier-Bessel theorem it follows that

$$f(\lambda) = \frac{\gamma_0}{\gamma_0 u_0} \int_0^\infty J_1(\lambda \rho') E_\rho(\rho', z) \rho' d\rho' \quad (37)$$

This equation for $f(\lambda)$ can then be substituted back into equation (35) to obtain an expression for $H_\phi^A(\rho, z)$ in terms of $E_\rho(\rho, z)$. It is also noted that $J_1(\lambda \rho)$ can be replaced by $\lambda \rho / 2$ as ρ tends to zero so that

$$\lim_{\rho \rightarrow 0} H_\phi^A(\rho, z) = \frac{\gamma_0}{2\gamma_0} \int_0^\infty \int_0^\infty \rho J_1(\lambda \rho') e^{-\gamma_0 z} u_0^{-1} \lambda^2 d\lambda E_\rho(\rho', z) \rho' d\rho' \quad (38)$$

and introducing Sommerfeld's Integral

$$\int_0^\infty e^{-\gamma_0 z} u_0^{-1} \lambda J_1(\lambda \rho') d\lambda = (\gamma_0^2 + \rho'^2)^{-1/2} e^{-\gamma_0 (\gamma_0^2 + \rho'^2)^{1/2} z} \quad (39)$$

the integration in equation (38) with respect to λ can now be carried out to give

$$\lim_{\rho \rightarrow 0} H_\phi^A(\rho, z) = -\frac{\gamma_0 \rho}{2\gamma_0} \int_0^\infty \frac{e^{-\gamma_0 (\gamma_0^2 + \rho'^2)^{1/2} z}}{\partial \rho' (\gamma_0^2 + \rho'^2)^{1/2}} E_\rho(\rho', z) \rho' d\rho' \quad (40)$$

Inserting this expression into equation (34) leads directly to equation (2) for the input impedance increment.

Appendix II

The Approximate Boundary Condition

The magnetic field in the ground outside the screen is a solution of the wave equation

$$(\Delta - \gamma^2) H_\phi(\rho, z) = 0$$

and therefore

$$H_\phi(\rho, z) = -\frac{\pi}{\gamma} \int_0^\infty u J_1(\lambda \rho) e^{-\gamma z} f(\lambda) d\lambda \quad (41)$$

for $z \geq 0$ and where $u = (\lambda^2 + \gamma^2)^{1/2}$.
The electric field is given by

$$E_\rho = -\frac{\pi}{\gamma} \int_0^\infty u J_1(\lambda \rho) e^{-\gamma z} f(\lambda) d\lambda \quad (42)$$

The binomial expansion of u is of the form

$$u = \gamma \left(1 + \frac{\lambda^2}{2\gamma^2} + \dots \right)$$

so that

$$E_\rho = -\frac{\pi}{\gamma} \left\{ \int_0^\infty \gamma J_1(\lambda \rho) e^{-\gamma z} f(\lambda) \lambda d\lambda + \int_0^\infty \frac{\lambda^2}{2\gamma} J_1(\lambda \rho) e^{-\gamma z} f(\lambda) \lambda^3 d\lambda + \dots \right\} \quad (43)$$

Now the differential equation for J_1 is given by

$$\left(\frac{\partial}{\partial \rho} \rho \frac{\partial}{\partial \rho} \rho - \lambda^2 \right) J_1(\lambda \rho) = 0 \quad (44)$$

so it readily follows that

$$E_\rho = -\gamma H_\phi - \frac{\pi}{2\gamma^2} \frac{\partial}{\partial \rho} \rho \frac{\partial}{\partial \rho} \rho H_\phi \pm \text{terms in } \gamma^{-4} \quad (45)$$

The second and remaining terms are negligible if the propagation constant of the ground is sufficiently large and if H_ϕ is not varying too rapidly with ρ . That is, ρH_ϕ should not change appreciably in a distance equal to $|\gamma^{-1}|$ in the radial direction.

Acknowledgement

We would like to express our appreciation to Mr. G.D. Monteath of the British Broadcasting Corporation and to Dr. F.R. Abbott, of the U.S. Navy Electronics Laboratory who communicated to us unpublished reports of their work. We are also indebted to Dr. W.J. Surtees and Mr. D.A. Trumpler who assisted us in the earlier phase of the investigation and to Mr. P.A. Field and Dr. W. Petrie who read the manuscript.

References

- (1) G.H. Brown, "General Considerations of Tower Antennas for Broadcast Use" Proc. of the I.R.F. 23, 311, 1935.
- (2) G.H. Brown, R.F. Lewis, J. Epstein, "Ground Systems as a Factor in Antenna Efficiency" Proc. of the I.R.F. 25, 753, 1937.
- (3) F.R. Abbott, "Design of Buried R.F. Ground Systems" Proc. of the I.R.F., 40, 846, 1952.

(4) A. Leitner and R.D. Spence, "Effect of a Circular Ground Plane on Antenna Radiation" Jour. App. Phys. 21, 1001, 1950.

(5) J.E. Storer, "Impedance of an Antenna over a Large Circular Screen" Jour. App. Phys. 22, 1058, 1951.

(6) G.D. Monteath, "Application of the Compensation Theorem to Certain Radiation and Propagation Problems" Proc. of the Inst. Elect. Engrs. 98, Pt. IV, 23, 1951.

(7) W.J. Surtees and J.R. Wait, "Impedance of a Top-Loaded Antenna of Arbitrary Length Over a Circular Grounded Screen" Jour. App. Phys. (in press).

(8) J.R. Wait, "Impedance of an Antenna Over a Circular Ground System" Radio Physics Lab. Report 19-0-4, 1953. (Can. Defence Research Board Project D48-95-55-07.)

(9) G.G. MacFarlane, "Surface Impedance of an Infinite Wire Grid at Oblique angles of Incidence" Jour. Inst. Elect. Engrs. 93, Pt. III A, 1523, 1946.

(10) B.L. Coleman, "Propagation of Electromagnetic Disturbances along a Thin Wire in a Horizontally Stratified Medium" Phil. Mag. 41, 276, 1950.

(11) E. Jahnke and F. Emde, "Tables of Functions" Dover Publications, New York, 1945.

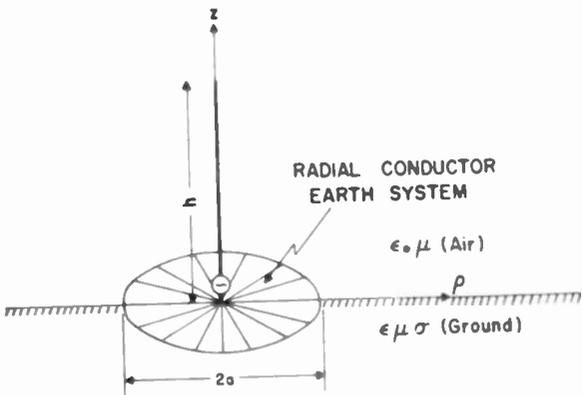
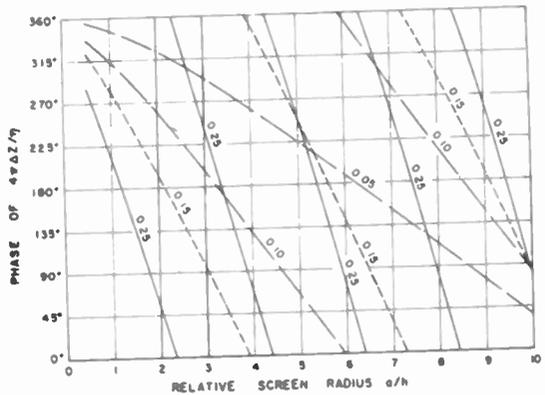
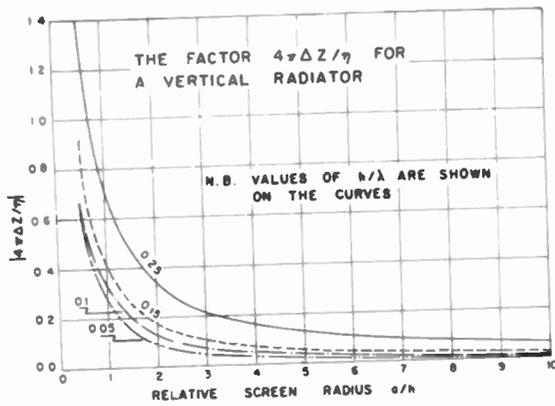


Fig. 1



2(b)



2(a)

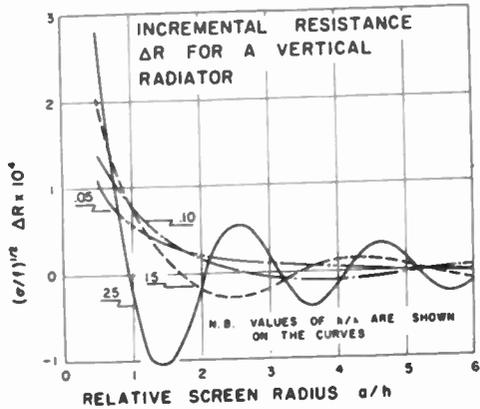


Fig. 3

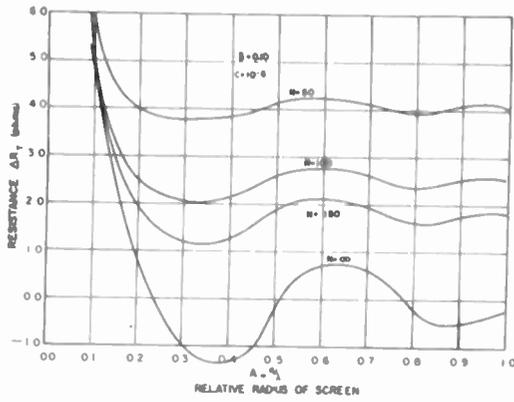


Fig. 4

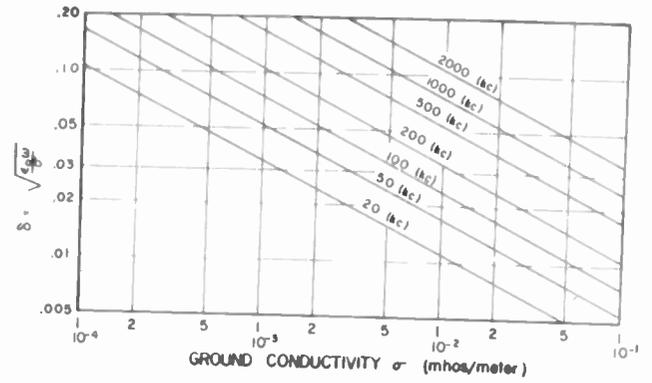


Fig. 8

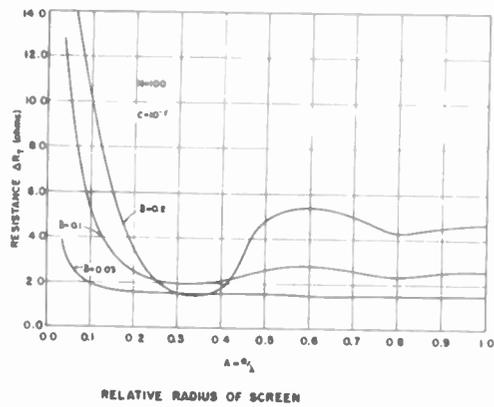


Fig. 5

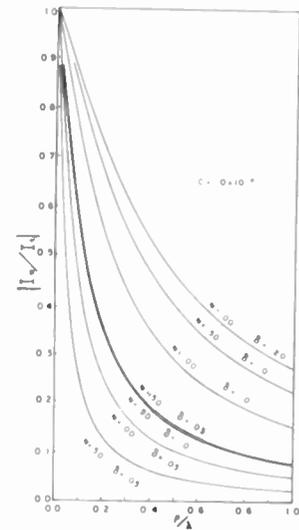


Fig. 9

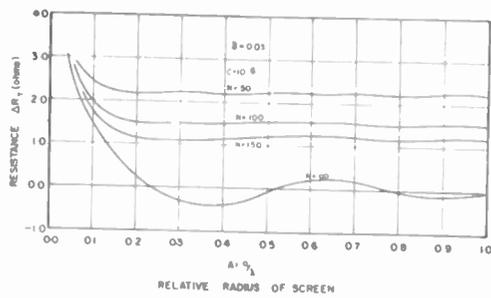


Fig. 6

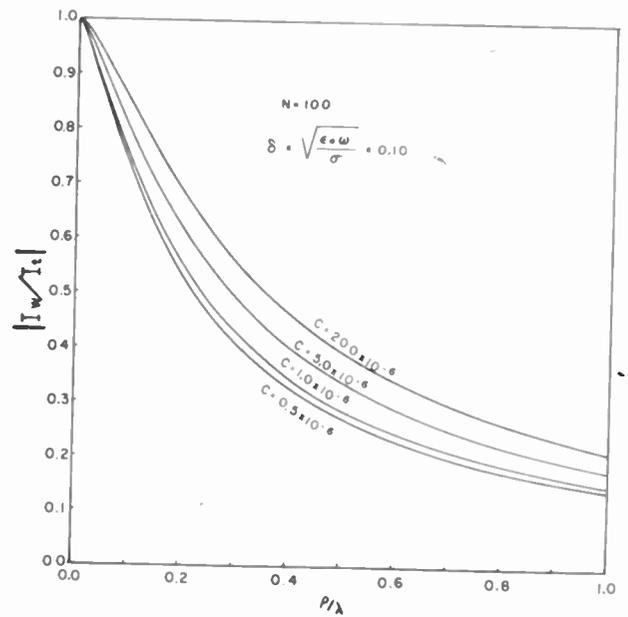


Fig. 10

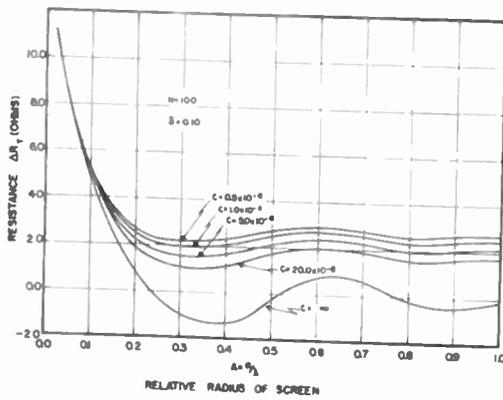


Fig. 7

SOME INFORMATION THEORY ASPECTS OF PROPAGATION
THROUGH TIME VARYING MEDIA

by

Joseph Feinstein
National Bureau of Standards
Washington 25, D. C.

Abstract

The channel capacity of a communications system which utilizes wave propagation through a time varying medium such as the ionosphere or troposphere is evaluated in terms of the statistical properties of the medium and of the noise. The signal fading in such a system reduces the capacity.

Information theory concepts are broadened to include the possibility of multiple reception at spaced receiving sites, and the consequent increase in theoretical channel capacity is computed as a function of the number of such sites. Current practices in the use of diversity reception and directional antennas are examined in the light of these results.

Introduction

The conventional treatment of the capacity of a communications channel given by information theory postulates steady signals contaminated by noise during the transmission process. In a wide class of systems, however, there is another effect which tends to degrade information: the signal fading associated with transmission through a time-varying medium. This is present, for example, in all systems utilizing space propagation of waves through the ionosphere or troposphere.

To compensate, as it were, for this effect the relatively large region of space occupied by the signal field when a propagation link is utilized in a communications system introduces additional degrees of freedom into the reception process. While semi empirical methods of taking advantage of these degrees of freedom, known as diversity, have long been known in the art, no standards are available against which their performance may be judged. It is the general purpose of this paper to present a fundamental approach to the problem of transmission through time varying media, utilizing the formalism of information theory. The problems treated have been limited to calculations of ideal channel capacity, rather than to a consideration of the effects of fading upon specific modulation systems. The results serve, however, to indicate the fundamental limitations present in this type of communications link.

In Part I channel capacity is evaluated for the conventional single point reception of a fading signal in the presence of noise; several types of fading distributions and fluctuation characteristics are considered. The generalization to multiple receiving elements is developed in Part II, and applied to an examination of directional and diversity reception.

Part I. Single Point Reception

1. Nature of the Received Signal

The physical distinction between fading and noise as mechanisms of signal degradation is evident from the fact that noise is received even in the absence of a transmitted signal while fading effects act to modulate the signal. The resultant amplitude received, y , when signal x is sent is of the form:

$$y = AM + N \quad (1)$$

where M is the fading modulation factor, and N is the noise. The statistical characteristics of the random variable M are, of course, related to the fluctuations of the transmission medium. The conditions which the probability distribution $p(M)$ must satisfy are:

$$\int_{-\infty}^{\infty} p(M) dM = 1 \quad (2a)$$

$$\int_{-\infty}^{\infty} M^2 p(M) dM = 1 \quad (2b)$$

The first follows from the definition of probability and the second from the requirement of constant average power. If we take $p(M)$ to be of normal form:

$$p(M) = \frac{1}{m\sqrt{2\pi}} e^{-\frac{(M-d)^2}{2m^2}} \quad (3)$$

$$\text{then to satisfy (2b), } d = \sqrt{1-m^2} \quad (4)$$

For $m \ll 1$, $d \approx 1$, and we obtain a Gaussian distribution with a mean of unity; for $m = 1$,

$$p(M) = \frac{1}{\sqrt{2\pi}} e^{-M^2/2} \quad (5)$$

The first of these agrees with experimental observations on ionospheric signals under conditions of shallow fading¹, while the second occurs under conditions of Rayleigh fading. In the intermediate

region the normal form no longer holds, but is replaced by an expression involving Bessel functions¹. We shall restrict our applications to the two limiting forms above since they serve to illustrate the types of behavior obtained and possess the merit of mathematical simplicity.

Consideration of the remaining property of M which we shall require, its rate of fluctuation with respect to the sequence of signals which constitute a message, will be deferred until section 5.

We make the conventional Gaussian white noise assumptions for N , so that:

$$p_N(N) = \frac{1}{\sigma_N \sqrt{2\pi}} e^{-\frac{N^2}{2\sigma_N^2}} \quad (6)$$

where σ_N^2 is the average noise power.

2. Channel Capacity Formulation Uncorrelated Fading

For a message consisting of n discrete signals we may write the expression of Shannon² for rate of information transmission in the form:

$$C' = \int \dots \int dA_1 \dots dA_n dy_1 \dots dy_n \times \\ \times P(A_1 \dots A_n | y_1 \dots y_n) \log \frac{P(A_1 \dots A_n | y_1 \dots y_n)}{P(y_1 \dots y_n)} \quad (7)$$

where conventional notation has been employed for the joint and conditional distributions. To convey maximum information the successive A 's should be independent of each other. Since the noise contributions to each signal are also independent we may write the joint probability distribution of all the A 's and y 's as a product of individual distributions for each pair (A, y) provided the fading factor M is uncorrelated on successive signals. Under these conditions (7) reduces to:

$$C' = n \iint dA dy P(A, y) \log \frac{P(A|y)}{P(y)} \quad (8)$$

and making use of the relations:

$$P(A, y) = p_A(A) P(A|y) \quad (9a)$$

$$P(y) = \int p_A(A) dA \quad (9b)$$

$$P(A|y) = \int p_M(M) p_N(y-AM) dM \quad (10)$$

we may proceed to evaluate C' once we have selected specific forms for p_A and p_M . To obtain the channel capacity the distribution function of the transmitted signal A should be such as to maximize C' . For the case of a non-fading signal in noise it has been established² that a Gaussian p_A is

optimum. It may be shown that this choice continues to yield maximum transmission rate in the case of shallow (displaced Gaussian) fading*. For Rayleigh fading, the derivation of the maximal form of p_A , subject to the condition of constant variance, is given in Appendix A; since this distribution is physically unrealizable we shall assume a Gaussian form for p_A here as well, so as to have a basis for inter-comparison of our results. We may not, however, call C' the channel capacity in this case. We now proceed to evaluate C' for the two cases.

3. Evaluation of Capacity - Uncorrelated Fading

For Gaussian fading, (10) becomes:

$$p(A|y) = \int_{-\infty}^{\infty} \frac{dM}{\sigma_M \sqrt{2\pi}} e^{-\frac{(M-1)^2}{2\sigma_M^2}} \frac{e^{-\frac{(y-AM)^2}{2\sigma_N^2}}}{\sigma_N \sqrt{2\pi}} \quad (11)$$

$$= \frac{1}{\sqrt{2\pi(\sigma_N^2 + A^2\sigma_M^2)}} e^{-\frac{(y-A)^2}{2(\sigma_N^2 + A^2\sigma_M^2)}} \quad (11a)$$

Substituting in (9):

$$P(y) = \int_{-\infty}^{\infty} \frac{dA}{\sigma_A \sqrt{2\pi}} e^{-\frac{A^2}{2\sigma_A^2}} e^{-\frac{(y-A)^2}{2(\sigma_N^2 + A^2\sigma_M^2)}} \quad (12)$$

In view of the fact that $m \ll 1$ for this formulation to apply we assume $\sigma_N^2 \gg m^2 \sigma_A^2$, i.e., the effect of noise predominates over that of fading. There an approximate evaluation of (12) leads to:

$$P(y) \approx \frac{1}{\sqrt{2\pi(\sigma_N^2 + \sigma_A^2[m^2\sigma_A^2])}} e^{-\frac{y^2}{2(\sigma_N^2 + \sigma_A^2[m^2\sigma_A^2])}} \quad (12a)$$

Inserting these quantities into (8), C' becomes, to the same order of approximation:

$$C' = \frac{n}{2} \log \left(1 + \frac{\sigma_A^2}{\sigma_N^2 + m^2 \sigma_A^2} \right) \quad (13)$$

The form of this result enables a simple interpretation to be drawn of the effect of shallow fading, viz. that the effective noise power is increased by m^2 x the average signal power, the total effective signal power remaining unchanged.

For Rayleigh fading, (10) becomes:

$$p(A|y) = \int_{-\infty}^{\infty} \frac{dM}{\sigma_M \sqrt{2\pi}} e^{-\frac{M^2}{2}} \frac{e^{-\frac{(y-AM)^2}{2\sigma_N^2}}}{\sigma_N \sqrt{2\pi}} \quad (13)$$

$$= \frac{1}{\sqrt{2\pi(\sigma_N^2 + \sigma_M^2)}} e^{-\frac{y^2}{2(\sigma_N^2 + \sigma_M^2)}} \quad (14a)$$

Substituting in (9):

* The form of eq. (12a) admits of this interpretation.

$$P(y) = \int_{-\infty}^{\infty} \frac{dA}{\sigma_A \sqrt{2\pi}} e^{-\frac{A^2}{2\sigma_A^2}} \frac{e^{-\frac{y^2}{2(\sigma_A^2 + \sigma_N^2)}}}{\sqrt{2\pi(\sigma_A^2 + \sigma_N^2)}} \quad (15)$$

If one were to assume the noise power to be much greater than the signal power, the same type approximation as was made in (12) could be introduced here, and would lead to the result $C' = 0$. However, since such an approximation is valid only for signal to noise ratios much less than unity, it is of dubious interest. Instead we make the more physically interesting assumption that the noise power is much less than the signal power, so that we may set $\sigma_N^2 = 0$, and solve for the effect of fading alone in the absence of noise. Eq. (15) then yields:

$$P(y) = \frac{1}{\pi\sigma_A} K_0\left(\left|\frac{y}{\sigma_A}\right|\right) \quad (15a)$$

where K_0 is the modified Hankel function. Evaluation of (8) then yields a pure numeric*, independent of the signal power as one might expect: $C'/n = 0.500$. One may express the effect of fading upon the rate of information transmission in terms of the signal to noise ratio which yields the same value of capacity by equating this result to $\frac{1}{2} \log(1 + \sigma_A^2/\sigma_N^2)$, noting that natural base logarithms are to be employed. This procedure gives

$\sigma_A/\sigma_N = 1.72$ as the equivalent signal to noise power ratio.

The assumption of uncorrelated M values for neighboring signals made in this section, represents the worst possible situation from the point of view of degradation of information. We now proceed to consider the case where M possesses some degree of correlation from point to point, and then to inquire into the factors which determine this correlation.

4. Evaluation of Capacity Correlated Fading

In the general case we may write for the conditional probability distribution of the A 's and y 's:

$$P(A_1, \dots, A_n | y_1, \dots, y_n) = \int dM_1 p_{M_1}(M_1) p_N(y_1 - A_1 M_1) \times \\ \times \int dM_2 p_{M_2}(M_2) p_N(y_2 - A_2 M_2) \dots \times \\ \times \int dM_n p_{M_n}(M_n) p_N(y_n - A_n M_n) \quad (16)$$

where the p_{M_k} express the dependence of a given M_k upon the preceding $(k-1)M$'s. It seems reasonable

* The $\int_0^{\infty} K_0(x) \log K_0(x) dx$ which arises here was evaluated numerically as -0.223 .

to take a multivariate normal form for the conditional distribution of M . In view of the complexities attending such a choice we make the further assumption that M is a first order Markoff process, thus restricting ourselves to bivariate distributions. For Gaussian fading then³:

$$p_{M_k}(M_1, \dots, M_{k-1} | M_k) = p_{M_k}(M_{k-1} | M_k) = \frac{e^{-\frac{[M_k - 1 - \rho(M_{k-1} - 1)]^2}{2\sigma^2(1-\rho^2)}}}{\sqrt{2\pi(1-\rho^2)}} \quad (17)$$

where ρ is the correlation coefficient between successive M 's. The evaluation of (16) with the form (17) has been relegated to Appendix B. For $\rho \ll 1$ one obtains a reduction in fading variance m^2 , by the factor $(1-\rho^2)$. Difficulties of integration have prevented evaluation for other ranges.

For $\rho \approx 1$, a possible choice for p_{M_k} may be obtained by taking M to be perfectly correlated in groups of q points, and uncorrelated between these groups. Then employing a delta function representation of p_{M_k} :

$$P(A_1, \dots, A_n | y_1, \dots, y_n) = P(A_1, \dots, A_q | y_1, \dots, y_q) \times \\ \times P(A_{q+1}, \dots, A_{2q} | y_{q+1}, \dots, y_{2q}) \dots \\ P(A_1, \dots, A_q | y_1, \dots, y_q) = \int dM_1 p_{M_1}(M_1) p_N(y_1 - A_1 M_1) \times \\ \times \dots \times \int dM_q \delta(M_q - M_1) p_N(y_q - A_q M_q) \quad (18)$$

For Gaussian fading an approximate evaluation of (19) leads for large q to a form similar to that obtained for uncorrelated fading but with an effective noise power given by (see Appendix C):

$$\sigma_N'^2 = \sigma_N^2 + \frac{m^2 \sigma_A^2}{1 + m^2 (q-1) \sigma_A^2 / \sigma_N^2} \quad (20)$$

The channel capacity is then found from (13) by replacing the fading depth factor m^2 , by

$\frac{m^2}{1 + m^2 (q-1) \sigma_A^2 / \sigma_N^2}$ The value to be chosen for q is related in a general way to the rate at which the correlation curve falls off with signal element separation.

