1962 IRE International Convention Record

PART 1

Sessions Sponsored by

IRE Professional Groups on

Antennas and Propagation

at

the IRE International Convention, New York, N.Y.

March 26-29, 1962

The Institute of Radio Engineers

1962 IRE INTERNATIONAL CONVENTION RECORD

An annual publication devoted to papers presented at the IRE International Convention held in March of each year in New York City. Formerly published under the titles CONVENTION RECORD OF THE I.R.E. (1953 & 1954), IRE CONVENTION RECORD (1955 & 1956), and IRE NATIONAL CONVENTION RECORD (1957, 1958, & 1959).

Additional copies of the 1962 IRE INTERNATIONAL CONVENTION RECORD may be purchased from the Institute of Radio Engineers, 1 East 79 Street, New York 21, N.Y., at the prices listed below.

Part	Sessions	Subject and Sponsoring IRE Professional Group	Prices for Members of Sponsoring Professional Group (PG), IRE Members (M), Libraries and Sub. Agencies (L), and Nonmembers (NM)			
			PG	M	L	NM
1	8, 16, 23	Antennas & Propagation	$ .70	$ 1.05	$ 2.80	$ 3.50
2	10, 18, 26, 41, 48	Automatic Control Circuit Theory	1.00	1.50	4.00	5.00
3	1, 9, 17, 25, 28, 33	Electron Devices Microwave Theory & Techniques	1.00	1.50	4.00	5.00
4	4, 12, 20, 34, 49	Electronic Computers Information Theory	1.00	1.50	4.00	5.00
5	5, 13, 15, 22, 29, 47, 54	Aerospace & Navigational Electronics Military Electronics Radio Frequency Interference Space Electronics & Telemetry	1.20	1.80	4.80	6.00
6	3, 11, 31, 35, 42, 45, 50, 52	Component Parts Industrial Electronics Product Engineering & Production Reliability & Quality Control Ultrasonics Engineering	1.40	2.10	5.60	7.00
7	30, 37, 43, 51	Audio Broadcasting Broadcast & Television Receivers	.80	1.20	3.20	4.00
8	7, 24, 38, 46, 53	Communications Systems Vehicular Communications	1.00	1.50	4.00	5.00
9	2, 19, 27, 32, 39, 40, 44	Bio-Medical Electronics Human Factors in Electronics Instrumentation Nuclear Science	1.20	1.80	4.80	6.00
10	6, 14, 21, 36	Education Engineering Management Engineering Writing & Speech	.80	1.20	3.20	4.00
		Complete Set (10 Parts)	$10.10	$15.15	$40.40	$50.50

Responsibility for the contents of papers published in the IRE INTERNATIONAL CONVENTION RECORD rests solely upon the authors and not upon the IRE or its members.

1962 IRE INTERNATIONAL CONVENTION RECORD

PART 1 - ANTENNAS AND PROPAGATION

TABLE OF CONTENTS

Page

Antennas - I
(Session 8: sponsored by PGAP)

Pattern Characteristics of an Antenna Focused in the Fresnel Region . . . . . J.J. Stangel and W.M. Yarnell 3
Polarization Tracking of Antennas. . . . . J.F. Ramsay, J.P. Thompson and W.D. White 13
Log Periodic Circuit Analysis (Abstract) . . . . . R. Mittra 43
Scanning Characteristics of Two-Reflector Antenna Systems . . . . . W.D. White and L.K. DeSize 44
Annular Slot Monopulse Antennas . . . . . K.C. Kelly and F.J. Goebels, Jr. 71

Propagation
(Session 16: sponsored by PGAP)

Radiation Characteristics of Slot Antennas Covered with a Plasma Slab in the Presence of a Static Magnetic Field Perpendicular to the Slot . . . . . H. Hodara 81
Trapped Waves in Varying Dielectric Media . . . . . J.E. Chin 99
Submerged Antenna Performance . . . . . W.L. Weeks and R.C. Fenwick 108
FM Backscatter Sounding as a Means for Monitoring Propagation Conditions During Short-Wave Broadcasts . . . . . R.B. Fenwick and O.G. Villard, Jr. 136
Precipitation Scatter as an Interference Source in Communication Satellite Systems . . . . . A.S. Dennis 145

Antennas - II
(Session 23: sponsored by PGAP)

ISCAN - Inertialess Steerable Communication Antenna . . . . . H. Brueckmann, J.R. Gruber, and C.A. Bramble 152
Superdirective Antenna Arrays for Improved VLF Reception (Abstract) . . . . . R.O. Schildknecht 164
The Coupling and Mutual Impedance Between Conical Log-Spiral Antennas in Simple Arrays . . . . . J.D. Dyson 165
On Random Removal of Radiators from Large Linear Arrays . . . . . T.M. Maher and D.K. Cheng 183
A Spacing Weighted Antenna Array . . . . . Y.T. Lo 191

PATTERN CHARACTERISTICS OF AN ANTENNA FOCUSED IN THE FRESNEL REGION*

John J. Stangel and William M. Yarnall
Surface Armament Division
Sperry Gyroscope Company
Division of Sperry Rand Corporation
Great Neck, L.I., New York

This paper is an expansion of the theory of stationary aperture focusing as developed by Wehner and Bickmore. The field of a circular antenna focused in the Fresnel region is approximated in terms of finite Hankel transforms of the aperture amplitude function. Although the peak gain of the focused antenna is comparable to the Fraunhofer gain of an equivalent uniformly illuminated aperture, the pattern characteristics undergo second-order perturbations. A series solution for the pattern through the focal plane is given for an amplitude distribution expressible as a power series. The effect of amplitude taper is studied.

Introduction

For years, the phenomenon of coherent antenna focusing at microwave frequencies was merely a scientific and mathematical curiosity without practical application. But recently, the anticipated requirements of space technology in a number of areas, including power transmission and communications, have stimulated renewed interest in the subject.[1] In addition, new procedures involving aperture focusing have been proposed for the convenient measurement of the radiation characteristics of modern, high-gain, large-aperture antennas.[2,3]

Aperture focusing is basically the technique of carefully controlling the phase on a radiating surface to accomplish optimum antenna efficiency in the near field. It is thus possible to achieve otherwise unattainable concentrations of r-f energy in space.

This paper is essentially an extension of the theory of stationary aperture focusing as developed by Wehner [4] and Bickmore. [5, 6] These authors have shown that, to a first order of approximation, the azimuth pattern through the focal plane is the same as the Fraunhofer pattern of an equivalent unfocused aperture. Starting with fundamental concepts of diffraction theory, this paper re-examines the characteristics of a focused circular antenna to evaluate the deviation from a perfect Fraunhofer analog.

Integral Formulation of the Diffraction Field

Let S represent the area of a plane, continuous, finite source in the x-y plane of the coordinate system depicted in Fig. 1, where the source point Q indicates an infinitesimal element of area in S. Then, the scalar diffraction field at the field point P due to the finite source is given by: [7]

$$U_p = \frac{1}{4\pi} \int F(\xi,\eta) \frac{e^{-jkr_1}}{r_1} \left[\left(jk + \frac{1}{r_1} \right) \hat{z}\cdot\hat{r}_1 + jk\, \hat{z}\hat{s} \right] dS \qquad (1)$$

where $\hat{z}$ is the unit vector is the z direction, $\hat{s}$ is the unit normal of the phase front at the aperture, k is the wave number, $2\pi/\lambda$, and the remaining constants are defined in Fig. 1. The complex aperture distribution, $F(\xi,\eta)$, consists of a phase term, $\Phi(\xi,\eta)$, and an amplitude term, $A(\xi,\eta)$:

$$F(\xi,\eta) = A(\xi,\eta)e^{-j\Phi(\xi,\eta)} \qquad (2)$$

If it is assumed that the distance from the aperture is large compared to the wavelength, then one may eliminate second-order terms in equation (1), and the field becomes:

$$U_p = \frac{jk}{4\pi} \int F(\xi,\eta) \frac{e^{-jkr_1}}{r_1} \left[(\hat{z}\cdot\hat{r}_1) + \hat{z}\cdot\hat{s} \right] dS \qquad (3)$$

We now make the usual Fraunhofer assumptions, that, as it affects intensity:

$$\frac{1}{r_1} \approx \frac{1}{r}$$

and as it affects phase:

$$r_1 \approx r - \sin\theta\,(\xi\cos\phi + \eta\sin\phi)$$

* The research reported in this paper was sponsored in part by the Air Force System Command, Rome Air Development Center, Griffiss Air Force Base, New York, under contracts AF 30(602)2307 and AF 30(602)2532.

and

$$\hat{z}\cdot\hat{r}_1 \approx \hat{z}\cdot\hat{r} = \cos\theta$$

where r, θ, and ϕ are the spherical coordinates of the field point, P. If we further substitute the polar coordinates ρ and ψ for the rectangular system (ξ, η), thereby implying an aperture of circular symmetry, the far-field pattern takes the form:

$$U_o = \frac{jk}{4\pi}\cdot\frac{e^{-jkr}}{r}\int_o^{2\pi}\int_o^a F(\rho,\psi)e^{jk\rho\sin\phi\cos(\psi-\phi)} \tag{4}$$

$$(\cos\theta + \hat{z}\cdot\hat{s})\rho\,d\rho\,d\psi$$

Furthermore, for an aperture distribution of uniform phase (i.e., $\Phi(\xi,\eta) = \gamma$ and $\hat{z}\cdot\hat{s} = 1$) and circularly symmetric amplitude (i.e., $A(\xi,\eta) = A(\rho)$, the above expression reduces to:

$$U_o = \frac{jk}{4\pi}\cdot\frac{e^{-j(kr+\gamma)}}{r}(1+\cos\theta)\int_o^{2\pi}\int_o^a \tag{5}$$

$$A(\rho)e^{jk\rho\sin\theta\cos(\psi-\phi)}\rho\,d\rho\,d\psi$$

But, by remembering that Bessel functions of the first kind may be formulated as definite intergrals by the following formula: [8]

$$J_\nu(x) = \frac{1}{2\pi j^\nu}\int_o^{2\pi} e^{jx\cos u}\cos(\nu u)\,du \tag{6}$$

where ν is the order of the function, and by remembering also that if g is a periodic function of u:

$$\int_\phi^{\tau+\phi} f(g)du = \int_o^{\tau} f(g)\,du$$

where τ is the period of g, one may readily express the rationship of equation (5) as a finite Hankel transform of the product, $\rho A(\rho)$:

$$U_o = \frac{jk}{2}\cdot\frac{e^{-j(kr+r)}}{r}(1+\cos\theta)\int_o^a \tag{7}$$

$$A(\rho)\,J_o(k\rho\sin\theta)\,\rho\,d\rho$$

Now, consider the solution of equation (3) for the field in the Fresnel zone. Then, we assume that $1/r_1 \approx 1/r$ as it affects amplitude, and $r_1 \approx r - \rho\sin\theta\cos(\psi-\phi) + 1/z\,\rho^2/2r$ as it affects phase. (According to Hansen and Bailin[9], the Fresnel approximation holds with reasonable accuracy to focal distances of $0.4\,a^2/\lambda$.) The field at P becomes:

$$U_\rho = \frac{jk}{4\pi}\cdot\frac{e^{-jkr}}{r}\int_o^{2\pi}\int_o^a$$

$$F(\rho,\psi)e^{jk\rho[\sin\theta\cos(\psi-\phi) - \rho/2r]} \tag{8}$$

$$[(\hat{z}\cdot\hat{r}_1) + (\hat{z}\cdot\hat{s})]\,\rho\,d\rho\,d\psi$$

We here observe that if focusing is to occur at a point P (r, 0, 0) on the z-axis of Fig. 2, the phase front in the vicinity of the aperture must form a segment of a sphere with center at P (r, 0, 0). For only then, according to geometrical optics, will waves emanating from all portions of the aperture attain a common phase at the focal point. To achieve this condition, it is necessary that the source points along a radius ρ in the aperture be advanced by a phase length s with respect to that at the orgin. From the sketch, it is immediately apparent that $(r+s)^2 = r^2 + \rho^2$, or that with the use of a binomial approximation, $s \approx \rho^2/2r$. Then:

$$\Phi(\rho) = -\frac{k}{2}\cdot\frac{\rho^2}{r} + \gamma$$

$$\hat{z}\cdot\hat{r}_1 = \frac{r}{r_1}\cos\theta \approx \cos\theta\left[1 + \frac{\rho}{r}\sin\theta\cos(\psi-\phi)\right]$$

and

$$\hat{z}\cdot\hat{s} = \left[1 + \frac{\rho^2}{r^2\cos^2\theta}\right]^{-1/2} \approx 1$$

Substituting these approximations into equation (8), the field becomes:

$$U = \frac{jk}{4\pi}\cdot\frac{e^{-j(kr+\gamma)}}{r}\int_o^{2\pi}\int_o^a$$

$$A(\rho)e^{jk\rho\sin\theta\cos(\psi-\phi)} \tag{9}$$

$$\left[(1+\cos\theta) + \frac{\rho}{r}\sin\theta\cos\theta\cos(\psi-\phi)\right]\rho\,d\rho\,d\psi$$

Comparison with equation (5) immediately shows that the field of a focused antenna has the form:

$$U_\rho = U_o + \delta U \tag{10}$$

where δU is a perturbation of the far field solution of the unfocused (uniform phase) aperture, U_o, and:

$$\delta U = \frac{jk}{4\pi} \cdot \frac{e^{-j(kr+\gamma)}}{r^2} \int_0^{2\pi} \int_0^a A(\rho) e^{jk\rho \sin\theta \cos(\psi - \phi)} \sin\theta \cos\theta \cos(\psi - \phi)\, \rho^2 d\rho\, d\psi$$

But, with the aid of equation (6) one obtains:

$$\delta U = \frac{-k}{2} \cdot \frac{e^{-j(kr+\gamma)}}{r^2} \sin\theta \cos\theta \int_0^a A(\rho) J_1(k\rho \sin\theta)\, \rho^2\, d\rho \qquad (11)$$

In general, for equations of the types shown as (7) and (11), the integral operator may be considered as defining a transformation of the amplitude function A (ρ) in Lebesque space where the kernel is a Bessel function. Then, if A (ρ) satisfies Dirichlet's conditions on the interval (0, a), the integral falls into the class of finite Hankel transforms. The properties of such transforms are studied in detail by Sneddon. [10] Depending on the nature of the amplitude function, it may be more convenient to redefine the problem for solution in an infinite domain. It may be necessary to replace the amplitude function by a discontinuous function, A' (ρ), which is defined as follows: A'(ρ) = A(ρ) in (0, a), A'(ρ) = 0 for ρ > a, and A'(ρ) = A' (-ρ). Then, the theory of infinite Hankel transforms, which is in a more advanced stage of development, may be applied. [11]

Antenna Gain

It is possible to draw some definite conclusions regarding the peak gain of a focused antenna without an explicit solution of the integral equations (7) and (11). We define the gain function as the ratio of the power density in a given direction to the average power density:

$$G(\theta, \phi) = \frac{4\pi P(\theta, \phi)}{P_T} \qquad (12)$$

where P (θ, ϕ) is the power radiated per unit solid angle in the (θ, ϕ) direction, and P_T is the total power radiated. P(θ, ϕ) is given by:

$$P(\theta, \phi) = \frac{1}{2}\sqrt{\frac{\epsilon}{\mu}}\, r^2 |U_p|^2 \qquad (13)$$

and the total power radiated is equal to the net power flow through the aperture [12] (i.e., the surface integral of the Poynting vector on the aperture):

$$P_T = \frac{1}{2}\sqrt{\frac{\epsilon}{\mu}} \int_0^{2\pi} \int_0^a |F(\rho, \psi)|^2 (\hat{1}_z \cdot \hat{1}_s)\, \rho\, d\rho\, d\psi \qquad (14)$$

If we assume, as previously, that $\hat{z} \cdot \hat{s} \approx 1$ for both the focused and the unforcused antennas, then, for the same power input, the net power radiated will be independent of focusing, and equation (14) becomes:

$$P_T = \pi \sqrt{\frac{\epsilon}{\mu}} \int_0^a |A(\rho)|^2 \rho\, d\rho \qquad (15)$$

In re-examing the results previously derived, we here note that the unperturbed (Fraunhofer) field of equation (7) is 90 degrees out of phase with the perturbation of relation (11). Then, using equation (10), we see that the absolute value of the radiation field has the following formulation:

$$|U|^2 = |U_o|^2 + |\delta U|^2$$

From equation (14), it follows that the gain of a focused antenna is expressible as the sum of an equivalent far-field gain plus a perturbation:

$$G(r, \theta, \phi) = G_o(\theta, \phi) + \delta G(r, \theta, \phi) \qquad (16)$$

where

$$\delta G(r, \theta, \phi) = \frac{2\pi}{P_T} \sqrt{\frac{\epsilon}{\mu}}\, r^2\, |\delta U|^2$$

or, substituting for δU:

$$\delta G(r, \theta, \phi) = \frac{\pi k^2}{8 r^2 P_T} \sqrt{\frac{\epsilon}{\mu}} \sin^2 2\theta \left[\int_0^a A(\rho) J_1(k\rho \sin\theta)\, \rho^2\, d\rho \right]^2 \qquad (17)$$

Evaluating this expression on the peak of the beam (i.e., θ = 0), one observes that $\delta G(r, 0, \phi) = 0$, and therefore the peak gain of an antenna is the same whether it is focused in the Fresnel zone or at infinity.

Series Formulation of the Amplitude Pattern

If the aperture distribution, $A(\rho)$, of equations (7), (11), and (14) is not readily expressible in closed form, or if its Hankel transform cannot be conveniently ascertained as a simple, analytic function, it may be possible to achieve a satisfactory solution by writing the aperture distribution as an infinite power series converging everywhere within the aperture boundary. Accordingly, we let:

$$A(\rho) = \sum_{m=o}^{\infty} a_m \rho^{2m} \tag{18}$$

where the a_m's constitute a set of constants satisfying Cauchy's convergence principle. Retention of only the even powers of p is consistent with the condition of circular symmetry.

We define the finite Hankel transform of order ν of a function $f(x)$ as the following integral:

$$\mathcal{H}_\nu [f(x),s] = \int_o^s f(x) J_\nu (x)\, x\, dx \tag{19}$$

Let $f(x)$ be represented by an infinite power series:

$$f(x) = \sum_{i=o}^{\infty} c_i x^i \tag{20}$$

Then:

$$\mathcal{H}_\nu[f(x), s] = \sum_{i=o}^{\infty} c_i \mathcal{H}_\nu[x^i, s] \tag{21}$$

Now consider the implications of the relatively common differential recurrence relation:[13]

$$d\left[x^{\rho+1} J_{\rho+1}(x)\right] = x^{\rho+1} J_\rho (x)\, dx \tag{22}$$

From this it follows that:

$$x^{i+1} J_\nu(x)dx = x^{i-\nu}\, d\left[x^{\nu+1} J_{\nu+1}(x)\right]$$

$$= d\left[x^{i+1} J_{\nu+1}(x) - (i-\nu)\, x^i\right]$$

And also, in a similar manner,

$$x^i J_{\nu+1}(x)dx = d\left[x^i J_{\nu+2}(x)\right] - (i-\nu-2)$$

$$x^{i-1} J_{\nu+2}(x)dx$$

And by further repeating this procedure, we may arrive at the series:

$$x^{i+1} J_\nu(x)dx = d\left[x^{i+1} J_{\nu+1}(x)\right]$$

$$-(i-\nu)\, d\left[x^i J_{\nu+2}(x)\right]$$

$$+(i-\nu)(i-\nu-2)\, d\left[x^{i-1} J_{\nu+3}(x)\right] - \cdots\cdots$$

$$\cdots + (-1)^{n+1}(i-\nu)(i-\nu-2)\cdots[i-\nu-2(n-1)]$$

$$d\left[x^{i-n+1} J_{\nu+n+1}(x)\right] + \cdots\cdots$$

Substituting this into the definition of the finite Hankel Transform of equation (19), we see that this series terminates if $(i-\nu)$ is either zero or an even integer (i.e., $i-\nu = 2N$, where $N = 0, 1, 2, \ldots\ldots$). Then:

$$\mathcal{H}_\nu(x^i,s) = s^{i+1} J_{\nu+1}(s) - (i-\nu)\, s^i J_{\nu+2}(s) +$$

$$+ (i-\nu)(i-\nu-2)\, s^{i-1} J_{\nu+3}(s) - \cdots\cdots$$

$$\cdots + (-1)^{\frac{i-\nu}{2}} (i-\nu)(i-\nu-2)\cdots 4.2\, s^{\frac{1}{2}(i-\nu+2)}$$

$$J_{\frac{i+\nu+2}{2}}(s).$$

for $\nu \geq 0$. Or,

$$\mathcal{H}_\nu(x^i,s) = \sum_{n=o}^{N} \frac{(-2)^n N!}{(N-n)!} s^{i+1-n} J_{\nu+n+1}(s) \tag{23}$$

We may now redefine the integral expression in equations (7) and (11) as finite Hankel transforms according to equation (19):

$$\int_o^a A(\rho) J_o(k\rho\sin\theta)\rho d\rho = \frac{1}{k^2\sin^2\theta}$$

$$\mathcal{H}_o[A'(x), k a \sin\theta]$$

and

$$\int_o^a A(\rho)\, J_1(k\rho\sin\theta)\, \rho^2 d\rho = \frac{1}{k^3 \sin^3\theta}$$

$$\mathcal{H}_1[x\, A'(x), k a \sin\theta]$$

where $x = k\rho\sin\theta$. Using the series nomenclature of equation (18) and the theorem of equation (21), we get:

$$\int_o^a A(\rho)\, J_o(k\rho\sin\theta)\,\rho d\rho = \sum_{m=o}^{\infty} \frac{a_m}{(k\sin\theta)^{2m+2}}$$

$$\mathcal{H}_o\left[x^{2m},\ k\sin\theta\right]$$

$$\int_o^a A(\rho)\, J_1(k\rho\sin\theta)\,\rho^2 d\rho = \sum_{m=o}^{\infty} \frac{a_m}{(k\sin\theta)^{2m+3}}$$

$$\mathcal{H}_1\left\{x^{2m+1},\ k\sin\theta\right\}$$

And, applying equations (23) it follows that:

$$\int_o^a A(\rho)\, J_o(k\rho\sin\theta)\rho d\rho = \sum_{m=o}^{\infty} a_m a^{2m+2} m!$$

$$\sum_{n=o}^{m} \frac{(-2)^n J_{n+G}(k a\sin\theta)}{(m-n)!\,(k a\sin\theta)^{n+1}} \qquad (24)$$

and,

$$\int_o^a A(\rho) J_1(k\rho\sin\theta)\rho^2 d\rho = \sum_{m=o}^{\infty} a_m a^{2m+3} m!$$

$$\sum_{n=o}^{m} \frac{(-2)^n J_{n+2}(k a\sin\theta)}{(m-n)!\,(k a\sin\theta)^{n+1}} \qquad (25)$$

These identities may now be employed to obtain series expansions for the field and gain expressions. Then, the Fraunhofer field of equation (7) becomes:

$$U_o = \frac{jk}{2}\cdot\frac{e^{-j(kr+\gamma)}}{r}(1+\cos\theta)\sum_{m=o}^{\infty} a_m a^{2m+2} m!$$

$$\sum_{n=o}^{m} \frac{(-2)^n J_{n+2}(k a\sin\theta)}{(m-n)!\,(k a\sin\theta)^{n+1}} \qquad (26)$$

and its related perturbation due to focusing in the Fresnel region is:

$$\delta U = \frac{-k}{2}\cdot\frac{e^{-j(kr+\gamma)}}{r^2}\sin\theta\cos\theta\sum_{m=o}^{\infty} a_m a^{2m+3} m!$$

$$\sum_{n=o}^{\infty} \frac{(-2)^n J_{n+2}(k a\sin\theta)}{(m-n)!\,(k a\sin\theta)^{n+1}} \qquad (27)$$

from equation (11). In addition, the perturbation of the gain in equation (17) becomes:

$$\delta G(r,\theta,\phi) = \frac{\pi k^2}{8 r^2 P_T}\sqrt{\frac{\epsilon}{\mu}}\sin^2 2\theta$$

$$\left|\sum_{m=o}^{\infty} a_m a^{2m+3} m!\ \sum_{n=o}^{m} \frac{(-2)^n J_{n+2}(k a\sin\theta)}{(m-n)!\,(k a\sin\theta)^{n+1}}\right|^2 \qquad (28)$$

These expressions are now in a form which is readily amenable to solution by digital computer for an arbitary aperture distribution.

Focused Antenna with Specialized Amplitude Distributions

Under the assumption of an aperture with a uniform illumination, the field equations (26) and (27) undergo extensive simplification. For such a case, the coefficients of the series (18) are defined as follows $a_m = a_o \neq 0$ for $m = O$ and $a_m = O$ for $m > O$. Then, the Fraunhofer field becomes:

$$U_o = \frac{jk a_o a}{2}\cdot\frac{e^{-j(kr+\gamma)}}{r}(1+\cos\theta)\,\frac{J_1(k a\sin\theta)}{k a\sin\theta} \qquad (29)$$

which is equivalent to Airy's classical diffraction formula for a circular aperture.[14] Similarly, the Fresnel field focusing correction takes the following form:

$$\delta U = \frac{-k a_o a^2}{2}\cdot\frac{e^{-j(kr+\gamma)}}{r^2}\sin\theta\cos\theta\,\frac{J_2(k a\sin\theta)}{k a\sin\theta} \qquad (30)$$

Again making use of the right angle phase relationship between the Fraunhofer field and the perturbation, we obtain the following expression for the normalized power patterns of a focused antenna:

$$|F(r,\theta)|^2 = \left|\frac{1+\cos\theta}{2}\cdot\frac{2J_1(k a\sin\theta)}{k a\sin\theta}\right|^2 + \left|\frac{a\sin\theta\cos\theta}{r}\cdot\frac{J_2(k a\sin\theta)}{k a\sin\theta}\right|^2 \qquad (31)$$

Or, if one approximates further that $\cos\theta \approx 1$,

$$|F(r,\theta)|^2 = \left|\frac{2J_1(k a\sin\theta)}{k a\sin\theta}\right|^2 + \left|\frac{\lambda}{2\pi r} J_2(k a\sin\theta)\right|^2 \qquad (32)$$

Figure 3 is a plot of the pattern of a uniformly illuminated focused antenna as described by

equation (32) with the focal distances $r = \infty$ (Fraunhofer) and $r = 10\lambda$. On the same coordinate axis is the unfocused pattern of a 10λ aperture at $r = 10\lambda$ as reported by Hansen and Bailin.[15]

To evaluate the effect of inducing an edge taper on the illumination function, one may proceed in like manner. If we assume the coefficients of the series of equations (18) as:

$$a_m = a_o \neq o \qquad m = o$$

$$a_m = \frac{\epsilon - 1}{a^2} a_o \qquad m = 1$$

$$a_m = o \qquad m > 1$$

then, the amplitude distribution in the aperture becomes:

$$A(\rho) = a_o \left[1 + \left(\frac{\epsilon - 1}{a^2} \right) \rho^2 \right] \tag{33}$$

where ϵ is the relative edge taper in voltage ratio and a_o is simply a normalizing constant. Then, substituting into equation (26), the far-field pattern takes on the following form:

$$U_o = \frac{jka_o a^2 \epsilon}{4} \cdot \frac{e^{-j(kr + \gamma)}}{r} (1 + \cos\theta)$$

$$\left[\frac{2J_1(ka\sin\theta)}{ka\sin\theta} + \frac{1 - \epsilon}{\epsilon} \cdot \frac{4J_2(ka\sin\theta)}{(ka\sin\theta)^2} \right] \tag{34}$$

and the perturbation due to focusing is:

$$\delta U = \frac{-ka_o a^3 \epsilon}{4} \cdot \frac{e^{-j(kr + \gamma)}}{r^2} \sin\theta \cos\theta$$

$$\left[\frac{2J_2(ka\sin\theta)}{ka\sin\theta} + \frac{1 - \epsilon}{\epsilon} \cdot \frac{4J_3(ka\sin\theta)}{(ka\sin\theta)^2} \right] \tag{35}$$

and, the expression for the normalized pattern with the usual small angle approximation has the following dependence:

$$|F(r,\theta)|^2 = \left[\frac{2\epsilon}{1+\epsilon} \right]^2 \left\{ \left| \frac{2J_1(ka\sin\theta)}{ka\sin\theta} + \frac{1-\epsilon}{\epsilon} \cdot \frac{4J_2(ka\sin\theta)}{(ka\sin\theta)^2} \right|^2 + \left| \frac{\lambda}{2\pi r} \left[2J_2(ka\sin\theta) + \frac{1-\epsilon}{\epsilon} \cdot \frac{4J_3(ka\sin\theta)}{ka\sin\theta} \right] \right|^2 \right\} \tag{36}$$

Equations (34), (35), and (36) are valid for all values of ϵ in the interval $1 \geq \epsilon > o$. For the case of an infinite edge tape (i.e., $\epsilon = o$), the normalized pattern, equivalent to equation (36), reduces to:

$$|F(r,\theta)|^2 = \left| \frac{8J_2(ka\sin\theta)}{(ka\sin\theta)^2} \right|^2 + \left| \frac{\lambda}{2\pi r} \cdot \frac{J_3(ka\sin\theta)}{ka\sin\theta} \right|^2 \tag{37}$$

We now recognize that the pattern may be written in the form:

$$|F(r,\theta)|^2 = |F_o(\theta)|^2 + \left| \frac{\lambda}{2\pi r} \delta F(\theta) \right|^2 \tag{38}$$

where the correction function, δF, is:

$$\delta F = \frac{2\epsilon}{1+\epsilon} \left[2J_2(ka\sin\theta) + \frac{1-\epsilon}{\epsilon} \cdot \frac{4J_3(ka\sin\theta)}{ka\sin\theta} \right] \tag{39}$$

for a circular aperture with an illumination taper given by ϵ. This correction function is plotted in Fig. 4 with ϵ as parameter. The effect of this focusing perturbation on the Fraunhofer pattern at a focal distance of 10λ is illustrated by the dashed curves in Figs. 5 and 6 for $\epsilon = 0$ and $1/2$ respectively.

Conclusion

The foregoing discussion establishes an extension to previous techniques for the computation of antenna patterns when the antenna is focused at a point in its Fresnel zone. The

pattern itself changes slightly from the polar pattern that would be observed were the antenna to be focused at infinity; the invariance of gain is established.

The Fresnel-zone focusing of antennas offers application at ranges less than the usual $(2D^2)/\lambda$ criterion, (for testing extremely large apertures) in that the observed Fresnel pattern may be corrected to a predicted far-field pattern with greater assurance, particularly in antenna systems designed for radio astronomy where the Fraunhofer region begins at ranges of hundreds of miles and are frequently not tested. Refocusing to a reduced range for the test can be accomplished with relative ease. The test itself can then give a measure of comparison to predicted performance.

References

1. J. Robieux, "Near Zone Power Transmission Formulas," Proc. IRE, June 1959, vol. 47, pp. 1161-63.

2. Robert W. Bickmore, "Fraunhofer Pattern Measurement in the Fresnel Zone," Canadian J. Phys., 1957, vol. 35, pp. 1299-1308; also Hughes Aircraft Co., Sci. Rep. No. 8, September 1955.

3. Alan F. Kay, "Far Field Data at Close Distances," Tech. Res. Group, Inc., New York, October 1956, Contr. No. AF 19(604)-1126.

4. R.S. Wehner, "Limitations of Focused Aperture Antennas," Rand Corp., Oct. 1949, Rep. No. RM-262.

5. Robert W. Bickmore, "On Focusing Electromagnetic Radiators," Canadian J. Phys., 1957, vol. 35, pp. 1292-1298.

6. Robert W. Bickmore, "A Focusing Criterion for Rectangular Apertures," Hughes Aircraft Co., Oct. 1954, Tech. Memo No. 376.

7. Samuel Silver, Microwave Antenna Theory and Design, McGraw Hill Book Company, Inc., New York, 1949, p. 170.

8. Philip M. Morse and Herman Feshbach, Methods of Theoretical Physics, McGraw Hill Booh Company, Inc., New York, 1953, p. 620 and p. 1323.

9. R.C. Hansen and L.L. Bailin, "Near Field Analysis of Circular Aperture Antennas," Hughes Aircraft Co., Sci. Rep. No. 3508/3, Aug. 1959, Cont. No. AF(19)604-3508.

10. I.N. Sneddon, "Functional Analysis," Handbuch der Physik, vol. II, Mathematical Methods II, Springer - Verlag, Berlin, 1955, p. 314.

11. Harry Bateman, Tables of Integral Transforms, vol. II, McGraw Hill Book Company, Inc., New York, 1954, pp. 3-92.

12. Samuel Silver, op. cit., p. 177.

13. Louis A. Pipes, Applied Mathematics for Engineers and Physicists, McGraw Hill Book Company, Inc. New York, 1946, p. 314.

14. Max Born and Emil Wolf, Principles of Optics, Pergamon Press, New York, 1959, p. 397.

15. R.C. Hansen and L.L. Bailin, op. cit.

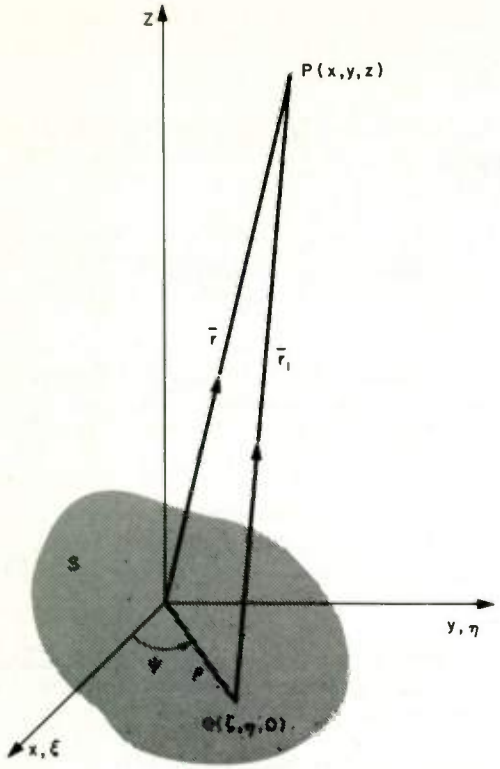


FIG.1 DEFINITION OF GENERAL FIELD COORDINATES

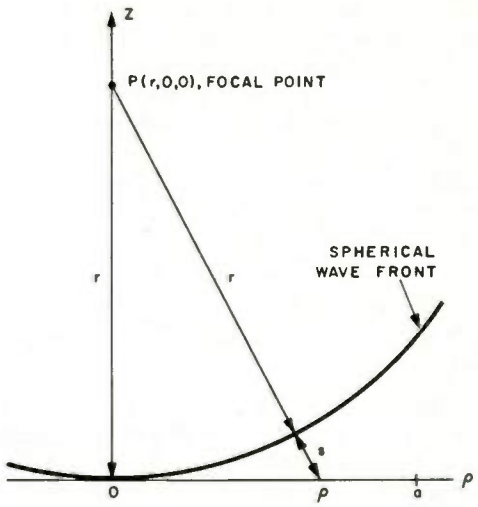


FIG.2 COORDINATE NOMENCLATURE FOR FOCUSED APERTURE

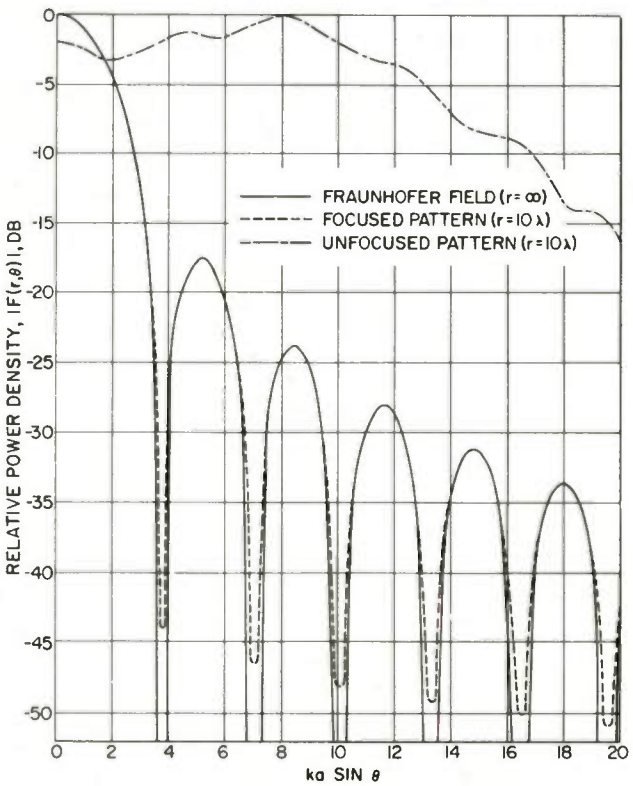


FIG. 3 PATTERNS OF A UNIFORMLY ILLUMINATED CIRCULAR APERTURE

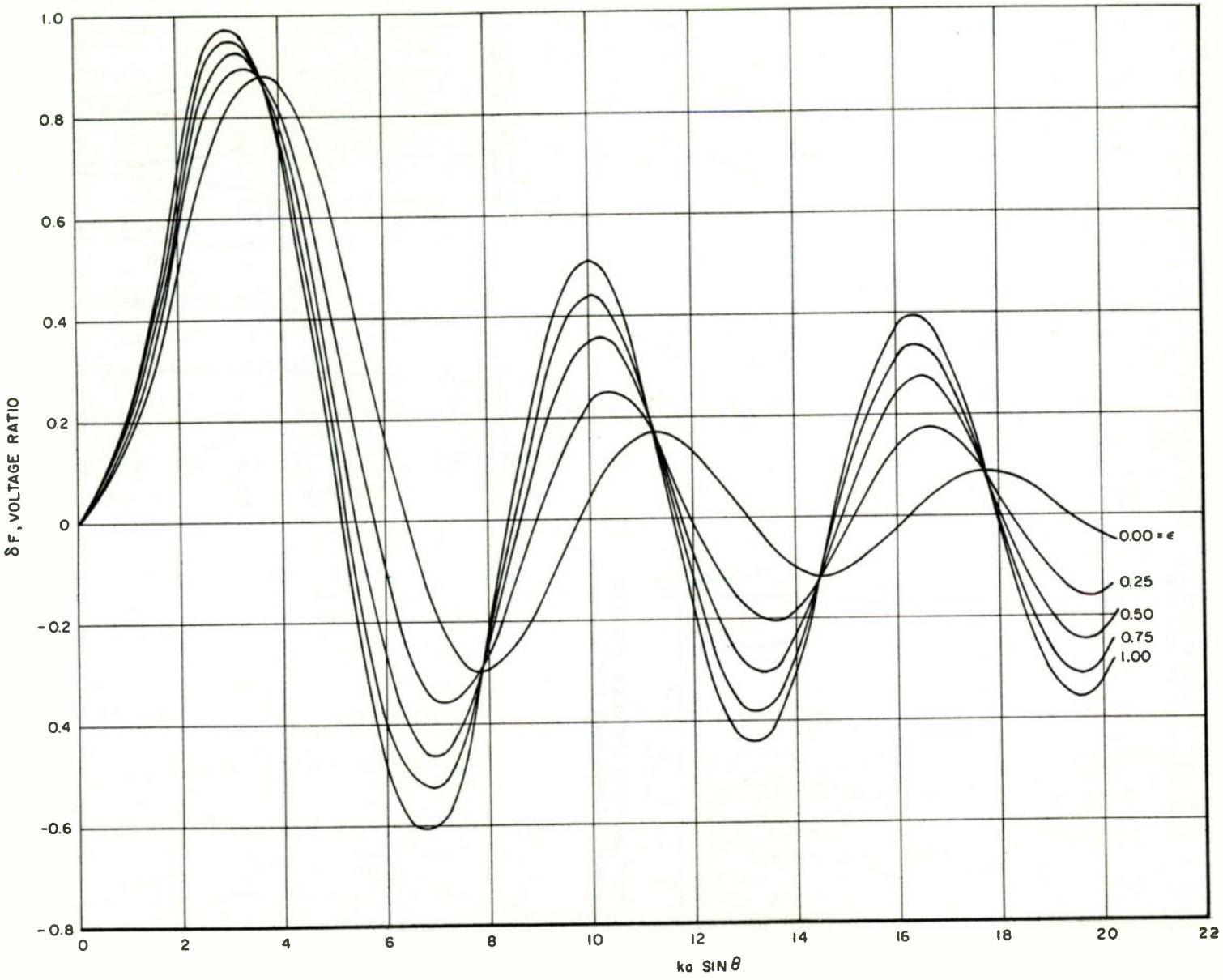


FIG. 4 PATTERN CORRECTION FACTOR FOR VARIOUS EDGE TAPERS

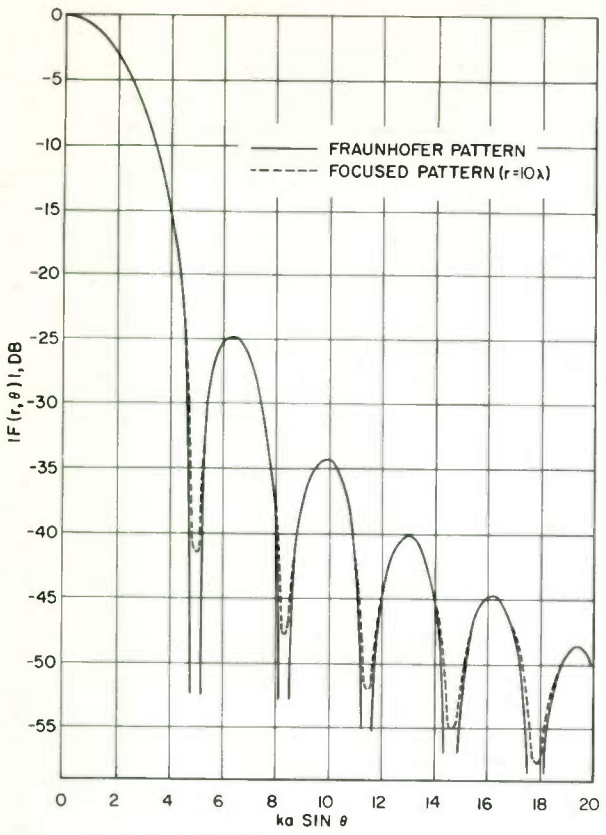


FIG. 5 PATTERNS OF A CIRCULAR APERTURE WITH INFINITE EDGE TAPER

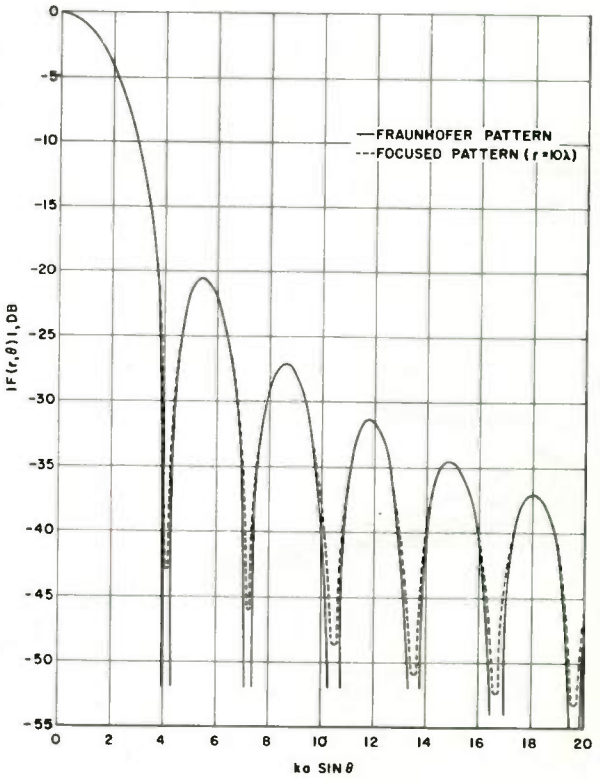


FIG. 6 PATTERNS OF A CIRCULAR APERTURE WITH A 6-DB EDGE TAPER

POLARIZATION TRACKING OF ANTENNAS

John F. Ramsay, James P. Thompson, and Warren D. White

Airborne Instruments Laboratory
A Division of Cutler-Hammer, Inc.
Deer Park, Long Island, New York

Summary

In Part I of this paper, the polarization characteristics of antennas using "geographical" analogies are reviewed. From the longitude polarization of a dipole and the latitude polarization of a loop, a vector description is given to the radiation from these basic elements. Combinations of dipoles, loops, or dipoles and loops can then be readily analyzed from simple complex vector formulas describing the amplitude, phase, and polarization of the far fields. In particular, turnstile-type combinations providing nominal circular polarization are discussed.

The vector technique is also applied to aperture antennas. It is shown that the obliquity factor of aperture antennas $(1 + \cos \theta)$ is a scalar constituent of an obliquity vector characterizing the radiated field, whether associated with linear or circular polarization. The vector antenna pattern for a linearly polarized horn antenna is derived and applied to the problem of the square horn fed by orthogonal TE modes in quadrature. Strong left circularly polarized lobes are radiated by a nominally right circular horn. The reduction of cross-polarization by providing a square horn with polarized teeth is described. The paper contains "geographical" drawings and sketches of 3 D contour plots of right and left circular component antenna patterns.

In Part II of this paper, "Punch-Through" problems associated with side-lobe suppression systems are discussed. The effectiveness of side-lobe suppression systems at any given angle of incidence can be determined by the measurement of punch-through probability for all possible polarizations of the input signal. For systems where only random linear polarizations need be considered, a rotating linearly polarized test signal is used, and the punch-through probability can be measured directly as a percentage of a complete rotation that produces an acceptable response.

For side-lobe suppression systems that must consider response to completely random polarized signals, a rotating linearly polarized source can again be used as the test signal. In this case, however, ellipticity measurements are obtained on each of the antennas separately, and this data can then be used to complete the punch-through probability by either analytic or plotting techniques.

Typical results obtained for an S-band circularly polarized side-lobe suppression system are presented which support the conclusion that in a multi-antenna side-lobe suppression system, careful selection of antenna types with regard to polarization tracking characteristics must be used in order to minimize the amount of unwanted signal reception or punch-through.

Introduction

Any antenna of reasonable directivity has side lobes. The reduction of side-lobe levels constitutes an important branch of antenna engineering. Yet, the best engineering leaves a residue side-lobe level. Even a very low side-lobe level can impair the operation of certain electronic systems, whether radar, communications, or other. Hence system engineers have devised side-lobe suppression systems that are mixed systems--partly wave, partly electronic. A common technique uses an auxiliary low-gain antenna for reception in conjunction with the high-gain directive antenna. Comparison of the signals received on the two antennas allows rejection of signals arriving from the side-lobe directions of the higher-gain antenna.

If the higher-gain antenna has a certain nominal polarization but, in its pattern, has cross-polarized side lobes at certain angles, and if the auxiliary low-gain antenna has the same nominal polarization as the high-gain antenna but has no cross-polarized components, the side-lobe suppression system will be inoperative at angles where the high-gain antenna is cross-polarized--that is, signals will punch-through. The probability that punch-through will occur in a side-

lobe suppression system using circularly polarized antennas is the subject of Part II of this paper. To secure effective side-lobe suppression, the two antennas must track in polarization over their joint field of view.

Part I of the paper reviews the basic polarization characteristics of certain elementary radiators. The latitude-longitude geographical approach to polarization is used to give a physical interpretation of complex vectors as used in the determination of how antennas generate cross-polarized components. It is shown that nominally circularly polarized antennas usually produce cross-polarized components, not only in their side-lobe regions but even in the main beam. An elementary explanation is also given of the general antenna pattern obliquity factor $(1 + \cos \theta)$.

PART I
POLARIZATION CHARACTERISTICS OF ANTENNAS

by

John F. Ramsay

A. Polarization Loci on Space Sphere

The far-field pattern of an antenna is obtained at any range greater than or equal to $2D^2/\lambda$, where D is the greatest dimension of the aperture and λ is the wavelength. If, therefore, the antenna is at the center of an imaginary sphere of a radius of at least $2D^2/\lambda$, a spherical surface is obtained on which the far field of the antenna can be mapped in relative amplitude, relative phase, and relative polarization. At any point on the sphere, the field has a complex amplitude and a state of polarization. Thus, if the field is probed by a linearly polarized detector, always oriented for maximum response, a complex amplitude is obtained having a state of polarization identical with that of the probe--that is, linear in a given orientation. If the probe (in the same element of the wavefront) is rotated through a right angle, in general, a second complex amplitude will be obtained, having an orthogonal linear state of polarization.

The two complex amplitudes obtained by sampling the field by orthogonal linearly polarized probes yield a complete description of the state of polarization at the given point on the sphere. If the amplitudes are numerically equal and possess a relative phase difference of 90°, the state of polarization at the point is circular; otherwise the state of polarization is linear or elliptical.

Alternatively, the field at a point can be sampled by circularly polarized probes, first by a right circularly polarized probe and second by a left circularly polarized probe. The two complex amplitudes obtained there again define the state of polarization at the point.

In general, the field can be sampled by elliptically polarized probes with orthogonal elliptical polarizations to yield the state of polarization. The techniques of linear or circular component sampling is by far the commonest in practice using two probes. Methods using three or more probes are also possible.

The state of polarization at a point on the sphere can be found by examining the field by a "matching" polarization in the probe. In this case, the state of polarization of the probe is the same as that of the incident wave; the amplitude is zero if the probe has a polarization orthogonal to the matching polarization.

On the surface of the space sphere, relative amplitudes or relative phases or relative polarizations can be mapped. The mapping of relative amplitudes yields the conventional contour plot of the antenna radiation pattern. The mapping of relative phases, though feasible, does not appear to have yet been adopted. The mapping of polarization is implicit in the mapping of amplitude (or phase) and can be made explicit. Thus, polarization can be mapped on the sphere independently of amplitude and phase.

For example, the polarization lines or loci of a vertical electric dipole in its far field are shown in Figure 1A as a set of longitudes. The axis of the dipole is the polar axis of the space sphere. At the poles the magnitude of the electric field is zero but at all other points a field exists polarized "longitudinally" as

shown. Note that near the poles the field of the vertical dipole is virtually horizontally polarized.

The vertical magnetic dipole, represented in Figure 1B by a small horizontal loop antenna, has as polarization loci a set of latitude circles; note that everywhere the polarization is horizontal. A vertical slot antenna in a vertical conducting plane has the same loci as the loop.

On a geographical globe, lines of longitude and lines of latitude intersect one another at right angles--that is, they are orthogonal. Hence an antenna consisting of a vertical electric dipole and a horizontal loop (Figure 1C), fed simultaneously, will generate on the space sphere the dipole longitudes of Figure 1A and the loop latitudes of Figure 1B as component polarization loci. At any point in space, two orthogonally polarized fields exist. The relative amplitudes can be made equal throughout space, since an infinitesimal dipole and an infinitesimal loop have identical space patterns. Since the amplitudes are equal but in quadrature phase, as the far field of a loop is in quadrature time phase with respect to that of a dipole fed in phase, the resultant polarization is circular at all points on the space sphere and has a constant sense of circularity.

If a vertical and a horizontal dipole are fed (or receive) together, the component loci are the intersecting pair of longitude sets shown in Figure 1D; it is evident that the constituent lines of each system are not orthogonal throughout space. Although the polarization lines are orthogonal in the principal planes, the antenna individual amplitude distributions do not match in these planes. If the two dipoles are fed with equal amplitudes in quadrature (the "turnstile" antenna), a circularly polarized radiation field is obtained at only two points on the space sphere--that is, only on the normal to the plane of the turnstile antenna. The exact patterns of the turnstile antenna are given in Section C.3 of Part I, and in Figure 6.

The turnstile antenna and the dipole/loop antenna can be termed biphase antennas. A three-phase antenna is the tripole shown in Figure 1E. Three small dipoles are spaced at 120° and commonly fed with equal amplitudes and rotating phase to produce circular polarization; however, circular polarization is produced only on the normal to the plane of the tripole. The three sets of oriented longitudes indicated in Figure 1E illustrate the onset of a pure polarization error caused by the tripole.

Mutually perpendicular loop antennas form a system that is the magnetic analog of the crossed electric dipole system. There are, of course, amplitude errors in the relative space patterns, since the characteristic doughnut patterns are mutually perpendicular. The polarization loci of the elements are orthogonal sets of latitudes as shown in Figure 1F; polarization errors are therefore to be expected, since individual polarization lines are not orthogonal in general.

The aforementioned geographical features of polarization loci assist in visualizing and analyzing the polarization characteristics of antennas.

B. Resolution of Polarizations

1. Reference Coordinate System. A right-handed Oxyz coordinate system will be used, with the origin at the center of the space sphere. Elementary radiators will be placed at the origin, and aperture antennas will have their apertures in the x-y plane (x horizontal, y vertical) and will be regarded as firing generally in the horizontal z-direction.

The intersections of the coordinate axes with the space sphere will be called poles, and the axes themselves polar axes, as required. Taking Oz as a polar axis of reference, the associated spherical polar coordinates are

$$\theta_z, \phi_z$$

and belong to an orthogonal coordinate system having longitudes and latitudes with z as the pole, as shown in Figure 2A. The radiation pattern data--amplitude, phase, and polarization--will be expressed in terms of θ_z, ϕ_z, sometimes written θ, ϕ. Ox and Oy can be taken as polar axes with associated spherical polar coordinates θ_x, ϕ_x and θ_y, ϕ_y, respectively. A unit vector $\overline{A}$ is specified in the three coordinate systems by

$$\overline{A} = \sin\theta_z \cos\phi_z \overline{x} + \sin\theta_z \sin\phi_z \overline{y} + \cos\phi_z \overline{z}$$

$$= \sin\theta_y \sin\phi_y \overline{x} + \cos\theta_y \overline{y} + \sin\theta_y \cos\phi_y \overline{z}$$

$$= \cos\theta_x \overline{x} + \sin\theta_x \cos\phi_x \overline{y} + \sin\theta_x \sin\phi_x \overline{z}$$

where $\overline{x}$, $\overline{y}$, $\overline{z}$ are unit vectors at the origin in the axial directions.

The following relations between the coordinate angles are given for reference:

$$\cos \theta_x = \sin \theta_y \sin \phi_y = \sin \theta_z \cos \phi_z$$

$$\cos \theta_y = \sin \theta_z \sin \phi_z = \sin \theta_x \cos \phi_x$$

$$\cos \theta_z = \sin \theta_x \sin \phi_x = \sin \theta_y \cos \phi_y$$

$$\sin \theta_x = \sqrt{(1 - \sin^2 \theta_y \sin^2 \phi_y)} = \sqrt{(1 - \sin^2 \theta_z \cos^2 \phi_z)}$$

$$\sin \theta_y = \sqrt{(1 - \sin^2 \theta_z \sin^2 \phi_z)} = \sqrt{(1 - \sin^2 \theta_x \cos^2 \phi_x)}$$

$$\sin \theta_z = \sqrt{(1 - \sin^2 \theta_x \sin^2 \phi_x)} = \sqrt{(1 - \sin^2 \theta_y \cos^2 \phi_y)}$$

At a point on the space sphere unit vectors in the tangent plane, $\bar{\theta}_z$, $\bar{\phi}_z$ directed to θ_z and ϕ_z increasing respectively, define an orthogonal coordinate system (Figure 2A). If complex amplitudes are known in the $\bar{\theta}_z$, $\bar{\phi}_z$ directions, the state of polarization is derived from these linear components, together with the complex resultant amplitude having that state of polarization. It is evident that determination of complex amplitudes in the $\bar{\theta}_z$, $\bar{\phi}_z$ directions amounts to resolving the field along a longitude-latitude spherical graticule having the z-axis as the polar axis.

2. Horizontal Dipole. A horizontal dipole, at the center of the space sphere, is oriented to lie on the x-axis (Figure 2B). A polarization longitude from dipole pole B contains a typical point A. Through A pass a longitude and a latitude referred to the propagation axis Oz, with unit vectors $\bar{\theta}_z$, $\bar{\phi}_z$ at A. The unit vector characteristic of the dipole polarization at A is $\bar{\theta}_x$. The unit vector $\bar{\theta}_x$ has to be resolved in the directions of the unit vectors $\bar{\theta}_z$, $\bar{\phi}_z$.

In the spherical triangle BAC, the angle C is a right angle and

$$a = BC = \phi_z$$

$$b = AC = 90^\circ - \theta_z$$

From the properties of spherical triangles

$$\tan A = \frac{\tan a}{\sin b} = \frac{\tan \phi_z}{\cos \theta_z}$$

Hence,

$$\sin A = \frac{1}{\sqrt{(1 + \cot^2 \phi_z \cos^2 \theta_z)}}$$

and

$$\cos A = \frac{1}{\sqrt{(1 + \tan^2 \phi_z \sec^2 \theta_z)}}$$

But resolving $\bar{\theta}_x$

$$\bar{\theta}_x = -\cos A \, \bar{\theta}_z + \sin A \, \bar{\phi}_z$$

Hence,

$$\theta_x = \frac{1}{\sqrt{(1 + \tan^2 \phi_z \sec^2 \theta_z)}}$$

$$(-\bar{\theta}_z + \tan \phi_z \sec \theta_z \, \bar{\phi}_z)$$

3. Vertical Dipole. A vertical dipole at the center of the space sphere is oriented to be on the y-axis (Figure 2C). A polarization longitude from the dipole pole B′ contains a typical point A′. Through A′ pass a longitude and a latitude referred to the propagation axis Oz, with unit vectors $\bar{\theta}_z$, $\bar{\phi}_z$ at A′. The unit vector characteristic of the dipole polarization at A′ is $\bar{\theta}_y$. The unit vector $\bar{\theta}_y$ has to be resolved in the direction of the unit vectors $\bar{\theta}_z$, $\bar{\phi}_z$.

In the spherical triangle B′A′C, the angle C is a right angle and

$$a' = CB' = 90^\circ - \phi_z$$

$$b' = A'C = 90^\circ - \theta_z$$

Then

$$\tan A' = \frac{\cot \phi_z}{\cos \theta_z}$$

giving

$$\sin A' = \frac{1}{\sqrt{(1 + \tan^2 \phi_z \cos^2 \theta_z)}}$$

$$\cos A' = \frac{1}{\sqrt{(1 + \cot^2 \phi_z \sec^2 \theta_z)}}$$

Since

$$\bar{\theta}_y = -\cos A' \, \bar{\theta}_z - \sin A' \, \bar{\phi}_z$$

then

$$\bar{\theta}_y = \frac{1}{\sqrt{(1 + \cot^2 \phi_z \sec^2 \theta_z)}}$$

$$(-\bar{\theta}_z - \cot \phi_z \sec \theta_z \, \bar{\phi}_z)$$

4. Vertical Loop. With dimensions small with respect to the wavelength, a loop antenna is placed at the center of the space sphere, the axis of the loop lying along the x-axis, which is the polar axis of the loop (Figure 2D). A typical polarization line of the loop is a vertical latitude circle; this latitude circle intersects a longitude and a latitude circle referred to O_Z as polar axis at a representative point on the sphere. The unit vector characteristic of the loop polarization is $\bar{\phi}_x$ which has to be resolved along $\bar{\theta}_z$ and $\bar{\phi}_z$.

The small latitude circle of the loop is orthogonal with the polarization longitude of an ϕ_x-directed dipole. The latter line of longitude makes an angle A with the $\bar{\theta}_z$ unit vector (Figure 2B). Hence, the small circle makes an angle 90° - A with the $\bar{\theta}_z$ direction. From Figure 2D

$$\bar{\phi}_x = -\cos(90^\circ - A)\bar{\theta}_z - \sin(90^\circ - A)\bar{\phi}_z$$

$$= -\sin A \, \bar{\theta}_z - \cos A \, \bar{\phi}_z$$

Using the values of sin A and cos A obtained previously for the horizontal dipole,

$$\bar{\phi}_x = \frac{1}{\sqrt{(1 + \cot^2 \phi_z \cos^2 \theta_z)}}$$

$$(-\bar{\theta}_z - \cot \phi_z \cos \theta_z \, \bar{\phi}_z)$$

5. Horizontal Loop. A loop antenna whose dimensions are small with respect to the wavelength is placed at the center of the space sphere, its axis lying along a polar y-axis. A typical polarization line of the loop is a small horizontal circle on the space sphere (Figure 2E). At a representative point on this small circle pass a longitude circle and a latitude circle having Oz as polar axis. The unit vector characteristic of the loop polarization is $\bar{\phi}_z$, which has to be resolved along $\bar{\theta}_z$ and $\bar{\phi}_z$.

The latitude circle of the loop is orthogonal with the polarization longitude of an Oy-directed dipole. The latter line of longitude makes an angle A′ with the $\bar{\theta}_z$ vector (Figure 2C). Hence, the horizontal latitude circle of the loop makes an angle 90° - A′ with the $\bar{\theta}_z$ direction. From Figure 2E

$$\bar{\phi}_y = \cos(90^\circ - A')\bar{\theta}_z - \sin(90^\circ - A')\bar{\phi}_z$$

$$= \sin A' \, \bar{\theta}_z - \cos A' \, \bar{\phi}_z$$

Using the values of sin A′ and cos A′ obtained previously for the vertical dipole,

$$\bar{\phi}_y = \frac{1}{\sqrt{(1 + \tan^2 \phi_z \cos^2 \theta_z)}}$$

$$(\theta_z - \tan \phi_z \cos \theta_z \, \bar{\phi}_z)$$

C. Vector Antenna Patterns

Antennas radiate beams. Even dipoles and loops can be regarded as radiating toroidal beams. Many attempts are made to devise antennas that do not radiate beams--that is, are isotropic sources. Omnidirectional coverage is frequently required by users of antennas. Electromagnetic waves, however, are polarized

and antennas are polarized because they radiate or receive polarized waves. Because antennas are polarized, omnidirectional coverage is impossible to obtain. It is possible, however, to make antennas that radiate in all directions but possess polarizations that vary from direction to direction. These "compromise" antennas can be of practical use; their optimum design and application require a knowledge of the polarization characteristics of antennas. The first feature is pure polarization, an almost geometrical property characteristic of the aperture and of the illumination on the space sphere.

Since antennas radiate beams, the energy density on the space sphere is not uniform and leads to the familiar space patterns of antenna measurements. The space patterns, however, cannot be fully specified without descriptions of the polarizations. Vectors are natural tools for describing polarization and hence apply suitably to space patterns. Vector radiation pattern descriptions will therefore be given for some elementary antennas.

1. Horizontal Dipole. The horizontal dipole has a polarization line that is a longitude with the x-axis as polar axis, represented by the unit vector $\overline{\theta}_x$. Along the longitude, the electric field varies as $\sin \theta_x$ for a short dipole or as $\cos(90^\circ \cos \theta_x)/\sin \theta_x$ for a half-wave dipole. (Short dipoles only will be treated in the following.) The electric field amplitude can be attached as a scalar multiplier to the electric field polarization vector to give a vector field pattern for the dipole as

$$\overline{E}(\theta{:}x) = \sin \theta_x \cdot \overline{\theta}_x$$

If the pattern is measured in the angular coordinates θ_z, ϕ_z with Oz as polar axis, then

$$\overline{E}(\theta_z, \phi_z) = \frac{\sqrt{(1 - \sin^2 \theta_z \cos^2 \phi_z)}}{\sqrt{(1 + \tan^2 \phi_z \sec^2 \theta_z)}}$$

$$(-\overline{\theta}_z + \tan \phi_z \sec \theta_z \overline{\phi}_z)$$

$$= -\cos \theta_z \cos \phi_z \overline{\theta}_z + \sin \phi_z \overline{\phi}_z$$

A vector pattern formulation of this type can be checked by physical considerations as follows. The dipole is lying along the horizontal x-axis. In the horizontal principal plane the pattern is $\sin \theta_x$ and in the vertical principal plane the amplitude is constant. In the horizontal principal plane (zx), $\phi_z = 0$ and θ_z = the pattern angle, with a polarization vector $\overline{\theta}_z$ (horizontal). In $\overline{E}(\theta_z, \phi_z)$ we therefore put $\phi_z = 0$, obtaining

$$E(\theta_z, \phi_z) = -\cos \theta_z \cdot \overline{\theta}_z$$

In the horizontal plane $\overline{\theta}_z = -\overline{\theta}_x$ and $\theta_z = 90^\circ - \theta_x$; hence

$$E(\theta_z, \phi_z) = \sin \theta \cdot \overline{\theta}_x$$

In the vertical principal plane (yz), $\phi_z = 90^\circ$, $\overline{\phi}_z$ exists but $\overline{\theta}_z$ does not; hence

$$E(\theta_z, \phi_z) = \overline{\phi}_z$$

There is no variation with the pattern angle θ_z and there is no $\overline{\theta}_z$ or vertical polarization. Only horizontal polarization exists ($\overline{\phi}_x$ at $\phi = 90^\circ$), for all θ_z pattern angles at a uniform amplitude of unity, which is characteristic of a dipole.

2. Vertical Loop. A loop antenna having its axis lying along Ox has the same pattern in space as a dipole, with the orthogonal polarization lines representing vertical latitudes. The vector pattern is

$$\overline{E}(\theta{:}x) = j \sin \theta_x \cdot \overline{\phi}_x$$

Referred to θ_z, ϕ_z pattern angles

$$\overline{E}(\theta_z, \phi_z) = j \frac{\sqrt{(1 - \sin^2 \theta_z \cos^2 \phi_z)}}{\sqrt{(1 + \cot^2 \phi_z \cos^2 \theta_z)}}$$

$$(-\overline{\theta}_z - \cot \phi_z \cos \theta_z \overline{\phi}_z)$$

$$= -j(\sin \phi_z \overline{\theta}_z + \cos \theta_z \cos \phi_z \cdot \overline{\phi}_z)$$

a formula verifiable by principal-plane considerations.

Similar formulas are obtainable for the vertical dipole and horizontal loop. These vector pattern expressions for the basic radiators enable vector patterns to be formulated for combinations of elementary radiators. Combinations that are normally expected to produce circular polarization are of special interest. Before obtaining vector patterns of circularly polarized antennas, it is necessary to discuss the complex feed amplitudes given to the elementary radiators in a combination.

3. Turnstile. Consider, for example, two crossed dipoles--one oriented along Ox the other along Oy--the phase center of each being at the origin. The polarization and pattern characteristics of each dipole have been discussed. When both are used in combination with a common feed, individual complex amplitudes are fed to each, differing in general. The vector radiation pattern must then be a function of the element amplitudes as well as the element current distributions and the element polarization.

Suppose the power fed to the x-directed element is X^2, the power fed to the y-directed element is Y^2, the total input power is $X^2 + Y^2$, and the fractional powers to the respective elements are

$$\frac{X^2}{X^2 + Y^2} \text{ and } \frac{Y^2}{X^2 + Y^2}$$

The current in the x-dipole will be proportional to $X/\sqrt{(X^2 + Y^2)}$ and will have a certain reference phase. The current in the y-dipole will be proportional to $Y/\sqrt{(X^2 + Y^2)}$, assuming identical dipoles, and will have a phase relative to the current in the x-dipole of δ. The normalized distributions of current in each dipole are $I(x)$ and $I(y)$: in the case of similar small dipoles $I(x)$ and $I(y)$ are the same constant--say unity.

The two-element antenna system can therefore be given a vector feed excitation

$$\overline{A}(x:y) = \frac{X}{\sqrt{(X^2 + Y^2)}} I(x)\overline{x} + \frac{Ye^{j\delta}}{\sqrt{(X^2 + Y^2)}} I(y)\overline{y}$$

$$= \frac{1}{\sqrt{\left[1 + \left(\frac{Y}{X}\right)^2\right]}} \left(\overline{x} + \frac{Y}{X} e^{j\delta}\overline{y}\right)$$

where the excitation ratio is $Y/X\ e^{j\delta}$ and $\overline{x}$, $\overline{y}$ are unit polarization vectors in the Ox and Oy directions. For electric dipoles, the unit vectors $\overline{x}$, $\overline{y}$ at the aperture transform to or generate the $\overline{\theta}_x$, $\overline{\theta}_y$ longitudes on the space sphere.

The vector radiation of two crossed dipoles fed with a complex excitation ratio is then

$$\overline{E}(\theta_z, \phi_z) = \overline{E}(\theta{:}x) + \frac{Y}{X} e^{j\delta}\overline{E}(\theta{:}y)$$

$$= -\cos\theta_z \cos\phi_z \overline{\theta}_z + \sin\phi_z \overline{\phi}_z + \frac{Y}{X} e^{j\delta} (-\cos\theta_z \sin\phi_z \overline{\theta}_z - \cos\phi_z \overline{\phi}_z)$$

$$= -\cos\theta_z(\cos\phi_z + \frac{Y}{X} e^{j\delta}\sin\phi_z)\overline{\theta}_z + (\sin\phi_z - \frac{Y}{X} e^{j\delta}\cos\phi_z)\overline{\phi}_z$$

Nominal circular polarization is produced conventionally for $X = Y$ and $\delta = -90^\circ$ (right circular polarization); for this case

$$\overline{E}(\theta_z, \phi_z) = -\cos\theta_z(\cos\phi_z - j\sin\phi_z)\overline{\theta}_z + (\sin\phi_z + j\cos\phi_z)\overline{\phi}_z$$

$$= e^{-j\phi_z}(-\cos\theta_z \overline{\theta}_z + j\overline{\phi}_z)$$

$$= \cos\theta_z e^{j(\pi - \phi_z)}\overline{\theta}_z + e^{j(\frac{\pi}{2} - \phi_z)}\overline{\phi}_z$$

On the space sphere $\overline{\theta}_z$ represents all the longitudes with O_z as polar axis, and $\overline{\phi}_z$ represents all the latitudes. The vector pattern equation describes how the field varies in amplitude and phase along the longitudes and latitudes when detected by a linear electric probe tangential to either. It is immediately obvious that on a given latitude the $\overline{\phi}_z$ amplitude is constant; there is only a phase rotation. For longitudinal polarization, however, the $\overline{\theta}_z$ amplitude falls away as $\cos\theta_z$, confirming that in the plane of the dipoles the polarization is in that plane only.

Although the amplitude distributions along the longitudes are the same, there

is a phase shift of the $\bar{\theta}_z$ patterns from longitude to longitude in the same direction as the phase rotation in the orthogonal polarization. The relative phase shift between components at a given point is -90°. The components are orthogonal in space. Hence the behavior of the resultant polarization is determined solely by the amplitude differences in the linear components. Any elliptical polarization on the space sphere has then no orientation--that is, major or minor axes are coincident with longitudes or latitudes. In the vector radiation patterns $\bar{E}(\theta_z, \phi_z)$, the phase factor that is ϕ-sensitive can be discarded, leaving the essential nature of the polarization as being given by

$$\bar{E}(\theta, \phi) = \cos\theta\, \bar{\theta} - j\bar{\phi}$$

the z subscripts having been dropped. The polarization is circular on the normal to the plane containing the dipoles, right circular on the positive z-axis, and left circular on the negative z-axis. In the plane of the dipoles the polarization is linear and in that plane. At points on the space sphere elsewhere, the polarization is elliptical and non-oriented with an axial ratio equal to cos θ_z. Contours of constant axial ratio are therefore latitude-like circles having the z-axis as common axis. For positive z, the polarization is right hand, and for negative z it is left hand.

The elliptical polarization radiated can be eliminated by substituting a loop (or slot) antenna for one of the dipoles. The coverage, however, is affected, but the polarization is circular everywhere in only one sense. This electromagnetic dipole was discussed earlier.

4. Horizontal Dipole and Vertical Loop. An x-directed dipole is encircled by a small loop having an x-directed axis. The loop is fed to yield an equal radiated power to that from the dipole. The vector antenna pattern of the system is

$$\begin{aligned}\bar{E}(\theta_z, \phi_z) &= -\cos\theta_z \cos\phi_z\, \bar{\theta}_z + \sin\phi_z\, \bar{\phi}_z \\ &\quad + j(\sin\phi_z\, \bar{\theta}_z + \cos\theta_z \cos\phi_z\, \bar{\phi}_z) \\ &= (-\cos\theta_z \cos\phi_z + j\sin\phi_z)\bar{\theta}_z \\ &\quad -j(-\cos\theta_z \cos\phi_z + j\sin\phi_z)\bar{\phi}_z \\ &= (-\cos\theta_z \cos\phi_z + j\sin\phi_z) \\ &\quad (\bar{\theta}_z - j\bar{\phi}_z)\end{aligned}$$

The vector in parentheses indicates that the polarization is right circular everywhere. The complex number in parentheses indicates that the circularly polarized field has an amplitude and a phase dependent on its coordinates. The amplitude pattern is

$$\sqrt{(\cos^2\theta_z \cos^2\phi_z + \sin^2\phi_z)} = \sin\theta_x$$

and the phase pattern is

$$-\tan^{-1}\left(\frac{\sin\phi_z}{\cos\phi_z \cos\theta_z}\right) = -\tan^{-1}(\tan\phi_z \sec\theta_z)$$

Had the field been measured in a $\bar{\theta}_x$, $\bar{\phi}_x$ coordinate system, the vector pattern would have been

$$\bar{E}(\theta_x, \phi_x) = \sin\theta_x(\bar{\theta}_x - j\bar{\phi}_x)$$

The amplitude pattern sin θ_x is therefore invariant with respect to the coordinate system of measurement. It is easy to show, however, that

$$\bar{\theta}_x - j\bar{\phi}_x = e^{-j\tan^{-1}(\tan\phi_z \sec\theta_z)}(\bar{\theta}_z - j\bar{\phi}_z)$$

Hence, the phase angle encountered in the field measurement in the $\bar{\theta}_z$, $\bar{\phi}_z$ coordinate system is caused by the rotation of the circularly polarized detector in passing from the $\bar{\theta}_x$, $\bar{\phi}_x$ system to the θ_z, ϕ_z system.

The previous examples, crossed-dipole turnstile and electromagnetic dipole, illustrate the application of the vector pattern formulation to "point source" circularly polarized elementary radiators. Other examples--such as tripoles and crossed loops--can be analyzed in a similar manner, and the techniques can be extended to more complicated radiators of the point source type and to arrays of such elements.

Aperture antennas are, however, frequently circularly polarized. What are the vector antenna patterns of such radiators? Reflector and lens antennas have polarization characteristics that can be complicated functions of the vector feed patterns and the polarization geometry of

the focusing elements. Small aperture antennas have polarization characteristics determined by both the aperture polarization and by diffraction at the boundary of the aperture and the immediate environment. Even a radome can affect the polarization performance.

5. Linearly Polarized Horn Antenna and its Obliquity Vectors. The most elementary aperture antenna is the pyramidal horn, long enough to yield negligible aperture phase error and having an aperture large enough to be free from edge diffraction. Horns larger than $3\lambda \times 3\lambda$ are adequate to be regarded as aperture antennas. Before examining circularly polarized horn antennas, it is advantageous to find a vector pattern for a linearly polarized horn.

A square aperture horn is taken at the center of the space sphere, the center of the aperture being at the origin and the aperture edges parallel to the Ox and Oy axes, respectively. The horn is fed with a TE_{10} waveguide mode; it is assumed that the mode amplitude distribution exists in phase at the mouth of the horn and that the aperture wave velocity is close to the free-space velocity of light. The electric field is polarized parallel to Ox, is uniform in the x-direction, and is cosinusoidal in the y-direction. The aperture edges are parallel to the directions of Ox, Oy respectively.

The wave launched from the aperture of the horn can be regarded as launched from an array of infinitesimal areas in the plane of the aperture, each of size dxdy. The polarization of the horn is then the polarization characteristic of an element of plane wave. Such an element can be simulated by an infinitesimal electric dipole in combination with an infinitesimal magnetic dipole (for example, a loop). Thus, a short dipole oriented along Ox and a small loop having its axis along Oy set at the center of the space sphere will generate the polarization loci of an element of wave from the horn (Figure 3A). It is clear that the resultant polarization will be neither that of the dipole alone nor that of the loop alone but will be a hybrid dipole-loop polarization. The far fields of a dipole and a loop are in time quadrature when these elements are fed in phase: for the dipole-loop simulation of an element of plane wave, the elements must, therefore, be fed in quadrature. Thus, using notations established above, the vector pattern of the dipole-loop combination is

$$\bar{E}(\theta_z, \phi_z) = \sin\theta_x \cdot \bar{\theta}_x - \sin\theta_y \cdot \bar{\phi}_y$$

The minus sign can be accounted for by considering the point on the z-axis: there, $\theta_x = 90^\circ = \theta_y$, $\bar{\phi}_y = -\bar{\theta}_x$, and the fields must add. Then, in terms of the $\bar{\theta}_z$, $\bar{\phi}_z$ coordinate system

$$\begin{aligned}\bar{E}(\theta_z, \phi_z) &= -\cos\theta_z \cos\phi_z\, \bar{\theta}_z + \sin\phi_z\, \bar{\phi}_z \\ &\quad - \cos\phi_z\, \bar{\theta}_z + \cos\theta_z \sin\phi_z\, \bar{\phi}_z \\ &= (1 + \cos\theta_z) \\ &\quad (-\cos\phi_z\, \bar{\theta}_z + \sin\phi_z\, \bar{\phi}_z)\end{aligned}$$

In the horizontal xy plane, $\phi_z = 0$, and the polarization is purely θ_z or horizontal, the pattern being

$$-(1 + \cos\theta_z)\bar{\theta}_z$$

In the vertical yz plane, $\phi_z = 90^\circ$ and the polarization is purely $\bar{\phi}_z$ or again horizontal, the pattern being

$$(1 + \cos\theta_z)\bar{\phi}_z$$

But $\phi_z = -\theta_z$ infinitesimally near the z-axis. Hence, the pattern amplitudes match.

The reader familiar with antenna pattern calculations using the Fourier transform will recognize $(1 + \cos\theta_z)$ as the "obliquity factor" that prefixes the integral.

In planes other than the principal planes, the dipole loop combination has polarization components in both the $\bar{\theta}_z$ and $\bar{\phi}_z$ directions. The vector pattern of the horn is then the Fourier transform of the illumination multiplied by the vector pattern of the wave element.

$$\bar{G}_x(u, v) = \text{sinc } u \; \text{cosk } v \; \bar{g}_x$$

where $\text{sinc } u = \dfrac{\sin u}{u}$;

$$\text{cosk } v = \left(\frac{\pi}{2}\right)^2 \frac{\cos v}{\left(\frac{\pi}{2}\right)^2 - v^2}$$

$$\text{and } u = \frac{\pi a}{\lambda} \sin\theta_z \cos\phi_z;$$

$$v = \frac{\pi a}{\lambda} \sin\theta_z \sin\phi_z$$

with $\bar{g}_x = (1 + \cos \theta_z)$

$(-\cos \phi_z \, \bar{\theta}_z + \sin \phi_z \, \bar{\phi}_z)$

If the polarization of the horn in the aperture plane is in the y-direction (Figure 3B), the vector pattern is

$$G_y(u, v) = \text{cosk } u \text{ sinc } v \, \bar{g}_y$$

where $\bar{g}_y = (1 + \cos \theta_z)$

$(-\sin \phi_z \, \bar{\theta}_z - \cos \phi_z \, \bar{\theta}_z)$

6. Circularly Polarized Horn Antenna and its Obliquity Vector. If an element of wavefront in an antenna aperture is circularly polarized, the element can be resolved into two orthogonally linearly polarized wave elements in time quadrature. The resulting vector pattern of the element is then, for right circular polarization (Figure 3C)

$$\bar{g}_R = \bar{g}_x - j \, \bar{g}_y$$

$$= -(1 + \cos \theta_z) \, e^{-j\phi_z} \, (\bar{\theta}_z - j\bar{\phi}_z)$$

The radiation pattern is everywhere right circularly polarized, has a maximum in the z-direction and a zero in the backward or negative z-direction, being a cardioid of revolution. Physically the pattern can be regarded as generated by a pair of orthogonal dipoles and a pair of orthogonal loops instead of by a pair of dipole loops as used in the wave element simulation.

The aperture of a horn antenna, however, where fed by orthogonal TE modes is not uniformly circularly polarized over all wave elements in the aperture. The constant-by-cosine property of each mode leads to one component being uniformly distributed, the orthogonal component being cosine distributed. Elliptical polarization is generated in the aperture (Figure 4); the horn radiation pattern cannot, therefore, be circularly polarized at all angles. If the aperture polarization is resolved into right and left circular components, then the right circular component distribution will generate a $\bar{g}_R$ pattern vector, whereas the left circular component distribution will produce a $\bar{g}_L$ field (Figure 3D) given by

$$\bar{g}_L = -(1 + \cos \theta_z) \, e^{j\phi_z} \, (\bar{\theta}_z + j\bar{\phi}_z).$$

The vector aperture distribution of the horn is

$$\bar{A}(x, y) = \cos \frac{\pi y}{a} \, \bar{x} - j \cos \frac{\pi x}{a} \, \bar{y}$$

where $\bar{x}$, $\bar{y}$ are unit vectors along Ox, Oy respectively. To resolve into circular components, unit right circular and unit left circular vectors are defined by

$$\bar{r} = \frac{1}{\sqrt{2}} \, (\bar{x} - j \, \bar{y})$$

$$\bar{\ell} = \frac{1}{\sqrt{2}} \, (\bar{x} + j \, y)$$

whence

$$\sqrt{2} \, \bar{A}(x, y) = (\cos \frac{\pi y}{a} + \cos \frac{\pi x}{a}) \, \bar{r}$$

$$+ (\cos \frac{\pi y}{a} - \cos \frac{\pi x}{a}) \, \bar{\ell}$$

Unit circular polarization vectors $\bar{r}$, $\bar{\ell}$ produce radiation vectors g_R, g_L in the far field; thus, the vector radiation pattern of the horn fed by orthogonal modes is

$$\bar{E}(\theta, \phi) = (\text{sinc } u \text{ cosk } v$$

$$+ \text{cosk } u \text{ sinc } v) \, \bar{g}_R$$

$$+ (\text{sinc } u \text{ cosk } v$$

$$- \text{cosk } u \text{ sinc } v) \, \bar{g}_L.$$

The principal plane patterns of the (u, v) transforms for right circular polarization are

$$G_R(u) = \text{sinc } u + \text{cosk } u$$

$$G_R(v) = \text{sinc } v + \text{cosk } v$$

and for left circular polarization are

$$G_L(u) = \text{sinc } u - \text{cosk } u$$

$$G_L(v) = \text{sinc } v - \text{cosk } v$$

The left circular component is zero if $u = v$--that is, $\phi = \pm 45^\circ$. The right circular diagonal pattern is then

$$G_R(w) = 2 \text{ sinc } w \text{ cosk } w$$

where $w = \sqrt{\left\{\frac{1}{2}(u^2 + v^2)\right\}}$

Principal-plane (v = o) and diagonal-plane (u = v) pattern cuts in the (u, v) coordinate system are shown in Figure 5. The highest levels of left circular polarization occur in regions near the first nulls of the right circular patterns, there being significant left circular polarization within the main beam of the right circular patterns. The maximum level of the left circular polarization is -16 db in the principal plane shown. The (1 + cos θ) polarization factor is, of course, not shown in the (u, v) plot.

7. Toothed Horn. The left circular polarization can be reduced by fitting teeth to the mouth of the horn, either inside or outside. The teeth may be made of wires or vanes. These fittings have no significant effect on the wave having a linear polarization perpendicular to their length; the teeth, however, short-circuit the field polarized parallel to them. The effective electrical aperture of a mechanically square aperture is then rectangular and that of a circular aperture is quasi-elliptical for each component wave. The E- and H-plane beamwidths of a given component linearly polarized wave can then be made equal at an appropriate level; the orthogonal component is similarly beamwidth-matched.

If the side of the square horn has dimensions a, the teeth (if internal) reduce the aperture to a dimension $a' \times a'$. The effective aperture for each mode is then $a \times a'$.

The vector pattern of the toothed horn is, therefore,

$$\bar{E}(\theta_z, \phi_z) = \text{sinc } u' \cos k\, v\, \bar{g}_x - j \cos k\, u \text{ sinc } v' \bar{g}_y$$

where $u' = \pi \frac{a'}{\lambda} \sin \theta_z \cos \phi_z$

$$v' = \pi \frac{a'}{\lambda} \sin \theta_z \sin \phi_z$$

The left circular component in the principal plane reduces to -28 db near the base of the right circular beam. At a wide angle, however, a side lobe of the order of -20 db appears. These values apply to the toothed horn having $a' = \frac{5}{7} a$.

D. Scalar Antenna Patterns

A scalar antenna pattern is the pattern taken on a pattern range using a specific polarization on the antenna remote from the antenna whose patterns are being measured. The remote antenna can be called the range antenna, and the measured antenna can be called the test antenna. For pattern measurements made at constant range, for example $2D^2/\lambda$, the test antenna can be regarded as being at the center of the space sphere, and the range antenna as a probe antenna on the surface of the space sphere. Either the test antenna can be rotated to point in any direction and the range antenna held at a fixed point on the space sphere (for example, on the z-axis), or the test antenna can be kept fixed with, say, its aperture in the xy plane, and the range antenna can be moved over the surface of the space sphere. Because the latter technique allows the polarization performance to be analyzed in terms of the motion of a point on the space sphere, whereas, the former involves antenna rotations and rotations of poles and polarization lines, the simpler method of holding the test antenna fixed and moving the range antenna will be adhered to here. Both methods are equivalent.

If the range antenna is linearly polarized on the peak of its beam, its polarization can be aligned with the unit $\bar{\theta}_z$ vector at any point on the sphere to yield scalar antenna patterns that are the coefficients of $\bar{\theta}_z$ in the expressions for the vector pattern. Similarly, if the range antenna polarization is matched to the unit vector $\bar{\phi}_z$, the scalar coefficients of $\bar{\phi}_z$ in the vector pattern are obtained. Thus, for a horizontal dipole the scalar patterns are

$$\left|E_\theta(\theta_z, \phi_z)\right| = \cos \theta_z \cos \phi_z$$

and

$$\left|E_\phi(\theta_z, \phi_z)\right| = \sin \phi_z$$

whereas for a horizontal loop

$$\left|E_\theta(\theta_z, \phi_z)\right| = \cos \phi_z$$

and

$$\left|E_\phi(\theta_z, \phi_z)\right| = \cos \theta_z \sin \phi_z .$$

For a horn polarized parallel to Ox

$$\left|E_\theta(\theta_z, \phi_z)\right| = \text{sinc } u \cos k\, v\, (1 + \cos \theta_z) \cos \phi_z$$

$$\left|E_{\phi}(\theta_z, \phi_z)\right| = \text{sinc } u \cos^k v \; (1 + \cos\theta_z) \sin\phi_z$$

where u, v are functions of θ_z, ϕ_z.

To secure the right circular pattern of a nominally right circularly polarized test antenna, the range antenna is right circularly polarized, since "like antennas see each other" and "unlike antennas are blind" when polarization is in question. The measured scalar pattern of the antenna is then obtained by multiplying the vector pattern scalarly by a left circular polarization unit vector, since the antennas face each other.

For example, a turnstile antenna of two dipoles generating right circular polarization on the positive z-axis has a vector pattern

$$\overline{E}(\theta_z, \phi_z) = e^{-j\phi_z} (-\cos\theta_z \, \overline{\theta}_z + j\overline{\phi}_z)$$

The measured right circular polarization pattern is then

$$E_R(\theta_z, \phi_z) = e^{-j\phi_z} (-\cos\theta_z \, \overline{\theta}_z + j\phi_z) \cdot (\overline{\theta}_z + j\overline{\phi}_z)$$

$$= e^{-j\phi_z} (-\cos\theta_z - 1)$$

$$= -e^{-j\phi_z} (1 + \cos\theta)$$

since

$$(a\overline{\theta}_z + b\overline{\phi}_z) \cdot (\overline{\theta}_z + j\overline{\phi}_z) = a + jb$$

defines scalar multiplication of two vectors, providing the so-called "dot" product, for which $\overline{\theta}_z \cdot \overline{\theta}_z = 1 = \overline{\phi}_z \cdot \overline{\phi}_z$.

The left circular scalar pattern is obtained from the vector pattern by multiplication by $(\overline{\theta}_z - j\overline{\phi}_z)$, yielding

$$E_L(\theta_z, \phi_z) = e^{-j\phi_z} (1 - \cos\theta_z)$$

$$= -e^{j(\pi-\phi_z)} (1 - \cos\theta_z)$$

showing the counter rotation of phase. If phase is not measured the right and left circular patterns are given by

$$E(\theta) = 1 \pm \cos\theta$$

Scalar patterns of the other circularly polarized antennas including the horns described can be similarly obtained by an application of the dot product, yielding contour plots as illustrated in Figure 6.

PART II

POLARIZATION PUNCH-THROUGH IN SIDE-LOBE SUPPRESSION SYSTEMS

by

James P. Thompson and Warren D. White

A. Introduction

This portion of the paper is concerned with the polarization punch-through problem associated with side-lobe suppression systems with antenna patterns of the type shown in Figure 7. In this type of system, a relatively narrow-beam antenna is used for the main acceptance antenna and a wide-beam antenna is used to reject all signals arriving from outside the desired acceptance angle. Rejection is accomplished by comparing amplitudes of the received signal in both channels. When the amplitude of the signal received in the wide-beam channel is greater than that received in the narrow-beam channel by some selected threshold value, the signal is rejected. The polarization punch-through problem associated with such a system is that at some angles outside the desired acceptance angle unwanted signals will be accepted by the system because of the polarization characteristics of the antennas at these angles. Furthermore, within the acceptance angle desired, signals may be rejected because of differences in the polarization of the two antennas.

This problem occurs because the polarization of the two antennas at various angles is never precisely the same. It is therefore possible for a signal outside the desired acceptance angle having an orthogonal polarization to the wide-beam antenna and some in-line polarization component to

the narrow-beam antenna to be accepted by the system. Conversely, it is possible for a signal inside the acceptance angle to produce little response in the narrow-beam antenna and a large enough response in the wide-beam antenna to cause the signal to be rejected. This latter problem is not usually found, however, since it is simpler to have polarization tracking of the two antennas over narrow angles close to boresight than it is for large angles off boresight. The point to be understood is that the condition for signal acceptance is a function of the polarization of the incident radiation and the polarization difference between the narrow- and wide-beam antennas.

To show how rejection action depends upon the polarization of the incident radiation, an example in which the antennas are assumed to be linearly polarized and the incident radiation is assumed to be randomly linearly polarized can be considered. For this example:

θ is defined as the angle between the polarization of the narrow-beam antenna and the incident radiation.

ϕ is defined as the angle between the polarization of the wide-beam antenna and the incident radiation.

P_N is the maximum amplitude of the signal produced at the narrow-beam antenna terminals by incident radiation.

P_W is the amplitude of the signal produced at the wide-beam antenna terminals by incident radiation.

The amplitudes of the signals in the two channels are $P_N \cos \theta$ and $P_W \cos \phi$.

Conditions for acceptance:

$$\frac{P_N \cos \theta}{P_W \cos \phi} \geq \text{acceptance threshold, } T_A$$

Conditions for rejection:

$$\frac{P_N \cos \theta}{P_W \cos \phi} \leq T_A$$

The acceptance threshold (T_A) is the minimum ratio of narrow-beam to wide-beam signal power at the respective antenna terminals required to achieve an accept condition. The actual threshold value that is used will be based on the allowable number of noise firings and the probability of signal detection that is required for a given system application.

If the narrow-beam and wide-beam antennas are identically polarized--that is, $\theta = \phi$--the condition for acceptance is independent of the polarization of the incident radiation and depends only on the ratio P_N/P_W.

If the antennas are not identically polarized, the acceptance curve becomes dependent on the polarization of the incident radiation. This is graphically represented in Figure 8, where it is seen that at a given angle of incidence and for orthogonally polarized antennas the polarization angles from 0 degrees to ψ degrees are accepted. Therefore, if random linear polarization is assumed, the probability of intercept at that angle of incidence is $\psi/90$. The same results would be obtained at other angles of incidence. Thus, at any given angle of incidence, if the incident signal is of random linear polarization, the system should be considered as having a probability of acceptance rather than a definite yes-no decision process. For angles of incidence close to boresight, this probability would be high, since the antenna polarizations would be nearly identical and P_N/P_W would be large because of antenna gain characteristics. Consequently, the plot of $P_N \cos \theta / P_W \cos \phi$ would be well above the threshold value (T_A) for almost all polarizations.

To expand the above considerations to circularly polarized antennas and randomly polarized input signals, it will be necessary to make use of the concept of the polarization sphere.[15] Any polarization can be represented by a unique point on this sphere. Linear polarizations are assigned to the equator, with the north and south poles being circular polarizations with right and left sense, respectively. Along lines of constant longitude are elliptical polarizations that vary from linear to circular with increasing latitude. The major axis of the elliptical polarizations are oriented in the direction of the linear polarization of the same longitude.

To obtain the probability of acceptance for any possible polarization of the input signal, it is necessary to find the fraction of the surface of the polarization sphere yielding antenna signal outputs whose ratio exceeds the required acceptance threshold (T_A). To do this, it is necessary to obtain some measure of the difference in polarization of the two antennas. This is done by taking ellipticity measurements consisting of the magnitude of response in the major and minor axes and relative orientation of the polarization ellipse formed by these axes at uniformly distributed points throughout the antenna patterns. A relative measure of the difference between two polarizations can be

obtained by the distance between the points on the polarization sphere. Therefore, by transforming the antenna polarization measurements to points on the polarization sphere, a measure of the difference between the polarizations is obtainable.

Once the measure of difference of polarization is found, it is possible to determine the probability of acceptance as a function of the polarization of the incident radiation. The details of the theory involved and methods for obtaining polarization punch-through probability measurements for the random linear polarization input and the completely random polarization input is described in the next sections.

B. Computation and Measurement of Punch-Through Probability for Random Linear Polarization

Since the usual method of antenna pattern measurement corresponds so closely to the situation encountered under the assumption that the incident polarization is linear with a random orientation, the computation of the punch-through probability for this case is rather simple. Several methods can be used.

1. Both antennas can be illuminated simultaneously and their outputs connected to a recording ratio meter. The percentage of one complete rotation of the incident polarization for which the required threshold ratio is exceeded is the punch-through probability at that angle of incidence. The same measurements can be taken at other angles of incidence to determine the punch-through probability at these angles.

2. Both antennas are illuminated simultaneously, this time with a pulsed source. The two antennas are connected to a threshold circuit and the output is applied to a counter. If a fixed number of pulses (say, for example, 10^4) are emitted while the source polarization rotates through one or more revolutions, the counter reading will give the punch-through probability at the measured angle of incidence directly.

3. Each antenna's response to a random linear polarization input is measured individually, and the response plot is made in rectangular coordinates with a logarithmic scale of signal strength as the ordinate and polarization angle as the abscissa. The two patterns thus obtained are then superimposed on a tracing table, one being shifted in amplitude relative to the other by the amount of the threshold setting. The portion of the plot for which the narrow-beam antenna response exceeds the wide-beam response is the punch-through probability at the angle of incidence measured. Similar measurements can be taken for other angles of incidence. This method has the advantage of not requiring any equipment not normally found on a pattern range.

In each of the previously mentioned methods, if the antennas are required to operate over a wide range of frequency, measurements at each angle of incidence will have to be made at several input frequencies since the antenna response is also a function of frequency.

C. Computation of Punch-Through Probability for a Completely Random Polarization Input

1. General Theory. Let the incident wave be represented by the complex vector[16]

$$\underrightarrow{E} = E_o\, e^{i\phi_u} \frac{1}{\sqrt{1 + |P|^2}} \left(\underrightarrow{u_1} + \vec{P}\, \underrightarrow{u_2}\right)$$

where $\underrightarrow{u_1}$ and $\underrightarrow{u_2}$ are any arbitrary pair of orthogonal unit vectors and,

$$E_o = \sqrt{\left|E_{u_1}^2\right| + \left|E_{u_2}^2\right|}$$

$$\phi_u = \text{Arg}\left\{\overrightarrow{E_{u_1}}\right\}$$

$$\vec{P} = \frac{\overrightarrow{E_{u_2}}}{\overrightarrow{E_{u_1}}}$$

E_o is a quantity related to the total energy of the wave and $\vec{P}_u$ is a complex quantity characterizing the nature of the polarization.

Similarly, let the equivalent heights of the narrow-beam and wide-beam antennas be represented by the complex vectors

$$\underrightarrow{N} = N_o\, e^{i\psi_N} \frac{1}{\sqrt{1 + |Q_N|^2}} \left(\underrightarrow{u_1} + \vec{Q}_N\, \underrightarrow{u_2}\right)$$

$$\underrightarrow{W} = W_o\, e^{i\psi_W} \frac{1}{\sqrt{1 + |Q_W|^2}} \left(\underrightarrow{u_1} + \vec{Q}_W\, \underrightarrow{u_2}\right)$$

where $\vec{Q}$ is the complex quantity characterizing the nature of the antenna polarization.

The output voltages of the two antennas can then be represented by the complex scalars

$$\vec{V}_N = \vec{E} \cdot \vec{N}^*$$

$$= E_o N_o \, e^{i(\phi - \psi_N)} \frac{1 + \vec{P}\,\vec{Q}_N^*}{\sqrt{(1 + |P|^2)(1 + |Q_N|^2)}}$$

$$\vec{V}_W = \vec{E} \cdot \vec{W}^*$$

$$= E_o W_o \, e^{i(\phi - \psi_W)} \frac{1 + \vec{P}\,\vec{Q}_W^*}{\sqrt{(1 + |P|^2)(1 + |Q_W|^2)}}$$

The ratio of these two quantities is

$$\vec{A} = \frac{\vec{V}_N}{\vec{V}_W} = \frac{N_o}{W_o} e^{i(\psi_W - \psi_N)} \sqrt{\frac{1 + |Q_W|^2}{1 + |Q_N|^2}} \left\{ \frac{1 + \vec{P}\,\vec{Q}_N^*}{1 + \vec{P}\,\vec{Q}_W^*} \right\}$$

Our problem of finding the polarization punch-through probability can now be stated as one of finding that portion of the polarization sphere representing the incident wave that will yield ratios satisfying the relation

$$|A| \geq T_A$$

The problem is considerably simplified if we make a transformation of the unit vectors to cause Q_W to vanish. Such a transformation is the following.

$$\vec{u}_1 = \frac{1}{\sqrt{1 + |Q_W|^2}} (u_1' - \vec{Q}_W \vec{u}_2')$$

$$\vec{u}_2 = \frac{1}{\sqrt{1 + |Q_W|^2}} (\vec{Q}_W^* \vec{u}_1' + u_2')$$

Our equation for $\vec{A}$ is then transformed to

$$\vec{A} = \frac{N_o}{W_o} e^{i(\psi_W' - \psi_N')} \frac{1 + \vec{P}'\vec{Q}_N'^*}{\sqrt{1 + |Q_N'|^2}}$$

where

$$\vec{P}' = \frac{\vec{P} - \vec{Q}_W}{1 + \vec{P}\,\vec{Q}_W^*}$$

and

$$\vec{Q}_N' = \frac{\vec{Q}_N - \vec{Q}_W}{1 + \vec{Q}_N\,\vec{Q}_W^*}$$

$\vec{P}'$ is the complex quantity characterizing the polarization of the input signal in the transformed coordinate system.

$\vec{Q}_N'$ similarly is the complex quantity characterizing the polarization of the narrow-beam antenna in the transformed coordinate system.

$\vec{Q}_W'^*$ and $\vec{Q}_N'^*$ are the orthogonal polarizations that produce no response in the wide-beam and narrow-beam antennas.

Since we are interested only in $|A|$, we do not concern ourselves with the transformed phase angles ψ_W' and ψ_N'. We can further simplify the expression by making the substitutions

$$\frac{N_o}{W_o} = T_o$$

$$\vec{Q}_N' = \rho e^{i\alpha}$$

$$\vec{A} = \vec{a}\, T_o \, e^{-i(\psi_W' - \psi_N')}$$

where ρ is the desired distance between the narrow-beam and wide-beam antenna polarizations in the P' plane and α represents the orientation of this distance line in the P' plane.

We then obtain

$$\vec{a} = \frac{1 + \vec{P}'\rho e^{-i\alpha}}{\sqrt{1 + \rho^2}}$$

and the condition for punch-through is

$$|a| \geq \frac{T_A}{T_o} = t$$

If we substitute

$$\vec{a} = te^{i\lambda}$$

and solve for $\vec{P}'$, we obtain an equation for the contour bounding the punch-through region in the P' plane

$$\vec{P}' = \frac{1}{\rho} e^{i\alpha} \left[te^{i\lambda} \sqrt{1 + \rho^2} - 1 \right]$$

If we now make a further transformation by rotating the P′-plane through the angle α, we can eliminate α from the contour equation and obtain an equation for the contour in this new P″ plane.

Let $\vec{P'} = P''e^{-i\alpha}$

we obtain

$$P'' = \frac{1}{\rho}\left[te^{i\lambda}\sqrt{1+\rho^2} - 1\right]$$

which is the equation for a circle centered on the point $P'' = -1/\rho$ and having a radius

$$R = \frac{t}{\rho}\sqrt{1+\rho^2}$$

Our problem now resolves itself into one of projecting this circle in the P″ plane onto the surface of the unit sphere and then determining the fraction of the surface of the sphere that is outside this circle. The situation is shown in Figure 9 and the details of the computations are shown in Figure 10. As can be seen, the area in question is that portion of the surface of the sphere bounded by a cone whose apex angle is $2(\beta - \gamma)$. If we were to rotate the sphere so that the axis of this cone were vertical, with its apex pointed down, we could then say the problem is equivalent to that of finding all the area below the parallel marking latitude $\frac{\pi}{2} - (\beta - \gamma)$. The fraction of the whole sphere that fits this description is the punch-through probability, p, when the incident radiation is uniformly distributed over the sphere.

$$p = \frac{1}{2}\left\{1 + \cos(\beta - \gamma)\right\}$$

If the incident radiation is not uniformly distributed over the sphere, this fraction would have to be adjusted accordingly.

Now from Figure 9, it can be seen that since the sphere is of unit diameter

$$\tan\beta = \frac{1}{\rho}\left(1 + t\sqrt{1+\rho^2}\right)$$

$$\tan\gamma = \frac{1}{\rho}\left(1 - t\sqrt{1+\rho^2}\right)$$

therefore

$$\tan(\beta - \gamma) = \frac{\tan\beta - \tan\gamma}{1 + \tan\beta\tan\gamma}$$

$$= \frac{2t}{1+t^2}\frac{\rho}{\sqrt{1+\rho^2}}$$

An analytic expression for ρ in terms of $\vec{Q_N}$ and $\vec{Q_W}$ is

$$\rho = \left|\vec{Q_N'}\right| = \left|\frac{\vec{Q_N} - \vec{Q_W}}{1 + \vec{Q_N}\,\vec{Q_W^*}}\right|$$

If $\vec{u_1}$ and $\vec{u_2}$ are taken to be unit vectors representing circular polarization, $\vec{Q_N}$ and $\vec{Q_W}$ can be determined from ellipticity pattern measurements.

Let r_N and r_W be the axial ratios and θ_N and θ_W be the major axis orientations of the narrow-beam and wide-beam antenna ellipticity patterns as shown in Figure 11. Then

$$\vec{Q_N} = \frac{1 - r_N}{1 + r_N}e^{-2i\theta_N}$$

$$\vec{Q_W} = \frac{1 - r_W}{1 + r_W}e^{-2i\theta_W}$$

Thus we have

$$\rho = \left|\frac{\frac{1 - r_N}{1 + r_N}e^{-2i\theta_N} - \frac{1 - r_W}{1 + r_W}e^{-2i\theta_W}}{1 + \left(\frac{1 - r_N}{1 + r_N}\right)\left(\frac{1 - r_W}{1 + r_W}\right)e^{-2i(\theta_N - \theta_W)}}\right|$$

If we let

$$\theta = \theta_W - \theta_N$$

$$\rho = \left|\frac{\frac{1 - r_N}{1 + r_N}e^{2i\theta} - \frac{1 - r_W}{1 + r_W}}{1 + \left(\frac{1 - r_N}{1 + r_N}\right)\left(\frac{1 - r_W}{1 + r_W}\right)e^{2i\theta}}\right|$$

After considerable manipulation, this expression reduces to

$$\rho = \left| \frac{-(r_N - r_W)\cos\theta + j(1 - r_N r_W)\sin\theta}{(1 + r_N r_W)\cos\theta - j(r_N + r_W)\sin\theta} \right|$$

$$= \sqrt{\frac{(r_N - r_W)^2 \cos^2\theta + (1 - r_N r_W)^2 \sin^2\theta}{(1 + r_N r_W)^2 \cos^2\theta + (r_N + r_W)^2 \sin^2\theta}}$$

Using the above equations for ρ and p with r_N, r_W and θ as input data, a computer program can be used to reduce the data. Figure 12 is a typical punch-through probability plot obtained from ellipticity measurements on S-band antennas. In this case, the narrow-beam antenna was a turnstile-fed end-fire cigar type and the wide-beam antenna was a simple turnstile antenna. Both antennas were circularly polarized.

2. Determination of ρ and p from Charts. ρ can also be determined graphically by means of transformations performed on the transmission line chart[17]. ρ is, in fact, the radius to the point representing the wide-beam antenna after we have rotated the sphere to bring the point corresponding to the narrow-beam antenna to the center of the chart.

To enable the complete punch-through problem to be solved graphically, a curve has been prepared (Figure 13) that is a plot of the punch-through probability, p, as a function of the threshold (t) for several values of ρ. To facilitate the use of this curve with a conventional transmission line chart, the curves are labeled not in terms of ρ but in terms of Z, the corresponding impedance at the standing-wave minimum on a 50-ohm transmission line. The relation between these quantities is

$$Z = 50\,\frac{1 - \rho}{1 + \rho}$$

$$\rho = \frac{1 - Z/50}{1 + Z/50}$$

$$Z_N = 50\,\frac{1 - \rho_N}{1 + \rho_N} = 50\,\frac{1 - |\vec{Q}_N|}{1 + |\vec{Q}_N|}$$

$$Z_W = 50\,\frac{1 - \rho_W}{1 + \rho_W} = 50\,\frac{1 - |\vec{Q}_W|}{1 + |\vec{Q}_W|}$$

since $$\vec{Q}_N = \frac{1 - r_N}{1 + r_N} e^{-2i\theta_N}$$

and $$\vec{Q}_W = \frac{1 - r_W}{1 + r_W} e^{-2i\theta_W}$$

For θ_N and $\theta_W = 0$ corresponding to the standing-wave minimums and substituting $\vec{Q}_N$ and $\vec{Q}_W$ into Z_N and Z_W, respectively, we obtain

$$Z_N = 50\, r_N$$

$$Z_W = 50\, r_W$$

To summarize the equation of the curve of Figure 13 is

$$p = \frac{1}{2}\left\{1 + \cos\,(\beta - \gamma)\right\}$$

where

$$\tan\,(\beta - \gamma) = \frac{2t}{1 - t^2}\,\frac{\rho}{\sqrt{1 + \rho^2}}$$

$$\rho = \frac{1 - Z'_W/50}{1 + Z'_W/50}$$

$t = T_A/T_o$ = normalized threshold ratio

$$T_o = \frac{N_o}{W_o} = \sqrt{\frac{M_N^2 + m_N^2}{M_W^2 + m_W^2}}$$

T_A = receiver threshold ratio

The following example illustrates the method.

Given a Priori	Value in Example
D_T = threshold (db) = 20 $\log_{10}$ (T)	D_T = 10 db
Measured from antenna response diagrams	
Narrow-beam antenna maximum M_N	M_N = 1.0
Narrow-beam antenna minimum m_N	m_N = 0.15
Wide-beam antenna maximum M_W	M_W = 0.6

Given a Priori	Value in Example
Wide-beam antenna minimum m_W	$m_W = 0.3$
Relative orientation θ	$\theta = 12$ degrees

d = relative orientation converted to wavelengths toward load, where 0.375λ represents a relative orientation of 0 degrees.

Compute

$Z_N = 50(m_N/M_N)$	$Z_N = 7.5$
$Z_W = 50(m_W/M_W)$	$Z_W = 25$
$d = 0.375 + \theta/360$	$d = 0.40833$
$T_o^2 = (M_N^2 + m_N^2)/(M_W^2 + m_W^2)$	$T_o^2 = 2.2722$
$t = T_A - 10 \log T_o^2$	$t = 6.435$ db

To find Z_W' (Figure 14):

1. Plot P_N by striking an arc from Z_N to the "positive phase angle" axis.
2. Draw a line from the center to d.
3. Plot P_W, striking an arc from Z_W to the line just drawn.
4. Transform P_W to P_W' by moving along a constant impedance line the same number of degrees required to move P_N to the center.
5. Find Z_W' by striking an arc from P_W' to the "impedance ohms" scale.

This yields a result of Z = 22.5.

The above assumes that both antennas will lie in the same hemisphere both before and after the transformation. This means, initially, that they both have the same sense of elliptical polarization and that they are more nearly similar than orthogonal. If this is not the case, we must think of the chart as two charts back to back, with one chart representing the front hemisphere and the other the rear. The impedance-ohms scale on the rear chart should be interpreted as having negative numbers.

To find punch-through probability, p:

In Figure 13 we proceed on the line corresponding to t = 6.4 db until we reach the curve for 22.5 ohms (interpolated). We then proceed downward and read p = 4.5 percent on the scale at the bottom.

3. Instrumentation for Taking Ellipticity Measurements. The system shown in Figure 15 can be used to simplify the taking of the ellipticity data required to compute the punch-through probability for the completely random polarization input. The data format obtained on the rectangular coordinate plotter is as shown in Figure 16. At each incidence angle measured, the plotter will produce a line for the response of each of the antennas. The top of the line represents the major axis response (M_W or M_N) and the bottom of the line represents the minor axis response (m_W or m_N). The orientation angle of the rotating linearly polarized incident polarization at the maximum response points provides the major axis orientation angles θ_W and θ_N.

D. Conclusion

It is concluded that in a multi-antenna side-lobe suppression system, careful selection of antenna types with regard to polarization tracking characteristics must be used to minimize the amount of unwanted signal reception or punch-through. A choice of antennas for a given application should be based on antenna characteristics required to satisfy the system requirement. In particular, the polarization tracking of antennas should be one of the basic factors affecting a choice of antenna type.

Acknowledgments

The authors gratefully acknowledge the contribution of many individuals who assisted in the preparation of the paper.

In the antenna analysis of Part I, George E. Skahill and Boris V. Popovich contributed in technique, computations, and pattern delineations; their checking of the MSS was particularly appreciated. The development of the system for the ellipticity measurements reported in Part II was carried out by George Belden.

References

1. S. C. Schelkunoff, "Electromagnetic Waves," Van Nostrand, New York, 1943.
2. J. D. Kraus, "Antennas," McGraw-Hill, 1950.
3. H. Jasik (Editor), "Antenna Engineering Handbook," McGraw-Hill, 1961.
4. D. W. Fry and F. K. Goward, "Aerials for Centimetre Wavelengths," Cambridge University Press, England, 1950.
5. I. Todhunter and J. G. Leathem, "Spherical Trigonometry," Macmillan, 1911.

6. H. P. Westman (Editor), "Reference Data for Radio Engineers," ITT, New York, 1960.

7. W. L. Barrow and F. M. Greene, "Rectangular Hollow-Pipe Radiators," Proc IRE, Vol 26, p 1498-1519, 1938.

8. S. G. Komlos, P. Foldes, and K. Jasinski, "Feed System for Clockwise and Counter Clockwise Circular Polarization," IRE Trans (PGAP), Vol 9, No. 6, p 577-578, November 1961.

9. J. J. Epis, "Compensated Electromagnetic Horns," Microwave Journal, Vol 4, No. 5, p 84-89, May 1961.

10. J. F. Ramsay, G. E. Skahill, and B. V. Popovich, "Polarization Characteristics of Antennas," Antenna Research and Development Symposium, University of Illinois, October 1960.

11. J. F. Ramsay, "Circular Polarization for C.W. Radar," Proceedings of a Conference on Centimetric Aerials, HMSO, London, England, 1952.

12. J. F. Ramsay, "Fourier Transforms in Aerial Theory," Marconi Review Monograph, 1948.

13. K. S. Kelliher, W. C. Scott, and N. Marchand, "Cross Polarization Effects on Antenna Radiation Patterns," IRE Conv Rec, Part I, p 153-159, 1956.

14. J. A. Steers, "Study of Map Projections," University of London Press, London, England, 1946.

15. G. A. Deschamps, "Geometrical Representation of the Polarization of a Plane Electromagnetic Wave," Proc IRE, Vol 39, No. 5, p 540-544, May 1951.

16. M. L. Kales, "Elliptically Polarized Waves and Antennas," Proc IRE, Vol 39, No. 5, p 544-549, May 1951.

17. V. H. Rumsey, "Transmission Between Elliptically Polarized Antennas," Proc IRE, Vol 39, No. 5, p 535-540, May 1951.

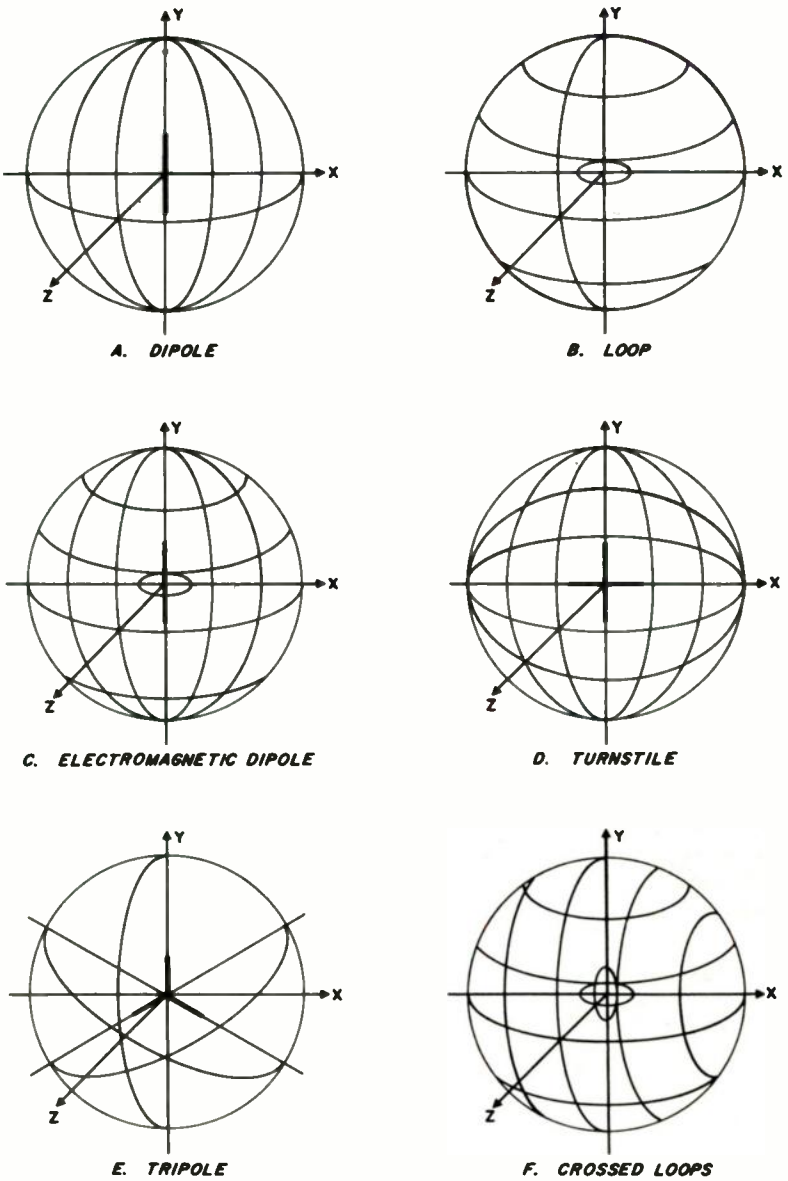
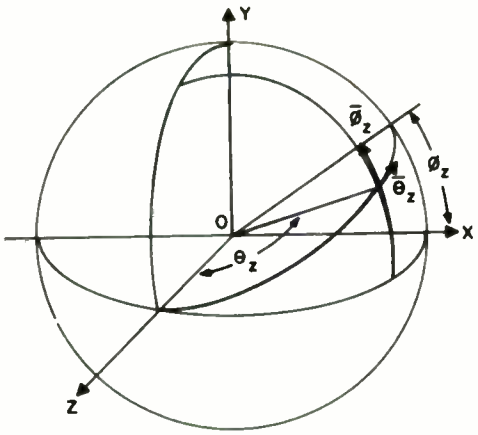
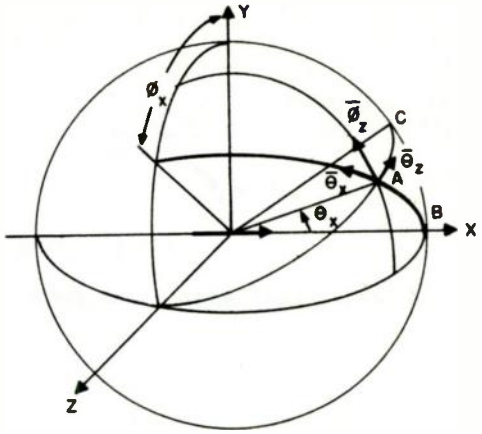


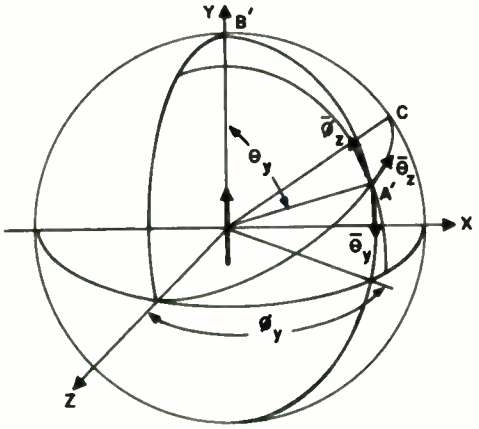
Fig. 1. Polarization characteristics of elementary radiators.



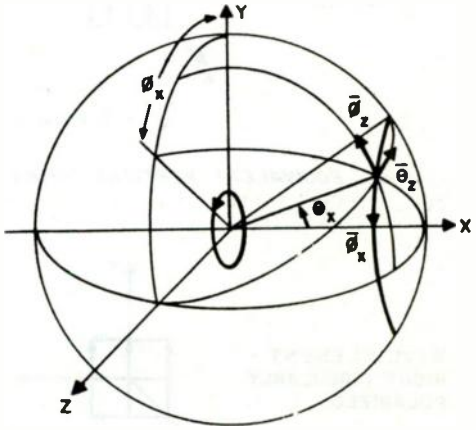
A. REFERENCE COORDINATE SYSTEM



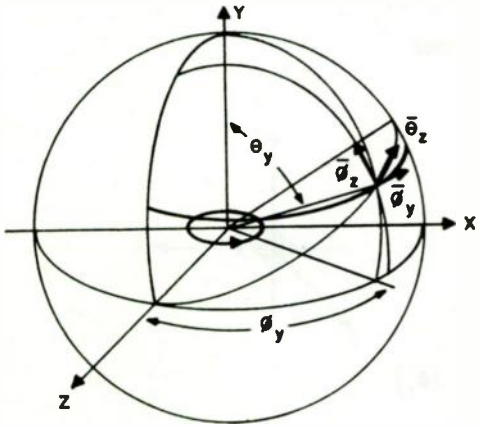
B. HORIZONTAL DIPOLE



C. VERTICAL DIPOLE



D. VERTICAL LOOP



E. HORIZONTAL LOOP

B. $\theta = \theta_z \quad \phi = \phi_z$

$$\bar{\theta}_x = \frac{1}{\sqrt{(1 + \tan^2 \phi \sec^2 \theta)}} \left\{ -\bar{\theta} + \tan \phi \sec \theta \, \bar{\phi} \right\}$$

C. $$\bar{\theta}_y = \frac{1}{\sqrt{(1 + \cot^2 \phi \sec^2 \theta)}} \left\{ -\bar{\theta} - \cot \phi \sec \theta \, \bar{\phi} \right\}$$

D. $$\bar{\phi}_x = \frac{1}{\sqrt{(1 + \cot^2 \phi \cos^2 \theta)}} \left\{ -\bar{\theta} - \cot \phi \cos \theta \, \bar{\phi} \right\}$$

E. $$\bar{\phi}_y = \frac{1}{\sqrt{(1 + \tan^2 \phi \cos^2 \theta)}} \left\{ \bar{\theta} - \tan \phi \cos \theta \, \bar{\phi} \right\}$$

Fig. 2. Resolution of polarizations.

WAVE ELEMENT
X-POLARIZED

$$\bar{g}_x = \left(1 + \cos\theta_z\right)\left(-\cos\phi_z\,\bar{\theta}_z + \sin\phi_z\,\bar{\phi}_z\right)$$

A. EQUIVALENT HORIZONTAL DIPOLE AND HORIZONTAL LOOP IN QUADRATURE

WAVE ELEMENT
Y-POLARIZED

$$\bar{g}_y = \left(1 + \cos\theta_z\right)\left(-\sin\phi_z\,\bar{\theta}_z - \cos\phi_z\,\bar{\phi}_z\right)$$

B. EQUIVALENT VERTICAL DIPOLE AND VERTICAL LOOP IN QUADRATURE

WAVE ELEMENT
RIGHT CIRCULARLY
POLARIZED

$$\bar{g}_R = -\left(1 + \cos\theta_z\right) e^{-j\phi_z}\left(\bar{\theta}_z - j\bar{\phi}_z\right)$$

C. (A) + (B) IN QUADRATURE

WAVE ELEMENT
LEFT CIRCULARLY
POLARIZED

$$\bar{g}_L = -\left(1 + \cos\theta_z\right) e^{j\phi_z}\left(\bar{\theta}_z + j\bar{\phi}_z\right)$$

D. (A) − (B) IN QUADRATURE

Fig. 3. The obliquity vector $(1 + \cos\theta)\,\bar{g}$.

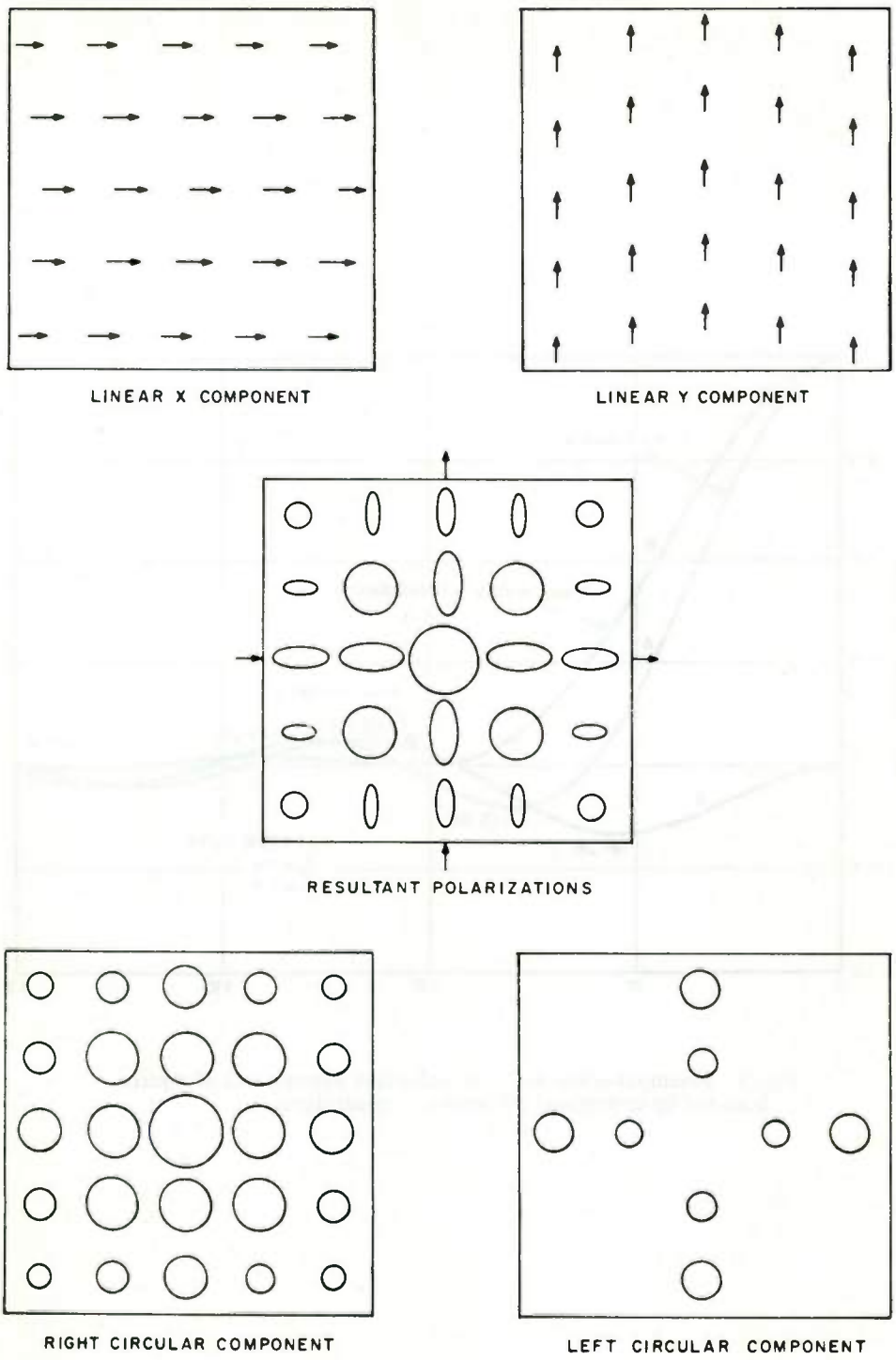


Fig. 4. Circularly polarized aperture components of square horn fed by orthogonal modes in quadrature.

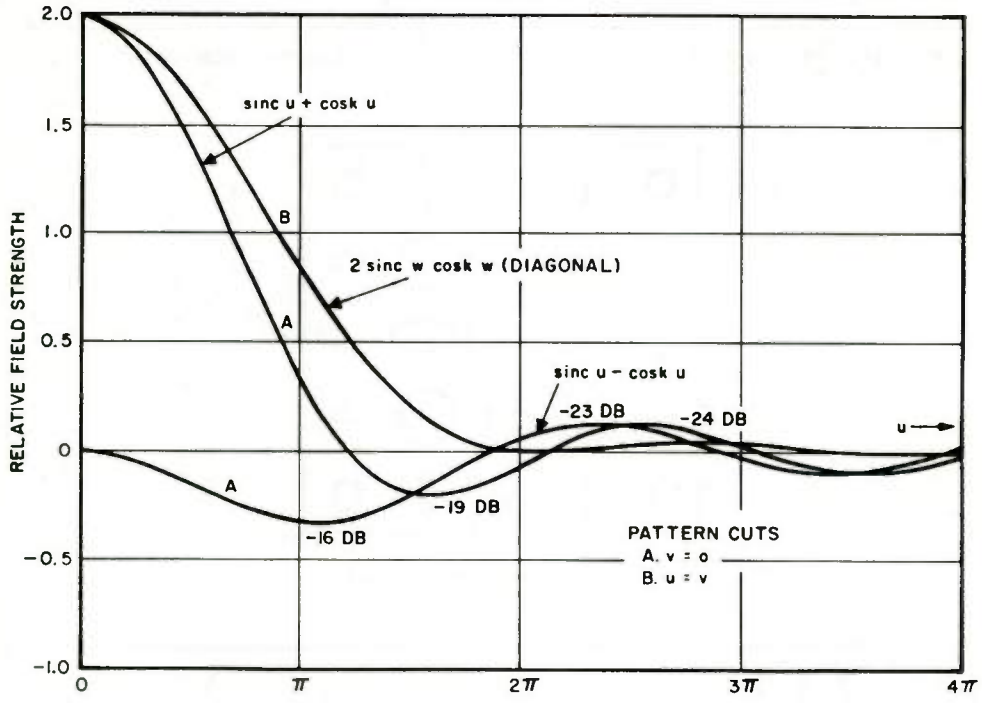
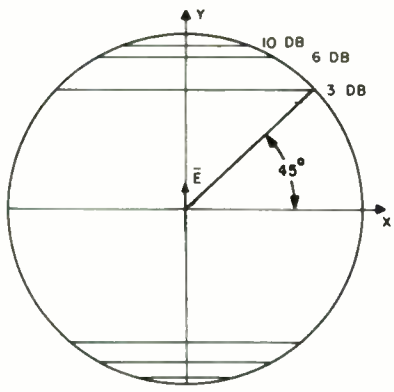
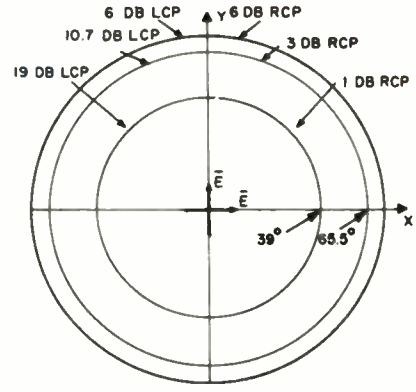


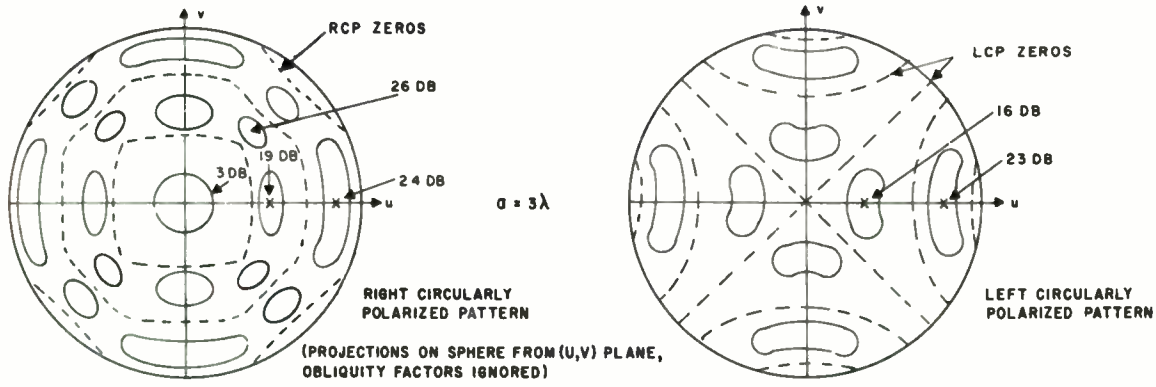
Fig. 5. Principal-plane and diagonal-plane pattern cuts of square horn fed by orthogonal TE modes in quadrature.



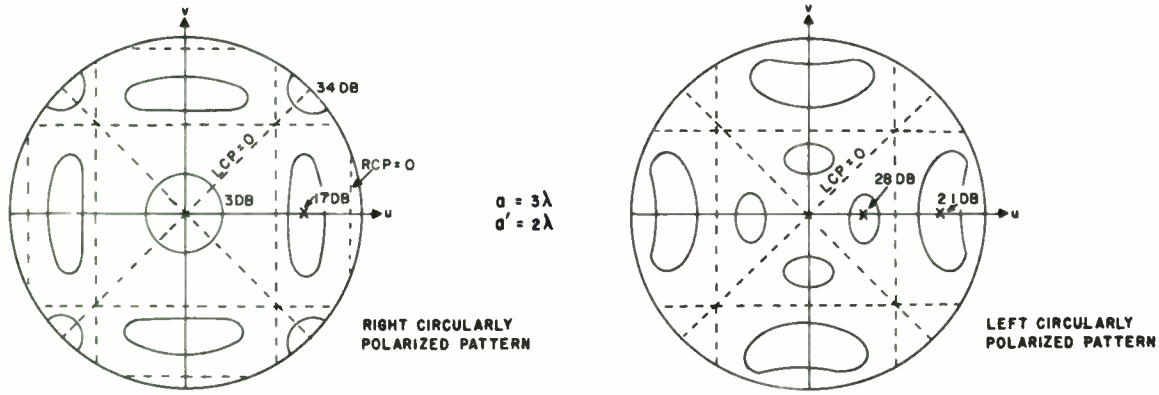
ELECTRIC, MAGNETIC, AND ELECTROMAGNETIC DIPOLES



TURNSTILE



"CIRCULARLY POLARIZED" HORN



TOOTHED HORN

Fig. 6. Contour patterns of right and left circular polarization components.

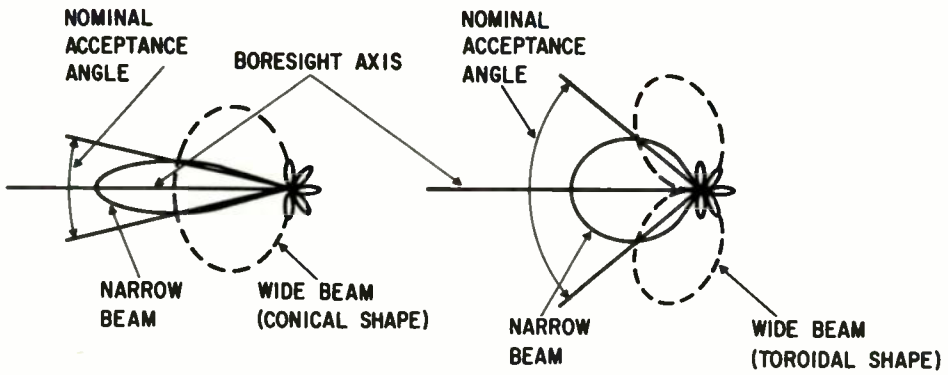


Fig. 7. Typical side-lobe suppression system antenna patterns.

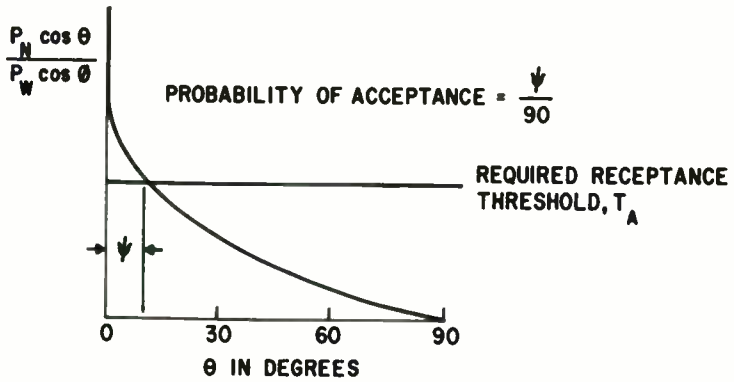


Fig. 8. $\left(\frac{P_n \cos\theta}{P_w \cos\phi}\right)$ vs polarization of incident radiation for orthogonal linearly polarized antennas.

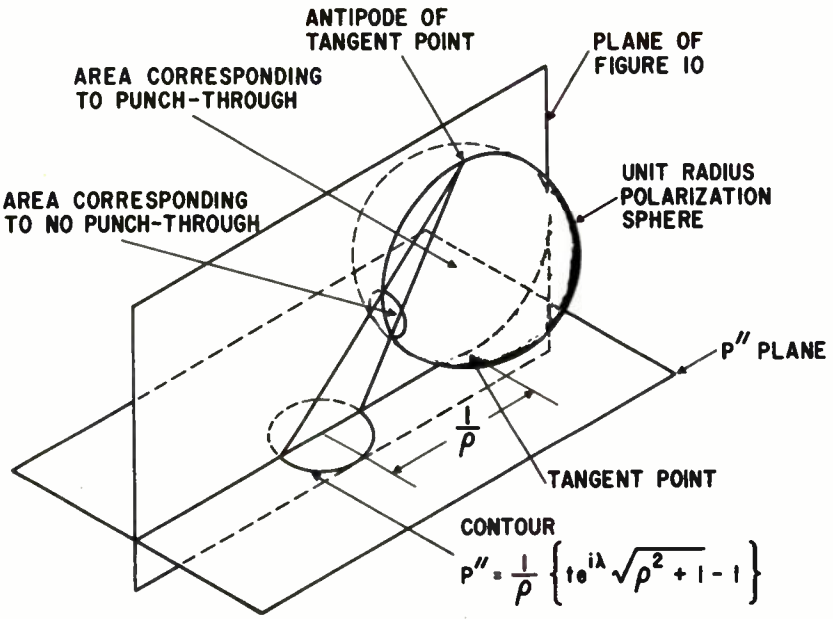


Fig. 9. Stereographic projection of punch-through contour from P'' plane to unit radius polarization sphere.

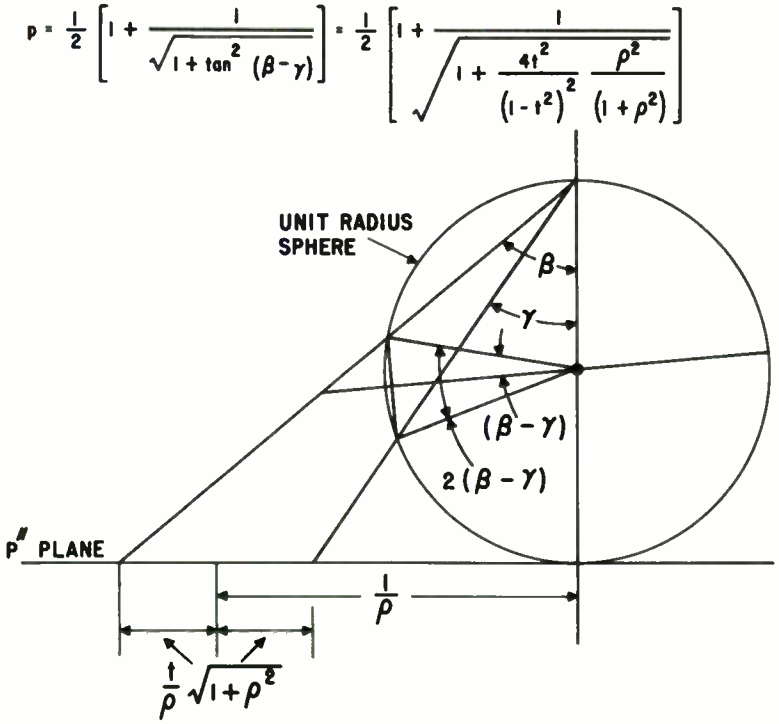


Fig. 10. Details of projection of punch-through contour from P'' plane to unit radius polarization sphere.

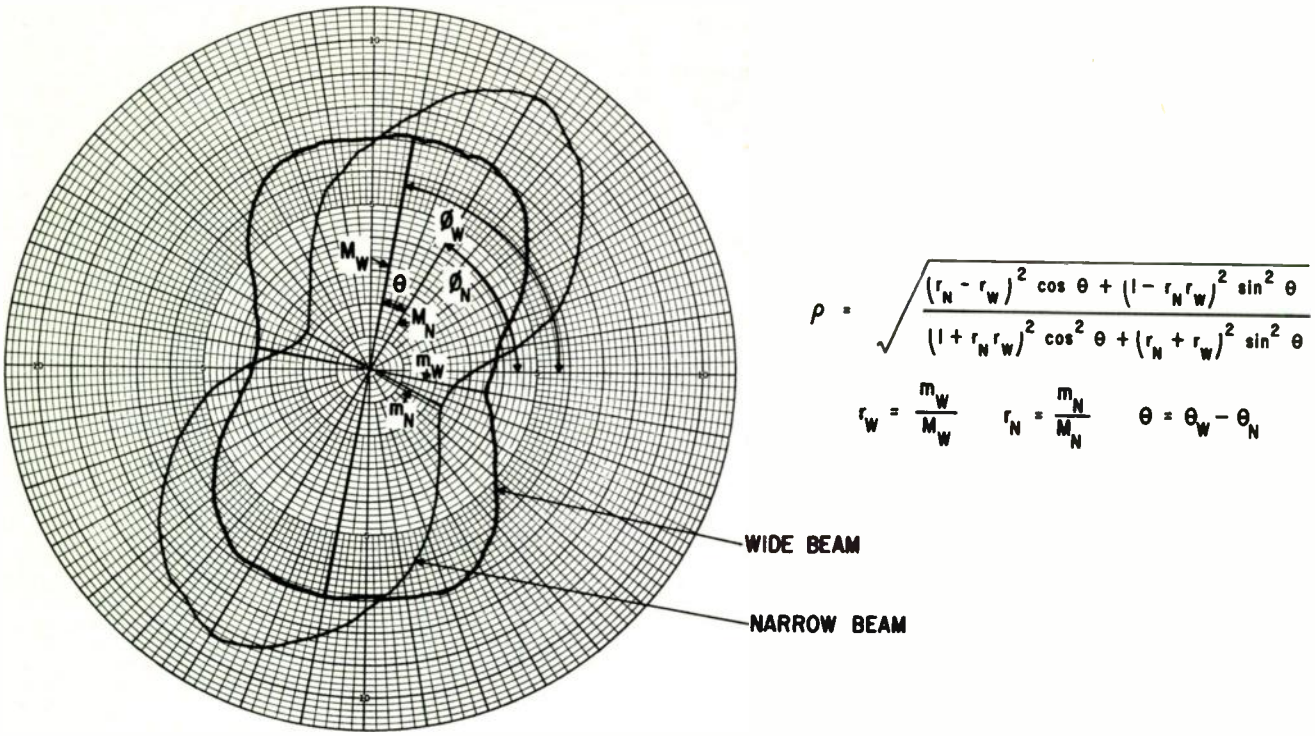


Fig. 11. Typical ellipticity patterns for circularly polarized wide-beam and narrow-beam antennas.

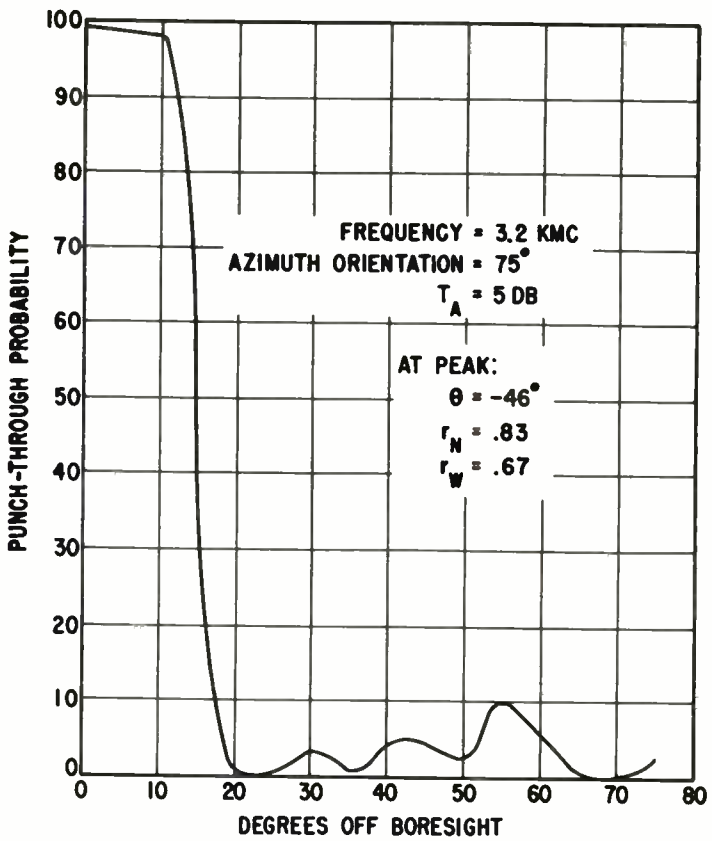


Fig. 12. Punch-through probability vs angle off boresight.

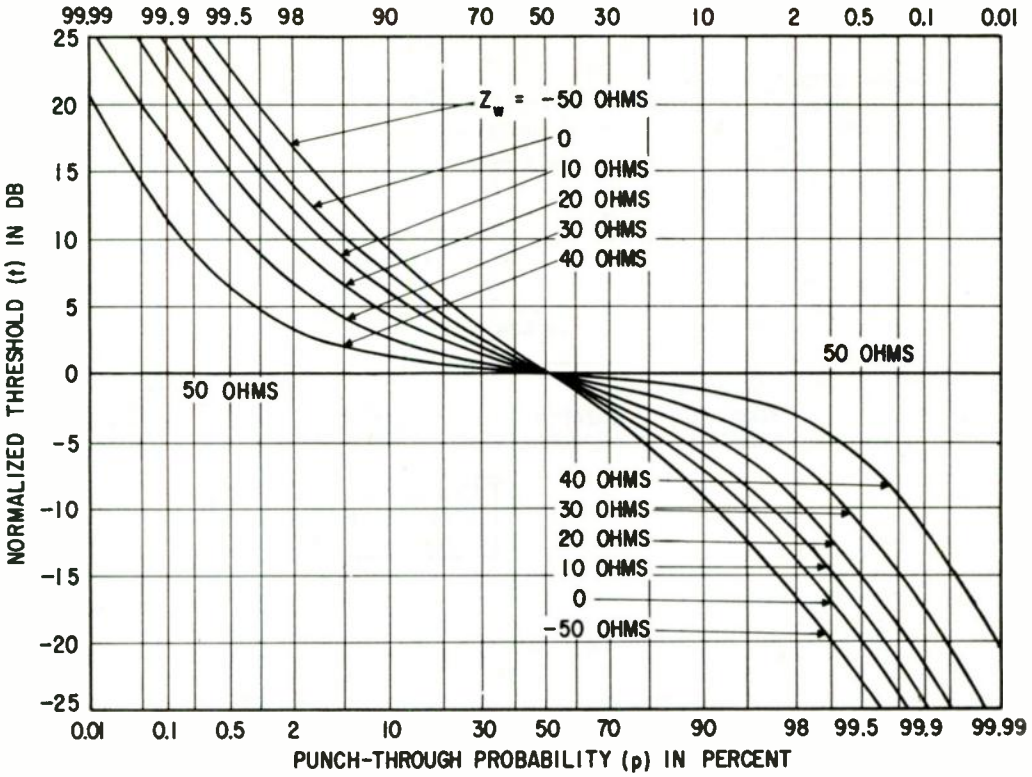
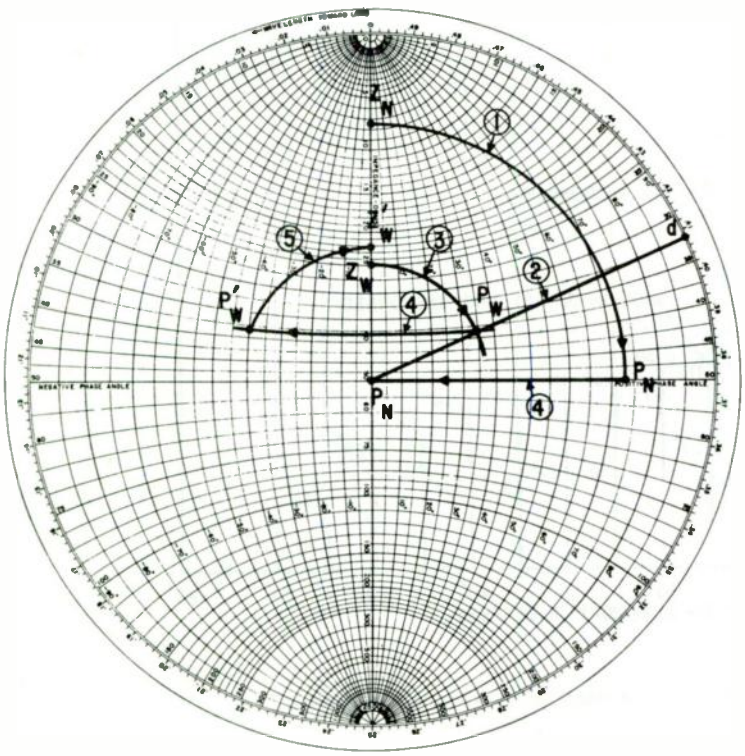


Fig. 13. Plot of punch-through vs normalized threshold for several values of Z_w.



$$\rho = \frac{1 - \frac{Z'_W}{50}}{1 + \frac{Z'_W}{50}} \qquad Z_W = 50\, r_W$$

$$Z_N = 50\, r_N \qquad d = .375 + \frac{\theta}{360}$$

Fig. 14. Illustration of use of Smith chart to obtain punch-through probability.