For the case of correlated Rayleigh fading, if the effect of the fading is reduced below that

of the noise as a result of the correlation

$[\sigma_n^2 \gg \sigma_a^2 (1-\rho^2)]$ then the above expressions are applicable, with $m = 1$. If, however, the fading still remains the dominant factor in degrading information we encounter some difficulty. For if noise is neglected then the δ function formulation of the preceding paragraph leads to an infinite channel capacity. This is not surprising in view of the infinitely fine gradations which may be distinguished in the ratio of two completely unperturbed signals.

For a bivariate normal distribution of P_{i_k} , and neglecting noise, correlated Rayleigh fading leads to:

$$P(R_1, \dots, R_n / y_1, \dots, y_n) = \frac{e^{-y_1^2/2R_1^2}}{R_1 \sqrt{2\pi}} \times \dots \times \frac{e^{-(y_n^2 - \rho y_1 y_n / R_1)^2}}{R_n \sqrt{2\pi(1-\rho^2)}} \quad (21)$$

Difficulties of integration prevent us from obtaining $P(y_1, y_2, \dots, y_n)$ explicitly. For $\rho \ll 1$, however, it is reasonable to assume that this y -distribution is not materially affected by the correlation. The increase in transmission rate over the uncorrelated case then takes the form:

$$\Delta C' = \frac{n-1}{2} \log \frac{1}{1-\rho^2} \quad (22)$$

5. Nature of Fading Correlation

In the preceding section we have introduced the symbol ρ to denote the correlation of the fading variable M between adjacent signal elements of a message. Now it is well known² that a message of duration T , and occupying bandwidth W , requires $2TW$ values, or signal elements for its specification. Consequently the signals corresponding to such a message are spaced at time intervals $\frac{1}{2W}$. Then if M does not vary with frequency over the band W , ρ is simply the autocorrelation function of M , evaluated for a time displacement $\frac{1}{2W}$. In the more general case, however, M may vary with frequency as well as with time; we therefore define:

$$C_0(\omega, \tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T M(\omega_0, t) M(\omega_0 + \omega, t + \tau) dt \quad (23)$$

One may obtain a crude value of ρ when selective fading is present by dividing the band W into segments of width p_0 , where p_0 is the frequency interval over which the fading remains fairly well correlated. Then in each of the subdivisions the sampling rate is $2p_0$ signals/second, so that ρ is again the autocorrelation function of M , but now evaluated for a time displacement $\frac{1}{2p_0}$. Since this time interval is greater than $\frac{1}{2W}$,

ρ will decrease, as expected. A more accurate approach has as its point of departure, the Fourier expression for a signal as the sum of its frequency components:

$$s(t) = \sum_k a_k \cos(\omega_k t + \phi_k) \quad (24)$$

where the frequencies are spaced at intervals $\frac{1}{T}$.

For subdivisions of bandwidth p , ρ is given by:

$$\rho_p = \lim_{T \rightarrow \infty} \frac{1}{T} \int dt \frac{s(t) M(\omega, t) s(t') M(\omega, t')}{s(t) s(t')} \quad (25)$$

where $t' = t + \frac{1}{2p}$; inserting (24):

$$\rho_p = \frac{\int_0^T dt \sum_{k, l} M(\omega_k, t) M(\omega_l, t') a_k \cos(\omega_k t + \phi_k)}{\sum_{k, l} a_k \cos(\omega_k t + \phi_k)} \times \frac{a_l \cos(\omega_l t' + \phi_l)}{a_l \cos(\omega_l t + \phi_l)} \quad (25a)$$

Since the ϕ_k are random and independent for transmission of maximum information, the sum of cross product terms arising from $\sum_{k \neq l} \dots$ may be taken as zero. Then:

$$\rho_p = \frac{\sum_k \cos(\omega_k \tau, \tau = \frac{1}{2p}) \cos \frac{\omega_k}{2p}}{\sum_k \cos \frac{\omega_k}{2p}} \quad (26)$$

Replacing the summation by an integration as we pass to the limit $T = \infty$, and maximizing with respect to p :

$$\rho = \frac{M_{\text{max}}}{(F)} \frac{\int_{-P/2}^{P/2} c_o(\omega, \tau = \frac{1}{2P}) \cos \frac{\omega}{2P} d\omega}{\int_{-P/2}^{P/2} \cos \frac{\omega}{2P} d\omega} \quad (26a)$$

As an example, we take

$$c_o(\omega, \tau) = e^{-\frac{\tau^2}{2\tau_o^2}} e^{-\frac{|\omega|}{P_o}} \quad (27)$$

and let $p = xP_o$; then

$$\rho(x) = \frac{e^{-\left(\frac{1}{2P_o\tau_o}\right)^2 x^2}}{1+4x^2} \left[\frac{2x}{\sin \frac{1}{4}} + e^{-\frac{x}{2}} - 2xe^{-\frac{x}{2}} \frac{1}{\cos \frac{1}{4}} \right] \quad (28)$$

The optimum fraction x is plotted as a function of the $2P_o\tau_o$ product in Figure 1a, and the corresponding maximum ρ is shown in Figure 1b. Similar curves are given for an M correlation of the form

$$c_o(\omega, \tau) = e^{-|\omega/\omega_o|} e^{-|\tau/\tau_o|} \quad , \text{ to illustrate}$$

the dependence of the results on the specific form assumed. These curves make it clear that there is an optimum manner of subdivision of a band $W, >$

P_o , for a given set of fading characteristics. It is also to be noted that as $2P_o\tau_o$ falls below unity, $\rho \rightarrow 0$, even for optimum subdivision. Consequently the results of section 1, which assumed no correlation between adjacent points, are valid for $2P_o\tau_o$, somewhat less than one.

Such small values of $P_o\tau_o$ raise certain questions regarding the operational significance of the definition (23), wherein an envelope comparison of two continuous signals, initially transmitted on frequencies W_o and W_o+W , is visualized. This requires that the bandwidth of the receivers be sufficient to admit the modulation frequencies produced by the fluctuation of M . But for a temporal correlation interval τ_o , modulation frequencies of order $\frac{1}{\tau_o}$ are present. Consequently separation

of the two signals becomes uncertain when their frequency difference falls to $\frac{1}{\tau_o}$; thus it becomes

impossible to distinguish frequency correlation intervals $P_o < \frac{1}{\tau_o}$, if both signals are received

at the same point in space. Conceptually, however, it is possible to visualize an ensemble of identically fluctuating systems so that the signals are separated physically. No limit would then exist on the detectable frequency separation. In any event the exact values of P_o and τ_o are no longer of importance once their product has fallen below unity. It is interesting to note that $2P_o\tau_o \approx 1$ has been identified by Gabor⁴ with a quantum, or elementary cell of information.

6. Finite Transmission Time Viewpoint

To realize channel capacity, infinite transmission time is required. It seems reasonable to view fading as destroying the channel coherence periodically, and therefore as acting to shorten the effective time of transmission of a message. One would expect this viewpoint to be applicable only when fading is not the dominant factor limiting channel capacity, and to be related to the delta function correlation assumption in section 4.

The following treatment takes as its point of departure the expression given by Shannon⁵ for the probability of uncertainty of each bit of a received message when a finite transmission time is employed:

$$p = 2^{-T\eta} \quad (29)$$

where $\eta = C-R$, C being the channel capacity and R the actual rate of information transmission. The equivocation produced by this uncertainty,

$$H = -\sum p_n \log p_n \quad (30)$$

where the summation is over all the possible states of the system, represents the diminution in transmission rate and is therefore equal to η , when expressed on a per second basis:

$$\eta = -2W [p \log p + (1-p) \log (1-p)] \quad (31)$$

Inserting (29), and setting $\eta = 2Wg$, so that g is reckoned on a bit basis:

$$g = [(2TW)g 2^{-(2TW)g} - \{1 - 2^{-(2TW)g}\} \log \{1 - 2^{-(2TW)g}\}] \quad (32)$$

We may now insert the correlation interval $\tau_o P_o$ for TW , and solve this transcendental equation numerically for g as a function of $2\tau_o P_o$.

The results have been plotted in Figure 2 in terms of the db reduction in signal to noise ratio below that of an ideal channel. Equation (20), with $m=1$ and $q = 2\tau_o P_o$ has been similarly plotted for comparison. The nature of the reasoning leading to (29) is such that one would expect the results to

be meaningful only for fairly large τ_{OP_0} . Comparison of these curves indicates that the two effects are only qualitatively similar. The reverse dependence upon initial signal to noise ratio is especially noteworthy.

Part II. Multiple Element Reception

1. Space Characteristics of the Signal Field

We turn now to systems of information transmission which utilize relatively unguided wave propagation through unconfined media. To evaluate the performance of such systems requires a generalization of the treatment in part I. To this end it is desirable to introduce the notion of angular power spectrum of received energy, and the related concept of a ground correlation pattern⁶. Briefly, the medium inhomogeneities whose fluctuations give rise to fading also tend to destroy the phase coherence which is responsible for confining the received energy to a cone of angle determined by the Fresnel zone geometry for the system. As a consequence power may be received over a very much wider, and fluctuating angle when such inhomogeneities are present; the angular power spectrum $P(\theta)$ gives the time average power received at angle θ , measured from the specular direction.

Because of the finite distance over which the fluctuations within the medium are correlated, the correlation between signals received at any two points near the ground will fall as the points are further separated. The correlation of the signals as a function of their spacing, known as the ground correlation pattern, has been shown to be⁶ the Fourier transform of the angular power spectrum. For an antenna with pattern factor $F(\theta)$ set noise N_s , and antenna noise field N_a /unit solid angle (assumed isotropic) the signal to noise ratio is given by:

$$\frac{S}{N} = \frac{\int P(\theta) F(\theta) d\Omega}{N_s + N_a \int F(\theta) d\Omega} \quad (33)$$

For antenna noise predominant over set noise this has a maximum value P_m/N_a , where P_m is the maximum power density in the spectrum. In practice of course there is not much point in obtaining greater directionality than that which corresponds to the angle between, say, the half power points of $P(\theta)$. The reason little further gain is obtainable is directly related to the fact that a more extensive structure will occupy a region greater than that over which the signal remains correlated, so that phase incoherence sets in. A new effect then arises, termed diversity reception.

We visualize the simultaneous reception of the transmitted signal on a set of elements, each of which may be a directional antenna (for optimum performance), the spacing between elements being greater than the correlation distance at the receiving site. Consequently the fading of the signals received on the individual elements is uncorrelated. Since the correlation distance of the noise field is generally much smaller than that of the signal

field, the noise component of the signals is also assumed uncorrelated. In this connection it should be noted that the use of a directional antenna may alter the fading characteristics of a received signal. A well known example of such behavior is found in the so called MUSA antenna, employed to select a single ionospheric mode so as to reduce the selective fading which arises from mode interference. This situation corresponds to the existence of several maxima in $P(\theta)$, one of these being selected by $F(\theta)$.

In the next section we deduce the increase in channel capacity which results from multiple reception, with our two types of fading, and with various resultant signal selection laws.

2. Channel Capacity With Diversity Reception

Give a set of ℓ independent signals: $y_1, y_2 \dots y_\ell$ corresponding to a transmitted signal A , we are faced with the problem of selecting a method of combining the individual signals. It would be desirable to choose this law of combination so as to minimize our error of determination of A . This problem is well known in statistics; for minimization of the variance of the distribution of the combined signal \bar{A} , the method of maximum likelihood is appropriate⁷. This method yields for a Gaussian distribution (displaced from the true value A) of the individual signals:

$$\bar{A} = \frac{1}{\ell} \sum_{k=1}^{\ell} y_k \quad (34)$$

i.e., the arithmetic mean of the instantaneous signals. For the case of Rayleigh fading accompanied by noise, as specified by eq. (14a), one finds:

$$\overline{A^2} = \frac{1}{\ell} \sum_{k=1}^{\ell} y_k^2 - \sigma_N^2 \quad (35)$$

a root mean square, or demodulated type of combination. We shall examine diversity systems obeying these optimum combinational laws, as well as certain other which are of interest.

We consider first the case in which each received message element is taken as the arithmetic mean \bar{y}_ℓ , of the ℓ individual signals $y_1, y_2 \dots y_\ell$ corresponding to a single transmitted signal A . Then for a single message element:

$$P(A/\bar{y}_\ell) = \int \dots \int dy_1 dy_2 \dots dy_{\ell-1} \times p(A/y_1) \cdot p(A/y_2) \dots p(A/\bar{y}_\ell - [y_1 + y_2 + \dots + y_{\ell-1}]) \quad (36)$$

For Gaussian fading this yields:

$$P(A/\bar{y}_\ell) = \left(\frac{2\pi(\sigma_N^2 + A^2)}{\ell} \right)^{-1/2} \exp \left\{ -\frac{\ell(\bar{y}_\ell - A)^2}{2(\sigma_N^2 + A^2)} \right\} \quad (37)$$

where (11a) has been employed for each $p(A/y)$. The reduction in variance by $\sqrt{\ell}$ leads to the increased capacity:

$$C'_\ell = \frac{n}{2} \log \left(1 + \frac{\sigma_A^2 \ell}{\sigma_N^2 + n^2 \sigma_A^2} \right) \quad (38)$$

for uncorrelated fading on successive transmitted message elements. This result may be viewed as a decrease in effective noise power by the factor ℓ .

Applying (36) to Rayleigh fading we get a similar reduction in variance:

$$P(A/\bar{y}_\ell) = \frac{1}{A} \sqrt{\frac{\ell}{2\pi}} \exp \left\{ -\frac{\ell \bar{y}_\ell^2}{2A^2} \right\} \quad (39)$$

but this does not lead to any increase in capacity because the transformation $u = y\sqrt{\ell}$ leaves the expression for C' invariant. Physically this result is traceable to the mean value of zero associated with this type of distribution.

Using the optimum combinatorial law for Rayleigh fading specified by (35), with $\sigma_N^2 = 0$, we obtain for the arithmetic mean of square law demodulated signals⁸:

$$P(\bar{x}_\ell) = \frac{\left(\frac{\ell}{2}\right)^{\ell/2}}{\Gamma\left(\frac{\ell}{2}\right)} e^{-\frac{\ell \bar{x}_\ell}{2}} \bar{x}_\ell^{\ell/2 - 1} \quad (40)$$

The conditional distribution $P(A/\bar{y}_\ell^2)$ is obtained from (40) by setting $x_\ell = \bar{y}_\ell^2/A^2$. The distribution of \bar{y}_ℓ^2 becomes:

$$P(\bar{y}_\ell^2) = \frac{1}{\sigma_A \sqrt{2\pi}} \int dA e^{-\frac{A^2}{2\sigma_A^2}} \cdot \frac{\left(\frac{\ell}{2}\right)^{\ell/2}}{A^\ell \Gamma(\ell/2)} e^{-\frac{\ell \bar{y}_\ell^2}{2A^2}} \times \left(\frac{\bar{y}_\ell^2}{A^2}\right)^{\ell/2 - 1} \quad (41)$$

$$= \frac{\ell^{\ell/4}}{2^{\ell/2} \Gamma(\ell/2)} e^{-\frac{1}{\sigma_A^2} \sqrt{\ell} \bar{y}_\ell^2} \left(\frac{\bar{y}_\ell^2}{\sigma_A^2}\right)^{\ell/4 - 1} F_2\left(\frac{\bar{y}_\ell^2}{\sigma_A^2}\right) \quad (41a)$$

for ℓ even, where $F_2(x) = 1$, $F_4(x) = 1 + \frac{1}{x}$,

$$F_6(x) = \frac{3}{x} + \frac{3}{x^2} + 1.$$

The integrals which arise in determining C' can be evaluated⁹ in terms of the gamma function and its logarithmic derivative, plus some numerical integration occasioned by the presence of the F function. It is because of the labor involved in these numerical integrations that ℓ has been taken up to six only. The resultant values of transmission rate are tabulated below; for comparison, the single element case which as might be expected yields the same value of C' for the variable y_1^2 , as for y_1 , the variable considered in part I, has been listed.

ℓ	C' units/message element
1	0.500
2	0.693
4	1.01
6	1.20

For large values of ℓ we may make use of the central limit theorem to replace (40) by the Gaussian distribution:

$$P(\bar{x}_\ell) \approx \sqrt{\frac{\ell}{4\pi}} e^{-\frac{\ell}{4} (\bar{x}_\ell - 1)^2} \quad (42)$$

so that

$$P(\bar{y}_\ell^2) = \int_{-\infty}^{\infty} \frac{dA}{\sigma_A \sqrt{2\pi}} e^{-\frac{A^2}{2\sigma_A^2}} \frac{1}{A^2} \sqrt{\frac{\ell}{4\pi}} e^{-\frac{\ell}{4} \left(\frac{\bar{y}_\ell^2 - A^2}{A^2}\right)^2} \quad (43)$$

for $\ell \gg 1$,

$$P(\bar{y}_\ell^2) = \frac{\left(\bar{y}_\ell^2\right)^{-1/2}}{\sigma_A \sqrt{2\pi}} e^{-\frac{\bar{y}_\ell^2}{2\sigma_A^2}} \quad (43a)$$

i.e., the distribution of \bar{y}_ℓ^2 approaches that of A^2 . The channel transmission rate under these conditions is given by:

$$C' = \int_0^{\infty} d(\bar{y}_\ell^2) \int_{-\infty}^{\infty} \frac{dA}{\sigma_A \sqrt{2\pi}} e^{-\frac{A^2}{2\sigma_A^2}} \cdot \frac{1}{A^2} \sqrt{\frac{\ell}{4\pi}} e^{-\frac{\ell}{4} \left(\frac{\bar{y}_\ell^2 - A^2}{A^2}\right)^2} \times \log \frac{\sqrt{\frac{\ell}{4\pi}} \frac{1}{A^2} e^{-\frac{\ell}{4} \left(\frac{\bar{y}_\ell^2 - A^2}{A^2}\right)^2}}{\frac{1}{\sigma_A} \sqrt{2\pi} \bar{y}_\ell^2 e^{-\frac{\bar{y}_\ell^2}{2\sigma_A^2}}} \quad (44)$$

$$= \frac{1}{2} \log \ell + \frac{E}{2} \quad (44a)$$

where E is Euler's constant. The equivalent signal to noise ratio increases approximately linearly with ℓ , the same type of increase found for the instantaneous mean of Gaussian fading. The corresponding increase in db of effective signal to noise ratio has been plotted as a function of ℓ in figure 3, for the two types of fading.

Finally, we consider the diversity system commonly employed today; this system selects for each message element the maximum signal present in the set of receiving elements. The conditional probability distribution of this signal, denoted by \hat{y}_ℓ is, for Rayleigh fading:

$$P(A/\hat{y}_\ell) = \frac{\ell}{A\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\hat{y}_\ell}{A}\right)^2} \left[\text{erf}\left(\frac{\hat{y}_\ell}{A\sqrt{2}}\right) \right]^{\ell-1} \quad (45)$$

It is possible to expand the error function in a Maclaurin series of which only the first term need be retained if ℓ is not too large. Consequently we replace (45) with:

$$P \approx \left(\frac{\hat{y}_\ell}{A\sqrt{2}}\right)^{\ell-1} \frac{e^{-\frac{1}{2}\left(\frac{\hat{y}_\ell}{A}\right)^2}}{A\sqrt{2}} P\left(\frac{\ell}{2}\right) \quad (45a)$$

For substitution $u = \ell \frac{\hat{y}_\ell}{A}$, (45a) becomes identical with (40). Consequently the transmission rates of the two systems are equal for corresponding values of ℓ , so long as the approximation of (45) by (45a) holds. This appears to be valid through $\ell = 4$. For higher values of ℓ a decrease in performance as compared to the optimum system specified by efficiency in the sense defined by Cramer approaches zero.

A word should be said regarding the relative effects of noise upon these two types of diversity. The maximum signal selection system is preferred in practice because of the deleterious effects of noise upon a straight average of demodulated signals. To my knowledge there is no system in use which attempts to subtract off this noise in accordance with the prescription given by (35), although it may be that the variability with time of σ_N^2 makes this course impractical. In any event it is apparent that the performance of the maximum selection system will deteriorate rapidly once signal to noise ratios in the vicinity of unity are reached.

Conclusion

In this paper it is hoped a start has been made on the analysis of communications systems in which signal fading occurs. The two types of fading distributions worked with, Rayleigh and Gaussian, were chosen to illustrate situations wherein fading is, respectively, a dominant and a minor factor; subsequent approximations made in carrying through the analysis were consistent with this aim. The

fundamental viewpoint of information theory was adopted throughout, so that consideration has been limited to channel transmission rate calculations.

For shallow fading it has been shown that the signal fluctuations give rise to an equivalent noise power. A quantitative evaluation of the rather strong degradation of information produced by single point reception of a Rayleigh fading signal has been presented, and the amelioration possible with diversity systems has been explored. In this connection it has been shown that for a small number of receiving elements, the maximum signal selection type of diversity performs as well as the optimum, mean square type. The relation between the spectral and temporal characteristics of fading, and the correlation of the fading modulation on adjacent message elements has been developed, and has been shown to lead to an optimum manner of band subdivision. Expressions for the increase in transmission rate produced by partial correlation of the fading on successive signals have been developed for a variety of special cases, and from several viewpoints, although the rather intuitive justification for some of the approximations made make these results trustworthy only as order of magnitude estimates.

With the exception of section 6 of part I all values of transmission rate are based upon the natural logarithm as a measure of information. The more physically meaningful equivalent db of signal to noise ratio has been employed on the curves.

Appendix A

Derivation of P_A Distribution Which Maximizes Transmission Rate in the presence of Rayleigh Fading

$$C' = \iint A dy \frac{P(A)}{P(y)} \log \frac{P(A/y)}{P(y)} \quad (1a)$$

$$= \int A P(A) \int dy P(A/y) \log P(A/y) - \int dy P(y) \log P(y) \quad (2a)$$

In accordance with the methods of the calculus of variations, we add as constraints $\mu \int P_A(A) dA - \lambda \int A^2 P_A(A) dA$ since we wish to maximize subject to the conditions:

$$\int P_A(A) dA = 1 \quad ; \quad \int A^2 P_A(A) dA = \sigma_A^2 \quad (3a)$$

For Rayleigh fading (in the absence of noise):

$$P(A/y) = \frac{1}{A\sqrt{2\pi}} e^{-y^2/2A^2} \quad (4a)$$

Differentiating with respect to P_A ,

$$\frac{\partial [C' + \text{Constraints}]}{\partial P_A} = \int dA (-\log A - \lambda A^2 - \mu) - \int dy [1 + \log P(y)] \frac{\partial P(y)}{\partial P_A} \quad (5A)$$

where

$$P(y) = \int \frac{dA}{A\sqrt{2\pi}} e^{-y^2/2A^2} P_A(A) \quad (6A)$$

so that

$$\frac{\partial P(y)}{\partial P_A} = \int \frac{dA}{A\sqrt{2\pi}} e^{-y^2/2A^2} \quad (7A)$$

we now set (5A) equal to zero, and interchange the order of integration in the last term on the right, ignoring questions of convergence for the moment.

$$\int dA [-\log A - \lambda A^2 - \mu - 1 - \frac{1}{A\sqrt{2\pi}} \int dy e^{-y^2/2A^2} \log P(y)] \quad (8A)$$

This has the solution

$$\log P(y) = -\mu - 1 - \lambda y^2 - \log y - \frac{1}{2}(\epsilon + \log 2) \quad (9A)$$

so that

$$P(y) = \frac{C}{y} e^{-\lambda y^2} \quad (10A)$$

where C and λ are to be chosen to satisfy eq. (3A). The y distribution diverges when integrated over the infinite domain as a result of the $(\frac{1}{y})$ factor. This is related to the interchange of limits performed above. We may proceed nevertheless by writing an exponent ν on this factor, and allowing ν to approach unity only in our final normalized expressions. To find P_A we must solve:

$$\int_{-\infty}^{\infty} \frac{dA}{A\sqrt{2\pi}} P_A(A) e^{-y^2/2A^2} = \frac{C}{y^\nu} e^{-\lambda y^2} \quad (11A)$$

By making the transformation $x = \frac{1}{A^2}$ the left hand side may be brought into the form of a Laplace transform.

$$\frac{1}{\sqrt{2\pi}} \int_0^{\infty} dx \left[\frac{1}{x} P_A\left(\frac{1}{\sqrt{x}}\right) \right] e^{-y^2/2x} = \frac{C}{y^\nu} e^{-\lambda y^2} \quad (12A)$$

which has the solution:

$$P_A(A) = \frac{C\sqrt{2\pi}}{2^{1/2} \Gamma(\frac{\nu}{2})} \cdot \frac{1}{A^\nu (1-2\lambda A^2)^{1-\nu}} \quad |A| < \frac{1}{\sqrt{2\lambda}}$$

$$= 0 \quad |A| > \frac{1}{\sqrt{2\lambda}} \quad (13A)$$

The conditions (5A) lead to the following values of parameters in the vicinity of $\nu = 1$:

$$\lambda = \frac{1}{\sigma_A^2 \Gamma(\frac{1-\nu}{2})} \quad C = \frac{1}{2 \Gamma(\frac{1-\nu}{2})} \left[\frac{2}{\sigma_A^2 \Gamma(\frac{1-\nu}{2})} \right]^{\frac{1-\nu}{2}} \quad (14A)$$

The leading term in the expression for channel capacity is $\frac{1}{2} \log P(\frac{1-\nu}{2})$ which becomes infinite for $\nu = 1$. Since $\lambda \rightarrow 0$ under these conditions, the value of A having infinite probability,

$\lambda = \frac{1}{2\lambda} \rightarrow \infty$. Consequently this distribution function is deemed physically unrealizable. It results from the lack of a constraint on the maximum value permitted for A .