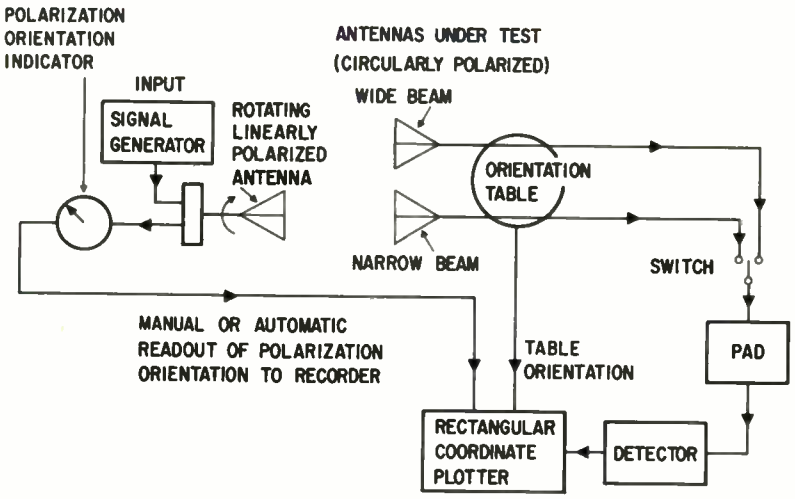


Fig. 15. Test instrumentation for taking ellipticity data.

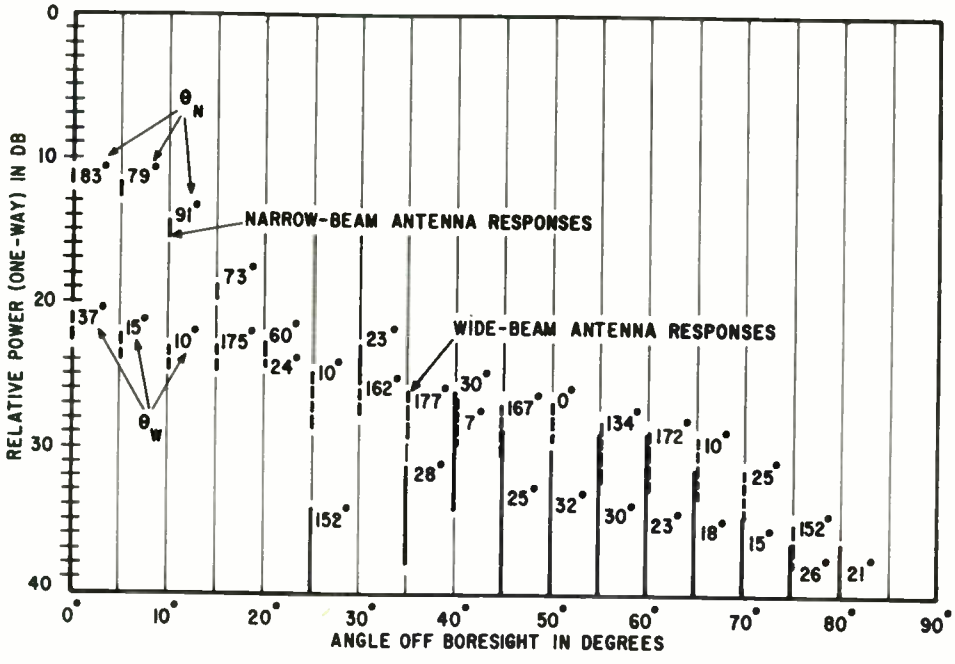


Fig. 16. Ellipticity data format obtained from rectangular plotter.

LOG PERIODIC CIRCUIT ANALYSIS

R. Mittra, Electrical Engineering Department
University of Illinois,
Urbana, Ill.

Abstract

The paper deals with the analysis of a class of log-periodic circuits. Closed form expressions are developed for the Foster type LP networks with or without losses. It is proved that the locus of the input impedance of these networks on a reflection coefficient chart, e.g., on the Smith chart, is a circle for networks of order one. It is shown that for a network of a general order the plot is not necessarily a circle.

The input impedance of a log periodically loaded transmission line is studied. It is found that for certain conditions the reflection coefficient plot is very close to a circle but in general it generates closed curves of various shapes.

SCANNING CHARACTERISTICS OF TWO-REFLECTOR ANTENNA SYSTEMS*

W. D. White and L. K. DeSize

Airborne Instruments Laboratory
A Division of Cutler-Hammer, Inc.
Deer Park, Long Island, New York

Summary

The off-axis properties of two-reflector systems of the Cassegrain and Schwarzschild type have been investigated on a theoretical basis to determine the allowable scan angle. Design parameters for systems of this type as a function of beamwidth, scan angle, magnification, and aperture block have been obtained and are presented in graphical form. In addition, a modification of the Schwarzschild system that involves overcompensating the reflecting surfaces has resulted in a larger allowable scan angle. The effects of astigmatism, location of the best focus for off-axis performance, and the effects upon the radiation patterns are presented in graphical form. A method of reducing the effects of aperture block by using the secondary reflector as a secondary source of radiation are also described.

I. Introduction

In recent years, with the advent of very high resolution antenna systems for radar, radio astronomy, and the tracking of space vehicles, an increasing amount of interest has been generated in two-reflector systems of the Cassegrain type. This type of system is characterized by the ability to place the feeds at the rear of the primary reflector and thus alleviate some of the problems associated with the transmission of high powers and the use of low-noise amplifiers on reception. Because of the extremely narrow beamwidth of these high-resolution systems, it is desirable to be able to scan the beam over some reasonable solid angle to ease the target acquisition problem. The off-axis or scanning characteristics of this type of system are then an important tool in determining their usefulness for specific applications.

At this point, it should be stated that by scanning characteristics we are talking about small physical angles (in the order of ± 5 degrees), though the actual number of beamwidths scanned might be large. In all two-reflector systems, the amount of scanning that can be obtained is limited by the geometry of the system--for example, an appreciable amount of the incident energy must be intercepted by the secondary reflector in order to obtain good beam formation. Thus, to obtain a reasonable scan angle, it is not desirable to use too high a magnification, since for a given primary reflector the higher the magnification the smaller the size of the secondary reflector. This is contrary to the usual requirement of minimum aperture block; thus, at best a compromise situation must exist between scan angle and aperture block.

Perhaps the most common of the two-reflector systems is the Cassegrain system, which uses a paraboloid as the primary reflector and a hyperboloid as the secondary reflector. The characteristics of this system have been discussed previously in the literature[1,2], and it is not the purpose of this paper to consider this system in detail but rather to use it for a comparison with the Schwarzschild system[3] which has better off-axis characteristics.

II. Schwarzschild System

In a superficial way, the Schwarzschild telescope resembles the more familiar Cassegrain telescope. Like the Cassegrain system, it is a two-mirror system in which rays arriving parallel to the axis are perfectly focused to a point. Unlike the Cassegrain telescope, however, the mirror cross sections are not simple conic sections but are transcendental curves. The design satisfies the Abbé sine condition and, therefore, the primary coma effects are eliminated. Thus, the performance of the Schwarzschild telescope for small finite feed displacements off axis is superior to that of the equivalent Cassegrain telescope.

In optics, the Schwarzschild system has been largely superseded by the Schmidt camera, because the Schmidt camera can provide a wider field of view. As the prototype for a scanning antenna, however, the Schwarzschild system has much to recommend it. Unlike the Schmidt camera and similar designs, it uses no refracting elements. Therefore, it is not subject to the power limitations associated with various forms of dielectric lenses. Although a single reflector can be made to satisfy the Abbé sine condition by using a stepped surface, such a design introduces an element of frequency dependence not present in the Schwarzschild design. Finally, as in the Cassegrain design, the focal surface can be placed in the rear of the primary reflector out of the beam-forming region. Consequently, bulky plumbing arrangements and elaborate forms of scanning apparatus can be used without interfering with the beam formation.

* The work reported in this paper has been sponsored by the Rome Air Development Center, Air Research and Development Command, under Contract AF 30(602)-1980.

Despite these advantages, the Schwarzschild design does not appear to have been widely used. In part, this may be because it provides a somewhat limited field of view as compared with such systems as the torus reflector. We believe, however, that at least in part, the Schwarzschild system has been neglected because of a lack of published information on its performance capabilities and suitable design procedures. The principal references have been published in connection with the development of the AN/MPG-1 radar. In this case, the system was successful in meeting its design objectives. The available literature, however, fails to give a design procedure that can be generalized to other applications.

The present study was undertaken to alleviate the lack of information available. Several Schwarzschild configurations were analyzed using a ray-tracing technique. As a result, we are able to draw certain general conclusions and suggest a design procedure.

A. General Conclusions

1. The Schwarzschild telescope is an excellent prototype for a scanning antenna that is to produce very narrow beams over small to moderate scan angles. Unlike the simple parabolic reflector, the performance of a Schwarzschild antenna cannot be evaluated simply in terms of the number of beamwidths in the allowable scan angle--that is, the statement that a given design will scan a 0.5-degree beam ±5 degrees does not mean that the same antenna will scan a 1-degree beam ±10 degrees. The same design will probably be capable of scanning a 0.1-degree beam more than ±2 degrees. The allowable scan angle decreases as the magnification is increased. The aperture block that is due to the secondary reflector also decreases with increasing magnification. Thus, when the required scan angle is small, favorable designs having little aperture block are possible.

Unlike Luneberg lenses, Schmidt cameras, and similar designs requiring refracting elements, the Schwarzschild telescope is not subject to the power limitation imposed by the presence of a refracting medium. Furthermore, unlike the stepped parabola designs, the Schwarzschild telescope satisfies the Abbé sine condition without introducing a bandwidth-limiting frequency dependence.

2. The Schwarzschild antenna will always produce better off-axis beams than the Cassegrain antenna. The Schwarzschild design gives a coma coefficient that approaches zero with zero slope on the axis. The corresponding Cassegrain antenna will have a coma coefficient approaching zero with a finite slope. The other aberrations appear to be similar between the two antennas. Thus, the Schwarzschild antenna will yield better off-axis performance than the corresponding Cassegrain antenna. In certain situations, the improvement may be slight but there will always be an improvement.

3. The Schwarzschild design is also applicable to moderately wide-angle scans in the Wheeler twist or the folded pillbox form. Designs in which the allowable scan angle is appreciable (say ±5 degrees or more) tend toward a configuration in which the secondary reflector is large enough to block an appreciable portion of the primary aperture. Such designs are still usable if the aperture block can be eliminated by independent means. Two systems for doing this are the Wheeler twist configuration and the folded pillbox design.

4. For very wide scan angles, the Schwarzschild design is not competitive with other antenna types. When an attempt is made to design a Schwarzschild antenna to cover a very wide scan angle (say around ±15 degrees) the resulting reflector design is so distorted as to make alternative systems such as spherical or torus antennas preferable.

B. Computation of Surface Contours

1. Abbé Sine Condition. The Schwarzschild telescope is a two-mirror system designed to satisfy the Abbé sine condition. Most discussions of the Abbé sine condition occur in the literature of geometrical optics. The following review is presented for the benefit of those unfamiliar with this literature.

Figure 1 shows the cross section of the Schwarzschild antenna. A ray, parallel to the x-axis and separated from it by distance H, is reflected first by the primary reflector and then by the secondary reflector. The angle between the final segment of the ray and the x-axis is α. For the system to satisfy the Abbé sine condition, the following relation should hold.

$$\frac{H}{\sin \alpha} = F$$

where F is a constant independent of H.

Figure 2 shows an expanded cross-section view of portions of the antenna. The reflector focuses rays that are parallel to the x-axis to a point. Stated another way--the reciprocity principle--a spherical wave emanating from the focal point will emerge from the system as a plane wave whose wavefront is normal to the axis. For a scanning antenna, if the source of the spherical wave is moved a small distance (y) normal to the x-axis, a new plane wave should emerge from the system with a tilt angle θ relative to the original plane wave. The path length of the ray shown in Figure 2 is shortened by the amount

$$\Delta p \cong y \sin \alpha, \qquad (1)$$

assuming that the displacement (y) is small relative to the dimensions of the system. If the emergent wave is a plane wave with the tilt angle θ, the path shortening will be

$$\Delta p = H \tan \theta \qquad (2)$$

Equating these two expressions for Δp, we obtain

$$y \sin \alpha = H \tan \theta \tag{3}$$

or

$$\frac{H}{\sin \alpha} = \frac{y}{\tan \theta} = F \tag{4}$$

This derivation is an approximation based on y being small relative to the dimensions of the system (θ is a small angle). When the deviations are substantial, superior results can sometimes be obtained by departing from a rigorous satisfaction of the Abbe sine condition.

2. Geometrical Relations. As seen in Figure 1, there are two segments of the ray between the focal point and the point of reflection from the primary reflector. We designate the lengths of these segments by ρ and ℓ. If the emergent wavefront is to be a plane wave, the following equation should hold.

$$\rho + \ell = t_1 + t_2 + x \tag{5}$$

From purely geometrical relations we can see that

$$\rho \cos \alpha - \ell \cos \alpha' = t_2 - t_1 + x \tag{6}$$

and

$$\rho \sin \alpha + \ell \sin \alpha' = H \tag{7}$$

Since the angle of reflection is equal to the angle of incidence, we can write

$$\alpha - i = i - \alpha' \tag{8}$$

and from the small detail of point A (Figure 1) we can see that

$$\tan i = \frac{d\rho}{\rho d\alpha} \tag{9}$$

Equations 4 through 9 now constitute a set of six equations relating seven variables (α, α', ρ, ℓ, x, H, and i) and three parameters (F, t_1, and t_2). We can eliminate all variables but ρ and α, obtaining thereby the equation

$$\frac{1}{\rho}\frac{d\rho}{ds} = -\frac{\rho + Fs - t_1 - F}{(1 - s)(t_1 - Fs)} \tag{10}$$

where

$$s = \sin^2 \frac{\alpha}{2}$$

Equation 10 then is a differential equation for the surface of the secondary reflector. The solution can be written as

$$\frac{1}{\rho} = \frac{s}{t_1} + \frac{1}{t_2}\left(1 - \frac{sF}{t_2}\right)^{\frac{F}{F-t_1}}\left(1 - s\right)^{-\frac{t_1}{F-t_1}} \tag{11}$$

We thus have a definitive equation for the secondary reflector in polar coordinates.

To compute the surface of the primary mirror, we first eliminate α' from equations 6 and 7 obtaining

$$\ell^2 = \rho^2 - 2\rho \cos \alpha \,(t_2 - t_1 + x) + (t_2 - t_1 + x)^2 + (F^2 - 2F\rho) \sin^2 \alpha$$

We then eliminate ℓ from this equation and equation 5. Making the substitutions,

$$H = F \sin \alpha$$

$$\sin \alpha = 2\sqrt{s(1 - s)}$$

$$\cos \alpha = 1 - 2s$$

and solving for x, we obtain

$$x = -t_2 + (1 - s)\frac{F^2 s + \rho(t_1 - 2Fs)}{t_1 - \rho s}$$

or substituting for ρ, its value, we obtain after some reduction

$$x = -t_2 + \frac{F^2}{t_1} s(1 - s) + t_2(1 - s)^{\frac{F}{F-t_1}}\left(1 - \frac{sF}{t_1}\right)^{\frac{F-2t_1}{F-t_1}} \tag{12}$$

whereas

$$H = F \sin \alpha = 2F\sqrt{s(1 - s)} \tag{13a}$$

Without giving the details of the derivation, note that the primary mirror can also be expanded in a Taylor series of the form

$$\frac{x}{F} = a_2 \frac{H^2}{F^2} + a_4 \frac{H^4}{F^4} + a_6 \frac{H^6}{F^6} + \ldots \tag{13b}$$

The first few coefficients are:

$$a_2 = \frac{m}{4}$$

$$a_4 = -\frac{1}{32}$$

$$a_6 = -\frac{m+5}{384}$$

$$a_8 = -\frac{2m^2 + \ldots}{6144}$$

$$a_{10} = -\frac{6m^3 + \ldots}{122,880}$$

$$a_{12} = -\frac{24m^4 + 290m^3 + 1555m^2 + 4850m + 8401}{2,949,120}$$

In this expression, m is the magnification and t_1 is set equal to t_2.

3. Relation of Parameters. A discussion of the relations between the various parameters of a Schwarzschild antenna will be facilitated by introducing the concept of the equivalent Cassegrain. Figure 3 shows a cross section through a generalized Cassegrain antenna. It consists of a parabolic primary mirror and a hyperbolic secondary. A ray arriving parallel to the axis and striking the primary reflector at point A is reflected toward the primary focal point B. The secondary mirror has a hyperbolic section and one of the two hyperbolic foci is coincident with the primary focal point at B. The ray is thus intercepted at point C and directed to the secondary focal point D. The equations for the reflecting surfaces may be written as

$$x_1 = \frac{H_1^2}{4f_1} \tag{14}$$

$$x_2 = \frac{f_3 - f_2}{2}\left\{\sqrt{1 + \frac{H^2}{f_2 f_3}} - 1\right\} \approx (f_3 - f_2)\left\{\frac{H^2}{4f_2 f_3} - \frac{H^4}{16 f_2^2 f_3^2} + \frac{H^6}{32 f_2^3 f_3^3} + \ldots\right\} \tag{15}$$

The focal lengths f_1, f_2, and f_3 are as shown in Figure 3. The equivalent focal length of the Cassegrain system is magnified by the ratio f_3/f_2 or in other words

$$f_{eq} = f_1 \frac{f_3}{f_2} = mf_1 \tag{16}$$

In the following, we shall limit ourselves to the case where the secondary focus lies on the surface of the primary reflector--that is, we limit ourselves to the case where $t_1 = t_2$ or b = 0. We can then write

$$f_3 + f_2 = f_1 = (m + 1) f_2$$

therefore

$$f_1 = \frac{f_{eq}}{m} \tag{17}$$

$$f_2 = \frac{f_1}{m+1} = \frac{f_{eq}}{m(m+1)} \tag{18}$$

$$f_3 = mf_2 = \frac{f_{eq}}{m+1} \tag{19}$$

In other words, the surfaces are completely defined in terms of the equivalent focal length and the magnification.

Since the Schwarzschild and the Cassegrain antennas differ only in the higher-order terms of the surface equations, we can define the two antennas as equivalent if they have the same surface curvatures on the axis and the same surface separations.

The parameters of the Schwarzschild antenna can then be written

$$F = f_{eq}$$

$$t_1 = t_2 = f_3 = \frac{f_{eq}}{m+1} \tag{20}$$

4. Curve Fitting. Since both reflector surfaces of the Schwarzschild system are described by irrational equations, they do not lend themselves to analytic ray-tracing procedures. Therefore, it is preferable to describe the surfaces by either of two polynomial approximations: (1) a Taylor series terminated after a finite number of terms, or (2) a polynomial that is a least-squares fit to a finite number of surface points computed by using the exact equations. Both methods have been tried. In general, the least-squares fit yields a polynomial that is a more accurate representation of the reflector surfaces over a finite extent than the representation from the truncated Taylor series.

C. Analysis of Performance

1. Ray Tracing. To evaluate the ability of an antenna reflector system to produce off-axis beams, one approach would be to trace rays outward from the focal point to a plane surface normal to the intended beam direction. The variation of the lengths of these rays would then constitute a

direct measure of the path error. This path error could then be multiplied by the appropriate constant to yield the phase error of the illumination pattern.

In optical ray tracing, the technique is usually the reverse of that described above. Rays are traced inward from the reference plane until they intersect an image plane. This reflects the fact that optical devices are normally judged in terms of image quality rather than in terms of equivalent beam pattern. An incidental advantage of the optical technique is that the effect of small focal plane shifts can be investigated without repeating the entire tracing operation.

The two techniques are related by the fact that a small displacement of the image point can be divided by the effective focal length of the system and the result interpreted as an angular aberration. If an inbound ray parallel to the central ray and displaced from it by a distance H strikes the focal plane at a point ϵ units from the central ray, then the angular aberration is $A_e = \epsilon/F$ radians. This means that if a ray is traced outward from the focal point emerging H units from the central ray, it will be inclined at an angle A_e relative to the central ray.

Since the wavefront is normal to the direction of propagation, the tilt of the wavefront is also A_e radians. Assuming that A_e is so small that $A_e = \sin A_e = \tan A_e$, then the total wavefront distortion or path-length error is obtained from the relation

$$\text{Path error} = \Delta p = \int A_e \, dH \cos \Theta$$

The cos Θ factor is introduced since dH is measured in the plane of the aperture, whereas the integration should take place in a plane normal to the beam.

In conducting this study, we were fortunate in having available the Elgeet Optical Program. This is a program written for the LGP-30 Computer by Mr. Gordon Spencer of the University of Rochester for the Elgeet Optical Company. The ray-tracing portion of this program is capable of handling any combination of rotationally symmetric optical surfaces, reflecting or refracting, and spheric or aspheric up to a total of 35 surfaces. Data for aspheric surfaces are inserted by giving the coefficients of a 10^{th}-order polynomial. Ray data are inserted by giving the coordinates of the ray intersection with a plane tangent to the first surface and the direction cosines of the ray. The output is in the form of the coordinates of the ray's intersection with the focal plane and data pertaining to the angle with which the ray strikes the focal plane.

2. Reduction Program. For optical applications, image plane data is usually sufficient; however, the antenna engineer is more interested in beam patterns and needs to know the phase error of the illumination. To provide this type of information, an optical reduction program was written. This program permits the image plane data to be translated in terms of aperture plane phase errors, and provides a means for evaluating the effect of focal plane shifts. The procedure is as follows.

The outputs of the ray-trace program in the form of curves of y versus H and tan α versus H are applied as the input to a least-squares adjustment program to obtain a polynomial approximation to these curves--that is,

$$y \cong B_0 + B_1 H + B_2 H^2 + \ldots + B_{10} H^{10} \qquad (21)$$

$$\tan \alpha \cong C_0 + C_1 H + C_2 H^2 + \ldots + C_{10} H^{10} \qquad (22)$$

The intercept of the shifted focal plane is

$$y' = y + \Delta F \tan \alpha \qquad (23)$$

where ΔF is the focal plane shift. The angular error is then

$$A_e = \frac{y' - y'_0}{F} \qquad (24)$$

where y'_0 is the value corresponding to the central ray. Therefore, a new polynomial is formed as

$$A_e \approx \sum_0^{10} D_i H^i, \qquad (25)$$

where

$$D_0 = 0$$

$$D_i = \frac{B_i + \Delta F C_i}{F}, \quad i \neq 0 \qquad (26)$$

Normally, ΔF is chosen so that $D_1 = 0$. This is the value that provides best focusing for rays near the center of the aperture. Where there is substantial spheric aberration, however, it is possible that better operation for the aperture as a whole can be obtained by using a slightly different value for ΔF. Therefore, the program is arranged so that it can be caused to compute its own focal position (setting $D_i = 0$) or it can compute with a focal position that is specified as part of the input data.

Integrating the angular aberration polynomial, we obtain another polynomial expression for the path error

$$\Delta P = \cos \Theta \, \Sigma E_i H^{i+1} \qquad (27)$$

where

$$E_i = \frac{D_i}{i + 1} \qquad (28)$$

3. Overcompensation. Figure 4 shows the type of result obtained from the reduction program. This is a curve of the path error as a function of position along the aperture for two different angles off axis.

The unit of measure used is the focal length of the primary reflector. For the configuration plotted, this is one-half of the effective focal length of the complete system. The curves obtained for the equivalent Cassegrain antenna are shown for comparison. The Cassegrain system is equivalent to the Schwarzschild system in the sense that it has the same reflector spacing and the same reflector curvatures at the apex. Curves are shown for the cases of 5-degree off-axis operation and 10-degree off-axis operation. The path error is less for the Schwarzschild system than for the Cassegrain system.

In the 5-degree case, the improvement of the Schwarzschild system over the Cassegrain system is much more pronounced than in the 10-degree case. In addition, for the 5-degree off-axis case, the shape of the Schwarzschild curve differs markedly from that for the Cassegrain; whereas the difference in shape is minor for the 10-degree curve. This suggests that, though the Abbe' sine condition ensures the absence of primary coma distortion for small displacements off axis, a 5-degree displacement is only marginally small and a 10-degree displacement is definitely not small.

The implication is that, for an antenna to scan a given finite sector, better results might be obtained by departing somewhat from the true Schwarzschild design. To test this hypothesis, the following procedure was tried.

1. A primary reflector contour was chosen in which the deviation from a parabolic shape was twice that of the Schwarzschild primary reflector.

2. A corresponding secondary reflector was computed that would maintain equal path lengths to the focal point for a plane wave coming in parallel to the axis. The resulting system is termed 100 percent overcompensated.

Path-error curves are shown in Figure 5 for a range of off-axis angles. From Figures 4 and 5 it can be seen that in the region of 5 degrees, both the Schwarzschild system and the overcompensated systems give asymmetrical path-error curves, but that the direction of the asymmetry is reversed. The curve for the Schwarzschild system is roughly symmetrical at a point well to the left of center. The overcompensated system is roughly symmetrical at a point well to the right of center. The implication is that, had we introduced less overcompensation, the curve could have been made symmetrical. As will be seen, this implication is borne out.

In Figure 5, the overcompensated antenna 10 degrees off axis is still asymmetrical to the left though somewhat nearer to a symmetrical pattern than the Cassegrain antenna, or true Schwarzschild antenna. The effect of varying the compensation is much less pronounced on the right side of the graph than on the left.

At first glance, Figure 5 may appear somewhat confused especially on the left side. Closer inspection, however, will show a definite pattern. At small angles off axis (2 degrees), the curves tend to go up on the left side and down on the right. This is the condition of the negative coma. As the angle increases (4 to 6 degrees) and because of spherical aberration, the left part of the curve continues to go up but the right part reverses. At about 8 degrees off axis, the effects of coma have almost disappeared and we have a nearly symmetrical picture in which the dominant component is spherical aberration. As we continue to go still farther off axis, positive coma is apparent; the right end of the curve continues to increase but the left part decreases. Near the center, the 8-degree curve is below the 6-degree curve, though they intersect farther out. The 10-degree curve is well below the 8-degree curve on the left side, even though it is still positive at the left end. If we continue farther, the left end of the curve will eventually go negative.

It may be questioned whether the overcompensation has accomplished much, even at 10 degrees off axis, since the right end is not affected much and the path error on the left (though of reversed sign) is still large. The answer is that a symmetrical path-error curve is desirable, since this can be compensated in some degree by focus shift. For these curves, the focal point was calculated by nulling D_1, the focus-error coefficient of the angular aberration. This ensures optimum focus for the rays near the aperture center. It does not, however, ensure best focus for the entire aperture. Later in this paper, we will present examples to show how a modification of the focus shift can be used to minimize the errors over the aperture.

Since the implication from Figure 5 is that at moderate angles off axis (say 5 degrees) somewhat less than 100-percent overcompensation would be desirable, curves were computed for a 50-percent overcompensated system with a magnification of 2. In this region, the 50-percent overcompensated system appears to be near to an optimum.

4. Effects of Magnification Changes. For a given application, we have a variety of Schwarzschild configurations that can be used. Given a particular aperture diameter, we can use a long focal length resulting in a large f_1/D ratio or we can use a shorter focal length. Given the focal length, we can use a relatively shallow primary reflector and a low magnification, or we can use a deeper primary and a higher magnification. While the magnification stays fixed, the effect of varying the focal length can be determined from curves of the type shown in Figure 5, by appropriately adjusting the scales. Changing the magnification, however, necessitates a new computation.

From one point of view, that of minimizing the aperture block that is due to the presence of the secondary reflector, it is desirable to use as high a magnification as possible. On the other hand, the magnification cannot be increased indefinitely. Figure 6 shows a family of Cassegrain antennas with the same parabolic primary reflector but different magnifications and thus different hyperbolic secondary reflectors. The dashed line shows that a ray arriving parallel to the axis is reflected to the same focal point regardless of the secondary reflector used. A ray (solid line) that makes a finite angle with the axis is reflected to different focal points depending on the magnification. Note, however, that the ray misses the smallest of the secondaries completely. This is an unsatisfactory condition. As we shall see later, the performance deteriorates rapidly as this condition is approached.

Figure 7 shows what we call the "critical angle." We define this angle as the angle off axis at which a ray strikes one edge of the primary reflector, crosses over, and barely strikes the opposite edge of the secondary reflector. In other words, when rays arrive at angles greater than the critical angle, the secondary reflector is completely missed by the rays. This angle decreases with increasing magnification and decreases as the primary reflector's focal length is shortened.

Although Figures 6 and 7 are shown for a Cassegrain system rather than for a Schwarzschild system, we can expect the same general trends in the Schwarzschild system as well, since the two systems are superficially alike. In themselves, these curves do not tell us the exact limitations on the performance of any given configuration; they do indicate a trend wherein maximum usable magnification decreases as the angle of desired coverage is increased.

To examine the problem in more detail, several cases were computed with a magnification of 4. Since 100-percent overcompensation seemed too much for a magnification of 2, we used smaller degrees of overcompensation. Accordingly, the first curves computed were for the true Schwarzschild system and for systems having 25- and 50-percent overcompensation. It turned out, however, that these overcompensated systems produced results varying but little from those of the true Schwarzschild system. Greater degrees of overcompensation were then tried, and performance curves were computed for 100- and 200-percent overcompensated systems. Figure 8 shows the path error curves for the 200-percent overcompensated system.

To some degree, these curves show the same trends as those shown in Figure 5 (magnification of 2), though the effects of spherical aberration are less prominent. Comparing Figure 5 with Figure 8 indicates that, for small angles, Schwarzschild system performance is relatively independent of magnification. At large angles, however, the curves deviate substantially on the left sides, though the right sides are still similar.

Figure 9 is a plot of the coma coefficients for various cases and indicates graphically the effect of magnification. This illustration is normalized in terms of the effective focal length--that is, all the systems plotted have an effective focal length of 10 units. Consider first the curves for: (1) the single parabola, (2) the Cassegrain system with a magnification of 2, and (3) the Cassegrain system with a magnification of 4. These three curves are tangent at origin. This indicates that, for small angles off axis, the performance of a Cassegrain antenna is a function only of the effective focal length and is independent of the magnification. As we go farther off axis, however, we find that the Cassegrain curves deviate from those of the parabola; the higher the magnification, the earlier this deviation becomes apparent. This was already known in a qualitative way, but it was not known that the marked deviations set in so early for moderate magnifications.

In both cases, the Schwarzschild curves approach the origin with zero slope, but the higher magnification results in substantially greater error even at relatively small angles. This implies that for a system of given effective focal length, it is better to use a long focal length primary reflector and small magnification rather than a short focal length primary reflector and greater magnification. This is counter to the requirement of using high magnification to minimize aperture blocking. In practical cases, a compromise must be made.

Although the two curves for 100-percent overcompensated systems start with the same negative slope, the higher-magnification curve reverses direction faster and intersects the axis again at a much smaller angle. The implication here is that, for a given coverage angle, much less overcompensation is required with a low magnification system.

Figure 10 shows the same information as does Figure 9, but here we have a constant primary focal length. (Each system plotted has a primary focal length of 5 units.) From this set of curves, it is apparent that holding the primary focal length constant and increasing the magnification does reduce coma in a Cassegrain system, though the reduction is not as great with large angles off axis as it is with small angles. For the Schwarzschild systems, the lower magnification results in somewhat less coma, though the curves are not greatly different.

5. Use of Very High or Very Low Magnifications. Although low magnifications tend to increase the size of the secondary reflector, there are at least two cases in which aperture blocking that is due to the presence of the secondary is not a problem. The first of these is the case where scanning is required in only one dimension. Here, we can use a folded pillbox type of design as was used in the AN/MPG-1.

The other method of avoiding aperture blocking is by using the Wheeler polarization twist principle. This system uses a grid-type secondary reflector that reflects one polarization but is transparent to the orthogonal polarization. The primary reflector is covered with a diagonal wire grid spaced one-quarter wavelength in front of the

main reflecting surface. This converts an incident wave that is horizontally polarized to a vertically polarized wave. Thus, if the secondary reflector is arranged to reflect vertical polarization and transmit horizontal polarization, and if the feed horn is vertically polarized, then a wave emerging from the horn will be reflected by the secondary reflector. Upon striking the primary, it will be converted to a horizontally polarized wave and, on reaching the secondary a second time, will pass through.

Because of these possibilities for using low magnifications, computations were made for the Schwarzschild antennas of m = 1 and of m = 1/2. For these low magnifications the deviation from simple conic surfaces is quite apparent. These structures are not usable for values of H/f_1 as large as for the case with structures having higher magnification. On the other hand, they may be usable at larger angles off axis.

Computations showed that the path error curves were more symmetrical than the corresponding curves for the higher magnification. This is an indication that spherical aberration is becoming relatively more important than coma. The question arises as to whether it might not be possible to compensate for this effect. By leaving the secondary mirror unchanged and deforming the primary, it is possible to introduce a certain amount of negative spherical aberration for the on-axis case and thereby decrease the maximum aberration at the limit of scan. In addition, the aberration remaining can be partially compensated by an appropriate focus shift.

In cases where the required scan angle is small (say 1 or 2 degrees), it may be possible to obtain a satisfactory solution to the aperture block problem simply by using a relatively high magnification (say 10, or so). To investigate this possibility, computations were made for a Cassegrain antenna having a magnification of 10.

The antenna computed had a primary f_1/D ratio of 0.25. At about 2.6 degrees off axis, a ray striking the edge of the primary reflector is just tangent to the secondary at a point near the apex. Figure 11 shows the path-error curves at several lesser angles. Note that the performance is good at angles of 1/2 degree and 1 degree off axis, but rapidly deteriorates as the critical angle is approached.

To accommodate rays as much as 2 degrees off axis, the diameter of the secondary reflector need only be about 14 percent of the primary diameter. Thus, the blocking would be only about 2 percent of the area of a circular aperture. At this angle, the path errors are such that the system could be used with a nominal beamwidth of 1 degree. If the scan is limited to 1 degree off axis, the path error is greatly improved, and the system can be used with a nominal beamwidth of about 0.15 degree.

6. Astigmatism and Focus Shift. In most of the foregoing discussion, we have treated the optics as a two-dimensional problem. This is a legitimate procedure when the system is to be used for one dimension of scan only, as was the case with the AN/MPG-1 antenna. When a full three-dimensional system is used, we can scan in two dimensions by appropriate motion of the feed horn. Here, however, another source of aberration must be considered. This additional aberration, which is called astigmatism, results from the fact that focusing in the direction of the scan does not occur at the same point as the focus in the orthogonal direction. The effect is shown in Figure 12, which illustrates a cross section through a 200-percent overcompensated system having a magnification of 4. Shown on this plot is the locus of the meridional focal points. Note that this locus is sharply curved. To check for astigmatic effects, ray-tracing computations were made for a fan of rays lying in a plane normal to the scan direction. This was done for fans at 4 degrees and 8 degrees off axis. The ray-tracing results were then processed in the normal manner, and the transverse focal points were computed. The results are shown as crosses in Figure 12. Note that the meridional and transverse focal points are indeed separate.

To evaluate the seriousness of the astigmatic effects, path-error computations were made with the focal point specified. Curves were computed for both meridional and transverse fans at the meridional focus, at the transverse focus, and at an intermediate point midway between. The results are shown in Figures 13 and 14. Figure 13 shows the results for the meridional fan 4 degrees off axis. Curve 1 is the path error at the meridional focus. This is the same as the 4-dgree curve of Figure 8. Curve 2 results when the focal point is shifted midway between the meridional and transverse foci, and curve 3 results when the focal point is at the transverse focus. Note that the maximum error is not increased as we go from curve 1 to curve 2. In fact, the optimum focus would appear to be a point intermediate between the meridional and average foci. Accordingly, curve 4 was computed, which is for a focal point one-quarter of the way from the meridional focus to the transverse focus. This point appears to be near the optimum focal point for the transverse fan.

Figure 14 shows the corresponding curves for the transverse fan. If we have a circular aperture, the focal point that minimizes the errors in the meridional fan is not the focal point that minimizes the errors in the transverse fan. The best compromise appears to result in a maximum path error about 50 percent greater than that computed for the meridional fan at the meridional focus. This is the case when we have a pencil beam scanned in two dimensions; however, if we had a rectangular aperture in which the aperture dimension normal to the direction of scan were less than one-half the aperture dimension in the direction of scan, then we could largely ignore the astigmatic effects--that is, astigmatism introduces a significant but not fatal effect in the case of a pencil beam, but can be largely ignored in the case of the usual type of fan-beam radar.

As previously indicated, at a magnification of 4 the Schwarzschild and the overcompensated systems have little advantage over the Cassegrain system when we are as far as 8 degrees off axis.

If 8-degree coverage is desired, a lower magnification should be used. In this case the coma effect is larger than the astigmatic effect, so that the optimum compromise focus is very near to the optimum meridional focus, and the maximum path error is about that occurring in the meridional fan.

Similar computations have been made for the 100-percent overcompensated system having a magnification of 2. At 4 degrees off axis, the average focus gives reasonably good results, the maximum path error in both the meridional and transverse fans being less than that for the meridional fan at the meridional focus. At 8 degrees off axis, the optimum focus is somewhat closer to the transverse than to the meridional focal point. Again, at the optimum focal point the maximum path error is somewhat less than that for the meridional fan at the meridional focus.

D. Design Criteria

1. Nominal Beamwidth. Although path-error curves of the type previously shown permit the evaluation of a given configuration when a design has been evolved, they are not in the most convenient form for designing an antenna to meet given specifications. For this purpose, we have prepared the curves shown in Figures 15 through 17. Roughly speaking, these curves convey the same information as the path-error curves, but it is plotted in a different manner. Note that the curves are plotted as curves of allowable scan angle versus nominal beamwidth. For each magnification, there are curves for several values of the f_1/D ratio. The nominal beamwidth is defined as 480 times the ratio of maximum path error to aperture diameter. The logic behind this definition is based on the assumption of a one-eighth wavelength allowable path error and an illumination that yields a beamwidth of $60\lambda/D$ degrees.* For example, if it is desired to radiate a 1-degree beam, the aperture should be about 60 wavelengths in diameter, and the maximum path error should be less than or equal to one-eighth wavelength or 1/480 of the aperture.

For some cases, particularly when it is desired to minimize side-lobe levels, it may be desirable to hold the path error to less than the one-eighth wave nominal tolerance. In other cases, it may be that the illumination pattern is such that the beamwidth is substantially different from $60\lambda/D$ degrees. These cases do not require that new curves be drawn. All that is necessary is that the nominal beamwidth used be related to the actual beamwidth by the equation.

$$B_n = \left(\frac{K_1}{K_2} \right) B_a \qquad (29)$$

where

B_a = actual beamwidth,

B_n = nominal beamwidth used with curves,

$K_1 = 8\ P_e/\lambda$,

$K_2 = B_a/(60\lambda/D)$,

P_e = path error allowed.

Suppose, for example, that it is desired to hold phase errors to a maximum of 30 degrees corresponding to a path error of $\lambda/12$. Thus,

$$K_1 = \frac{8}{12} = \frac{2}{3}$$

Further suppose that the illumination pattern used yields a beamwidth of $80\lambda/D$ degrees. Thus,

$$K_2 = \frac{80}{60} = \frac{4}{3}$$

We then have

$$B_n = \frac{1}{2} B_a$$

In other words, if the actual beamwidth desired is 1 degree, we look up the curve for a 0.5-degree beam. For example, in Figure 16 we see that with a magnification of 2 and an f_1/D ratio of 0.5, we can scan about ± 4.4 degrees. If the f_1/D ratio is increased to 0.667, the allowable scan range is increased to ± 5.9 degrees. If the required scan were, say ± 5 degrees, we could conclude that the ratio f_1/D should be greater than 0.5 and need not be as large as 0.667. A value of 0.55 would probably be adequate.

2. Choice of Magnification. While the previously outlined procedure can be used to determine the proper f_1/D ratio, once the magnification is fixed, determining the optimum magnification is a somewhat more complex process. It is unclear as to just what feature of the overall design should be optimized. To illustrate the problem, one example has been worked out in some detail. Suppose that the antenna is required to scan through an angle of ± 5 degrees, and that the nominal beamwidth should be 0.25 degree. Consider first, the case of a magnification of 1/2. From the design curves, we pick off points of beamwidth versus f_1/D that permit us to draw a curve of beamwidth versus f_1/D ratio for a 5-degree scan angle. Interpolating graphically, we estimate that $f_1/D = 1.46$ is required for a 0.25-degree beamwidth. We treat the other magnifications in a similar manner, thereby obtaining a curve of magnification versus f_1/D ratio for a beamwidth of 0.25 degree and a scan angle of 5 degrees.

If we were attempting to minimize f_1, the resulting curve would indicate that a magnification of 2.5 or so would be an optimum; however, there is no reason a priori to assume that f_1 is the quantity that should be minimized to obtain the most desirable design. Table I gives several other pertinent dimensions normalized in terms of unit aperture diameter.

* This beamwidth is achieved with about 10 db of cosine taper in a rectangular aperture.

Table I

Dimensions of Schwarzschild Antenna

Magnification	Primary Focal Length	Effective Focal Length	Reflector Separation	Secondary Diameter
1/2	1.46	0.73	0.487	0.905
1	0.94	0.94	0.47	0.685
2	0.73	1.46	0.487	0.455
4	0.86	3.44	0.688	0.372

Nominal beamwidth 0.25 degree, required scan angle ±5 degrees, aperture diameter 1 unit.

The diameter of the secondary reflector is the diameter required to intercept all of the rays at the maximum off-axis angle. At the higher magnifications, only a small portion of this diameter is being used at any one time.

In some cases, the choice of magnification can be determined by the fact that a certain effective focal length is desired to match a particular feed structure. In other cases, it may be desired to minimize the diameter of the secondary reflector and thereby minimize the aperture block. Were it not for the requirement of covering a finite sector of scan, the minimum aperture block would correspond to a very high magnification degenerating in the limit to a simple parabolic primary and a point secondary. Figure 18 shows how the diameter of the actual secondary required varies with magnification; there is a minimum in the curve at a magnification of about 3.5. This then would be the optimum magnification for this example if aperture block were the main consideration.

One more feature of magnification will be noted; in all cases except the case of magnification of 4, the focal region is substantially smaller than the secondary reflector. Thus, in these cases, interference to beam formation due to the presence of the feed structure can be expected to be minor in comparison with that due to the aperture blocking. In the case of a magnification of 4, however, the situation is reversed and the focal region becomes larger than the secondary reflector. This is another factor that tends to limit the maximum magnification used in a given application.

3. Effects of Overcompensation. The curves of Figures 15 through 17 give design data for the true Schwarzschild antenna that satisfies the Abbe sine condition. Figures 19 through 22 give comparable data for certain overcompensated systems. Note that these curves are characterized by a hump at small angles of scan--that is, at small angles off axis, these antennas will not support as narrow a beamwidth as the true Schwarzschild antenna. This is the region where the overcompensation produces negative coma. Note, however, that at larger angles off axis, after we are over the hump, the overcompensated antennas will support smaller beamwidths. For example, Figure 16 shows that at 6 degrees off axis, a Schwarzschild antenna with a magnification of 2 and an f_1/D ratio of 0.667 will support a nominal beamwidth of 0.52 degree, whereas Figure 19 shows that a 50-percent overcompensated system will support a beamwidth of about 0.4 degree, an improvement of about 23 percent. On the other hand, Figure 20 shows that 100-percent overcompensation is too much for this case, since the same f_1/D will only support a nominal beamwidth of 0.84 degree or more at 6 degrees off axis; however, if it were required that we cover, say ±9 degrees of scan, 100 percent overcompensation yields better results than are obtained with either the true Schwarzschild antenna or the 50-percent overcompensated system. In general, the benefits to be derived by overcompensating are of a somewhat secondary but not insignificant nature. It is reasonable to base a design procedure on finding the first rough approximation to the desired parameters without taking account of the possible benefits that might accrue from the use of overcompensation. After the preliminary choice of parameters is made, we can then investigate the possibilities of further improvement of performance through the use of overcompensation.

4. Effects of Astigmatism. Figures 15 through 17 and 19 through 22 do not reflect the effects of astigmatism. Therefore, the question arises as to what extent, if any, a design based on these illustrations is deteriorated by astigmatic effects. If a fan beam is being radiated, the focus may be chosen as optimum in the narrow direction of the beam, and the astigmatism is unlikely to affect the other direction very much in the usual case--that is, astigmatism can be largely ignored for fan beams.

When a pencil beam is desired, it is found that astigmatism is a minor but not negligible effect. Except when coma is nulled (by the use of overcompensation), the dominant aberration appears to be the coma. In the cases checked, the maximum path error at the optimum focus was somewhat less than the meridional path error from the meridional focus, except in the vicinity of a coma null. Thus, we believe that if the preliminary choice of parameters is based on Figures 15 through 17, the resulting design will be satisfactory despite the effects of astigmatism.

5. Design Procedure Summary. Ideally, it would be desirable to base a design procedure on a set of curves that show a direct relation between the design parameters and the pertinent performance characteristics, such as gain and side-lobe level. This has not proved to be feasible in the present case. For a given configuration, the relation between maximum path error and f_1/D ratio can be shown by a single curve, and is independent of the operating frequency. The relation between side-lobe level and f_1/D ratio, however, involves another variable and requires a whole family of curves, one for each beamwidth. The relation between path error and side-lobe level cannot be simply expressed, since it depends on the shape of the path-error curve.

Additional factors that prevent us from going directly from performance characteristics to final design are the variables introduced by the possibility of overcompensation and the need to consider astigmatism in some cases. Thus, the procedure used must involve a certain amount of "cut and try operation," if precise control of the performance characteristics is desired. A tentative path-error tolerance must be selected; then use the curves of Figures 15 through 17 to arrive at an initial design. This design performance can then be computed in detail, including the effects of astigmatism and overcompensation, if applicable. After the resulting radiation patterns have been computed, the performance can be evaluated and the path-error specification revised as necessary. This process is then repeated until a satisfactory design evolves.

Despite these limitations, we believe that the procedure is useful. We believe that the first cut based on Figures 15 through 17 will be close enough to the final design to decide whether the overall configuration will be satisfactory, or whether a different type of antenna should be investigated.

Attempts to express the data of Figures 15 through 17 in a single analytic formula have had only limited success. The formula,

$$B = \left(\frac{\Theta D}{10 f_1}\right)^2$$

where

B = nominal beamwidth in degrees and

Θ = allowable scan angle in degrees,

results in curves that are roughly similar but substantially different in detail. It is not known to what extent this is due to the inherent complexity of the problem, and to what extent it is due to the computation inaccuracies.

III. Side-Lobe Reduction

In all two-reflector systems, including the Cassegrain and Schwarzschild systems, the goemetry of the system is such that the secondary reflector presents an obstacle to incoming or outgoing energy. The effect of such an obstacle is to alter the amplitude illumination across the antenna aperture, resulting in a radiation pattern that has higher side lobes and a lower gain than that of the aperture with the obstacle removed. Both of these effects, a higher side-lobe level and a lower gain, are undesirable and tend to put severe limitations on the use of two-reflector systems.

For some applications, it is possible to reduce the size of the secondary reflector to the point where the effects of aperture blocking are negligible. In other applications, however, notably wide-angle scanning, a large secondary reflector is usually necessary to obtain adequate performance. Since it is impossible to physically eliminate the blocking obstacle, a method must be found to negate its blocking effect without compromising its performance as a secondary reflector. If the aperture distribution of the unblocked aperture can be duplicated with the blocked aperture, then the radiation patterns of both apertures will be identical. From an economical standpoint, it is also desirable to utilize the secondary reflector in any method to reduce the side-lobe level. With this in mind, four methods have been considered to reduce the increased side-lobe levels that are due to aperture blocking. All of the methods use some modification of the secondary reflector. The four methods are:

1. Serrated secondary reflector,

2. Hyperboloidal lens with grid-type secondary reflector,

3. Bootlace secondary reflector,

4. Semi-bootlace secondary reflector.

All of the above methods were moderately successful in reducing the side lobes due to the aperture block of the secondary reflector. Rather than going into the details of each technique, we will briefly discuss the simplest technique--the semi-bootlace technique. In this technique, a small element is placed on the front of the secondary reflector and fed by means of coaxial line through an amplifier and a phase shifter to a second element feeding the rear of the secondary reflector as a secondary aperture. By varying the phase shifter and the gain of the amplifier, the illumination across the total aperture is adjusted to fill in the null due to the blocking of the secondary reflector.

Typical results using this technique are shown in Figures 23 and 24. Figure 23 shows the measured radiation pattern of a Cassegrain system having a secondary reflector 0.21 the diameter of the primary reflector, and Figure 24 shows the measured radiation pattern of the same system using the secondary reflector as a secondary source. From the figures it can be seen that an improvement of about 8 db in side-lobe level is obtained. It is interesting to note that the second side lobe rather than the first side lobe was the limiting factor on the lowest obtainable side-lobe level. It is thought that this is the result of forward spillover from the feed.

IV. Conclusions

In general, the Schwarzschild antenna fits best in applications where the required scan angle as measured in degrees is relatively small, though it may be considerable when measured in beamwidths. An example would be a tracking radar having a fractional degree beamwidth where a limited rapid scan action is desired to provide ease of acquisition, and to relieve the response time requirements on the drive system. Other applications, such as pincushion trackers, artillery spotting radars, damage-assessment radars, etc., readily present themselves. In the nonradar field, there are satellite and space communications problems in which the antenna requirement is similar to that of a radar tracker, and in which a limited rapid scan action would be

desirable. Another case in which Schwarzschild antennas of this nature might prove useful is in scatter-communications links, where the direction of arrival is somewhat uncertain because of the nature of the propagation medium. In a conventional system, the usable antenna gain is limited as a result of this uncertainty. With a Schwarzschild antenna, it would be possible to use a beamwidth narrower than the uncertainty, and to use a diversity system based on direction of arrival.

In applications where more substantial coverage sectors are required, it may still be possible to use the Schwarzschild principle, provided that the aperture-blocking problem can be solved either through the use of the Wheeler twist principle, or through the use of the folded pillbox configuration. Although there are other types of optical systems, such as the Luneberg lens and the Schmidt camera, which give wider fields of view, the Schwarzschild antenna may still be preferable because it does not require any refracting media with its consequent power limitations. Other systems, such as spherical reflector and torus types, can provide wide coverage without refracting media, but for small to moderate coverage sectors, the Schwarzschild antenna is expected to be more compact.

In applications such as general surveillance systems where relatively large coverage sectors are required, the Schwarzschild antenna is not competitive. Spherical and torus types have no coma regardless of the coverage angle. Although they must have large f_1/D ratios to keep the spherical aberration within bounds, the required f_1/D ratio is independent of the sector covered. The Schwarzschild antenna, though it can be quite compact for small angle coverage, requires larger and larger f_1/D ratios as the coverage angle is increased. The exact break point will depend on the detailed system requirements. For very narrow beamwidths, the range of application for the Schwarzschild antenna will be greater than when the beamwidth is larger. Under no circumstances, however, does it appear that the Schwarzschild antenna would be applicable to a large, such as 40 degrees, field-of-view requirement.

In the past, attempts to use the Schwarzschild design have been hampered by a lack of published information from which design parameters and performance characteristics could be readily obtained. Precise data for a particular application will still require a substantial amount of computation. The curves in this paper, however, provide a means for estimating the parameters required to meet a given set of performance specifications. If safety factors are included in the performance specifications, they can be used as the basis for a final design. If the most efficient design is desired, however, it appears that the design procedure should be an iterative process. The initial design is based on the curves in this paper and an assumed path-error specification. The performance of the initial design is then investigated in detail including the effects of overcompensation and astigmatism, if applicable. The initial path-error specification is then revised as necessary and the process repeated.

Attempts have been made to devise a simple algebraic formula relating the pertinent design considerations. To date, however, these attempts have not yielded a usable result. Apparently, the problem is sufficiently complex so that it defies representation by a simple formula.

Several methods of reducing side-lobes due to aperture blocking in two-reflector systems have been suggested and demonstrated, however, much remains to be done in improving these techniques and devising methods that will work as the beam is scanned.

V. References

1. L. K. DeSize, D. J. Owen, and G. E. Skahill, "Final Report--Investigation of Multibeam Antennas and Wide-Angle Optics," Airborne Instruments Laboratory, Report No. 7358-1, Astia No. AD237854, Contract AF 30(602)-1980, January 1960.
2. P. W. Hannan, "Microwave Antennas Derived from the Cassegrain Telescope," PGAP Transactions, Vol AP-9, No. 2, p 140-153, March 1961.
3. K. Schwarzschild, Abhandl. Ges. Wiss., Göttingen Math. Physik, Kl. 2, 1905.

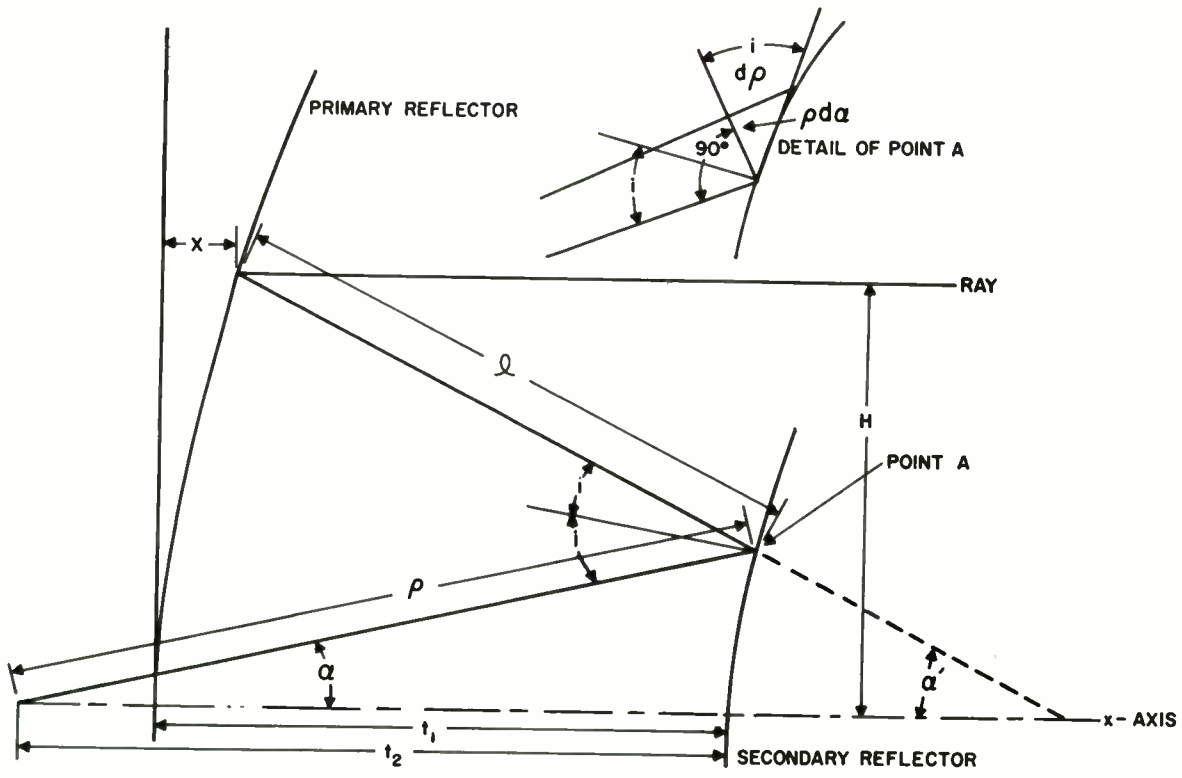


Fig. 1. Cross section of generalized Schwarzschild antenna.

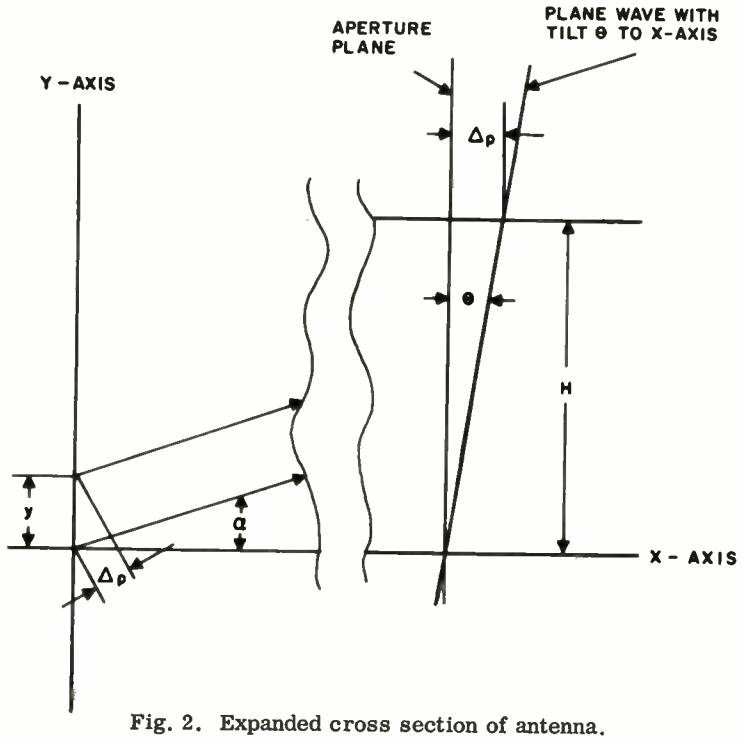


Fig. 2. Expanded cross section of antenna.

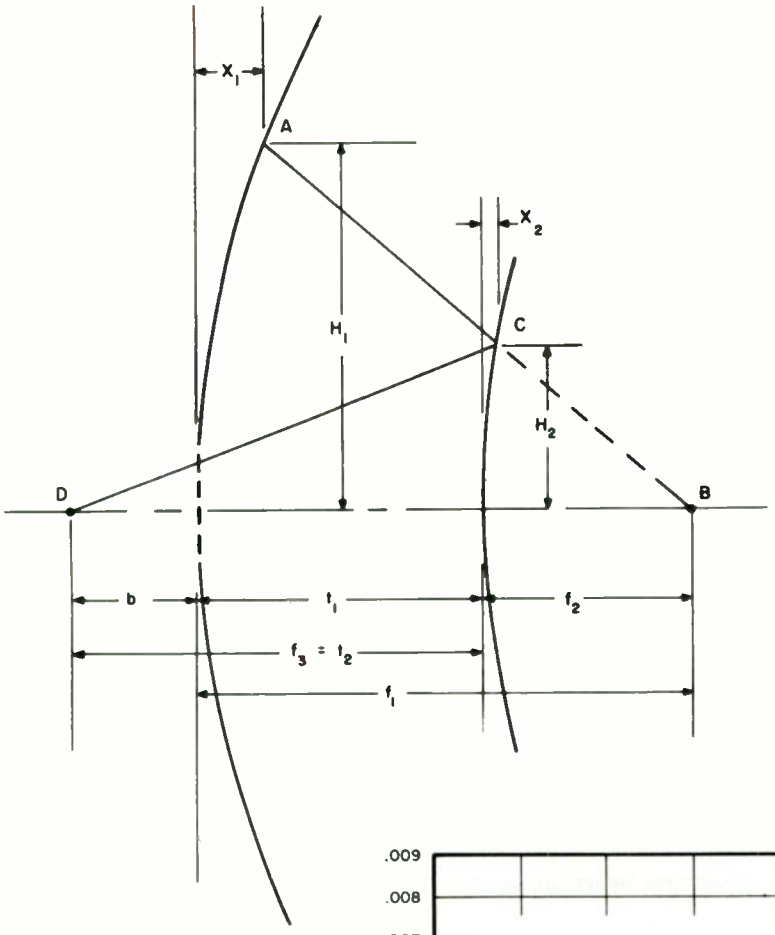


Fig. 3. Generalized Cassegrain antenna.

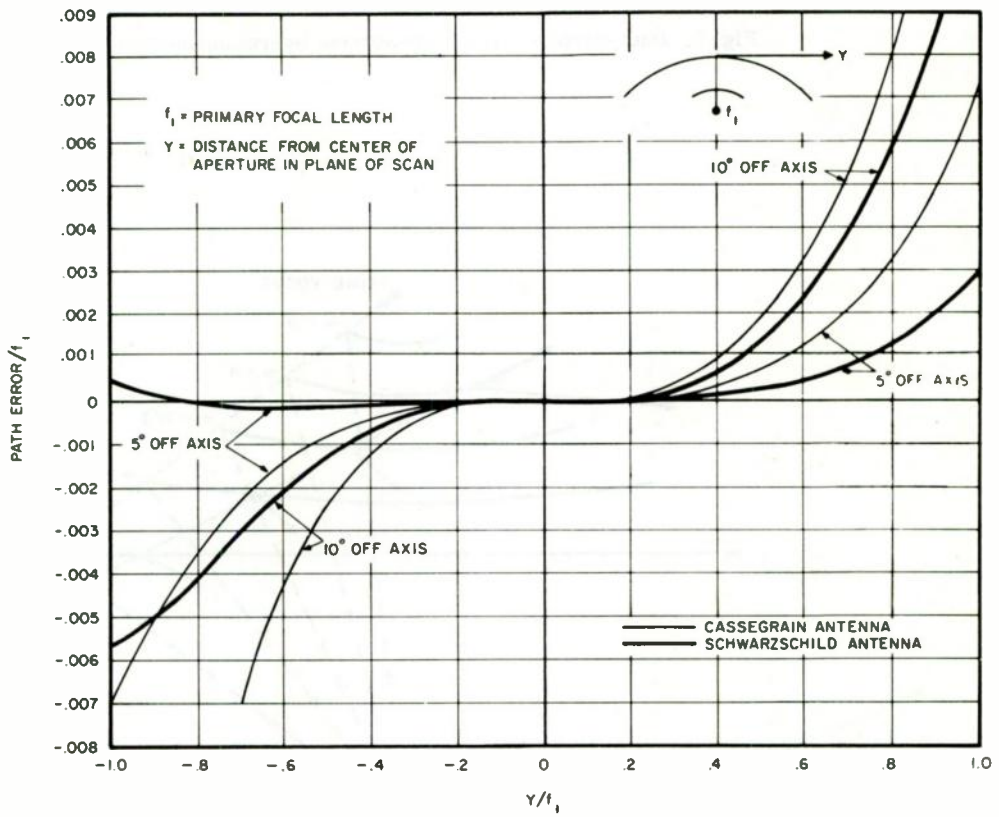


Fig. 4. Performance of Cassegrain antenna vs Schwarzchild antenna. (m = 2)

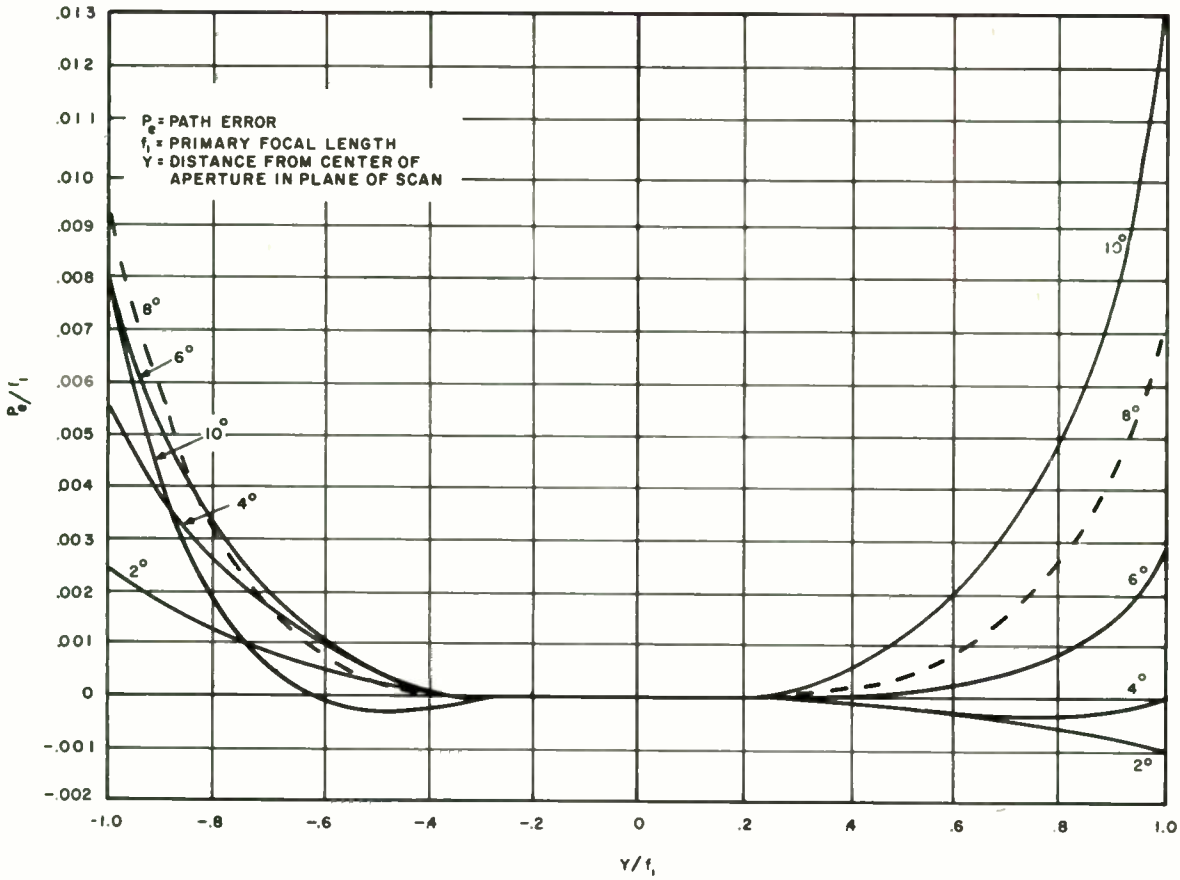


Fig. 5. Path-error curves, 100-percent overcompensated system (m = 2).

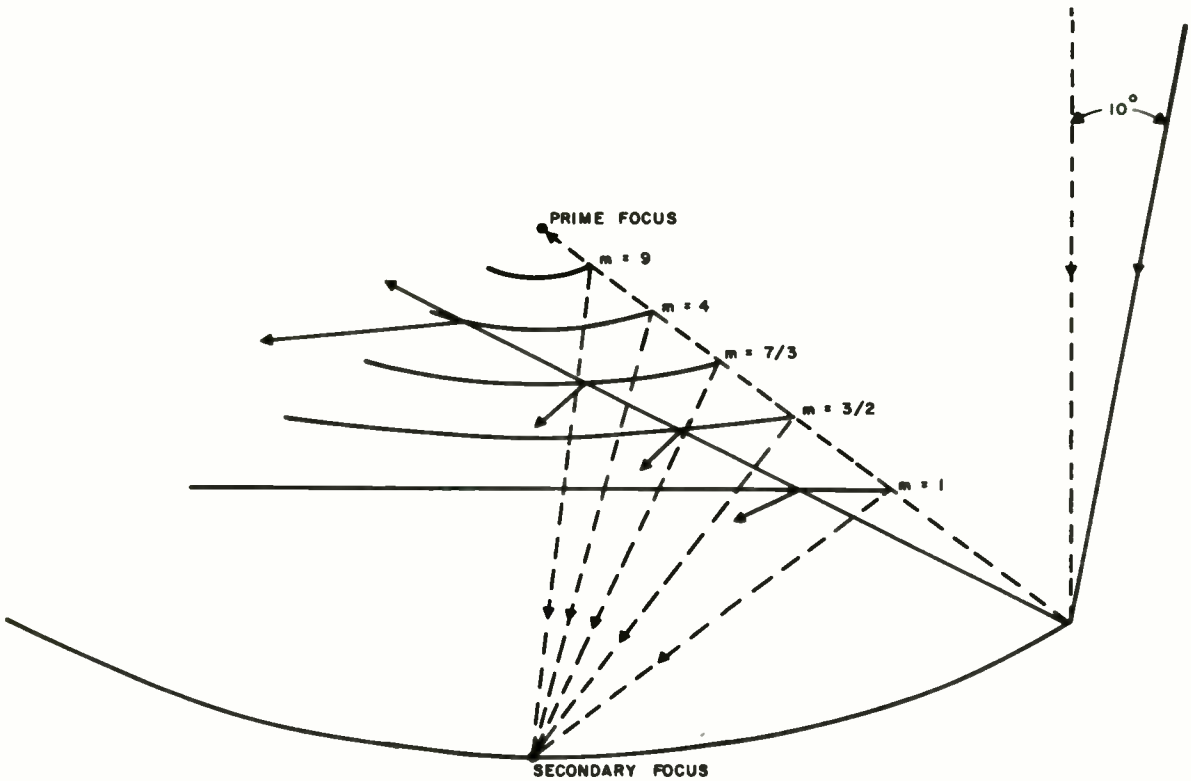


Fig. 6. Family of Cassegrain systems having a common parabolic primary reflector.

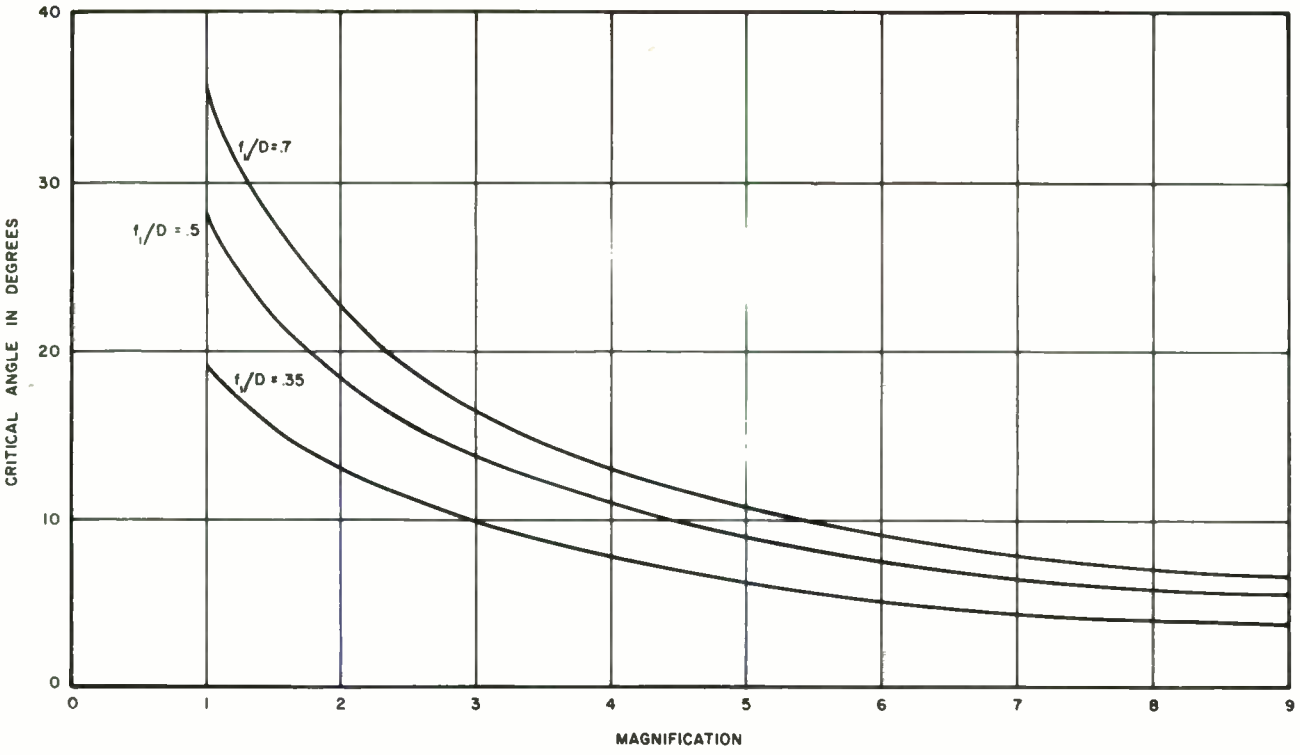


Fig. 7. Critical angle vs magnification as a function of f_1/D ratio of primary reflector.

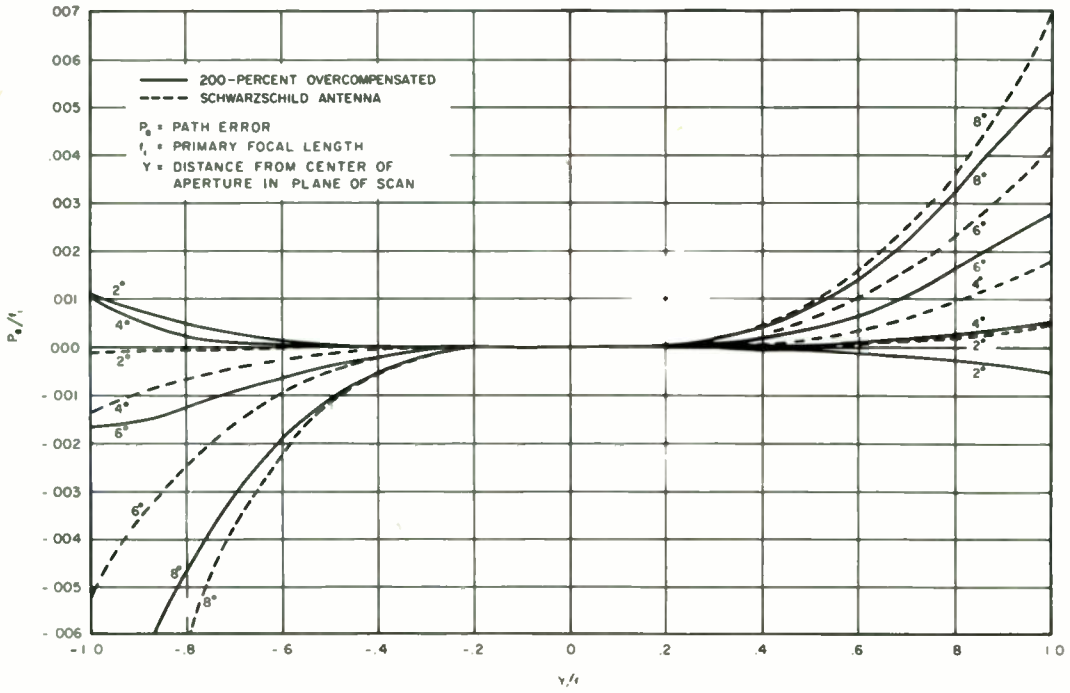


Fig. 8. Path-error curves, 200-percent overcompensated system (m = 4).

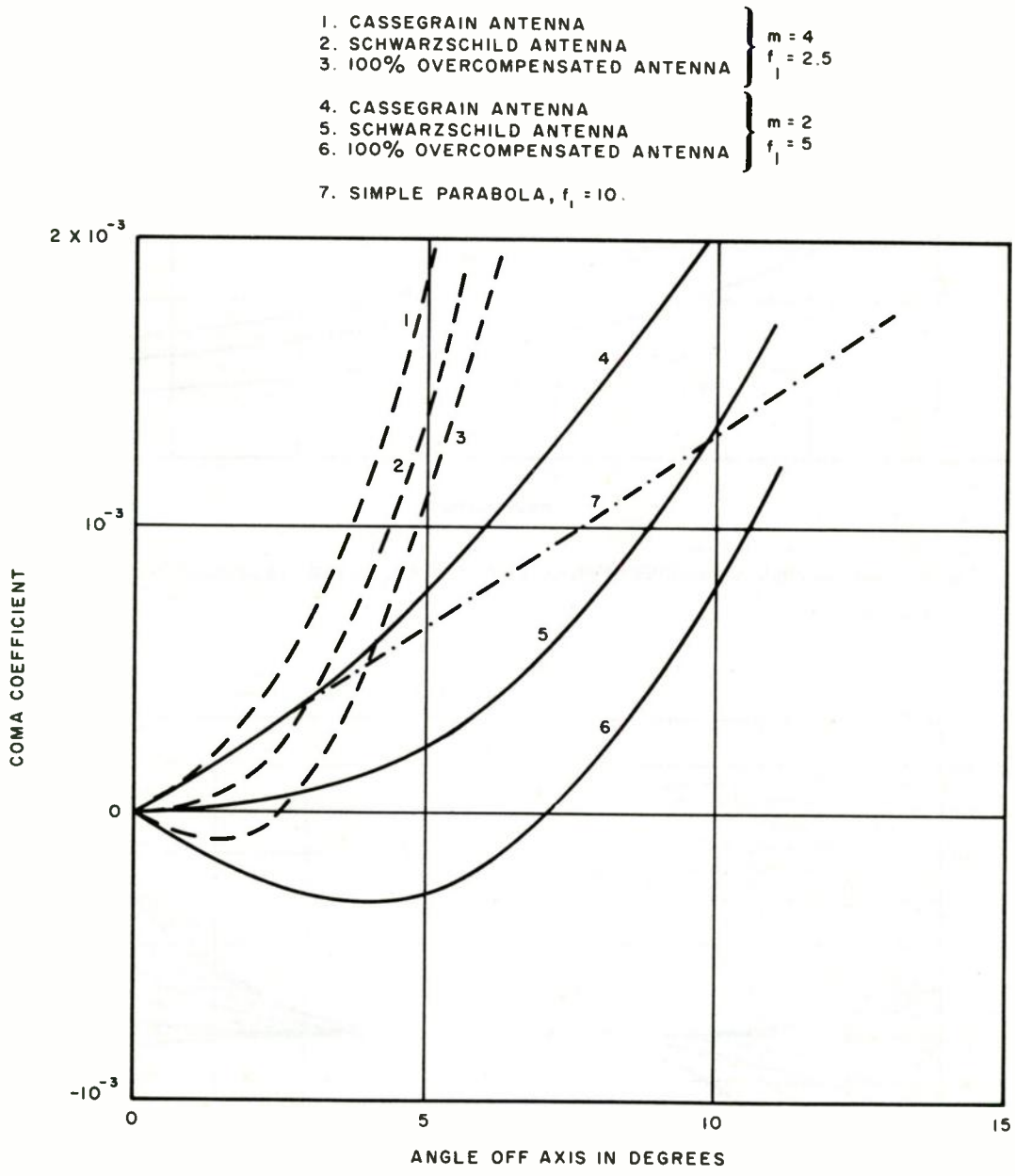


Fig. 9. Coma coefficient vs angle off axis (normalized for effective focal length of 10 units).

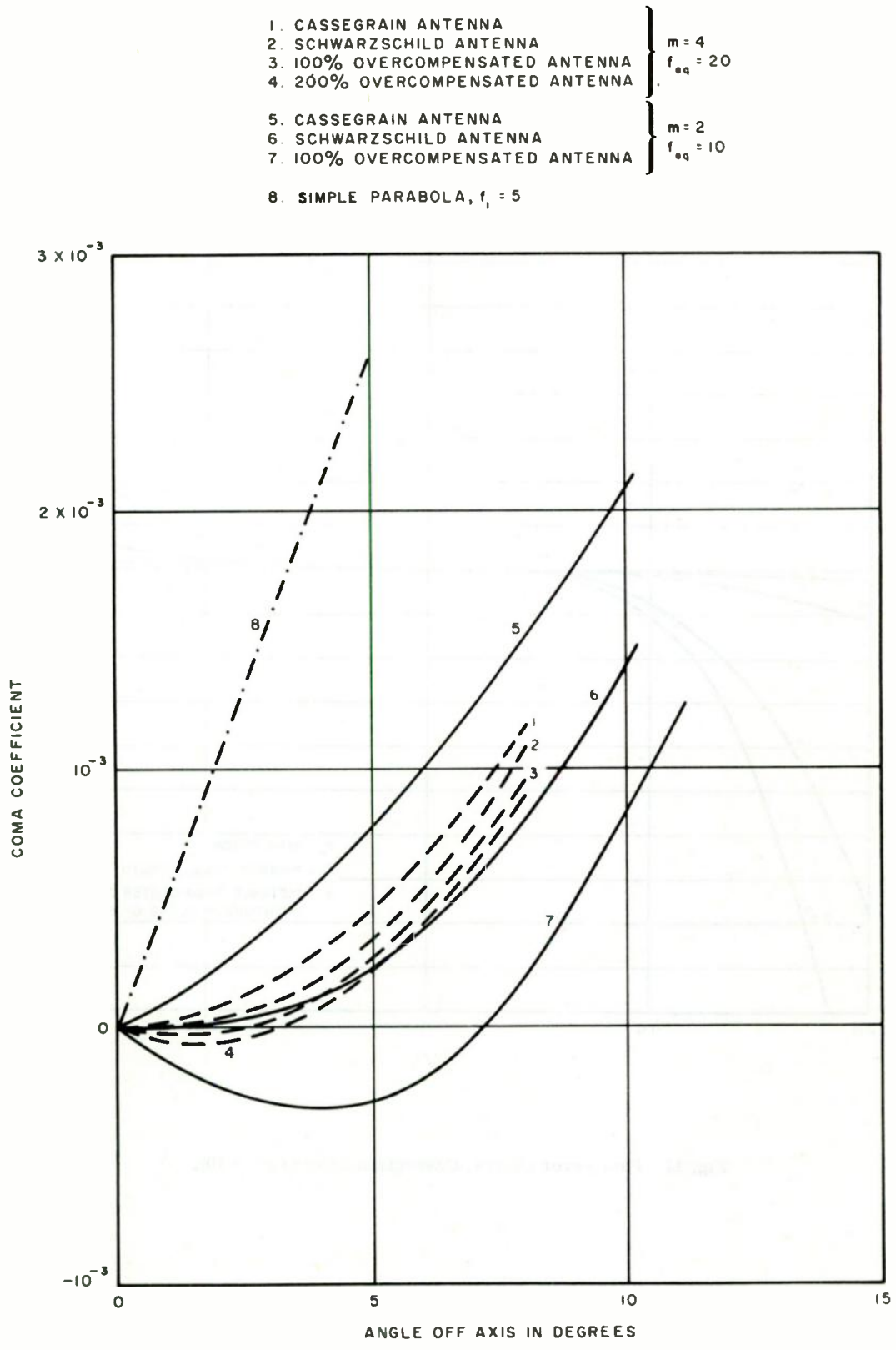


Fig. 10. Coma coefficient vs angle off axis (normalized for primary focal length of 5 units).

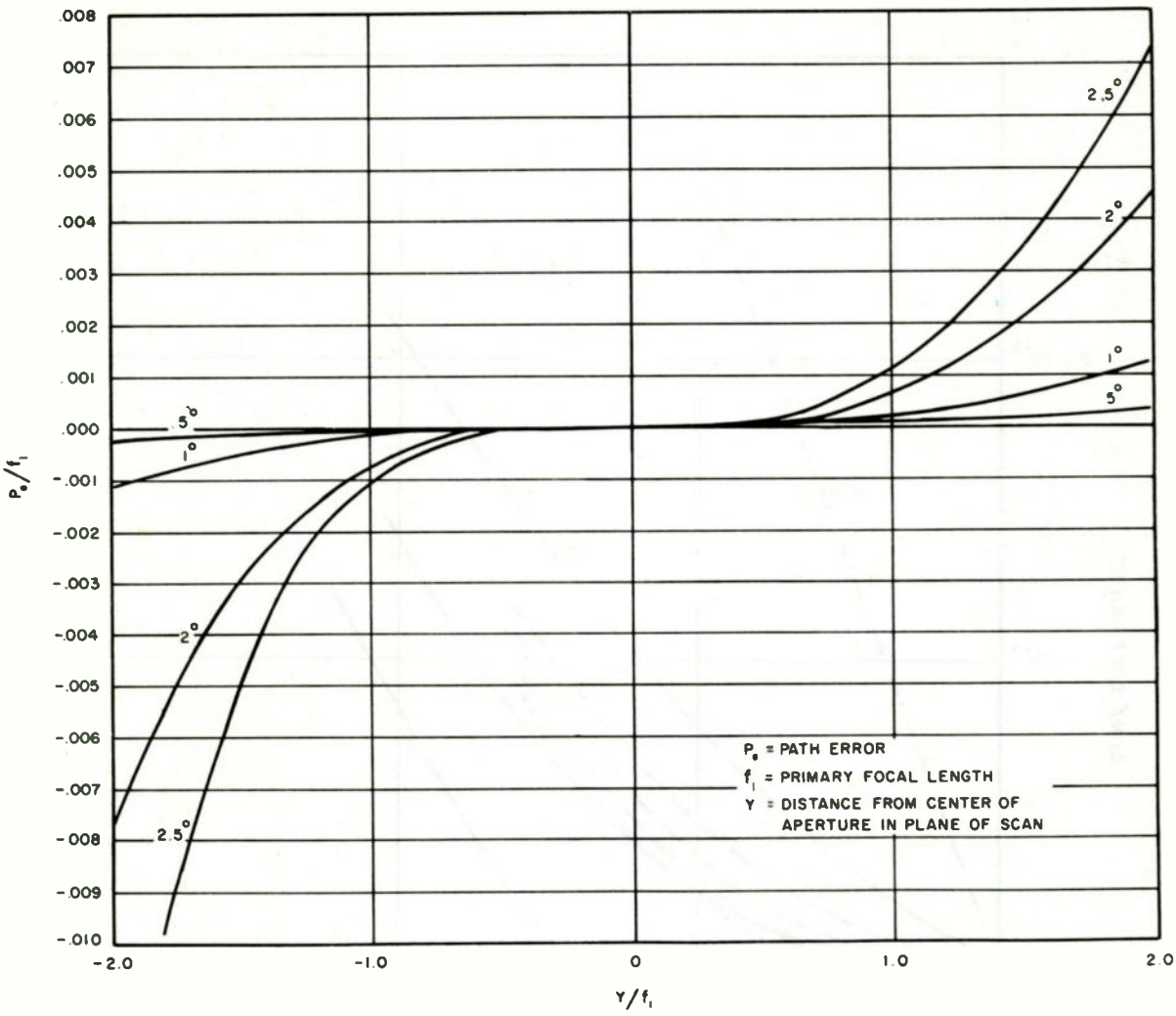


Fig. 11. Path-error curves, Cassegrain antenna (m = 10).

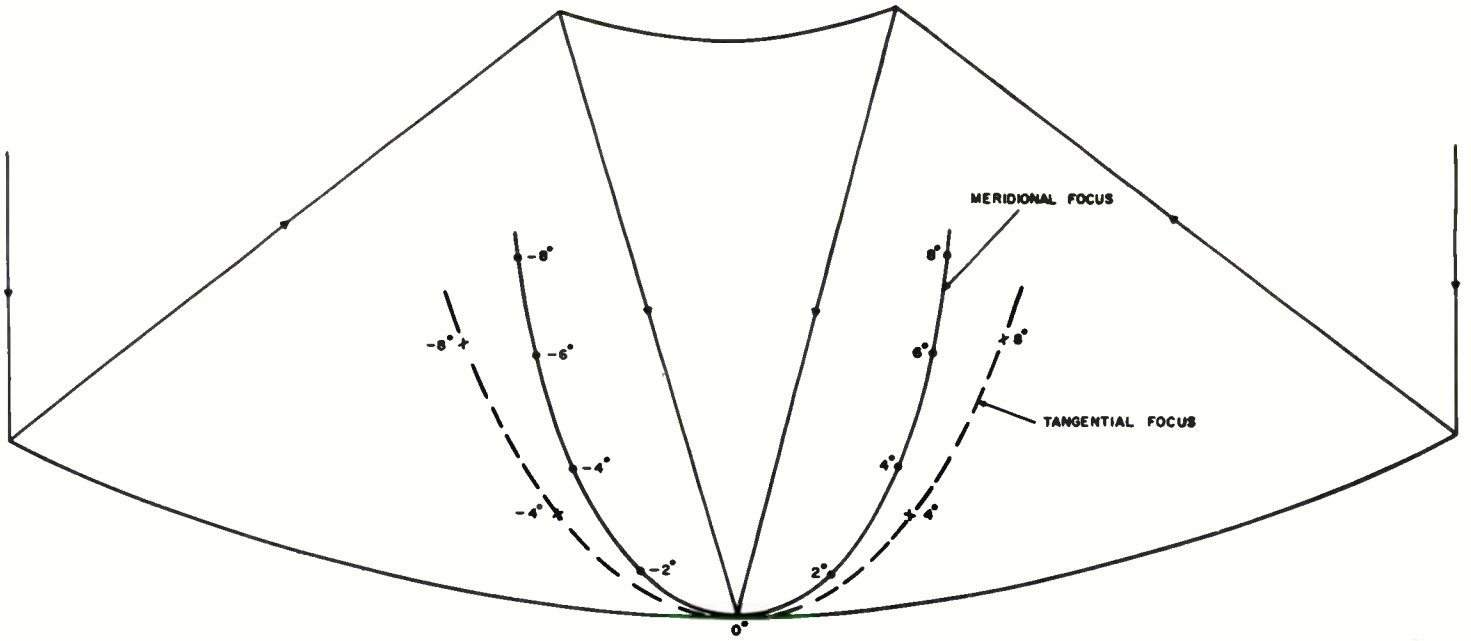


Fig. 12. Cross section showing focal points of 200-percent overcompensated system (m = 4).

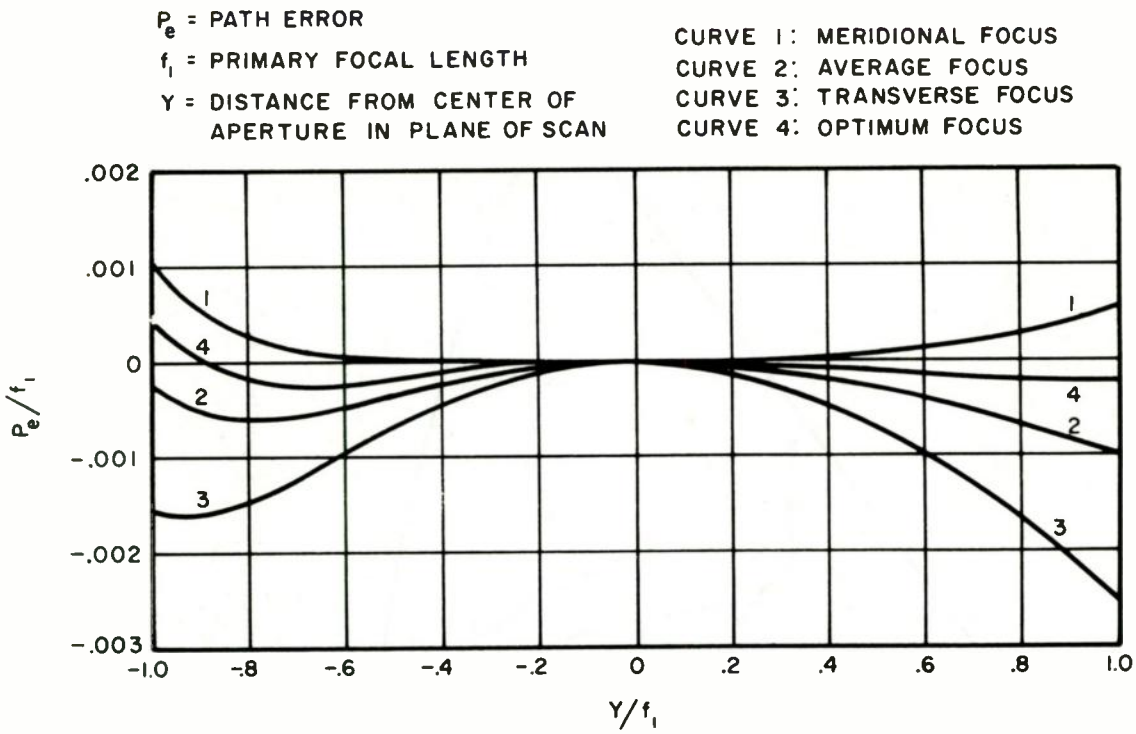


Fig. 13. Path-error curves, meridional ray fan, 200-percent overcompensated system (m = 4) 4 degrees off axis.

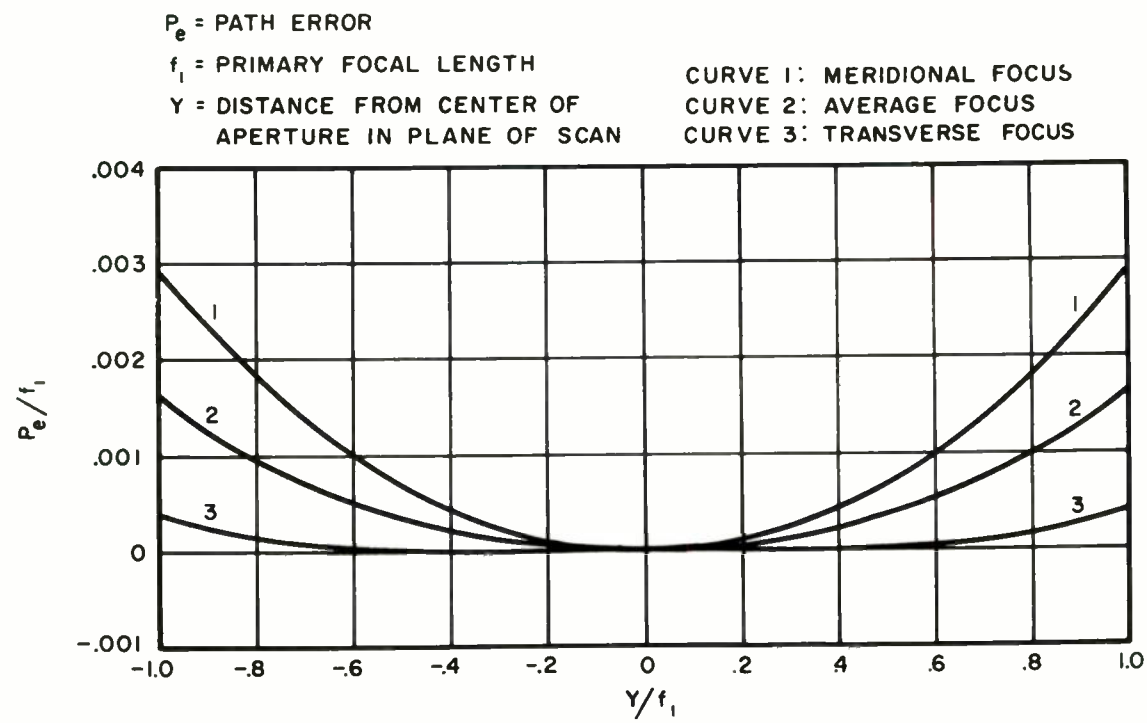


Fig. 14. Path-error curves, transverse ray fan, 200-percent overcompensated system (m = 4) 4 degrees off axis.

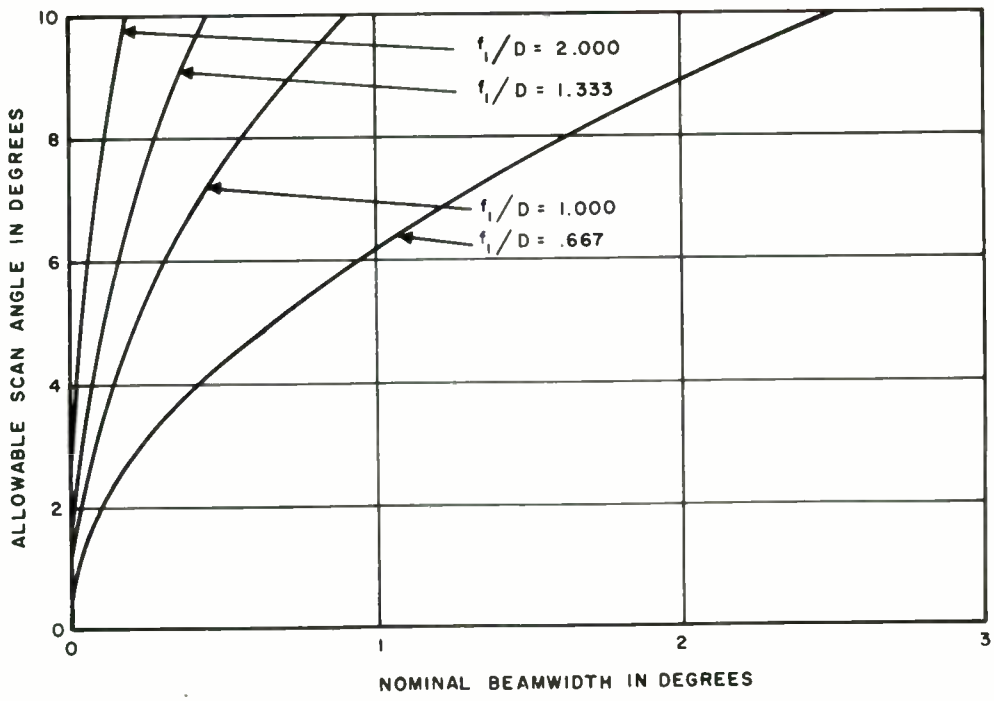


Fig. 15. Allowable scan angle vs beamwidth, case m = 1.

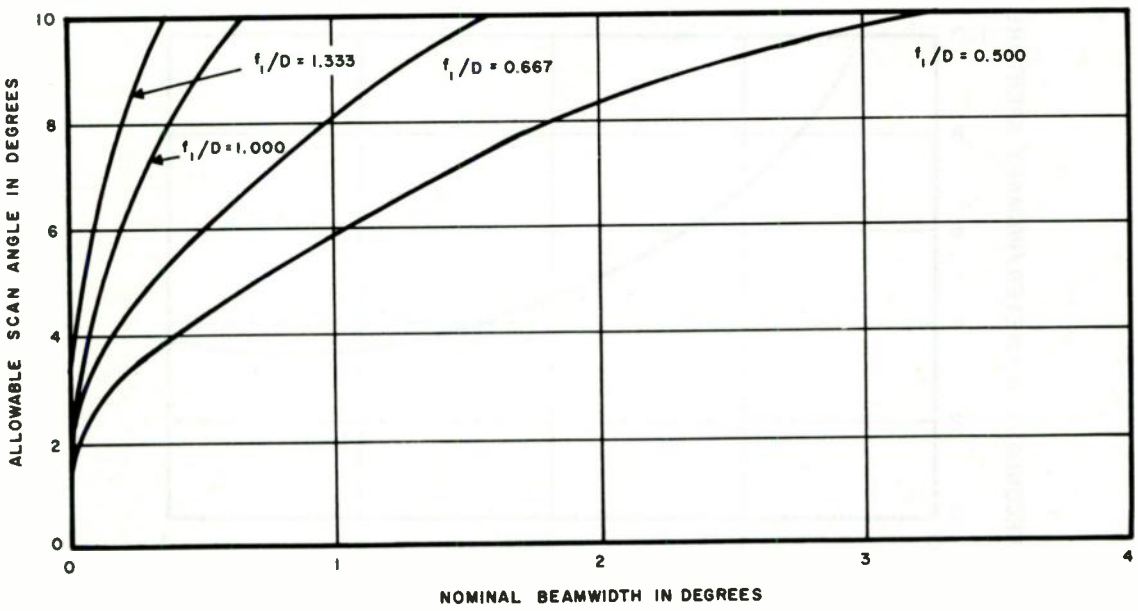


Fig. 16. Allowable scan angle vs beamwidth, case m = 2.

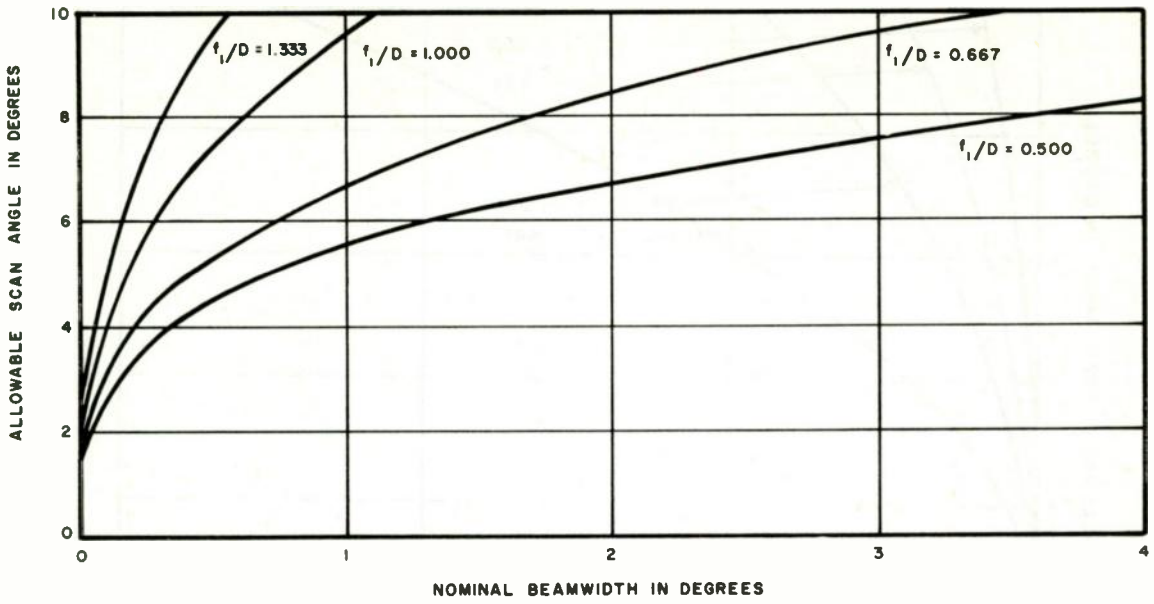


Fig. 17. Allowable scan angle vs beamwidth, case m = 4.

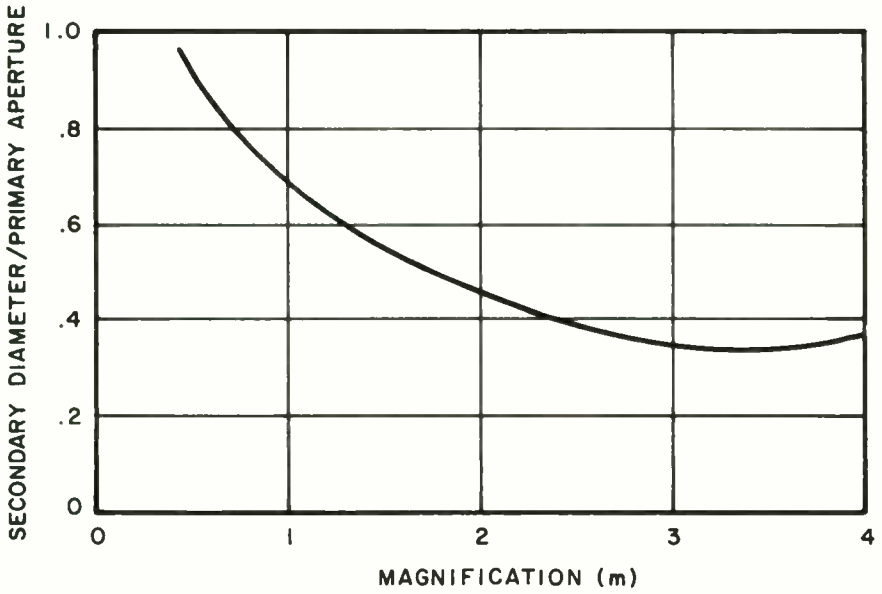


Fig. 18. d as a function of m.

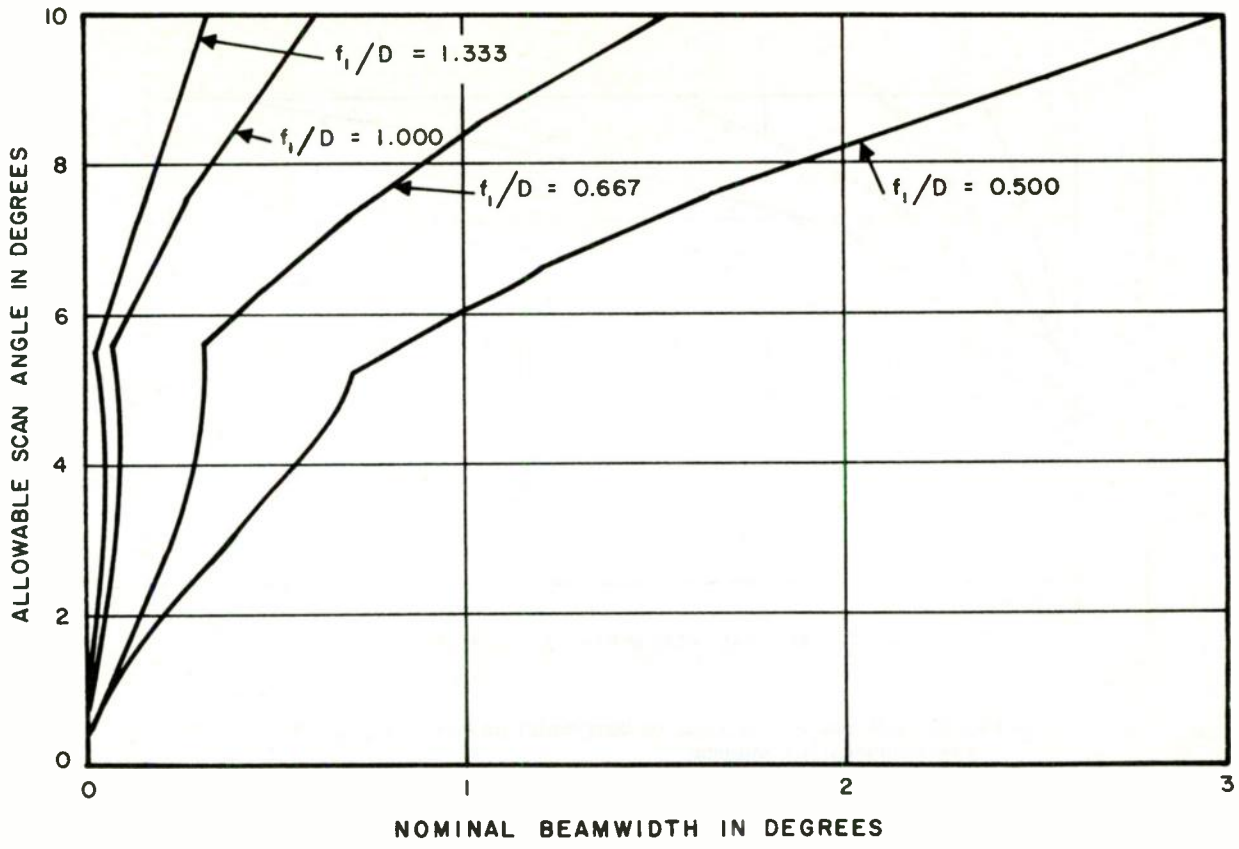


Fig. 19. Allowable scan angle vs beamwidth (m = 2), 50-percent overcompensated antenna.

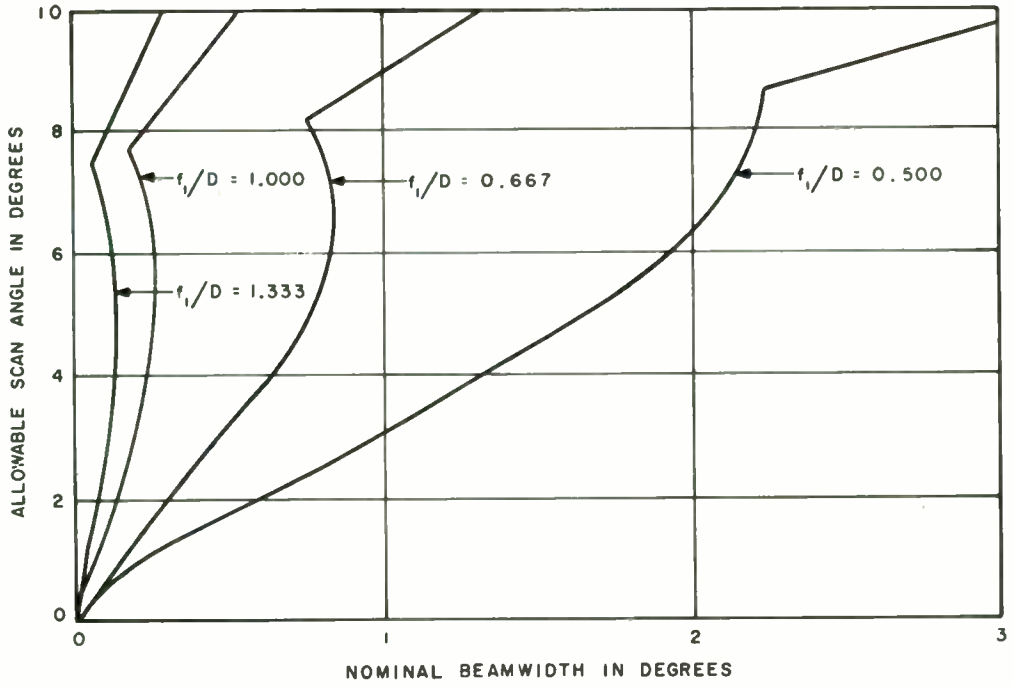


Fig. 20. Allowable scan angle vs beamwidth (m = 2), 100-percent overcompensated antenna.

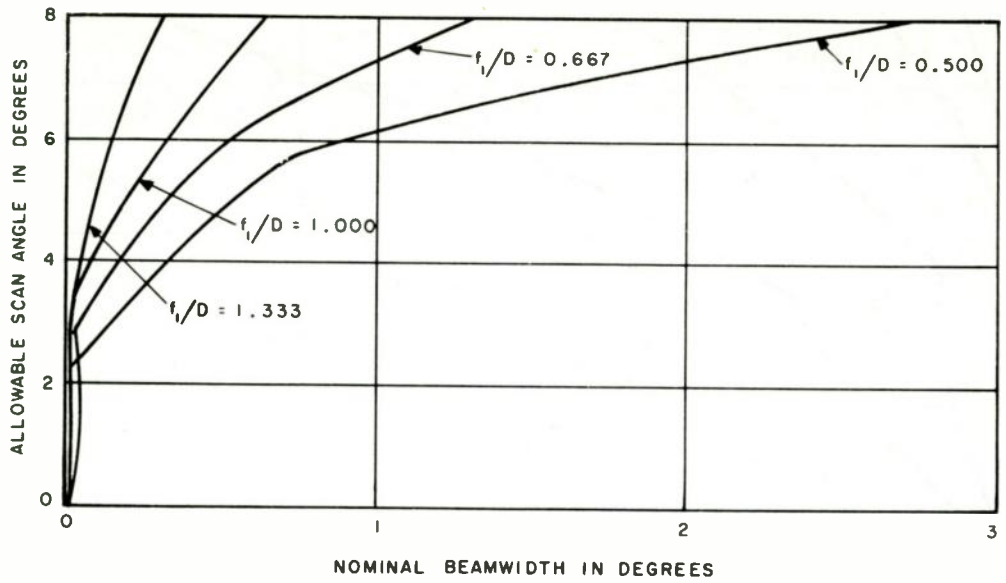


Fig. 21. Allowable scan angle vs beamwidth (m = 4), 100-percent overcompensated antenna.

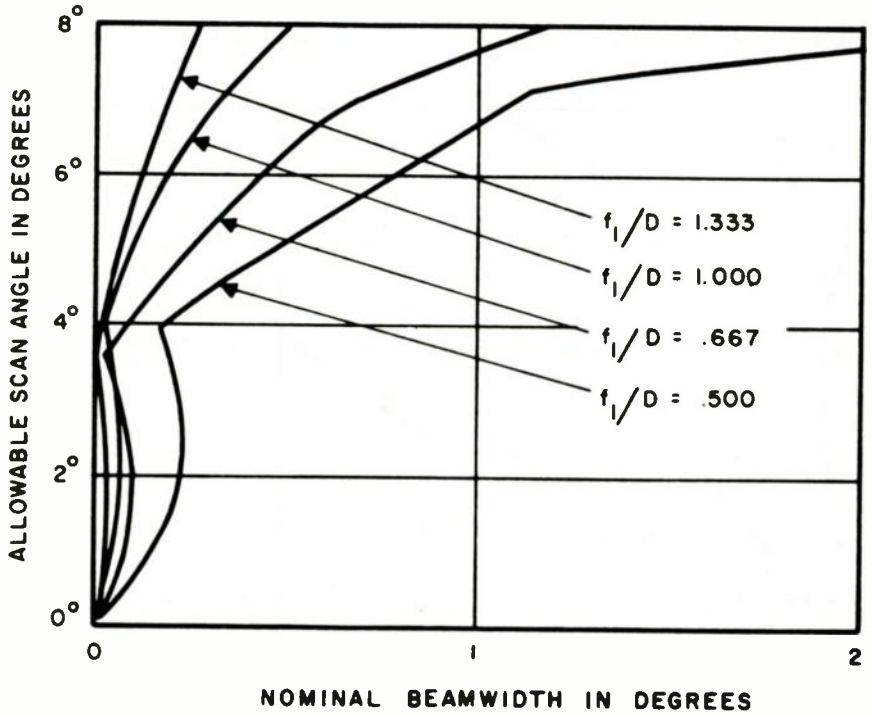


Fig. 22. Allowable scan angle vs beamwidth (m = 4), 200-percent overcompensated antenna.

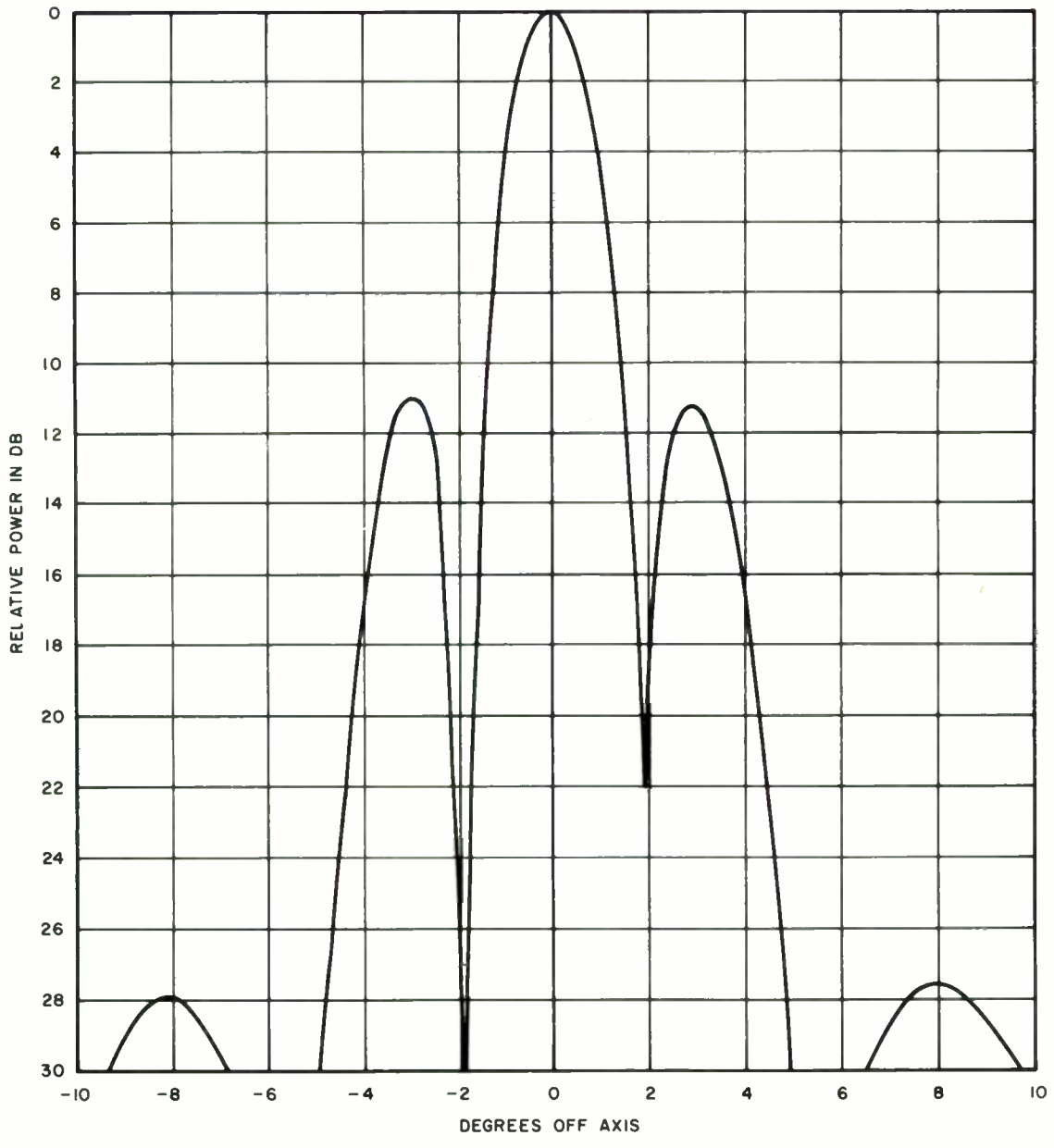


Fig. 23. Radiation pattern of normal Cassegrain at X-band.

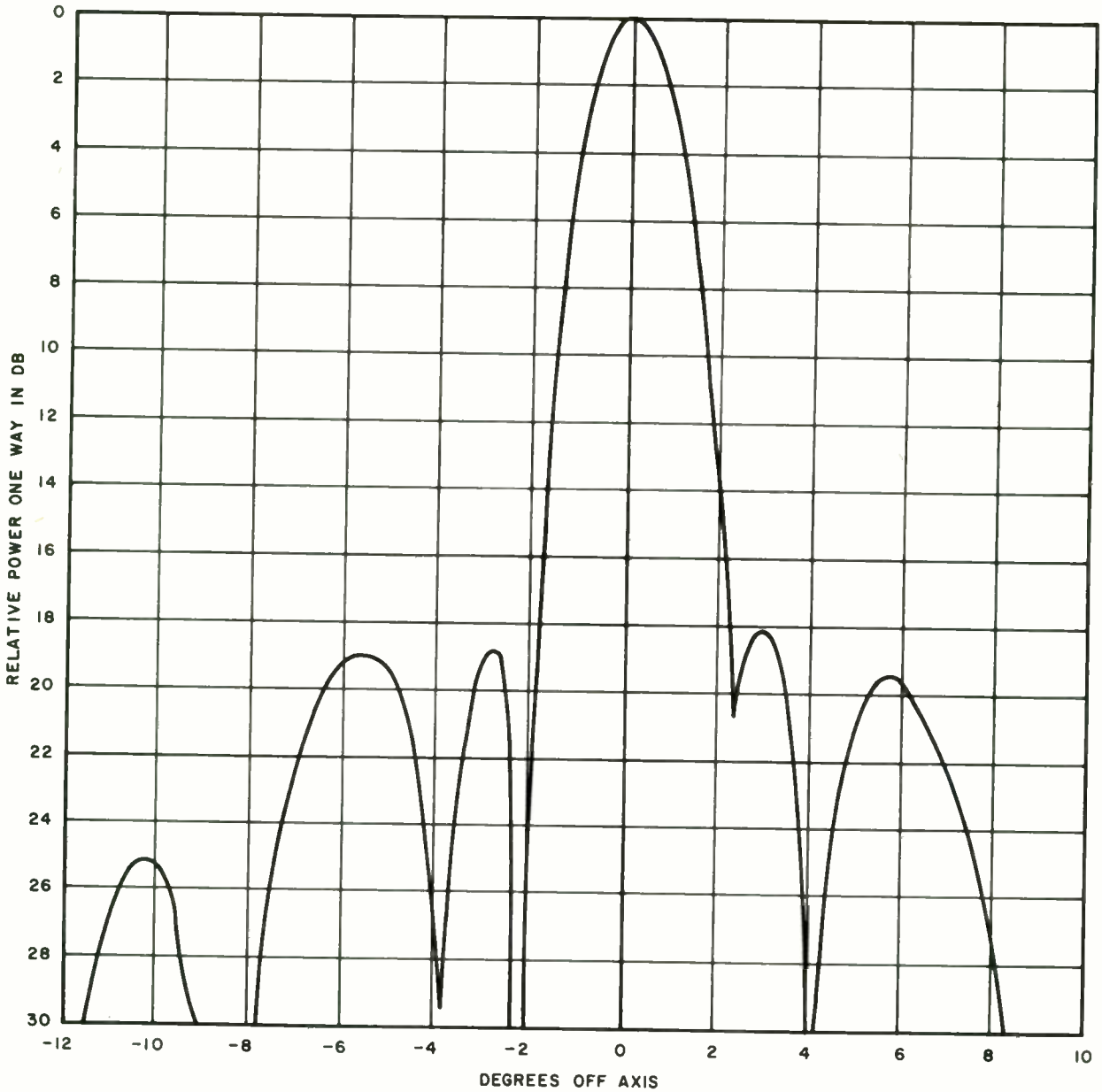


Fig. 24. Radiation pattern using semi-bootlace system at X-band.

ANNULAR SLOT MONOPULSE ANTENNAS

K. C. Kelly and F. J. Goebels, Jr.
Hughes Aircraft Company
Culver City, California

Summary

This paper describes the design and performance of annular slot planar arrays which generate pencil beams and monopulse tracking beams in both principal planes. These planar arrays consist of a transmission line in the form of a radial waveguide whose upper plate contains annular groups of linear slots. Monopulse operation is obtained by dividing the transmission line into four independently excited quadrants, each of which is propagating a field configuration identical to the dominant H_{01} radial waveguide mode. The antennas radiate linear polarization. In contrast with many other types of two-dimensional waveguide slot arrays, the antennas can be fabricated inexpensively. Small diameter versions are shown to possess the following electrical and mechanical advantages when compared to a dish antenna of comparable size: 1) better gain through increased aperture efficiency, 2) improved backlobe suppression, and 3) compact size and lower weight. The experimental design was done at X-band.

Introduction

This paper is the third in a series[1,2] which illustrates the utilization of annular slots in the design of planar array antennas for various functions. This interest in annular slots persists because the radial waveguide fed annular slot antenna has numerous fabrication advantages when compared to planar arrays formed by a multiplicity of slotted rectangular waveguides. To date however, these radial waveguide antennas fail to promise the aperture efficiencies that have been achieved at Hughes in a decade of studies, experiments, and production on slotted rectangular waveguide arrays. Circular apertures of any extent may be generated since the radius at which the radial waveguide is terminated is a variable under the control of the designer. Annular slots, or their approximation in the form of ringed groups of short linear slots, may be concentrically arrayed to fill one plate of the chosen extent of a radial waveguide. Radiation pattern synthesis with concentric arrays of annular slots has been treated extensively by Bickmore,[3] and solutions for the modes and impedances of a radial waveguide are widely available.[4] Thus, attainment of particular antenna performance characteristics is a matter of determining, (1) the waveguide mode to be employed, (2) the arrangement of the short slots which build up each annulus, and (3) the slot conductances.

The studies reported here were directed towards the development of broadside pencil beam antennas which could also provide difference patterns in two orthogonal planes. These antenna pattern characteristics form the basis for monopulse tracking systems. The generation of a sum beam and two difference beams is also necessary for a variety of synthetic conical lobing systems for tracking radars. Linearly polarized antennas are treated in this paper.

The following sections describe the simplest radial waveguide modes which lend to these objectives. Also described are the feeding techniques used, the impedance relations which apply for this non-uniform transmission line, and the theoretical predictions and experimental results obtained in the program. One feature of the results is the use of non-resonant annular slots to circumvent the usual radial waveguide array problem of a restriction to full waveguide wavelength radial spacing between the annular slots of a broadside antenna. Improved aperture efficiencies were realized with the half waveguide wavelength radial spacings that were made possible by the use of alternations in the phase angles of the non-resonant annuli.

The aperture efficiency obtained, combined with the reduced antenna weight, moment, and backlobe radiation make these antenna attractive for replacement of small dish antennas with front mounted monopulse feeds.

Theoretical Considerations

The Waveguide

A radial waveguide is a non-uniform transmission line which may be described in terms of a ρ, ψ, z polar coordinate system. Two pertinent examples of radial waveguides are the full cylindrical and 90° cylindrical sector regions shown in Figure 1. The aperture to be generated must contain four independent sectors to obtain the three sets of aperture phasing which produce the sum, azimuth difference, and elevation difference beams. Therefore, attention is immediately restricted to the 90° sectoral radial waveguides. For ease of launching and for preservation of the mode, it is best to employ the dominant mode if suitable. On examination it is found that the dominant E mode cannot be fed at $\rho < \frac{\lambda}{\pi}$. This fact

limits freedom of placement of the innermost annulus of a concentric array of annular slots. This restriction does not exist for the dominant H mode in a 90° radial waveguide sector. Thus, the dominant H mode was adopted in this study.

The fields of the dominant H mode in a 90° radial waveguide sector are given by[4]

$$\begin{aligned} \bar{E}_\psi &= \frac{2V}{\pi\rho} \sin \frac{\pi}{b} z \\ \bar{H}_z &= \frac{2I}{b} \sin \frac{\pi}{b} z \\ \bar{H}_\rho &= -j\eta \frac{V}{\pi\rho} \frac{\lambda}{b} \cos \frac{\pi}{b} z \end{aligned} \qquad (1)$$

where

V is the rms mode voltage,
I is the rms mode current,
b is the plate separation in the z direction,
λ is the free space wavelength,
η is the intrinsic admittance of the medium.

The time dependence exp (jωt) is suppressed here and throughout this discussion. Solution of the transmission line equations for this mode shows I and V to be the Hankel functions

$$\begin{aligned} I &= H_o^{(2)}(K_{10}\rho) \\ V &= -j \frac{1}{Y} H_1^{(2)}(K_{10}\rho) \end{aligned} \qquad (2)$$

in an infinite waveguide. The fields in an infinite guide are then,

$$\begin{aligned} \bar{E}_\psi &= -j \frac{2k}{\eta b K_{10}} H_1^{(2)}(K_{10}\rho) \sin \left(\frac{\pi}{b} z\right) \\ \bar{H}_z &= \frac{2}{b} H_o^{(2)}(K_{10}\rho) \sin \left(\frac{\pi}{b} z\right) \\ \bar{H}_\rho &= -\frac{2\pi}{b^2 K_{10}} H_1^{(2)}(K_{10}\rho) \cos \left(\frac{\pi}{b} z\right) \end{aligned} \qquad (3)$$

with Y, the characteristic admittance of the mode, replaced by its value

$$Y = \eta \frac{b}{\pi\rho} \frac{K_{10}}{k} \qquad (4)$$

where

$$K_{10} = \sqrt{k^2 - \left(\frac{\pi}{b}\right)^2}$$

is the H_{01} mode propagation constant and

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

is the free space propagation constant. The important features of Equation (3) are: (1) the only currents in the top and bottom plates of the radial waveguide are purely circumferential since these currents are derived from $\hat{e}_z \times \bar{H}_\rho$, and (2) these ψ directed currents are independent of ψ.

The Annular Slot and Its Radiation Pattern

Figure 2 illustrates the appearance of four slotted sectoral waveguides combined to form a circular aperture with a single quasi-annular slot. Consider the case for which the quadrants one and four are 180° out of phase with quadrants two and three to cause the currents at radius ρ to have the instantaneous directions shown by the arrows near the periphery. Further, note that all slots are excited with the same amplitude since the amplitude of the interrupted currents is independent of ψ. Also, slot lengths, L, and inclinations, C, are constant (except for mirror symmetry in inclination directions). Then it follows, for the illustration in Figure 2, that all vertical field components from the aperture reinforce on the axis, and all horizontal components cancel. The annulus produces linear polarization in the vertical plane in this instance.

Experience and analysis shows that the far field pattern of closely spaced short linear slots (arc spacing $<0.7\lambda$) in the quasi-annulus of Figure 2 is not significantly unlike that of a continuous annular slot. Therefore, for the case in Figure 2, the electric field in the equivalent continuous annular aperture is described by

$$\bar{E}_{slot} \alpha \hat{e}_\rho \frac{\sin \psi}{/\sin \psi/} \cos C + \hat{e}_\psi \frac{\cos \psi}{/\cos \psi/} \sin C \qquad (5)$$

Applying Bickmore's equations[5] for the $\hat{e}_\rho$ components and deriving, by duality, equations for the $\hat{e}_\psi$ components, the far field voltage patterns in the R, θ, φ space coordinate system of Figure 3 are

$$\begin{aligned} \bar{E}_\theta &\alpha \sum_{m=0}^{\infty} \left\{ \frac{k\rho \cos C}{(2m+1)} J'_{(2m+1)}(k\rho \sin\theta) + \frac{(-1)^m \sin C}{\sin\theta} J_{(2m+1)}(k\rho \sin\theta) \right\} \cos(2m+1)\varphi \\ \bar{E}_\varphi &\alpha \sum_{m=0}^{\infty} \left\{ \frac{\cos C}{\sin\theta} J_{(2m+1)}(k\rho \sin\theta) + \frac{(-1)^m k\rho \sin C}{(2m+1)} J'_{(2m+1)}(k\rho \sin\theta) \right\} \cos\theta \sin(2m+1)\varphi \end{aligned} \qquad (6)$$

with the exp (-jkR) and 1/R factors suppressed.*

*The theoretical difference patterns were also derived from Bickmore's results.

Here J and J' are the Bessel function of the first kind and its derivative, respectively. Equation (6) gives the sum beam pattern from a single element. The radiation field is linearly polarized parallel to the plane $\varphi = C^{\circ}$. When $C = 45^{\circ}$, the H plane and E plane beamwidths are both equal to $\sim 20.4^{\circ} \frac{\lambda}{\rho}$ and the first sidelobes are 7.3 db down.

Arrays

The sidelobe suppression and aperture efficiency are improved, with respect to a single annulus, when several annular slots are concentrically arrayed. In that case, the far field pattern becomes

$$\bar{E}_{\theta} \alpha \sum_{n=1}^{N} \sum_{m=0}^{\infty} A_n \exp \left[j\Delta_n \right] \left\{ \frac{k\rho_n \cos C_n}{(2m+1)} J'_{(2m+1)}(x_n) + \frac{(-1)^m \sin C_n}{\sin \theta} J_{(2m+1)}(x_n) \right\} \cos (2m+1)\varphi \qquad (7)$$

$$\bar{E}_{\varphi} \alpha \sum_{n=1}^{N} \sum_{m=0}^{\infty} A_n \exp \left[j\Delta_n \right] \left\{ \frac{\cos C_n}{\sin \theta} J_{(2m+1)}(x_n) + \frac{(-1)^m k\rho_n \sin C_n}{(2m+1)} J'_{(2m+1)}(x_n) \right\} \cos \theta \sin (2m+1)\varphi$$

where

$x_n = k\rho_n \sin \theta$,

A_n = voltage amplitude of nth annulus,

Δ_n = phase of nth annulus,

N = number of annuli.

Gain, Sidelobes, and Cross Polarization

Ideally, an annular slot for a linearly polarized pencil beam should have an excitation with the form[6]

$$\bar{E}_{slot} \alpha \hat{e}_{\rho} \sin \psi + \hat{e}_{\psi} \cos \psi \, . \qquad (8)$$

In contrast with the form given in Equation (5), this form produces a linearly polarized beam with no cross polarized lobes. The far field pattern of the ideal linearly polarized excitation may be derived from Equations (6) or (7) by taking just one term, the m = 0 term, in each of the infinite summations indicated. The effects of these other terms, for the case where $C = 45^{\circ}$, may be summarized by the following:

1. First sidelobes are -7.3 db instead of -7.9 db, for a single annulus;
2. First sidelobes tend towards -15.8db instead of -17.6db, for a uniform amplitude array of several appropriately spaced annular elements;
3. Cross polarized lobes are approximately -20db instead of $-\infty$ db;
4. The gain is reduced by 0.8db.

Generally, these consequences are not serious in view of the advantages in the simplicity of the structure. Techniques for obtaining the ideal annular slot excitation[6] did not appear to lend themselves to the monopulse objective of this program.

The Tilt Angle "C"

The peak values of the ψ and ρ components of Equation (8) are equal. The same is true of Equation (5) if $C = 45^{\circ}$. Since an approximation to Equation (8) is sought, the computations reported above have emphasized $C = 45^{\circ}$. The case of $C = 45^{\circ}$ produces equal beamwidths in the E and H planes. Furthermore, this case would produce maximum gain but for the fact that the H plane pattern differs from the E plane pattern by a multiplicative term cosine θ. Maximum gain occurs at approximately 48° for an array of closely spaced in-phase annular slots with uniform excitation.

The gain, first sidelobe level and beamwidth are plotted versus the angle C in Figure 4. The theoretical values given apply to an array of closely spaced annular slots ($< 0.7\lambda$ radial spacing) with a uniform amplitude distribution.

A final comment on Equations (5) and (8); analysis of measured radiation patterns of a single annulus strongly suggest that mutual coupling causes the behavior of the slot system pictured in Figure 2 to resemble that of Equation (8) more nearly than that of Equation (5). This is reasonable in view of the fact that coupled systems generally cannot contain sharp discontinuities across regions.

Design and Experimental Results

Wavelength Radial Spacing

To demonstrate the theories stated above, a series of X-band annular slot antennas were constructed. The units were approximately 7.5 wavelength in diameter. Standing wave array design was employed in order to avoid the losses associated with the power absorbed by a matched termination used in traveling wave arrays. Under this condition the radial dependence of the standing wave electric field (E_{ψ}) in the H_{01}

radial waveguide is given by $J_1(K_{10}\rho)$, the Bessel function of the first kind. When a short circuit is placed in the radial waveguide so that $J_1(K_{10}\rho) = 0$, (ρ_s being the radius of a cylindrical short circuit) then the radii at which the Bessel function of the second kind $N_1(K_{10}\rho_n) = 0$, designates the radii for the maxima of the circumferential currents in the plates of the radial waveguide. However, the currents at the radii which correspond to odd roots of N_1 are 180° out of phase with those at the even root radii. It is not possible to use the inclination of the short slots as a means to obtain in-phase radiation from out of phase currents (as is commonly practiced in the design of linear arrays on rectangular waveguide) since the inclination of the linear slots comprising an annulus are constrained by the radiation pattern requirement for each annular group. Thus, only three resonant annular slots were employed in the initial 7.5λ aperture design. An annular slot was located at each of the three in-phase roots of N_1 which are available in the 7.5λ diameter air-filled radial waveguide. For uniform aperture illumination and C = 45°, this case produces a theoretical gain figure of 22.5 db.

In order to design for a uniform amplitude distribution using these non-uniform transmission lines, two admittance expressions are needed to obtain the condition for correct power division between successive annuli. One expression gives the radial dependence for the characteristic admittance of these sector transmission lines for any radius, and the other, the input-output admittance equation. These are respectively,

$$Y_o(\rho) = \eta \frac{bK_{10}^2}{2k} \left[J_o^2(K_{10}\rho) + N_o^2(K_{10}\rho) \right] \qquad (9)$$

$$Y'_{in}(\rho) = \frac{Y_{in}(\rho)}{Y_o(\rho)} = \frac{Y'_{out}(\rho_o)\, Ct(x,y) + j\mathcal{J}(x,y)}{j\, Y'_{out}(\rho_o) + \mathcal{J}(x,y)\, ct(x,y)} \qquad (10)$$

where

$$x = K_{10}\rho, \quad y = K_{10}\rho_o$$

$$\mathcal{J}(x,y) = \frac{J_o(x)N_o(y) - N_o(x)J_o(y)}{J_1(x)N_1(y) - N_1(x)J_1(y)}$$

$$ct(x,y) = \frac{J_1(x)N_o(y) - N_1(x)J_o(y)}{J_o(x)N_o(y) - N_o(x)J_o(y)}$$

$$Ct(x,y) = \frac{J_1(y)N_o(x) - N_1(y)J_o(x)}{J_1(x)N_1(y) - N_1(x)J_1(y)}$$

Thus, $Y'_{in}(\rho)$ is the normalized admittance occurring at radius ρ when a normalized admittance $Y'_{out}(\rho_o)$ is placed at ρ_o where $\rho_o > \rho$. By appropriately employing these equations, the required admittance values for a uniform amplitude distribution were obtained for each array.

A limited slot conductance measurement program sought to obtain the appropriate conductance values for the slot configuration shown in Figure 2. The test apparatus consisted of a 90° radial waveguide sector on which the slot plates were measured; a transition section which provided a workable admittance transition from radial waveguide to standard rectangular waveguide; and a rectangular waveguide admittance measuring stand. The equivalent network parameters of the equations which describe the admittances in the radial waveguide in terms of the admittance in the rectangular waveguide were evaluated and provided the means for obtaining annular slot data. Each of the four sectors performs the same electrical function so all data was taken on one quadrant. Ground planes were erected to simulate mutual coupling between quadrants.

The primary factors which control the annular slot admittance or coupling are the angle of inclination (Angle "C") that the linear slots in an annulus make with the circumferential current tangent, the number of short slots in an annulus, and slot length. Two characteristics of the data aided in the acquisition of the annular slot admittance values. First, the fact that the variation in annular slot conductance, for small changes in angle "C", was similar to that of angle variations with an edge slot in the narrow wall of a rectangular waveguide[7]; thus, the power radiated could be predicted in terms of the equation for a resonant length edge slot. Secondly, the phase variations of each annular slot admittance as a function of slot length closely resembles that of a displaced longitudinal shunt slot in the broadwall of a rectangular waveguide; hence, the curves for the latter could be used to acquire the necessary phase value for each annulus. By utilizing these relationships the collection of slot data was facilitated.

As discussed previously, the aperture of one array consisted of three in-phase resonant annular slots. The annular slots were positioned at radial distances which corresponded to the odd roots of $N_1(K_{10}\rho)$; these were $k\rho_1 = 3.123$, $k\rho_2 = 12.218$ and $k\rho_3 = 21.175$ on the waveguide for which $k/K_{10} = 1.421$. In designing for a uniformly illuminated aperture the relative power to be radiated for each annulus becomes

$$P(\rho_1) = 1.000$$
$$P(\rho_2) = 3.912$$
$$P(\rho_3) = 6.780 \qquad (11)$$

since the amount of power radiated per annulus is proportional to its radius. Through application of (9), (10) and (11), and subject to the condition that the radial waveguide be matched at the

input, the required absolute conductances for each annular slot were determined to be

$$G(\rho_1) = 65.94\ (10^{-6}\text{mhos})$$
$$G(\rho_2) = 71.44\ (10^{-6}\text{mhos})$$
$$G(\rho_3) = 71.70\ (10^{-6}\text{mhos}).$$

Initial data suggested that smaller conductances might be more readily obtained so a new conductance value for each annular slot was then computed (for an input vswr of 1.2) yielding conductance values of

$$G(\rho_1) = 56.29\ (10^{-6}\text{ mhos})$$
$$G(\rho_2) = 60.99\ (10^{-6}\text{mhos})$$
$$G(\rho_3) = 61.21\ (10^{-6}\text{mhos})$$

The set of conductance values corresponding to a vswr of 1.2 was not reached in the time alloted for this phase of the experimental program. Figure 5 summarizes the values that were used in an array design. Subsequent data has shown that the desired values may be more closely obtained. Furthermore, these values may be obtained with the angle "C" more closely constrained to 45°.

The array designed in this phase of the program is shown in Figure 6. Each quadrant is fed by a magnetic dipole (a slot) at its apex. These are fed in turn by a coax line in a manner similar to that shown in the early prototype in Figure 7. The principal plane radiation patterns are shown in Figure 8. The antenna gain was determined to be 20.5 db, or 2 db below theoretical for the case. One decibel of this may be accounted for by losses in the feed structure. The second db loss is probably accounted for by the fact that the slot conductance value objectives were not realized at the time of the array's construction.

The front to back ratio (backlobe suppression) was found to be better than -45 db.

One-half Wavelength Radial Spacing

The three ring aperture described above has an outer ring radius of 21.175 radians. A uniformly illuminated circular aperture of this size has a theoretical gain of 26.5 db. A gain of 26.2 db is predicted by Equation (7) for the case of in-phase annular slots at each of the five roots of N_1 within the aperture (C_n's = 45°, A_n all equal). The fact that the currents at odd and even roots of N_1 are in phase opposition could not be overcome by staggering the tilts of the short slots. As described earlier, the slot tilts are firmly tied to the shape of the radiation patterns. In this matter, the rectangular waveguide planar array has a significant advantage since staggered slot tilts, or staggered slot offsets are routinely employed to achieve in-phase radiation with half-wave element spacing.

Three methods were considered for including two more annular slots in the aperture. One was dielectric loading to produce nine roots of N_1 in the given aperture, and thus have five in-phase roots for slot location. Another was the inclusion of obstacles with a transmission coefficient of $1\angle\pi$ and located at the roots of J_1 between the outer four roots of N_1. Though possible, these two methods were rejected because of weight and complexity, respectively. The third approach combined the phase of the radiators and the phase of the exciting currents in a manner which reduced the phase error. Since the phase of the currents of the successive roots of N_1 are given by

$$\alpha_n = \exp\left[j(n-1)\pi\right] \tag{12}$$

one may choose the admittances of successive annular slots to be non-resonant and to have alternating phase angles, β_n, of

$$\beta_n = \exp\left[-j(-1)^n \gamma\right] \tag{13}$$

The net phase radiated by the nth annular slot will then be

$$\begin{aligned} A_n &= \exp\left[j\left\{(n-1)\pi - (-1)^n\right\}\right] \\ &= j\exp\left[j(-1)^n\left(\tfrac{\pi}{2} - \gamma\right)\right] \end{aligned} \tag{14}$$

So, non-resonant annular slots having admittances with phase angle γ with successively positive and negative sign will combine with the π phased successive roots of N_1 to produce cophasality to plus and minus the complement of γ. Figure 9 illustrates the gain versus the non-resonance angle, γ. Obviously, $\gamma = 90°$ is impractical but with $\gamma = 60°$ the real part of the admittance is sufficiently above zero to obtain radiation and a reasonable input vswr. It should be noted that the large susceptances tend to cancel at the input as a result of the alternating signs. The theoretical patterns shown in Figure 10 illustrates the fact that the main beam and first sidelobes are fairly insensitive to γ. The far out sidelobes account for the loss in antenna gain for small values of γ.

In view of the foregoing, slot data was gathered to construct a five annulus experimental unit with $\gamma = 60°$. The theoretical gain for this case is 24.2 db when all C_n's are equal to 45°. The admittance values sought were

$$Y_1 = (39.5 + j84.7)\ (10^{-6}\text{mhos})$$
$$Y_2 = (42.4 - j60.5)\ (10^{-6}\text{mhos})$$
$$Y_3 = (42.8 + j91.8)\ (10^{-6}\text{mhos})$$
$$Y_4 = (42.9 - j61.3)\ (10^{-6}\text{mhos})$$
$$Y_5 = (43.0 + j92.1)\ (10^{-6}\text{mhos})$$

These values actually represent phase angles of 65° for n = 1,3,5 and -55° for n = 2,4. The shorter linear slots gave large phase angles more readily. The far field for this case is identical to that for $\gamma = 60^\circ$. To lessen the time to prototype construction, a unit was constructed when the slot data was approximately proportional to the desired values, to wit,

$$Y_1 = (11.7 + j49.4)\ (10^{-6}\text{mhos})$$
$$Y_2 = (16.9 - j18.4)\ (10^{-6}\text{mhos})$$
$$Y_3 = (14.4 + j40.8)\ (10^{-6}\text{mhos})$$
$$Y_4 = (11.5 - j13.1)\ (10^{-6}\text{mhos})$$
$$Y_5 = (15.2 + j65.0)\ (10^{-6}\text{mhos})$$

The innermost slot presented the greatest difficulty and its geometry departed from that of Figure 2.

The radiation patterns shown in Figure 11 compare the theoretical and experimental patterns for the five annulus array of non-resonant radiators pictured in Figure 11. The antenna gain was found to be 22.0 db after correction for vswr and for sum-difference circuit losses.

Conclusions

In the course of this program it has been demonstrated that monopulse antennas are readily designed around the use of a radial waveguide. Improvements in the match between theoretical and experimental patterns can be expected when improved slot design data is compiled. The radial waveguide appears to lead to the most simple mechanical configuration of the several waveguide types which may be used for a circular or nearly circular slot array. Productized versions of the experimental antennas described herein have been produced inexpensively with the aid of modern metal forming techniques and chemical milling of the slots. Weights in the order of 100 grams have been achieved with thin metal versions of a 7.5λ diameter X-band antenna.

With respect to small monopulse dish antennas, lightweight radial waveguide monopulse antennas offer a modest improvement in aperture efficiency, a marked improvement in back lobe suppression, a marked reduction in moment about the gimbal axes, and are competitive in fabrication costs. Other programs in which the authors are involved firmly establish the fact that various forms of rectilinear waveguide fed arrays offer the ultimate in aperture efficiencies in the design of both large and small monopulse antennas. Fabrication costs are greater than costs for radial waveguide antennas, however.

Acknowledgements

The authors gratefully acknowledge the extensive help received from Messrs. S. Stevens, D. L. Loyet, and J. L. Novak in the performance of the experimental program. Mr. G. R. Sturiale's studies on fabrication techniques produced several ingenious solutions and so enhanced the worth of the electromagnetic theory findings. Mr. E. Levitan deserves a note of thanks for his direction of the programming of the numerous gain and pattern computations which were essential to the execution of the program. A final note of thanks goes to Mrs. Marie English for her painstaking preparation of the manuscript.

References

1. K. C. Kelly, "Recent Annular Slot Array Experiments", IRE National Convention Record, Vol. 5, Part I, pp. 144-151, March 1957.

2. F. J. Goebels Jr. and K. C. Kelly, "Arbitrary Polarization from Annular Slot Planar Antennas", IRE Trans., Vol. AP-9, pp. 342-9, July 1961.

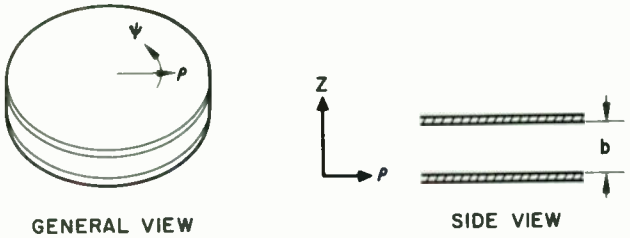
3. R. W. Bickmore, "Radiation Pattern Synthesis with Annular Slots", University of California ERL Report No. 112, Series 60, May 1954.

4. N. Marcuvity, Editor, "Waveguide Handbook", McGraw-Hill Book Company Inc., pp. 95-96, 1951.

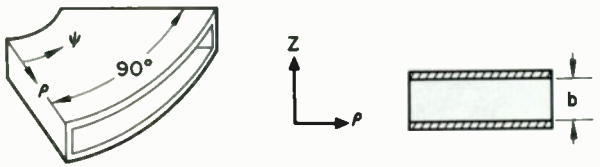
5. R. W. Bickmore, Op. Cit., pg. 2.

6. F. J. Goebels, Jr. and K. C. Kelly, Op. Cit., pg. 344.

7. I. P. Kaminow and R. J. Stegen, "Waveguide Slot Array Design", Technical Memorandum 348, Hughes Aircraft Company, Culver City, July 1954.



(a) RADIAL WAVEGUIDE OF CYLINDRICAL CROSS SECTION



(b) 90° RADIAL WAVEGUIDE SECTOR

Fig. 1. Two types of radial waveguides.

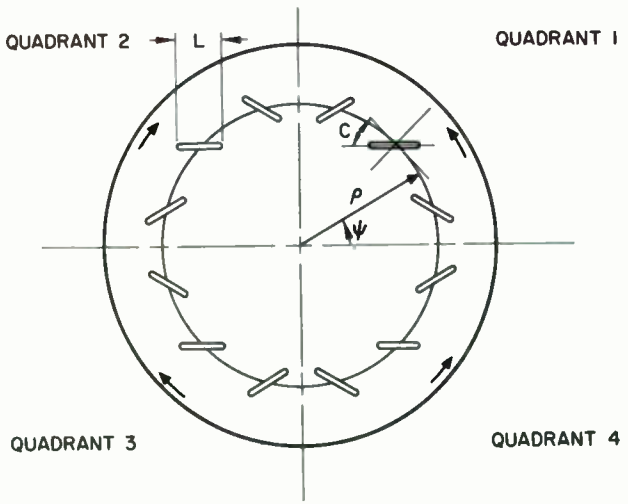


Fig. 2. A typical quasi-annular slot element.