Appendix B

Upon inserting (17), eq. (16) becomes:

$$P(A_1, \dots, A_n | y_1, \dots, y_n) = \int_{-\infty}^{\infty} \frac{dM_1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(M_1-1)^2}{2(1-\rho^2)}} \frac{1}{\sigma_N \sqrt{2\pi}} e^{-\frac{(y_1 - A_1 M_1)^2}{2\sigma_N^2}}$$

$$\times \int_{-\infty}^{\infty} \frac{dM_2}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(M_2-1-\rho(M_1-1))^2}{2(1-\rho^2)}} \dots \int_{-\infty}^{\infty} \frac{dM_n}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(M_n-1-\rho(M_{n-1}-1))^2}{2(1-\rho^2)}}$$

(1B)

Integrating over M_n first we obtain the factor:

$$\frac{e^{-\frac{(y_n - A_n)^2}{2\sigma_n^2}}}{\sqrt{2\pi(\sigma_n^2 + A_n^2(1-\rho^2))}} \exp \left[\frac{\left\{ \frac{\rho(M_n - 1)}{m\sqrt{2(1-\rho^2)}} + \frac{m\sqrt{\rho^2}}{\sigma_n^2\sqrt{2}} A_n(y_n - A_n) \right\}^2}{1 + \frac{A_n^2 m^2}{\sigma_n^2} (1-\rho^2)} \right] \quad (2B)$$

We may now proceed with either of two approximations. For $\rho \ll 1$, if we neglect the factor $\rho(M_n - 1)$, then we obtain a simple Gaussian distribution for $(y_n - A_n)$ with variance

$$\sigma^2 = \sigma_n^2 + m^2 \sigma_n^2 (1-\rho^2) \quad (3B)$$

and succeeding integrations are independent.

Alternatively, for $\rho \approx 1$, if we neglect

$$\frac{A_n^2 m^2}{\sigma_n^2} (1-\rho^2) \quad \text{compared to unity, then after}$$

expanding the exponent in brackets, retaining the first two terms and neglecting the last, successive integrations may be carried through yielding upon summation the conditional distribution of uncorrelated fading, multiplied by the factor:

$$\exp \left[\frac{m^2}{\sigma_n^2} \sum_{k=1}^n \sum_{l>k}^n \rho^{l-k} A_k A_l (y_k - A_k)(y_l - A_l) \right] \quad (4B)$$

It has not been found possible to perform the integrations required to evaluate channel capacity in this case.

Appendix C

For Gaussian fading eq. (19) takes the form:

$$P = \int_{-\infty}^{\infty} \frac{dM}{\sqrt{2\pi}} e^{-\frac{(M-1)^2}{2\sigma_n^2}} \left(\frac{1}{\sigma_n \sqrt{2\pi}} \right)^g e^{-\sum_{k=1}^g \frac{(y_k - MA_k)^2}{2\sigma_n^2}} \quad (1C)$$

$$= \left(\frac{1}{\sigma_n \sqrt{2\pi}} \right)^g e^{-\frac{1}{2\sigma_n^2} \sum_{k=1}^g (y_k - A_k)^2} \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}} e^{-x^2 \left[\frac{1}{2\sigma_n^2} + \frac{1}{2\sigma_n^2} \sum_{k=1}^g A_k^2 \right]} \times e^{\frac{x}{\sigma_n^2} \left[\sum_{k=1}^g A_k (A_k - y_k) \right]} \quad (2C)$$

$$= \left(\frac{1}{\sigma_n \sqrt{2\pi}} \right)^g \frac{1}{\sqrt{1 + \frac{m^2}{\sigma_n^2} \sum_{k=1}^g A_k^2}} \exp \left[-\frac{1}{2\sigma_n^2} \sum_{k=1}^g (y_k - A_k)^2 \right] \times \exp \left\{ \frac{-\frac{m^2}{\sigma_n^2} \sum_{k=1}^g A_k (A_k - y_k)}{1 + \frac{m^2}{\sigma_n^2} \sum_{k=1}^g A_k^2} \right\} \quad (3C)$$

The double sum $\sum_k \sum_{l \neq k} A_k A_l (A_k - A_l)(A_k - y_l)$ tends to cancel if g is sufficiently large, since each of its individual factors is independent and random. There remains:

$$P = \left(\frac{1}{\sigma_n \sqrt{2\pi}} \right)^g \frac{1}{\sqrt{1 + \frac{m^2}{\sigma_n^2} \sum_{k=1}^g A_k^2}} \exp \left\{ -\frac{1}{2\sigma_n^2} \sum_{k=1}^g (y_k - A_k)^2 \left[\frac{1 + \frac{m^2}{\sigma_n^2} \sum_{k=1}^g A_k^2}{1 + \frac{m^2}{\sigma_n^2} \sum_{k=1}^g A_k^2} \right] \right\} \quad (4C)$$

$$= \left(\frac{1}{\sigma_n \sqrt{2\pi}} \right)^g \frac{1}{\sqrt{1 + \frac{m^2}{\sigma_n^2} \sum_{k=1}^g A_k^2}} \exp \left\{ -\frac{1}{2\sigma_n^2} \sum_{k=1}^g (y_k - A_k)^2 \left[\frac{1 + \frac{m^2}{\sigma_n^2} \sum_{k=1}^g A_k^2}{1 + \frac{m^2}{\sigma_n^2} \sum_{k=1}^g A_k^2} \right] \right\} \quad (5C)$$

For large g we may set $\sum_{k=1}^g A_k^2 \approx g \sigma_A^2$; this yields:

$$P = \left(\frac{1}{\sigma_n \sqrt{2\pi}} \right)^g \frac{1}{\sqrt{1 + \frac{m^2 g \sigma_A^2}{\sigma_n^2}}} \exp \left\{ -\frac{\sum_{k=1}^g (y_k - A_k)^2}{2\sigma_n^2 \left[1 + \frac{m^2 g \sigma_A^2}{\sigma_n^2} \right]} \right\} \quad (6C)$$

Since the normalizing factor $\left(\frac{1}{\sigma_n \sqrt{2\pi}} \right)^g$ approaches unity at large g , as does the bracketed exponential factor which modifies σ_n^2 , the interpretation given in eq. (20) of the text follows.

References

1. McNicol, R. W. E., Proc. I.E.E. III 96, 517 (1949).
2. Shannon, C. E., Bell Syst. Tech. J. 27, 635 (1948).
3. Lawson, J. L. and Uhlenbeck, G. E., "Threshold Signals" (1950), p. 49.
4. Gabor, D., J.I.E.E., III 93, 429 (1946).
5. Shannon, C. E., loc. cit., p. 413.
6. Feinstein, J., Trans. I.R.E. PGAP II, No. 1 (1954).
7. Cramer, H., "Mathematical Methods of Statistics", (1946), p. 498.
8. *ibid.*, p. 236
9. Grobner, W. and Hofreiter, N., "Integratafel", part II.

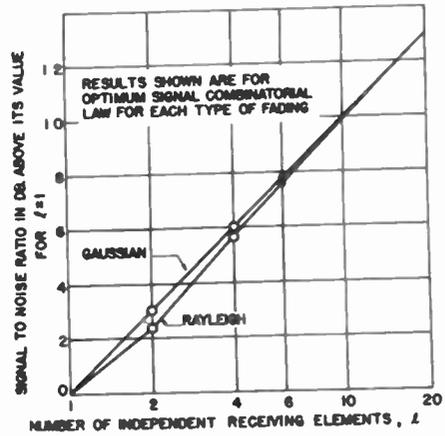
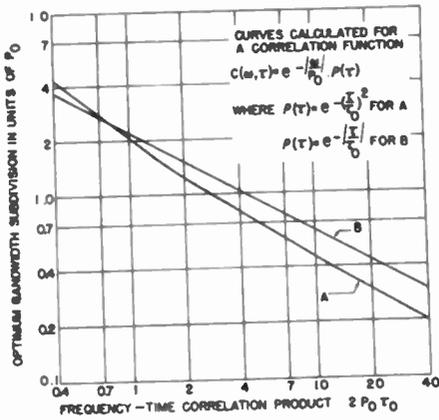


Fig. 2
Reduction in channel capacity produced by a finite number of message elements.

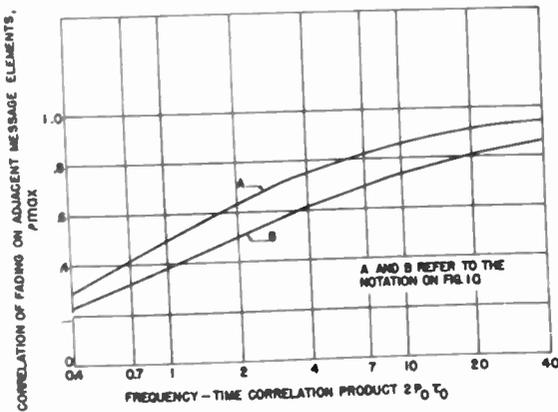


Fig. 1
(a) Optimum bandwidth subdivision in the presence of selective fading; (b) fading correlation for the optimum case.

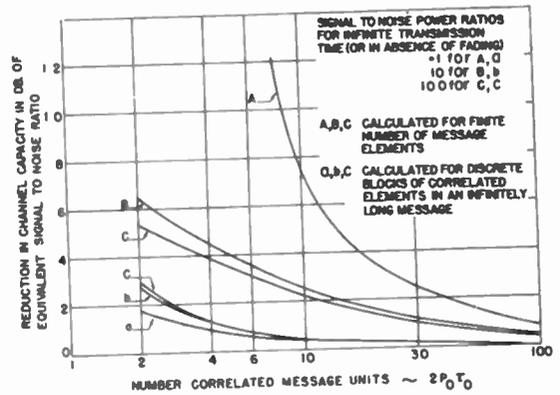


Fig. 3
Increase in effective signal-to-noise ratio produced by diversity reception.

COMPARATIVE 100 Mc MEASUREMENTS AT DISTANCES FAR BEYOND

THE RADIO HORIZON

Albrecht P. Barsis

National Bureau of Standards

Boulder, Colorado

Abstract

Results of 100 Mc measurements are evaluated in terms of distributions of hourly medians of transmission loss. Transmitters were located at elevations of approximately 6200, 8800, and 14100 feet above mean sea level on the eastern slope of the Rocky Mountains. Receiving sites were located at distances of 230 and 400 miles from the transmitters. Various types of antennas were employed for transmitting and receiving. The results of studies of fading rate and fading range are reported for the 400 mile site by comparing signals received simultaneously on two different antennas. These results are compared to those expected from the application of the tropospheric scattering theory.

1. Introduction

The Tropospheric Propagation Research Section of the National Bureau of Standards has conducted a long-range recording program on frequencies in the 100 to 1000 Mc range over a path extending eastwards from Cheyenne Mountain near Colorado Springs, Colorado. ¹ As a part of this program, receiving stations were set up and operated at distances far beyond the radio horizon in order to investigate the character of long distance tropospheric wave propagation. This paper presents some of the results obtained during two recording periods: one during August of 1952, chosen to represent a typical summer condition, and the other during February of 1953, chosen to represent a typical winter condition.

For the experiments described here the frequency of 100 Mc was used. All transmissions were continuous wave and horizontally polarized.

A fixed 2 kw transmitter located at the Cheyenne Mountain Summit site was used alternately with a mobile 1 kw transmitter which was located most of the time at Camp Carson near the base of Cheyenne Mountain, but during August 1952 was also brought to the top of Pikes Peak. The antenna elevations above mean sea level are 8805 ft. for the Cheyenne Mountain Summit site, 6260 ft. for the Camp Carson site, and 14115 ft. for Pikes Peak. The receiving site at Garden City, Kansas is approximately 225 miles from the transmitters, and at an elevation of 2860 ft. above mean sea level. Corresponding data for the Anthony, Kansas receiving site are 400 miles, and 1335 ft., respectively. Fig. 1 is a pictorial representation of the path showing also other sites not considered in this paper.

From August 6 to 13th, 1952, the transmissions were alternated every hour between the transmitter on Cheyenne Mountain and the transmitter at Camp Carson. After the mobile transmitter was moved to the top of Pikes Peak, transmissions from this point and Cheyenne Mountain were also alternated every hour. This covered the period from August 14 to August 20th.

For the 1953 recording period (February 20-27) only the Camp Carson site was used alternately with Cheyenne Mountain, and the individual transmission periods were two or three hours. The following is a list of all recording periods and paths used in this study.

Aug.6-13, 1952 Cheyenne Mt. -Garden City
Cheyenne Mt. -Anthony
Camp Carson-Garden City
Camp Carson-Anthony

Aug. 14-20, 1952	Cheyenne Mt. -Garden City Cheyenne Mt. -Anthony Pikes Peak - Garden City Pikes Peak - Anthony
Feb. 20-27, 1953	Cheyenne Mt. -Garden City Cheyenne Mt. - Anthony (Rhombic) Cheyenne Mt. -Anthony (Yagi) Camp Carson - Garden City Camp Carson - Anthony (Rhombic) Camp Carson -Anthony (Yagi)

Receiving Antennas	Gain in decibels relative to isotropic radiator
Garden City Dipole	2.15
Anthony Rhombic	16.3 (Aug. '52) and 18.5 (Feb. '53)
Anthony Yagi	11.8

3. Basic Transmission Loss and Angular Distance

It has been found convenient to express radio propagation data in terms of basic transmission loss and a parameter Θ which is termed angular distance. The concept of Basic Transmission Loss is explained by reference to Fig. 2 which shows the relationship of various system units, all of which are measured or expressed in decibels. The uncorrected transmission loss, L'_m , is the quantity actually measured which is the ratio of the power leaving the transmitter to the power available at the receiver terminals. The other quantities shown may be derived therefrom if the line losses of the transmitting and receiving systems are known, and if the measured free-space gains of the antennas are assumed to be valid for the particular transmission path. Basic transmission loss is seen to be the sum of the measured transmission loss corrected for line losses and the antenna gains, and thereby constitutes the transmission loss of a system in which the transmitting and the receiving antenna are both considered to be isotropic. The quantity B may be termed basic transmission loss in free space, and depends only on distance and frequency. ^{2/} For an attenuation relative to free space A decibels, basic transmission loss is the sum of B and A .

If it is desired to transform basic transmission loss to decibels above one microvolt per meter of received field (for one kilowatt of effective radiated power), the relation

$$F = 139.4 + 20 \log f_{mc} - L_B$$

may be used.

The angular distance Θ is best explained by reference to Fig. 3 which shows Θ for the actual paths under study, as well as for a smooth, spherical earth. The terrain profile

2. Transmitting and Receiving Antennas

As one of the principal purposes of this paper is to compare performance of various antenna systems, a detailed description of the antennas used is necessary.

The Cheyenne Mountain installation employs a 90 degree corner reflector fed by a folded dipole. This antenna is mounted on the side of a 100 ft. steel tower, and by its location on the sheer slope of the mountain. simulates an airborne transmitter.

At the Camp Carson site a 10-wavelength rhombic antenna was constructed about 40 ft. above ground level. Due to space limitations on Pikes Peak a 5-element Yagi antenna was employed there at about 20 ft. above ground level.

The Garden City receiving site is equipped with a 100 Mc dipole antenna approximately 18 ft. above ground level. The Anthony receiving site has a rhombic antenna similar to the one used for transmitting at Camp Carson. During February 1953 the Yagi antenna described above was also used for reception of the 100 Mc signals at Anthony.

The following tabulation shows the measured gain in the horizontal direction of maximum directivity of all antennas in decibels relative to an isotropic radiator. Actually the antennas were measured by comparison with a dipole which has a nominal gain of 2.15 decibels relative to an isotropic radiator.

Transmitting Antennas	Gain in decibels relative to isotropic radiator
Cheyenne Mt. Corner Reflector	10.0
Camp Carson Rhombic	18.1
Pikes Peak Yagi	8.5

in the great circle plan between transmitting and receiving antennas is plotted on the basis of an earth radius four-thirds times the actual radius thereby taking into account standard atmospheric refraction. The angle between lines drawn from the transmitting and receiving antennas tangent to the actual horizons is the angle Θ , and it may be easily seen that it is also a measure of the scattering region which is thought to be responsible for the existence of signals far beyond the radio horizon.^{3/} Fig. 3 also illustrates the change in location of the scattering region with the height of the transmitting or receiving terminal. The term "angular distance" for the angle Θ is derived from the fact that for a smooth spherical earth it is proportional to the distance between horizons.

4. Analysis of Hourly Median Data

The 100 Mc data for the recording periods listed above were analyzed principally to yield hourly median values of basic transmission loss for each path. In addition thereto, a 48 hour period (February 24-25, 1953) has been selected as a basis for more detailed studies of the simultaneous reception on the Rhombic and Yagi antenna at Anthony.

For each path and each recording period a cumulative distribution of hourly medians was computed and plotted on logarithmic probability paper. Samples of such distributions are shown on Fig. 4 illustrating the distribution of hourly medians at Garden City for the period February 20-27, 1953. In our studies we chose the difference of the values exceeded by 90% and 10% of all hourly medians (sometimes called the "interdecile range") as a measure of the variations of transmission loss. These values were determined for all recording periods and paths. The 50% value and the interdecile range are plotted versus the angular distance Θ on Fig. 5. The trend of the 50% values (or overall medians) serves to illustrate the usefulness of the Θ concept in evaluating long-distance propagation data. The trend shown indicates that there is a pronounced seasonal effect with increasing Θ . This seasonal effect is even more striking in a study of the range of hourly medians. The interdecile range shown varies from 4 to 6 decibels in February to more than 15 decibels in August, with its maximum appearing for a value of Θ appropriate to the Garden City receiving site.

The number of hourly medians measured during the recording periods described in this paper was not sufficient to permit an evaluation of diurnal variations.

5. Comparison of signals received simultaneously on the Rhombic and Yagi antennas at Anthony

a. Detailed Study of the February 24-25 Data

For the period of February 24-25 the records from the receivers connected to the Rhombic and the Yagi antennas were analyzed using a basic period of five minutes. For each of these periods a cumulative time distribution of transmission loss levels was obtained and plotted on specially designed graph paper (as shown on Fig. 6 for the hour midnight to 1:00 AM (February 25). Inspection of this figure shows that the distributions plotted in this manner approximate straight lines with negative slopes close to unity. The graph paper has been designed in a way so that a Rayleigh distributed variable will appear as a straight line having a slope of -1 . It is thereby shown that transmission loss levels for short periods of time are Rayleigh distributed. Thus the signal received by the antenna appears to be the vector sum of a large number of components of random phase with its total energy constant over the short time period considered^{2/} - a result which is now commonly ascribed to scattering within the volumes defined on Fig. 3. The average of the twelve 5-minute distributions is also shown on Fig. 6.

The median for each of the 5-minute distributions served as a reference level for determining fading rate. Of many possible definitions it was found convenient to define fading rate as the number of times per minute the signal trace crosses its median level with positive slope. This is also the number of signal excursions per minute above the median level. Fading rate values obtained in this way were averaged for each hour of record. Fig. 7 shows the average fading rates plotted versus time for the four paths. The appearance of the graph suggests that the fading rates are correlated, and calculations show that for the signals received simultaneously on the Rhombic and the Yagi antennas the correlation coefficient of the average hourly fading rate is 0.847 for Cheyenne Mountain transmissions,

and 0.890 for transmissions originating from Camp Carson.

If the overall two-day average fading rate is plotted versus the combined transmitting and receiving antenna gains (see Fig. 8) the appearance of the graph shows the fading rate as a function of the gains as well as a function of Θ . It has been shown by S. O. Rice that the fading rate of the received signal is a function of the r. m. s. velocity of the scattering elements drifting relatively to each other within the scattering volume. ^{4,5} In reference 4, Rice gives a relation between the drift velocity of scatterers and the fading rate of the received signal on the assumption of small scale turbulence. Although the atmosphere does not satisfy the condition of small scale turbulence, we may nevertheless get an order-of-magnitude estimate of the drift velocity of the scatterers by using Rice's formula. The observed fading rate may be used to compute the r. m. s. drift velocity in accordance with the following expression (based on the assumption that the signal is Rayleigh-distributed),

$$u = \frac{\lambda N}{177 \sin \frac{\Theta}{2}}$$

where u is in meters per second, N is the observed fading rate per minute, λ is the wave length in meters, and Θ is the angular distance defined above. The observed average fading rates are tabulated below, together with the r. m. s. drift velocities derived therefrom.

Path	N fades/ min.	Θ milli- radians	u meters/ sec.
Cheyenne Mt. - Rhombic	5.1	58.5	2.95
Cheyenne Mt. -Yagi	5.2	58.5	3.01
Camp Carson- Rhombic	3.6	68	1.79
Camp Carson-Yagi	3.9	68	1.94

This tabulation shows that the drift velocities derived from the fading rates are smaller for the Camp Carson path which corresponds to a scattering volume at higher altitude. (See Fig. 3) The distinction between simultaneous data from the two receiving antennas is less pronounced if fading rate only is considered.

b. Comparison of hourly medians received simultaneously on the Anthony antennas

In comparing hourly median values of transmission loss received simultaneously on the Rhombic and Yagi antenna, all available data for the entire February 20-27 recording period have been considered. It should be kept in mind that the two antennas are at approximately the same height above ground, and are spaced about 60 ft. apart horizontally in a way so that interaction between the antennas is minimized. Fig. 9 shows the relation between the hourly medians received on the two antennas in the form of two scatter diagrams - one for Cheyenne Mountain and one for Camp Carson transmissions. Each point on each of the diagrams represents one hourly median with its abscissa corresponding to the transmission loss received on the Rhombic, and its ordinate corresponding to the transmission loss received simultaneously on the Yagi. The difference in transmission loss (or ratio of received fields) should be constant depending on the difference in antenna gain and line losses, and is represented by straight lines drawn on the graph. Inspection of the graphs shows, however, that there is considerable scattering of points and the lines do not seem to represent an average drawn through the points. The computed correlation coefficients are 0.331 for the Cheyenne Mountain data and 0.510 for the Camp Carson data. The additional departure of the theoretical relationship shows that the measured free-space gain values are not realized over long distance transmission paths. ³ A cumulative distribution of the differences in median values is shown on Fig. 10. Normal distributions are approximated with an interdecile range of 6.0 decibels for the Cheyenne Mountain transmissions, and 4.0 decibels for the Camp Carson transmissions.

The degree of correlation between hourly median values is thereby found to be much less than between average hourly fading rates. Also, the correlation for the Camp Carson medians is slightly better than the one for the Cheyenne Mountain medians. If this difference is significant it points to the fact that Camp Carson transmissions are affected by a scattering volume at higher altitude for which drift velocities are shown to be smaller by interpretation of the fading rate data. The scattering volume associated with the Camp Carson transmissions

may also be assumed to be reduced in size due to the higher gain transmitting antenna.

6. Conclusions

The results of the measurements described in this paper are divided into two parts. The first part concerns the behavior of hourly median values of transmission loss for various paths at typical summer and winter recording periods, and the second part concerns detailed investigation of the fields received simultaneously on two antennas.

Results of the first part are summarized in Fig. 5. The median hourly transmission loss value tends to increase with increasing angular distance Θ and shows a seasonal trend of less increase in summer than in winter. Its variance is substantially higher in summer, but tends to decrease for larger values of Θ .

Results of the second part show good correlation between fading rates of signals received simultaneously on two antennas at essentially the same height and the same distance from the transmitters; however, the correlation between median values of trans-

mission loss received at the two antennas is substantially less than the correlation of fading rates.

References

1. "Tropospheric Propagation Research", NBS Technical News Bulletin, Vol. 36, No. 8, pp. 120 - 124, August 1952.
2. K. A. Norton, "Transmission Loss in Radio Propagation", Proc. I. R. E., Vol. 41, No. 1, pp. 146 - 152, January 1953.
3. J. W. Herbstreit, K. A. Norton, P. L. Rice, and G. E. Schafer, "Radio Wave Scattering in Tropospheric Propagation", 1953 I. R. E. National Convention Record.
4. S. O. Rice, "Statistical Fluctuations of Radio Field Strength Far Beyond the Horizon", Proc. I. R. E., Vol. 41, No. 2 pp. 274 - 281, February 1953.
5. S. O. Rice, "Mathematical Analysis of Random Noise", Bell Telephone System Monograph G-1589, 1944-45.