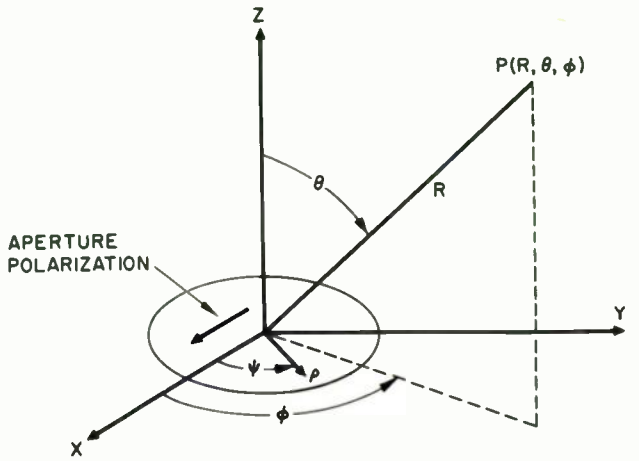


Fig. 3. Annular slot aperture and space geometry.

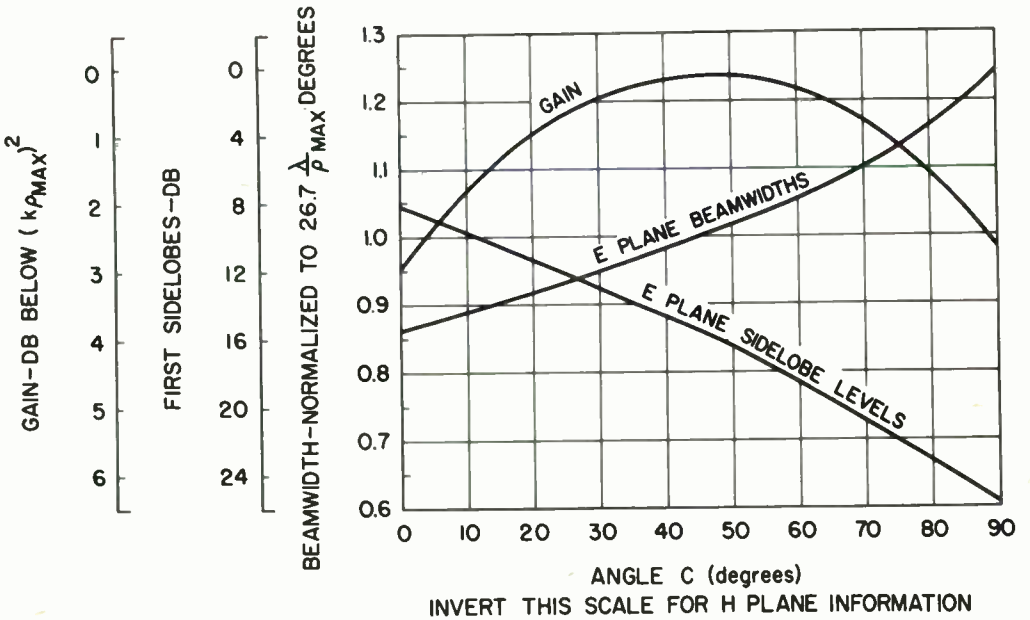
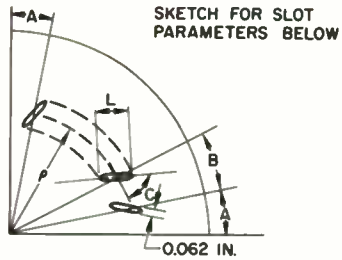


Fig. 4. Gain, sidelobe, and beamwidth variation with short slot tilt angle, C.



kρ (RADIANS)	NUMBER OF SLOTS	ANGLE "A"	ANGLE "B"	ANGLE "C"	SLOT LENGTH "L" (WAVELENGTHS)	SLOT ADMITTANCE (MICROMHOS)
3.123	1	45° 0'	—	20° 32'	0.571	35.51 + j 35.48
12.218	5	9° 0'	18° 0'	37° 3'	0.508	57.24 + j 5.214
21.175	8	5° 37'	11° 15'	60°	0.498	61.21 + j 5.408

Fig. 5. Slot admittance data for three annular slot array.

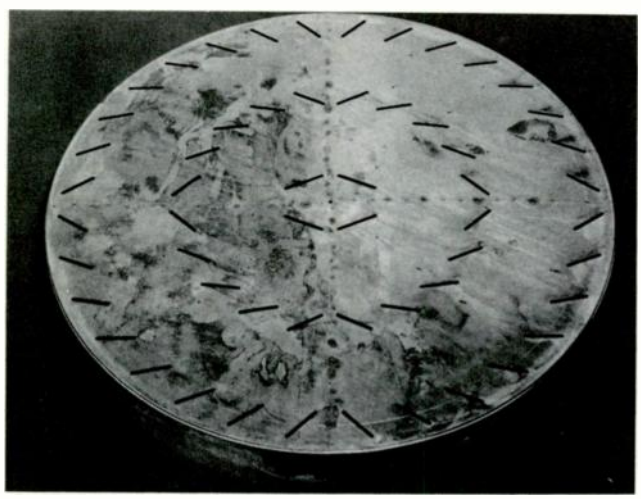


Fig. 6. Productized three annular slot monopulse antenna.

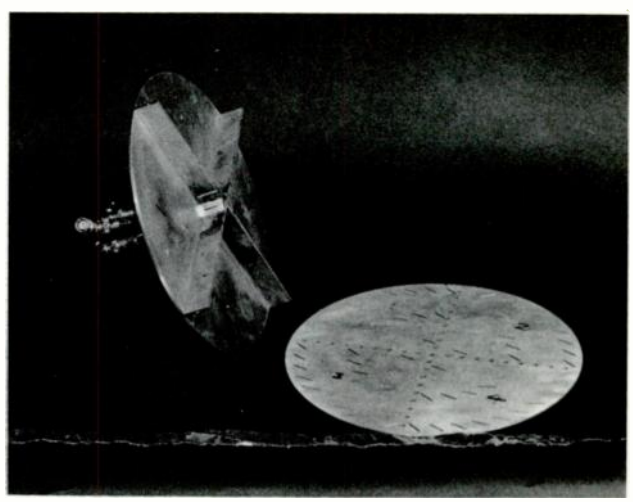


Fig. 7. Disassembled prototype annular slot monopulse antenna showing coax feeds.

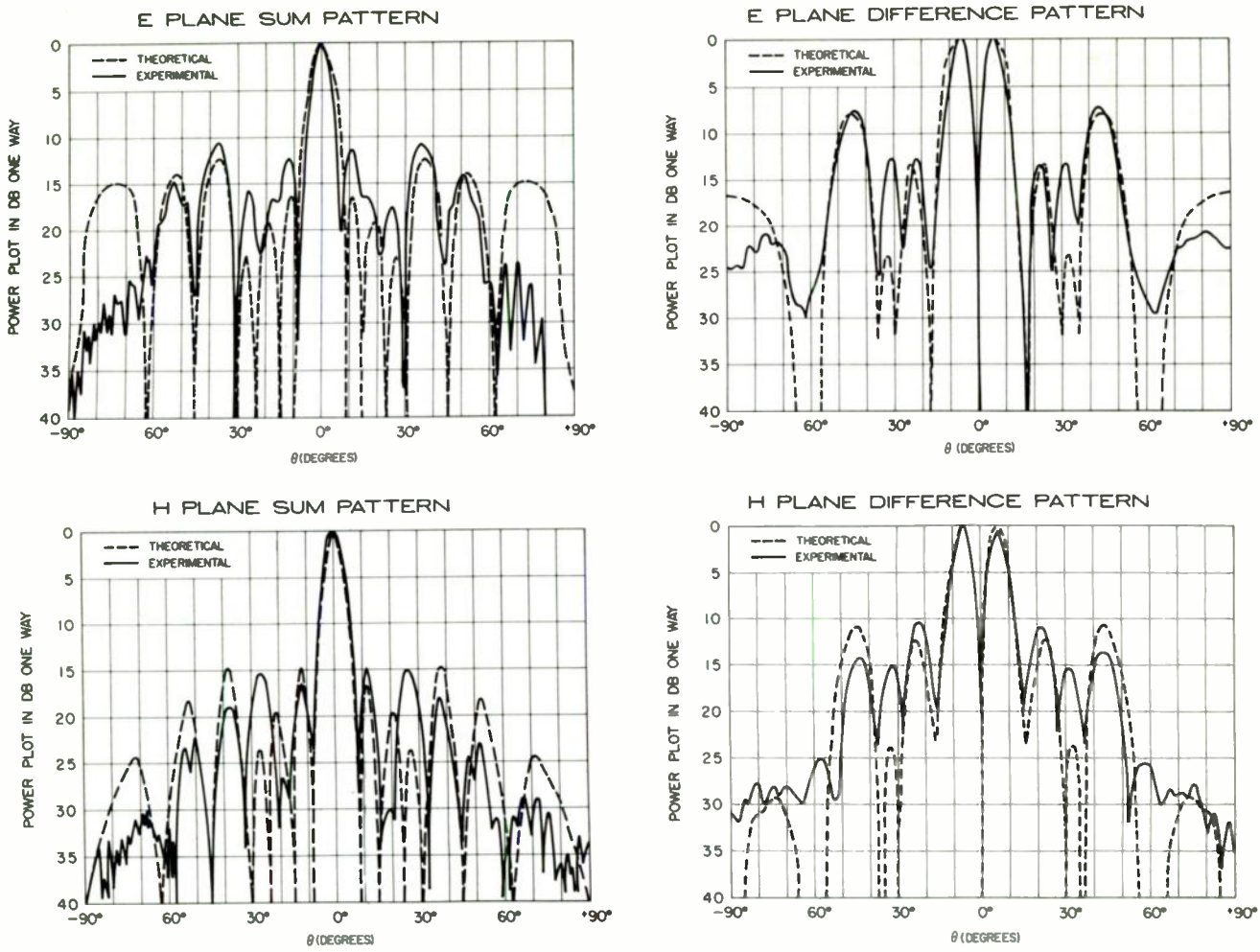


Fig. 8. Comparison of theoretical and experimental radiation patterns of three annular slot array.

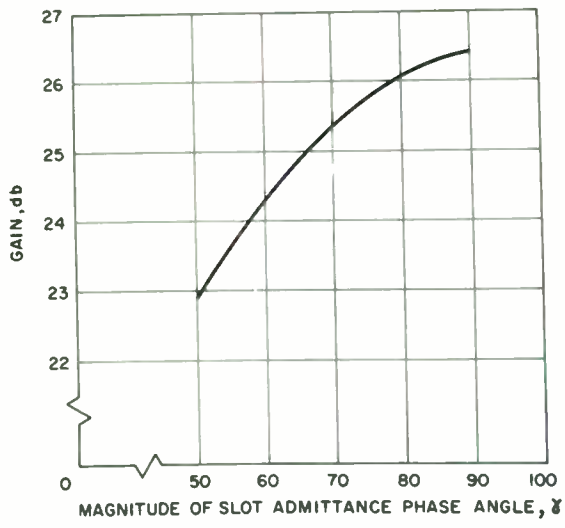


Fig. 9. Theoretical gain of five element array as a function of annular slot admittance phase angle.

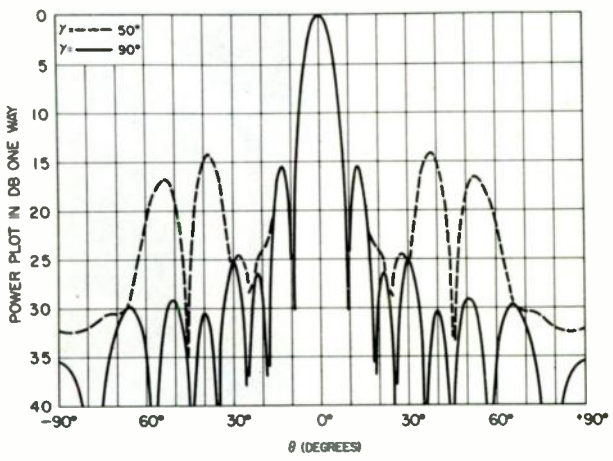


Fig. 10. Theoretical radiation patterns of five element array for $\gamma = 90°$ and $50°$ with $C = 45°$.

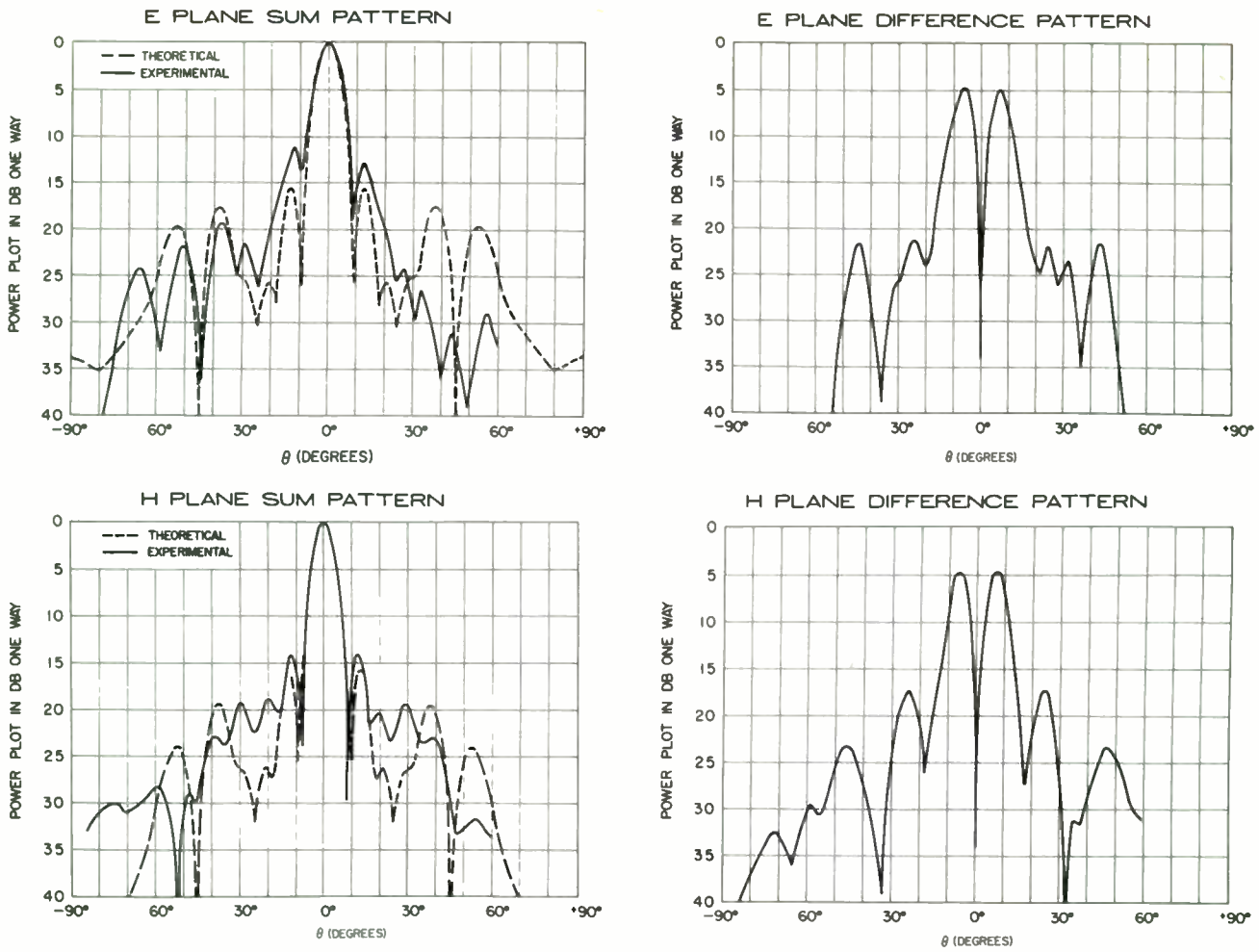


Fig. 11. Comparison of theoretical and experimental patterns of five annular slot array.

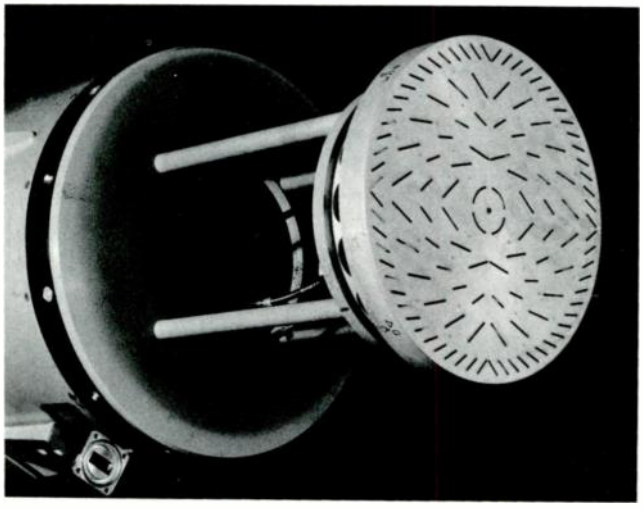


Fig. 12. Productized five annular slot array.

RADIATION CHARACTERISTICS OF SLOT ANTENNAS COVERED WITH A PLASMA SLAB IN THE PRESENCE OF A STATIC MAGNETIC FIELD PERPENDICULAR TO THE SLOT

H. Hodara
The Hallicrafters Company
Chicago, Illinois

Summary

It is shown that the presence of a gyrotropic lossless plasma slab covering a rectangular slot radiator does not alter significantly the radiation pattern in the vertical plane when the static magnetic field is perpendicular to the slot. The plasma covered antenna remains omnidirectional and the radiated power is not significantly reduced when the cyclotron frequency exceeds the operating frequency by one order of magnitude ($\omega_b^2 >> \omega^2$).

It is also found that the matching between the antenna and its surrounding is sharply improved, if the antenna is covered with such a slab, at critical combinations of the plasma slab parameters and the operating frequency; this finding may be useful for plasma diagnostics, in order to determine the slab width L, the plasma frequency ω_p, and the collision frequency ν, by varying ω_b and ω.

Introduction

The radiation characteristics of slot antennas covered with a gyrotropic plasma slab are analyzed in this paper.

The slot is assumed to be supported by a perfect conductor, infinite in extent in the x-y plane; the plasma is rendered gyrotropic by means of a uniform static magnetic field impressed perpendicularly to the slot (Figure 1). An expression is first developed for the radiation fields for the case of a slot of arbitrary geometry and excitation. The radiation fields are then evaluated explicitly for a rectangular slot across which an electric field in the TE_{10} mode is impressed rigidly.

Radiation patterns for some values of the plasma slab parameters are calculated for the case of an infinitely long narrow slot excited by a constant voltage V across it.

Derivation of the Vector Wave Equation in Anisotropic Plasmas

The steady state matrix form of Maxwell's equations is:

$$(1) \qquad \nabla \times \begin{vmatrix} \mathbf{E} \\ \mathbf{H} \end{vmatrix} = j\omega \begin{vmatrix} -\mu_v \mathbf{H} \\ \epsilon_v \mathfrak{E}_r \cdot \mathbf{H} \end{vmatrix}$$

where μ_v and ϵ_v are the vacuum permeability and permittivity respectively and ω is the wave frequency in radians per second. The relative dyadic permittivity $\mathfrak{E}_r$ has components given by the following matrix in the case of a z-directed static magnetic field $\mathbf{B}_o$:

$$(2) \qquad \left| \mathfrak{E}_r \right| = \begin{vmatrix} \epsilon_1 & \epsilon_2 & 0 \\ -\epsilon_2 & \epsilon_1 & 0 \\ 0 & 0 & \epsilon_3 \end{vmatrix}$$

The matrix components in terms of the plasma frequency ω_p, cyclotron frequency ω_b, and the collision frequency ν are given by

$$(3) \qquad \begin{vmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{vmatrix} = \begin{vmatrix} 1 \\ 0 \\ 1 \end{vmatrix} + \frac{\omega_p^2}{j\omega} \begin{vmatrix} \dfrac{\nu + j\omega}{\omega_b^2 + (\nu + j\omega)^2} \\ \dfrac{-\omega_b}{\omega_b^2 + (\nu + j\omega)^2} \\ \dfrac{1}{\nu + j\omega} \end{vmatrix}$$

Eliminating $\mathbf{H}$ in favor of $\mathbf{E}$ in (1) gives the vector wave equation for the electric field

(4) $$\nabla \times (\nabla \times \mathbb{E}) - \beta_v^2 \mathbb{\epsilon}_r \cdot \mathbb{E} = 0$$

where β_v, the wave number in vacuum is given by

(5) $$\beta_v^2 = -j\omega\mu_v j\omega\epsilon_v$$

Resolution of the Vector Wave Equation into Scalar Equations

Mathematical simplicity for the geometry under consideration dictates the choice of Cartesian coordinates, x,y,z. Expansion of the vector wave equation (4) in terms of rectangular coordinates using (2) yields:

(6) $$\begin{vmatrix} \partial_y^2 + \partial_z^2 + \beta_v^2\epsilon_1 & -(\partial_x\partial_y - \beta_v^2\epsilon_2) & -\partial_x\partial_z \\ -(\beta_v^2\epsilon_2 + \partial_x\partial_y) & \partial_x^2 + \partial_z^2 + \beta_v^2\epsilon_1 & -\partial_y\partial_z \\ -\partial_x\partial_z & -\partial_y\partial_z & \nabla_T^2 + \beta_v^2\epsilon_3 \end{vmatrix} \begin{vmatrix} E_x \\ E_y \\ E_z \end{vmatrix} = \begin{vmatrix} 0 \\ 0 \\ 0 \end{vmatrix}$$

where ∇_T, the transverse operator, is:

(7) $$\nabla_T = 1_x\partial_x + 1_y\partial_y$$

Solution of the Vector Wave Equation

Eliminating E_x and E_y in favor of E_z in (6) yields a fourth order partial differential equation. This equation is readily obtained by rewriting (6) as follows:

(8) $$\begin{vmatrix} 0_{11} & 0_{12} & 0_{13} \\ 0_{21} & 0_{22} & 0_{23} \\ 0_{31} & 0_{32} & 0_{33} \end{vmatrix} \begin{vmatrix} E_x \\ E_y \\ E_z \end{vmatrix} = \begin{vmatrix} 0 \\ 0 \\ 0 \end{vmatrix}$$

where the differential operators 0_{ij} in (8) are identical with the corresponding terms in (6) .

The first two rows of (8) give:

(9) $$\begin{vmatrix} 0_{11} & 0_{12} \\ 0_{21} & 0_{22} \end{vmatrix} \begin{vmatrix} E_x \\ E_y \end{vmatrix} = \begin{vmatrix} -0_{13} \\ -0_{23} \end{vmatrix} E_z$$

yielding for E_x and E_y:

(10a) $$DE_x = (0_{12}0_{23} - 0_{13}0_{22})\,E_z$$

(10b) $$DE_y = (0_{21}0_{13} - 0_{23}0_{11})\,E_z$$

D is the determinant of (9) :

(11) $$D = 0_{11}0_{22} - 0_{12}0_{21}$$

Substitution of (10a) and (10b) into the third row of (8) yields after multiplication by the differential operator D from (11) :

(12) $$\{ 0_{31}(0_{12}0_{23} - 0_{13}0_{22}) + 0_{32}(0_{21}0_{13} - 0_{23}0_{11}) + 0_{33}(0_{11}0_{22} - 0_{12}0_{21}) \}\, E_z = 0$$

Expansion, manipulation, and simplification of (12) in terms of the explicit expressions for the 0's from (6) yield the desired differential equation for E_z.

(13) $$\{ \epsilon_1\nabla_T^4 + (\epsilon_1 + \epsilon_3)\partial_z^2\nabla_T^2 + \beta_v^2(\epsilon_T^2 + \epsilon_1\epsilon_3)\nabla_T^2 + 2\beta_v^2\epsilon_1\epsilon_3\partial_z^2 + \epsilon_3\partial_z^4 + \beta_v^4\epsilon_T^2\epsilon_3 \}\, E_z = 0$$

where

(14) $$\epsilon_T^2 = \epsilon_1^2 + \epsilon_2^2$$

(15) $$\nabla_T^4 = \partial_x^4 + 2\partial_x^2\partial_y^2 + \partial_y^2$$

The method of separation of variable does not work for (13) because of the cross-term $\partial_z^2\nabla_T^2$. This difficulty is eliminated by taking the double Fourier transform of E_z with respect to x and y, namely

(16a) $$\bar{\bar{E}}_z(k,p,z) = \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_z e^{-ikx} e^{-ipy}\, dx\, dy$$

The two bars above $\bar{\bar{E}}_z$ indicate double Fourier transform. k and p are wave numbers in the s and y direction respectively. The inverse transform is defined as:

(16b) $$E_z(x,y,z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \bar{\bar{E}}_z(k,p,z)\, e^{ikx} e^{ipy}\, dk\, dp$$

Applying (16a) to (13), noting that

(17a) $$\left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \nabla_T^2 E_z e^{-ikx} e^{-ipy}\, dx\, dy = -(k^2+p^2)\bar{\bar{E}}_z$$

(17b) $$\left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \nabla_T^4 E_z e^{-ikx} e^{-ipy}\, dx\, dy = (k^2+p^2)^2\bar{\bar{E}}_z$$

and letting

(18) $$u^2 = k^2 + p^2$$

gives

(19) $$\left\{\partial_z^4 + 2\beta_v^2\left(\epsilon_1 - \frac{\epsilon_1+\epsilon_3}{2\epsilon_3}\frac{u^2}{\beta_v^2}\right)\partial_z^2 + \beta_v^4\left(\epsilon_T^2 - \frac{\epsilon_T^2+\epsilon_1\epsilon_3}{\epsilon_3}\frac{u^2}{\beta_v^2} + \frac{\epsilon_1}{\epsilon_3}\frac{u^4}{\beta_v^4}\right)\right\}\bar{\bar{E}}_z = 0$$

Solution of (19) by the method of characteristic roots gives, after substitution of

(20) $$\gamma = j\beta = \partial_z(\ )$$

in (19), the following solution for $\bar{\bar{E}}_z$

(21) $$\bar{\bar{E}}_z = \sum_{n=-2}^{n=2}{}^{\neq 0}\, e_{nz}(k,p)\, e^{i\beta_n z}$$

e_{nz} is a mode amplitude; substituting (21) in (19) yields the eigen values for β_n:

(22) $$\frac{\beta}{\beta_v} = \pm\left\{\left[\epsilon_1 - \frac{\epsilon_1+\epsilon_3}{2\epsilon_3}\frac{u^2}{\beta_v^2}\right] \pm \sqrt{\left[\frac{\epsilon_1-\epsilon_3}{2\epsilon_3}\frac{u^2}{\beta_v^2}\right]^2 - \epsilon_2^2\left[1 - \frac{u^2}{\beta_v^2\epsilon_3}\right]}\right\}^{\frac{1}{2}} \quad (1)$$

There are four values to β_n given by the four possible sign combinations. Substituting (21) into (16b) gives the solution for E_z

(23) $$E_z(x,y,z) = \sum_{n=-2}^{n=2}{}^{\neq 0} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e_{nz}(k,p)\, e^{ikx} e^{ipy} e^{i\beta_n z}\, dk\, dp = \sum_{n=-2}^{n=2}{}^{\neq 0}\, e_{nz}$$

Equation (23) is given the following physical interpretation: The field components are made up of two pairs of sets of modes, set 1 and set 2. To every pair of values k,p corresponds two modes in each set, a forward mode β_{-n} and a backward mode β_n, where:

(24) $$\beta_{-n} = -\beta_n \quad (n = 1,2)$$

The mode members of each set have a respective phase velocity given by $v_{1,2} = \frac{\omega}{\beta_{1,2}}$. ω is the angular frequency of the electromagnetic wave. The mode numbers k, p take on continuous real values between $-\infty$ and $+\infty$. They form two continuous spectra corresponding to the two orthogonal transverse direction x and y, since the geometry of the problem is unbounded in this direction.(2)

Expressions for the Transforms of the Field Components

The previous derivation shows that all field components are of the same form as (23). In particular

$$H_z(x,y,z) = \sum_{\substack{n=-2 \\ n\neq 0}}^{n=2} \int_{-\infty}^{\infty}\int h_{nz}(k,p)\, e^{jkx} e^{jpy} e^{j\beta_n z}\, dk\, dp \tag{25a}$$

where

$$h_{nz}(k,p) = \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty}\int H_{nz}(x,y,z)\, e^{-jkx} e^{-jpy} e^{-j\beta_n z} \tag{25b}$$

Because of the linearity of Maxwell's equations, each component of the sum (24a) and (23) satisfies (1). Substituting (16) and (25) for arbitrary suscript n into (1) yields:

$$pe_{nz} - \beta_n e_{ny} = -\omega\mu_v h_{nx} \tag{26a}$$

$$\beta_n e_{nx} - ke_{nz} = -\omega\mu_v h_{ny} \tag{26b}$$

$$ke_{ny} - pe_{nx} = -\omega\mu_v h_{nz} \tag{26c}$$

$$kh_{ny} - ph_{nx} = \omega\epsilon_v\epsilon_3 e_{nz} \tag{26d}$$

$$ph_{nz} - \beta h_{ny} = \omega\epsilon_v(\epsilon_1 e_{nx} + \epsilon_2 e_{ny}) \tag{27a}$$

$$\beta h_{nx} - kh_{nz} = \omega\epsilon_v(-\epsilon_2 e_{nx} + \epsilon_1 e_{ny}) \tag{27b}$$

The transverse mode amplitudes can be expressed in terms of the longitudinal ones through the set of equations (26)

$$\begin{bmatrix} e_{nx} \\ e_{ny} \\ h_{nx} \\ h_{ny} \end{bmatrix} = -\frac{1}{u^2} \begin{bmatrix} \frac{k}{\beta_n}\beta_c^2 & -p\omega\mu_v \\ \frac{p}{\beta_n}\beta_c^2 & k\omega\mu_v \\ p\omega\epsilon_v\epsilon_3 & k\beta_n \\ -k\omega\epsilon_v\epsilon_3 & p\beta_n \end{bmatrix} \begin{bmatrix} e_{nz} \\ h_{nz} \end{bmatrix} \tag{28}$$

where

$$\beta_c^2 = \beta_v^2\epsilon_3 - u^2 \tag{29}$$

Substituting (28) and (29) in (27) gives the determinantal equation,

$$\begin{bmatrix} -\beta_n\beta_v\epsilon_2^2 u^2 & \left[-(\beta_v^2\epsilon_1 - \beta_n^2)u^2 + (\beta_v^4\epsilon_T^2 - 2\beta_n^2\beta_v^2\epsilon_1 + \beta_n^4)\right]\omega\mu_v \\ \left[-(\beta_v^2\epsilon_T^2 - \beta_n^2\epsilon_1)u^2 + \epsilon_3(\beta_v^4\epsilon_T^2 - 2\beta_n^2\beta_v^2\epsilon_1 + \beta_v^4)\right]\omega\epsilon_v & \beta_n\beta_v^2\epsilon_2 u^2 \end{bmatrix} \begin{bmatrix} e_{nz} \\ h_{nz} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{30}$$

Equation (30) shows that e_{nz} and h_{nz} are not independent. The coupling between the longitudinal modal components is due to the anisotropy of the medium. If the determinant of (30) is equated to zero, the resulting determinantal equation yields, of course, β's identical with (22).

At this point it will be enlightening to re-derive (22) as follows. Since e_{nz} and h_{nz} are related algebraically through (30), let

(31) $h_{nz} = Y_n e_{nz}$

Y_n is a function of u, β_n, ω and the constitutive parameters of the medium; it has the dimension of admittance, h_{nz} and e_{nz} have the dimensions of current times distance and voltage times distance respectively. For each mode, that is for each pair of values of k,p and a given value of n, a model can be conceived consisting of a transmission line of characteristic admittance Y_n with voltage e_{nz} and current h_{nz}.

Substituting (31) into (30) yields two equations for Y_n,

(32a) $$Y_n = - \frac{\omega\epsilon_v\left[\epsilon_3(\beta_v^4\epsilon_T^2 - 2\beta_n^2\beta_v^2\epsilon_1 + \beta_n^4) - (\beta_v^2\epsilon_T^2 - \beta_n^2\epsilon_1)u^2\right]}{\beta_n\beta_v^2\epsilon_2 u^2}$$

(32b) $$Y_n = - \frac{\beta_n\beta_v^2\epsilon_2 u^2}{\omega\mu_v\left[(\beta_v^2\epsilon_1 - \beta_n^2)u^2 - (\beta_v^4\epsilon_T^2 - 2\beta_n^2\beta_v^2\epsilon_1 + \beta_n^4)\right]}$$

Equating the RHS of (32) gives the determinantal equation for β which becomes after a few manipulations

(33) $$\left\{\beta_v^2\epsilon_3\right\}\left\{\beta_n^4 - 2\beta_n^2\beta_v^2\epsilon_1 + \beta_v^4\epsilon_T^2\right\}\left\{\beta_n^4 - 2\beta_n^2\left[\beta_v^2\epsilon_1 - \left(\frac{\epsilon_1+\epsilon_3}{2\epsilon_3}\right)u^2\right] + \left[\beta_v^4\epsilon_T^2 - \beta_v^2 u^2\left(\frac{\epsilon_T^2}{\epsilon_3} + \epsilon_1\right) + \frac{\epsilon_1}{\epsilon_3}u^4\right]\right\} = 0$$

The first curled bracket of (32) is in general different from zero. Equating the second curled bracket to zero yields the two possible values of β for uniform plane waves in a gyrotropic plasma with a static magnetic field parallel to the Poynting vector,

(34a) $$\left.\frac{\beta_n^2}{\beta_v^2}\right|_{\text{plane wave}} = \epsilon_1 \pm i\epsilon_2$$

Equating the third curled bracket to zero,

(34b) $$\beta_n^4 - 2\beta_n^2\left[\beta_v^2\epsilon_1 - \left(\frac{\epsilon_1+\epsilon_3}{2\epsilon_3}\right)u^2\right] + \left[\beta_v^4\epsilon_T^2 - \beta_v^2\left(\frac{\epsilon_T^2}{\epsilon_3} + \epsilon_1\right)u^2 + \frac{\epsilon_1}{\epsilon_3}u^4\right] = 0$$

gives the determinantal equation for β in its most general form for the problem on hand. (34b) is identical to (19) if (20) is used. The resultant expression (22) for β is repeated for convenience,

(35) $$\frac{\beta_n^2}{\beta_v^2} = \left[\epsilon_1 - \frac{\epsilon_1+\epsilon_3}{2}\frac{u^2}{\beta_v^2\epsilon_3}\right] \pm \sqrt{\left[\frac{\epsilon_1-\epsilon_3}{2}\frac{u^2}{\beta_v^2\epsilon_3}\right]^2 - \epsilon_2^2\left[1 - \frac{u^2}{\beta_v^2\epsilon_3}\right]}$$

Note that (35) reduces to (33) when $u = 0$, i.e. in the case of two single uniform plane waves. Thus (34) is the general determinantal equation since it contains the second bracket of (33).

The expressions (31a) and (31b) for the mode admittance Y are greatly simplified by using (5), (34) and the following relation

(36) $$\epsilon_3\left[\beta_n^4 - 2\beta_n^2\beta_v^2\epsilon_1 + \beta_v^4\epsilon_T^2\right] = u^2\left[\beta_v^2(\epsilon_T^2 + \epsilon_1\epsilon_3) - \beta_n^2(\epsilon_1+\epsilon_3) - u^2\epsilon_1\right]$$

obtained from (34b), thus

(37a) $$Y_n = \frac{\epsilon_3\beta_n^2 - \epsilon_1\beta_c^2}{\epsilon_2\beta_n\omega\mu_v}$$

Further manipulations yield the alternate form

(37b) $$Y_n = \frac{\beta_c^2}{\beta_n} \frac{\omega \epsilon_v \epsilon_2}{\beta_v^2 \epsilon_1 - (u^2 + \beta_n^2)}$$

Note that

(37c) $$Y_{-n} = -Y_n$$

Substituting (30) into (27) and using the relation for the wave admittance of vacuum,

(38) $$Y_v = \sqrt{\frac{\epsilon_v}{\mu_v}}$$

gives the expressions for the transverse field components in terms of e_{nz}

(39) $$\begin{vmatrix} e_{nx} \\ e_{ny} \\ h_{nx} \\ h_{ny} \end{vmatrix} = -\frac{e_{nz}}{u^2} \begin{vmatrix} k\beta_c \frac{\beta_c}{\beta_n} - p\beta_v \frac{Y_n}{Y_v} \\ p\beta_c \frac{\beta_c}{\beta_n} + k\beta_v \frac{Y_n}{Y_v} \\ p\beta_v \epsilon_3 Y_v + k\beta_n Y_n \\ -k\beta_v \epsilon_3 Y_v + p\beta_n Y_n \end{vmatrix}$$

The transforms of the total longitudinal fields are obtained by applying (16a) to (23)

(40a) $$\bar{\bar{E}}_z(k,p,z) = \sum_{n=-2}^{n=2} {}_{n\neq 0}\, e_{nz} e^{j\beta_n z} = \sum_{n=-2}^{n=2} {}_{n\neq 0}\, \bar{\bar{E}}_{nz}(k,p,z)$$

Similarly

(40b) $$H_z(k,p,z) = \sum_{n=-2}^{n=2} {}_{n\neq 0}\, Y_n h_{nz} e^{j\beta_n z} = \sum_{n=-2}^{n=2} {}_{n\neq 0}\, Y_n \bar{\bar{E}}_{nz}(k,p,z)$$

after using (31). The transforms of the total transverse field components are obtained by means of (39),

(41) $$\begin{vmatrix} \bar{\bar{E}}_{nx}(k,p,z) \\ \bar{\bar{E}}_{ny}(k,p,z) \\ \bar{\bar{H}}_{nx}(k,p,z) \\ \bar{\bar{H}}_{ny}(k,p,z) \end{vmatrix} = -\frac{1}{u^2} \sum_{n=-2}^{n=2} {}_{n\neq 0}\, e^{j\beta_n z} e_{nz} \begin{vmatrix} k\beta_c \frac{\beta_c}{\beta_n} - p\beta_v \frac{Y_n}{Y_v} \\ p\beta_c \frac{\beta_c}{\beta_n} + k\beta_v \frac{Y_n}{Y_v} \\ p\beta_v Y_v \epsilon_3 + k\beta_n Y_n \\ -k\beta_v Y_v \epsilon_3 + p\beta_n Y_n \end{vmatrix}$$

Isotropic Case

When the anisotropy of the medium is removed by letting ω_b approach zero, the expressions for the modal admittances simplify considerably.

Let Y be the admittance of the unbounded isotropic plasma

(42) $$Y = \sqrt{\frac{\epsilon_3 \epsilon_v}{\mu_v}}$$

then a few algebraic manipulations on (37) show that

$$(43)\qquad \lim_{\omega_b \to 0} \begin{vmatrix} Y_{-2} \\ Y_{-1} \\ Y_1 \\ Y_2 \end{vmatrix} = Y \begin{vmatrix} i \\ -i \\ i \\ -i \end{vmatrix}$$

Note that positive and negative values of jY_n for $n < 0$ in the limit case ($\omega_b = 0$) do not correspond to forward and backward waves but instead correspond to the first set and second set of modes respectively. It will be recalled that these two sets of modes are the ones that reduce to right handed and left handed circularly polarized waves in the cases of uniform plane wave propagation.

In the isotropic case

$$(44)\qquad \lim_{\omega_b \to 0} \beta_n = \beta = \sqrt{\beta_v^2 \epsilon_3 - u^2}$$

and (40) reduces to

$$(45a)\qquad E_z(k,p,z) = (e_{1z} + e_{2z})\, e^{j\beta z} + (e_{-1z} + e_{-2z})\, e^{-j\beta z}$$

$$(45b)\qquad H_z(k,p,z) = Y(e_{1z} - e_{2z})\, e^{j\beta z} + Y(e_{-1z} - e_{-2z})\, e^{-j\beta z}$$

The factors preceding the exponential of (45) are the backward and forward modal amplitudes for isotropic media,

$$(46a)\qquad e_{1z} + e_{2z} = e_z^+$$

$$(46b)\qquad e_{-1z} + e_{-2z} = e_z^-$$

$$(46c)\qquad e_{1z} - e_{2z} = h_z^+$$

$$(46d)\qquad e_{-1z} - e_{-2z} = h_z^-$$

e_z^-, h_z^- are forward waves; e_z^+, h_z^+ are backward waves.

Substituting (46) into (45) yields

$$(47a)\qquad E_z(k,p,z) = e_z^+ e^{j\beta z} + e_z^- e^{-j\beta z}$$

$$(47b)\qquad H_z(k,p,z) = h_z^+ e^{j\beta z} + h_z^- e^{-j\beta z}$$

Note that in the isotropic unbounded case e_z^+, e_z^-, h_z^+, h_z^- are independent until boundary conditions are imposed. Furthermore in a problem where TM modes are absent, $e_{\pm 1z} = e_{\pm 2z}$.

As a special case, the forward transverse field modal components for unbounded vacuum are derived as these expressions will be needed later on. In vacuum,

$$(48)\qquad \beta_c^2 \longrightarrow \beta_o^2 = \beta_v^2 - u^2$$

$$(49)\qquad \beta_n \longrightarrow -\beta_o$$

$$(50)\qquad Y_n \longrightarrow Y_v$$

Denoting the forward mode amplitudes by an "o" suscript and substituting (46), (48), (49), into (39) for $n = -1, -2$ gives

$$(51)\qquad \begin{vmatrix} e_{ox} \\ e_{oy} \\ h_{ox} \\ h_{oy} \end{vmatrix} = \frac{1}{u^2} \begin{vmatrix} k\beta_o & p\frac{\beta_v}{Y_v} \\ p\beta_o & -k\frac{\beta_v}{Y_v} \\ -p\beta_v Y_v & k\beta_o \\ k\beta_v Y_v & p\beta_o \end{vmatrix} \begin{vmatrix} e_{oz} \\ h_{oz} \end{vmatrix}$$

Boundary Conditions

The problem of interest consists of an arbitrary radiating aperture in a perfect plane conductor of infinite extent in the x-y directions located at z = 0. The plane conductor is covered with a slab of anisotropic plasma of width L and situated in vacuum. Hence, the secondary boundary of the problem consists of a plane interface between the gyrotropic plasma and vacuum at z = L. (Figure 1)

In order to solve for the field components throughout space: 1) the field components in the plasma slab must be matched to the vacuum field components at z = L, 2) the plasma field components in the slab must be matched to the applied or exciting field over the conductor. The conductor is excited with an applied electric field of component $E_x^a(x,y,0)$ and $E_y^a(x,y,0)$ over the aperture; naturally the electric field components vanish over the conductor; E^a and H^a cannot be specified independently over the plane z = 0 since H is determined by E from Maxwell's equation. This is analogous to the network problem in which if the voltage source (applied E-field) and the network parameters (geometry of the region) are specified, the current (H-field) is then determined.

The boundary conditions will be applied to the Fourier transforms of the field components.

Boundary Conditions at z = L

Using (39), (40), and (51), the two longitudinal and four transverse modal components are matched at z = L,

$$(52a)\qquad \sum_{\substack{n=-2 \\ n\neq 0}}^{n=2} \epsilon_3 e_{nz} e^{j\beta_n L} = e_{oz} e^{-j\beta_o L}$$

$$(52b)\qquad \sum_{\substack{n=-2 \\ n\neq 0}}^{n=2} Y_n e_{nz} e^{j\beta_n L} = h_{oz} e^{-j\beta_o L}$$

$$(52c)\qquad \sum_{\substack{n=-2 \\ n\neq 0}}^{n=2} k\beta_c \frac{\beta_c}{\beta_n} e_{nz} e^{j\beta_n L} - \sum_{\substack{n=-2 \\ n\neq 0}}^{n=2} p\beta_v \frac{Y_n}{Y_v} e_{nz} e^{j\beta_n L} = -k\beta_o e_{oz} e^{-j\beta_o L} - p\frac{\beta_v}{Y_v} h_{oz} e^{-j\beta_o L}$$

$$(52d)\qquad \sum_{\substack{n=-2 \\ n\neq 0}}^{n=2} p\beta_c \frac{\beta_c}{\beta_n} e_{nz} e^{j\beta_n L} + \sum_{\substack{n=-2 \\ n\neq 0}}^{n=2} k\beta_v \frac{Y_n}{Y_v} e_{nz} e^{j\beta_n L} = -p\beta_o e_{oz} e^{-j\beta_o L} + k\frac{\beta_v}{Y_v} h_{oz} e^{-j\beta_o L}$$

$$(52e)\qquad \sum_{\substack{n=-2 \\ n\neq 0}}^{n=2} p\beta_v \epsilon_3 Y_v e_{nz} e^{j\beta_n L} + \sum_{\substack{n=-2 \\ n\neq 0}}^{n=2} k\beta_n Y_n e_{nz} e^{j\beta_n L} = +p\beta_v Y_v e_{oz} e^{-j\beta_o L} - k\beta_o h_{oz} e^{-j\beta_o L}$$

$$(52f)\qquad \sum_{\substack{n=-2 \\ n\neq 0}}^{n=2} -k\beta_v \epsilon_3 Y_v e_{nz} e^{j\beta_n L} + \sum_{\substack{n=-2 \\ n\neq 0}}^{n=2} p\beta_n Y_n e_{nz} e^{j\beta_n L} = -k\beta_v Y_v e_{oz} e^{-j\beta_o L} - p\beta_o h_{oz} e^{-j\beta_o L}$$

Only four out of the six equations in the set (52) are independent. Any two equations can be expressed as a linear combination of the others, thus:

$$(53a)\quad p(52e) - k(52f) = \sum_{\substack{n=-2 \\ n\neq 0}}^{n=2} u^2\beta_v \epsilon_3 Y_v e_{nz} e^{j\beta_n L} = u^2\beta_v Y_v e_{oz} e^{-j\beta_o L}$$

$$(53b)\quad p(52c) - k(52d) = \sum_{n=-2}^{n=2}{}_{n\neq 0} \; -u^2\beta_v \frac{Y_n}{Y_v} e_{nz} e^{j\beta_n L} = -u^2 \frac{\beta_v}{Y_v} h_{oz} e^{-j\beta_o L}$$

Any four equations or four linear combinations of the above set (52) constitute a set of independent equations. Thus using (53a) with (52e) or (52b) gives

$$(54a)\quad \sum_{n=-2}^{n=2}{}_{n\neq 0} \; \beta_n Y_n e_{nz} e^{j\beta_n L} = -\beta_o h_{oz} e^{-j\beta_o L}$$

Similarly using (53b) with (52e) or (52d) yields

$$(54b)\quad \sum_{n=-2}^{n=2}{}_{n\neq 0} \; \frac{\beta_c^2}{\beta_n} e_{nz} e^{j\beta_n L} = -\beta_o e_{oz} e^{-j\beta_o L}$$

Equations (52a) , (52b) , (53a) , and (54b) form one of the possible sets of four equations at the boundary z = L. This set will be used in the solution of the problem and is repeated below for convenience:

$$(55a)\quad \sum_{n=-2}^{n=2}{}_{n\neq 0} \; \frac{1}{\beta_n} e^{j\beta_n L} e_{nz} + \frac{\beta_o}{\beta_c^2} e^{-j\beta_o L} e_{oz} = 0$$

$$(55b)\quad \sum_{n=-2}^{n=2}{}_{n\neq 0} \; \beta_n Y_n e^{j\beta_n L} e_{nz} + \beta_o e^{-j\beta_o L} h_{oz} = 0$$

$$(55c)\quad \sum_{n=-2}^{n=2}{}_{n\neq 0} \; e^{j\beta_n L} e_{nz} - \frac{e^{-j\beta_o L}}{\epsilon_3} e_{oz} = 0$$

$$(55d)\quad \sum_{n=-2}^{n=2}{}_{n\neq 0} \; Y_n e^{j\beta_n L} e_{nz} - e^{-j\beta_o L} h_{oz} = 0$$

Boundary Conditions at z = 0

Let the Fourier transforms of the applied electric fields be

$$(56a)\quad \bar{\bar{E}}_x^a = \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_x^a(x,y,0)\, e^{-jkx} e^{-jpy}\, dk\, dp$$

$$(56b)\quad \bar{\bar{E}}_y^a = \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_y^a(x,y,0)\, e^{-jkx} e^{-jpy}\, dk\, dp$$

Equating the transforms of the transverse electric field components, $\bar{\bar{E}}_x$ and $\bar{\bar{E}}_y$, from (41) to (56a) and (56b) respectively, gives after setting z = 0

$$(57a)\quad \sum_{n=-2}^{n=2}{}_{n\neq 0} \; \left[k\beta_c \frac{\beta_c}{\beta_n} - p\beta_v \frac{Y_n}{Y_v}\right] e_{nz} + u^2 \bar{\bar{E}}_x^a = 0$$

$$(57b)\quad \sum_{n=-2}^{n=2}{}_{n\neq 0} \; \left[p\beta_c \frac{\beta_c}{\beta_n} + k\beta_v \frac{Y_n}{Y_v}\right] e_{nz} + u^2 \bar{\bar{E}}_y^a = 0$$

The set (57) is equivalent to the following set obtained by a linear combination of the above equations

$$(58a)\qquad \sum_{\substack{n=-2\\ n\neq 0}}^{n=2} \frac{\beta_c^2}{\beta_n}\, e_{nz} + (p\,\bar{\bar{E}}_y^a + k\,\bar{\bar{E}}_x^a) = 0$$

$$(58b)\qquad \sum_{\substack{n=-2\\ n\neq 0}}^{n=2} \beta_v \frac{Y_n}{Y_v}\, e_{nz} + (k\,\bar{\bar{E}}_y^a - p\,\bar{\bar{E}}_x^a) = 0$$

The set of six equations (55a,b,c,d) and (58a,b) will yield the transforms of the field components in both regions in terms of $\bar{\bar{E}}_x^a$ and $\bar{\bar{E}}_y^a$. For convenience, the set is repeated below in matrix form

$$(59)\qquad \begin{vmatrix} \frac{e^{i\beta_{-2}L}}{\beta_{-2}} & \frac{e^{i\beta_{-1}L}}{\beta_{-1}} & \frac{e^{i\beta_1 L}}{\beta_1} & \frac{e^{i\beta_2 L}}{\beta_2} & \frac{\beta_o}{\beta_c^2}e^{-i\beta_o L} & 0 \\ \beta_{-2}Y_{-2}e^{i\beta_{-2}L} & \beta_{-1}Y_{-1}e^{i\beta_{-1}L} & \beta_1 Y_1 e^{i\beta_1 L} & \beta_2 Y_2 e^{i\beta_2 L} & 0 & \beta_o e^{-i\beta_o L} \\ e^{i\beta_{-2}L} & e^{i\beta_{-1}L} & e^{i\beta_1 L} & e^{i\beta_2 L} & -\frac{e^{-i\beta_o L}}{\epsilon_3} & 0 \\ Y_{-2}e^{i\beta_{-2}L} & Y_{-1}e^{i\beta_{-1}L} & Y_1 e^{i\beta_1 L} & Y_2 e^{i\beta_2 L} & 0 & -e^{-i\beta_o L} \\ \frac{\beta_c^2}{\beta_{-2}} & \frac{\beta_c^2}{\beta_{-1}} & \frac{\beta_c^2}{\beta_1} & \frac{\beta_c^2}{\beta_2} & 0 & 0 \\ \beta_v\frac{Y_2}{Y_v} & \beta_v\frac{Y_{-1}}{Y_v} & \beta_v\frac{Y_1}{Y_v} & \beta_v\frac{Y_2}{Y_v} & 0 & 0 \end{vmatrix} \begin{vmatrix} e_{-2z} \\ e_{-1z} \\ e_{1z} \\ e_{2z} \\ e_{oz} \\ h_{oz} \end{vmatrix} = \begin{vmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -(p\,\bar{\bar{E}}_y^a + k\,\bar{\bar{E}}_x^a) \\ -(k\,\bar{\bar{E}}_y^a - p\,\bar{\bar{E}}_x^a) \end{vmatrix}$$

Solution of the Far-Field Components

Let Δ be the determinant of the square matrix (59). $\Delta_{e_{oz}}$ is the determinant of the same matrix after substituting for its fifth column the column matrix of the RHS of (59). Similarly $\Delta_{h_{oz}}$ is obtained after substituting the same column matrix into the sixth column of the original matrix.

The solution for Δ, $\Delta_{e_{oz}}$ and $\Delta_{h_{oz}}$ is tedious but straightforward (an understatement!) the results are given below:

$$(60)\qquad \Delta = -\frac{4}{\epsilon_3}\frac{\beta_v}{Y_v}\frac{e^{-i2\beta_o L}}{\beta_1\beta_2\beta_c^2}(\beta_1 Y_1 - \beta_2 Y_2)\left\{\beta_o(\beta_1 Y_1 - \beta_2 Y_2)\,G_1 G_2 + F^2(\beta_1 Y_1 c_1 G_2 - \beta_2 Y_2 c_2 G_1)\right\}$$

$$(61)\qquad \Delta_{e_{oz}} = \frac{4e^{-i\beta_o L}}{\beta_1\beta_2}(\beta_1 Y_1 - \beta_2 Y_2)\left\{-\frac{\beta_v}{Y_v}V_x\left[(\beta_1 Y_1 c_1 - \beta_2 Y_2 c_2) + i\beta_o(s_1 Y_1 - s_2 Y_2)\right] + \beta_c^2 V_y\left[(c_1 - c_2) + i\beta_o\left(\frac{s_1}{\beta_1} - \frac{s_2}{\beta_2}\right)\right]\right.$$

$$(62)\qquad \Delta_{h_{oz}} = \frac{4}{\epsilon_3}\frac{e^{-i\beta_o L}}{\beta_1\beta_2}(\beta_1 Y_1 - \beta_2 Y_2)\left\{\frac{\beta_v}{Y_v}V_x\beta_1 Y_1\beta_2 Y_2\left[\epsilon_3\frac{\beta_o}{\beta_c^2}(c_1 - c_2) + i\left(\frac{s_1}{\beta_1} - \frac{s_2}{\beta_2}\right)\right] + \beta_c^2 V_y\left[\epsilon_3\frac{\beta_o}{\beta_c^2}(\beta_1 Y_1 c_2 - \beta_2 Y_2 c_1) + i(\beta_1 Y_1\frac{s_2}{\beta_2} - \beta_2 Y_2\frac{s_1}{\beta_1})\right]\right.$$

e_{oz} and h_{oz} are given by

$$(63a)\qquad e_{oz} = b_c^2\epsilon_3 e^{ib_o l}\,\frac{\frac{V_x}{\beta_v}\left\{(b_1 y_1 c_1 - b_2 y_2 c_2) + ib_o(s_1 y_1 - s_2 y_2)\right\} - \frac{V_y}{\beta_v}b_c^2\left\{(c_1 - c_2) + ib_o\left(\frac{s_1}{b_1} - \frac{s_2}{b_2}\right)\right\}}{b_o(b_1 y_1 - b_2 y_2)\,g_1 g_2 + f^2(b_1 y_1 c_1 g_2 - b_2 y_2 c_2 g_1)} = \frac{\Delta_{e_{oz}}}{\Delta}$$

$$(63b)\qquad h_{oz} = -Y_v b_c^2 e^{ib_o l}\,\frac{\frac{V_x}{\beta_v}b_1 y_1 b_2 y_2\left[\epsilon_3\frac{b_o}{b_c^2}(c_1 - c_2) + i\left(\frac{s_1}{b_1} - \frac{s_2}{b_2}\right)\right] + \frac{V_y}{\beta_v}b_c^2\left[\epsilon_3\frac{b_o}{b_c^2}(b_1 y_1 - b_2 y_2) + i\left(b_1 y_1\frac{s_2}{b_2} - b_2 y_2\frac{s_1}{b_1}\right)\right]}{b_o(b_1 y_1 - b_2 y_2)\,g_1 g_2 + f^2(b_1 y_1 c_1 g_2 - b_2 y_2 c_2 g_1)} = \frac{\Delta_{h_{oz}}}{\Delta}$$

The following notation has been introduced in the above equations:

(64a) $V_x = (p\,\bar{\bar{E}}_y^a + k\,\bar{\bar{E}}_x^a)$

(64b) $V_y = (k\,\bar{\bar{E}}_y^a - p\,\bar{\bar{E}}_x^a)$

(65a) $s_n = \sin \beta_n L$

(65b) $c_n = \cos \beta_n L$

(66) $\beta_c = b_c \beta_v$

(67a) $\beta_o = b_o \beta_v$

(67b) $l = \beta_v L = \dfrac{2\pi L}{\lambda_v}$

(68) $Y_n = y_n Y_v$

(69a) $G_n = \epsilon_3 \beta_o c_n + j \dfrac{s_n}{\beta_n} \beta_c^2 = g_n \beta_v$

(69b) $F^2 = \beta_c^2 - \epsilon_3 \beta_o^2 = f^2 \beta_v^2$

The far-field components are given by the inverse transforms of e_{oz} and h_{oz}. Thus

(70) $$H_{oz} = \int_{-\infty}^{\infty}\int h_{oz}(k,p)\, e^{jkx} e^{jpy} e^{-j\beta_o z}\, dk\, dp$$

The evaluation of the inverse transform is readily obtained by the method of stationary phase. The method is described elsewhere[3] and a physical interpretation for it has been given in a previous paper.[2] Because (70) is a double Fourier transform, the method of stationary phase must be applied twice in succession. This is best done with the following change of variables.

(71) $$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = r \begin{bmatrix} c_\phi s_\theta \\ s_\phi s_\theta \\ c_\theta \end{bmatrix} \qquad (c_\phi = \cos\phi,\ s_\theta = \sin\theta,\ \text{etc.})$$

(72) $$\begin{bmatrix} k \\ p \end{bmatrix} = u \begin{bmatrix} c_v \\ s_v \end{bmatrix} \qquad (c_v = \cos v,\ s_v = \sin v)$$

Using (48), (71) into (70) and integrating over the u-v plane defined by (72) gives

(73) $$H_{oz}(x,y,z) = \int_0^{\infty} e^{-jr\, c_\theta \sqrt{\beta_v^2 - u^2}}\, u\,du \int_0^{2\pi} h_{oz}(u,v)\, e^{jr\{u s_\theta (c_\phi c_v + s_\phi s_v)\}}\, dv$$

Let,

(74) $X_1(u,v) = u s_\theta (c_\phi c_v + s_\phi s_v)$

Define,

(74a) $\dfrac{dX}{dv} = X_{1_v}$

(74b) $\dfrac{d^2X}{dv^2} = X_{1_{vv}}$

(74c) $v = v_s$

is the value of v that makes $X_{1_v} = 0$.

As shown elsewhere[2]

(75) $$\int_0^{2\pi} h_{oz}(u,v)\, e^{jr\,X(u,v)}\, dv \approx h_{oz}(u,v_s)\sqrt{\frac{j2\pi}{rX_{1_{vv}}(u,v_s)}}\, e^{jrX(u,v_s)} + \text{high order terms in } \frac{1}{r}$$

Defining

(76) $$X_2(u,v_s) = X_1(u,v_s) - c_\theta \sqrt{\beta_v^2 - u^2}$$

Substituting the RHS of (75) into (73) and integrating according to the rule of (75) gives

(77) $$H(x,y,z) = \frac{j2\pi}{r} \frac{u_s h_{oz}(u_s,v_s)}{\sqrt{X_{1_{vv}}(u_s,v_s)\, X_{2_{uu}}(u_s,v_s)}}\, e^{jrX_2(u_s,v_s)}$$

It is easily shown that

(78a) $$u_s = \beta_v s_\theta$$

(78b) $$v_s = \phi + \pi$$

(78c) $$\beta_{os} = \beta_o(u_s,v_s) = \beta_v c_\theta$$

(79a) $$X_{1_{vv}}(u_s,v_s) = \beta_v s_\theta^2$$

(79b) $$X_{2_{uu}}(u_s,v_s) = \frac{1}{\beta_v c_\theta^2}$$

which yields the final results:

(80) $$\begin{vmatrix} E_{oz}(x,y,z) \\ H_{oz}(x,y,z) \end{vmatrix} = \frac{(2\pi)^2}{\lambda_v r} c_\theta\, e^{-j(\beta_v r - \frac{\pi}{2})} \begin{vmatrix} e_{oz}(k_s,p_s) \\ h_{oz}(k_s,p_s) \end{vmatrix}$$

(81) $$\begin{vmatrix} k_s \\ p_s \end{vmatrix} = -\beta_v s_\theta \begin{vmatrix} c_\phi \\ s_\phi \end{vmatrix}$$

The far-field components are obtained by means of (63) and (81) into (80).

Radiation Power Density

The radiated power density is given by Poynting vector

(82) $$\mathcal{W} = \frac{1}{2}\mathcal{R}(\mathbb{E} \times \mathbb{H}^*)$$

"$\mathcal{R}$" stands for the "real part of" and "*" for conjugate.

Expanding (82) in rectangular coordinates, using expression (51) in conjunction with (80) for the rectangular components and changing to spherical coordinates r,θ,z yields after a few manipulations

(83) $$\mathcal{W}_r = \frac{1}{2}\left[\frac{(2\pi)^2 c_\theta}{r\lambda_v}\right]^2 \left(Y_v \left|\frac{e_{oz}(k_s,p_s)}{s_\theta}\right|^2 + Z_v \left|\frac{h_{oz}(k_s,p_s)}{s_\theta}\right|^2\right)$$

where 1r is the unit vector in the r-direction and

(84) $$Y_v = \frac{1}{Z_v}$$

Radiation from a Rectangular Slot Excited in the TE_{10} Mode

As an example of the applications of the results (80) and (81) to the evaluation of radiation fields, consider a rectangular slot in which the source distribution is forced to be sinusoidal. The long dimension of the slot extends from

$y = -\frac{b}{2}$ to $y = \frac{b}{2}$ and the short dimension from $x = -\frac{a}{2}$ to $x = \frac{a}{2}$. The exciting field is

(85a) $$E_x^a = \frac{V}{a} \cos \frac{\pi}{b} y$$

(85b) $$E_y^a = 0$$

The Fourier transform of the applied fields are

(86a) $$\bar{\bar{E}}_x^a = \frac{V}{2\pi} \frac{\sin \frac{ka}{2}}{\frac{ka}{2}} \frac{b \cos \frac{pb}{2}}{\pi^2 - (pb)^2}$$

(86b) $$\bar{\bar{E}}_y^a = 0$$

Substitution of (81) into (86a), using (64a) gives V_x and V_y

(87a) $$V_{x_s} = -\frac{V}{2\pi} \left[\frac{\sin (\beta_v s_\theta c_\phi \frac{a}{2})}{\beta_v s_\theta c_\phi \frac{a}{2}}\right] \left[\frac{b \cos(\beta_v s_\theta s_\phi \frac{b}{2})}{\pi^2 - (\beta_v s_\theta s_\phi b)^2}\right] (\beta_v s_\theta c_\phi)$$

(87b) $$V_{y_s} = \frac{V}{2\pi} \left[\frac{\sin (\beta_v s_\theta c_\phi \frac{a}{2})}{\beta_v s_\theta c_\phi \frac{a}{2}}\right] \left[\frac{b \cos (\beta_v s_\theta s_\phi \frac{b}{2})}{\pi^2 - (\beta_v s_\theta s_\phi b)^2}\right] (\beta_v s_\theta s_\phi)$$

The radiation fields are then obtained by substituting (81) and (87) into (63). The radiated power follows from substitution of (63) into (83).

For a narrow slot,

(88) $$\frac{k_s a}{2} << 1 ;$$

then in the $\phi = 0$ plane,

(89) $$V_{x_s} \approx -\beta_v s_\theta \frac{V}{2\pi} \frac{b}{\pi^2}$$

It will be seen in the next section that the transform of an applied constant E_x^a-field across an infinitely long narrow slot is identical with (89) except for one constant factor $\frac{b}{\pi^2}$; hence the patterns of the infinite slot and the rectangular slot are similar in the $\phi = 0$ plane.

Radiation from an Infinitely Long Narrow Slot

For this case

(90) $$\bar{\bar{E}}_y = 0$$

(91) $$p = 0$$

(92) $$\bar{\bar{E}}_x \rightarrow \bar{E}_x = \frac{V}{2\pi}$$ (single bar indicates transform with respect to x only)

Substituting (90), (91), (92) into (64) gives

(93a) $$V_x = k \frac{V}{2\pi}$$ (note that V_x in this case has no longer the dimension of volts)

(93b) $$V_y = 0$$

Substituting (93) in (63) yields the transforms of the radiation fields

(94a) $$e_{oz} = e^{jb_o l} \frac{V}{2\pi} \frac{k}{\beta_v} b_c^2 \epsilon_3 \frac{(b_1 y_1 c_1 - b_2 y_2 c_2) + j b_o (s_1 y_1 - s_2 y_2)}{b_o (b_1 y_1 - b_2 y_2) g_1 g_2 + f^2 (b_1 y_1 c_1 g_2 - b_2 y_2 c_2 g_1)}$$

(94b) $$h_{oz} = -e^{jb_o l} Y_v \frac{V}{2\pi} \frac{k}{\beta_v} \quad b_1 y_1 b_2 y_2 \frac{e_3 b_o (c_1 - c_2) + j b_c^2 (\frac{s_1}{b_1} - \frac{s_2}{b_2})}{b_o (b_1 y_1 - b_2 y_2) g_1 g_2 + f^2 (b_1 y_1 c_1 g_2 - b_2 y_2 c_2 g_1)}$$

Evaluation of $E_{oz}(x,y,z)$ and $H_{oz}(x,y,z)$ by the stationary phase method shows the following changes in equation (80) :

(95a) $$r \longrightarrow \rho = \sqrt{x^2 + z^2}$$

(95b) $$\frac{(2\pi)^2}{r} \longrightarrow \sqrt{\frac{(2\pi)^2}{\rho}}$$

(95c) $$e^{j\frac{\pi}{2}} \longrightarrow e^{j\frac{\pi}{4}}$$

thus giving

(96) $$\begin{vmatrix} E_{oz}(x,z) \\ H_{oz}(x,z) \end{vmatrix} = \sqrt{\frac{(2\pi)^2}{\lambda_v \rho}} \; c_\theta \; e^{-j(\beta_v \rho - \frac{\pi}{4})} \begin{vmatrix} e_{oz}(k_s) \\ h_{oz}(k_s) \end{vmatrix}$$

where

(97) $$k_s = -\beta_v \sin\theta$$

Note that in this 2-dimensional problem (where the geometry is dependent of y-direction), k_s is obtained from its 3-dimensional counterpart by simply making $\phi = 0$. Thus it is seen that the radiation from rectangular apertures (3-dimensional problem) has a pattern in the $\phi = 0$ plane similar to the pattern obtained from the infinitely long and narrow slot.

The radiated Poynting vector becomes after substituting (94) and (97) in (83)

(98) $$\mathcal{W} = \frac{1_\rho}{2} \left[\frac{(2\pi)^2}{\rho\lambda_v}\right] c_\theta^2 \left(Y_v \left|\frac{e_{oz}(k_s)}{s_\theta}\right|^2 + Z_v \left|\frac{h_{oz}(k_s)}{s_\theta}\right|^2\right)$$

1_ρ is the unit vector in the ρ-direction in ρ,θ,z system of coordinates (Figure 2). This expression can be reduced to

(99) $$\mathcal{W} = 1_\rho \pi_\rho = \frac{1_\rho}{2} \frac{(2\pi)^2}{\rho\lambda_v} \left(Y_v \left|e_{oy}(k_s)\right|^2 + Z_v \left|h_{oy}(k_s)\right|^2\right) c_\theta^2$$

by means of (51) in the case of $p = 0$. The first term in (99) is the contributions from TE-modes and the second term is the contribution from TM-modes.

Calculations of Radiation Patterns for the Infinitely Long Narrow Slot Antenna

To facilitate the radiation pattern calculations, the bracketed quantities in (94) are defined as normalized radiation intensities N_h and N_e. Thus (94) become

(100a) $$\left|e_{oz}(k_s)\right| = \frac{V}{2\pi} \frac{k_s}{\beta_v} \left|N_h(k_s)\right|$$

(100b) $$\left|h_{oz}(k_s)\right| = Y_v \frac{V}{2\pi} \frac{k_s}{\beta_v} \left|N_e(k_s)\right|$$

Expressing $e_{oy}(k_s)$ and $h_{oy}(k_s)$ in terms of e_{oz} and h_{oz} through (51) for the case $p = 0$ and substituting into (99) gives

(101) $$\mathcal{W}_\rho = \frac{1}{2} \frac{Y_v V^2}{\rho\lambda_v} \left(\left|N_h(k_s)\right|^2 + \left|N_e(k_s)\right|^2\right) c_\theta^2$$

A normalized gain function $G(\theta)$ is defined as

(102) $$G(\theta) = \frac{\pi_\rho}{\frac{1}{2}\frac{Y_v V^2}{\rho \lambda_v}}$$

Thus,

(103) $$G(\theta) = \left(|N_h|^2 + |N_e|^2\right) c_\theta^2$$

evaluated for $k_s = -\beta_v s_\theta$.

The gain function defined by (103) although quite complicated reduces to a simple expression under the following restrictions

(104a) $$\nu = 0$$

(104b) $$\omega^2 << \omega_b^2 << \omega_p^2$$

(104c) $$\frac{\omega_p}{\sqrt{\omega\omega_b}} 2\pi \frac{L}{\lambda_v} >> 1$$

Under these restrictions (103) becomes

(105) $$G(\theta) \approx \frac{(1 + c_\theta^2) c_\theta^2}{\left|(1 + c_\theta^2)\left(\frac{s + c}{b}\right) - j2c_\theta\left(c - \frac{s}{b^2}\right)\right|^2}$$

where the various symbols in the above expressions are given by

(106a) $$b = \frac{\omega_p}{\sqrt{\omega\omega_b}}$$

(106b) $$s = \sin bl$$

(106c) $$c = \cos bl$$

(106d) $$l = \frac{L}{\lambda_v} 2\pi$$

$G(\theta)$ is plotted in Figures 3, 4, and 5 for several values of c and s. In Figure 3, $G(\theta) = 1$ corresponds to the radiation pattern of an infinitely long narrow slot in the absence of plasma. Figures 3 and 4 show that the power radiated by the slot antenna over a broad angle is not significantly reduced in the presence of a lossless gyrotropic plasma slab. Figure 5 shows that the radiated power is significantly increased (as much as 27 db) for some critical combination of slab width and plasma, cyclotron and signal frequencies. This occurs at

(107) $$\frac{\omega_p}{\sqrt{\omega\omega_b}} 2\pi \frac{L}{\lambda_v} = \left(N + \frac{1}{2}\right)\frac{\pi}{2} \qquad N = \text{positive integer} >> 1$$

This increase in radiated power over the free space antenna pattern can be explained on the basis that $G(\theta)$ is derived for a constant applied voltage V. In the absence of plasma, the input impedance Z_{in} of an infinitely long narrow slot is highly reactive (capacitive); hence for a given applied voltage V, the radiated power P_r is given by:

(108) $$P_r = \frac{V^2}{|Z_{in}|^2} R_r$$

where R_r is the radiation resistance.

The presence of a gyrotropic plasma slab over the antenna although changing both the radiation resistance and the capacitive reactance may decrease the reactance in such a way that the ratio $\frac{R_r}{|Z_{in}|^2}$ increases over its free space value. The net result is then an increase in radiated power since V is fixed. Hence for the proper combination of plasma parameters, a gyrotropic slab with a static magnetic field perpendicular to the slot may improve the matching between a long narrow slot radiator and its free space environment.

According to the stationary phase analysis[2], the radiation field at one point is determined essentially by a plane wave originating at the slot and arriving along the radial direction θ locating that point.

As θ departs from 0° the radiation pattern should approach a null rapidly; this is due to the fact that as θ increases the plane wave along this direction encounters increasing attenuation due to a lengthening of its path through the plasma and approaching a mode of propagation transverse to the magnetic field.[4] However, because of the assumption (104), in particular $\omega^2 << \omega_b^2$, it turns out that the component of the cyclotron frequency along the direction of propagation, namely $\omega_b \cos\theta$, is higher than ω for angles below 85°, and plane wave propagation along this angle can still be considered to be in the quasi-longitudinal mode.

Conclusions

It is shown that the presence of a gyrotropic lossless plasma slab covering a rectangular slot radiator does not alter significantly the radiation pattern in the $\emptyset = 0$ plane when the static magnetic field is perpendicular to the slot. The plasma covered antenna remains omnidirectional and the radiated power is not significantly reduced when the cyclotron frequency exceeds the operating frequency by one order of magnitude ($\omega_b^2 >> \omega^2$).

It is also found that the matching between the antenna and its surrounding is sharply improved, if the antenna is covered with such a slab, at critical combinations of the plasma slab parameters and the operating frequency; this finding may be useful for plasma diagnostics, in order to determine L, ω_p, and ν by varying B_o and ω.

Acknowledgments

This work was made possible through joint cooperation and sponsorship from the Hallicrafters Company and the Physical Science Laboratory of New Mexico State University under Bell Telephone Laboratories Contract No. 27495-42.

References

1. This equation had also been derived by E. Arbel, "Radiation from a Point Source in an Anisotropic Medium," Research Report PIBMRI-861-60, 2 November, 1960, Polytechnic Institute of Brooklyn, Microwave Research Institute, Brooklyn, New York.

2. Hodara and Cohn, "Radiation Characteristics of Long Narrow Slot Antennas in Lossy Anisotropic Plasma with a Static Magnetic Field Parallel to the Slot," Proceedings of the 17th Annual National Electronics Conference, Chicago, Illinois; October, 1961.

3. C. M. Knop, "The Radiation Fields from a Circumferential Slot on a Metal Cylinder Coated with a Lossy Dielectric," IRE PGAP; November, 1961.

4. H. Hodara, "The Elimination of the Re-entry Radio Blackout by Means of a Static Magnetic Field," Proceedings of the IRE, December, 1961.