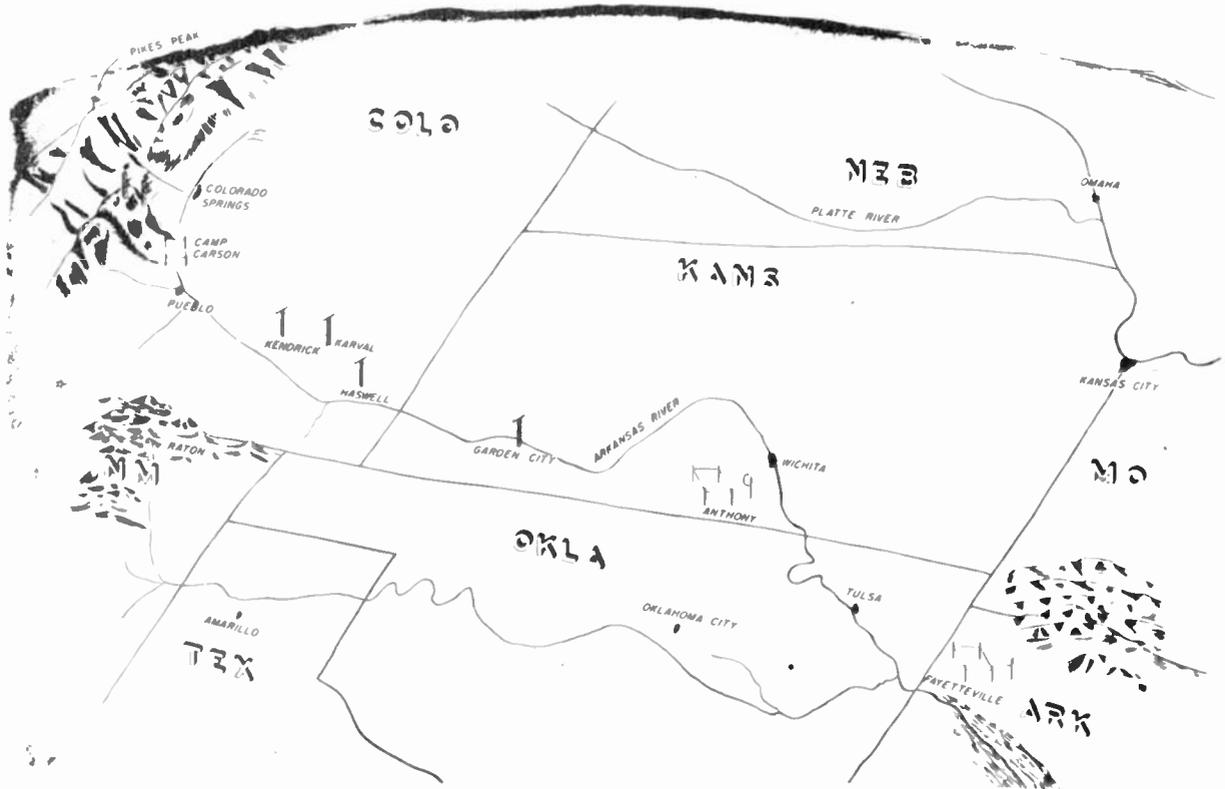
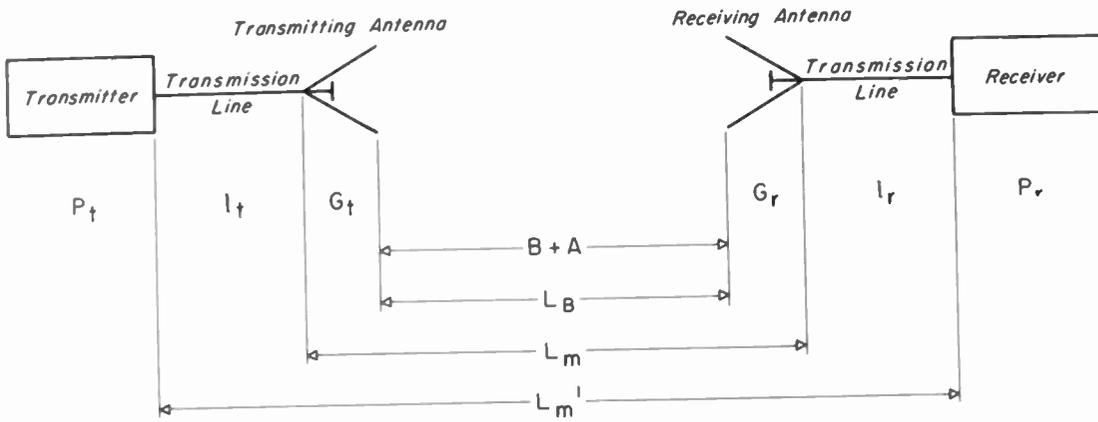


Fig. 1



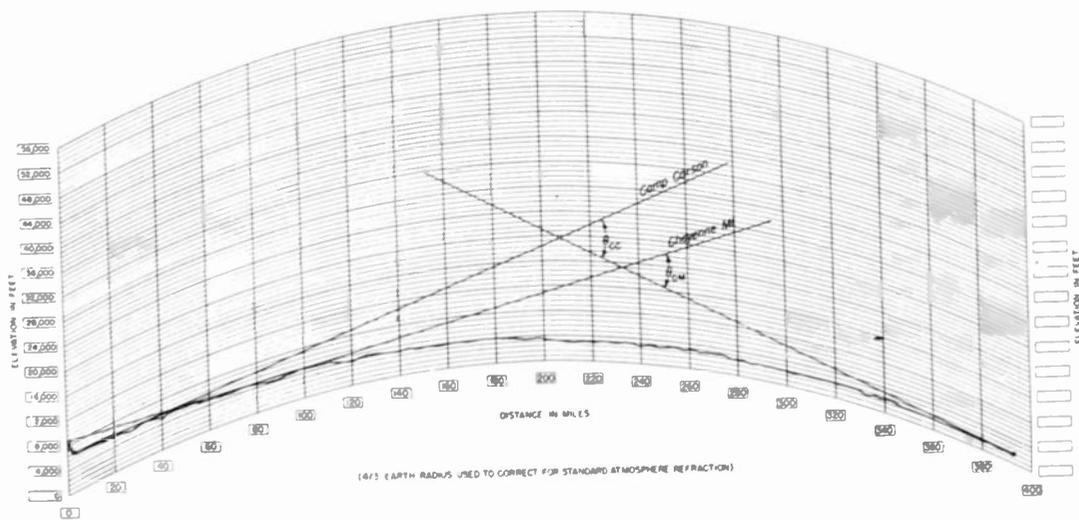
$$L_m = L_m' - (l_t + l_r)$$

$$L_B = L_m + (G_t + G_r)$$

$$A = L_B - B$$

$$B = 20 \log \left(\frac{4\pi d}{\lambda} \right)^2 = 36.581 + 20 \log d_{mi} + 20 \log f_{mc}$$

Fig. 2
Pictorial representation
of system units.



DERIVATION OF θ FOR SMOOTH EARTH

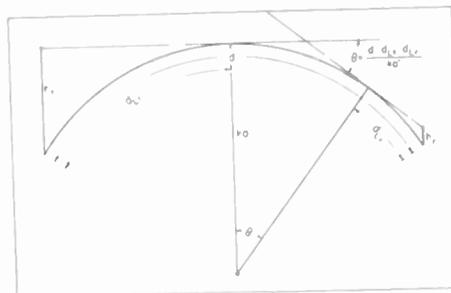


Fig. 3 - The parameter θ in tropospheric propagation.

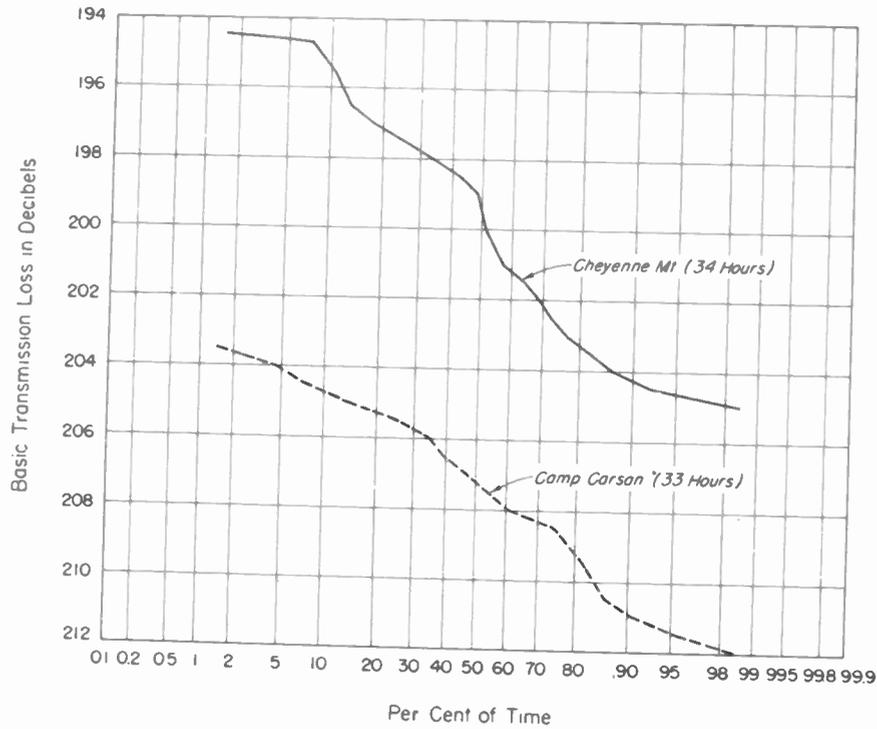


Fig. 4
Sample distribution of hourly
medians, February 20-27, 1953;
Garden City - 100 mc.

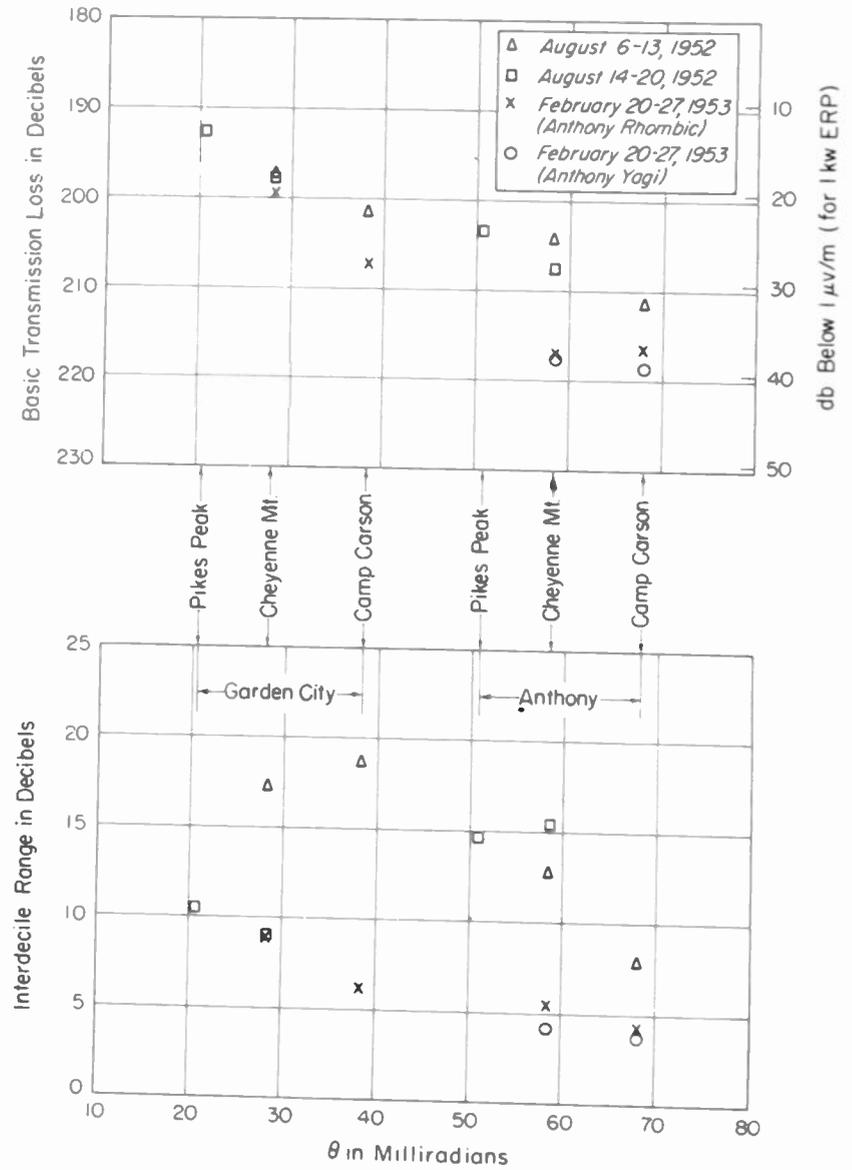


Fig. 5
Graph of measured
medians and interdecile ranges
for $f = 100$ mc.

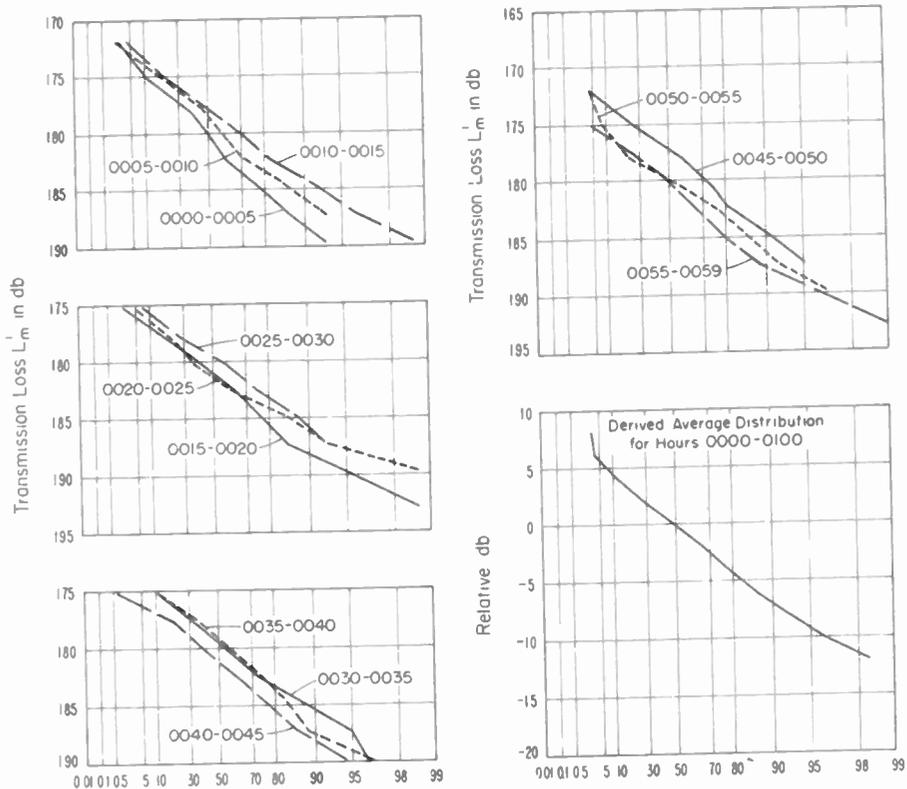


Fig. 6 (left) - Sample distributions of instantaneous transmission loss levels; Camp Carson to Anthony rhombic, 100 mc - February 25, 1953.

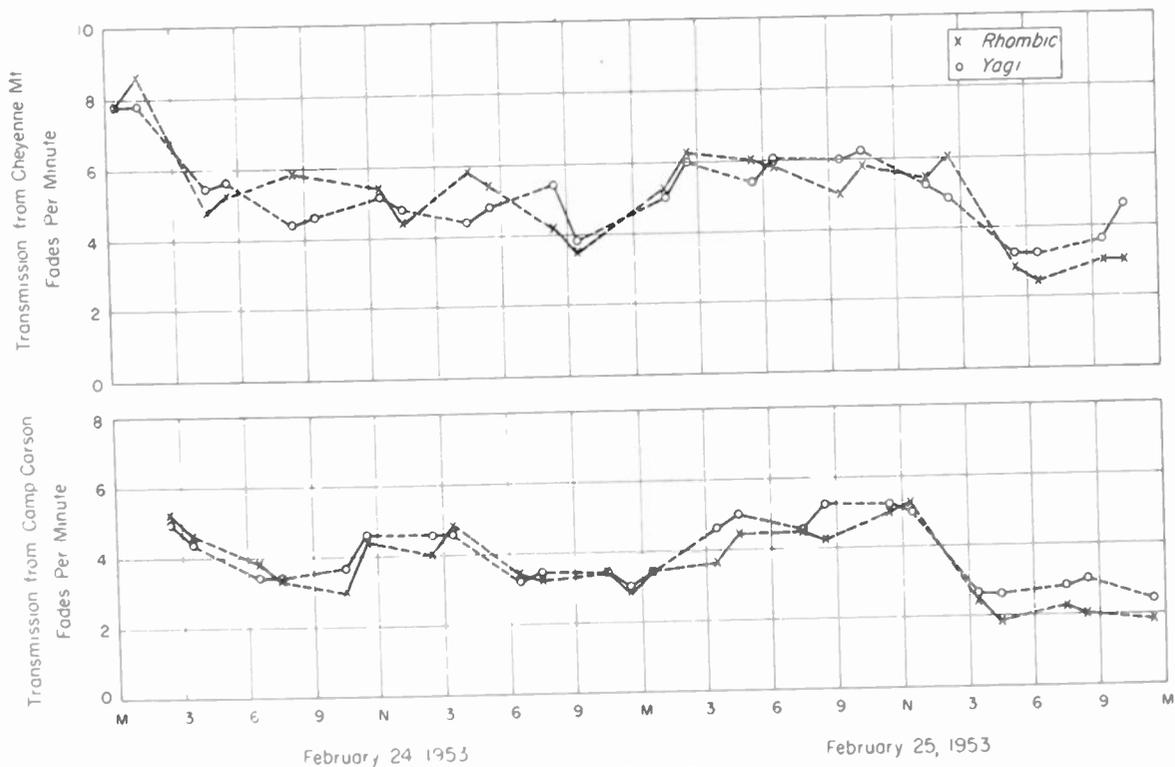


Fig. 7
Average hourly fading rate of 100-mc signal; Anthony, Kansas - February 24-25, 1953.

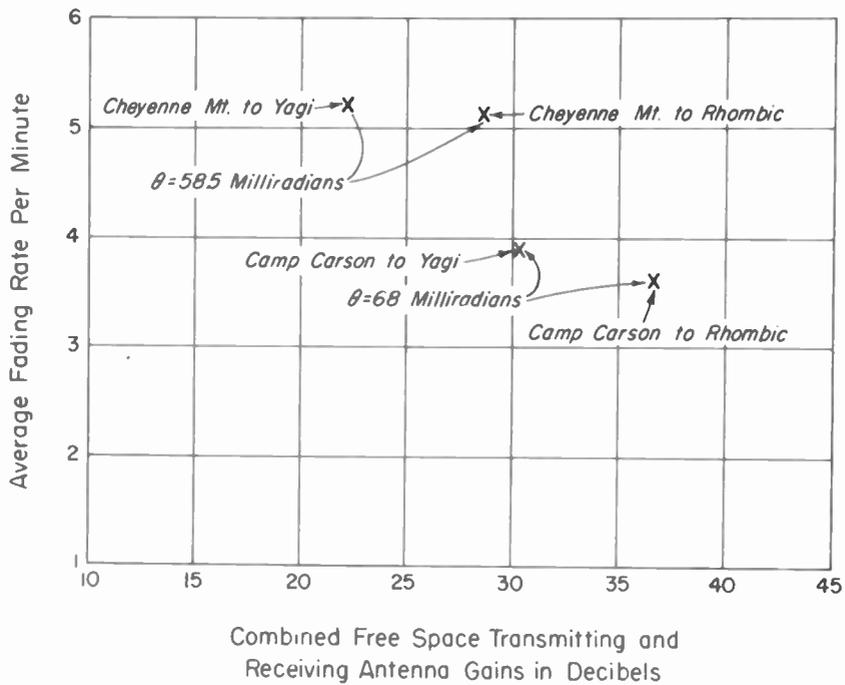


Fig. 8
Graph of average fading rate vs. antenna gain.

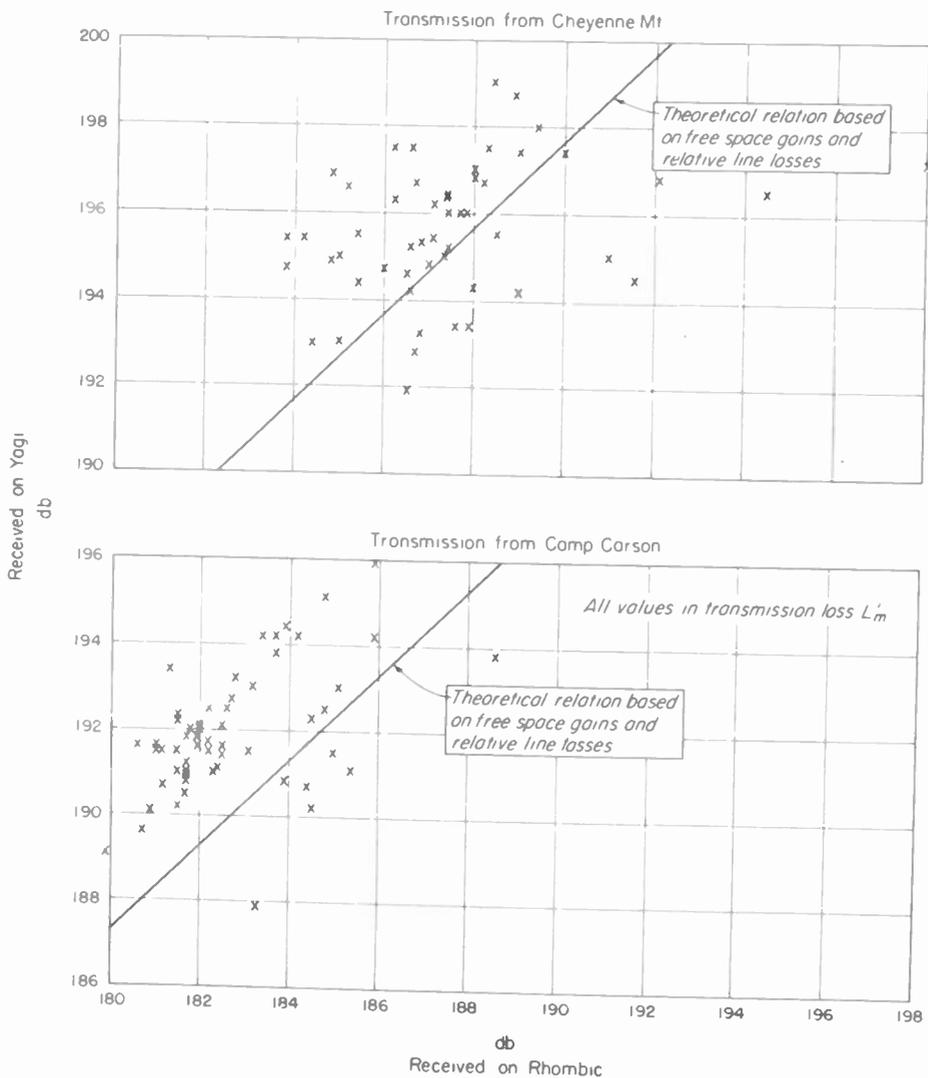


Fig. 9
Correlation diagram of hourly median values.

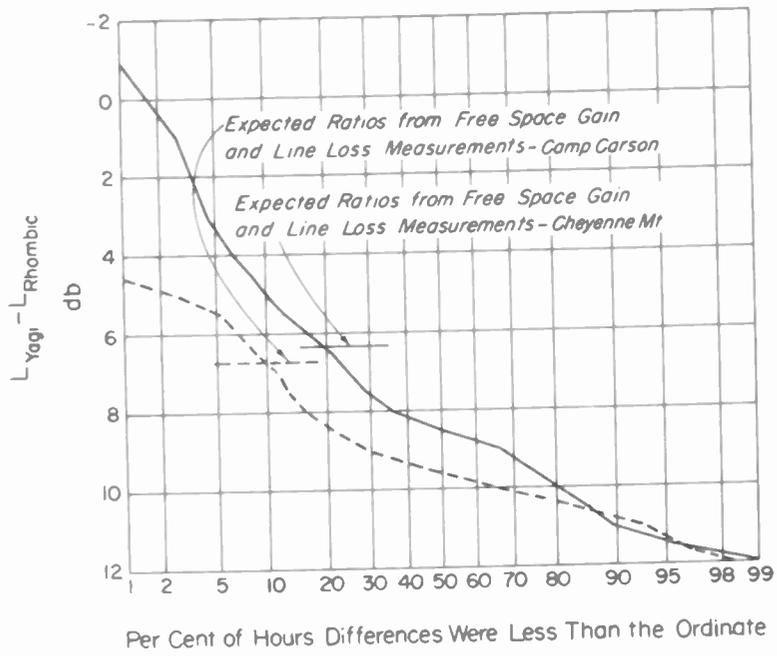


Fig. 10
 Distribution of difference
 in hourly median transmission
 loss received on rhombic and Yagi;
 Anthony, Kansas, 100 mc -
 February 20-27, 1953.

THE MEASUREMENT OF THE POLARIZATION OF RADIO WAVES REFLECTED
FROM THE IONOSPHERE AT NON-VERTICAL INCIDENCE

G. T. Inouye
Cruft Laboratory, Harvard University
At Present with Technicolor Motion Picture Corporation
Hollywood, California

Summary

Measurements of direction of arrival as well as polarization have been made on 9.1 mc/s pulsed transmissions over a 1300 km path from Glenville, North Carolina to Lexington, Massachusetts. The vertical and horizontal angles of arrival were required in order to determine the polarization of the downcoming wave from that measured at the receiving site. Therefore, the direction of arrival was measured in addition to the polarization. Equipment for the continuous automatic recording of the information was constructed; the directional data being obtained in the form of two phase measurements, and the polarization data as phase and amplitude ratio measurements. Data were recorded simultaneously for each pulse of the series of pulses arriving for each one transmitted. Samples of the data are shown, and the correlation with the magnetic-ionic theory is discussed.

Introduction

Due to the presence of the earth's magnetic field in the ionosphere, an incident radio wave is split into an ordinary and an extra-ordinary component, each of which propagates through the ionosphere according to its characteristic index of refraction. The polarizations as well as the indices of refraction of these two components at any point on their trajectories are determined by the following: (1) the direction of the wave normal, (2) the magnitude and direction of the earth's magnetic field, (3) the density of free electrons, (4) the frequency of the transmitted signal, (5) the frequency of collisions of the free electrons with surrounding air molecules, and (6) whether the component is ordinary or extra-ordinary.

In particular, Booker¹ has shown that the limiting polarization of the downcoming wave is determined at the lower edge of the ionosphere where the density of free electrons is very small. Hence, the direction of the wave normal applicable to the computation of the polarization of the downcoming wave is obtained by measuring the vertical angle of arrival. In addition, the vertical angle of arrival enters into the polarization measurement in the angular relationship between the downcoming wave and the loop-type receiving antennas used for the measurement. Note that the horizontal angle of arrival also enters into this consideration if the signal deviates appreciably from the great circle path. Finally, the vertical angle of arrival enters into the ground reflection

coefficients which also modify the measured polarization.

Theoretical Background

The computation of the trajectory, and in particular the vertical angle of arrival, from vertical height measurements at vertical incidence, is usually performed by neglecting the earth's magnetic field, and also the collisions of the free electrons with surrounding air molecules which mainly affects the absorption. For transmissions over short distances the assumption of a flat earth is valid and the computations are straightforward. For more oblique paths, the curvature of the earth must be taken into account and the problem is considerably more difficult. Waterman² has investigated the approximations made by various workers in solving the problem and has come to the conclusion that a numerical solution was most appropriate. Figures 1 and 2, taken from his report, are curves of vertical angles of arrival and transmission time delay as functions of the maximum electron density in the F₂-layer. These curves were computed by the Wave Propagation group at Cruft Laboratory for the 1300 km path between Glenville, North Carolina and Cambridge, Massachusetts. This particular path was chosen so that the ionospheric sounding station of the National Bureau of Standards at Ft. Belvoir, Virginia lies at the midpoint. The abscissas in Figures 1 and 2 have been normalized with respect to the signal frequency, f , and is in terms of F :

$$F = \frac{f}{f_c}, \quad F_c^2 = \frac{N_o e^2}{M \epsilon_0} \quad (1)$$

The critical frequency, f_c , is proportional to the square-root of the maximum electron density, N_o , in the layer. Hence, the abscissa of the tip of the "nose" of the curves at the extreme right corresponds to the minimum value of N_o for which transmission occurs. As N_o increases, the abscissa goes towards the left, and there are two diverging branches, the upper of which describes the Pedersen ray which has a higher angle of arrival and a greater transmission time delay than the lower ray.

The effect of the earth's magnetic field may now be reintroduced as a first approximation as small frequency corrections,

$$f_o = f \sqrt{1 + |Y_L|}, \quad \text{and} \quad f_x = f \sqrt{1 - |Y_L|} \quad (2)$$

for the ordinary and extra-ordinary components. The frequency normalized longitudinal component of the gyro-magnetic frequency, γ_L , is

$$\gamma_L = \frac{eB^0 \cos \theta}{2\pi f m}, \quad (3)$$

where B^0 is the earth's magnetic field and θ is the angle between the wave normal and the direction of B^0 . The gyro-magnetic frequency, $\frac{eB^0}{2\pi m}$, is about 1.46 mc/s for the United States. The question arises as to the value of θ to be used in the frequency corrections, since it varies along the trajectory. Taking the maximum value of $\cos \theta$, γ_L is 0.16 for the signal frequency of 9.1 mc/s used, and the correction amounts to 8% at most. Thus for a given signal frequency and electron density, the abscissa, F , is shifted slightly to the right in Figures 1 and 2 for ordinary rays, and to the left for extra-ordinary rays. Hence the sequence of arriving signals in the order of increasing vertical angle of arrival and increasing time delay is (1) extra-ordinary lower ray, (2) ordinary lower ray, (3) ordinary Pedersen ray, and (4) extra-ordinary Pedersen ray. Because of the frequency correction there is a small value of N_0 for which extra-ordinary rays are transmitted, but no ordinary rays. Thus, the first signals to come through in the morning and the last to come through at night are extra-ordinary rays.

The downcoming magneto-ionic components may be described as being elliptically polarized in general. They have the property that, looking in the direction of propagation, the ellipse of the ordinary ray is traced out with a counter-clockwise sense of rotation, and the extra-ordinary ray with a clockwise sense of rotation if the angle, θ , between the wave-normal and the earth's magnetic field is acute upon emerging from the ionosphere. The senses of rotation are reversed if θ is obtuse, and in any case, are always of the opposite senses for the two components. When θ is 90° , both polarizations are linear, and when θ is zero or 180° , both are circular.

In Figure 3 are curves showing the polarization:

$$R' = \frac{H_{\parallel}}{H_{\perp}} = - \frac{E_{\perp}}{E_{\parallel}} = P' e^{j\phi'} \quad (4)$$

as a function of the vertical angle of arrival, ψ , for various directions of propagation, B , measured from magnetic North. These have been computed for a signal frequency of 9.135 mc/s, a gyro-magnetic frequency of 1.46 mc/s, and a magnetic dip or inclination of $68^\circ 30'$. The effect of collisions between electrons and surrounding air molecules has been assumed to be negligible. The subscripts, \parallel and \perp , indicate the components of the H or E fields parallel and perpendicular to the plane of propagation, that is, the plane of the great circle passing through the receiver and transmitter. For the path used ($B = 68^\circ 30'$ fortuitously happened to be equal to the magnetic dip), the polarization of both components is nearly circular since the amplitude ratio, P_0' , varies from 0.8 to

1.0, and the phase angle, ϕ_0' , from 70° to 90° for the ordinary ray. The polarization of the extra-ordinary ray is obtained from the relations

$$P_{X'} = \frac{1}{P_0'}; \quad \phi_{X'} = \phi_0' + \pi. \quad (5)$$

Measuring Techniques

The requirement that the directions of arrival be measured in addition to the polarization was pointed out in the previous section. This was accomplished by two phase difference measurements of the outputs of two pairs of spaced antennas. The polarization was determined by measuring the amplitude ratio and phase difference of the outputs of two loop antennas with their axes both horizontal and mutually perpendicular and as close to the ground as possible. The orientation of these antennas is shown in Figure 4. The direction measuring antennas are located so that the pair in line with the great circle path measures the vertical angle of arrival directly if the signal suffers no appreciable lateral deviation. The other pair, located on a perpendicular line, gives a measure of the lateral deviation.

The polarization, $Q'e^{j\phi'}$ as measured by the loop antennas oriented as shown in Figure 4, computed from the polarization of the downcoming curve, is shown in Figure 5 as a function of the vertical angle of arrival, ψ . These curves are obtained from those of Figure 3 by adding the ground-reflected components to the downcoming components to obtain the total H field measured by the loop antennas. Also included are curves which would have been obtained, had the ground been a perfect reflector with infinite index of refraction, $\mu = \infty$, and also if the downcoming waves had been perfectly circular in polarization, $R = \pm j$. The actual index of refraction, $\mu = \sqrt{81 - j58}$, was computed from the measured ground constants of 81 for the relative dielectric constant, and 29.3×10^{-14} emu for the conductivity.

The technique used for measuring the phase difference of the outputs of the various antennas in determining the direction of arrival and the polarization is essentially a substitution and null method. A block diagram of the system with a vectorial representation of the voltages is shown in Figure 6. The phase of one channel is shifted continuously with respect to the other channel at a constant rate of 16 seconds per revolution by a synchronous motor. Taking the difference between the detected r.f. sum and the detected r.f. difference, the output is a pulse envelope whose sign changes from + to - and from - to + for each rotation of the phase shifter; the nulls occurring when the r.f. phases are in quadrature. The measurement is completed by introducing a signal of known phase from a local pulsed transmitter and noting the difference in dial settings of the calibrated phase shifter for corresponding transitions in the signs of the reference and the unknown signals. The reference

pulse transmitter is located about three wavelengths away in the direction of the remote transmitter. Therefore the local signal is a reference for zero lateral deviation, zero vertical angle of arrival, zero polarization phase angle, and zero db polarization amplitude ratio. The data are actually recorded by photographing a fast horizontal sweep of about 1500 microseconds duration on a cathode ray tube, intensity modulated by the video difference signal described above. Since we are working with pulsed signals, the reference pulse is put at the beginning of the sweep and measurements are made on the entire train of pulses which occupy the remainder of the sweep. This fast horizontal sweep is also swept across the screen vertically in synchronism with the shaft of the phase shifter. We thus obtain a frame calibrated vertically in degrees of phase and horizontally in relative transmission time delay. These frames are indexed horizontally along the roll of photographic paper or film, giving a semi-continuous measure of the phase.

To record the polarization amplitude ratio, the gain of one channel is varied with respect to the other by a potentiometer ganged with the phase shifter shaft. The r.f. sum and difference circuits are disabled, but the video difference is still applied to the intensity grid of the cathode ray tube. With the same presentation as for the phase measurement, the amplitude ratio is determined by noting the relative heights in the frame where the reference and the unknown signals change signs.

Results

The data taken in the first part of the project were recorded on 35mm film and decoded on a microfilm reader. An example of the records obtained is shown in Figure 7. On these measurements, readings of the horizontal angle of arrival, ϕ_0 , vertical angle of arrival, ϕ_ψ , polarization phase angle ϕ_p , and polarization amplitude ratio, ϕ_A , were made sequentially for three minutes out of every ten minutes. Since each frame required 16 seconds, two or three readings were averaged for each point plotted. The first trace of alternate light and dark sections at the left of each frame are calibration markers for the rotation of the shaft of the phase shifter. The non-linearity of the vertical sweep was due to the potentiometers used, since the calibration markers were taken directly off the shaft of the synchronous motor. The phase shift was linear with respect to shaft rotation. The second trace from the left in each frame is the reference pulse, while the third and fourth traces are the E- and F-layer signals. The fifth trace, which appears diffuse and irregular, is an abnormal reflection and can be seen to be definitely arriving from other than the great circle path in the ϕ_0 frames. The transmission time delay record taken at Cruft Laboratory on the same day is shown at the top of Figure 7. The Bureau of Standards has reported this particular day as being a magnetically stormy day, and the time delay record shows loss of transmission between 1350 and 1450 EST and from 1700 EST on.

In the latter part of the project the data were recorded on 3 5/16 inch wide photographic paper instead of film, doing away with the necessity for a microfilm reader. Instead of recording the four phase angles sequentially they were recorded at different times. A linear wire potentiometer was obtained for the vertical sweep generator, and the frame indexed along the photographic paper only in between frames so that the frames were recorded vertically instead of on a slant. An example of the type of record obtained is shown in Figure 8. This particular set of records shows the horizontal angle of arrival and its behavior near the end of transmission as the electron density is decreasing in the evening. From the point of view that the edge of the layer is travelling westwards, following the sun, it is surprising to note in how short a time--only a matter of minutes--the signal shifts towards the west before giving out. The maximum amount of lateral deviation before the signal disappears is about 5 degrees.

An example of the data obtained is shown in Figure 9. The horizontal angle of arrival was found to fluctuate in a random manner about a mean value with amplitudes of a few degrees with periods ranging from 10 to 30 minutes, and the vertical angle of arrival, likewise, with somewhat larger amplitudes. Observations of similar phenomena have been reported by the British workers, E. N. Bramley and W. Ross³. Further examples of the fluctuations in the horizontal angle of arrival are shown in Figure 10. The deviation of the mean value from the great circle path of a few degrees is presumably an error in the laying out of the receiving antennas although this was checked with a U.S. Geodetic Survey map. Going back to Figure 9, the value of the vertical angles of arrival computed from the Bureau of Standards reports of critical frequencies and the curves of Figure 1 are indicated with squares. These tend to be about 5 degrees lower than the observed values, although some records were obtained where the correlation was much closer. One possible explanation of the discrepancy may be in the particular model of the ionosphere assumed in the computation of Figure 1.

The polarization was found to be as expected from the theoretical considerations whenever the various modes of transmission were separated in time of arrival. Referring to Figure 2, the relative time delay of the ordinary and extraordinary lower ray for pulses of about 100 microseconds width is insufficient to resolve these two components except at the very tip of the "nose". The components of the Pedersen ray are resolved, but these are more highly attenuated except near the "nose". Fortunately, the conditions of propagation are such that the attenuation is small at just these times. Thus the polarization amplitude ratio, Q' , was found to fluctuate widely above and below zero db for the lower ray indicating phase interference between the ordinary and extra-ordinary components, but to be near unity ratio for the Pedersen ray, as predicted. The polarization phase angle, q' , also fluctuated widely for the lower ray, but showed

some tendency toward negative values indicating a predominance of the ordinary component as against the extra-ordinary. No statistical treatment of the data was attempted.

The ability of the extra-ordinary ray to sustain transmission with small electron densities insufficient for ordinary ray transmission is strikingly brought out by the polarization phase angle records shown in Figure 11. On these records phase angles below the reference line indicate negative values of q' and hence ordinary rays, while phase angles above the line indicate extra-ordinary rays. The reference lines are drawn from the reference signal which appears at the left side of each frame. In Figure 11a, for instance, the lower ray fluctuates until about 2012 EST while the Pedersen ray exhibits well behaved ordinary and extra-ordinary components until the ordinary component, which has merged with that of the lower ray, disappears at 2015 EST. Thus, a gap appears between the extra-ordinary lower and Pedersen rays, and these gradually merge together until they are indistinguishable from each other at 2020 EST. The remaining signal which lasts until 2025 EST has the polarization of an extra-ordinary ray.

The project was supported jointly by the Navy Department (Office of Naval Research), Signal Corps of the U.S. Army, and the U.S. Air Force under Contract N5-ori-76, T.O. 28. The aid and cooperation of Professor H. R. Mimno, Mr. J. A. Pierce, and the members of the Wave Propagation group at Cruft Laboratory is gratefully acknowledged.

References

1. H. G. Booker, "Oblique Propagation of Electromagnetic Waves in a Slowly Varying Non-Isotropic Medium", Proc. Roy. Soc. 155, 235, (1936).
2. A. T. Waterman, Jr., "Ray Theory Applied to a Spherical Ionosphere", Cruft Laboratory Technical Report No. 142, (1952).
3. E. N. Bramley and W. Ross, "Measurement of the Direction of Arrival of Short Radio Waves Reflected at the Ionosphere", Proc. Roy. Soc. 207, 251, (1951).

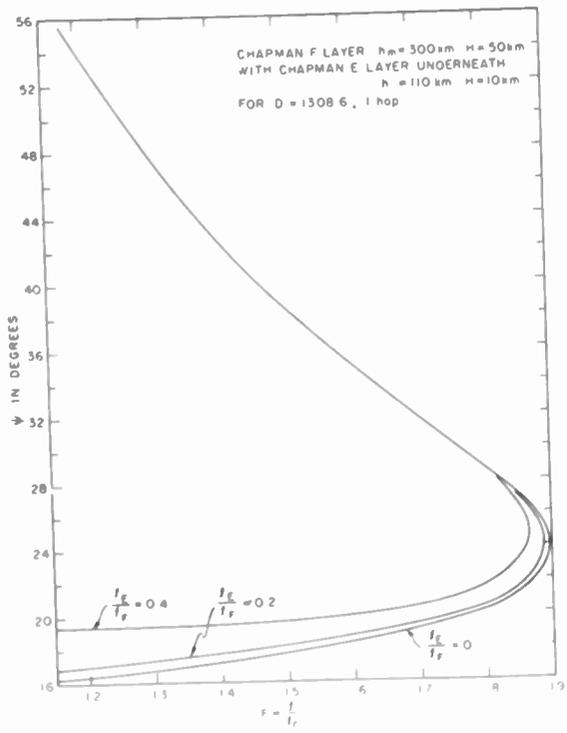


Fig. 1
The vertical angle of arrival on the Glenville-Cambridge path.

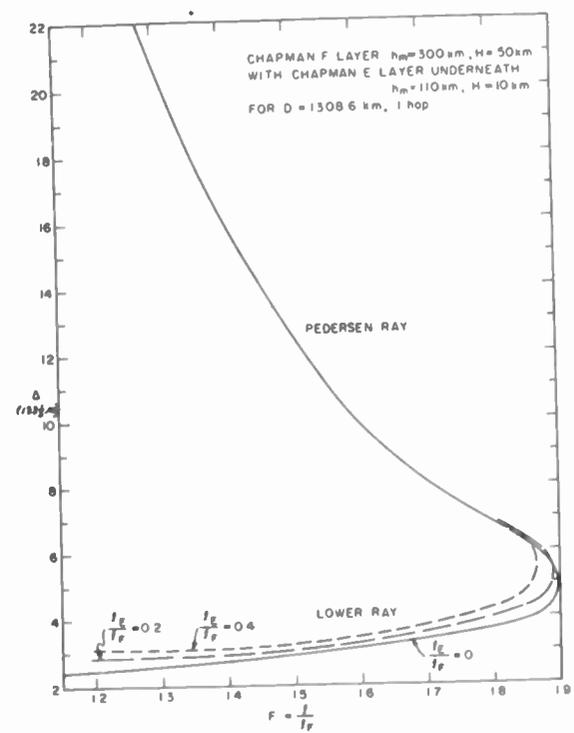


Fig. 2
Transmission delay-time on the Glenville-Cambridge path.

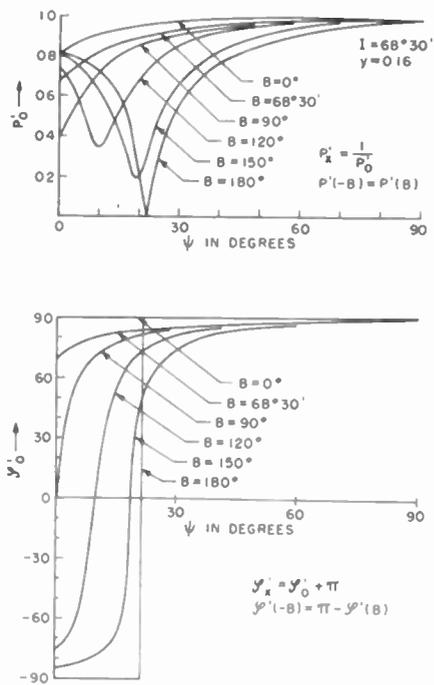


Fig. 3
The polarization $R' = P'e^{j\theta'}$ of the downcoming wave as a function of the vertical angle of arrival.

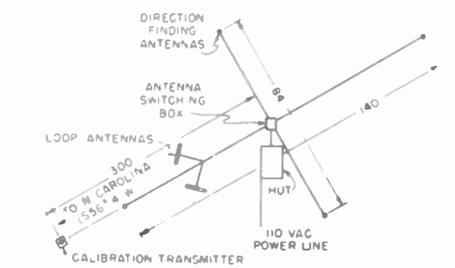


Fig. 4
Plan view of entire antenna system.

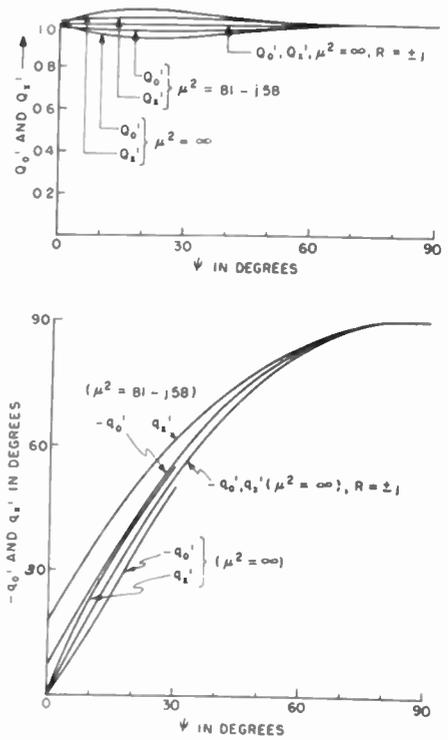


Fig. 5
The polarization $Q'e^{j\theta'}$ in the rotated ground plane coordinates $B = 68 + 30j$.

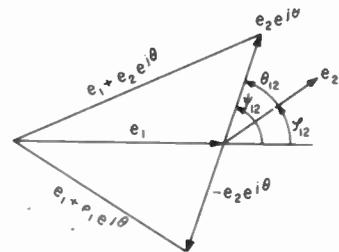


Fig. 6(a)
The vectors in the complex plane.

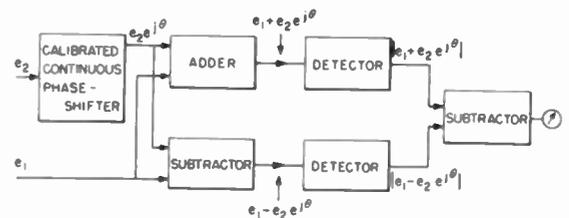


Fig. 6(b)
Block diagram of phase-measuring system.

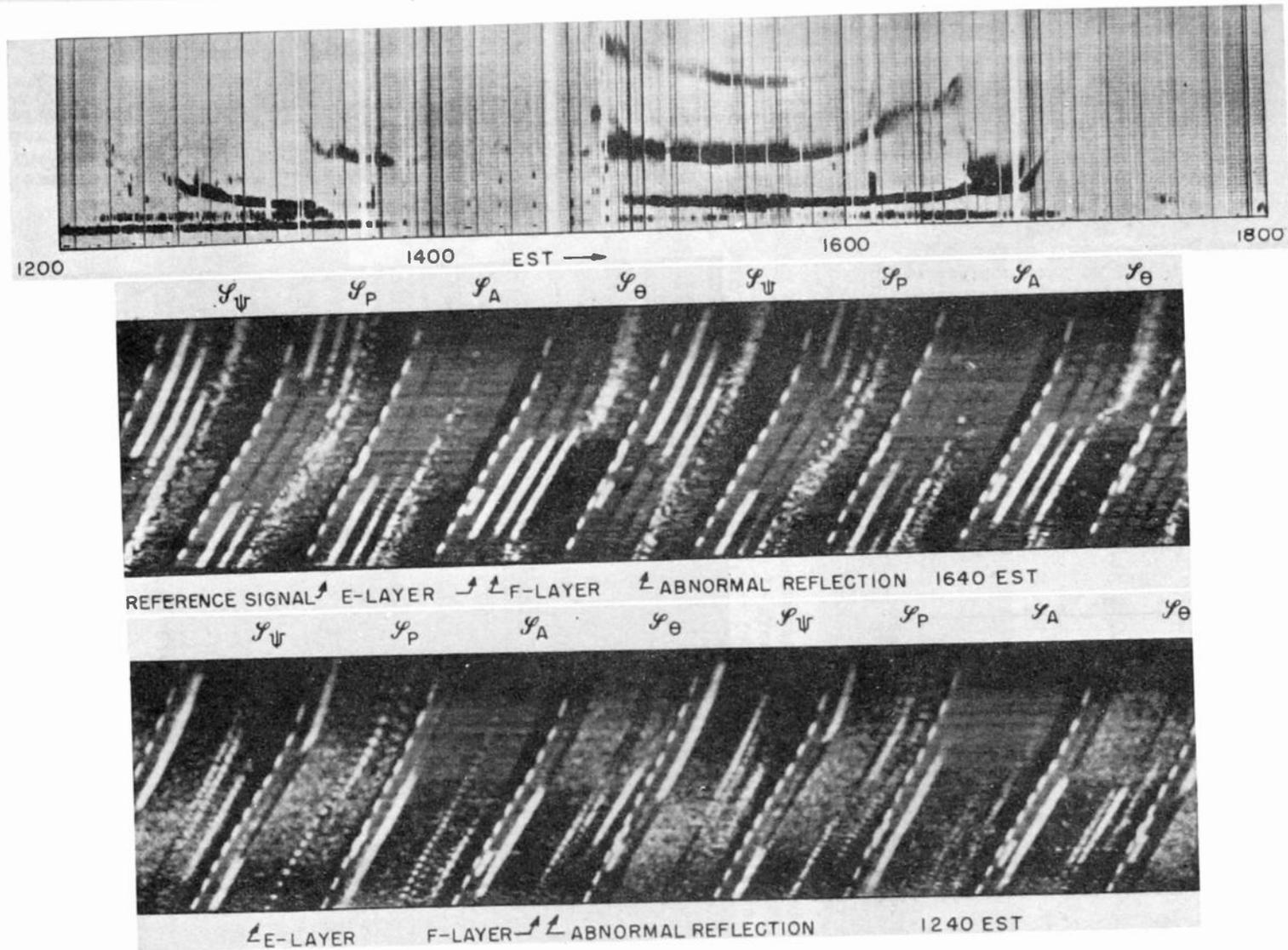


Fig. 7

Time-delay and nearly simultaneous 35mm records on a magnetically disturbed day, Sept. 25, 1951. Note the difference in ϕ_{θ} between the abnormal reflections and the normal E and F-layer signals. Also the diffuseness of the abnormal signal at 1640, and the rapid fluctuations of the polarization of the E- and F-layers at 1240.

The disappearance of all signals between 1350 and 1500 is due to high absorption, and not due to loss of synchronization.

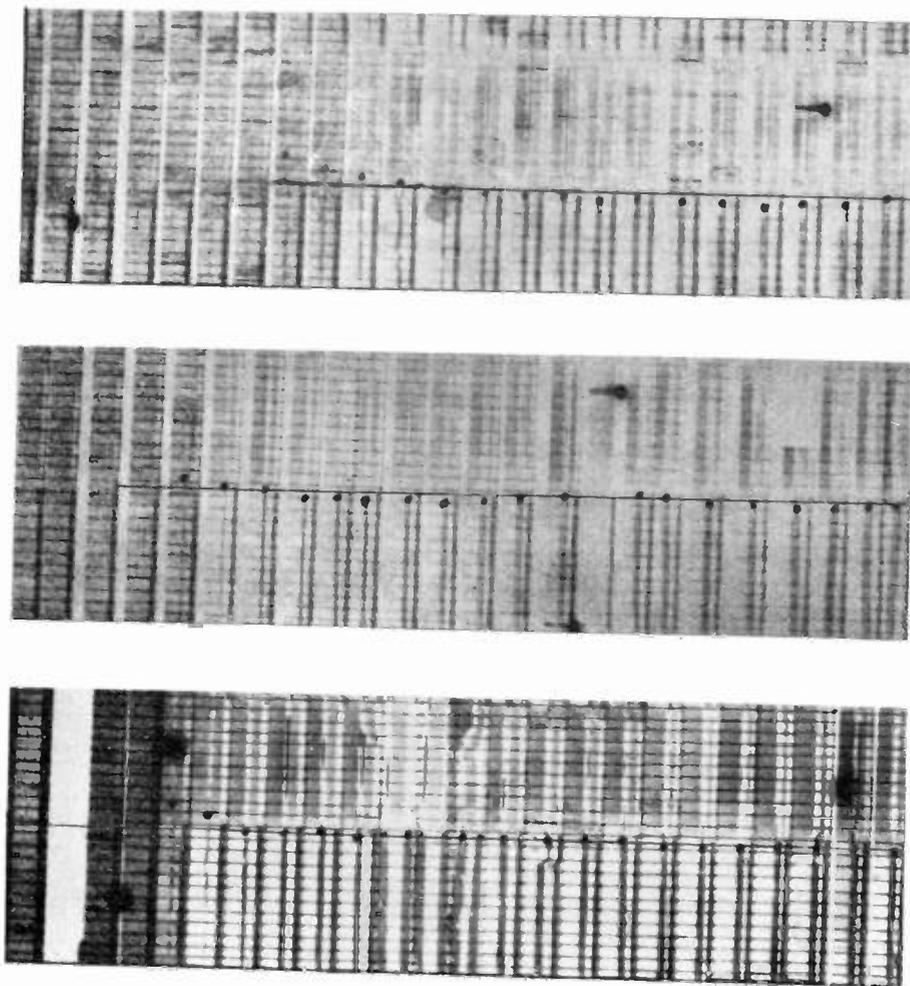


Fig. 8
Three examples of the deviation of the horizontal angle of arrival to the North-West as the F-layer signal dies at night. Note the irregular nature of the signal at these times, and the short interval of time during which this deviation occurs. Each division represents approximately 3° of azimuth.

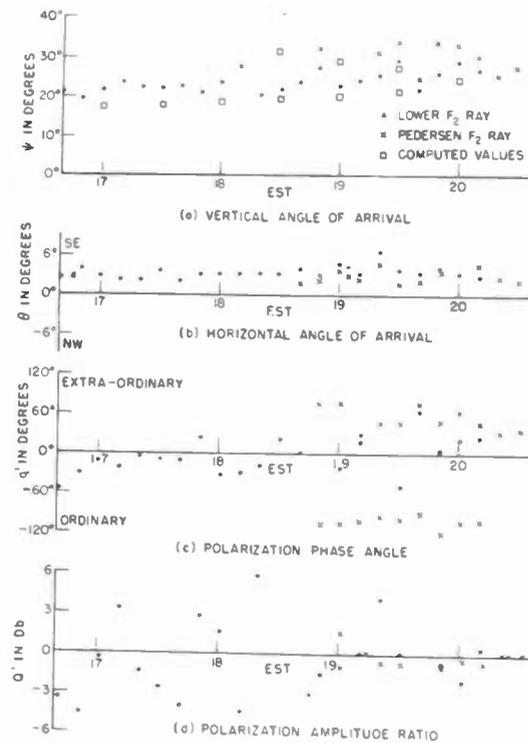


Fig. 9
Simultaneous direction of arrival and polarization data on the 9.1 mc/s signal on the Glenville-to-Lexington path (Nov. 19, 1951).