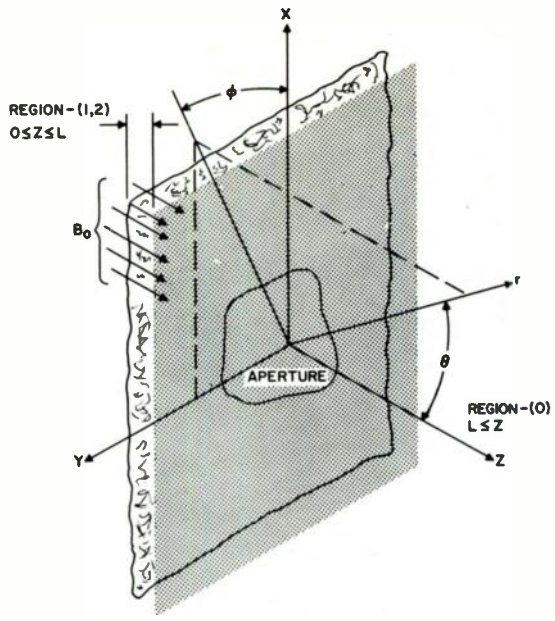


Figure 1.

ARBITRARY APERTURE SLOT RADIATOR COVERED WITH GYROTROPIC PLASMA (STATIC MAGNETIC FIELD PERPENDICULAR TO PLANE OF APERTURE).

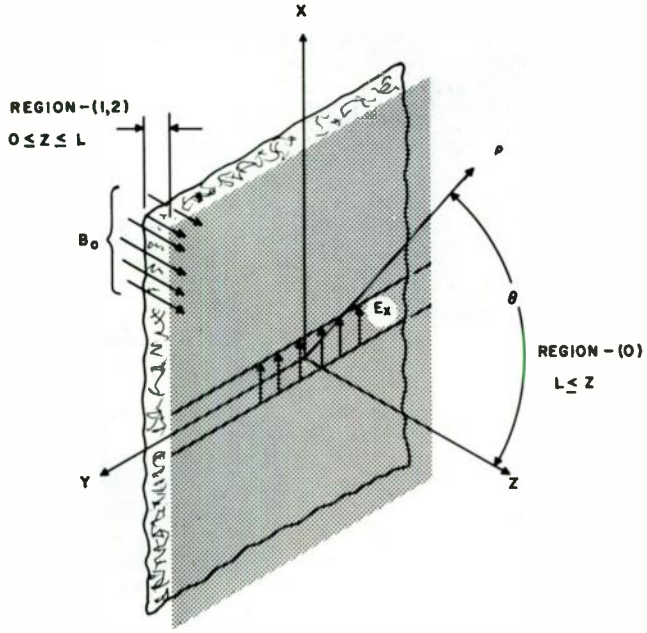


Figure 2.

INFINITE SLOT RADIATOR COVERED WITH GYROTROPIC PLASMA (STATIC MAGNETIC FIELD PERPENDICULAR TO THE SLOT).

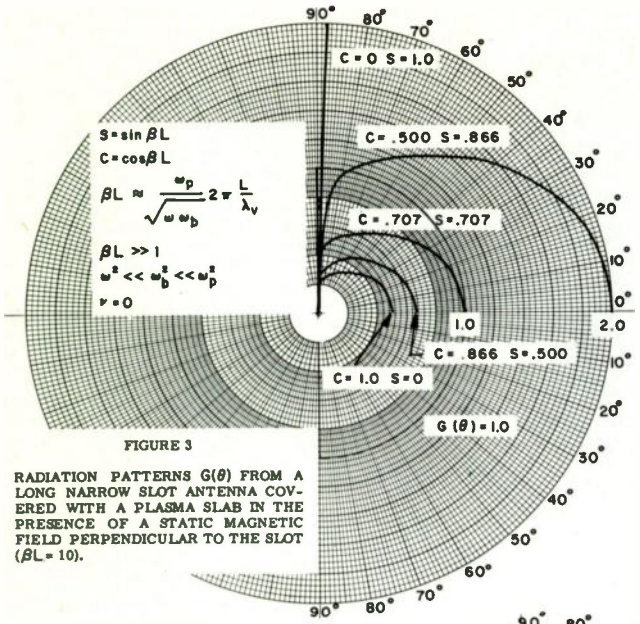


FIGURE 3

RADIATION PATTERNS G(θ) FROM A LONG NARROW SLOT ANTENNA COVERED WITH A PLASMA SLAB IN THE PRESENCE OF A STATIC MAGNETIC FIELD PERPENDICULAR TO THE SLOT (βL = 10).

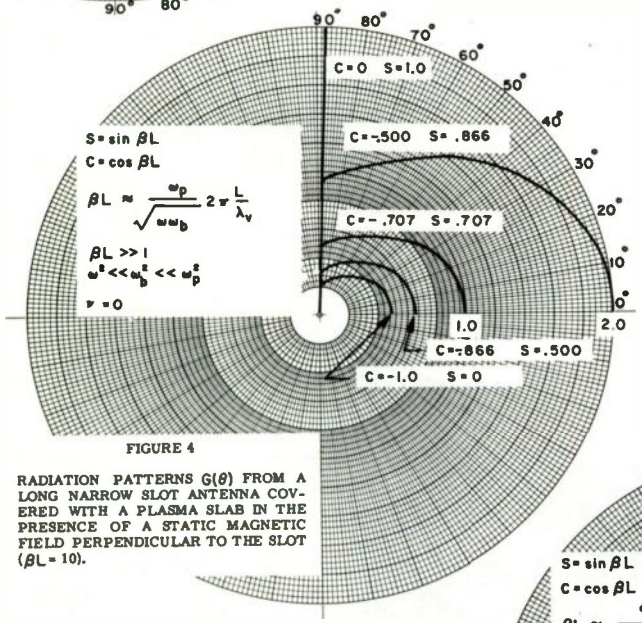


FIGURE 4

RADIATION PATTERNS G(θ) FROM A LONG NARROW SLOT ANTENNA COVERED WITH A PLASMA SLAB IN THE PRESENCE OF A STATIC MAGNETIC FIELD PERPENDICULAR TO THE SLOT (βL = 10).

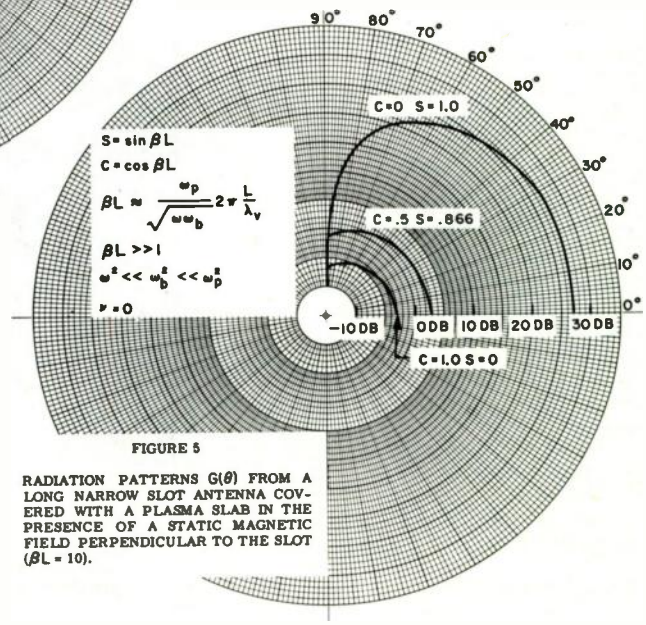


FIGURE 5

RADIATION PATTERNS G(θ) FROM A LONG NARROW SLOT ANTENNA COVERED WITH A PLASMA SLAB IN THE PRESENCE OF A STATIC MAGNETIC FIELD PERPENDICULAR TO THE SLOT (βL = 10).

TRAPPED WAVES IN VARYING DIELECTRIC MEDIA

John E. Chin
Air Armament Division
Sperry Gyroscope Company
Division of Sperry Rand Corporation
Great Neck, L.I., New York

In this paper the propagation characteristics of a variable dielectric medium are investigated. Specifically, the medium is an infinite dielectric slab imbedded in free space but whose dielectric constant merges continuously into the free space value at the boundaries while attaining a maximum value at the center of the slab. In solving Maxwell's Equations in such a medium, it is shown that certain solutions of the fields in the medium reveal the existence of trapped modes within the slab. The solutions are expressible in terms of hypergeometric functions which simplify to hyperbolic functions for the lowest ordered modes. The governing conditions which establish the existence of these modes are also used to determine the appropriate cutoff frequency of each mode.

I. Introduction

It is fairly well known that an infinite slab of dielectric of a finite thickness will guide electromagnetic energy. The guiding effect or trapping of these waves within the slab is due to the total reflection of the waves at the free-space dielectric interface. For a fixed thickness, there exists a frequency below which total reflection at the interface is no longer possible, so that energy is transmitted across the interface. Above this frequency, the energy is considered trapped within the slab.

If the slab of dielectric has a variable dielectric constant such that there is no distinguishable boundary between free-space and the dielectric, it is not quite as obvious that energy will be trapped. In a variable medium, however, there will be a continuous bending of the rays, and it is this gradual refraction which plays the same role as total reflection does at the interface of a distinguishable slab.

In this paper, the trapping of electromagnetic energy by a smoothly varying dielectric slab is investigated by actually solving Maxwell's equations in the medium. The particular medium studied has a dielectric variation based upon the Epstein-Eckart potential "well" that was introduced in the mid-1930's because Schroedinger's Wave Equation could then be solved in terms of known functions.[1]

It will be shows that, under certain conditions, solutions of the wave equation in such a medium can be obtained which possess the characteristics of trapped waves within the slab. These solutions also show that these trapped wave are modal in character with each trapped mode having a particular cutoff frequency.

In part II, the wave equation in the medium is derived from Maxwell's equations and is then transformed into a form for which solutions are known to exist. This form is the hypergeometric equation whose solutions are hypergeometric functions.

In part III, certain particular solutions of the hypergeometric equation are shown to be representative of trapped waves if certain restrictions are imposed upon the parameters of these solutions. These restrictions form the basis for determining the cutoff frequencies of the modes and also permit calculation of the phase velocities of the trapped waves.

In the last section, the first three modes are derived and the field equations are reduced to hyperbolic functions.

II. Derivation of the Wave Equation

Maxwell's equations (in m.k.s. units) in a medium are:

$$\nabla \times \bar{H} = \frac{\partial \bar{D}}{\partial t} = K_e \epsilon_o \frac{\partial \bar{E}}{\partial t} \quad (1a) \qquad \nabla \cdot \bar{D} = \nabla \cdot K_e \epsilon_o \bar{E} = 0 \quad (1c)$$

$$\nabla \times \bar{E} = -\frac{\partial \bar{B}}{\partial t} = -K_m \mu_o \frac{\partial \bar{H}}{\partial t} \quad (1b) \qquad \nabla \cdot \bar{B} = 0 \quad (1d)$$

where $\bar{H}$, $\bar{B}$, $\bar{D}$, and $\bar{E}$ are the usual field vectors; ϵ_o and μ_o are, respectively, the permittivity and permeability of free space; and K_e and K_m are, respectively, the dielectric constant and relative permeability of the medium.

Operating on Eq. (1b), and making use of Eq. (1a) and (1c), we obtain the vector wave equation:

$$-\nabla\left(\frac{1}{K_e}\nabla K_e \cdot \bar{E}\right) - \nabla^2\bar{E} = -K_m K_e \epsilon_o \mu_o \frac{\partial^2 \bar{E}}{\partial t^2} \qquad (2)$$

In order to simplify the problem, assume that K_e is a function of one variable ($K_e = K_e(x)$) and that waves are polarized in the plane perpendicular to the x-direction. Under these conditions:

$$\nabla K_e \cdot \bar{E} = 0$$

so that Eq. (2) becomes:

$$\nabla^2\bar{E} - K_m K_e \epsilon_o \mu_o \frac{\partial^2 \bar{E}}{\partial t^2} = 0 \qquad (3)$$

If $\bar{E}$ is assumed to have a harmonic time variation, Eq. (3) becomes:

$$\nabla^2\bar{E} + K_m K_e k_o^2 \bar{E} = 0 \qquad (4)$$

where use has been made of:

$\epsilon_o\mu_o = \frac{1}{C^2}$, $k_o = \frac{\omega}{C} = \frac{2\pi}{\lambda_o}$, and $k_m = 1$ for most materials,

and where C is the velocity of light, and λ_o is the free space wavelength of the wave. For Cartesan coordinates, there is no loss in generality if the y-axis is chosen such that it is coincident with the vector $\bar{E}$. Then $\bar{E} = jEy(x,z)$, where j is unit vector in y-direction.

If:

$$Ey = \Omega(z)\,\Phi(x) \qquad (5)$$

Eq. (4) results in two equations upon separation of the variables:

$$\frac{1}{\Omega}\frac{d^2\Omega}{dz^2} = -k_z^2 \text{ and } \frac{1}{\Phi}\frac{d^2\Phi}{dx^2} + k_o^2 K_e = k_z^2 \qquad (6)$$

where k_z^2 is the separation constant.

The first of Eq. (6) yields immediately:

$$\Omega(z) = e^{jk_z z} \qquad (7)$$

while the second of Eq. (6) is:

$$\frac{d^2\Phi}{dx^2} + k_o^2 K_e \Phi = k_z^2 \Phi \qquad (8)$$

At this point, the form of the dielectric variation is identified to be:

$$K_e(x) = 1 + K\,\mathrm{sech}^2\frac{x}{L} \qquad (9)$$

As illustrated in Fig. 1

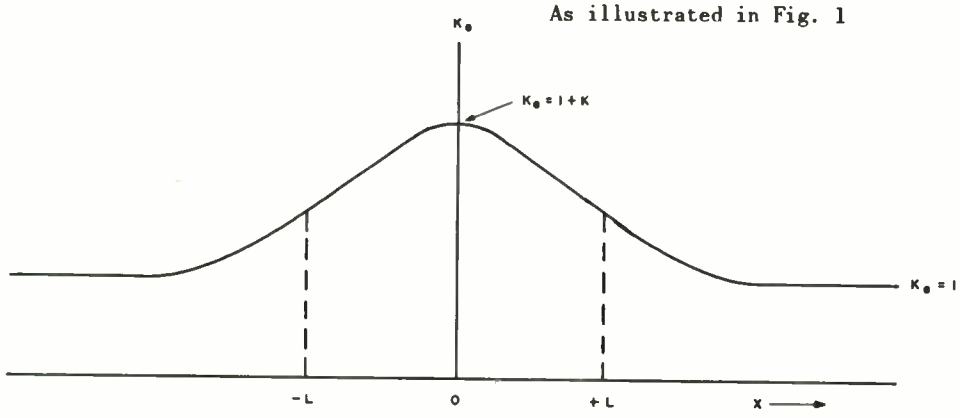


FIG.1 PLOT OF K_e AS FUNCTION OF X

If the variable x is transformed by:

$$\xi = -e^{-\frac{2x}{L}}$$

Eq.(8) becomes:

$$\xi^2 \frac{d^2\Phi}{d\xi^2} + \xi \frac{d\Phi}{d\xi} + \left[- \frac{k_o^2 L^2 K \xi}{4(1-\xi)^2} + \frac{k_o^2 L^2}{4} - \frac{k_z^2 L^2}{4} \right] \Phi = 0 \tag{10}$$

In this equation, let the dimensionless quantity

$$\frac{L^2 k_o^2}{4} - \frac{L^2 k_z^2}{4} = \lambda \tag{10a}$$

and let

$$\frac{k_o^2 L^2 K}{4} = H$$

Equation (10) is now rewritten:

$$\xi^2 \frac{d^2\Phi}{d\xi^2} + \xi \frac{d\Phi}{d\xi} + \left[\lambda - \frac{H\xi}{(1-\xi)^2} \right] \Phi = 0 \tag{11}$$

If we assume $\Phi = (1-\xi)^{\alpha} \xi^{\beta} f$ and substitute into Eq. (11), we will have:

$$\xi(1-\xi) \frac{d^2 f}{d\xi^2} + [2\beta + 1 - (2\beta + 2\alpha + 1)\xi] \frac{df}{d\xi} + \alpha(\alpha + 2\beta) f = 0 \tag{12}$$

provided the following definitions are made:

$$\beta^2 = -\lambda \text{ and } \alpha = \frac{1}{2} [1 - \sqrt{1 + 4H}] \tag{13}$$

Equation (12) is known as Gauss' Equation or the hypergeometric equation[2], which is:

$$\xi(1-\xi) \frac{d^2 f}{d\xi^2} + [c - (a + b + 1)\xi] \frac{df}{d\xi} - abf = 0 \tag{14}$$

where

$$\begin{aligned} a &= \alpha \\ b &= 2\beta + \alpha \\ c &= 2\beta + 1 \end{aligned} \tag{15}$$

There are 24 particular solutions to this equation. The solutions, however, are not all valid for all values of the variable, ξ. Since the range of interest encompasses all values of x from $-\infty$ to $+\infty$, the solutions which can be used for evaluating the trapped waves must be valid over the range of ξ from $-\infty$ to 0. Forsyth[3] lists all 24 solutions and the range of convergence for each. It is from among these that the solutions corresponding to the trapped waves must be found.

III. Solutions Corresponding to Trapped Waves

For waves that are trapped within the slab, the solution must decay exponentially for large $|x|$. The solutions which have this asymptotic behavior would correspond to the trapping of energy within the slab with no energy transfer in the x-direction. Hence the problem is to find, among the 24 particular

solutions to Eq. (14), one that will decay exponentially for large $|x|$ and is valid over the range from $-\infty < x < \infty$ or $-\infty < \xi < 0$.

A solution to Eq. (14) which is valid for $-\infty < \xi < 0$ and $2 < \xi < +\infty$ is:[4]

$$f_4 = (1-\xi)^{-b} F\left[b,\ c-a;\ b+1-a;\ \frac{1}{1-\xi}\right] \tag{16}$$

where F [b, c - a; b + 1 - a; 1/1 - ξ] is a hypergeometric series, defined according to:

$$F[a,\ b;\ c;\ x] = 1 + \frac{ab}{1.c}x + \frac{a(a+1)b(b+1)}{1.2c(c+1)}x^2 + \frac{a(a+1)(a+2)b(b+1)(b+2)}{1.2.3\ c(c+1)(c+2)}x^3 + \cdots \tag{17}$$

Substituting in Eq. (16) the values of a, b, and c as defined in Eq. (15), and recalling that:

$$\Phi = (1-\xi)^{\alpha}\,\xi^{\beta} f \tag{18}$$

we have:

$$\Phi = \xi^{\beta}(1-\xi)^{-2\beta} F\left[2\beta+\alpha,\ 2\beta+1-\alpha;\ 2\beta+1;\ \frac{1}{1-\xi}\right] \tag{19}$$

If this is to be the solution of the trapped waves, it must reduce asymptotically to exponentially decaying functions. At $x = -\infty$, or $\xi = -\infty$ we note that Eq. (19) reduces to:

$$\Phi_{\xi\to-\infty} = \xi^{\beta}(-\xi)^{-2\beta} = e^{-j2\pi\beta}\left[-e^{-\frac{2x}{L}}\right]^{-\beta} \tag{20}$$

since F [a, b; c; 0] = 1.

Equation (20 is an exponentially decreasing function for negative values of x provided β is a positive real quantity, which is possible for $k_z > k_o$, since from Eq. (13):

$$\beta^2 = -\lambda = \frac{L^2 k_z^2}{4} - \frac{L^2 k_o^2}{4} \tag{21}$$

At $x = +\infty$, or $\xi = 0$, Fq. (19) cannot be used because F [a, b; c; 1] does not converge. It can, however, be expressed as the sum of two hypergeometric series that do converge for small ξ. Using the known formula, Eq (16) may be written:

$$f_4 = \frac{\Gamma(1-c)\Gamma(1+b-a)}{\Gamma(1-a)\Gamma(1+b-c)} f_1 - \frac{\Gamma(c)\Gamma(1-c)\Gamma(b+1-a)}{\Gamma(2-c)\Gamma(c-a)\Gamma(b)} e^{j\pi(c-1)} f_5{}^{5} \tag{22}$$

where

$$f_1 = (1-\xi)^{-a} F\left[a,\ c-b;\ c;\ \frac{\xi}{\xi-1}\right]$$

$$f_1 = (1-\xi)^{-\alpha} F\left[\alpha,\ 1-\alpha;\ 2\beta+1;\ \frac{\xi}{\xi-1}\right] \tag{23}$$

and

$$f_5 = \xi^{1-c}\,(1-\xi)^{c-b-1}\,F\left[b+1-c,\ 1-a;\ 2-c;\ \frac{\xi}{\xi-1}\right]$$

$$f_5 = \xi^{-2\beta}\,(1-\xi)^{-\alpha}\,F\left[\alpha,\ 1-\alpha;\ 1-2\beta;\ \frac{\xi}{\xi-1}\right] \tag{24}$$

Hence, Eq. (19) becomes:

$$\Phi = \xi^{\beta}\,(1-\xi)^{\alpha}\,f_4 = \frac{\Gamma(-2\beta)\Gamma(2\beta+1)}{\Gamma(1-\alpha)\Gamma(\alpha)}\,\xi^{\beta}\,F\left[\alpha,\ 1-\alpha;\ 2\beta+1;\ \frac{\xi}{\xi-1}\right]$$

$$-\frac{\Gamma(2\beta+1)\Gamma(-2\beta)\Gamma(2\beta+1)}{\Gamma(1-2\beta)\Gamma(2\beta+1-\alpha)\Gamma(2\beta+\alpha)}\,e^{j2\pi\beta}\,\xi^{-\beta}\,F\left[\alpha,\ 1-\alpha;\ 1-2\beta;\ \frac{\xi}{\xi-1}\right] \tag{25}$$

From relations involving Gamma Function, the coefficients of f_1 and f_5 in Eq. (22) can be reduced to:

$$\frac{\Gamma(-2\beta)\Gamma(2\beta+1)}{\Gamma(1-\alpha)\Gamma(\alpha)} = -\frac{\sin\pi\alpha}{\sin 2\pi\beta} \tag{26}$$

and

$$\frac{\Gamma(2\beta+1)\Gamma(-2\beta)\Gamma(2\beta+1)}{\Gamma(1-2\beta)\Gamma(1+2\beta-\alpha)\Gamma(2\beta+\alpha)} = -\prod_{n=o}^{\infty}\left[1+\frac{\alpha}{2\beta+n}\right]\left[1-\frac{\alpha}{1+2\beta+n}\right] \tag{27}$$

(See Appendix.)

Substituting these into Eq (25), we obtain:

$$\Phi = -\frac{\sin\pi\alpha}{\sin 2\pi\beta}\left[-e^{-\frac{2x}{L}}\right]^{\beta} F\left[\alpha,\ 1-\alpha;\ 2\beta+1;\ \frac{\xi}{\xi-1}\right]$$

$$+\prod_{n=o}^{\infty}\left[1+\frac{\alpha}{2\beta+n}\right]\left[1-\frac{\alpha}{1+2\beta+n}\right] e^{j2\pi\beta}\left[-e^{-\frac{2x}{L}}\right]^{-\beta} F\left[\alpha,\ 1-\alpha;\ 1-2\beta;\ \frac{\xi}{\xi-1}\right] \tag{28}$$

With this form, it is possible to investigate the asymptotic character of Φ as $x\to\infty$ or $\xi\to 0$. Recalling that β is a positive real number as required for an exponential decay at $x\to-\infty$, Eq. (28) shows that as $x\to+\infty$ or $\xi\to 0$, the first term is an exponential decay while the second term is an increasing exponential. Since the trapped waves must decay for large $|x|$, the second term must vanish. By virtue of the infinite product in the coefficient of the second term, it will vanish for

$$\frac{\alpha}{2\beta+n} = -1 \text{ or } \alpha = -2\beta-n,\ n = 0,\ 1,\ 2,\cdots \tag{29}$$

Under these same conditions,

$$\frac{\sin\pi\alpha}{\sin 2\pi\beta} = (-1)^n.$$

Substituting for α and β from Eq. (10a), (13), and (21), Eq. (29) is:

$$\frac{1}{2}\left[1 - \sqrt{1 + k_o^2 L^2 K}\right] = -L\sqrt{k_z^2 - k_o^2} - n \tag{29a}$$

Rewriting this leads to:

$$\sqrt{1 + k_o^2 L^2 K} = 2Lk_o\sqrt{\left(\frac{k_z}{k_o}\right)^2 - 1} + 2n + 1 \tag{30}$$

where $n = 0, 1, 2, \cdots$

Equation (30) thus represents the criterion for trapped waves to exist in the dielectric slab when $k_z > k_o$. When $k_z = k_o$, the fields no longer decay exponentially and this may be considered as the cutoff frequency of the wave. If $k_z = k_o$, Eq. (30) becomes:

$$[k_o]_{\text{cutoff}} = k_{c_n} = \frac{2}{L}\sqrt{\frac{n(n+1)}{K}} \tag{31}$$

where k_{c_n} is the free space phase constant of the cutoff frequency for a particular value of n.

Thus it is seen that for each integral value of n, there is a distinct cutoff frequency for the trapped wave which is indicative of its modal character.

The separation constant k_z is related to the phase velocity of the trapped modes. For a particular frequency related to a particular k_o, only one k_z can exist in a certain mode. To find its value, Eq. (30) must be solved for k_z. Since the equation is transcendental; a graphical solution is indicated.

In Fig. 2, both the left-hand side and right hand side of Eq. (30) are plotted along the ordinate as a function of k_oL. The solid curves represent the left-hand side for different values of K, while the dotted lines represent the right-hand side for different values of n. It is noted that the dotted lines are for $k_z/k_o = 1$, which intersect the solid curves at values of k_oL corresponding to the cutoff value $k_{c_n}L$. If a frequency above cutoff is considered, the dotted lines have a positive slope corresponding to a fixed value of k_z/k_o. As an example, assume that a frequency corresponding to k_1L is trapped in the n = 1 mode, and that it is desired to determine its phase velocity, $V_1 = \omega/k_z$. Assume also that K = 2. Referring to Fig. 2, extend k_1L to point B. Since the mode is one which corresponds to n = 1, the line AB is drawn. From Eq. (30) the slope of AB is:

$$2\sqrt{\left(\frac{k_z}{k_o}\right)^2 - 1}, \tag{32}$$

and solving for k_z, we have:

$$k_z = k_1\sqrt{\left(\frac{\text{Slope AB}}{2}\right)^2 + 1}. \tag{33}$$

IV. Equations of the Trapped Modes

Since it has been shown that Eq. (19) represents the trapped modes for $\alpha = -2\beta - n$. this relation may be substituted into Eq. (19) to give:

$$\Phi = \left[-e^{-\frac{2x}{L}}\right]^{\beta} \left[1 + e^{-\frac{2x}{L}}\right]^{-2\beta} F\left[-n,\ 4\beta + 1 + n;\ 2\beta + 1;\ \frac{1}{1 + e^{-\frac{2x}{L}}}\right] \tag{34}$$

Since the first parameter of the hypergeometric series is a negative integer, it can be seen from the definition of the hypergeometric series (Eq. 17) that the series terminates after a finite number of terms. For n = 0, or the lowest mode:

$$\Phi = \left[-e^{-\frac{2x}{L}}\right]^{\beta} \left[1 + e^{-\frac{2x}{L}}\right]^{-2\beta} \tag{35}$$

which simplifies to:

$$\Phi = \left[-\frac{1}{4}\operatorname{sech}^2\frac{x}{L}\right]^{\beta} \tag{36}$$

The complete solution for the field is then, from Eq. (5), (7), and (36):

$$Ey_{n=0} = e^{jk_z z}\left[-\frac{1}{4}\operatorname{sech}^2\frac{x}{L}\right]^{\frac{\alpha}{2}} \tag{37}$$

Similarly, for n = 1:

$$\Phi = \left[-e^{-\frac{2x}{L}}\right]^{\beta} \left[1 + e^{-\frac{2x}{L}}\right]^{-2\beta} \left[1 - \frac{2}{1 + e^{-\frac{2x}{L}}}\right] \tag{38}$$

and

$$Ey_{n=1} = -e^{jk_z z}\left[-\frac{1}{4}\operatorname{sech}^2\frac{x}{L}\right]^{\frac{\alpha+1}{2}} \tanh\frac{x}{L} \tag{39}$$

For n = 2:

$$\Phi = \left[-e^{-\frac{2x}{L}}\right]^{\beta} \left[1 + e^{-\frac{2x}{L}}\right]^{-2\beta} \left[1 - \frac{2(4\beta+3)}{2\beta+1}\,\frac{1}{1 + e^{-\frac{2x}{L}}} + \frac{2(4\beta+3)}{2\beta+1}\,\frac{1}{\left(1 + e^{-\frac{2x}{L}}\right)^2}\right] \tag{40}$$

and

$$Ey_{n=2} = e^{jk_z z}\left[-\frac{1}{4}\operatorname{sech}^2\frac{x}{L}\right]^{\frac{\alpha+2}{2}} \left[1 - \frac{4\beta+3}{4\beta+2}\operatorname{sech}^2\frac{x}{L}\right] \tag{41}$$

The approximate plots of the three lowest modes are shown in Fig. 3.

V. Conclusions

The analysis shows that a variable dielectric slab will guide waves in much the same manner as a uniform dielectric slab in free space. These guided waves are trapped within the slab according to distinct modes. Each mode possesses a cutoff frequency that is dependent upon the maximum value of dielectric constant and the width of the slab. The guiding of the waves in the medium resembles that of other dielectric guides in that frequencies well above cutoff are more contained within the slab. In addition to this resemblance, the phase velocity of the trapped wave decreases with increasing frequency.

Acknowledgment

The author is deeply indebted to Professor J. Shmoys at the Polytechnic Institute of Brooklyn. It was Professor Shmoys who first suspected the existence of the trapped modes and suggested this problem. Without his help and guidance, this paper would not have been possible.

Appendix

Simplification of the coefficieints of Fq. (25)

Since:

$$\Gamma(1+z) = z\Gamma(z) \tag{A1}$$

and

$$\Gamma(z)\Gamma(-z) = -\frac{\pi}{z}\csc(\pi z) \tag{A2}$$

and

$$\Gamma(z)\Gamma(1-z) = \pi\csc(\pi z) \tag{A3}$$

Then:

$$\frac{\Gamma(-2\beta)\Gamma(2\beta+1)}{\Gamma(\alpha)\Gamma(1-\alpha)} = \frac{2\beta\Gamma(-2\beta)\Gamma(2\beta)}{\Gamma(\alpha)\Gamma(1-\alpha)} = -\frac{\csc(2\pi\beta)}{\csc\pi\alpha} = -\frac{\sin\pi\alpha}{\sin(2\pi\beta)} \tag{A4}$$

From Eq. (A1), (A2), and (A3), we also obtain:

$$\frac{\Gamma(1+z)\Gamma(-z)}{\Gamma(1-z)} = -\Gamma(z) \tag{A5}$$

Applying Eq (A5):

$$\frac{\Gamma(1+2\beta)\Gamma(-2\beta)\Gamma(1+2\beta)}{\Gamma(1-2\beta)\Gamma(1+2\beta-\alpha)\Gamma(2\beta+\alpha)} = -\frac{\Gamma(2\beta)\Gamma(1+2\beta)}{\Gamma(2\beta+\alpha)\Gamma(1+2\beta-\alpha)} \tag{A6}$$

But:[6]

$$\frac{\Gamma(Z_1)\Gamma(Z_2)}{\Gamma(Z_1+Z_3)\Gamma(Z_2-Z_3)} = \prod_{n=o}^{\infty}\left[1+\frac{Z_3}{Z_1+n}\right]\left[1-\frac{Z_3}{Z_2+n}\right] \tag{A7}$$

so that:

$$\frac{\Gamma(1+2\beta)\Gamma(-2\beta)\Gamma(1+2\beta)}{\Gamma(1-2\beta)\Gamma(1+2\beta-\alpha)\Gamma(2\beta+\alpha)} = \prod_{n=o}^{\infty}\left[1+\frac{\alpha}{2\beta+n}\right]\left[1-\frac{\alpha}{1+2\beta+n}\right] \tag{A8}$$

References

1. H. Eyring, J. Walter, and G. Kimball, *Quantum Chemistry* (New York: Wiley and Sons, Inc., 1944), p. 311.

2. A.R. Forsyth, *A Treatise on Differential Equations* (London: Macmillan Co. Ltd., 1956; 6th Edition), p. 209.

3. Ibid., pp. 215-216.

4. Ibid., pp. 218.

5. A. Erdelyi, ed., *Higher Transcendental Functions*, Vol. I (New York: McGraw-Hill Book Co., Inc., 1953), p. 107.

6. Ibid., pp. 3-5.

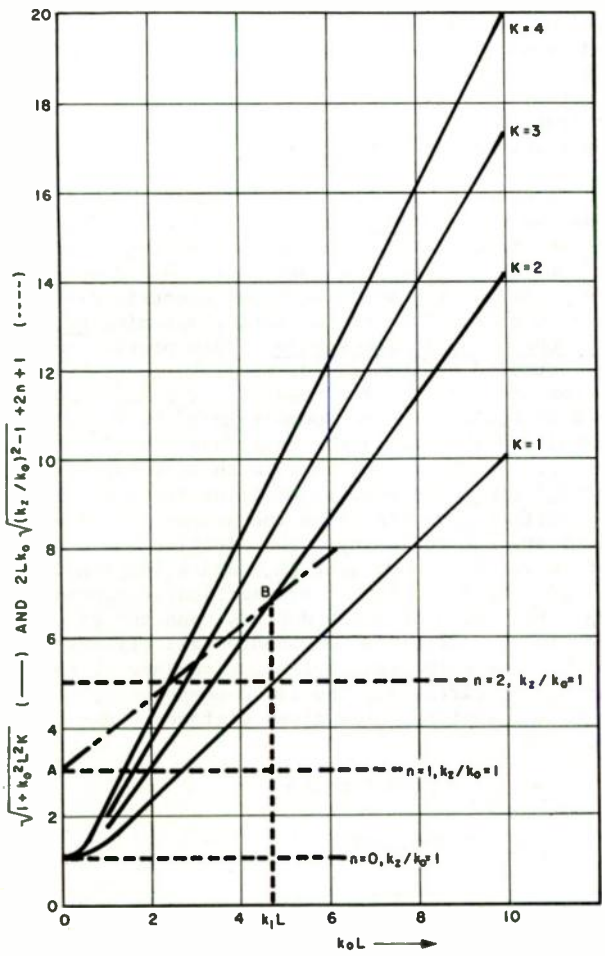


FIG.2 GRAPHICAL SOLUTION FOR TRAPPED WAVES

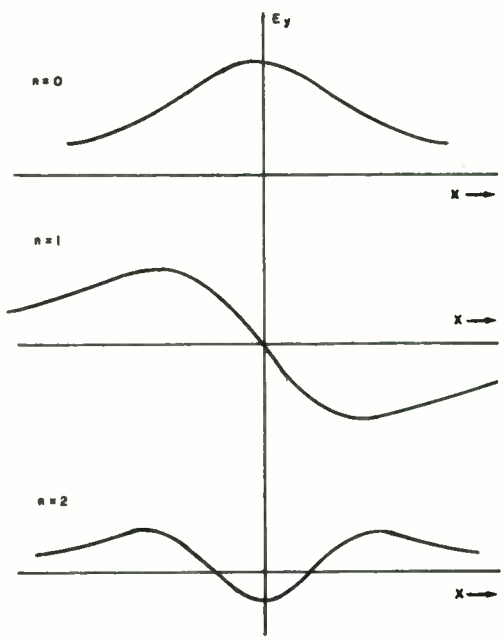


FIG.3 APPROXIMATE PLOT OF THE THREE LOWEST MODES

SUBMERGED ANTENNA PERFORMANCE

W. L. Weeks and R. C. Fenwick
Collins Radio Company
Richardson, Texas

Summary

This paper gives the results of a theoretical and experimental study of the characteristics of antennas in lossy environments, such as wires buried in the earth or sea. A criterion for the evaluation of such antennas (for surface waves or any other single mode of propagation), called relative communication efficiency, is defined and a mathematical formula for its calculation is derived. A number of specific evaluations are presented, with emphasis on horizontal wire antennas. The results of experiments to determine the input impedance and current distribution on wires of different sizes and different insulation thicknesses are summarized and compared to a simple theory. The results of measurements of the field strength of horizontal wires and arrays are presented and compared to the theoretical results. The signal to noise ratios observed with submerged receiving antenna structures are discussed.

1. Introduction

It is sometimes impossible or impractical to employ the best antennas for a given communication link. This may be so because of military or aesthetic requirements and/or structural difficulties. It may even be necessary to submerge the antenna structure into a lossy medium, although in the absence of other requirements this step is clearly inadvisable. In such instances it is especially important to have a precisely defined criterion for comparing one possible antenna structure to another. The object of this report is to define such a criterion, to illustrate the use of this criterion in a few common situations, and to present the evaluations and characteristics of basic subsurface antenna types. The results are applicable to such diverse problems as very low frequency communications, submarine communications, hardened communications and airport communications.

If the antenna is under the surface of the earth or sea, fields are established in the medium (earth or sea) and these propagate outward. However, in the absence of some fortuitous earth duct propagation, the attenuation of the spherical waves traveling in direct paths is very much greater than the 6 db/octave value which characterizes free space. The energy is dissipated by the conductivity of the earth or sea, and since, for example, the added attenuation of direct waves at 100 kc/sec is typically in the range from 175 to 17500 db/mile, the signals must reach the receiving site by some mechanism other than straight line propagation through the earth.

2. Modes of Propagation

The main modes of propagation in a communication system in which one or both terminals is submerged are 1) ground or surface wave propagation (to ground stations) 2) sky wave propagation 3) VLF and ELF mode propagation (which we regard as the general case of types 1) and 2), and 4)"Direct" paths (to aircraft). These modes of propagation have been individually examined in detail[5]. While there are a few performance features which are peculiar to certain modes (especially the horizontally polarized sky waves and direct waves), the important trends in the performance are the same for all modes of propagation. Consequently, for economy, we will confine most of our attention here to the ground wave, or surface wave mode of propagation. As a matter of fact, in certain communications systems the surface wave mode may be the only important mode of propagation.

In order to attach a precise meaning to the mode, we will define the surface wave fields to be the fields (at say 10λ or more distance) on or not far under the surface of the earth which are produced by a point vertical electric dipole on the surface of the earth, assuming no atmosphere and no ionosphere. This problem has been studied by Sommerfeld, Weyl, Burrows, Norton and others. The results are summarized in Stratton's[1] book, in Sommerfeld's[2] book and by Wait[3,4] who treats the horizontal dipole source. To the extent that earth curvature can be neglected, the most significant features of the surface wave field for the present considerations are the following: The electric field has both a vertical component (E_z) and a longitudinal (in the direction of propagation) component (E_ρ), the ratio of which depends upon the ground constants. All field components fall off exponentially with depth below the surface of the earth. Specifically, the field components, beneath the surface, at a given distance, have the form

$$E_{ze} \doteq Cn^{-1}\, e^{-jnkz}\, e^{-jk\rho}$$

$$E_{\rho e} \doteq Ce^{-jnkz} e^{-jk\rho} \qquad (1)$$

where

$n = \sqrt{\epsilon_r - j\frac{\sigma}{\omega\epsilon_0}}$, the complex index of refraction of earth

$k = 2\pi / \lambda_{air}$

z = depth below the surface (positive downward)

C = a complicated quantity involving distance and ground constants.

It follows from (1) that, beneath the surface, the ratio of components at a given position is

$$\frac{E_{ze}}{E_{\rho e}} = \frac{1}{n} . \qquad (2)$$

However, from the preceding references or the simple argument which follows, the ratio of the field components in the air is

$$\frac{E_{za}}{E_{\rho a}} = n , \qquad (3)$$

The simple argument goes as follows: tangential E must be continuous at the surface so $E_{\rho a} = E_{\rho e}$ at the surface. But normal D must also be continuous, so $\epsilon_a E_{za} = \epsilon_e E_{ze}$. Dividing the latter equation by the former

$$\frac{E_{za}}{E_{\rho a}} = \frac{\epsilon_e E_{ze}}{\epsilon_a E_{\rho e}} = n^2 \frac{E_{ze}}{E_{\rho e}} \qquad (4)$$

so that given equation (2), equation (3) follows. The form of equations (1), (2), and (3) is responsible for several important features in the antenna evaluations.

The fact, as stated above, that the mode of propagation is defined by a consideration of the fields produced by a particular point source placed at the receiving location is central to the discussion. Such a device is also very useful in the evaluation of which of various modes of propagation is likely to be of most importance. But mainly it serves as an anchor point in what sometimes seems to be a sea of confusion.

3. Criterion for the Comparison of Antennas: Relative Communication Efficiency

When an antenna is in a lossy environment it is difficult to apply the usual concepts of gain and efficiency to describe the performance. For this reason it has been found quite helpful to define another, related, quantity in order to compare the performance of antennas in lossy environments. The criterion for comparison will be called relative communication efficiency, (RCE).

The relative communication efficiency is defined in the obvious way. Given that the input powers to a pair of antennas, s, and t, (standard and test) are equal, the definition is the ratio

$$RCE_{t,s} = \frac{\text{Power density at Receiving Position when t is transmitting}}{\text{Power density at Receiving Position when s is transmitting}} .$$

Usually, if an antenna comparison is desired (as opposed to a system comparison), the power density referred to above is that in a given or desired mode of propagation. Then

$$RCE_{t,s} = \frac{P_{dm,t}}{P_{dm,s}} \qquad (5)$$

where the notation $P_{dm,t}$ means power density in the desired mode when the test antenna, t, is transmitting, and $P_{dm,s}$ is defined similarly. There is obviously another perfectly equivalent expression for the relative communication efficiency. It is the ratio of the power inputs to antennas s and t which is necessary to produce the same power density in the desired mode at the receiving position:

$$RCE_{t,s} = \frac{P_{\text{input, s}}}{P_{\text{input, t}}} \qquad (6)$$

In practice, for convenience in the determination of transmitter power requirements, the standard antenna is often taken to be a quarter wave vertical monopole, so that standard propagation charts and calculations can be employed.

4. Reciprocity Theorem and Derivation of Basic Formulas

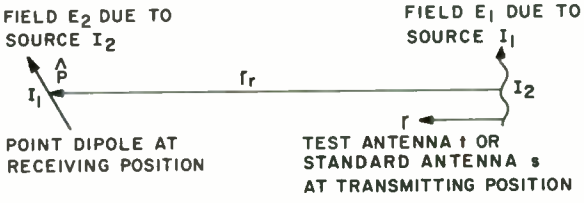
It seems that the easiest way to obtain a useful formula for comparing antennas is to employ the reciprocity theorem. It will be recalled that the following theorem can be deduced vigorously from Maxwell's equations

$$\iiint (E_2 \cdot J_1 - H_2 \cdot M_1)\,dv = \iiint (E_1 \cdot J_2 - H_1 \cdot M_2)\,dv \qquad (7)$$

In this formula, source 1 consisting of electric currents J_1 and magnetic currents M_1 generates field E_1, H_1, while source 2 consisting of J_2 and M_2 generates the field E_2, H_2. The materials must be linear and isotropic, and the sources must be finite and operate at the same frequency, but otherwise the theorem is perfectly general. For brevity, in this report we will exclude magnetic current sources from the discussion and we will almost always be concerned with the currents on the wires. In such a case the reciprocity theorem is

$$\int E_2 \cdot I_1(\xi_1)\,d\xi_1 = \int E_1 \cdot I_2(\xi_2)\,d\xi_2 \qquad (8)$$

where ξ_1 and ξ_2 are variables which measure length along the thin line currents, I_1 and I_2.



As we will use this theorem, the source I_2 will be the antenna under consideration. In the comparison process, it will be first antenna t and then antenna s. The theorem will be useful to us in that the field E_2 (which is that due to either the test or the standard antenna) at the receiver can be found by the simple expedient of regarding I_1 as a point dipole at the receiving position $I_1(\xi) = \hat{p}\,\delta(r-r_r)$, where r_r is the location of the receiving position, and $\hat{p}$ is a unit vector showing the orientation of the point source. When this source is used in (8),

the result gives an expression for the component of E_2 in the direction of $\hat{p}$, at the receiving position:

$$E_{2p}(r_r) = \int E_1 \cdot I_2(\xi_2)\, d\xi_2 \qquad (9)$$

To repeat, E_1 is the field at the position of I_2 which is generated by a unit point dipole in whatever environment actually exists (and I_2 is the transmitting current distribution of first antenna t, and then antenna, s.). This quantity E_1 is known, and it defines the mode of propagation which we are concerned with in the comparisons. Normally (although not necessarily) this will be the same mode of propagation for the antennas under comparison. We can apply equation (9) twice, once for the test antenna and once for the standard antenna and obtain the following two equations:

$$\begin{aligned} E_t &= \int E_1 \cdot I_t\ (\xi_t)\, d\xi_t \\ E_s &= \int E_1 \cdot I_s\ (\xi_s)\, d\xi_t \end{aligned} \qquad (10)$$

where E_t and E_s are respectively the desired components of the fields at the receiving position produced by the test and standard antennas.

Making use of equation (10), it is quite easy to obtain a formula for the relative communication efficiency. Since the power density in the desired mode is proportional to the square of the field strength, if we require that the power densities produced by antennas t and s be the same at the receiving position, this implies $|E_t|^2 = |E_s|^2$ and therefore

$$\left|\int E_1 \cdot I_t(\xi_t)\, d\xi_t\right|^2 = \left|\int E_1 \cdot I_s(\xi_s)\, d\xi_s\right|^2 \qquad (11)$$

Now the condition can be satisfied in general only if the current amplitudes are different, which means that the input powers will be different. The ratios of power inputs can be obtained with the aid of the familiar relationship

$$P_{input} = \tfrac{1}{2}\left|I_{in}\right|^2 R_{in} \qquad (12)$$

and the definitions

$$i_t(\xi_t) = \frac{I_t(\xi_t)}{I_{in,t}}\ ,\quad i_s(\xi_s) = \frac{I_s(\xi_s)}{I_{in,s}}\ . \qquad (13)$$

With the definitions (13), the condition (11) can be written

$$\begin{aligned} &\left|I_{in,t}\right|^2 \left|\int E_1 \cdot i_t(\xi_t)\, d\xi_t\right|^2 = \\ &\left|I_{in,s}\right|^2 \left|\int E_1 \cdot i_s(\xi_s)\, d\xi_s\right|^2 \end{aligned} \qquad (14)$$

and then, with (12), the input powers introduced into the expression and the ratio of these obtained so that from (6), the equation is

$$RCE_{t,s} = \frac{P_{input,s}}{P_{input,t}} = \frac{R_{in,s}}{R_{in,t}} \frac{\left|\int E_1 \cdot i_t(\xi_t)\, d\xi_t\right|^2}{\left|\int E_1 \cdot i_s(\xi_s)\, d\xi_s\right|^2}\ . \qquad (15)$$

This is the basic formula for the comparison of a pair of antennas. It is equivalent to the standard methods for comparison of antennas in free space or lossless environments (gain and efficiency) but it is applicable to any situation for which the reciprocity theorem holds (linear, isotropic media).

Since this criterion has not often been used in the literature, it is worthwhile to apply it to a familiar example for purposes of illustration. Let us therefore compare a $\frac{\lambda}{4}$ vertical monopole with a good ground system to a $\frac{\lambda}{20}$ monopole with the same ground system, for communication in the surface wave mode of propagation. The current distribution (as a transmitter) on the quarter wave vertical is essentially sinusoidal:

$$I_s = I_{in,s} \cos \frac{2\pi}{\lambda} z$$

and the surface wave field produced by a distant point dipole, E_{1z}, is constant over the wire. The integral for the quarter wave vertical is then

$$\left|\int E_1 \cdot i_s(\xi_s)\, d\xi_s\right|^2 = \left|E_{1z} \int_0^{\frac{\lambda}{4}} \cos\frac{2\pi z}{\lambda}\, dz\right|^2 = \left|\frac{E_{1z}\lambda}{2\pi}\right|^2$$

In the short monopole, the current is linear to a good approximation:

$$I_t = I_{in,t}\ (1 - \frac{20}{\lambda} z)$$

so the integral in question is

$$\left|\int E_1 \cdot i_t(\xi_t)\, d\xi_t\right|^2 = \left|E_{1z} \int_0^{\frac{\lambda}{20}} (1 - \frac{20}{\lambda} z)\, dz\right|^2 = \left|E_z \frac{\lambda}{40}\right|^2 .$$

It follows then that RCE for the short vertical compared to the $\frac{\lambda}{4}$ vertical is

$$RCE_{t,s} = \frac{R_{in,\frac{\lambda}{4}}}{R_{in,\frac{\lambda}{20}}} \left|\frac{\pi}{20}\right|^2 .$$

Now a quarter wave vertical with a good ground system may have an input resistance of about 37 ohms of which perhaps 1 ohm arises from ohmic losses. The short vertical might have an input resistance of 2 ohms of which about 1 ohm is due to ohmic loss. The relative communication efficiency of the short monopole compared to the quarter wave vertical is then

$$RCE_{t,s} \doteq .46\ .$$

If the system were constructed so that the ohmic losses in both structures were negligible, then $R_{in\ \frac{\lambda}{4}} \approx 36$, $R_{\lambda\ \frac{\lambda}{20}} \approx 1$

so that

$$RCE_{t,s} \doteq .91 \text{ (no losses)}$$

With this assumption, the antennas are perfectly efficient so the value .91 for the RCE represents the gain of the $\frac{\lambda}{4}$ vertical over the $\frac{\lambda}{20}$ vertical.

5. General Features of Antenna Performance

Effect of Orientation and Position

The main factor which determines antenna performance in surface wave communication systems is the relative size of the field components of the surface wave and this, mathematically, depends on the position of the complex index of refraction factor, n, in the equation (see equations 2 and 3). To illustrate this point, consider the performance of very short thin antennas with uniform current along their lengths. Let the vector $\hat{\underline{a}}$ designate the orientation of the antennas. With such short antennas, the integrals in (15) are trivial and that equation gives the result

$$RCE_{t,s} = \frac{R_{in,s}}{R_{in,t}} \left| \frac{E_1 \cdot \hat{a}_t}{E_1 \cdot \hat{a}_s} \right|^2$$

The comparison of a short vertical and short horizontal in air can be readily made (t implies horizontal, s implies vertical). The quantity $E_1 \cdot \hat{a}_t = E_{\rho a} \cos \phi$, where ϕ is the angle between the horizontal antenna and the direction to the receiving site (location of the point vertical, J_1). This leads to the result

$$RCE_{t,s} = \frac{R_{in,s}}{R_{in,t}} \left| \frac{E_{\rho a} \cos\phi}{E_{za}} \right|^2 \doteq$$

$$\frac{R_{in,s}}{R_{in,t}} \left| \frac{\cos\phi}{n} \right|^2 \quad \left\{ \begin{array}{l} \text{short horizontal} \\ \text{to short vertical} \\ \text{at the surface} \end{array} \right\} \qquad (16)$$

where the last step follows from equation (3). The horizontal antenna is most effective when it lies along the line to the receiving antenna, but since $R_{in,s}$ and $R_{in,t}$ could be comparable in size, the horizontal is much less effective than the vertical since the complex index of refraction $|n|^2 \doteq \frac{\sigma}{\omega\epsilon}$ is typically quite large. If we assume that the input resistances are the same, the relative communication efficiency, in db, in the end-fire direction is given approximately by the equation

$$RCE_{t,s} \cong -42.5 + 10 \log \frac{f_{kc}}{\sigma_{mm}} \text{ db} \begin{cases} \text{short horizon-} \\ \text{tal to short} \\ \text{vertical wire} \\ \text{at surface} \end{cases} \qquad (17)$$

(σ_{mm} is conductivity in millimhos per meter.) It is to be emphasized that results (16) and (17) apply to short antennas at or just above the surface.

If both antennas are under the surface, the conclusion is opposite. If we maintain the notation t for horizontal, s for vertical, then the relationship is

$$RCE_{t,s} = \frac{R_{in,s}}{R_{in,t}} \left| \frac{E_{\rho e} \cos\phi}{E_{ze}} \right|^2 \doteq \frac{R_{in,s}}{R_{in,t}} \left| n \cos\phi \right|^2$$

$$= |n \cos\phi|^2 \quad \left\{ \begin{array}{l} \text{short horizontal to} \\ \text{short vertical, both} \\ \text{under the surface} \end{array} \right\} \qquad (18)$$

Equation (18) shows that if the antennas must be buried, the horizontal antenna is better than the vertical one by the same factor that the vertical is better than the horizontal when both are just above the surface. It is rather surprising to realize that burying a vertical antenna and keeping its orientation vertical results in about 85 db loss in efficiency.

Effect of Depth

Of course, burying an antenna to greater depths decreases its effectiveness in surface wave communication. A quantitative evaluation can be made by computing the relative communication efficiency of a pair of short horizontal antennas, one at the surface, and one at a depth h. The expression, (1), for the surface wave field can be employed in (15) as follows (t implies at depth, s implies at surface)

$$RCE_{ts} = \frac{R_{in,s}}{R_{in,t}} \left| \frac{E_{1\rho}(o)}{E_{1\rho}(h)} \right|^2 = \frac{R_{in,s}}{R_{in,t}} \left| e^{-jnkh} \right|^2$$

$$= \frac{R_{in,s}}{R_{in,t}} e^{\frac{-2h}{\delta}} \quad \begin{array}{l} \text{(Antenna at depth h,} \\ \text{relative to antenna} \\ \text{at the surface)} \end{array} \qquad (19)$$

where δ is the skin depth. In decibels, this is

$$RCE_{t,s} = -8.686 \frac{h}{\delta} + 10 \log \frac{R_{in,s}}{R_{in,t}} \text{ (db)} \qquad (20)$$

This result holds for any type of antenna having a well defined depth and this loss appears at the transmitter and the receiver if both are buried. Resistance variation with depth is discussed in section 7.

In the foregoing, the actual value of RCE is most strongly influenced by the position of the complex index of refraction ($|n|^2 \doteq \frac{\sigma}{\omega\epsilon_0}$) in the equation and by the attenuation with depth. Both of these factors appear in the same way for all of the modes of propagation mentioned above, so the general features are the same for all modes of propagation.

6. System Considerations

For the purpose of comparing systems, the familiar concept of system loss, which is defined

$$\text{System Loss} = \frac{\text{Power delivered to the receiving antenna load}}{\text{Power input to transmitting antenna}}$$

can be employed to compare systems. If two systems, s, and t, have the same power input, the relative system loss is

$$\text{Relative system loss} = \frac{P_{rec,t}}{P_{rec,s}} \qquad (21)$$

where $P_{rec,t}$ is the power delivered to the load connected to the receiving antenna in system, t, and $P_{rec,s}$ is defined similarly (same input power to both systems).

Although not necessary, it is helpful at this point to introduce the concept of receiving antenna cross sectional area, A_{rec} and to show that the ratio of the receiving cross sections for a pair of antennas is the same as their relative

communication efficiency (RCE) defined earlier. The cross section is defined as follows

$A_{rec} = \frac{P_{rec}}{P_{inc}}$, where P_{inc} is the power density in the incident field, so the ratio of receiving cross sections of a pair of antennas in the same incident field is

$$\frac{A_{rec,t}}{A_{rec,s}} = \frac{P_{rec,t}}{P_{rec,s}} \qquad (22)$$

The power received can be calculated from an equivalent circuit for the receiving antenna. If the load is a conjugate match for the antenna, the power delivered to the load is

$$P_{rec} = \frac{1}{8} \left(\frac{V_{oc}}{R_a} \right)^2 . \qquad (23)$$

To demonstrate the equivalence to RCE, it will be recalled (or can be shown) that the open circuit voltage produced in wire antenna by a field E_1 is

$$V_{oc} = \int E_1 \cdot i(\xi) \, d\xi \qquad (24)$$

where $I_{i(\xi)} = \frac{I(\xi)}{I_{in}}$. Therefore if we evaluate (23) using V_{oc} from (24) for each antenna, t, and s, and so replace $P_{rec,t}$ and $P_{rec,s}$ in (22), the result is

$$\frac{A_{rec,t}}{A_{rec,s}} = \frac{R_{in,s} \left| \int E_1 \cdot i_t(\xi_t) d\xi_t \right|^2}{R_{in,t} \left| \int E_1 \cdot i_s(\xi_s) d\xi_s \right|^2} \qquad (25)$$

A comparison of this equation to equation (15) shows that the ratio on the right is the relative communication efficiency defined above so that

$$\frac{A_{rec,t}}{A_{rec,s}} = RCE_{t,s}$$

which was to be demonstrated.

Now if we compare two systems, s and t, the relative system loss, if both transmitting antennas have the same power input, is

$$\text{Relative system loss}_{t,s} = \frac{P_{inc,t}}{P_{inc,s}} \frac{A_{rec,t}}{A_{rec,s}} = RCE_{t,s}\Big|_{trans} \frac{A_{rec,t}}{A_{rec,s}} = RCE_{t,s}\Big|_{trans} \times RCE_{t,s}\Big|_{rec} \qquad (26)$$

In words, whether or not the systems employ the same antennas as receivers and transmitters, the relative system loss is the product of the relative communication efficiency of the transmitting antennas times the relative communication efficiency of the receiving antennas. This result greatly enhances the utility of the specific values of RCE which are presented in this and a companion report.

7. Theoretical Model for Antenna with Comparison to Experiment

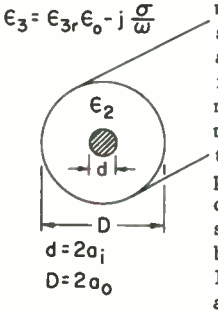
It should be clear from what has been said above, and, in particular from equation (15), that in order to compare antennas by the RCE criterion, we must know (1) the current distributions on the antenna when it is transmitting (2) the input resistance of the antenna. These quantities can, at least, be determined experimentally (and this has been done). Moreover fortunately, a relatively simple theory gives good approximations for certain simple types.

The Buried Wire as a Lossy Transmission Line

If the antenna consists of an insulated wire which is buried deeply enough (how deep is "enough" is discussed later in this section) in the earth or sea, a good theoretical model for the structure is a transmission line which consists of an insulated perfect conductor immersed in an infinite lossy medium. The input impedance to and current distribution on such a line can be determined from standard transmission line theory, once the propagation constants have been determined.

Propagation Constants

Consider a structure which consists of a perfectly conducting wire of diameter d, surrounded by an insulator which has a diameter D and dielectric constant ϵ_2, immersed in an infinite lossy medium. Clearly, there are many possible modes of propagation on this structure. The propagation constants for any of these modes can be found by solving the electromagnetic boundary value problem in cylindrical coordinates. This is a straightforward, if somewhat complicated and tedious task, and the resulting equation for the propagation constants of the TM modes is



$$\frac{H_n^{(2)}(k_3 a_o)}{H_n^{(2)\prime}(k_3 a_o)} = \frac{k_2 \epsilon_3 \left[J_n(k_2 a_i) N_n(k_2 a_o) - N_n(k_2 a_i) J_n(k_2 a_o) \right]}{k_3 \epsilon_2 \left[J_n(k_2 a_i) N_n'(k_2 a_o) - N_n(k_2 a_i) J_n'(k_2 a_o) \right]} \qquad (27)$$

where $k_2 = (\gamma_z^2 - \gamma_2^2)^{\frac{1}{2}}$ $k_3 = (\gamma_z^2 - \gamma_3^2)^{\frac{1}{2}}$

$\gamma_2^2 = -\omega^2 \mu \epsilon_2$, $\gamma_3^2 = -\omega^2 \mu \epsilon_3$,

$k_2^2 + \gamma_2^2 = k_3^2 + \gamma_3^2$,

γ_z is the propagation constant along the line, a_i and a_o are the inner and outer radii of the insulator, J_n, N_n are the Bessel functions of 1st and 2nd kind, respectively, and $H_n^{(2)} = J_n - jN_n$. The lowest order of these TM modes is the

one which passes into the familiar TEM mode for coaxial lines as the conductivity of medium III goes to infinity. In general, the roots to equations (27) must be found graphically or by trial and error (with a digital computer) but the computations are involved since the arguments of the Bessel functions are complex numbers. For many interesting values of d, D and σ, equation (27) can be approximated, incorporating the small argument approximations for the Bessel functions. If this is done, and the added restriction is imposed that the conductivity is high enough so that $\gamma_3 \gg \gamma_z$ or $k_3 \doteq \sqrt{\omega\mu\sigma}\angle -45^\circ$, the following equation for the propagation constant may be derived (δ is the skin depth in medium three).

$$\frac{\gamma_z}{\gamma_2} \doteq \left(1 - \frac{j\pi/4 + \ln\sqrt{2}(.89)\, a_o/\delta}{\ln a_o/a_i}\right)^{\frac{1}{2}} \qquad (28)$$

The overall approximation should be acceptable if a_o/δ is less than about .1 since the Hankel functions will then have an argument of less than about .3.

Figs. 1 and 2 display the real and imaginary parts of the propagation constant,

$$\frac{\gamma_z}{\beta_D} = \frac{\alpha_z}{\beta_D} + j\frac{\beta_z}{\beta_D}$$

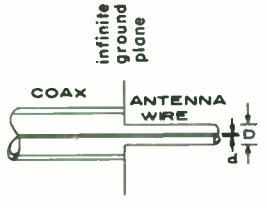
relative to the intrinsic propagation constant for the insulating sheath

$$(\gamma_2 = j\beta_D = j\omega\sqrt{\mu\epsilon_2}),$$

as determined by the approximate equation, (28). Note that the phase constant is typically a factor of two or more different (greater) from β_D, which itself may be a factor of two or so greater than free space. For this reason, the resonant lengths of buried wires (antennas) are typically quite different from those of antennas in free space.

Input Impedance

To employ the foregoing transmission line model as an antenna we must identify the feed region and, probably, introduce some kind of termination. For example, we can represent each half of a symmetrical dipole antenna by the device of introducing an infinite ground plane perpendicular to the wire, and imagine that an ordinary coaxial cable is connected to the ground plane and the center conductor. This specifies the feed region. As a termination, a free end of the wire can be insulated from the lossy medium, or connected to the lossy medium (either directly or through another circuit element such as a choke). In lieu of good information about the impedance of the termination if connected directly, we can introduce theoretical models which remove the difficulty. To represent the "grounded" state, we imagine that another infinite conducting plane is introduced, and the center wire is joined to it. To represent the "open" state, we imagine a perfect magnetic conducting plane inserted so as to make a perfect open circuit. In either case the structure is essentially idealized to a coaxial transmission line with a thick lossy outer conductor, with either a perfect open, or perfect short at tne free end.



FEEDPOINT MODEL CUTAWAY VIEW

To the extent that the transmission line model is valid, the input impedance and current distribution can be determined from lossy transmission line theory, once the complex propagation constant has been determined. The input impedance to the line section is

$$Z_{sc} = Z_o \tanh \gamma \ell \qquad \text{(grounded end)} \qquad (29)$$

$$Z_{oc} = Z_o \coth \gamma \ell \qquad \text{(insulated end)} \qquad (30)$$

where ℓ is the length of the line section (half length of dipole antenna), $\gamma_z = \alpha_z + j\beta_z$ is the complex propagation constant, and Z_o is the characteristic impedance (V/I on an infinite line)

$$Z_o \cong \left(\frac{1}{2\pi} \ln \frac{a_o}{a_i}\right) \frac{\gamma_z}{j\omega\epsilon_2} \qquad (31)$$

where a_o and a_i are the outer and inner radii of the insulator and ϵ_2 is the dielectric constant of the insulator. Identities 655.3 and 655.4 in Dwight's tables can be employed to separate the real and imaginary parts of the input impedance:

$$\tanh \gamma \ell = \frac{\sinh 2\alpha_z \ell + j \sin 2\beta_z \ell}{\cosh 2\alpha_z \ell + \cos 2\beta_z \ell}$$

$$\coth \gamma \ell = \frac{\sinh 2\alpha_z \ell - j \sin 2\beta_z \ell}{\cosh 2\alpha_z \ell - \cos 2\beta_z \ell}$$

so

$$R_{sc} = \frac{60}{\sqrt{\epsilon_r}} \ln \frac{a_o}{a_i} \frac{\frac{\beta_z}{\beta_o} \sinh 2\alpha_z \ell + \frac{\alpha_z}{\beta_o} \sin 2\beta_z \ell}{\cosh 2\alpha_z \ell + \cos 2\beta_z \ell} \qquad (32)$$

$$X_{sc} = \frac{60}{\sqrt{\epsilon_r}} \ln \frac{a_o}{a_i} \frac{\frac{\alpha_z}{\beta_o} \sinh 2\alpha_z \ell + \frac{\beta_z}{\beta_o} \sin 2\beta_z \ell}{\cosh 2\alpha_z \ell + \cos 2\beta_z \ell} \qquad (33)$$

$$R_{oc} = \frac{60}{\sqrt{\epsilon_r}} \ln \frac{a_o}{a_i} \frac{\frac{\beta_z}{\beta_o} \sinh 2\alpha_z \ell - \frac{\alpha_z}{\beta_o} \sin 2\beta_z \ell}{\cosh 2\alpha_z \ell - \cos 2\beta_z \ell} \qquad (34)$$

$$X_{oc} = \frac{60}{\sqrt{\epsilon_r}} \ln \frac{a_o}{a_i} \frac{-\frac{\alpha_z}{\beta_o} \sinh 2\alpha_z \ell - \frac{\beta_z}{\beta_o} \sin 2\beta_z \ell}{\cosh 2\alpha_z \ell - \cos 2\beta_z \ell}, \qquad (35)$$

where

ϵ_r is the relative permittivity of the insulator. The propagation constant is given in the section above.

These equations, (33) and (35), show that the condition for resonance is, $\pm \sin 2\beta_z \ell = \frac{\alpha_z}{\beta_z} \sinh 2\alpha_z \ell$, where the plus and minus signs give short circuit resonances and open circuit resonances respectively. The input impedance for short lengths is of special interest. This is discussed in a later section.

Current Distribution

The next item of interest is the current dis-

tribution on such submerged wires. If we call the current at the input I_o, the following expressions give the current distributions on end fed wires, as determined by conventional lossy transmission line theory:

$$I_{sc}(z) = \frac{I_o}{1+e^{-2\gamma\ell}}\left[e^{-\gamma z} + e^{\gamma(z-2\ell)}\right] \text{ end grounded} \quad (36)$$

$$I_{oc}(z) = \frac{I_o}{1-e^{-2\gamma\ell}}\left[e^{-\gamma z} - e^{\gamma(z-2\ell)}\right] \text{ end insulated} \quad (37)$$

In many practical structures, the attenuation constant, α_z, is relatively quite small. In such a case, the approximation $e^{\gamma z} \doteq e^{j\beta z}(1+\alpha_z)$, and similar forms can be employed to give the following approximate expressions for the current distribution:

$$I_{sc}(z) = \frac{I_o e^{-\gamma\ell}}{1+e^{-2\gamma\ell}}(2)\left[\cos\beta_z(\ell-z)+j\alpha_z(\ell-z)\sin\beta_z(\ell-z)\right] \quad (38)$$

$$I_{oc}(z) = \frac{I_o e^{-\gamma\ell}}{1-e^{-2\gamma\ell}}\,2\left[j\,\sin\beta_z(\ell-z)+\alpha_z(\ell-z)\cos\beta_z(\ell-z)\right] \quad (39)$$

From these expressions it is quite clear that if $\alpha_z \ell$ is small, then the current distribution is sinusoidal, to a good approximation. If, in addition, the antenna is quite short, the wire having a grounded end has almost a uniform current distribution while the wire having an insulated end has almost a linear current distribution.

Modeling Considerations and Experimental Facility

The propagation constant solution given by equation (27) is exact for the ideal medium. The evaluation of it however involves approximations. The impedance expressions are also exact for the idealized terminations. In practice, such ideal terminations can not be or are not employed. Consequently, the validity of the foregoing representations has to be verified. Effects of proximity to surface and other environmental details should be (and were) studied.

Controlled field experiments with antennas buried in the earth are inconvenient and expensive. Field experiments in the ocean are also troublesome (especially working in the state of Iowa). Consequently, a modeling procedure was adopted. A modeling facility based on accepted modeling theory was constructed. Basically, the facility consists of a large tank of water with a means for varying the salt concentration in order to vary the conductivity of the medium. Fig. 3 is a photograph of the facility including the device for holding the wire at specified positions and the attachments for current distribution measurements.

Comparison of Measured and Calculated Impedances and Current Distributions

Figs. 4 to 7 give some typical values of input impedance to center fed dipoles, including both calculated and experimental values. The calculations are based on equations (32)-(35) for the impedance and (28) for the propagation constant, which have been programmed for a digital computer (Fortran). Figure 8 gives a comparison of calculated and measured current distribution on an end fed wire operated between its third and fourth resonances. In view of the approximations, and deficiencies in the model as discussed below, it is seen that the agreement is very close, and we conclude that the theory gives a very useful engineering approximation.

The main deficiencies in the theoretical model as applied to the modeling facility are 1) the wire is assumed to be at infinite depth and 2) the terminating impedance is assumed to be either a perfect open or a perfect short circuit. These points are discussed below. Another deficiency of the theoretical model is that no provision is made for the inhomogeneities which characterize most installations in soil. The performance details are quite sensitive to the characteristics of the material immediately adjacent to the wire, except in the case of short grounded-end wires.

Effect of Proximity to the Surface

All of the theoretical models described in this section are based on the idea that the antenna is far enough under the surface that any effect of the interface is negligible. The interface does have an influence, but experimentally it is found that the changes are not drastic as long as the antenna is <u>in</u> (as opposed to over) the lossy medium. Fig. 9 gives the effect of depth on resonant frequency and input resistance. The measured values are expressed relative to those at large depths.

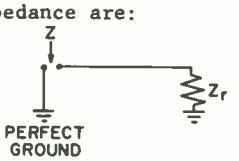
Once the antenna is <u>out</u> of the lossy medium, the properties change very rapidly. Fig. 10 shows the resistance and reactance of a 15 ft. antenna at different (small) heights above salt water. It is interesting to note that most of the changes take place within a distance of $\pm$ one skin depth from the water surface.

Effect of Terminating Impedance

The effect of a non-ideal terminating impedance can be included in the theory. The corrected expressions for the impedance are:

$$Z = \frac{Z_r Z_{oc} + Z_o^2}{Z_{oc} + Z_r}$$

or $$Z = Z_o^2\,\frac{Z_r + Z_{sc}}{Z_o^2 + Z_r Z_{sc}}$$



where Z_o is the wire characteristic impedance, Z_r is the terminating impedance, and Z_{oc} and Z_{sc} are the input impedances where $Z_r = \infty$ and $Z_r = 0$ respectively.

The change brought about by the inclusion of a resistive termination instead of a short circuit is illustrated in Figs. 4 and 5. The resistance of the "ground stake" was computed from standard formulas. The data indicate that for precision, the resistance of the ground termination should

be included in the calculation.

Effect of Actual Ground

In the theoretical models, it is usually necessary to imagine that the actual earth is characterized by a single value of conductivity (and dielectric constant). In practice, this is not the case. The conductivity varies with location, depth, moisture state and season, to name a few factors. Of course, the effects due to the latter two factors, moisture content and season, are greatest when the antenna is on the surface, with a thin insulating sheath (small $\frac{D}{d}$.)

An experimental determination of the effect of season and moisture content under such conditions was made with a 1500 ft. wire lying on Iowa earth (in a soy bean field). The impedance of the wire under different environmental conditions is shown in Figs. 11 and 12.

8. Characteristics and Comparison for some Specific Antenna Types

We will now summarize the performance characteristics of several specific antenna types.

Short Antennas

A subsurface dipole antenna is electrically "short" at frequencies much below its resonant frequency $f_{\frac{\lambda}{2}}$. In practice dipole antennas can be considered short at frequencies below one half of the resonant frequency if the ends are open, and below one quarter of this resonant frequency if the ends are shorted to ground. Formulas for the impedance and current distribution can be derived by employing small argument approximations in equations (29) - (39) above. If this is done, the result for the impedance of short grounded end types agrees with the results published earlier by Moore[9] and Wheeler[10], as given below.

The input resistance of a short dipole or monopole (end fed wire) with grounded ends is given by

$$R_o = \frac{1}{8}\omega\mu L = 30\pi^2 \frac{L}{\lambda} = .3\, f_{mc} L_{ft}$$

and reactance by

$$X_{sc} = \frac{4R_o}{\pi} \ln 1.588 \frac{\delta}{d} .$$

In practice, the resistance of the grounds, R_g, will increase the input resistance so that

$$R = R_o + 2R_g .$$

If the grounds are single vertical rods of length ℓ and radius a, and $\ell >> a$, for d.c.[8]

$$R_g = \frac{1}{2\pi\sigma} (\ln \frac{4\ell}{a} - 1) .$$

It is interesting that reactance of short grounded-end wires is not a function of insulation thickness or permittivity.

Approximate equations for the short open ended dipole are

$$R \doteq \frac{1}{24}\omega\mu_o L \doteq \frac{f_{mc} L_{FT}}{10}$$

and

$$X \doteq \frac{-7.2 \times 10^{10}}{\epsilon_r \omega L} \ln\frac{D}{d} \doteq \frac{1200\,\pi \ln \frac{D}{d}}{\epsilon_r R} .$$

These equations are not as useful as the ones for grounded ends; however, because the actual terminating impedance changes the input value and the terminating impedance is much more difficult to estimate.

The current distribution on short antennas is nearly uniform or decays almost linearly, depending respectively on whether the ends are grounded or ungrounded. Attenuation or end effects are seldom if ever great enough to seriously alter the current distribution from that which would be obtained by considering the wire and medium as a lossless coaxial transmission line with phase constant β_z.

Relative communication efficiency of short wires is given in Table I. In Table II measured RCE values are compared to calculated values.

Resonant Monopoles and Dipoles

In the manner of lossy transmission lines, subsurface wires exhibit resonances which present attractive impedance properties. The lowest order resonance is of primary interest since it requires shorter length wires and in many cases gives efficiency and pattern characteristics superior to those of longer wires operated at higher order resonances.

The general theoretical expressions for input impedance and measured impedances are given in section 7. Theoretically, for a half-wave, open-ended dipole at depth

$$R_{\lambda/2} = .15 L_{ft} f_{\lambda/2} \text{ mc} \quad \text{(dipoles)} \qquad (40)$$

(half for monopoles at $f_{\lambda/4}$)

if $\sinh \alpha_z L \cong \alpha_z L$, which is often the case, or should be. The resonant frequency $f_{\lambda/2}$ occurs approximately where $\beta_z L = \pi$ for dipoles or where $\beta_z L = \pi/2$ for monopoles. In theory, at the frequencies defined by these criteria

$$X = -\frac{\alpha_z}{\beta_z} R$$

but it is often the case that $\alpha_z/\beta_z << 1$. In practice equation (40) nearly always has given values lower than those measured, both at the surface and at depth, which has also been the case with the more exact analysis. Values of $R_{\lambda/2}$ measured in the laboratory typically run 25 to 30% higher than those given by (40).