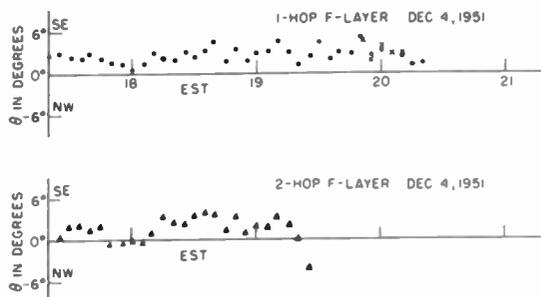
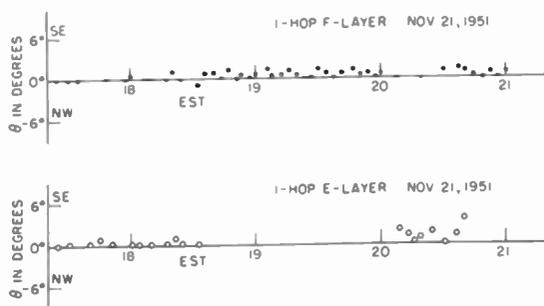
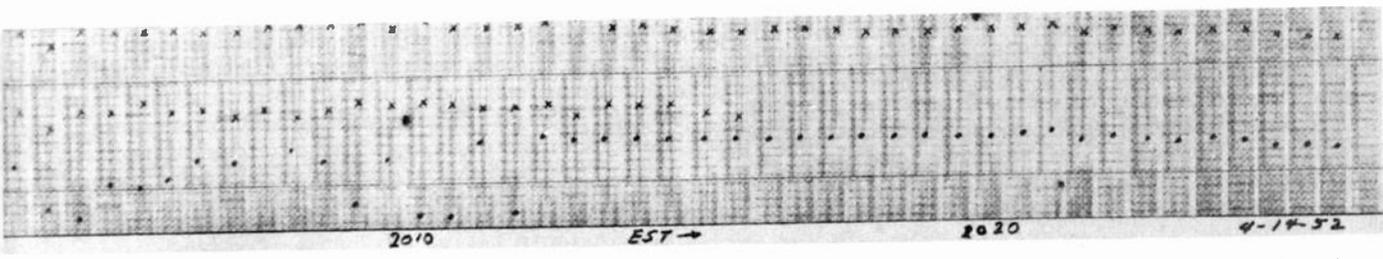
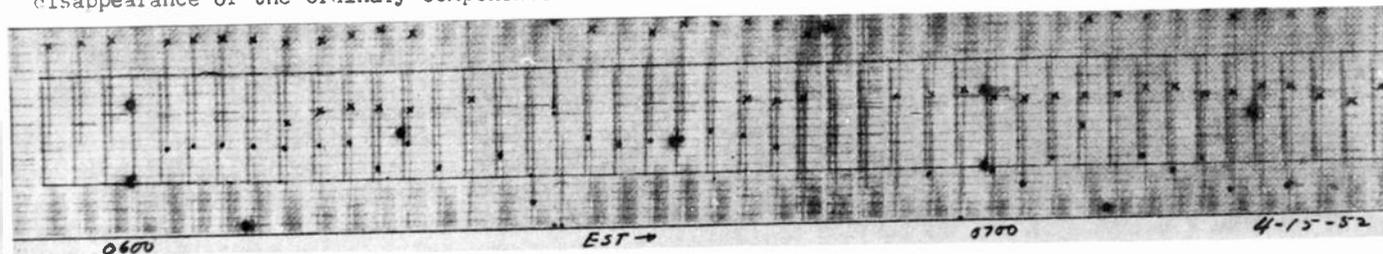


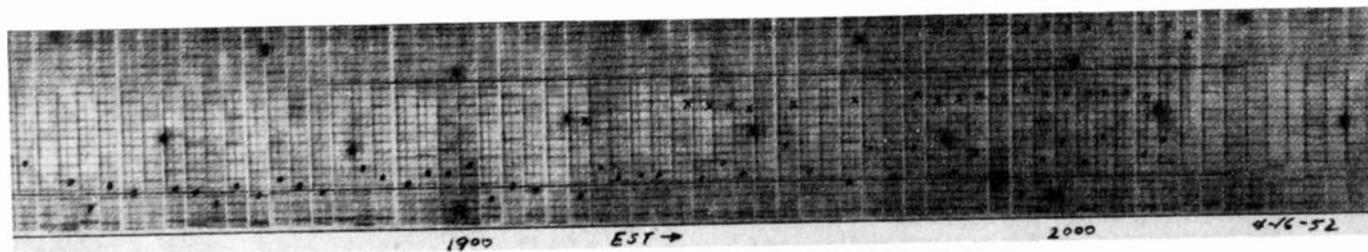
Fig. 10 - Evening horizontal angles of arrival of the 9.1 mc/s signal.



(a) The upper horizontal inked line is the zero phase-angle reference for the Pedersen F-layer signals, and the lower line, the reference for the lower F-layer signal. Each marker indicates 15° of phase. Readings are taken every $\frac{1}{2}$ minute on this record, or four times as frequently as on the others. Note disappearance of the ordinary components at 2015 while the extra-ordinary components remain until 2025.



(b) This is a morning build-up record showing the extra-ordinary ray starting at 552. The splitting of the signal into lower and Pedersen branches is first discernible at 601, and the ordinary appears between the two extra-ordinary signals at 611. A few minutes later, the two magneto-ionic components of the lower F-layer signal are practically coincident, and phase interference causes the phase-angle reading to become highly variable.



(c) Again, note the variability of the F-layer signals, and the comparative steadiness of the Pedersen F-layer signals. Also, the greater intensity of the ordinary Pedersen ray as compared to the extra-ordinary component.

• Lower F-layer signal X Pedersen F-layer signals

Fig. 11
Polarization phase-angle records taken on photographic paper.

FCC RULES AND PROPAGATION DATA

E. W. Allen, Chief Engineer
Federal Communications Commission
Washington 25, D. C.

Summary -- The engineering which is incorporated into the FCC Rules is designed to provide a satisfactory administrative tool for station assignment and licensing purposes. With the rapid growth of television, it was found necessary to include in the Rules a national assignment plan based upon available propagation and equipment data and upon estimates of average service areas. The data and procedures used for the development of the mileage spacings employed in the assignment plan; for the development of the propagation curves; and for the estimates of field strengths required for various grades of service are examined and discussed. An appraisal is made of present progress in UHF and of needed technical developments to permit the UHF to provide the necessary service.

Background

Subsequent to World War II the growing demands for television service made it necessary for the Federal Communications Commission to provide for the assignment of additional television channels to the various cities of the United States, so that opportunities would be provided for the establishment of additional television stations. Extensive public hearings beginning in May 1948, revealed that the potential demands for new television stations could not be met by the assignment of only the 12 channels available in the VHF portion of the radio spectrum. It therefore became necessary to provide for the allocation and assignment of 70 additional channels in the UHF portion of the spectrum. A national assignment plan, including both VHF and UHF channels, was developed and was issued by the Commission in April 1952.¹ Under the plan, specific channel assignments were made to each of some 1300 cities and towns, which can accommodate about 600 VHF stations and 1400 UHF stations.

In making these assignments, certain spacings between them were maintained in order to provide for each station a reasonable freedom from interference from stations located in other cities. These spacings were determined from estimates of the probable future service areas of television stations, as indicated by the existing knowledge of the characteristics of VHF and UHF radio waves and by the existing and expected developments in television transmitting and receiving equipment.

In order to obtain propagation and engineering information on which assignments could be based, the FCC called an Engineering Conference in November 1948. From the conference there was formed an Ad Hoc Propagation Committee, which examined the available VHF propagation data and developed methods for portraying it in a form

useful for assignment purposes, defined methods for specifying grades of television service, and made studies to show how effective use could be made of the channels available for television assignments.² The studies of this Committee were confined to the VHF, so that it became incumbent upon the staff of the FCC to assess the available UHF information, and to put it into summary form, so that it would be useful in the preparation of a coordinated VHF-UHF assignment plan and in the formulation of assignment and licensing rules, under which the plan could be placed in operation.

UHF Propagation Data

An analysis of four UHF surveys made in the vicinity of New York City and Washington, D. C., and available at the time of the issuance of the Commission's Notice of Further Proposed Rule Making on July 12, 1949 indicated that the median fields for distances out to about 20 miles from the transmitters could be expected to be substantially equal to the median fields in the lower VHF. A later and more comprehensive study of 18 UHF surveys, as shown in Table I, containing some data out to a distance of about 30 miles, verified this figure³ so that the Commission's Rules adopted in April, 1952 provided for the use of the same median field strength curves for the lower VHF channels 2-6 and the UHF channels 14-83.

TABLE I

Median Ratios of Plane Earth Fields to Measured Fields in db.				
1	2	3	4	5
Survey	Bu	F	Survey Medians	Radial Medians
Kansas City	Midland	507	19	12-20
Wash., D. C.	Philco	505	22	
Wash., D. C.	RCA	505	23	17-34
New York City	DuMont	612	28	27-31
New York City	RCA	510	20	16-26
New York City	CBS	490	16	
Harrisburg	McNary	515	23	12-31
Scranton	McNary	515	35	19-40
Easton	McNary	515	24	11-20
Reading	McNary	515	23	12-37
Cedar Rapids	Collins	410	19	
Pittsburgh	Westinghouse	503	27	26-30
San Francisco	Adair	600	24	
Bridgeport	Zenith	535	26	
Nashville	WSM	600	31	28-36
Ft. Wayne	Westinghouse	508	15	
Bridgeport	NBC	535	23	18-34
New York City	Bell Lab.	456	33	
Range			16-35	11-40
Weighted Median			22	

It would seem to be desirable to restate what is meant by "median field strength" within the context of the Commission's Rules and the studies which led up to them. The median field strengths shown by the curves in the Commission's Rules are the median values of measurements taken during the surveys of several stations. The curves cannot be expected to predict with accuracy the median field strength to be expected in a small area or even for the whole area covered by a single station. The lack of accuracy in the use of the median curves for predicting the coverage of individual stations is stated in Volume II of the Report of the Ad Hoc Committee and in Section 3.683 of the FCC Rules.

Reference to Table I shows this fact rather clearly. The weighted median of the ratios of plane earth to measured fields for all eighteen station surveys was 22 db. (Column 4) However, the range of median ratios for individual station surveys was from 16 to 35 db. Thus for complete station surveys, the median ratios ranged from 6 db above to 13 db below the median for all stations. Similarly, the median ratios for individual survey radials (Column 5) ranged from 11 to 40 db, or from +11 to -18 db compared to the overall median. Measurements for sectors of individual radials show an even wider spread, so that the probable error in the use of median curves to predict service contours for a particular station is quite high. More recent data, such as provided by Taylor at Portland, Oregon, and by Epstein and Peterson at Jersey City, fall within the ranges included in Table I, and the present curves as indicative of median values, can be said to be consistent with these data.^{4,5,6}

Some dissatisfaction has arisen as a result of the failure of the FCC curves to yield reliable predictions of individual station coverage. Perhaps some of the difficulties have arisen as a result of the failure to appreciate the nature and purpose of the FCC curves; namely, to provide a station assignment tool rather than a method for the detailed and accurate prediction of the coverage of individual stations. I should like to emphasize this distinction and to caution engineers to use all of the information at their command to assure themselves that both VHF and UHF television stations will cover the desired areas when placed in operation.

The recent investigation by Epstein and Peterson⁶ has shown that the probable error of predicting service contours can be measurably decreased by using diffraction formulae to estimate the effects of first order roughness such as shown on available topographic maps. Values are also given for the effects of residences and trees, so that more realistic estimates of the fields available at receiving locations can be made. Similar approaches to the treatment of systematic effects of terrain have previously been advocated by Bullington.⁷ In areas for which relief maps are available, shadowgraphs

such as described by Taylor⁴ should be used. Where conditions warrant, such refined predictions should be supplemented by measurements on existing or experimental transmitters in the area under study. The importance of such detailed studies to assure proper antenna siting can be realized when it is stated that for the worst radial examined in the above 18 surveys, about sixty times as much power is required in order to bring its median value up to the average for all radials. We at the Commission realize the seriousness of the difficulties confronting engineers in the solution of station siting and coverage problems, owing to the scarcity of dependable data and the pioneering state of measurement and estimation procedures. In an effort to assist, the Commission authorized its Chief Engineer to form the Radio Propagation Advisory Committee for the purpose of evaluating available data and recommending procedures and standards in these matters. This Committee consists of engineers from interested government agencies and from the radio industry. It welcomes the support of interested persons and also any information or data which will contribute to the solution of these problems. The Committee is concerned not only with the problems of coverage but also with the evaluation of the effects of interference from distant stations. I shall not discuss this latter matter, as it will be taken up in a succeeding paper by Mr. Herbstreit.

Field Strengths Required for Television Service

TABLE II

Grade B Service - Field Strengths Required to Overcome Receiver Noise -

Factor	(db above 1 microvolt per meter)		
	Channels 2-6	Channels 7-13	Channels 14-83
(1) Thermal noise	7	7	7
(2) Receiver noise figure	12	12	15
(3) Peak S/N ratio	<u>30</u>	<u>30</u>	<u>30</u>
(4) Receiver terminal voltage		49	52
(5) Transmission line loss	1	2	5
(6) Antenna factor	<u>-9</u>	<u>0</u>	<u>3</u>
(7) Local field strength	41	51	60
(8) 50% terrain factor	0	0	0
(9) 90% time factor	<u>6</u>	<u>5</u>	<u>4</u>
(10) Median field strength	47	56	64

In the absence of interfering fields, the field strengths required for television service are determined by the sensitivity and noise figures of available television receivers, and

the characteristics of available antennas and transmission lines. In developing the rules on which the station assignment plan was based, certain assumptions were made as to these characteristics for the typical television receiver and receiving installation which could be expected to be in use in the reasonably near future. These assumptions can be examined in some detail by reference to Table II from which were derived the field strength requirements for Grade B service⁸ which appear in the present rules.

(1) The thermal noise value, expressed in db above one microvolt at 300 ohms circuit impedance, is computed from the circuit temperature and is still valid.

(2) Receivers with a 12 db noise figure for channels 2-6 were being manufactured at the time of preparing the table in March, 1951. The assumptions for channels 7-13 and 14-83 were based on expected progress in receiver development. These values have been published for comment by the FCC on two occasions and have received general support. In a recent paper presented by Pan in Toronto⁹ the following ranges of noise figures were given for present receivers using grounded-grid circuits: channels 2-6, 4-7 db; channels 7-13, 6-10 db; channels 14-83, 12-21 db. It will be seen that some or all of the receivers in each group have lower noise figures than the figures assumed in Table II. However, in order to meet the assumption of 15 db for UHF receivers, the spread of noise figures must be greatly compressed so that the typical receiver approaches the characteristics of the best present receiver. The lesson to manufacturers is plain—build a UHF receiver with the lowest practical noise figure which the state of the art permits. Even then, UHF receivers will be inferior in this respect to VHF receivers.

Pan investigated the UHF receiver noise figures resulting from the use of available electron tubes as radio frequency amplifiers. He found that the planar type triodes resulted in a significant improvement in noise figure, 3-6 db, but they are too expensive for commercial use. The somewhat less expensive pencil triode gave some improvement over the best crystal mixer circuit at the lower end of the UHF band but none at the upper end. Available miniature triodes were unstable and were found to have up to about 2 db worse noise figures than the best crystal mixers. Thus there is urgently needed a new tube which can operate well as a low noise UHF amplifier and will be reasonable in cost. Such a tube would greatly assist in solving the problems of overloading the mixer circuit by strong signals and of reducing oscillator radiation.

(3) The 30 db peak visual carrier to R.M.S. noise ratio was chosen as a result of laboratory tests and observations which have been made independently at several laboratories. I believe that present judgment supports this value as an index of reasonable picture quality.

(4) Combining items (1), (2) and (3) yields the following values of required voltage in db above one microvolt across a 300 ohm receiver input, 49, 49, and 52 dbu.

(5) The transmission line was assumed to be 50 feet of 300 ohm line. Presently manufactured tubular lines of reasonable price can meet the assumed loss of 5 db at UHF even under wet conditions. Some open wire lines can do better by a db or two, and it would seem that this is an item to which particular attention should be given in further development and in manufacturing, and in which care should be exercised by the service man in both choice and installation. It should prove to be a relatively cheap way of retrieving a few sorely needed db.

(6) The antenna factor, as listed in Table II, is the reciprocal of the conventional "effective length" of the antenna. Antennas of 6 db gain, compared to a half-wave dipole, are assumed for VHF channels 2-13 and of 13 db gain for UHF channels 14-83. The type of antenna assumed for the UHF is not stated, but this gain can be achieved by a rhombic of reasonable size. More recently other types of antennas, for example, double bow-tie antennas in corner reflectors, have been developed to yield about 13 db across the band. The assumptions do not seem to be unreasonable for rural conditions. However, the local fields may vary so widely over relatively small distances, that some pains must be taken to assure that the antenna is properly placed for best results. There are also some locations in which high antenna gain cannot be realized by reason of distortion of the field. It is unknown as to whether such locations are sufficiently numerous as to form a serious impairment to television service.

(7) Based on the foregoing assumptions, local field strengths at the receiving antenna of 41 dbu (41 db above 1 microvolt per meter, which is equal to 112 uv/m), 51 dbu (354 uv/m) and 60 dbu (1000 uv/m) are required to produce a picture having a 30 db S/N ratio. Taylor, in his survey of UHF TV Station KPTV at Portland, Oregon,⁴ found that a 66 dbu (2 mv/m) local field provided a satisfactory picture. The antenna used was a bow-tie in a corner reflector, which provided about 4 db less gain than the above double bow-tie. The transmission line was in excess of 70 feet in length. Although the line loss and the noise figure of the receiver are not stated, we may assume that they were of good quality. Thus the acceptable values found by Taylor are reasonably consistent with the tabulated value of 60 dbu.

(8) The specification for Grade B service requires that the local field strength required for a satisfactory picture quality be available to 50% of possible receiver locations, so that no correction for terrain needs to be added in order to relate the local required field to the median or 50% terrain fields employed in the

field strength curves.

(9) The specification for Grade B service requires that the local field strength required for a satisfactory picture quality be available for at least 90% of the time at the best 50% of receiver locations at the outer limits of the service area. Available information on fading within line of sight indicates that it increases rather uniformly with increasing distance but is substantially independent of the frequency. The fading factor in Table II decreases with increasing frequency since the distance to the outer limits of the Grade B service area decreases. A recent reexamination of these figures in light of additional information indicates that the tabulated figures are somewhat generous, and could reasonably be reduced about 1 db to 5, 4 and 3 db, respectively.

(10) An overall appraisal of the factors which contribute to the required median field strengths for Grade B service indicates that, while lower field strengths might provide acceptable service at VHF, principally because of better receiver noise figures than the assumed values, the UHF field strengths cannot be lowered at this time. Further improvements in antennas, transmission lines and receiver noise figures must be made before this can be done.

The field strengths required in urban areas, in the absence of interference external to the receiver, were arrived at by a similar process, the principal differences residing in the assumptions of a typical urban antenna of 8 db gain, and of a 6 db terrain factor to provide service to 70% of receiver locations rather than 50%. Also an assumption was made as to the fields required to overcome noise and interference external to the receiver, which was likely to exist at typical urban locations, so that a final UHF value of 74 dbu was chosen as probably sufficient for both purposes. It has been quite true to date that the UHF television channels are fairly free of interference but it has been our experience that as services grow in the higher bands of frequencies, interference grows along with them in spite of concerted efforts to mitigate it. The median value of 74 dbu should be sufficient to override 30 to 45 dbu of sine wave interference, depending upon the location of the interference within the channel.

A 1 kilowatt transmitter with 60 db of harmonic suppression, and with the very conservative assumption of no antenna gain at the harmonic frequency, can produce a harmonic field of 43 dbu at one mile. Some current UHF television receivers can produce fields in excess of this value at 1000 feet. These figures reveal the importance of solving these problems at their sources and of maintaining the proper spacings between stations when their frequencies are such as to give rise to interference problems, and let me remind you that this problem is compounded

when receivers have numerous spurious responses. The UHF assignment plan is designed to minimize the effects of some of these and of radiation from the heterodyne oscillator when receivers employ 41 Mc IF strips and use the fundamental of the oscillator frequency. No protection is given to receivers of different design. Unless spurious responses and oscillator radiation can be properly controlled, a large percentage of any receivers having a nonconforming design will be in serious trouble when the UHF television and neighboring bands become more fully occupied.

UHF Transmitter Power Requirements

An examination of the Grade B service fields in Table II shows that the required UHF fields are 17 db higher than the fields required in the lower VHF. This corresponds to a power ratio of 50 to 1. Thus, even in favorable terrain or under conditions which will permit a choice of antenna site so that field strengths approximating those shown by the FCC curves are obtained over the area to be served, the UHF station will require much higher power to provide service areas and quality of service comparable to VHF stations. The majority of UHF stations now use 1 kw transmitters with antenna gains up to about 14 db. Some few have transmitters with power up to about 12 kw (11 dbk). With the high antenna heights which have been used in an effort to improve coverage, null fill-in, directionalizing and beam tilting practices have been resorted to in an effort to improve the near-in fields in populated areas. The development of higher transmitter power is essential to the successful operation of UHF television stations in competition with VHF stations. But in the course of developing higher power, let us not forget that this also holds the threat of higher levels of interference to television and other services which may be affected. Thus the development and adoption of adequate suppression measures for spurious and harmonic emission must proceed hand in hand with the development and use of higher power. Recently, suggestions have been made that booster stations, operating on the same channel as the parent station, or satellite stations, operating on different channels, be used to supplement coverage in difficult areas. If it develops that such operation should become widespread, changes in the channel assignment pattern may be required, involving closer spacing of transmitters. If so, it may prove to be impractical to give the added protection now found in the assignment plan with respect to oscillator radiation, image responses, intermodulation, etc. This means that a solution will depend upon improved receiver characteristics, including the development of a practical UHF amplifier.

Conclusion

While the progress to date in the develop-

the characteristics of available antennas and transmission lines. In developing the rules on which the station assignment plan was based, certain assumptions were made as to these characteristics for the typical television receiver and receiving installation which could be expected to be in use in the reasonably near future. These assumptions can be examined in some detail by reference to Table II from which were derived the field strength requirements for Grade B service⁸ which appear in the present rules.

(1) The thermal noise value, expressed in db above one microvolt at 300 ohms circuit impedance, is computed from the circuit temperature and is still valid.

(2) Receivers with a 12 db noise figure for channels 2-6 were being manufactured at the time of preparing the table in March, 1951. The assumptions for channels 7-13 and 14-83 were based on expected progress in receiver development. These values have been published for comment by the FCC on two occasions and have received general support. In a recent paper presented by Pan in Toronto⁹ the following ranges of noise figures were given for present receivers using grounded-grid circuits: channels 2-6, 4-7 db; channels 7-13, 6-10 db; channels 14-83, 12-21 db. It will be seen that some or all of the receivers in each group have lower noise figures than the figures assumed in Table II. However, in order to meet the assumption of 15 db for UHF receivers, the spread of noise figures must be greatly compressed so that the typical receiver approaches the characteristics of the best present receiver. The lesson to manufacturers is plain-build a UHF receiver with the lowest practical noise figure which the state of the art permits. Even then, UHF receivers will be inferior in this respect to VHF receivers.

Pan investigated the UHF receiver noise figures resulting from the use of available electron tubes as radio frequency amplifiers. He found that the planar type triodes resulted in a significant improvement in noise figure, 3-6 db, but they are too expensive for commercial use. The somewhat less expensive pencil triode gave some improvement over the best crystal mixer circuit at the lower end of the UHF band but none at the upper end. Available miniature triodes were unstable and were found to have up to about 2 db worse noise figures than the best crystal mixers. Thus there is urgently needed a new tube which can operate well as a low noise UHF amplifier and will be reasonable in cost. Such a tube would greatly assist in solving the problems of overloading the mixer circuit by strong signals and of reducing oscillator radiation.

(3) The 30 db peak visual carrier to R.M.S. noise ratio was chosen as a result of laboratory tests and observations which have been made independently at several laboratories. I believe that present judgment supports this value as an index of reasonable picture quality.

(4) Combining items (1), (2) and (3) yields the following values of required voltage in db above one microvolt across a 300 ohm receiver input, 49, 49, and 52 dbu.

(5) The transmission line was assumed to be 50 feet of 300 ohm line. Presently manufactured tubular lines of reasonable price can meet the assumed loss of 5 db at UHF even under wet conditions. Some open wire lines can do better by a db or two, and it would seem that this is an item to which particular attention should be given in further development and in manufacturing, and in which care should be exercised by the service man in both choice and installation. It should prove to be a relatively cheap way of retrieving a few sorely needed db.

(6) The antenna factor, as listed in Table II, is the reciprocal of the conventional "effective length" of the antenna. Antennas of 6 db gain, compared to a half-wave dipole, are assumed for VHF channels 2-13 and of 13 db gain for UHF channels 14-83. The type of antenna assumed for the UHF is not stated, but this gain can be achieved by a rhombic of reasonable size. More recently other types of antennas, for example, double bow-tie antennas in corner reflectors, have been developed to yield about 13 db across the band. The assumptions do not seem to be unreasonable for rural conditions. However, the local fields may vary so widely over relatively small distances, that some pains must be taken to assure that the antenna is properly placed for best results. There are also some locations in which high antenna gain cannot be realized by reason of distortion of the field. It is unknown as to whether such locations are sufficiently numerous as to form a serious impairment to television service.

(7) Based on the foregoing assumptions, local field strengths at the receiving antenna of 41 dbu (41 db above 1 microvolt per meter, which is equal to 112 uv/m), 51 dbu (354 uv/m) and 60 dbu (1000 uv/m) are required to produce a picture having a 30 db S/N ratio. Taylor, in his survey of UHF TV Station KPTV at Portland, Oregon,⁴ found that a 66 dbu (2 mv/m) local field provided a satisfactory picture. The antenna used was a bow-tie in a corner reflector, which provided about 4 db less gain than the above double bow-tie. The transmission line was in excess of 70 feet in length. Although the line loss and the noise figure of the receiver are not stated, we may assume that they were of good quality. Thus the acceptable values found by Taylor are reasonably consistent with the tabulated value of 60 dbu.

(8) The specification for Grade B service requires that the local field strength required for a satisfactory picture quality be available to 50% of possible receiver locations, so that no correction for terrain needs to be added in order to relate the local required field to the median or 50% terrain fields employed in the

field strength curves.

(9) The specification for Grade B service requires that the local field strength required for a satisfactory picture quality be available for at least 90% of the time at the best 50% of receiver locations at the outer limits of the service area. Available information on fading within line of sight indicates that it increases rather uniformly with increasing distance but is substantially independent of the frequency. The fading factor in Table II decreases with increasing frequency since the distance to the outer limits of the Grade B service area decreases. A recent reexamination of these figures in light of additional information indicates that the tabulated figures are somewhat generous, and could reasonably be reduced about 1 db to 5, 4 and 3 db, respectively.

(10) An overall appraisal of the factors which contribute to the required median field strengths for Grade B service indicates that, while lower field strengths might provide acceptable service at VHF, principally because of better receiver noise figures than the assumed values, the UHF field strengths cannot be lowered at this time. Further improvements in antennas, transmission lines and receiver noise figures must be made before this can be done.

The field strengths required in urban areas, in the absence of interference external to the receiver, were arrived at by a similar process, the principal differences residing in the assumptions of a typical urban antenna of 8 db gain, and of a 6 db terrain factor to provide service to 70% of receiver locations rather than 50%. Also an assumption was made as to the fields required to overcome noise and interference external to the receiver, which was likely to exist at typical urban locations, so that a final UHF value of 74 dbu was chosen as probably sufficient for both purposes. It has been quite true to date that the UHF television channels are fairly free of interference but it has been our experience that as services grow in the higher bands of frequencies, interference grows along with them in spite of concerted efforts to mitigate it. The median value of 74 dbu should be sufficient to override 30 to 45 dbu of sine wave interference, depending upon the location of the interference within the channel.

A 1 kilowatt transmitter with 60 db of harmonic suppression, and with the very conservative assumption of no antenna gain at the harmonic frequency, can produce a harmonic field of 43 dbu at one mile. Some current UHF television receivers can produce fields in excess of this value at 1000 feet. These figures reveal the importance of solving these problems at their sources and of maintaining the proper spacings between stations when their frequencies are such as to give rise to interference problems, and let me remind you that this problem is compounded

when receivers have numerous spurious responses. The UHF assignment plan is designed to minimize the effects of some of these and of radiation from the heterodyne oscillator when receivers employ 41 Mc IF strips and use the fundamental of the oscillator frequency. No protection is given to receivers of different design. Unless spurious responses and oscillator radiation can be properly controlled, a large percentage of any receivers having a nonconforming design will be in serious trouble when the UHF television and neighboring bands become more fully occupied.

UHF Transmitter Power Requirements

An examination of the Grade B service fields in Table II shows that the required UHF fields are 17 db higher than the fields required in the lower VHF. This corresponds to a power ratio of 50 to 1. Thus, even in favorable terrain or under conditions which will permit a choice of antenna site so that field strengths approximating those shown by the FCC curves are obtained over the area to be served, the UHF station will require much higher power to provide service areas and quality of service comparable to VHF stations. The majority of UHF stations now use 1 kw transmitters with antenna gains up to about 14 db. Some few have transmitters with power up to about 12 kw (11 dbk). With the high antenna heights which have been used in an effort to improve coverage, null fill-in, directionalizing and beam tilting practices have been resorted to in an effort to improve the near-in fields in populated areas. The development of higher transmitter power is essential to the successful operation of UHF television stations in competition with VHF stations. But in the course of developing higher power, let us not forget that this also holds the threat of higher levels of interference to television and other services which may be affected. Thus the development and adoption of adequate suppression measures for spurious and harmonic emission must proceed hand in hand with the development and use of higher power. Recently, suggestions have been made that booster stations, operating on the same channel as the parent station, or satellite stations, operating on different channels, be used to supplement coverage in difficult areas. If it develops that such operation should become widespread, changes in the channel assignment pattern may be required, involving closer spacing of transmitters. If so, it may prove to be impractical to give the added protection now found in the assignment plan with respect to oscillator radiation, image responses, intermodulation, etc. This means that a solution will depend upon improved receiver characteristics, including the development of a practical UHF amplifier.

Conclusion

While the progress to date in the develop-

ment of UHF television equipment is not as great as had been anticipated during the formulation of the television plan, nevertheless, continuing progress is being made. Discussions such as are to take place during this symposium serve to analyze the existing problems and their interrelation, and should be of assistance in serving to focus our efforts on the solution of the more urgent ones.

References

1. Federal Communications Commission; Sixth Report and Order in dockets 8736, 8975, 8976, and 9175; Federal Register, Part II; p. 3905; May 2, 1952.
2. Report of the Ad Hoc Committee; Federal Communications Commission; Vol. I; May 31, 1949; Vol. II; July 7, 1950.
3. H. Fine; UHF Propagation Within Line of Sight; Federal Communications Commission Mimeo. 65409; (T.R.R. 2.4.12) June 1, 1951.
4. J. P. Taylor; UHF in Portland; Broadcast News, Vol. 71; Sept.-Oct., 1952.
5. J. P. Taylor; More Reports on UHF Coverage; Broadcast News; Vol. 72; Jan.-Feb., 1953.
6. J. Epstein and D. W. Peterson; An Experimental Study of Wave Propagation at 850 Mc; Proc. IRE; Vol. 41, No. 5; pp 595-611; May, 1953.
7. K. Bullington; Radio Propagation at Frequencies Above 30 Mc; Proc. IRE.; Vol. 35, No. 10; pp 1122-1136; Oct., 1947.
8. Federal Communications Commission; Third Notice of Further Proposed Rule Making in dockets 8736, 8975, 8976 and 9175, Appendix B; Federal Register, Part II; p 3702; April 7, 1951.
9. W. Y. Pan; Investigation of UHF Amplifier Tubes; Unpublished paper presented at Fall Meeting of RETMA, RTMA (Canada) and IRE; Oct. 28, 1953.

PROPAGATION IN THE UHF-TV BAND

J.W. Herbstreit
National Bureau of Standards
Boulder, Colorado

ABSTRACT

The Central Radio Propagation Laboratory of the National Bureau of Standards has been conducting a program of research at frequencies of 418 and 1,046 mc in conjunction with an extensive 100 and 200 mc program of measurements throughout the United States. This research program has revealed many aspects of the frequency dependence of propagation from 100 to 1,000 mc, including the attenuation with distance and the magnitude of signal strength variations. Reception of 1,046-mc transmissions 400 miles from Cheyenne Mountain, Colo. has been found possible at all times. By far the most important factor determining the available

signal power available to the receiver in the uhf band is the effective absorbing area of the receiving antennas. This is illustrated by the use of the transmission loss concept in presenting the results of propagation studies. Transmission loss and its variability versus distance will be presented for a number of frequencies in the vhf and uhf bands as derived from the National Bureau of Standards propagation research program and an interpretation of the results will be given in terms of expected service and interference ranges in the uhf band.

OVERCOMING THE LINE-OF-SIGHT SHIBBOLETH WITH THE AIR AND HIGH POWER

Thomas J. Carroll
Massachusetts Institute of Technology
Cambridge, Massachusetts

1. Introduction

These observations on propagation well beyond the horizon as related to the new UHF television band are best introduced by an astonishingly prophetic quotation from the last paper of Marconi, after whom this session hall has been appropriately named. He delivered the paper on December 2, 1932. In the introduction, Marconi said:¹ "Electromagnetic waves under one meter in length are usually referred to as Quasi-Optical waves, the general belief being that with them communication is possible only when the two ends of the radio circuit are within visual range of one another; and, that consequently their usefulness is defined by that condition. Long experience has, however, taught me not always to believe in the limitations indicated by purely theoretical considerations or even by calculations, for these -- as we well know -- are often based on insufficient knowledge of all the relevant factors, but, in spite of adverse forecasts, to try out new lines of research, however unpromising they may seem at first sight." The paper itself describes Marconi's experiments in the UHF band off the Italian coast during the summer of 1932, which extended to distances of 168 miles by the way, despite the low powers of a few watts available in those pioneering days. On the last page of this remarkable paper is this prophetic paragraph: "In regard to the limited range of propagation of these microwaves, the last word has not been said. It has already been shown that they can travel around a portion of the earth curvature, to distances greater than had been expected, and I cannot help but remind you that at the very time when I first succeeded in proving that electric waves could be sent and received across the Atlantic Ocean in 1901, distinguished mathematicians were of the opinion that the distance of communications, by means of electric waves, would be limited to a distance of only about 165 miles." The reference is to a 1902 paper by Poincaré, the great French mathematician. We are now waking up to the fact that Marconi was right 22 years ago about the propagation of UHF considerably beyond the horizon, even as he had been right about transatlantic wireless in 1901, and the usefulness of short waves propagated via the ionosphere in the early nineteen twenties.² There is even a curious similarity between the time intervals it took theory to understand Marconi's original experimental discoveries about transatlantic propagation of long waves and his last work on beyond-the-horizon propagation of ultra-high frequencies. Although his transatlantic experiments were made in 1901, it was not until 1924 that it was proved to everyone that the ionosphere was probably responsible for the success of his 1901 defiance of the accepted line-of-sight views about radio

propagation at the turn of the century. Likewise at UHF, his experiments on propagation well beyond the horizon were made in 1932, and again more than two decades have passed before the reasons why they worked are beginning to be understood.

This paper proposes to give a simple account of the reasons for this weak and fading, but reliable, propagation of both VHF and UHF well beyond the horizon, and to indicate some of the probable implications for the future of UHF television, especially when the megawatt of effective radiated power which is permitted by the FCC rules goes on the air, although such powers are still technically "just around the corner". In my view, the word "bust" in the title of this symposium "UHF Television - Boom or Bust" is singularly inappropriate at the present time, especially from the viewpoint of the Professional Group on Antennas and Propagation. For a welcome change, the "new look" of the laws of nature concerning radio propagation is decidedly optimistic concerning the service which may ultimately be realized from high-power transmitters at points well beyond the line-of-sight. Of course, no one doubts that high power helps also within the horizon. A quotation of the nineteenth century humorist Artemus Ward is relevant: "It ain't so much what we don't know that gets us into trouble, it's what we know that ain't so." After almost a third of a century of effort, people in the middle thirties began to think too dogmatically that they understood fully and completely the theory of propagation around the curved earth, and the theoretical predictions were very gloomy about the prospects of useful propagation even from high-power sources much beyond line-of-sight. Alas, careful experimental checks were not carried out, although we now can see by hindsight that even the experiments of the early 1930's just after Marconi's did not fit the curved earth theory developed later. It now turns out that the oversight in these calculations, the things we knew that weren't so, lay in neglect of the air itself, its presence, its occasional abnormal stratification and perhaps, at times, its turbulence. It is both amusing and sobering to have to admit that so simple a thing as the presence of the earth's normal air as a dielectric envelope is one of the simple things overlooked even in a subject which has received as much theoretical attention as radio wave propagation. The air is a tenuous medium, of density about 1 gram per liter, but the total mass of the atmosphere is something like one-millionth of the mass of the earth, and many tons of gaseous dielectric are present above any radio path between terminals located in the

air ocean. Yet -- theoretical predictions about propagation beyond the line-of-sight have been repeatedly and overconfidently made by what is essentially an airless earth theory. It is now clear that the airless earth theory can be shown to be wrong for small signal levels by experiments performed in the early thirties, both well within and well beyond the horizon. The theory of diffraction around a curved earth was developed later, but never checked against the early experiments. The whole matter seems to be an excellent example of the necessity of forcing theory and experiment to check up on each other in scientific matters, each to prevent the other from believing things firmly that are not so. Frictions are undoubtedly generated when the disciplines of theory and experiment are commingled, but forcing them to spy on one another to detect errors is absolutely essential to the progress of science. This very field of radio wave propagation can serve as exhibit A of the dangers of letting the theorists and experimenters drift apart into separate compartments.

II. Mechanisms of Propagation into the Deep Shadow

Airless Earth

Even when the earth be idealized to a perfectly reflecting sphere with no air, it is by no means easy to calculate how radio waves diffract into the shadow of the earth bulge, even if the presence of the earth's envelope of air dielectric is completely neglected because the earth sphere is so large measured in radio wavelengths. But by the middle thirties, the problems were conquered of getting a numerical answer to the diffraction problem. Below the horizon the theory predicted a rapid exponential drop, running from about .55 db/mi at 50 Mc to 1.2 db/mi at 500 Mc. These figures take no account of the air except to use an earth radius of $4/3$ the true one, to allow for the downward refraction of the normal air. Under normal conditions this prediction of airless $4/3$ earth theory checks with experiment on the attenuation rates in db/mi below the horizon, but only for a few tens of miles until the field is of the order of -40 db or 1% of the free space inverse distance fields. Then at all frequencies the field strength distance curve becomes much less rapidly attenuated with distance, of the order of a few tenths of a db/mi at all frequencies: VHF, UHF, and higher. There is only a slight increase in attenuation rate from VHF to UHF. Furthermore, the absolute value seems roughly the same fraction of the free space field for all frequencies. While these weak fields fade at a rate up to a few cycles per second, there has not yet been reported any difficulty in their ability to carry information. Numerous observations indicate that signal-to-noise ratio is the main problem. Evidence can even be found in the literature prior to the last war for something missing in airless earth theory in experiments done over flat fields or water surfaces at distances up to a mile or so, where the measured fields were

sufficiently weak compared to free space to show up the fact that something was missing from the airless earth calculations. Since the really high power transmitters of FM and TV stations have come on the air, measurements have accumulated continuously all over the world showing the inadequacy of the predicted fields beyond the horizon at all times and places. The best and almost the only high power measurements beyond the horizon on UHF are those made by Gerks about five years ago and published in the November 1951 IRE Proceedings. If you wish to know what fields to expect more than a few tens of miles beyond the horizon at UHF, consult Gerks' curves, and by all means throw the $4/3$ airless earth diffraction calculations of the textbooks into the ash can. The textbooks give hopelessly too small an answer.

Uniform Air Layer

The easiest way to see what went wrong with airless earth theory is to add a layer of air to the earth in the simplest manner possible, as a layer which has the same constant density and index at all heights as it has at the surface, and would be therefore about 7.5 km thick. Radio waves proceeding outward as they leave a transmitting antenna in such a homogeneous air layer will be refracted to travel more nearly horizontally as they pass into the vacuum above the air. There will be also, however, downward reflected waves associated with the air-vacuum transition. This reflection may be calculated from Maxwell's equations by a process of matching solutions at the boundary, but it is more meaningful to consider the whole effect as a result of the presence of the polarizable air molecules, oxygen, nitrogen and water vapor principally, throughout the volume of the air layer. These molecules become each a tiny retransmitter of coherent radio waves under the polarizing influence of the wave from the transmitter, and it is the coherent addition of the primary wave and all the secondary waves radiated by the elements of the dielectric which are responsible for three things: (a) the slowing down of the speed of the resultant wave in the air by 3 to 4 hundredths of a percent; (b) the changed direction and speed of the outgoing wave in vacuum above the air; (c) and the weak downcoming reflected wave in the air connected with the inhomogeneity of space which is only partly filled with air. This homogeneous layer model of the troposphere can be calculated to reflect rather more signal than is usually observed beyond the horizon, but it serves well to illustrate what is omitted in the textbook theories of normal propagation, which takes no account of the air except for its refractive bending effect embodied in the trick of using a fictitious $4/3$ earth radius instead of the real one. Whenever an electromagnetic wave is forced to propagate from one medium to another there are inevitable reflection effects as well as refraction effects, and the reflective effects are entirely neglected in conventional $4/3$ earth theory.

Non-uniform Layers

Actually, under the influence of the gravity and the gas laws, the earth's air layer constitutes a non-uniform layer, in which the index of refraction decreases gradually to unity at great heights, rather than sharply at 7.5 km. Because air is a compressible gas, the pressure and density of the gas must be greatest near the surface where the air must bear the weight of all the air above. This non-uniformity of the normal atmosphere complicates the calculation greatly, but does not alter the principle illustrated by our earlier example of propagation into vacuum from our uniform homogeneous layer. A wave trying to escape to outer space by propagating through such a stratified inhomogeneous layer suffers feeble partial reflections as well as refractive bending of the major portion of the energy which escapes. This effect is enhanced for waves passing through at small glancing angles especially

Leaky Modes and the Inhomogeneous Air Layer³

Propagation in an inhomogeneous medium in which the velocity of propagation changes with height, is a complicated matter, just because Maxwell's equations for such a medium do not permit the propagation of progressive sinusoidal waves. Although the air is only slightly inhomogeneous, we must always remember that with high effective radiated powers and sensitive receiving systems commonly used nowadays, the question being asked is a very refined one indeed. The theory is being asked to calculate well beyond the horizon received powers of the order of 10^{-20} of that transmitted. The basic reason earlier theory is breaking down is its incapability of calculating the very small received powers which may now be transmitted and can be detected well beyond the horizon. When we recall that it took a third of a century to calculate diffraction around an airless earth, one might feel pessimistic about being able to say anything very exact about the effect of even the idealized inhomogeneous atmospheric envelope as well. Fortunately, the theory of how a linearly graded slab of dielectric surrounding a curved earth affects the allowed modes of propagation was rather exhaustively and exactly worked out by a group at Radiation Lab during the last war, in connection with quite another problem, the propagation of microwaves in surface ducts. Full account was taken of the complicated wave functions in the inhomogeneous air layer. If the normal air be idealized to be a linearly graded slab of dielectric, in which the index decreases linearly from the surface to vacuum value of unity around 30,000 ft., then a set of leaky modes appear which are transitional between the airless earth and the case of atmospheric ducts which would support unattenuated modes. These modes have attenuation rates in db/mi only about 1/3 to 1/10 of the lowest mode of the airless earth theory as we pass from 50 Mc to 3000 Mc. Thus the experimentally observed attenuation rates of a few tenths of a db/mile deep in the shadow region can be reliably attributed merely to the presence of the air layer. Furthermore, the

absolute value of the weak field supported by these modes whose airless earth leakage is somewhat suppressed by the presence of the air layer, is of the same order of magnitude as the observed average field out to about 400 miles at VHF, UHF, and SHF frequencies. In addition, these relatively slowly attenuated leaky modes supported by the idealized linearly graded air layer have the very desirable property of giving the same answer as the ordinary $4/3$ earth calculation within and just beyond the horizon, where only the refractive effect of the air layer is important. Thus, these new calculations to include the effect of an idealized air layer not only give about the right answer well beyond the horizon where the older theory give a hopelessly small answer, but also agrees with the old theory at closer ranges where the right answer can be approximately calculated neglecting the effect of the air. The conventional textbook story explains normal propagation within the horizon in terms of a direct and surface reflected wave. Just beyond the horizon, the field is described in terms of a single leaky mode, which attenuates rather rapidly because there is no upper boundary to prevent it. When the field has weakened to very roughly 1% of the free space field, then the effect of the earth's air dielectric envelope must be taken account, and the normal partial reflections which must occur just because the air layer is an inhomogeneous medium partially suppresses the outward leakage of energy, and causes the further attenuation of the weak and fading but omnipresent field to be of the order of a few tenths of a db/mi at all VHF and UHF frequencies.

Some authorities have distrusted this calculation at first because the atmosphere of the model is chopped off at 30,000 ft., and all air is neglected above this level. Recently, however, a few modes have been computed for models of the air with curved index profiles which decrease monotonically to vacuum at great heights, and have zero index gradient at the height where the waves emerge into the vacuum above. The attenuation rates and absolute values of these modes are of the same magnitude as the more completely known modes of the linearly tapered layer. The all important distinction is not in the shape of the index profile, but rather in the importance of making the proper distinction in the mathematics between the inhomogeneous air layer and the homogeneous void beyond.

III. Tropospheric Layers and Ducts

The idealized stratification we have just been discussing with the density and index of refraction decreasing gradually but smoothly upward according to a simple law, is essentially a stratification of the air caused by gravity itself. The air density and pressure is greatest at the surface because of the weight of the rest of the atmosphere above. Occasional peculiarities of the temperature and moisture distribution with height sometimes occur, and cause spectacular

increases in fields well beyond the horizon. If the index gradient over any height interval exceed 4 times normal, then this layer constitutes a so-called duct and if the wavelength is short enough, waves can be propagated between antennas located in or near such a duct almost as in free space, as if the earth were not there. The contribution of such stratification in the troposphere different from the normal gravitational stratification is capable of causing at times signals as high as free space and as low as the feeble guided wave levels due to the normal troposphere. Recent investigations of the atmospheric index structure with the airborne microwave refractometer have shown that these index stratifications aloft of appreciable magnitude may well be much more prevalent even in temperate zones than had hitherto been thought.

IV. Atmospheric Turbulence

About five years ago, when it was at long last realized that the measured fields well beyond the horizon were much stronger on all frequencies at all times and places than could be explained by $4/3$ airless earth diffraction calculations, or by any reasonable omnipresent ducts or elevated layers, there began to be discussed the possibility that scattering from blobs of turbulence in the upper atmosphere to distances well beyond the horizon might be part of the explanation of the observed omnipresent field not otherwise explained at the time. It is hypothesized that the index of refraction in a blob of air either a few centimeters or hundreds of meters in dimension may differ from moment to moment from the index of the surrounding air by a part in a million or less. This inhomogeneity would scatter out of a passing radio wave some energy which would not be so scattered if the air were smooth. One possible cause of such turbulence is just the blowing of the winds in our air ocean which we know is not idyllically static in time. Some rough estimates may be made of the amount of this turbulence from observation of the fluctuation in position and intensity of star images in astronomy. The difficulty in assessing the importance of this mechanism of turbulent scattering lies in the slender knowledge we have of the size and intensity of the blobs, their distribution with height especially in the upper air, and changes with time and season and geographical location. Much work is going on trying to determine by direct or indirect means just what contribution turbulent scattering makes to the field well beyond the horizon. It is important to realize that this contribution of turbulence is additive to the ones we have already discussed: the omnipresent effect of the presence of an idealized air layer, and the occasional strong effects of abnormal layers superimposed on the idealized smoothly tapered air layer. It is important to realize that even if there is no turbulence or no strong stratification other than that caused by gravity, then there still should exist a weak field well beyond the horizon which we can now calculate as well as measure, weak but not so hopelessly weak as to be useless, as we used to think erroneously. If turbulence is present of

sufficient strength to give a contribution stronger than this, so much the better. If strong super-refracting layers occur, that too is probably fine, except on the unlucky chance that they help the undesired signal preferentially above the desired signal. Psychologically, potential users of this kind of high power propagation well beyond line-of-sight should not embrace the misconception that the fields would vanish if for some reason the air high over the middle of the path should ever just quiet down and be free of turbulence and abnormal stratification. Short of the loss of the earth's air envelope, there would still be the normal air itself to give this weak propagation. It is a psychological help to realize that one does not have to rely on the refinement of imagining the air always to be turbulent in order to account for reception of a certain minimum weak signal. Actually loss of this minimum average signal would only occur if the earth's air envelope over the path were destroyed.

V. Implications for TV

These weak and fading, but omnipresent fields well beyond the horizon at UHF and VHF should hardly be imagined to contribute to the primary service area of a station, even when UHF stations are able to boost their effective radiated powers to the one megawatt level legally permitted them. Sufficiently large antennas and low noise figure receivers may be too expensive for average fringe area home receiving sites, but it is certainly too early to say how much people want TV and may be willing to spend to receive pictures well beyond line-of-sight. I am reminded of the strange and wondrous antenna rigs which dot the landscape atop homes well outside the primary service areas of American TV centers. One point should be mentioned in this connection. For propagation well beyond the horizon, antenna height is unimportant once the antenna is a few wavelengths above ground. The side yard is probably as good as the roof, and this factor may ultimately affect the willingness of the consumer to purchase antennas of perhaps a square meter or more of effective area at UHF. Initially at least, the first TV application of this kind of propagation well beyond the horizon will probably be in applications like community antenna systems, or satellite retransmissions on vertical polarization, or station-to-station relaying of programs. In all these applications it would seem that the cost of the big antennas and carefully engineered low-noise receiving systems might be economically justified, thus enabling these weak but omnipresent signals well beyond the horizon to be put to use. Such systems might be too expensive for most fringe area individual homes. In any case, the line-of-sight limitation which is usually thought to apply more harshly to UHF than to VHF is simply not so. With the higher power permitted to UHF, the advantage may ultimately prove to be the other way. Whether the fading of these weak fields will impose any bandwidth limitations on a wide band service such as television with its 4 Mc band requirement remains to be determined. Both the literature and

theory allow us to be optimistic about bandwidth, except at times when two or more mechanisms are operating with equal strength.

VII. Conclusion

From the viewpoint of our developing understanding of radio wave propagation, the second word in the title of this symposium "Boom or Bust" seems highly inappropriate at the present moment. Propagation theory is just now beginning to understand high power propagation of UHF as well as of VHF well beyond the horizon as a normal property of the troposphere, either its presence, occasional unusual stratification or turbulence, or a bit of all three. We now realize that Marconi was right in 1932 when he first observed the phenomenon, even as he was earlier in 1901 with his 1000 meter transatlantic transmissions and with short waves in the early twenties. It was the failure to check our theories by experiments and vice-versa which has held up our understanding of this for several decades. Direct test of some of the potentialities of high power UHF-TV transmissions has only become possible quite recently, and can be accomplished very soon. Actually, however, a new UHF broadcast service is being inaugurated with far better understanding of the laws of UHF propagation than was

available at the opening of AM broadcasting in the early twenties, or FM broadcasting in the thirties. The difficulties of equipment are similar to those which pioneers in any new band experience, all of which ingenuity and perseverance and energy can surmount. Propagation specialists now realize how wrong was the earlier pessimism about service beyond line-of-sight of high power transmitters both at VHF and UHF. The earth bulge is not as potent an impediment to the propagation as was once thought, because of its air envelope, the full effect of which theorists were careless enough to neglect. In a word, what had the earlier tropospheric propagation theory neglected? Answer: the troposphere!

References

- (1) G. Marconi, "Radio Communication by Means of Very Short Electric Waves" Proceedings, Royal Institution of Great Britain, vol. 27, pp. 509-544, 1932.
- (2) E. H. Armstrong, "The Spirit of Discovery, An Appreciation of the Work of Marconi" Elec. Eng., vol. 72, pp. 670-676, August 1953.
- (3) T. J. Carroll and R. M. Ring, "Normal Tropospheric Propagation of Short Radio Waves Well Beyond the Horizon" MIT, Lincoln Laboratory Technical Report No. 38.