Resonant resistance $R_{\lambda/2}$ for grounded-end dipoles is a function of end grounding resistance, and is higher with better grounds. Figure 4 demonstrates the effect. Resonant frequency is relatively independent of the grounding resistance.

In almost all cases it should be quite adequate to assume a cosinusoidal current distribution on open-ended, half-wave dipoles and quarter-wave monopoles. Figure 13 gives a typical resonant distribution. No measurements of relative phase have been made, but for reasonably small α_z the

phase is theoretically nearly uniform.

Relative communication efficiency of resonant types is given in Table I. In Table II measured RCE values are compared to calculated values.

Traveling-Wave Monopoles and Dipoles

When sufficient attenuation occurs or a wire is terminated so that the reflection from its end is negligibly small, the wire is said to be operating as a traveling-wave (TW) antenna. Such antennas possess very great bandwidth at the expense of length or efficiency. The length required is determined by the value of α_z on the wire in its environment at the frequency of operation. If 20 db is specified as the required attenuation, then the required length is

$$L_{TW} = \frac{2.3}{\alpha_z} , \qquad (41)$$

measured from the feed point. A TW dipole would thus have a length $2L_{TW}$. Lengths half of that given by equation (41) may be acceptable in practice.

The TW antenna has almost unlimited bandwidth at frequencies above that where (41) applies. The input impedance is primarily resistive, but exhibits a small capacitive reactance component. The resistance has a mean value of

$$R = \frac{60}{\sqrt{\epsilon_r}} \ln\left(\frac{D}{d}\right) \frac{\beta_z}{\beta_D} \quad \text{(monopoles)} \qquad (42)$$

(twice this for dipoles)

and the reactance has a mean value of

$$X = -\frac{60}{\sqrt{\epsilon_r}} \ln\left(\frac{D}{d}\right) \frac{\alpha_z}{\beta_D} \quad \text{(monopoles)} \qquad (43)$$

(twice this for dipoles)

For the lengths given by (41) the input impedance is theoretically very close to the values given by (42) and (43). For shorter lengths the resistance and reactance oscillate around the values given by (42) and (43). A 348-foot monopole in sea water gave measured impedances close to theoretical when operated in the traveling-wave mode.

Equations (36) and 37) give current distribution for TW operation if $\alpha_z \lambda$ is assumed very large. The result is

$$I = I_o e^{-\gamma x}$$

The Relative Communication Efficiency expressions are given in Table I.

9. Directional Properties

The dominant feature in the radiation pattern is the factor cosØ, where Ø is the angle measured from the end of the wire, as indicated in equation (16) for a short antenna. In general, for single wires, the azimuthal variation is included in the RCE calculation by using $\cos\emptyset |E_{\rho_a}| e^{+j\beta\cos\emptyset\xi}$ for "E" in $\int E\cdot i(\xi)\,d\xi$ over the wire, where Ø = 0 off the unfed end of monopoles. The practical result is that "β_o" can be replaced by "$\beta_o\cos\emptyset$" in RCE expressions derived without including the cosØ terms. (or "F" by "F cosØ" in Table I)

10. Arrays

Crossed Elements

It is possible to operate submerged antennas in crossed arrays. The pattern depends on the phase and amplitude of the input current to the crossed elements. To illustrate this point, consider a pair of short elements with uniform current distributions forming a "ground cross." Introduce coordinate axes, x, y along the crossed dipoles, with unit vectors x and y. Then the total current for the crossed dipole array is

$$i = i_x \hat{x} + e^{j\Psi} i_y \hat{y}$$

where Ψ is the phase difference between the elements. To find the open circuit voltage or evaluate RCE, we need the expression $E\cdot i$. If the incoming wave makes an angle Ø with the x-axis, the quantity is

$$E_1 \cdot i = E_1 i_x \cos\emptyset + E_1 i_y e^{j\Psi} \sin\emptyset$$

As an example let the amplitudes be the same $i_x = i_y = 1$. Then

$$\int E_1 \cdot i\, d\xi = E_1 \lambda (\cos\emptyset + e^{j\Psi}\sin\emptyset) .$$

The magnitude of the open circuit voltage squared, or the factor in the RCE expression is then

$$\left|\int E_1 \cdot i\, d\xi\right|^2 = \left|E_1\lambda\right|^2 (1 + 2\cos\Psi\sin\emptyset\cos\emptyset) .$$

This expression is essentially the pattern of the crossed dipoles. It shows that if the phase difference is an odd multiple of $\frac{\pi}{2}$, the pattern is omnidirectional in the horizontal plane.

Parallel Elements

It is also possible to construct an array with parallel elements. If the elements of the array are short, and the array is not so large that the field amplitude at a distant field point is different for different elements, then it can be shown that the space factor in the horizontal plane (surface waves) is the same as the space factor of the elements in free space. The difference from free space is, that because they are located in a lossy medium, the individual elements can be located physically much more closely together without mutual impedance interactions. Thus if the overall array size is small compared to a free space wavelength, the pattern of the array is very nearly the same as the pattern of a single element. The open circuit voltage expression is

$$\int E_1 \cdot i_a\, d\xi \doteq E_\rho N\lambda \cos\emptyset \quad \text{(series connection)}$$
$$\text{or} \qquad \doteq E_\rho \lambda \cos\emptyset \quad \text{(parallel connection)}. \qquad (44)$$

Since the pattern of an array is the same as that of a single element, it might seem that a small array has no advantage over a single element. This is not so, however, as can be illustrated by calculating the relative communication efficiency for an array compared to a single element. For the parallel connected array, it follows from (15), (16) and (44) that the RCE of a

small array compared to a single element (current i_s) is

$$RCE = \frac{R_1}{R_N} \frac{\left|\int E_1 \cdot i_a d\xi\right|^2}{\left|\int E_1 \cdot i_s d\xi\right|^2} = \frac{R_1}{R_N}$$

Thus, if the elements can be situated so that their mutual resistance is negligible, the resistance of the N parallel connected elements is $R_N \doteq \frac{R_1}{N}$ and so the relative communication efficiency is proportional to the number of elements.

In view of this, a very important piece of information is the mutual impedance of the elements as a function of their spacing in a lossy medium. This problem has been attacked theoretically and experimentally. Measured values of the impedance between two identical, parallel, side by side, resonant dipoles is shown in Fig. 14. The dipoles were at least three skin depths under the surface. Fig. 15 shows the mutual impedance of similar dipoles close to the surface. From these figures it appears that the mutual resistance is negligible when the spacing is something over two skin depths. (Since a skin depth can be a small fraction of a free space wave length, this implies that many elements can be employed and still have an array which does not violate the small array criterion.)

The unique aspect of subsurface arrays is that up to a certain width with a certain minimum spacing, an efficiency gain of N times is achieved with N elements with essentially no change in horizontal plane pattern (directivity). Studies indicate that a width of approximately 0.2 free space wavelength with element spacing of two skin depths are accurate limits for this condition (elements being resonant dipole elements near the surface). It is also worthy of note that the relative communication efficiency in the direction off the ends of the elements is increased by a factor of N no matter how wide the array is made (neglecting feed line problems). This observation can be important in point to point communication systems.

The efficiency increase of small arrays has been verified experimentally by field strength measurements. The results are included in Table II, which also gives a comparison of experimental and calculated values of RCE for other antenna types.

11. Near Field Considerations

In some of the applications of submerged antennas, the frequency may be so low that the transmitter and received are separated by a distance of the order of a wavelength or less. If this is so, many of the foregoing results are not applicable, since they are based on expressions for the distant fields. The same general procedure can be followed if the antennas are nearby, but near field expressions must be employed. Some specific results have been obtained[5] based on Wait's[4] formulas for the near fields of horizontal point dipoles.

12. Noise Considerations

As the foregoing discussions indicate, submerged antenna structures are quite inefficient. If a submerged antenna is employed as a receiving antenna, however, the inefficiency may not be a major consideration. This is so because the comparative worth of receiving antennas is determined by the signal to noise ratio and impedance which they present to the receiver input. In the low frequency bands, the most important sources of noise are usually electrical disturbances (storms) in the atmosphere. Most of the noise fields reach the receiver via a propagation mode in which the fields are related as indicated in equations (2) and (3).

Moreover, the noise sources are typically effectively distributed almost uniformly in the horizontal plane. Consequently the response of submerged horizontal antennas in these noise fields is different from that of the vertical antennas, because of the difference in the horizontal plane patterns. If the submerged antenna is designated, t, and the vertical antenna designated, s, as in the discussion above, it is easy to show that the ratio of the noise powers (generated by uniformly distributed noise sources) delivered to a matched load is

$$\frac{P_{rec,t,noise}}{P_{rec,s,noise}} = \frac{1}{2\pi} \int_0^{2\pi} RCE_{t,s}(\phi)\, d\phi$$

The main item of interest is the signal to noise ratio. The relative signal to noise ratio can be written

$$\frac{\left(\frac{S}{N}\right)_t}{\left(\frac{S}{N}\right)_s} = \frac{P_{rec,t,signal}}{P_{rec,s,signal}} \times \frac{P_{rec,s,noise}}{P_{rec,t,noise}} =$$

$$RCE_{t,s}(\phi_s) \cdot \frac{2\pi}{\int_0^{2\pi} RCE_{t,s}(\phi)\, d\phi} ,$$

where ϕ_s is the azimuthal bearing to the signal source relation to the end of the horizontal antenna. For many submerged structures, the field pattern varies as $\cos\phi$. ($\cos^2\phi$ or $\cos^3\phi$ for some, e.g. grounded end dipole at second resonance, Table I). In this case the integral

$$\int_0^{2\pi} RCE_{t,s}(\phi)\, d\phi \doteq \pi\, RCE_{t,s,max}$$

Then

$$\frac{\left(\frac{S}{N}\right)_t}{\left(\frac{S}{N}\right)_s} = 2\, \frac{RCE_{t,s}(\phi_s)}{RCE_{t,s,max}}$$

The conclusion is that if the submerged horizontal antenna is oriented so that the maximum in its pattern is directed toward the signal source, then the horizontal antenna delivers twice the signal to noise ratio as the vertical antenna, independent even of the depth of submersion (since the noise and signal are equally attenuated). This is true (only) as long as atmospheric noise greatly exceeds antenna noise, which is commonly the case at low frequencies.

The noise power delivered by a submerged hori-

zontal wire antenna is implied by the foregoing equation:

$$P_{rec,pwr,noise} = \frac{1}{2} RCE_{hor,vert} \cdot P_{rec,vert,noise}$$

The noise power of an array is given by the same type of expression provided the pattern of the array is substantially cosinusoidal. If the pattern is different, the term RCE must be divided by the directive gain (relative to the cosine pattern).

In practice, noise levels measured with submerged antennas have been found to be lower than those expected on the basis of the discussion above, even with a very small depth of submersion. The reason for this, especially in stormy periods, may be that some of the noise on a vertical antenna may be impulsive, due to charged particles striking the antenna. Of course, such charged particles cannot directly strike a submerged antenna.

References

1. Stratton, J.A., Electromagnetic Theory, pg 573-587, McGraw Hill, 1941
2. Sommerfeld, A., Partial Differential Equations, pg 246-257, Academic Press, 1949
3. Wait, J.R., J. of App. Phys. 24 pg 958-959, 1953
4. Wait, J.R., Canadian Jour. of Physics, 39, 1017-1028, July 1961
5. Weeks, W.L. & Fenwick, R.C. "Flush or Buried Antenna Structures," Collins Radio Company Research Report #257, Feb. 1962
6. For example, Schelkunoff, S.A., "Electromagnetic Waves," pg 49, D. Van Nortrand, New York, 1943
7. Sinclair, G., Proc. IRE 36, 1364-1370, Nov. 1948
8. Sunde, E.D., "Earth Conduction Effects in Transmission Systems," D. Van Nortrand, New York, 1949
9. Moore, R.K., "Theory of Radio Communication between Submerged Submarines," Ph.D. Dissertation at Cornell University, 1951
10. Wheeler, H.A., N.B.S. Journal of Research 65D pg 89-91, Feb. 1961

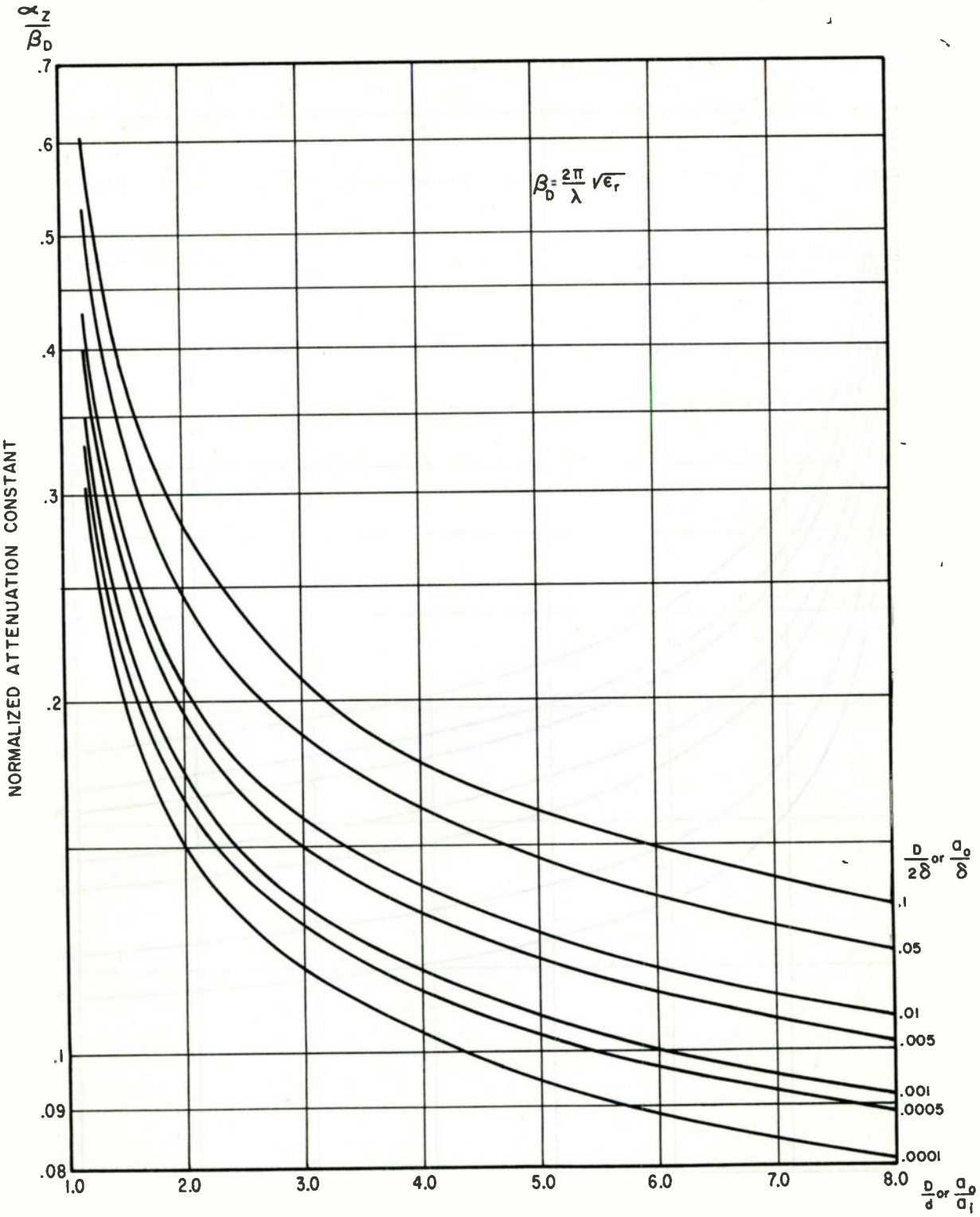


Fig. 1. Normalized attenuation constant on insulated wires in a conducting medium.

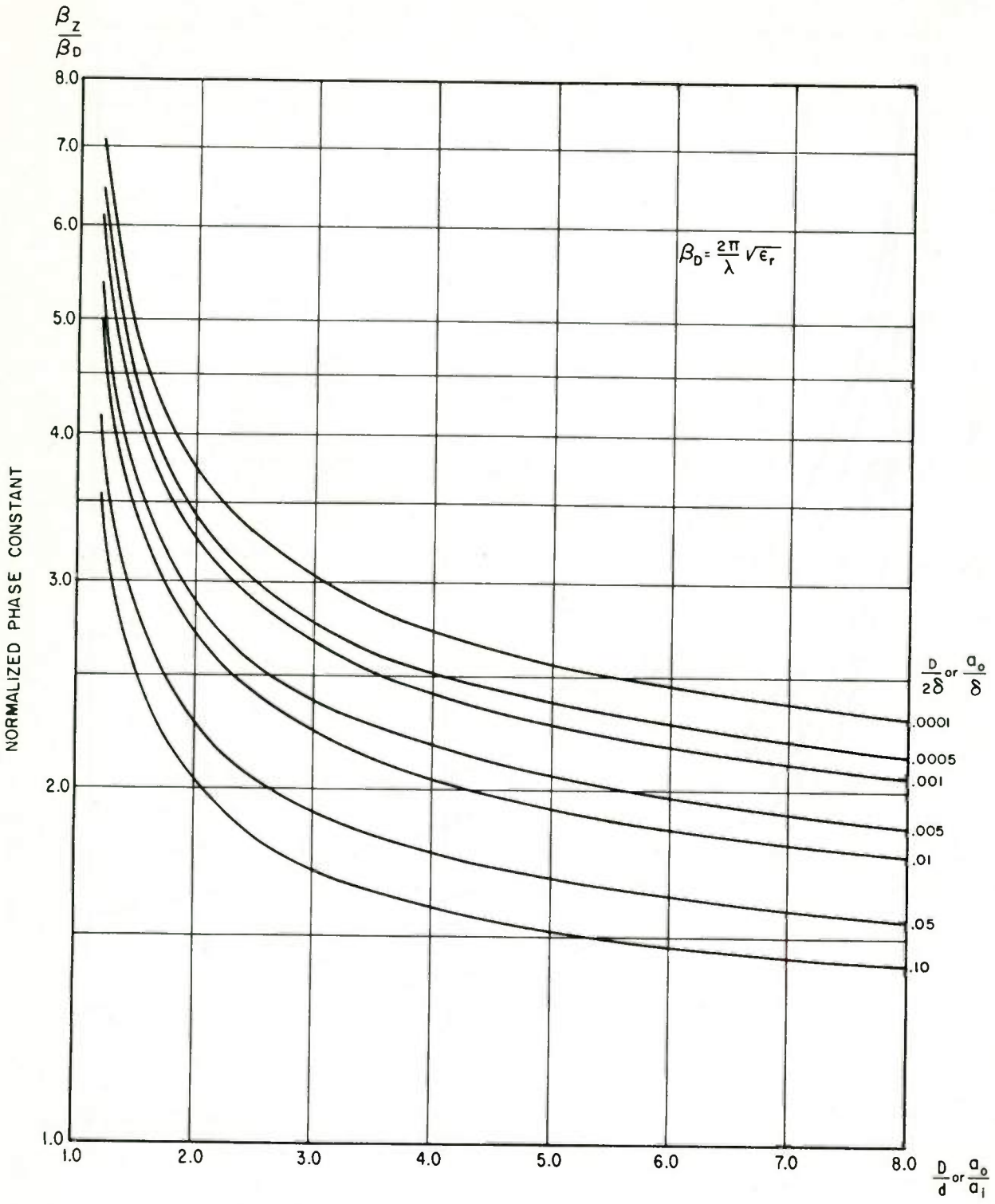


Fig. 2. Normalized phase constant on insulated wires in a conducting medium.

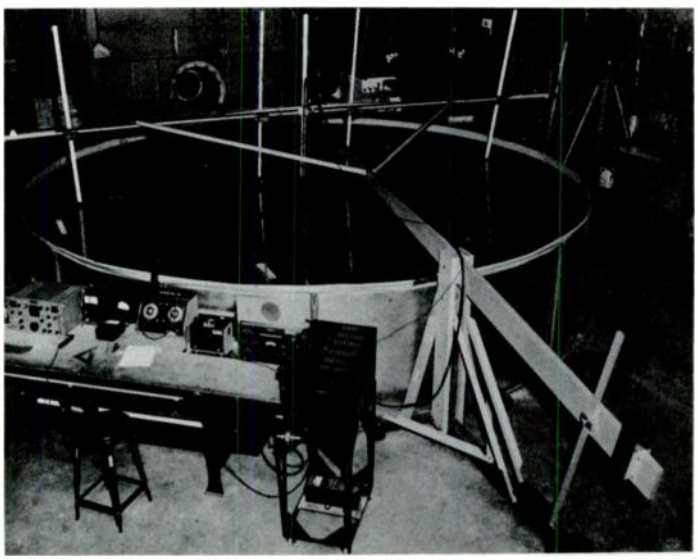


Fig. 3. Antenna modeling facility utilizing a salt water tank.

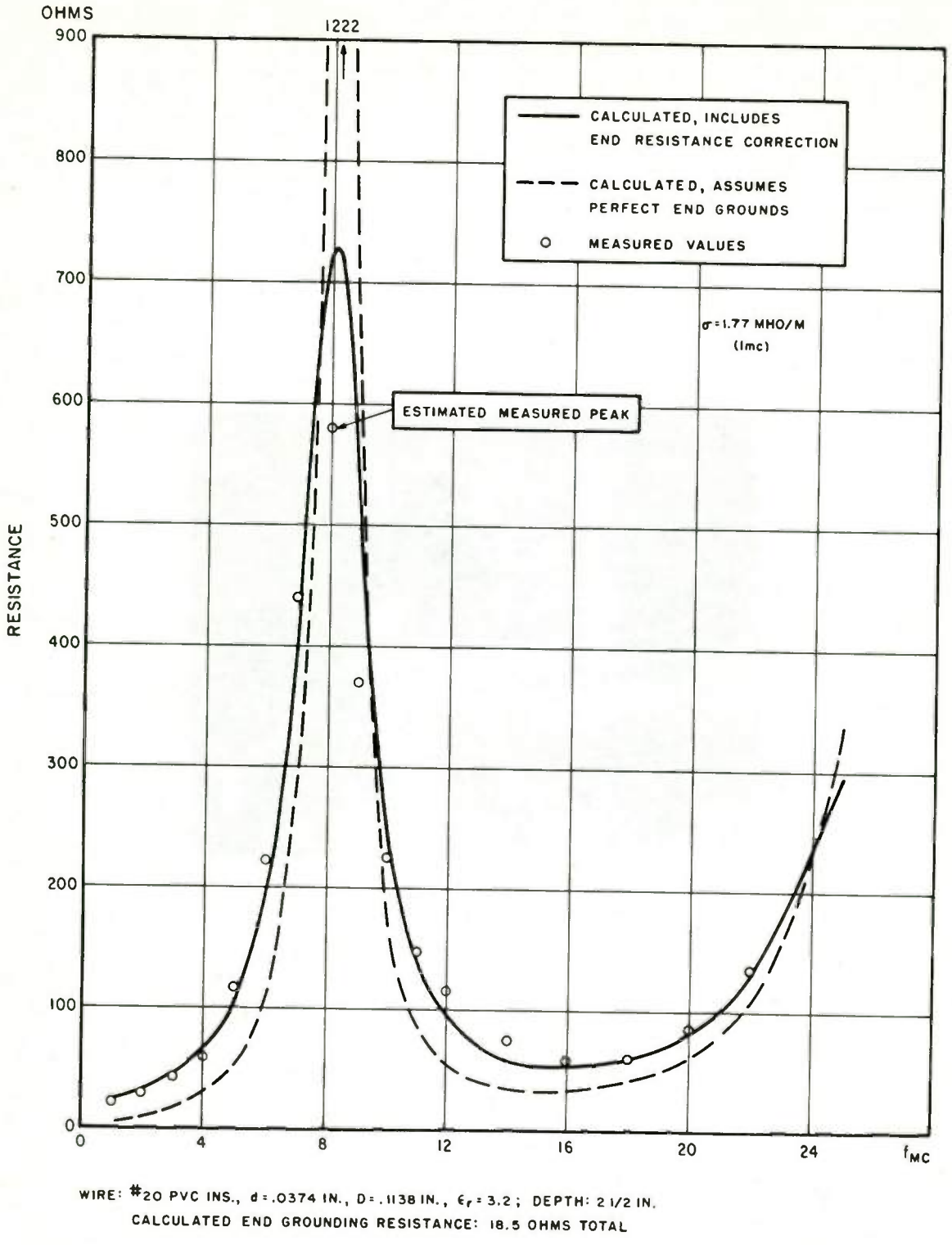


Fig. 4. Comparison of measured and theoretical resistance of a grounded end 15 foot dipole in salt water.

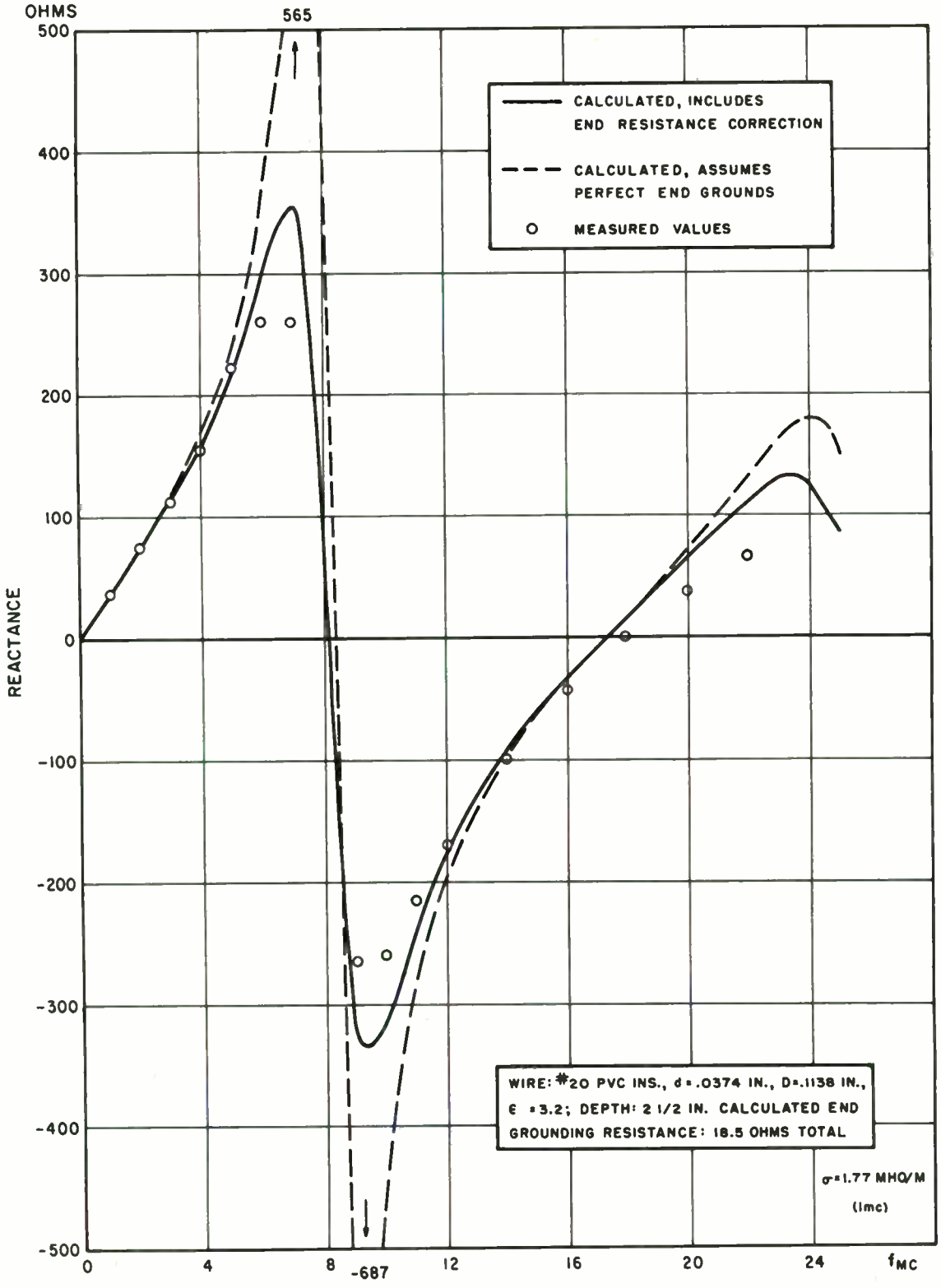


Fig. 5. Comparison of measured and theoretical reactance of a grounded end 15 foot dipole in salt water.

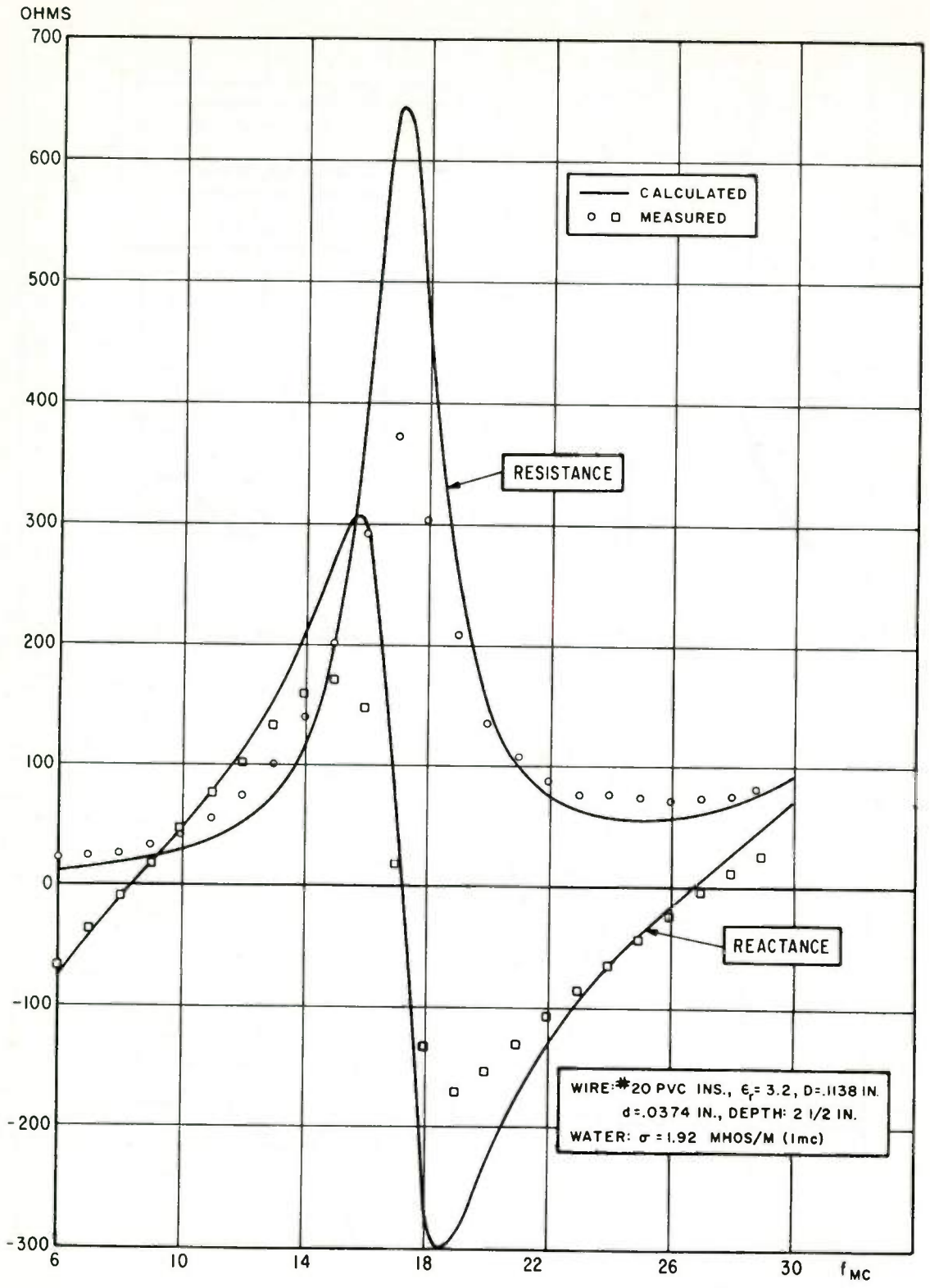


Fig. 6. Comparison of measured and theoretical impedance of an open ended 15 foot dipole in salt water.

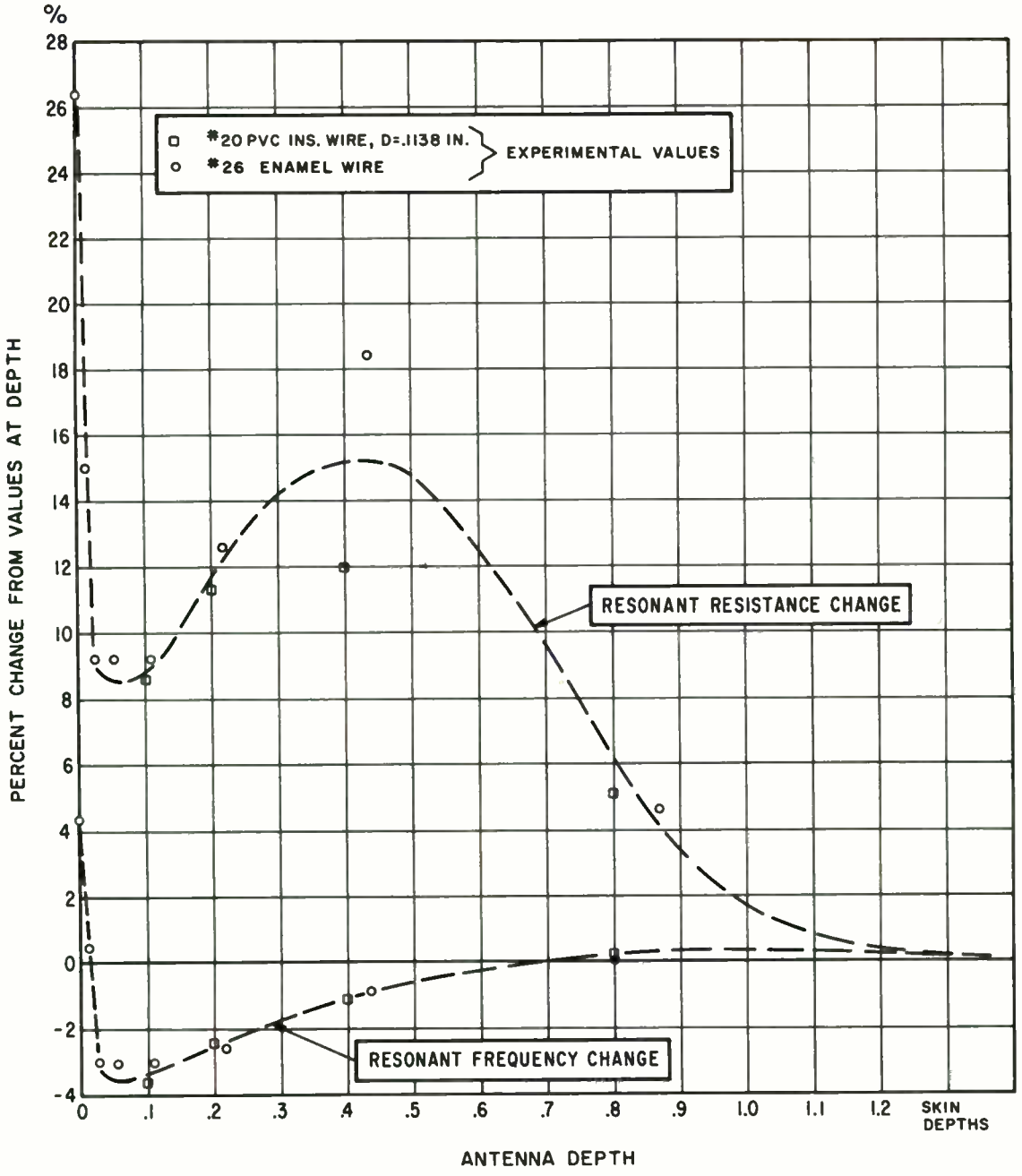


Fig. 9. Measured change of resonant resistance and frequency of a 15 foot dipole vs depth in salt water.

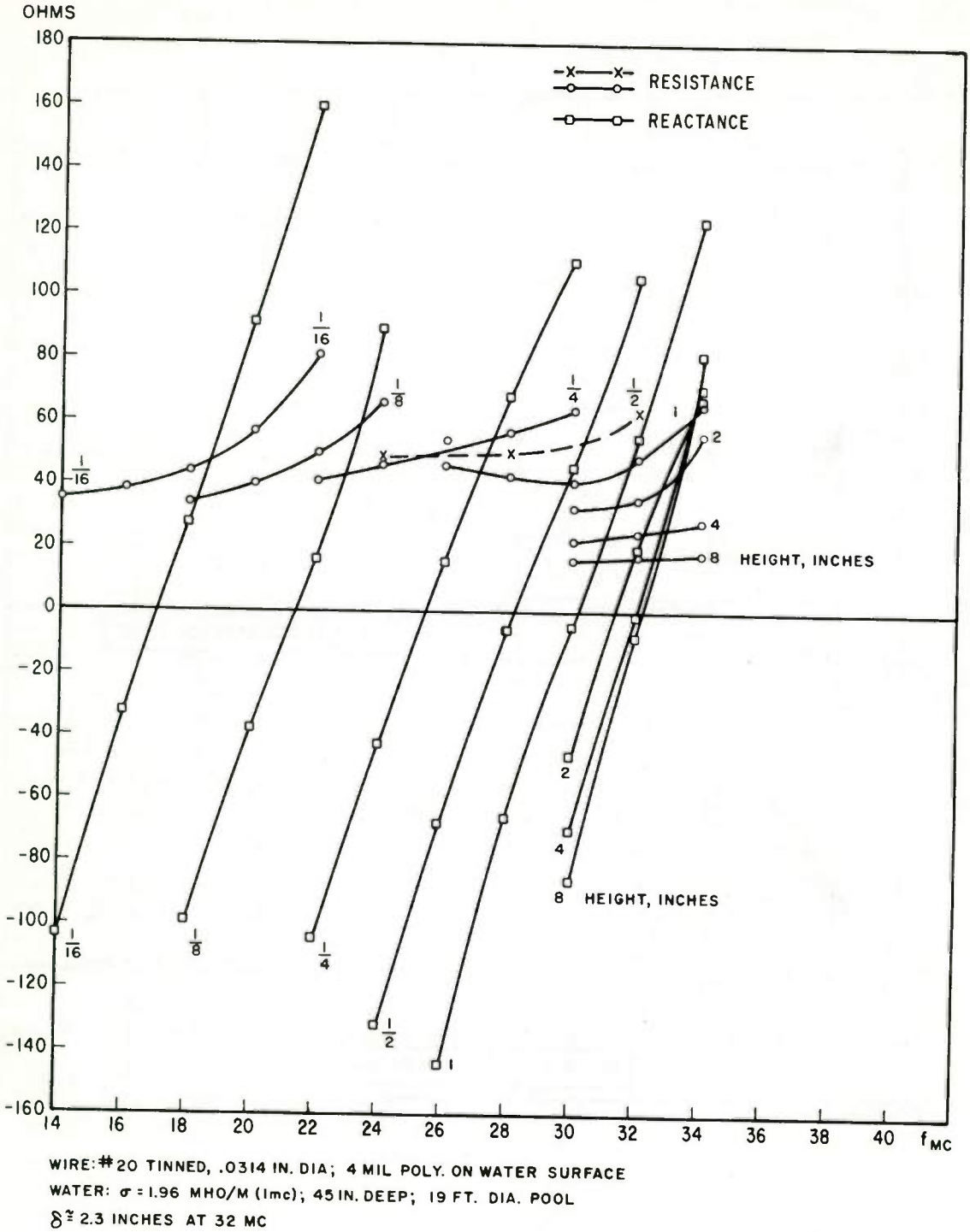


Fig. 10. Measured impedance of a 15 foot dipole vs height above salt water.

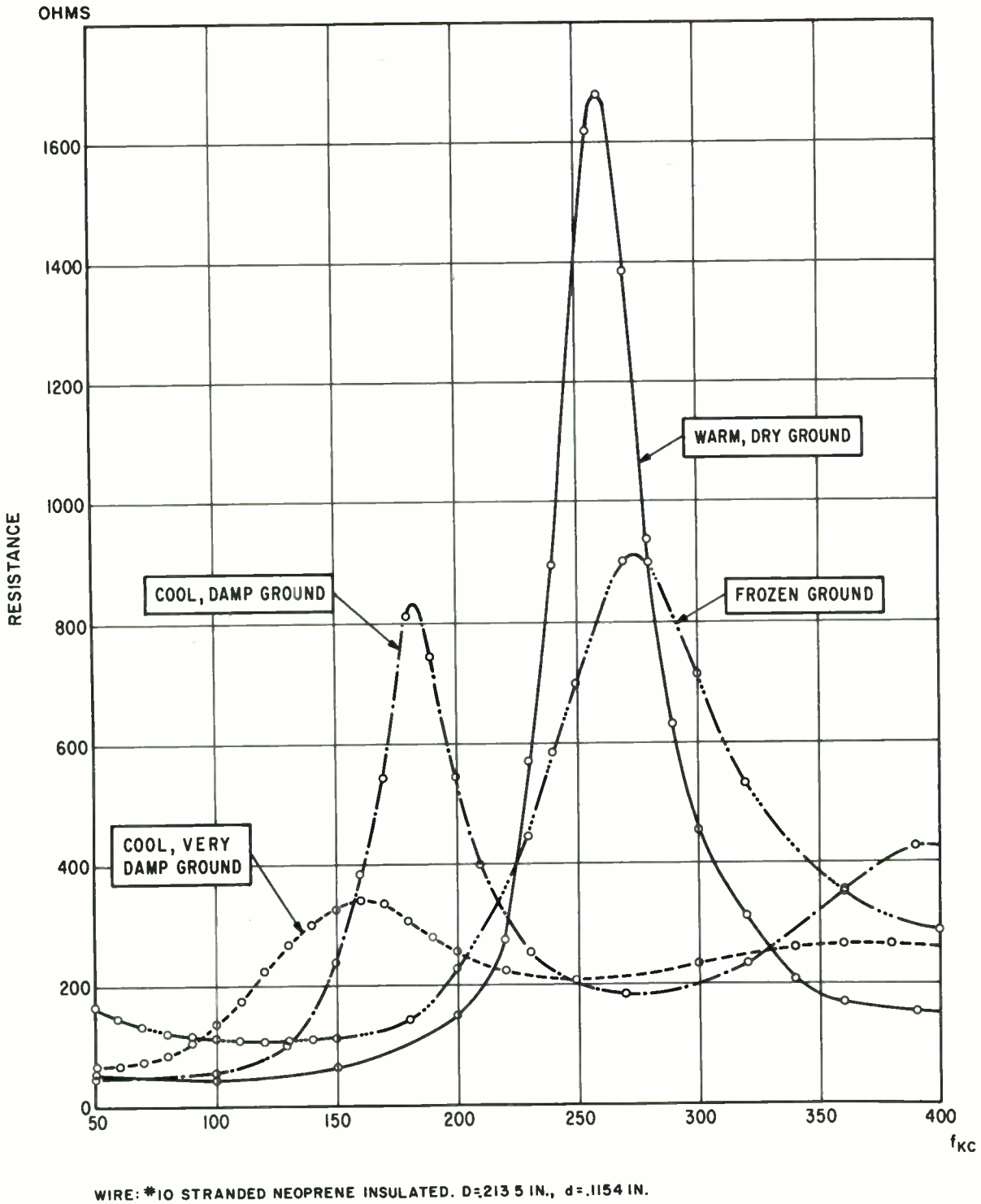


Fig. 11. Resistance of a 1500 foot dipole lying on the ground.

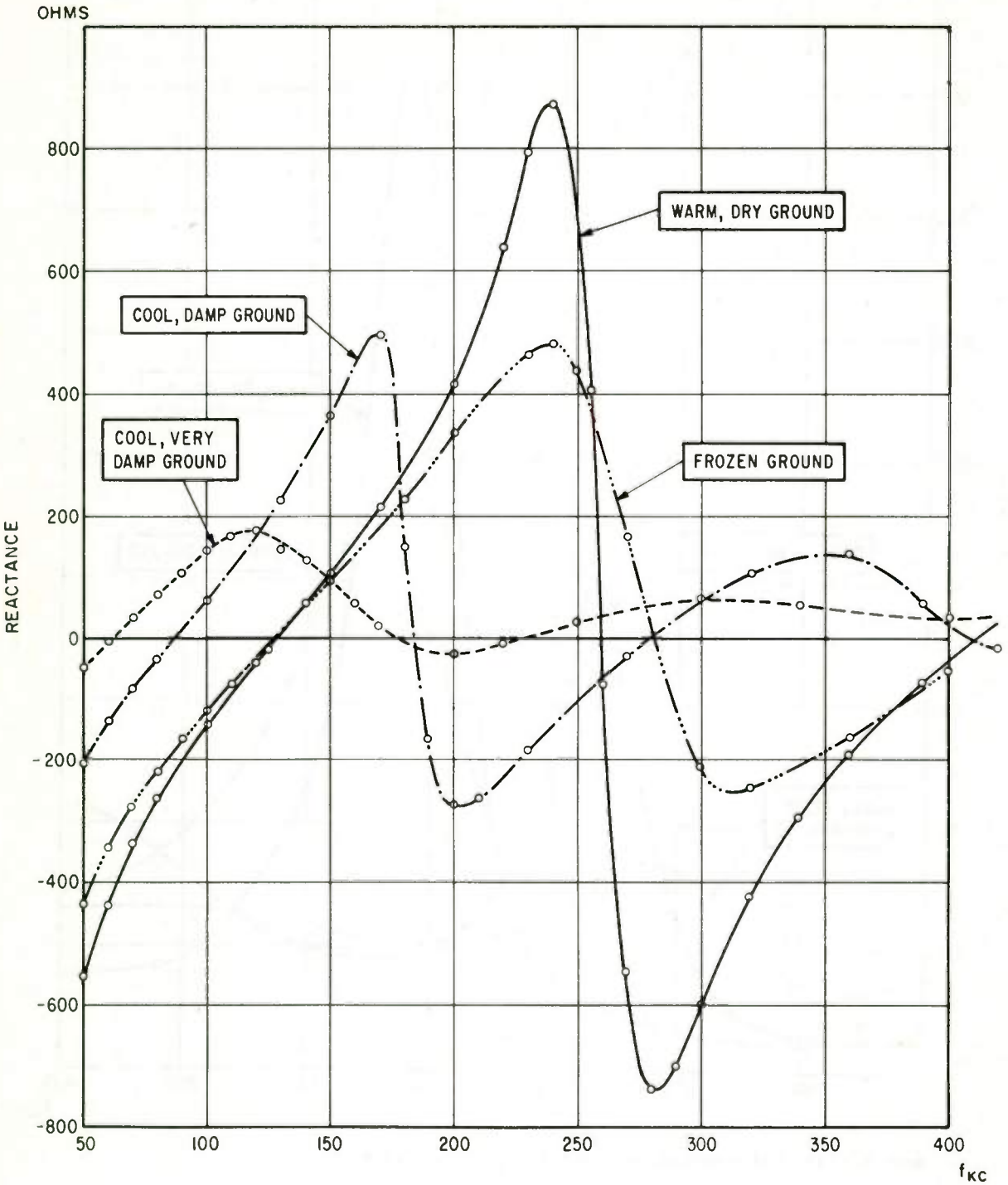


Fig. 12. Reactance of a 1500 foot dipole lying on the ground.

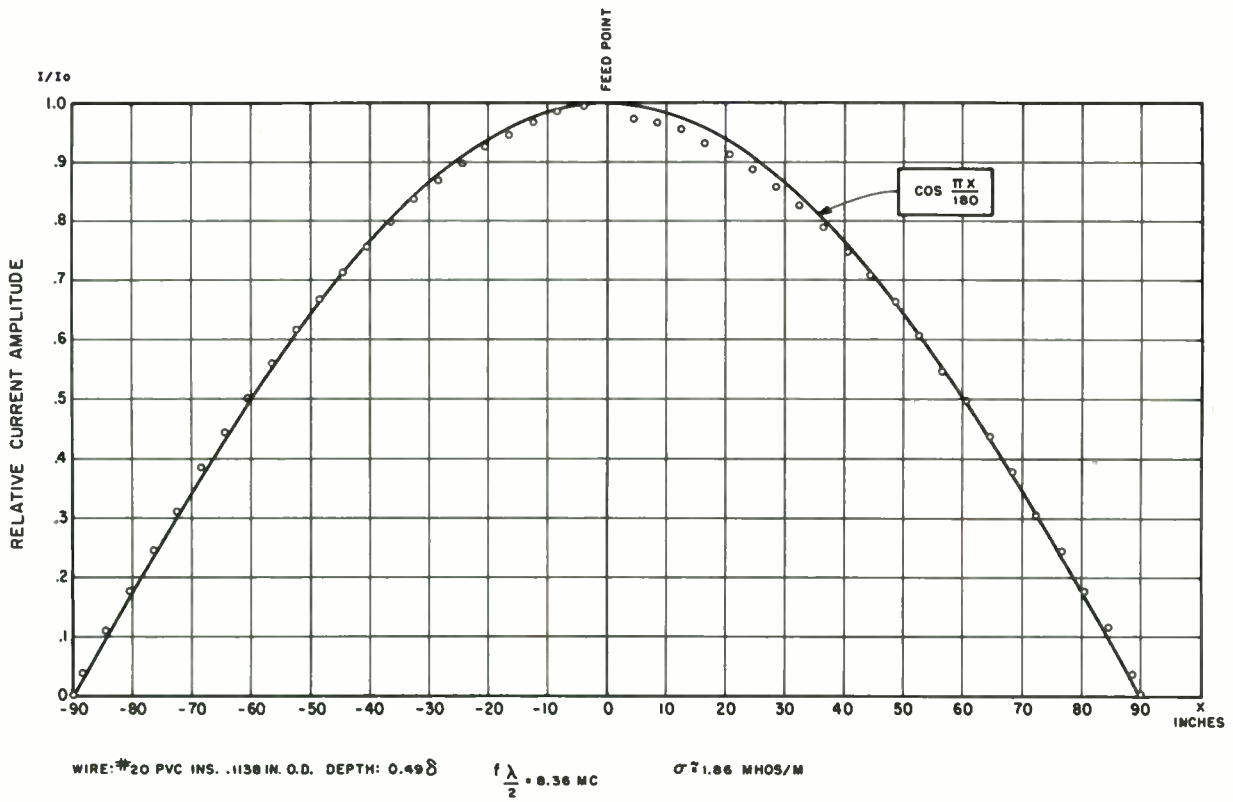


Fig. 13. Measured current distribution on a resonant dipole in salt water compared to a cosine curve.

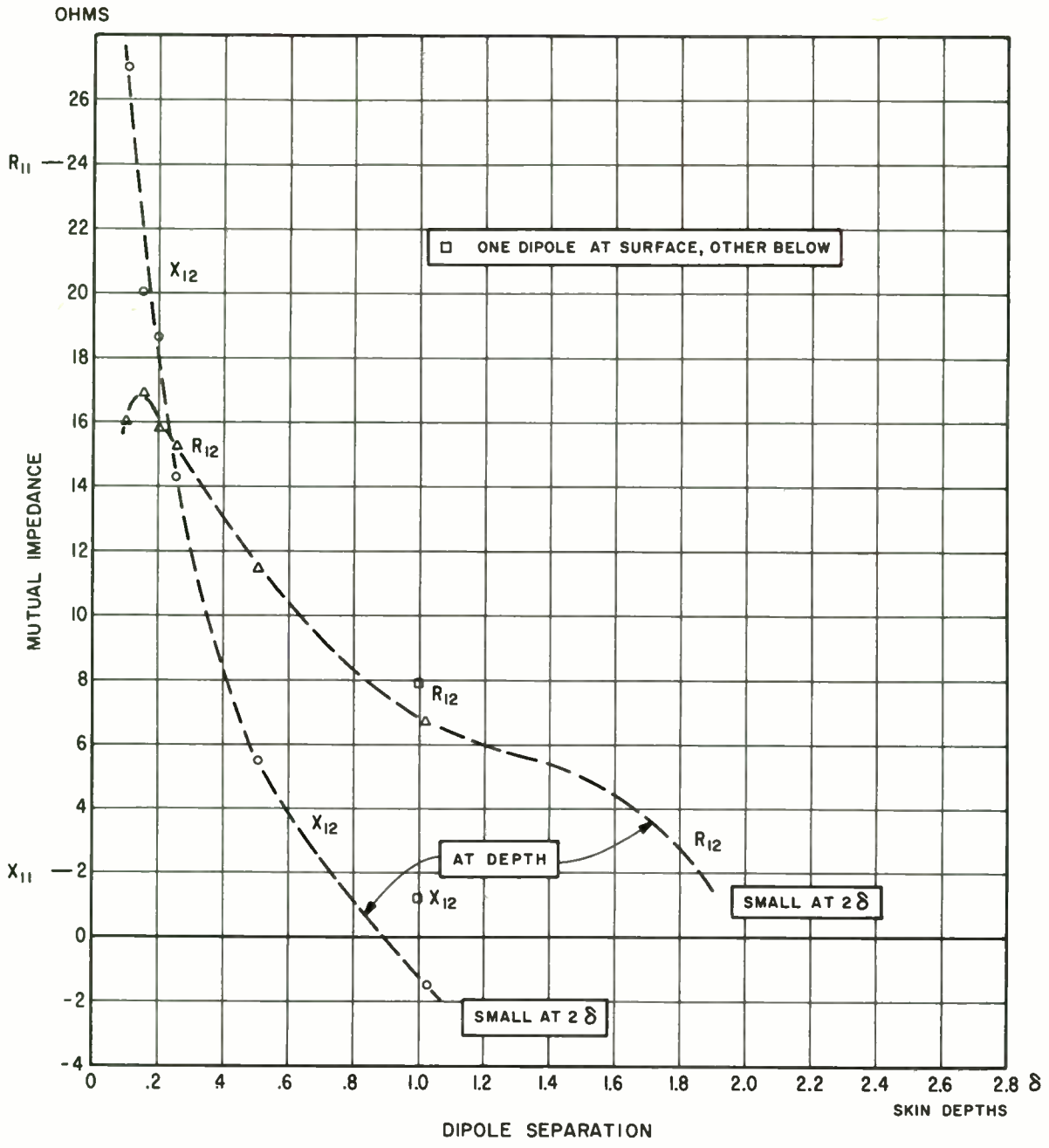


Fig. 14. Mutual impedance of parallel side-by-side half-wave dipoles in salt water.

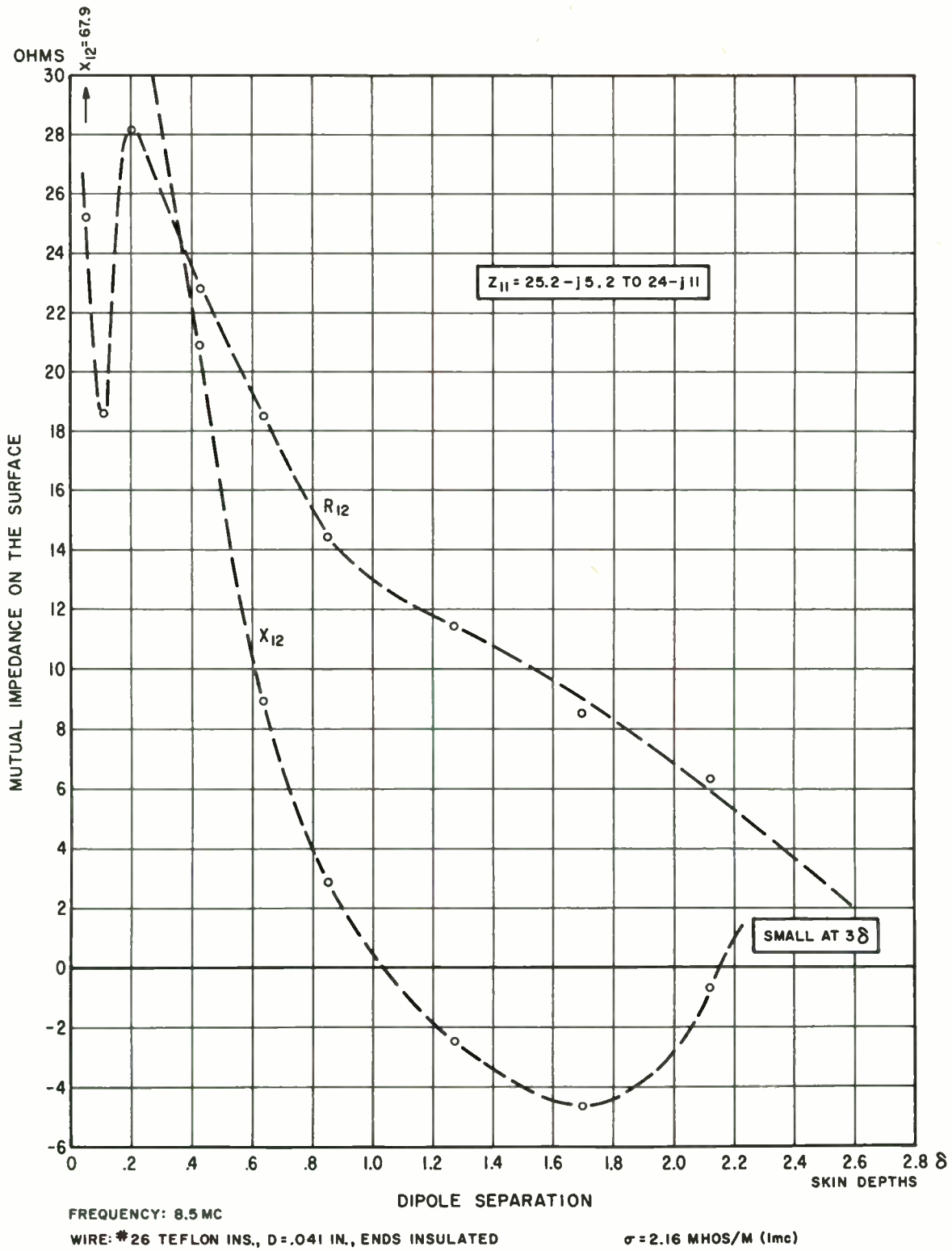
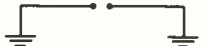
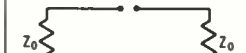


Fig. 15. Mutual impedance of parallel side-by-side half-wave dipoles on the surface of salt water.

Table I. Relative Communication Efficiency of Basic Subsurface Antennas. Reference: Perfect Quarter Wave Vertical

LENGTH \ TYPE						
SHORT $L \ll \lambda_{air}$	$\frac{9}{R}\left\|\frac{I}{\eta}\right\|^2(\beta_0 L)^2$ $f < f_{1/2}$	$\frac{9}{R}\left\|\frac{I}{\eta}\right\|^2(\beta_0 L)^2$ $f < f_{1/2}$	$\frac{36}{R}\left\|\frac{I}{\eta}\right\|^2(\beta_0 L)^2$ $f < f_{1/4}$	$\frac{36}{R}\left\|\frac{I}{\eta}\right\|^2(\beta_0 L)^2$ $f < f_{1/4}$	$\frac{36}{R}\left\|\frac{I}{\eta}\right\|^2 \frac{\beta_0^2}{\left\|2\alpha_z+i2\beta_z+i(\beta_z-\beta_0)L+i(\beta_z+\beta_0)e^{-[\alpha_z+i(\beta_z-\beta_0)]L}\right\|^2}{\beta_0^2+\alpha_z^2-\beta_z^2+i2\alpha_z\beta_z}$	$\frac{36}{R}\left\|\frac{I}{\eta}\right\|^2 \beta_0^2 \left\|\frac{e^{-[\alpha_z+i(\beta_z-\beta_0)]L}-1}{-\alpha_z-i(\beta_z-\beta_0)}\right\|^2$
INTERMEDIATE (Any Length Where $\alpha_z L$ Is Small)	$\frac{144}{R}\left\|\frac{I}{\eta}\right\|^2\left(\frac{F}{1-F^2}\right)^2\left(\frac{\cos\frac{\beta_0 L}{2F}-\cos\frac{\beta_0 L}{2}}{\sin\frac{\beta_0 L}{2F}}\right)^2$	$\frac{36}{R}\left\|\frac{I}{\eta}\frac{F}{1-F^2}\left(\frac{e^{j\beta_0 L}-jF\sin\frac{\beta_0 L}{F}-\cos\frac{\beta_0 L}{F}}{\sin\frac{\beta_0 L}{F}}\right)\right\|^2$	$\frac{144}{R}\left\|\frac{I}{\eta}\right\|^2\left[\frac{I_{max}}{I_0}\frac{F}{1-F^2}\left(\sin\frac{\beta_0 L}{2F}-F\sin\frac{\beta_0 L}{2}\right)\right]^2$	$\frac{36}{R}\left\|\frac{I}{\eta}\right\|^2\frac{I_{max}}{I_0}\frac{F}{1-F^2}\times\left[\sin\frac{\beta_0 L}{F}+jF\left(e^{j\beta_0 L}-\cos\frac{\beta_0 L}{F}\right)\right]^2$		
FIRST ORDER RESONANT	$\frac{144}{R}\left\|\frac{I}{\eta}\right\|^2\left(\frac{F}{1-F^2}\cos\frac{\pi F}{2}\right)^2$	$\frac{36}{R}\left\|\frac{I}{\eta}\right\|^2\left(\frac{F}{1-F^2}\right)^2(1+F^2-2F\sin\frac{\pi F}{2})$	$\frac{144}{R}\left\|\frac{I}{\eta}\right\|^2\left[\frac{I_{max}}{I_0}\frac{F}{1-F^2}(1-F\sin\frac{\pi F}{2})\right]^2$	$\frac{36}{R}\left\|\frac{I}{\eta}\right\|^2\left(\frac{I_{max}}{I_0}\frac{F}{1-F^2}\right)^2(1+F^2-2F\sin\frac{\pi F}{2})$		
TRAVELING WAVE	$\frac{144}{R}\left\|\frac{I}{\eta}\frac{\beta_0(\alpha_z+j\beta_z)}{\beta_0^2+\alpha_z^2-\beta_z^2+j2\alpha_z\beta_z}\right\|^2$	$\frac{36}{R}\left\|\frac{I}{\eta}\frac{\beta_0}{\alpha_z+j(\beta_z-\beta_0)}\right\|^2$	$\frac{144}{R}\left\|\frac{I}{\eta}\frac{\beta_0(\alpha_z+j\beta_z)}{\beta_0^2+\alpha_z^2-\beta_z^2+j2\alpha_z\beta_z}\right\|^2$	$\frac{36}{R}\left\|\frac{I}{\eta}\frac{\beta_0}{\alpha_z+j(\beta_z-\beta_0)}\right\|^2$		
SECOND ORDER RESONANT	$\frac{144}{R}\left\|\frac{I}{\eta}\right\|^2\left(\frac{I_{max}}{I_0}\frac{F}{1-F^2}\right)^2(1+\cos\pi F)^2$	$\frac{144}{R}\left\|\frac{I}{\eta}\right\|^2\left(\frac{I_{max}}{I_0}\frac{F}{1-F^2}\right)^2(1+\cos\pi F)$	$\frac{144}{R}\left\|\frac{I}{\eta}\right\|^2\left(\frac{F^2}{1-F^2}\sin\pi F\right)^2$	$\frac{144}{R}\left\|\frac{I}{\eta}\right\|^2\left(\frac{F^2}{1-F^2}\right)^2(1+\cos\pi F)$		

NOTES: 1. R IS TOTAL INPUT RESISTANCE.
2. L IS TOTAL LENGTH.
3. DEPTH LOSS $e^{-2h/\delta}$ IS OMITTED.
4. TO INCLUDE AZIMUTHAL DEPENDENCE SUBSTITUTE $\beta_0 \cos\phi$ FOR β_0 AND $F\cos\phi$ FOR F. $\phi = 0$ FOR RADIATION OFF THE UNFED END OF END FED WIRES.
5. $F = \frac{\text{RESONANT FREQUENCY}}{\text{RESONANT FREQUENCY IN AIR}} = \frac{\beta_0}{\beta_z}$
6. I_0 IS FEEDPOINT CURRENT.
7. f_1 IS THE FIRST ORDER RESONANT FREQUENCY.
8. EXCEPT FOR TRAVELING WAVE OPERATION THE EFFECT OF α_z ON CURRENT DISTRIBUTIONS IS ASSUMED SMALL.
9. $\beta_0 = 2\pi/\lambda_{air}$

Table II. Summary of RCE Measurements. Reference: Perfect Quarter Wave Vertical

ANTENNA TYPE	ASSUMED SOIL CONDUCTIVITY mhos/m	MEASURED RCE	CALCULATED RCE USING MEASURED R	MEASUREMENT METHOD AND ASSUMPTIONS
1500 foot dipole at 125kc open ends (first resonance at 137kc)	10^{-2}(Cedar Rapids)	-38 db	-36.6 db	F.S. Measurements; 1500 ft. dipole used for transmitting
348 foot end fed, grounded end, at 24kc. Electrically short	4.8(Pt.Loma) ocean	-79.6 db	-79.4 db (using calculated impedance)	Reception of station NPG; F.S. measured with whip. Azimuthal orientation not precise.
100 foot end fed, grounded end, at 342kc, first resonance at 970kc	$4x10^{-2}$(Dallas)	-44.1 db	-45.5 db	Reception of station DAL, F.S. measured using similarly oriented loop. Uniform current assumed.
As above, 210kc	$4x10^{-2}$(Dallas)	$\cong$ -48 db (signal weak)	-50.2 db	As above, station unknown.
As above, open ended, at 342kc	$4x10^{-2}$(Dallas)	-49.7 db	-51.4 db	Method as above, linear current assumed.
As above, grounded end, at 820kc	$4x10^{-2}$(Dallas)	-38.6 db	-38 db[1] (using calculated α_z)	Methods as above; station WBAP. Measured shortening factor used.
As above, open ended	$4x10^{-2}$(Dallas)	-39.7 db	-39 db	As above.
As above, grounded end, at 860kc	$4x10^{-2}$(Dallas)	-40.0 db	-37 db[1] (using calculated α_z)	Method as above, station unknown.
As above, open end	$4x10^{-2}$(Dallas)	-41.7 db	-38 db	As above
4 element array of 50 ft. dipoles spaced 14' apart, at 4.8mc	-	+6.5 db relative to a single dipole element	+6 db	Average current of array elements was made same as single element. F.S. Measurements were made at 3λ off the antenna ends. Mutual coupling was small. 6db and 3db assumed power increases have been subtracted.
2 element array, as above	-	+3.5db relative to a single element	+3 db	

[1]This assumes the measured resistance, which is lower than would be the case with perfect grounds; the assumed current neglects the effect of finite grounding resistance, the inclusion of which would tend to decrease the calculated efficiency.

FM BACKSCATTER SOUNDING AS A MEANS FOR MONITORING PROPAGATION CONDITIONS DURING SHORT-WAVE BROADCASTS

Robert B. Fenwick
O. G. Villard, Jr.

Radioscience Laboratory
Stanford University
Stanford, California

Summary

Employment of megawatt power levels in short-wave broadcasting, the declining sunspot cycle, and the growing number of broadcasters competing for spectrum space make it increasingly desirable for transmitter personnel to know when unusually low MUF's prevent a given broadcast from reaching its destination. Affected transmitters can be either shut off, thus saving running costs, or retuned to propagating frequencies. Backscatter sounding is known to be a satisfactory means for determining propagation conditions along the transmission path, but owing to the practical problems of integrating such sounding into broadcasting operations, has not been appreciably used as yet. It is shown that application of the FM ranging principle to backscatter sounding is feasible, permitting the broadcasting carrier itself to be used for sounding purposes without disturbance to transmitter operating conditions or appreciable annoyance to the distant listener. It is necessary that an associated receiving site within the ground-wave range of a given transmitting plant be available. Useful sounding measurements can be made with the aid of equipment normally found in any receiving station. In fact, an operator can gauge skip distance by estimating the pitch of an audible tone. On balance, this technique appears to have significant practical advantages in comparison with the pulse or modulation-correlation alternatives.

Introduction

During much of the time, in international short-wave broadcasting, transmissions are beamed towards a specific "target", which may represent a comparatively localized area. (For example, the B.B.C. aims transmissions at the East Coast of the U.S.A., while simultaneously broadcasting in the same waveband--but using different antenna direction--towards Central America.) According to present-day practice, broadcasting schedules are set up months in advance in accordance with predicted ionospheric conditions. To ensure that the listener receives a signal, in the face both of normal departures from predicted conditions, and ionospheric storms, it is customary to transmit the same program simultaneously in at least two, and sometimes three, adjacent wavebands. Thus the listener can choose whichever frequency reaches him best. From the standpoint of the broadcaster, this comparatively uneconomical procedure is suitable so long as the cost of running parallel transmissions is not exorbitant, and so long as there exists an adequate supply of frequency assignments.

This state of affairs is now in the process of being changed by pressure on frequency assignments caused both by the increasing number of nations desiring to conduct shortwave broadcasting, and by the decrease in MUF's accompanying the approaching sunspot minimum. In addition, many broadcasters are now raising power to the quarter or half megawatt level,[1] so that power and tube charges are becoming an increasingly large fraction of the total running expense. (The staff needed to operate a 250-kilowatt station is normally no larger than that needed for a 50-kilowatt counterpart.)

In these circumstances it is desirable to consider the use of backscatter sounding as a means for determining, at the transmitter, whether a given broadcast is reaching its destination. The feasibility of such a procedure is well established[2,3], but the most practical technique for incorporating such sounding into normal broadcasting operations has not yet been determined, and the only use of backscatter operations thus far appears to have been experimental. It is the purpose of this article to propose the use of the FM radar technique for backscatter sounding, as an alternative to the pulse and correlation radar techniques which have already been explored.[4,5] It is believed that the FM technique has certain practical advantages and that it should not be difficult to conduct such soundings in a routine manner as a part of normal broadcasting operations.

Uses of Backscatter Sounding in Broadcasting

Backscatter sounding can assist practical broadcasting operations in the following three main ways (presented in the order of increasing technical complexity):

a) It can determine when a given frequency has ceased to propagate at all (i.e., F4000 MUF is less than the operating frequency).

b) It can determine whether a broadcast aimed at a comparatively nearby target is "skipping over" that target, owing to lower-than-expected ionospheric bending power, and

c) It can--at least during the nighttime hours--determine whether transmission to a target at a given multi-hop distance is comparatively strong.

It should be pointed out that the above determinations are by no means infallible; backscatter sensing can be "spoofed" by unusual ionospheric conditions such as SID's and sporadic E. However, backscatter can be expected to give correct answers during a high percentage of the time.

With the aid of this information, a broadcast operator can do the following:

a) If a given broadcast is clearly missing its target, or if--due to lowered MUF's--transmission at that frequency has ceased entirely, the transmitter might as well be shut down so as to save power.

b) If alternative, lower frequencies are available, the transmitter can, of course, be retuned to such a frequency and transmission resumed. At the present time, broadcast schedules are worked out in advance by international consultation so as to avoid mutual interference, and no provisions are made for alternate frequencies in the event of poor ionospheric conditions. Although providing alternate frequencies will undoubtedly be difficult, owing to the spectrum shortage, there seems to be no reason, in principle, why it could not be done.

c) If no lower, alternative frequencies are available, it may be possible to retune the transmitter whose frequency has failed to a frequency which is propagating and run it in parallel with an existing transmission on that frequency, either for the purpose of increasing the total power radiated to a given target area, or for the purpose of widening the effective target area being covered at a given power level.

d) If a broadcast is going out at two different frequencies at different power levels, it may be possible to rearrange the transmitters to advantage. For example, a normal procedure might be to use 50 kilowatts at 15 Mc and a 250-kilowatt transmitter at 17 Mc. If the backscatter shows that 17 Mc has failed, due to low MUF's, but that 15 Mc is getting through, it will obviously be of advantage to use the high power at 15 Mc.

Periods of exceptionally low MUF generally last from several hours to days; hence operators have adequate time in which to reschedule the deployment of transmitters.

e) In the event that transmission is to multi-hop distances, and control of the vertical directivity of the transmitting beam is available, it may be possible, using backscatter as the indicator, to maximize the power delivered to a distant target area. This may be done by adjusting the vertical antenna pattern so as to concentrate the radiated power in the range of vertical angles most effective at any given time in delivering signal to the distant receiver.

Practical Problems of Measuring Backscatter at Broadcasting Plants

At first thought it might seem that the best technique for measuring transmission conditions by means of backscatter would be to install a separate sounder, either fixed or variable-frequency. However, the difficulty immediately arises that if such a sounder is installed at a transmitting plant, it will be blocked by the outgoing broadcasts. On the other hand, if it is installed at an associated receiving station, interference to receiving operations can be expected from the sounding transmissions. To install and man an entirely separate sounding station would be prohibitively costly.

Because broadcasting beams are comparatively narrow, and propagation conditions are not, in general, the same in all directions from a given location, it is desirable to use the broadcasting antenna for sounding, rather than a less-expensive, less-directive sounding antenna. Since broadcasting beam antennas are large and expensive structures, to duplicate them for sounding purposes would be uneconomical.

Three techniques for obtaining range information are well known: pulsing, frequency modulation (FM), and modulation-correlation. The first and last of these have been previously explored for sounding purposes[4,5], but to the best of the authors' knowledge FM has not.

It would be a convenience if reception for sounding purposes could be conducted at the transmitting plant. Unfortunately, this does not seem possible unless it were acceptable to pulse simultaneously all the transmissions being radiated by a given station. Only in this way could blocking of a sounding receiver be prevented.

Comparison of Sounding Techniques

Since most transmitting plants at the present time have nearby associated receiving stations, for relaying purposes, it seems wisest to locate sounding receivers at these receiving sites.