A COMPARISON OF ANTENNA PROBLEMS AT UHF & VHF TV

Lloyd O. Krause
General Electric Company
Syracuse, N.Y.

Summary

Television transmitting and receiving antennas are components of a major intelligence transfer system only small portions of which are subject to man's control. The information is carried between antennas by propagating waves which are affected by the propagation media. The main propagation phenomena are considered briefly, but not in detail, to point up how wavefront variations occur.

These usually are greater at UHF than at VHF and make the large capture area UHF receiving antenna impractical in many cases. The design and effective application of receiving antennas is quickly discussed in the light of these problems.

UHF transmitting antenna design is then reviewed. It is approached from the standpoint of requiring an effective size near the VHF antenna, in order to partly offset the inapplicable receiving antenna. The resulting many wavelength UHF apertures then require new feeding techniques. These are briefly described. It is also pointed out how the resultant high gains require care in application, with perhaps necessary contouring.

The manner in which the same FCC curves are used to predict coverage for channels 14-83 as for channels 2-6 despite differences in propagation characteristics is explained. A single set of measured UHF field intensities is shown.

In conclusion, adequate-signal UHF results in superb picture quality because of less selective distortion because of the smaller percent bandwidth. Man-made noise is almost nil. It is only necessary for the receiving end economically to approach as near VHF performance as the transmitting end has, and many of the UHF problems will vanish.

Introduction

The antennas used for transmitting and receiving television signals are an important part of a major intelligence system. The transmitting antenna launches the signal into space, to meet with all kinds of obstacles before it arrives at the welcoming receiving antenna. The receiving antenna gathers in the signal and delivers it to the receiver.

The antenna system comprises the transmitting transmission line, the transmitting antenna, the propagating media consisting of space and the earth, the receiving antenna, and the receiving trans-

mission line. All portions of this system perform differently at different frequencies. UHF frequencies are from 2-1/2 to 20 times as high as VHF frequencies. Naturally, then, some variations in performance are to be expected.

Some parts of this system are partially subject to man's control. Unfortunately, most of the system is not. Therefore, those parts that are need be given careful consideration for optimum performance. In the following some important factors in propagation and antenna elements of the system are discussed. The transmitting antenna system is considered in greatest detail, since this is one part most controllable by man. The variable factors between UHF and VHF are compared.

System Efficiency

It is the prime objective of the transmitting station to transfer as much energy as possible to as many receivers as possible. For practical reasons, this objective can be achieved only at extremely low power-efficiency. For example, suppose a station is delivering 1 mv peak at the terminals of a million receivers, each having an impedance of 300 ohms. The total peak power being delivered is 3.33 milliwatts. Such a station is probably radiating an actual peak power of at least 20 kw, so the overall r-f efficiency is only $.16 \times 10^{-4}\%$. However, this is not the way to look at the situation. Only when considered in the light of the efficiency of the transfer of intelligence is the circumstance seen to be an acceptable and practical one.

Nevertheless, this illustration does point out how important it is to capture as large a portion of the radiated energy as is possible with the receiving antenna. For practical reasons it is not possible ever to capture much of the energy. But every improvement made either in transmitting or receiving antennas is important. From the receiving antenna standpoint, the antenna should be made to have as large an effective capture area as possible. The transmitting antenna must concentrate the transmitted energy into as confined a beam as possible so that a given size receiving antenna can capture a larger portion of the transmitted energy. These matters will be discussed more later. Propagation factors influence UHF to a greater extent than VHF. Good antenna design assumes even greater importance, therefore, at UHF, though paramount design should always be used for best results.

The propagation factors are briefly

discussed to permit a better insight on their influence to the importance of antenna design.

Propagation Factors

An approach to the study of propagation usually is made beginning with the ideal case of free space. The other influencing factors of smooth earth, rough, statistical earth, diffraction, and refraction are then introduced as modifying factors. The free space case is an ideal never approached in any practical broadcast television antenna system. In the free space case the field intensity at any point is $E = 137.6 (kw)^{1/2} / d$ mv/meter, where kw is effective kilowatts being radiated toward the point, and d is in miles. That is, one kw of power from a dipole will produce 137.6 millivolts per meter maximum intensity at one mile. Every time the distance is doubled the signal drops to one-half, or 6 db.

For the free-space case, the ratio of received to transmitted power, P_R/P_t , is worth a brief discussion. The expression is:

$$\frac{P_R}{P_t} = \frac{G_R G_t}{1.52 \times 10^8 d^2 \lambda^2} = \frac{G_R G_t}{16.9 \times 10^2 d^2 F^2}$$

d = distance, miles F = frequency, megacycles

λ = wavelength, meters G_R and G_t are power gains over a dipole radiator.

The factor in the denominator on the right of the equation may be defined as the free-space attenuation factor between dipole antennas. Note that it is directly proportional to the square of the frequency. However, G_R and G_t are each proportional to frequency squared for a properly designed structure of fixed physical area. On this basis, the ratio of P_R/P_t would be proportional to frequency-squared, indicating better transmission at UHF than at VHF. However, the practical problem of maintaining the effectiveness of a certain sized physical area as the frequency is increased usually precludes such performance.

The free-space attenuation factor between dipole antennas may be written in logarithmic form as:

$$\text{Attenuation, db,} = 32.3 + 20 (\log d + \log F)$$

d = distance, miles
F = frequency, megacycles.

This free space signal is first of all modified by the reflected signal from the ground which arrives at the same point in different phase from the direct signal, and combines to give the total. The reflection coefficient of dielectric earth, for television frequencies may be considered as -1. The phase of this reflected signal depends on the heights of the transmitting and receiving antennas, and the distance between them, since this determines the path length difference. With these

factors taken into account, the total signal at a point in space is shown in Fig. 1, where the phase angle ϕ is equal to $4\pi h_1 h_2 / d$, resulting from a difference in path length of $2h_1 h_2 / d$ due to the antenna heights h_1 and h_2 separated a distance d. All units must be consistent.

Note that ϕ is directly proportional to frequency, and if pure smooth earth theory worked out in practice, the propagation at UHF would exceed that at VHF. This can be deduced from the fact that at UHF the receiving and transmitting antennas are effectively higher, since they are more wavelengths above ground.

It should be noted that the field resulting from the sum of the direct and reflected wave will be oscillatory about the free space value, exceeding it by 6 db on the crests, and going to minimum cusps of low value depending on the exact magnitude of the reflected wave. The oscillation will occur with height at a fixed distance, or will occur with distance at a fixed height. Usually, in the distances and heights of interest, the field intensity is being read on the downward slope of the last possible oscillation.

Besides the direct and reflected wave, there also exists a ground wave, which is caused by currents set up in the ground under the region of the receiving antenna by reflection of other direct waves at this point. Usually the effect of ground waves is quite negligible for receiving antennas more than a wavelength above ground.¹ Thus at all television frequencies with the usual assumed height of 30 feet for the receiving antenna, the ground wave effect is quite secondary.

Diffraction² causes the transmitted wave to tend to "flow" around the earth, as well as around other large objects in its path. The phenomenon of diffraction is caused by boundary currents, which set up new fields, and by redivergence of the main field propagating past the edge. The distance rate at which the total diffraction wave dies out apparently increases rapidly with frequency. Hence, especially at UHF, the diffraction wave dies out very quickly with distance beyond line of sight.

When the objects in the system become large compared to wavelength, they act more as shadows, with a smaller percentage of the total incident energy diffracting around the object due to edge currents and fields. Thus diffraction is not as effective at UHF in causing signal fill behind large obstacles.

The wave is not propagating through true free-space, or a vacuum. It is propagating through the atmosphere, which has an average dielectric constant greater than unity. Not only that, but this dielectric constant varies with height and time, as influenced by the water vapor content, as well as other foreign matter and temperature and pressure variation. The result is a change in re-

fractive index, which causes the wave to bend, just as a lens bends light.

On the average, the dielectric constant decreases with height, ever so slightly. Actually, a decrease in dielectric constant of 2.4×10^{-8} per foot of height causes the radio wave to bend on a radius equal to $4/3$ that of the earth. In other words, if the earth were $4/3$ larger, radio waves would remain parallel to it and follow it around due to refraction.¹ Sometimes the dielectric constant varies in sundry and anomalous fashion. Under certain conditions, this can cause ducting, and trapping, with unusual propagation results.

As mentioned earlier, large objects will cause shadows except for the effects of diffraction. Under certain conditions, diffraction over a sharp edge, such as that of a mountain, will result in less signal loss than would occur if the mountain were not there.

Besides these, there are the effects of absorption,³ in which the wave energy is dissipated in the medium as it passes through it, for example through a tree with leaves. At UHF this phenomenon is sometimes startling. However, it is outside the scope of this paper to analyze in detail the complex theory supporting the propagation phenomena discussed above.

Receiving Antenna Design

It is the job of the receiving antenna to accept the wave energy impinging upon it and deliver it with a minimum of distortion to the receiver, over the maximum possible bandwidth. The process is the inverse of transmission, but obeys the same basic laws. The receiving antenna must usually perform reasonably satisfactorily over a large number of channels. This performance includes both satisfactory impedance match and pattern shape.

Because the transmission line is usually short for a receiving antenna installation, the impedance match need not approach the high quality required for transmission. Nevertheless, the best possible is desirable to prevent signal distortion over the band. These problems are about equal at VHF or UHF, with UHF having a slight advantage because of a smaller percent bandwidth per channel.

Before the receiving antenna can deliver energy to the receiver, it must intercept it, collect it, and then deliver it. It is usually assumed that the receiving antenna is immersed in an ideal field for determining its characteristics. The ideal wave front is one that has uniform field intensity both in amplitude and phase over an area of about twice that of the effective capture area of the antenna. The type of field distribution assumed will radically effect the receiving pattern characteristic, as is well known from the reciprocal transmitting antenna. It is the fact that the receiving antenna very seldom actually is

or can be placed in the type of wave front for which it is designed that causes the single greatest difficulty in adequate antenna structures for UHF television.

The capture area of a receiving antenna is equal to:

$$A_R = \frac{G_R \lambda^2}{7.66}$$

G_R = power gain over dipole

λ = wavelength in meters

A_R = capture area in meters²

and the power delivered to a matched load from such a receiving antenna is

$$P_R = S A_R$$

P_R is the available power from the receiving antenna and is the power delivered to a matched load, watts

S is the plane wave power density, watts/meter², = $\frac{E^2 \times 10^{-6}}{120 \pi}$

A_R is as above

E = field intensity in mv/meter.

For the same type of antenna structures having the same gain, the capture area varies inversely as (frequency)². This means that a half-wave dipole at low channel VHF has about 100 times the capture area of a half-wave dipole in the mid-UHF range. Thus, it is capable of delivering 10 times the signal or 20 db more, to the receiver, for the same field strength. (Note that this agrees with the commonly accepted technique of making the delivered voltage proportional to the length of the dipole. However, the area concept is much broader in scope and application, and is to be preferred, since arrays of dipoles, or whatever type of structure being used, can then be compared on an effective area basis.) This difference in effective area is portrayed in Fig. 2, not to scale.

In order to have our UHF receiving antenna deliver the same signal to the receiver as our VHF antenna does, we must make it approach the VHF antenna in physical size. In the process of increasing the size, electrical as well as mechanical problems are encountered. The various elements in the array must be properly intra-connected as well as physically supported. If the horn-type approach is used then wind loading requires a heavy mechanical design. All this means added expense, and unfortunately, is no guarantee that any more signal will be delivered than would be delivered by a simple half-wave dipole. This is because the large area, or large aperture antenna, can produce the expected results only if it is excited by a uniform plane wave front, the condition for which it would normally be designed.

Suppose, as an example, we briefly and simply analyze a two element stack which, because of field distortion, has the lower and upper elements excited with arbitrary fields. (We shall neglect mutual impedance effects.) The array, with the applicable schematics for analysis, is shown in Fig. 3. By analysis, the power delivered to the load is:

$$P_L = \frac{(\bar{e}_1 \neq \bar{e}_2)^2 Y_a^2 Y_L}{2 (Y_a \neq Y_L)^2}$$

$$\text{If } Y_a = Y_L, \quad P_L = \frac{(\bar{e}_1 \neq \bar{e}_2)^2}{8Y_a}$$

$$\text{If } \bar{e}_1 = \bar{e}_2, \quad P_L = \frac{e^2}{2Y_a}$$

$$\text{If } \bar{e}_1 = -\bar{e}_2, \quad P_L = 0$$

On a pattern basis, the second case of $e_1 = -e_2$, would correspond to a pattern null in the particular direction from which such fields were arriving.

The effective application of a large aperture UHF receiving antenna requires a knowledge of the problems that field distortion present. Usually, such antennas may work fairly well in open spaces with good intervening terrain, because then there is some chance that the wave front is reasonably good. However, in city or other built up areas, where much of the signal may be due to diffraction and reflection effects, the chances of a suitable wave front are small, and the high-gain receiving antenna may be anything but. For this reason, backed by practical experience, most UHF antennas are not basically of high-gain, or large area, construction. Usually, an attempt is made to achieve only mediocre gain, perhaps by use of a simple mesh reflector. There is no doubt that a high-gain antenna would perform as the manufacturer advertises, except for difficulty in proper application.

To illustrate the effect of the capture area of the receiving antenna, a look at the free space equation will help:

$$\frac{P_r}{P_t G_t} = \frac{A_r}{19.8 \times 10^8 d^2}$$

where G_t = gain of transmitting antenna over a dipole.

P_t = transmitted power, watts

$P_t G_t$ = effective radiated power, watts.

A reduction in A_r requires an increase in $P_t G_t$ in the same proportion to maintain P_r . Hence it is very important that G_t be maintained proportionately higher at UHF than at VHF.

Transmitting Antenna Design

Since the UHF receiving antenna cannot usually

be given an effective area equal to that of a VHF antenna, it makes it all the more important that the transmitting antenna at least begin to achieve physical apertures approaching that of VHF transmitting antennas. As a fixed physical aperture is maintained and the frequency is increased, the gain of the antenna increases, and the beam becomes narrow. Thus a good UHF transmitting antenna requires proper illumination of a large physical aperture to produce the high gain; this in turn introduces some problems of application in that the narrow beam may require contouring for proper coverage. Fortunately, both these problems have been practically solved, and numerous successful UHF transmitting antenna installations are now scattered about the country.

The usual transmitting antenna is intended to radiate an omnidirectional pattern. This leaves only the vertical height of the antenna as a controllable dimension. For a uniformly illuminated vertical aperture the maximum theoretical power gain over a dipole is 1.24 times the aperture in wavelengths. The resulting vertical pattern has a $\sin x/x$ signal distribution, with a maximum side lobe level of 22%.

For practical reasons, this theoretical distribution is not obtained and the resulting gain is always less. Roughly, an achievement of slightly over 80% of the uniform illumination value is fairly good, which makes a practical gain equal to the aperture in wavelengths a good measuring stick. This is especially true as higher gains are strived for, since then smaller deviations from perfection result in net greater loss of gain. This is because the main beam becomes sufficiently narrow that a small increase in side lobe energy represents a fair percentage of the total energy.

Thus one of the big problems in design of UHF antennas having physical apertures approaching that of VHF antennas is obtaining proper current illumination of that aperture. There are two general approaches, that of a traveling current, and that of a resonant current system. The former has broader impedance bandwidth characteristics; the latter has broader phase bandwidth characteristics. A combination of the two basic methods may be used, with juggling of the relative impedance and phase bandwidths until adequate results are achieved. An important facet of the feed system used is the power handling capacity attainable in the limited physical dimensions of the structure. Present commercial UHF antennas on the market use both techniques to varying degrees. Sufficient literature is available on these structures so that further details need not be discussed here.

For a VHF antenna of adequate physical aperture, it is usually feasible to feed the energy over the aperture in a number of lumps small enough to permit individual cable feed to each element. This is a partial necessity, because of the greater percent bandwidth at VHF. Also, by using equal length cables, phasing is independent of frequency. However, very high gain antennas for VHF, especially

remains uniform once adequately established.

Conclusion

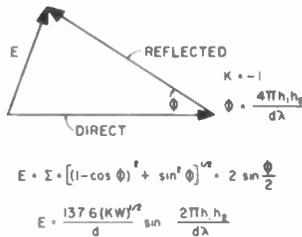
We have compared some of the antenna problems between VHF and UHF-TV. Little or nothing can be done about certain disadvantages of UHF propagation, except perhaps higher towers in some cases. It is difficult to install a receiving antenna of a large effective capture area at UHF. The UHF transmitting antennas do approach VHF transmitting antennas in effectiveness, and the practical limit in power gains has been achieved.

The greatest single improvement remaining to be done at UHF is a reduction in the receiver noise figure. When this can be done economically,

to permit receiving performance approaching that at VHF, most of the UHF coverage problems will cease to exist.

References

1. Kenneth Bullington, "Radio Propagation at Frequencies above 30 Megacycles," Proceedings of the I.R.E., October, 1947.
2. M. Katzin, "Ultra-High Frequency Propagation," Proceedings Radio Club of America, September, 1939.
3. E. A. Slusser, "Predicting Performance of UHF and SHF Systems," Electronics, June, 1951.



$$E = \Sigma \cdot \left[(1 - \cos \phi) + \sin^2 \phi \right]^{1/2} \cdot 2 \sin \frac{\phi}{2}$$

$$E = \frac{137.6 (KW)^{1/2}}{d} \sin \frac{2\pi h_1 h_2}{d\lambda}$$

Fig. 1

How the reflected wave effects the net field strength.

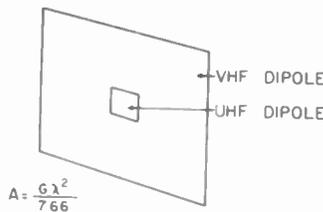


Fig. 2

Relative capture areas of vhf and uhf dipoles.

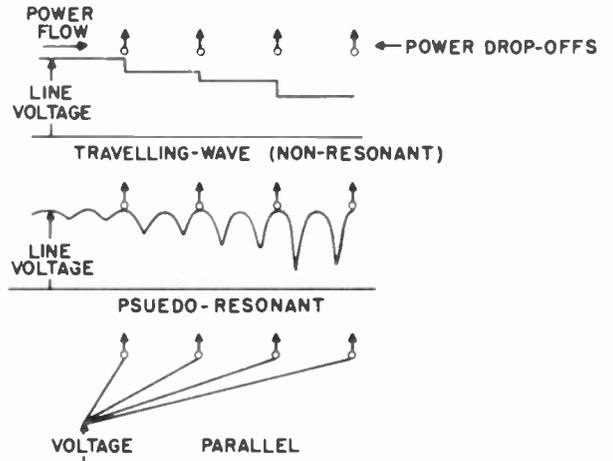


Fig. 4

The primary types of antenna feeds. Combinations may be used.

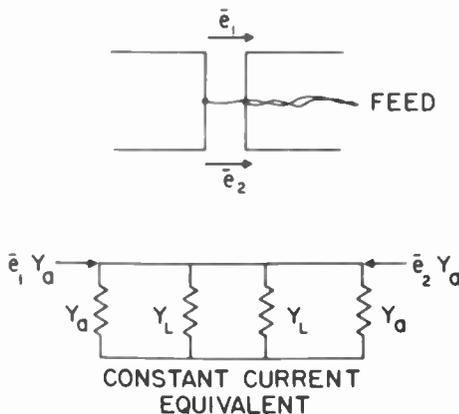


Fig. 3

A simple two-dipole receiving array with its circuit equivalent.

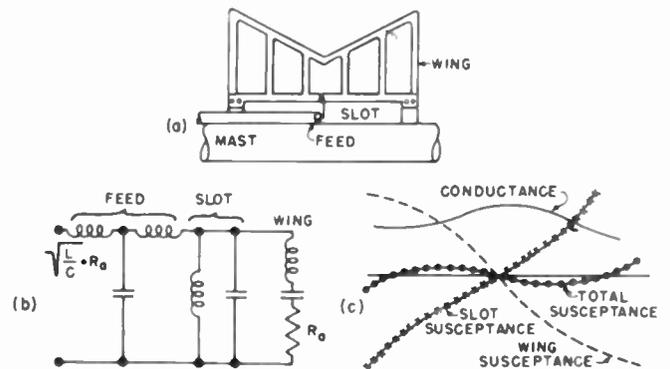


Fig. 5

The circuitry of a uhf batwing feed, showing susceptance compensations.

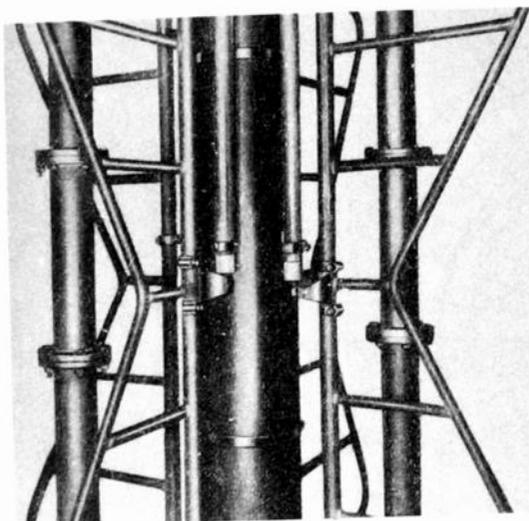


Fig. 6

Photograph of a typical batwing feed connection.

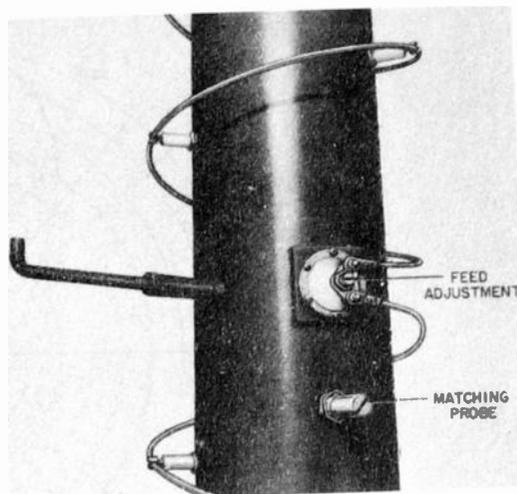


Fig. 8 - Photograph of a uhf helical antenna feed.

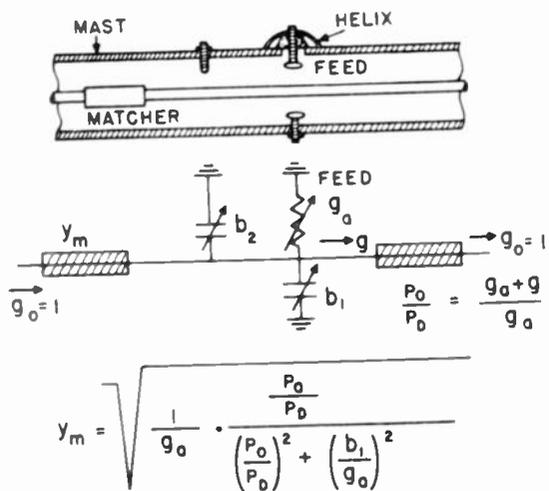


Fig. 7

Schematic of a typical coupling used in the helical uhf antenna with matching equation.

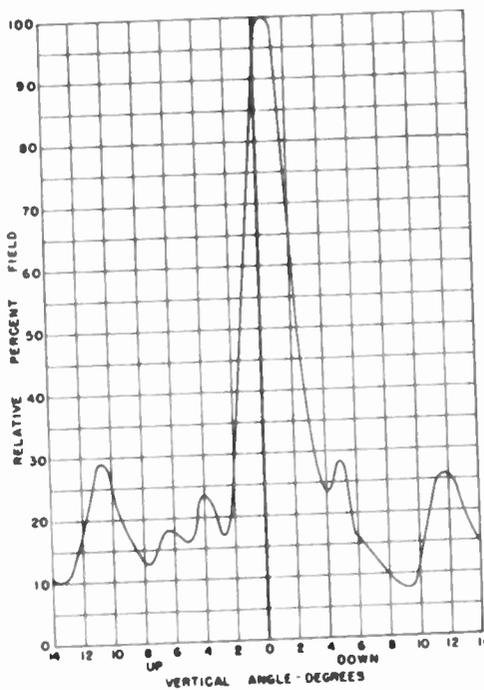


Fig. 9 - A contoured uhf antenna pattern.

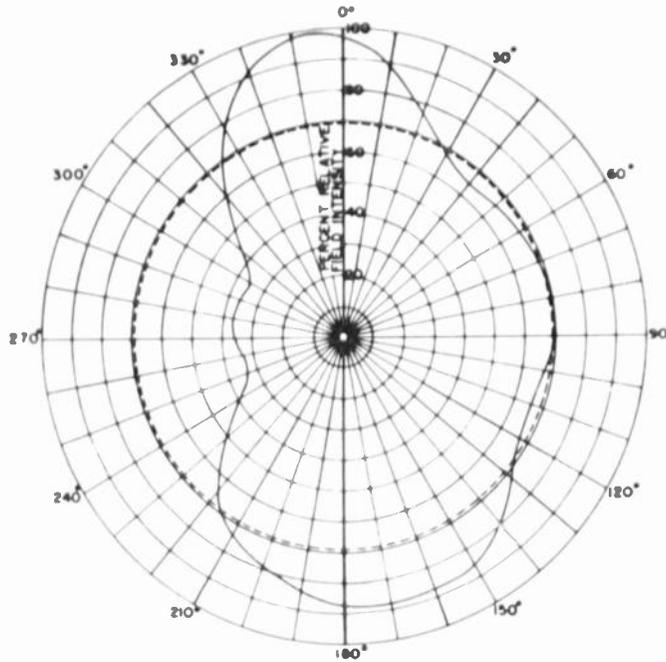


Fig. 10 - A directionalized uhf antenna pattern.

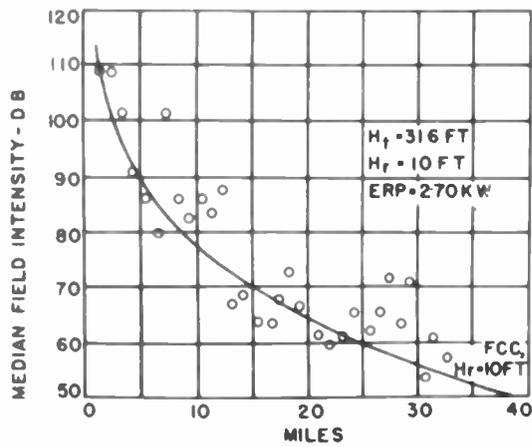


Fig. 11
 Typical median field intensity curve
 taken at WKAB-TV, channel 48.