In choosing a technique for ranging purposes, it is necessary to take the following considerations into account:

(1) The effect of sounding on normal transmitting operations.

(2) The effect of sounding on the distant listener.

(3) The length of time required to obtain a measurement, and

(4) The relative complexity of the receiving and recording equipment required.

Items (1) through (3) are important, because it is desirable to sound comparatively frequently throughout the broadcast day. Brief intervals of sounding during program breaks for station identification should suffice, since the time constant of the F layer for normal electron-density changes is on the order of one hour.

A comparison of the three techniques is given in summary form in Table I. It can be seen that on-off pulsing at the transmitter presents adjustment complications which, while not insuperable, are nonetheless troublesome. From this viewpoint, either correlation or FM sounding is certainly preferable, since the former offers no transmitter complications at all, and the latter can be carried on without disturbing the normal amplitude modulation of the station in any way.

Description of FM Sounding

A wide variety of modulating waveforms could be applied to amplitude-modulated carriers for

Table I Comparison of the Pulse, FM, and Modulation-Correlation Techniques for Obtaining Ionospheric Backscatter Information During Short-Wave Broadcasts

	PULSE RADAR	FM RADAR	MODULATION-CORRELATION RADAR
Effect on transmitter operation	Troublesome--requires change of bias levels, disconnection of modulator; changes load on the power supply.	No effect	No effect
Effect on distant listener during sounding intervals	Troublesome--pulses very annoying owing to receiver gain increase due to AGC action.	No change in average gain level, but FM will produce a perceptible "growl".	No effect
Relative complexity of receiving and recording arrangements required	Requires receiver modified for pulse reception; oscilloscope; camera or special magnetic tape for recording.	Uses unmodified receiver; standard tape recorder for recording; operator's ear gives first-order answer.	Comparatively complex; requires either variable delay device for correlation, or multiple delay devices.
Length of time required for measurement	Very short--instantaneous measurements possible	Very short--instantaneous measurements possible.	May be long if program material does not contain suitable frequency components.

ranging purposes. For example, an audio modulating impulse could be applied to the carrier through the ordinary modulator, so that ranging could be done by means of the sideband energy. However, this technique would produce an uncomfortably strong audio signal at the listener's receiver, and is accordingly undesirable.

Since home-type short-wave receivers are designed to respond to AM, it is possible to reduce annoyance caused to the listener by making the ranging modulation a frequency modulation of the transmitted carrier. Although various alternatives are possible, it is convenient from the receiving equipment standpoint to frequency-modulate the transmitter with a triangular waveform, and observe at a nearby site the audio-frequency beat note generated when the backscattered energy combines with some of the outgoing signal. The distribution in slant range of the backscattered energy can then be determined from a measurement of the frequency spectrum of the audio-frequency beat. The strongest backscatter returns from a slant range close to that of the edge of the skip zone, for the usual situation in h-f broadcasting where the operating frequency is well above the f2-layer critical frequency. Thus a first-order one-hop skip distance determination can be made simply by taking note of the strongest component of the audio-frequency beat signal, and estimating or measuring its pitch. This component will also--in most cases--be the lowest-pitch component present.

For the case of a single target having no appreciable depth in slant range, FM radar operation is as illustrated in Fig. 1. The transmitter frequency variation and the time variation of the beat frequency f_b generated by a target at a range of r kilometers, are shown. It will be observed that there is a "turn-around" transient generated each time the direction of the transmitter frequency change reverses. In practice, the duration of this transient is so short that the effect on the ear (or on a frequency analyzer) is negligible. The over-all effect is as if a continuous tone were present.

Additional targets at slant ranges other than the one shown will produce additional audio-frequency tones having pitches proportional to the target slant ranges.

Choice of Frequency-Modulation Parameters

The following considerations determine the choice of parameters for frequency modulation of the above type for the broadcast-sounding application. First, the echo slant range interval of greatest practical interest extends from a minimum range of about 750 km, to a maximum of approximately 3500 km for one-hop propagation. (If multi-hop working is of interest, the maximum range will be extended accordingly.) Second, during the sounding intervals, it is desirable to confine the bandwidth of the transmission--insofar as this is possible--to the normal channel width of about 5 kcs, in order to minimize adjacent-channel interference. Third, it is desirable that the frequency modulation not be any more noticeable at the distant receiver than is necessary.

The average home short-wave receiver can be assumed to have an IF passband whose response is comparatively flat over an interval 5 kc wide. In spite of this, even a small frequency modulation of the transmitted carrier will cause an audio output to appear at the receiver because of non-uniformities in the transmission-versus-frequency

characteristic of the ionospheric circuit, caused principally by the effect of multipath transmission. (These same transmission non-uniformities are responsible for selective fading.) The voltage waveform which appears at the second detector of a distant receiver will consist of the fundamental of the transmitter sweep frequency and its harmonics. If the fundamental sweep frequency can be kept low, not too much energy will fall within the audio passband of a typical receiver, which normally has a low-frequency cutoff on the order of 100 cps.

As a consequence of these considerations a reasonable choice would be to sweep 2000 cps repetitively with a triangular waveform of frequency 2.5 cps. This represents sweeping at the rate of 10,000 cps[2]. The round trip time delay associated with a typical backscatter echo at 1000 km slant range (about 750 km actual range, assuming F2-layer propagation) is 1/150 second. In this time, the transmitted frequency would have shifted some 67 cps; thus the audio beat note corresponding to an isolated target at that slant range would consist of bursts of 67 cps tone lasting for a time interval equal to approximately one-fifth of a second. Each burst would contain roughly thirteen cycles. There would be "gaps" between bursts--each 1/150 second long--due to the round trip travel time of the echo energy, as shown in Fig. 1. The width of the gap corresponds to roughly 1/2 cycle at 67 cps.

An echo at double the slant range produces a beat note of twice as high a frequency. The propagation time delay is doubled (1/75 sec), but this is still small compared with the time of one sweep (1/5 sec). Thus the total number of cycles per burst has been increased.

An echo at 4000 km slant range (normally the maximum for one-hop backscatter) will produce a tone consisting of bursts of 268 cps energy, intertupted by 1/37 sec gaps. Thus the audio-frequency waveform will consist of repetitive 46 cycle pulses of 268 cps energy separated by gaps whose width is approximately 7 cycles of 268 cps energy.

The wider the gap in the bursts of audio tone, the more difficult it will be for a spectrum analyzer to determine the pitch of that tone with accuracy. This circumstance sets an upper limit on sweep rate for targets at a given range, if the interval in the spectrum over which frequency is to be swept is held constant.

Thus in practice a compromise has to be found between frequency interval swept (which determines adjacent-channel interference, and--in part--the audibility of the FM at the distant receiver) and sweep rate, which determines the pitch of the resulting tone, and its degree of contamination by turn-around transients. (Sweep rate will also affect annoyance given to the distant listener and --to some extent--transmitted bandwidth.)

In the preceding example, an audio beat note of 67 cps corresponded to a slant range of 1000 km. This is an inconveniently low frequency. By recording the signal on a standard magnetic tape recorder at 1-7/8 i.p.s., and playing back at 7-1/2 i.p.s., this frequency can be increased to the readily-audible value of 268 cps. 4000-km energy then produces a tone of 1072 cps.

Degree of Annoyance Given to a Distant Listener

As mentioned earlier, non-uniformities in the transmission versus frequency curve of the ionosphere circuit will cause an effective increase in the fading rate of the signal, as received by the distant listener, during frequency modulation. This will show up as a low frequency "growl" with an audibility depending on the number of fading nulls present at the receiver per sweep of the transmitted signal.

The practical effect of this may be estimated from the following. A low-power 12 Mc experimental c-w transmission from Puerto Rico was frequency modulated at a 6 cps rate with 660 cps total deviation. This transmission was monitored at Stanford, California (5750-km distant), and at Smyrna, Georgia (2420-km). In each case the signal was translated down to the audio frequency range, and frequency-amplitude-time plots made. Examples are shown in Fig. 2.

Each frequency sweep provides one vertical line on the record. The intensity of the line is determined by the signal amplitude at each frequency represented along the line. Dark areas are intervals of low signal amplitude.

It is found that approximately one null per sweep is normally present on the longer path, and appreciably less than one null per sweep on the shorter path. Assuming one null per sweep as a representative number, a signal frequency-modulated over a 2 kc interval might average 3 nulls per sweep. If the sweep rate is 5 per second, the fundamental frequency produced at the distant receiver detector would be about 15 cps. Taking harmonics of this waveform into account, frequencies up to 150 cps may be expected to appear in the output of the receiver.

Occasionally, up to 3 nulls per sweep have been observed in the 660 cps band on the longer path. The record of Fig. 2a is unusual, and shows two. For the Puerto Rico to Georgia path, between zero and one null per sweep is normal. Thus the record in Fig. 2b is typical, with slightly less than one null per sweep.

It seems clear from these results that the distant listener will be conscious of the FM, and that the noticeability of the FM will be greater the longer the path. However, such sounding should not be too annoying if carried on for 10 to 15 seconds at a time, at 15-minute intervals.

Performance of a Typical Installation in the Presence of Ionospheric Backscatter

Operation of the FM ranging system with backscatter as the target was tested at a Stanford University field site at 1529 PST on 3 October 1961. A radio frequency of 15.3 Mcs was used. Transmitting and receiving antennas were 3 element Yagis directed north. The transmitter output power was 600 watts, and the receiving site was 6 miles north of the transmitting site.

The FM deviation in the experimental system was 2 kc, but the modulation waveform was a 7-1/2 cps triangular waveform instead of the 2.5 cps waveform as shown in Fig. 1. The effect of this difference is to increase the pitch of the audio frequencies obtained by a factor of three, and to increase the effective width of the turn-around gaps by a factor of three. A frequency-amplitude-time plot of the audio spectrum obtained is shown in Fig. 3. Note that the turn-around transients are not severe enough even in this case to deteriorate appreciably the quality of the record.

To the ear, the signal of Fig. 3 sounds noisy, but it none-the-less has a clearly-defined pitch.

Frequency-Modulating the Transmitter

Any of the standard frequency-modulating techniques can be used to obtain the desired waveform. In the experimental case described above a reactance-tube modulator, driven by a Hewlett-Packard waveform generator, was used to modulate a variable frequency oscillator. In the case of short-wave broadcasting, where crystal control is necessary, an arrangement such as that illustrated in Fig. 4 could be used. The output of a crystal oscillator is mixed with a low-frequency variable oscillator which is frequency-modulated in the desired way. Extreme linearity of frequency versus time is not required in the system proposed.

Receiver Considerations in the FM System

In FM backscatter sounding, the receiver must be able to produce a beat between the transmitted signal (normally available by ground-wave transmission between transmitter and receiver), and backscattered energy. In principle, a single receiver and antenna can perform this function, provided that the ground-wave in this antenna happens to be appreciably stronger than the backscatter. If the ground-wave is not adequately strong, backscatter components at different ranges will intermodulate, thus giving rise to spurious audio-frequency components in the receiver output. Too much ground-wave pickup will cause loss of sensitivity in the receiver, since the over-all gain must then be reduced in order to avoid overloading of the diode detector or the last IF stage.

If the normally-used antenna at the receiving station picks up too weak a ground-wave, selection of an antenna headed more nearly toward the transmitter may improve the situation. The strength of the backscatter coming back along the direction of fire of the transmitted beam is normally so great that no particular receiving antenna directivity is required. For multi-hop sounding, however, it may be desirable to use a receiving antenna directed in the same azimuth as the transmitter beam. In this event, if the ground-wave is not strong enough, it may be possible to add a second antenna, beamed directly at the transmitter, in which ground-wave pickup is maximized. (Use of vertical polarization may be helpful.)

The output of the "reference"--or ground-wave--antenna may be mixed with that of the "signal"--or backscatter--antenna at the receiver input terminals, as shown in Fig. 5. The over-all layout is shown in Fig. 6.

On the other hand if the ground-wave signal at the receiver is too strong, it may be possible to employ an antenna having a null in the direction of the local transmitter, while maintaining a reasonable gain in the direction from which the backscatter is arriving. If this fails, more separation between receiving and transmitting stations may be required.

It is desirable--although not necessary--for the receiving signal antenna of Fig. 5 to be directive to control ground-wave signal strength, and to suppress backscatter from undesired directions caused by transmitting antenna side-lobes. (To the extent that the transmitting antenna is undirectional, the backscatter received will be from the target area.)

Probably the most difficult situation arises when the receiving station is located "behind" the transmitting antenna, and the ground-wave is too strong. This situation will require the signal and reference antennas to be pointing in the same direction, toward the transmitter. In this case, discrimination between the backscattered energy and the ground-wave energy can be done on the basis of vertical angle of arrival and/or polarization. In a situation where the receiving station is in the direction of the transmitter beam, and the ground-wave is strong, the reference receiving antenna should have a high front-to-back ratio.

Reducing Intermodulation and Estimating Ground-Wave/Backscatter Intensity Ratio

At the expense of some equipment complication, the problem of obtaining the correct amount of demodulating-signal injection can be largely overcome by the scheme shown in Fig. 7. (This arrangement was actually used in the experimental test of Fig. 3, although the simpler one-receiver technique has also given good results.) In this arrangement, the antenna aimed at backscatter (upper channel) is designed to pick up as little ground-wave as possible. The ground-wave channel (the lower one) has its gain adjusted to prevent overloading.

An especially simple, practical embodiment of Fig. 7 employs two inexpensive receivers, Collins type 75S1, which happen to have conversion oscillator voltages brought out to external terminals.

The conversion oscillators of the reference receiver are disabled; the reference receiver mixers are fed by the signal receiver oscillators. The antenna for the signal receiver (upper channel) was directed toward the distant area illuminated by the transmitting antenna, while the reference receiver antenna (lower channel) was directed toward the transmitter. IF output of the reference receiver replaced the BFO input to the product detector of the signal receiver, thereby providing a strong demodulating voltage that was relatively free from backscatter components. The output of the product detector is then essentially free from intermodulation components, as may be seen in Fig. 3.

The amount of ground-wave signal present on a particular antenna at a given receiving site within the skip-zone of the transmitter may be readily estimated by noting the fading characteristics of the received signal as displayed by the usual "S" meter. If the signal fades regularly to zero, there is comparatively little ground-wave present. On the other hand, if the range of signal amplitude variation during fading is limited, the backscatter is comparatively weak compared with any ground-wave simultaneously present. This determination can be made comparatively quickly in practice, since measurements have shown that the spectral width of ground backscatter is in the order of one to two cycles per second[6]. Moreover, the center of the backscatter spectrum is often displaced from the transmitted frequency by one or two cps[7]. These circumstances imply that a combination of ground-wave signal plus backscatter should fade through the normally-expected range of values within a few seconds' time.

Methods of Analyzing the Receiver Audio Output

In general, increasing amounts of information can be obtained from the audio output of an FM system at progressively increasing complexity and cost. The simplest possible analyzer is the human ear, which can tell whether backscatter is present or not, and estimate its approximate range. Considerably more precise information can be determined with the aid of visual or graphical spectrum analyzers, which provide quantitative information on the range-amplitude structure of the backscatter. If the received tone is too low-pitched to be heard, a magnetic tape recording of the tone may be made and speeded up on playback. The record of Fig. 3 was obtained by recording the receiver output on a $150 home-type tape recorder, and later analyzing it on a Kay Electric Company "Sonalyzer" spectrum analyzer. In the analysis the tape was speeded up 8 times to bring the spectrum within range of the instrument. However, the frequency scale shown in the figure corresponds to real time. It would be desirable, in broadcasting, to tape-record the received signal for "first order" aural analysis immediately, and to have the records on hand for later, more comprehensive, analysis by spectrum analyzer at some central location.

A useful method of analyzing the backscatter record is to pass the signal into a manually-tunable selective audio amplifier. The range interval(s) of maximum backscatter amplitude (presumably, the boundaries of the region(s) where the transmission is being heard the strongest) can be rapidly and easily determined in this way.

Another method of determining the range of the strongest component of the backscatter is to use an audio frequency meter. (The strongest component normally is found at a slant range close to that of the skip distance.) As a test of this, the signal illustrated in Fig. 3 was fed to a Hewlett-Packard 500B frequency meter. A comparatively steady reading, having a value very close to that of the strongest component of the backscatter determined from the sound spectogram, was obtained. Such a meter is inherently insensitive to other frequency components which may be present, provided only that they are weaker than the one being registered.

Conclusions

It has been shown that the FM ranging principle can be applied to backscatter sounding, and that this technique should be useful for sounding at brief intervals during short-wave broadcasting in order to obtain propagation data. FM ranging has the following significant practical advantages in comparison with the pulse or modulation-correlation alternatives:

1) No change of d-c operating conditions is required at the transmitter. (Many complications arise if the transmitter must be pulsed.)

2) FM modulation causes much less annoyance to the distant broadcast listener than would pulse modulation. (The nature of the effect of the FM on the listener's receiver has been assessed with the aid of records from actual FM tests over typical h-f braodcasting distances. It is concluded that the listener annoyance caused by such modulation, as received in a standard AM receiver, would be small.)

3) Much simpler receiving and recording arrangements are required for FM sounding than for either the pulse or modulation-correlation schemes. The operator can use an unmodified receiver; his ear can provide first-order answers. The simplest type of magnetic tape recorder suffices to record the information.

4) The FM system can obtain data nearly instantaneously, as can the pulse radar. The correlation radar may require a longer time if the modulating waveform does not have suitable frequency components.

In applying the FM principle to broadcasting, it is convenient to frequency-modulate the transmitter with a triangular waveform. Most broadcasting plants have associated receiving stations within ground-wave range. At such a station the beat note generated when backscattered energy combines with some of the ground-wave signal can be heard in a standard receiver. The resulting tone is "noisy", but nevertheless has a fairly well-defined pitch, since most backscatter will be concentrated at a range corresponding very nearly to the one-hop skip distance. Methods for controlling the groundwave signal needed for demodulating the backscatter have been shown.

Either the operator's ear, a simple tunable selective audio amplifier, or a direct-reading frequency meter can furnish first-order information on skip distance. Range-time records of pulsed-radar-like quality can be obtained by means of a standard audio spectrum analyzer. Actual FM system tests have shown that "turn-around" transients do not appreciably affect the final frequency analysis.

Experience suggests that a 2 kc deviation at a rate of 2.5 cps (triangular modulation) provides the best balance between information gained and annoyance caused to the listener. If the sounding modulation occurs during those intervals when program modulation is momentarily absent, there is no possibility of exceeding the spectral bandwidth normally occupied by the station.

On balance, FM backscatter sounding appears preferable to the alternatives in many respects.

Achnowledgement

The authors are indebted to Dr. Braulio Dueno of the University of Puerto Rico, Mayaguez, Puerto Rico, and to Mr. Clair Powell, Mr. Joseph Hawkins, Mr. Larry Peden, and Mr. T. A. Croft of the Radioscience Laboratory, Stanford University. Their assistance was very helpful.

This work was supported by the Office of Naval Research under contracts Nonr 225(24) and Nonr 225(33).

References

1. E. T. Martin and G. Jacobs, "The International Broadcasting System of the Voice of America," presented at IRE International Convention, New York; March, 1961.
2. O. G. Villard, Jr. and A. M. Peterson, "Scatter-sounding, a Technique for Study of the Ionosphere at a Distance," IRE Trans. on Antennas and Propagation, Vol. AP-3, pp. 140-142; 1952.
3. R. Silberstein, "Sweep Frequency Backscatter--Some Observations and Deduction," IRE Trans. on Antennas and Propagation, Vol. AP-2, pp. 56-63; April, 1954.
4. E. D. R. Shearman, "An Investigation of the Usefulness of Backscatter Sounding in the Operation of h-f Broadcast Services," Proc. Inst. Elect. Engrs., Vol. 108B, pp. 361-374; February 1961.
5. L. C. Edwards and D. A. Hedlund, "Cozi Oblique Incidence Ionospheric Soundings Using Normal Communications Transmissions," IRE Trans. on Communications Systems, Vol. CS-8, pp. 160-164; September 1960.
6. P. W. Carlin, "A Technique for Displaying the Time Variation of the Spectral Distribution of the Fading Fluctuations and Doppler Shifts of Ionospherically-Propagated Ground Backscatter in Selected Small Range Intervals," Radio Propagation Laboratory, Stanford University, Stanford, California, TR No. 5, July 31, 1959. Contract Nonr 225(33).
7. H. Hoogasian, D. B. Odom, and J. F. Roche, "Investigation of Variations in the Ionosphere with the use of Phase Coherent Backscatter Techniques," presented at URSI-IRE fall meeting, Austin, Texas; October 23, 1961.

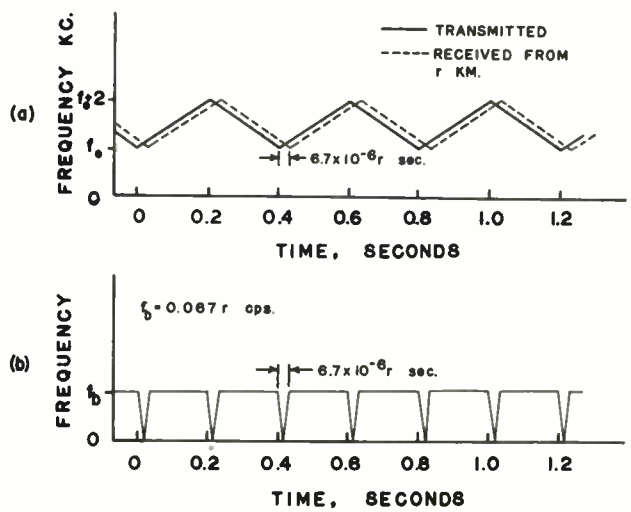


Fig. 1. (a) Waveform of transmitted frequency as a function of time suitable for an FM ionospheric radar. Also shown is received frequency as a function of time for a single-range target at r kilometers slant range. (b) Audio output frequency as a function of time for a receiver receiving both waveforms in (a).

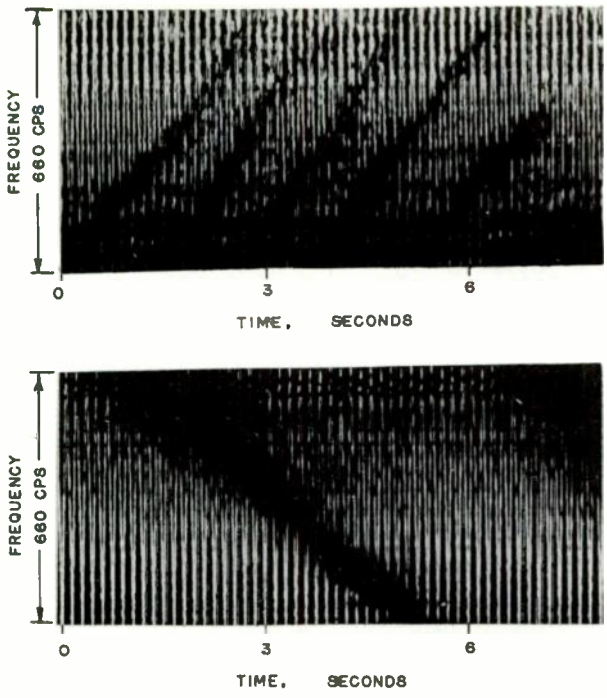


Fig. 2. (a) Frequency-time-amplitude record of FM signal transmitter at 12 Mc over 5750 km path. (b) Similar record taken for a 2420 km path.

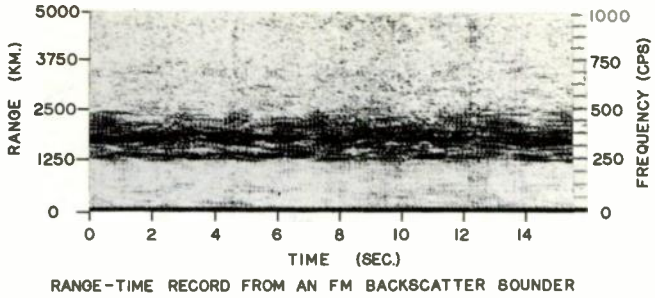


Fig. 3. Spectrum analysis of an experimental FM ionospheric backscatter sounding.

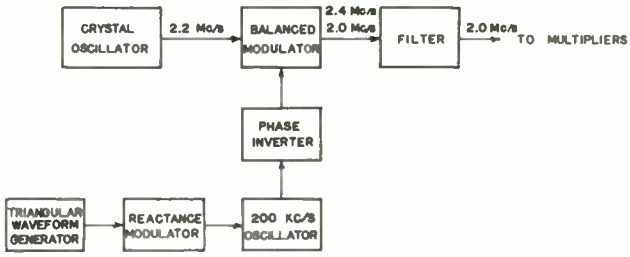


Fig. 4. Block diagram of a system providing stable triangular frequency modulation yet with crystal control.

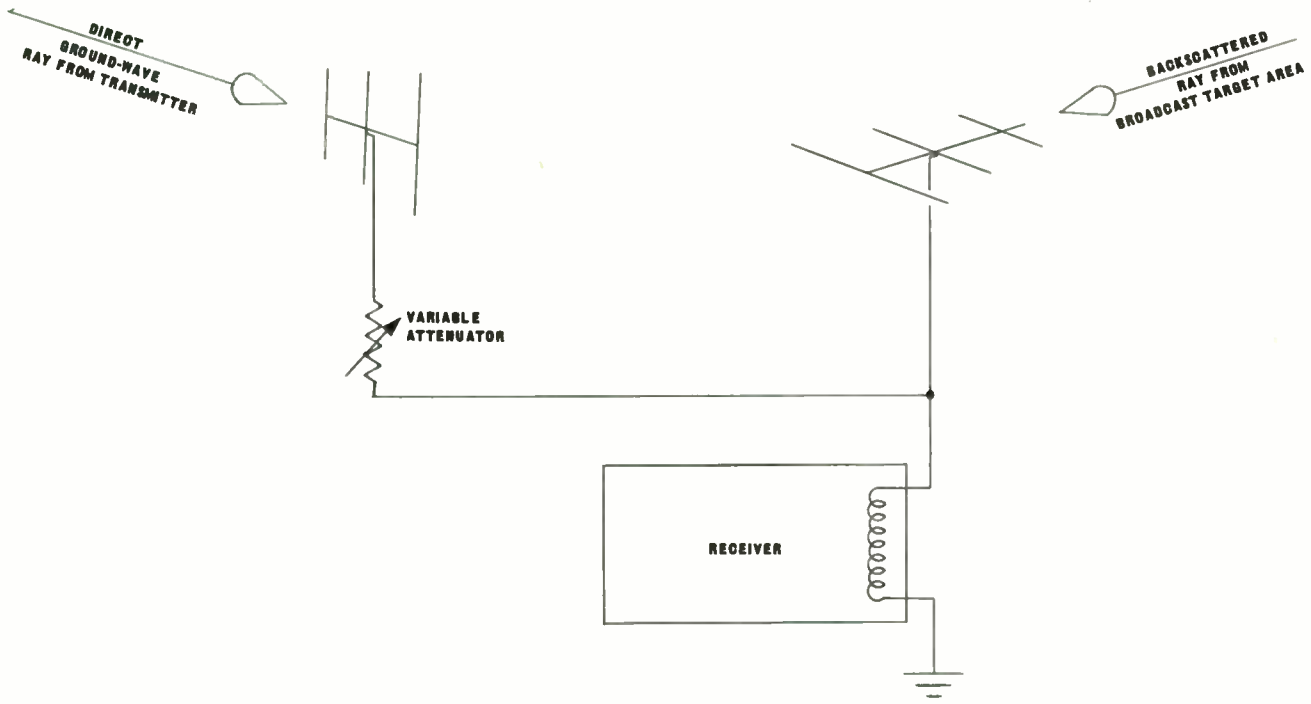


Fig. 5. Schematic diagram of a simple method for combining ground-wave and backscattered wave using one receiver.

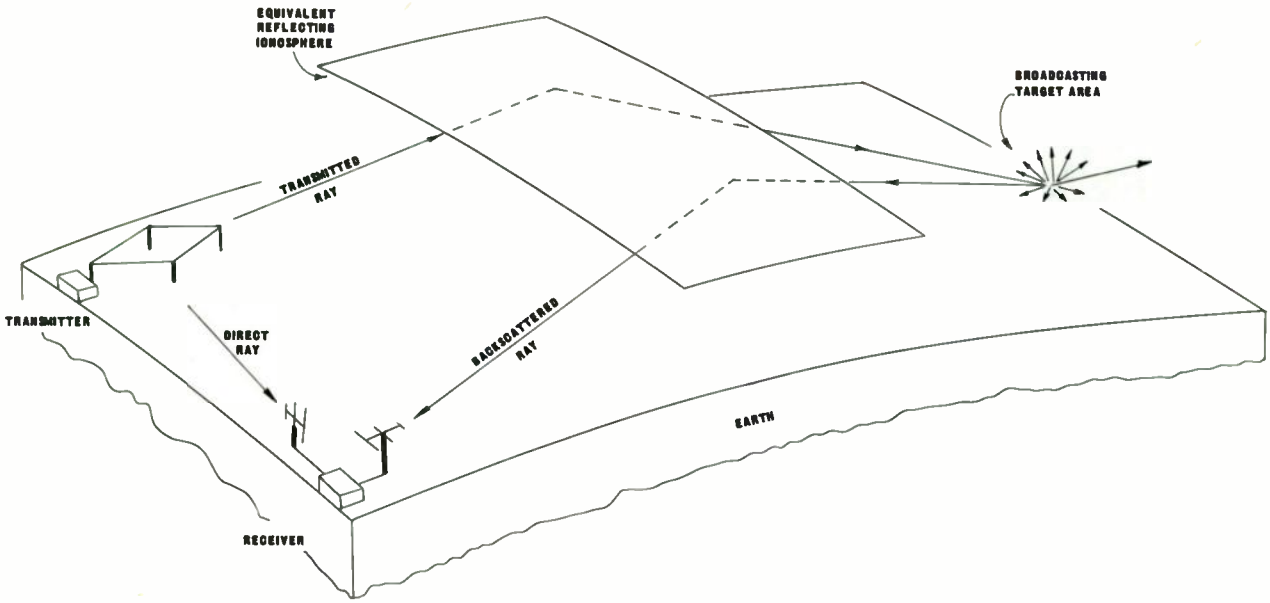


Fig. 6. Typical FM backscatter system geometry.

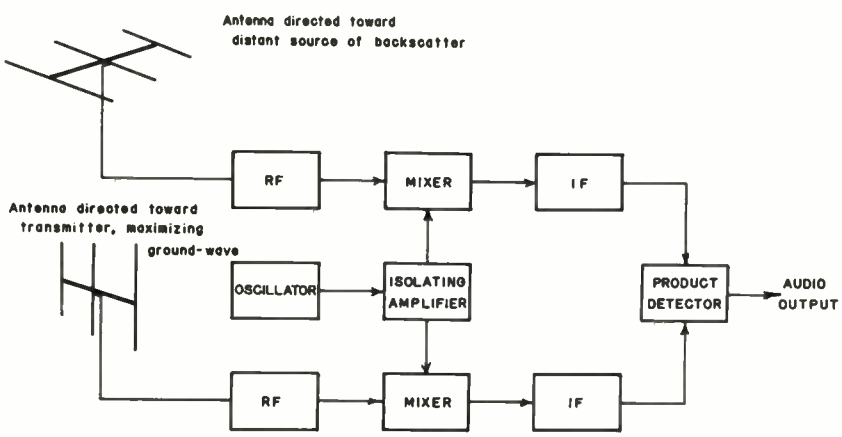


Fig. 7. More elaborate method for obtaining proper ground-wave demodulation at receiving station of an FM ionospheric radar.

PRECIPITATION SCATTER AS AN INTERFERENCE SOURCE IN COMMUNICATION SATELLITE SYSTEMS

Arnett S. Dennis
Stanford Research Institute
Menlo Park, California

Abstract

Communication satellite systems operating at frequencies as high as 3 Gc are now in the active planning stage. A number of recent papers have dealt with the interference produced at the ground receiving sites of such systems by tropospheric scatter from sources below the radio horizon. However, theoretical and experimental data from the field of radar meteorology show that scattering by precipitation particles at 3 Gc can easily exceed the tropospheric scatter component by several orders of magnitude. The precipitation scatter is approximately isotropic in most cases and hence the scattering volume need not be on the Great Circle path from the interfering source to the receiver site to be effective. Minimum separations for satellite terminals and interference sources computed considering only tropospheric scatter are therefore completely inadequate to guard against precipitation scatter.

Introduction

At the present time satellites are being considered as active relay stations for long-range communication systems. The large bandwidth required to make such systems feasible has led to suggestions that they be operated in the SHF region, with frequencies from 2 to 10 Gc appearing most favorable at present.[1] Because of the low power outputs of satellite transmitters, very sensitive receivers will be required at the ground terminals, with some manufacturers suggesting effective noise temperatures of less than 100°K. At such receivers interference problems will be acute. Several writers have analyzed in detail the potential interference arising from tropospheric scatter of signals originating below the radio horizon.[1,2]

Further consideration indicates that at the frequencies being considered precipitation scatter will be a more serious problem than tropospheric scatter. Indeed, Gordon concluded as early as 1952 that forward scatter from heavy rain could exceed tropospheric scatter signals even at wavelengths over 1 meter.[3] Since then an experiment in eastern Canada using 11-cm equipment has shown forward scatter from rain exceeding the tropospheric scatter component by as much as 15 db on occasion.[4]

It is the purpose of this paper to apply what is presently known about the scattering of SHF radio waves by precipitation in an attempt to predict interference levels at the ground terminals of communication satellite systems.

Review of Scattering Theory

Most theoretical treatments of the scattering of radio waves by precipitation have been based on a treatment given by Gustav Mie in 1908 of the scattering of a plane wave by a sphere.[5] The solution consists of an infinite series of spherical Bessel functions involving the complex dielectric constant of the sphere and the function α, defined as $\pi D/\lambda$, where D is the diameter of the sphere and λ is the wavelength of the incident radiation. The series of Bessel functions can be considered as a series of electric and magnetic multipole terms.[6] For $\alpha << 1$ the electric dipole term predominates and the formula for the scattering cross-section, Q_s, reduces to the Rayleigh formula

$$Q_s = \frac{2\pi^5 |K|^2}{3\lambda^4} \cdot D^6 \qquad (1)$$

The scattering cross-section is defined as the area which, when multiplied by the flux density of the incident wave, yields the total energy scattered by the particle. K is defined by the equation

$$K = \frac{m^2 - 1}{m^2 + 2} \qquad (2)$$

where m is the complex refractive index of the particle. For the frequency range from 2 to 10 Gc, $|K|^2$ is near 0.93 for water and 0.20 for ice.

As a Rayleigh scatterer functions as an electric dipole, the scattering is not isotropic; those regions at right angles to the incident electric vector are favored. These necessarily include a radar set viewing such particles. Therefore the concept of a back-scattering cross-section is introduced. The back-scattering cross-section, σ, of a particle is defined as the scattering cross-section of an isotropic scatterer which would return the same amount of power to the radar. For the Rayleigh case it is given by

$$\sigma = \frac{\pi^5 |K|^2}{\lambda^4} \cdot D^6 \qquad (3)$$

As individual precipitation particles show some independent motions within a storm, the radar return from an array of precipitation particles is an incoherent fluctuating signal whose mean intensity is proportional to the sum of the back-scattering cross-sections of the individual particles.[7] In the Rayleigh case, η, the radar

reflectivity per unit volume, is given by

$$\eta = \frac{\pi^5 |K|^2}{\lambda^4} \cdot Z \qquad (4)$$

where Z is the sum of the sixth powers of the particle diameters per unit volume. For cases where the relationship $\alpha << 1$ is not fulfilled, there is no simple relationship between η and the particle diameters.

Radar Reflectivity of Various Forms of Precipitation

Reflectivity of Rain

The Rayleigh approximation is valid for all but the largest raindrops at frequencies up to 10 Gc. As Eq. (4) shows, the radar reflectivity is a function of drop-size distribution as well as of the amount of liquid water present per un volume. However, there is a tendency for drop-size distribution to be determined by rainfall rate, and hence Z can be expressed in terms of rainfall rate R. One frequently quoted empirical relationship, based on drop-size distributions obtained in the Montreal area, yields, upon rearrangement,

$$\eta \text{ (rain)} = 6.9 \times 10^{-12} R^{1.6} (f_{Gc})^4 \qquad (5)$$

where η is expressed in m^{-1}, R in mm hr^{-1}, and f_{Gc} denotes the frequency in gigacycles.[6] This function is shown in Fig. 1 for rainfall rates of 1, 5, and 25 mm hr^{-1}.

Field measurements in the 3-Gc region have shown reflectivities around $10^{-8.5}$ m^{-1} in steady light rain and up to 10^{-6} m^{-1} in heavy thunderstorm rain,[8] but some hail may have been involved in the latter case. At 10 Gc the reflectivities are correspondingly higher. These results are in fair agreement with predicted values, but there is an unexplained tendency for measured values to run several decibels below those computed from drop-size distributions.[6]

Reflectivity of Snow

The derivation of a simple Z - R relationship for snow is complicated by the irregular shapes and wide variations in size and dielectric properties exhibited by snowflakes. A treatment of the problem and some experimental data are given in a recent article by Gunn and Marshall.[9] Their suggested formula for the reflectivity of dry aggregate snowflakes at a precipitation rate of 1 mm hr^{-1} is shown in Fig. 1.

In general, the radar reflectivity of dry snow is comparable to that of rain yielding the same rainfall rate, but snowstorms seldom yield rainfall rates (melted) greater than 3 to 4 mm hr^{-1}. Melting snow yields stronger radar echoes, which reach a maximum of 6 to 12 db above those in the rain below the melting level.[10] The enhanced radar echoes, known as the bright band, occur in a layer some 300 m thick. The bright band is most important for nearly horizontal beams. Polarization effects are present in both dry and melting snow, but more important in the case of melting snow.[11]

Reflectivity of Hail

In the frequency range which we are considering (2 - 10 Gc), the Rayleigh approximation breaks down for all hailstones except very small ones. Theoretical treatments of scattering by hail must be given in terms of the Mie scattering functions.

The calculation of the higher order multipole terms of the Mie expressions for Q_s and σ is complicated by the fact that the complex refractive index of water is a function of wavelength and temperature. Nevertheless, some progress has been made through the use of modern computers.[12,13] The contributions of the higher order multipoles increase with increasing α. Whenever the frequency of the incident radiation matches the frequency of one of the sphere's natural modes of electromagnetic oscillation, resonance effects appear. A plot of σ vs α shows a very complicated behavior, with numerous peaks and troughs due to interference among the various multipole contributions. Experimental verification of this behavior has been obtained by Atlas et al.[14] and by Gerhardt et al.[15]

Recently the computer programs have been extended to scattering through angles of less than π. The scattering patterns for large spheres are very complicated, with numerous marked side lobes. Herman and Battan find that, in certain size ranges, the power density scattered forward into a cone with scattering angles less than 30° exceeds the back-scattered power by more than 10 db.[16] When it is considered that hailstones come in a variety of shapes and have internal variations in dielectric properties, it is apparent that extrapolations of measured scattering cross-sections with respect to either scattering angle or frequency are not, in general, reliable. However, the findings of Herman and Battan on the preponderance of forward scatter for large hailstones are likely to find experimental confirmation, on the average.

Hail usually falls from thunderstorms accompanied by heavy rain showers, so the return from the hail cannot be measured separately. Recent measurements of the radar reflectivity of hailstorms in Arizona, New England, and Great Britain have yielded values of η up to $10^{-6.5}$ m^{-1} at 3 Gc and $10^{-2.5}$ m^{-1} at 10 Gc.[17,18,19,20] These experimental results are entered on Fig. 1. It should

be noted that they were not adjusted upward to allow for attenuation between the radar and the storms under observation.

Geometry of Scattering Problem

Continuous Wave Transmission

In order to compute the power received at an antenna through precipitation scatter, one must consider the geometry of the problem as well as the scattering cross-sections involved.

If there are scatterers in the volume common to the main beams of the transmitting and receiving antennas, their contribution to the scattered signal will usually be predominant. A convenient formula for this case with isotropic scatterers in the common volume is

$$P_r = p\,\bar{\eta}\,V\,\frac{\Omega_r}{4\pi} \qquad (6)$$

where P_r is the power received,*

p is the incident flux per unit area in the common volume,

$\bar{\eta}$ is the mean reflectivity of the scattering array per unit volume,

V is that part of the common volume occupied by the scatterers,

and Ω_r is the solid angle subtended by the receiving antenna at the common volume.

In satellite systems with sensitive ground receivers, side lobes can be important, especially where intense but localized storms are in progress. In such a case Eq. (6) can be conveniently rewritten as

$$P_r = p\,\bar{\eta}\,V \cdot \frac{g_r\lambda^2}{(4\pi R_r)^2} \qquad (7)$$

where $\bar{\eta}$ is the mean reflectivity of the volume (say, that of a shower) supplying the signal,

V is the volume occupied by the scatterers,

g_r is the gain of the receiving antenna in the direction of the scatterers,

R_r is the distance from the scatterers to the receiving antenna,

and the other symbols are as defined previously.

Pulsed Transmission

It is incorrect to treat cases with pulsed transmission in terms of mean power, because errors can be produced by pulses of interference even when the mean signal-to-noise ratio is large.

For cases with short pulses and sufficiently extensive arrays of scatterers, the pulse length will determine the size of the region contributing to the scatter signal received at a given instant. Let a pulse of duration τ be emitted from an antenna, starting at time $t = 0$. Then at any time $t > \tau$ an illuminated volume of length $c\tau$ is to be found moving out along the transmitter beam, where c is the speed of light. Figure 2 shows a plan view of the relationships among a transmitter T, the transmitted beam, and a receiver R outside the beam. At a given instant, let the illuminated volume by given by I_1I_2, with beamwidth considerations neglected. The contributing region is given at that instant by C_1C_2, where $C_1R = C_1I_1$ and $C_2R = C_2I_2$. From geometrical considerations

$$C_1C_2 = I_1I_2 + C_2R - C_1R$$

and to a first approximation

$$C_1C_2 = \frac{c\tau}{1 - \cos\theta} \qquad (8)$$

The speed with which the contributing region moves out along the beam varies as its length so that, neglecting attenuation and variations in the particle array, the total amount of energy scattered per unit length into unit solid angle remains constant. The power arriving at the receiver, however, varies as the length of the contributing region. Thus the scattered power must be divided by $(1 - \cos\theta)$ to yield P_r, the peak power received at R via scattering through the angle θ. In addition, the free-space losses between the contributing region and the received must be considered. From Fig. 2 it is seen that R_r is given by $d'/\sin\theta$, where d' is the perpendicular distance between the receiver and the beam. For the case where the pulse length determines the extent of the contributing region, we have, then,

$$P_r = \frac{pAc\tau}{1 - \cos\theta} \cdot \eta \cdot \frac{g_r}{(4\pi d')^2}\sin^2\theta \qquad (9)$$

where A is the cross section of the array of illuminated scatterers.

Differentiation of this function with respect to θ, assuming g_r a constant, shows that P_r has its maximum at θ equal to 0. There is no maximum at θ equal to $\pi/2$, as one might expect at first glance. This behavior appears to be verified in the photographic records of rain-scattered pulses obtained by Doherty and Stone.[4]

* Strictly speaking, the received signal is incoherent with P_r representing the mean power received.

Numerical Examples

General Comments

Although there are several criteria which must be satisfied if communication satellites are to share frequencies with existing services, we shall limit our discussion to the interference produced at a ground terminal of a satellite system by existing radio systems.

The seriousness of the interference produced by precipitation scatter in a given communications link will depend upon the signal-to-noise ratio, the modulation system employed, and other factors beyond the scope of this paper. In our numerical examples we shall follow the procedure used in Ref. 1, where the interference power level is compared with the existing noise level, rather than with the desired signal. Making the desired signal strong enough to override the total receiver noise will suppress any noise-like signal which is itself below the total noise level.

The total receiver noise consists of contributions picked up at the antenna and internally-generated noise. These can conveniently be described in terms of an antenna temperature, T_a, and an effective receiver noise temperature, T_e, according to

$$P_n = (T_a + T_e)\, k\, b \qquad (10)$$

where P_n is the total noise power,

k is Boltzmann's constant.

and b is the effective receiver bandwidth.

T_a and T_e are expressed in degrees Kelvin. T_a involves at various times noise contributions from galactic sources, the sun, black-body radiation from the earth, attenuation noise from precipitation, and so on.[21] T_a can be as low as 15 - 30°K under optimum conditions, but during rainy weather it can rise to 100 - 150°K.[22] Since T_e for maser amplifiers can be reduced to 15 - 30°K, it appears that T_a rather than T_e will be the limiting factor on system performance in many cases. As a reasonable estimate, we shall use 100°K for the sum $(T_a + T_e)$ and 10 Mc for the effective receiver bandwidth in making our comparisons. Then, from Eq. (10),

$$P_n \doteq -139 \text{ dbw} \qquad (11)$$

Scattering from Stratiform Precipitation

Consider the receiving antenna in a situation where uniform light precipitation extends from the ground up to height z_1 over a large area. Let the precipitation be illuminated uniformly above some height, z_o, by an interference source below the horizon. The power received is independent of the receiving antenna pattern as long as all the gain is directed through the illuminated layer. The situation is shown in Fig. 3. The received power, worked out on the basis of either Eq. (6) or Eq. (7), is

$$P_r = \frac{p\,\eta\,\lambda^2}{4\pi}\,(z_1 - z_o)\sec\psi \qquad (12)$$

where ψ is the angle between the receiving antenna beam and the vertical, and all other symbols are as previously defined. Equation (12) is derived assuming a flat earth and will not hold for values of ψ approaching 90°.

Assume that the interfering source is a back lobe of 0 db gain on a troposcatter transmitter 200 km away, emitting 20 kw in a 10-Mc band at 3 Gc (see Fig. 4). In this case z_o would be approximately 2.5 km. With moderate rain (5 mm hr^{-1}) extending up to 5 km over the receiver site, we would have for a vertical beam, using Eq. (12) and Fig. 1,

$$P_r = \frac{(10^{-7.4}\ \text{w m}^{-2})(10^{-8.1}\ \text{m}^{-1})(0.1\ \text{m})^2(2.5\ \text{km})}{4\pi}$$

$$= -152 \text{ dbw}$$

Comparison with Eq. (11) shows that the interference from this troposcatter transmitter is 13 db below the noise level.

Scattering by Convective Storms

Convective storms are characterized by strong vertical motions, high rainfall rates, and large raindrops, and hence by high radar reflectivity. This is particularly true of hail-bearing thunderstorms, which sometimes have their maximum reflectivity at heights of 5 to 10 km, and extend upwards to 15 or 20 km.[18] It should not be thought that hailstorms are rare occurrences; evidence is accumulating that a majority of thunderstorms in the continental United States are characterized by hail aloft at some stage during their existence.[23,24]

Let us now consider the case of the troposcatter transmitter discussed in the last sub-section, but replace the layer of precipitation by a thunderstorm with reflectivity 10^{-7} m^{-1} in the receiver beam, extending up to 12 km. In this case P_r is found to be approximately - 136 dbw, or 3 db above the total receiver noise in our example. Thus, with a thunderstorm in the receiver beam, forward scatter could be the limiting factor on system performance, even with the interference source below the horizon and pointing away from the receiver site. It should be noted that this example used a frequency of 3 Gc; at higher frequencies the scatter interference would be greater.

As a further example, consider the case shown in Fig. 5. Here a thunderstorm is located 200 km

from a carefully engineered receiving antenna, whose gain in the direction of the storm is -10 db. Using an effective earth radius of 1.5, which is typical of warm, humid air masses, to account for refractive effects, we find that those portions of the storm above 3.5 km are above the radio horizon at the receiver site. (This derivation assumes no terrain or artificial shielding.) The thunderstorm is illuminated by the main beam (25 db gain) of an airborne C-band weather radar 500 km away, which is emitting 1-μsec pulses of 20 kw peak power at 5.3 Gc. A straightforward calculation shows that p, the incident flux density at the storm, is -47 dbw m^{-2}. As a typical thunderstorm is some 10 to 15 km in diameter and extends upward about the same distance, a reasonable estimate of the illuminated cross-sectional area more than 3.5 km above ground is 100 km^2. Then, using Eq. (9), we find P_r, the interfering power received at R, to be -128 dbw for isotropic scatterers with η equal to 10^{-5} m^{-1}. During the time (order of 10 μsec) the contributing region is passing through the storm, it produces an interfering signal 11 db above the prevailing noise background.

From this it appears that the radar's side and back lobes could produce additional significant contributions by illuminating any other storms which happened to be within the radio horizon of the receiver. In a typical summer shower situation, the emission of each pulse from the radar would be followed by the arrival of numerous bursts of interference spread out over a period of several hundred microseconds. Since the pulse repetition period of a radar is of the same order, interference could be virtually continuous.

In the above no allowance has been made for attenuation, which could largely eliminate the contributions from certain storms if other intense storms lay between them and the receiver site. However, we note again that the radar data given above were not corrected for attenuation either, and so the results should be comparable. Furthermore, we have not allowed for the tendency for hail to scatter more strongly forward than backward, which could add several decibels to the received power. Polarization has also been neglected, but for scattering through an angle of 20°, changes of polarization would introduce no more than a fraction of a decibel change in the received power. It should be noted, too, that all examples given used typical rather than extreme values for transmitter power, radar reflectivity of precipitation particles, and storm dimensions.

Comparison with Tropospheric Scatter

In the last example given, the peak power received via the radar's main beam was computed at -128 dbw. Allowing for the antenna gains assumed, the basic transmission loss is

$$L_b = P_t - P_r + G_t + G_r$$

$$= 43 - (-128) + 25 + (-10) \text{ db}$$

$$= 186 \text{ db}$$

where P_t is the transmitter power,

G_t is the transmitter antenna gain,

and G_r is the (realized) gain of the receiving antenna.

Fig. V-1 of Ref. 1 (extrapolated) indicates that the basic transmission loss over a 700-km path at 5.3 Gc for troposcatter propagation is near 265 db. In our example, the power received through the tropospheric scatter medium would be reduced another 20 db or so, since only a side lobe of the radar antenna pattern is directed toward the receiver, and tropospheric scatter propagation is limited for practical purposes to Great Circle paths. In such a case the tropospheric scatter component can safely be treated as negligible compared to the precipitation scatter component.

Conclusions

Any high-output source of microwave radiation can create interference at a ground terminal of a communication satellite system, if it illuminates precipitation particles above the radio horizon of the receiving antenna. Because the scattering approaches isotropic, interference is not limited to those cases where the storms lie on the Great Circle path from the source to the receiver. As thunderstorms frequently extend to 10 - 20 km above ground, it appears that, in areas where they occur, high-power ground-based transmitters cannot share spectrum space in the centimetric region with communication satellite systems, unless spatial separations of the order of 1,000 km are imposed. For airborne equipment the required separations are necessarily greater.

References

1. "Frequency Allocations for Space Communications," a Report of the Joint Technical Advisory Committee, IRE-EIA; March 1961.

2. J. J. Downing, N. N. Berger, F. F. Fulton, Jr., and D. P. Harris, "Interference Considerations for Communications Satellites," Proc. Western Electronic Show and Convention, San Francisco, Calif.; Aug. 1961.

3. W. E. Gordon, "A Comparison of Radio Scattering by Precipitation and by a Turbulent Atmosphere," Proc. 3rd Weather Radar Conf., pp. F17-F24, McGill University, Montreal, Quebec; Sept. 1952.

4. L. H. Doherty and S. A. Stone, "Forward Scatter from Rain," Trans. IRE, PGAP-8, pp. 414-418; July 1960.

5. G. Mie, "Beitrage zur Optik truber Medien, speziell Kolloidaler Metallosungen," Ann. Physik,

vol. 25, pp. 377-445; 1908.

6. K. L. S. Gunn and T. W. R. East, "Microwave Properties of Precipitation Particles," Quart. J. Roy. Meteorol. Soc., vol. 80, pp. 522-545; Oct. 1954.

7. J. S. Marshall and W. Hitschfeld, "Interpretation of the Fluctuating Signal from Randomly Positioned Scatterers: Part I," Can. J. Phys., vol. 31, pp. 962-995; 1953.

8. D. E. Kerr, Propagation of Short Radio Waves, McGraw-Hill Book Company, Inc., New York, N.Y.; 1951.

9. K. L. S. Gunn and J. S. Marshall, "The Distribution with Size of Aggregate Snowflakes," J. Meteorol., vol. 15, pp. 452-461; Oct. 1958.

10. L. J. Battan, Radar Meteorology, pp. 85-90, Univ. of Chicago Press; 1959.

11. R. E. Newell, S. G. Geotis, M. L. Stone, and A. Fleisher, "How Round are Raindrops?" Proc. 5th Weather Radar Conf., pp. 261-268, Asbury Park, N.J.; Sept. 1955.

12. B. M. Herman and L. J. Battan, "Calculations of Mie Back-Scattering of Microwaves from Ice Spheres," Proc. 8th Weather Radar Conf., pp. 187-193, San Francisco, Calif.; 1960.

13. J. J. Stephens, "Radar Cross-Sections for Water and Ice Spheres," J. Meteorol., vol. 18, pp. 348-359; June 1961.

14. D. Atlas, W. G. Harper, F. H. Ludlam, and W. C. Macklin, "Radar Scatter by Large Hail," Proc. 8th Weather Radar Conf., pp. 1-8, San Francisco, Calif.; April 1960.

15. J. R. Gerhardt, C. W. Tolbert, S. A. Brunstein, and W. W. Bahn, "Experimental Determinations of the Back-Scattering Cross-Sections of Water Drops and of Wet and Dry Ice Spheres at 3.2 Centimeters," J. Meteorol., 18, pp. 340-347; June 1961.

16. B. M. Herman and L. J. Battan, "Calculations of the Total Attenuation and Angular Scatter of Ice Spheres," Proc. 9th Weather Radar Conf., pp. 259-265, Kansas City, Mo.; Oct. 1961.

17. R. J. Donaldson, "Thunderstorm Reflectivity Structure," Proc. 8th Weather Radar Conf., pp. 115-125, San Francisco, Calif.; April 1960.

18. R. J. Donaldson, "Radar Reflectivity Profiles in Thunderstorms," J. Meteorol., 18, pp. 292-305; June 1961.

19. S. G. Geotis, "Some Radar Measurements of Hail," Proc. 9th Weather Radar Conf., pp. 133-138, Kansas City, Mo.; Oct. 1961.

20. K. A. Browning and F. H. Ludlam, "Radar Analysis of a Hailstorm" Tech. Note 5, Contract AF 61(052)-254, Air Force Cambridge Research Center; 1960.

21. R. C. Hansen, "Low Noise Antennas," Microwave J., 2, No. 6, pp. 19-24; June 1959.

22. D. C. Hogg and R. A. Semplak, "The Effect of Rain and Water Vapor on Sky Noise at Centimeter Wavelengths," Bell System Tech. J., 40, pp. 1331-1348; Sept. 1961.

23. H. T. Harrison and W. B. Beckwith, "Studies on the Distribution and Forecasting of Hail in Western United States," Bull. Am. Meteorol. Soc., 32, pp. 119-131; April 1951.

24. T. J. Henderson, "Northeast Colorado Hail Suppression Project: Final Report--1959," Weather Modification Co., San Jose, Calif.; Feb. 1960.

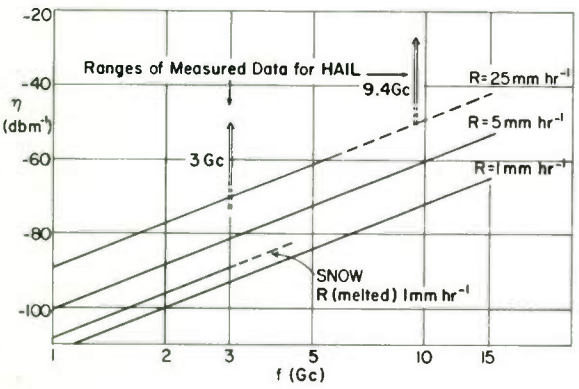


Fig. 1. Radar reflectivity per unit volume as a function of frequency.

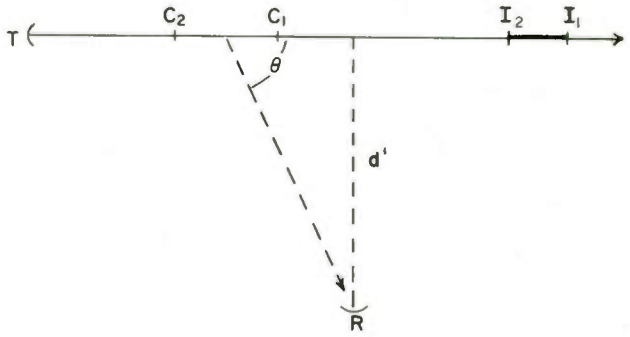


Fig. 2. Relationship between illuminated volume and contributing region.

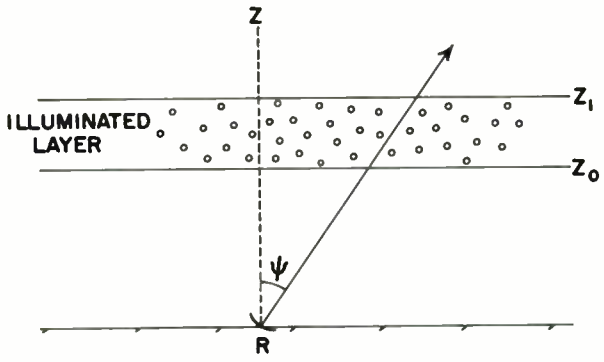


Fig. 3. Scattering from stratiform precipitation.

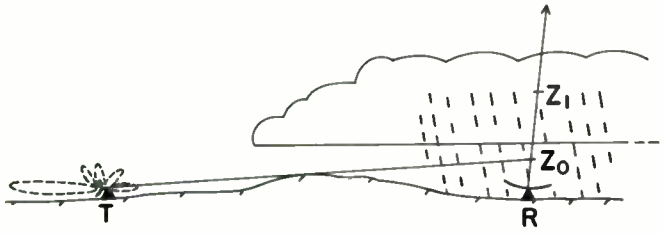


Fig. 4. Illumination of stratiform precipitation by back lobe of a tropospheric scatter transmitter.

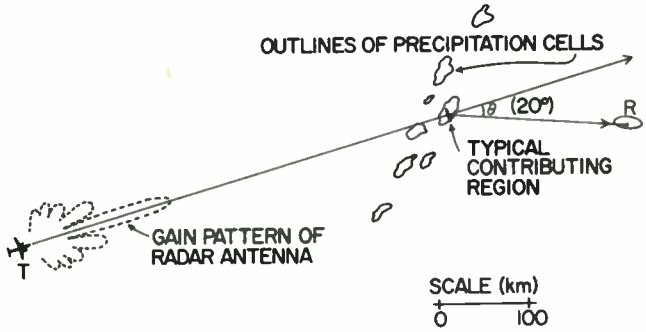


Fig. 5. Coupling of receiver to air-borne radar by precipitation cells.

ISCAN - INERTIALESS STEERABLE COMMUNICATION ANTENNA *

Dr. H. Brueckmann
U.S. Army Signal Research and Development Laboratories
Fort Monmouth, New Jersey

J. R. Gruber and C. A. Bramble
Avco Corporation
Cincinnati, Ohio

Summary

The main purpose of the new antenna system, described herein, is to enhance high-frequency communications by reducing the effects of multipath and interference. The experimental ISCAN system, designed for reception of long distance h-f communications, consists of two principal parts: a linear array of vertical dipoles over 6300 feet long and an electronic beam forming matrix. A single, fixed-beam radiation pattern is available at each of the 14 terminals of this matrix. The beams have the shape of coaxial conical shells and adjacent beams overlap each other at the half-power points. These beams are formed by the use of tapped delay cables which is an inherently broadband method. Inertialess scanning of an antenna beam is achieved by rapid switching of the 14 beams in sequence using electronic gating circuitry. This effectively scans the antenna beam up to 30 degrees off the array axis. A visual oscilloscope display is provided to show the relative outputs of the multiple beams at all times.

Introduction

Directive antennas of several types have been available for use in the high-frequency range for a number of years. Curtain arrays and long wire radiators such as rhombics have been used extensively for long-range communications and shortwave broadcasting. The beamwidth of a fixed-beam antenna cannot be made too narrow or it will not encompass all directions of arrival of the major modes from a distant h-f station. If, however, the antenna beam can be steered fast enough, it can be made more directive until the beam is so narrow it selects one mode only.

Best known among the antennas with fairly high directivity which have been successfully steered are the curtain array, Wullenweber, and the rhombic array used in the MUSA system.[1] Each these antennas has certain beam steering restrictions: the steering angle of the curtain array is fairly limited, the Wullenweber steers only in azimuth and the MUSA system steers only in elevation. These three steerable beam systems have several common limitations, namely, they are large, costly, and the beam steering is relatively slow. It may well be too slow for optimum reception of long-distance h-f signals.

The purpose of the ISCAN (Inertialess Steerable Communication Antenna) development [2, 3, 4] was to produce a steerable h-f array, of lower cost and with improved performance compared to the MUSA system, particularly for optimizing long range global communications and rejecting interference.

The ISCAN system represents an economical h-f receiving array with multiple, highly directive beams having the shape of conical shells. These beams overlap each other to cover a given solid angle in space and can be continuously monitored by separate receivers. At the same time the outputs of these beams can be switched sequentially by electronic means to a single receiver giving, in effect, very rapid beam scanning of a conical sector in space.

Theory of Operation

The present experimental ISCAN system consists of the antenna array, cabling, and associated electronic circuitry for generation and control of the multiple antenna beams as well as a visual display for monitoring the detected outputs of the multiple antenna beams. The frequency band of operation is 12 to 18 mc, which was chosen to suit the particular communication circuit and time of experimentation (1961-62) and represents a good compromise between performance and cost. There is no fundamental frequency range restriction imposed by technical limitations. The present system is used for reception only, but it could be adapted for transmission, also, by suitable modification.

* Work performed under Signal Corps Contract DA-26-039-SC-78293. The views of the authors do not purport to reflect the position of the Department of the Army or the Signal Corps.

The antenna is a linear array 96 wavelengths long (at 15 mc) which generates multiple beams covering angles of about 5° to 30° off the array axis. It is composed of 24 vertical, center-fed dipole antenna elements mounted on 24 wooden poles, as shown in figure 1, which are non-uniformly spaced and aligned to form an array 6,312 feet long. The system is installed at Windmill Point, Maryland, with the array axis aimed toward Central Europe. The array forms 14 fixed beams which have the shape of coaxial conical shells. Adjacent beams overlap each other at the three db points as shown pictorially in figure 2. Inertialess scanning of the antenna beam is achieved by rapid switching to each of the 14 beams in sequence using electronic gating circuitry which effectively scans the antenna beam up to 30 degress off the array axis.

The formation of multiple beams is achieved by the use of tapped delay lines inserted in the signal paths from the 24 antenna elements to the summing point (combiner) of the array. These delay lines compensate for the individual delays of the signals in space. A particular beam is formed when the delays are adjusted such that the antenna element signals, for that angle of arrival, reach the summing point at precisely the same time and are added vectorially in phase. The differential time delays generated in space between antenna elements depend only upon the angle of arrival and not upon the frequency of the received signal. Therefore the ISCAN steering system, which utilizes time delay rather than phase shift, is inherently broadband.

The geometry used for analysis of the time delays required for the array elements is given in figure 3. The array elements are arranged linearly and numbered 1 through 24. A signal arriving at an angle φ with respect to the array axis will arrive first at element No. 24. This same wavefront will reach element No. 1 at a later time which is proportional to the distance, $D_1 \cos \varphi$. If the signal induced in element No. 24 is allowed to travel in a coaxial cable toward element No. 1 the signal will arrive at a point A at the same time the parallel wavefront arrives at element No. 1. The distance of point A from element No. 24 is

$$D_A = \frac{v}{c} D_1 \cos \varphi \qquad (1)$$

where

- D_1 = distance of element No. 1 from No. 24
- c = velocity of light
- v = velocity of propagation in the cable.

If the remaining distance between point A and element No. 1 is equally divided, it is possible to locate a summing point, as in figure 3, which receives the signals induced in elements No. 1 and No. 24 at precisely the same time.

Signals induced in the remaining 22 elements will arrive, via the direct cable route to the summing point, earlier than the signals from elements 1 and 24. They experience less delay either because their path length is shorter or because they utilize less cable at the slower velocity of propagation. For this reason, the proper amount of fixed delay cable is added to the cabling from each element such that the total travel time of each signal from the parallel wave front location at element 24 equals the time delay of the interconnecting cable from element 24 to the summing point. The total amount of cable required from each element to the summing point is determined from the following time relationship

$$Tnc = \frac{D_{sp}}{v} - \frac{D_n \cos \varphi}{c} \qquad (2)$$

where v, c, and Φ are as defined above, and

- Tnc = time delay of cable connecting element n to the summing point,
- Dsp = distance of summing point from element No. 24, and
- Dn = distance of element n from element 24.

The maximum value of φ is 30 degrees for the ISCAN array and was chosen to accommodate the largest expected downcoming angle of arrival of the ionospherically reflected h-f signals. Equation (2) shows that, for smaller values of φ, less delay is required in each of the element signal paths if the position of the summing point is fixed. It is therefore possible to position taps along a portion of each delay cable so that additional beams can be formed between zero and 30 degrees. An analysis [5] reveals that the total amount of cable, including element transmission lines and delay cable, is minimized by locating the summing point near antenna element No. 5. Location of the summing point near the array center would require about 60 percent more cable.

A simplified diagram which illustrates the beam forming method is given in figure 4. Beams in five different directions are formed simultaneously by the use of multiple-tapped delay lines in each channel. Each group of taps shown joined together in the diagram is connected to a power adder or combiner. The outputs of the five combiners, A, B, C, D, and E represents five antenna beams displaced from one another by the desired angular increments. By use of r-f gates, electronically controlled, at the outputs of each combiner the antenna beams can be sequentially sampled at a very rapid rate for signal comparison.

This produces, in effect, inertialess beam steering. Figure 4 shows only six antenna elements and five beams, whereas the ISCAN array consists of 24 elements and generates 14 beams overlapping at the half-power points.

Figure 5, the overall system diagram, shows all the major components and the interconnecting r-f cable. The control point for the array is located near antenna element No. 5 and the actual tapped delay cable is designed for beam steering about element No. 13. These two parameters were so chosen to minimize the total amount of tapped delay cable as well as system coaxial cable.

The Antenna Array

Array Design

The physical area occupied by the antenna is very small since the interconnecting cable is buried under ground and the only parts visible are the 24 self-supported wooden poles on which the dipoles are mounted. The antenna elements are non-uniformly spaced and constitute an unfilled array in which the element spacing varies from about 200 to 350 feet that is, about 3 to 5 wavelengths at mid-band. The use of an unfilled array results in a considerable reduction in system cost since a conventional filled array with half-wavelength spacing would have about 192 dipole elements plus all the additional cable and fittings associated with the dipoles. The non-uniform spacing, wherein element position is determined by a polynominal relationship, is a new design concept of Avco which permits control of pattern sidelobes in the important spatial sector (within 30 degrees of the array axis). The overall length of the array was chosen to provide beams narrow enough to resolve multipath or interfering signals. The two lowest beams are only 4.0 degrees apart in elevation angle and the two highest are 1.2 degrees apart.

Figure 6 illustrates the position and relative beamwidth of the 14 overlapping beams as they exist in the elevation plane. The amplitude reduction at the lower elevation angles is caused by operation over soil of average conductivity. Since the array generates beams in the form of conical shells the azimuth plane beam patterns decrease in beamwidth in a similar manner as the off-axis angle increases. A typical calculated beam pattern showing sidelobe structure is given in figure 7. This pattern shows the sidelobe control provided up to angles of about 40 degrees off axis which is considered very adequate for the intended application of the array. The sidelobes at more than 40 degrees off axis average about 13 db below the main beam.

The actual directivity of the unfilled array may approach 18.9 db above isotropic (including 5.1 db element directivity) assuming that the signals from all 24 elements add coherently and the noise contributions add randomly. The directivity, calculated from patterns and assuming perfectly conducting ground, averages about 18.5 decibels above isotropic. Neglecting the effect of finite ground conductivity the directivity of all 14 beams is nearly the same because each fills about the same solid angle in space.

An effort was made to insure that the sidelobe levels, within 30 degrees of the array axis, were kept below a reasonable level. Statistical calculations were made to establish tolerances for phase and amplitude errors in the array so that these did not contribute sidelobes above -20 db*. Adjustments for both phase and amplitude were provided in each element signal path to keep errors within the permissible tolerances. Actual field tests performed after installation proved that the array could be adjusted to operate within these limits.

Antenna Elements

Each of the twenty-four elements of the ISCAN array consists of a vertically polarized center-fed dipole. Each dipole element, 29' 2" long overall, is supported by insulating rods on a wooden telephone pole, as in figure 8. An air-core auto-transformer, connected across the dipole ends at the feedpoint, transforms the dipole impedance down by a factor of about one-half. The impedance is resistive and 35 ohms at 14.2 mc. The intentional mismatch to the cable of 73 ohms characteristic impedance serves to reduce mutual coupling effects between elements and to improve "phase-tracking" over the frequency range.

The feedline coaxial cable is routed down the inside of the lower dipole tube where it joins the coaxial cable choke at the bottom of the lower element. The cable choke is resonant at approximately 17 mc and effectively isolates the dipole from the feed cable and ground, thereby eliminating the need for a ground screen or radials. Both the cable choke and matching transformer are encapsulated with a low-loss RTV silastic for protection against moisture and weathering.

Effectiveness of the center-fed dipole and isolation choke were verified by impedance measurements on the highest and lowest elements of the array. Phase tracking of these two elements was calculated from the impedance data and figure 9 shows that the relative phase shifts of these two elements differ by less than 7 degrees across the 12 to 18 mc band.

Element Cables

The transmission line, which couples each

*Except for the two sidelobes adjacent to the main beam.

antenna radiator to the control point, is one-half inch, haberline jacketed, 70-ohm, Foamflex coaxial cable (made by Phelps-Dodge). This cable is a semi-rigid coaxial line with a solid copper center conductor, foamed polyethelene dielectric, and aluminum outer conductor. It has a nominal attenuation of 0.32 db per 100 feet (15 mc) and a velocity of propagation of 0.82. This cable was chosen for its low attenuation and good temperature characteristics which allow it to be used in extended lengths without the need of pre-amplifiers at each antenna element. The cable was buried in a trench three feet deep running the length of the array to prevent ground plane distortions and to provide a fairly constant temperature environment for the cables during seasonal changes.

About 56,000 feet of this Foamflex cable is required for the element cabling and another 22,000 feet of the same 70-ohm cable, without the haberline jacket, is used for the fixed and tapped delay line. The total cost of cable is only about 10 percent of the total cost of the experimental antenna.

Beam Forming System

Control Point

All of the system components are located at the control position, near element No. 5, inside a semi-trailer van which is temperature controlled. These components include the fixed and tapped delay cables, the r-f driver amplifiers, tap amplifiers, combiners, programmer, and visual display arranged schematically as in the system diagram of figure 5. The general layout of the control van, figure 10, shows how the fixed and tapped delay cables are arranged.

The r-f driver amplifiers are installed between the interconnecting element cables and their associated fixed delay cables in order to compensate for the individual attenuation of the various lengths of cable and for the differences in effective height of the individual dipoles. The largest element cable has an attenuation of about 18 db. Also, isolation amplifiers (tap amplifiers) are provided at each tap of the tapped delay cables to reduce standing wave effects on the cables due to tap loading. Each combiner utilizes a specific tap on each cable to collect the 24 separate signals which are combined to form one of the 14 beams. Any of the combiners can then be connected to an ordinary communication receiver for continuous use of a given beam. Since the beams are broad-band, any number of communication receivers can be connected to any one of the 14 beams through the use of multicouplers; however, provision for just three receivers has been incorporated in the present ISCAN system. The use of multiple receivers permits simultaneous reception of a number of communication channels at various frequencies and of various modes of propagation at each frequency.

R-F Driver Amplifiers

The transmission lines from the antenna elements, upon entering the control van, are connected to separate r-f driver amplifiers. These amplifiers provide terminations for the element lines and their outputs drive the fixed delay cables. The amplifiers are wide band, covering 10 to 20 mc, and gain control is achieved by altering the amount of negative feedback around the the second stage which results in a significant reduction in intermodulation distortion. The feedback, which is applied through a variable capacitor, allows for the gain variation with a nearly constant phase shift. The 24 driver amplifiers are constructed on rack-mounted chassis, two amplifiers per chassis, and placed in a fan-ventilated, rack-type cabinet along with a power supply and 24 variable delay lines as shown in figure 11. The delay lines provide fine adjustment of the fixed delay for each signal path and are equivalent to approximately 40 feet of the Foamflex cable.

Typical performance of the driver amplifiers is listed below:

1. Noise Figure - Average 5.7 db at various gains over the passband.
2. Intermodulation Distortion - Third order products 50 db below two 25 mv signals induced in the antenna.
3. Input VSWR - Less than 1.5 from 12 to 18 mc.
4. Output VSWR - Less than 2.0 from 12 to 18 mc.
5. Gain at 15 Megacycles - Continuously variable from 10 to 37 db.
6. Phase Shift Relative phase shift between two amplifiers at various gains over the passband $\pm 4^{\circ}$.

Tap Amplifiers and Combiner

Figure 12 shows the construction of a tap amplifier chassis which contains 14 Type 2N706 grounded emitter amplifiers mounted on a single printed circuit board. The tap amplifiers provide isolation and drive signals through the inter-connecting cables to the combiner. Transistors were chosen to conserve space and power along with the need for a relatively high input impedance to allow direct connection to the delay cable without causing appreciable discontinuity. The tap amplifiers are capacitively coupled from the cable through a variable capacitor which is used as a gain adjustment since the outputs of all the 336 tap amplifiers must be equal. The cable tap itself consists of a direct connection to the inner conductor by means of a small screw which is supported and insulated from

the outer conductor, by a Rexolite head. The aluminum block, which supports the tap amplifier chassis in figure 12, also clamps the 14 coils of each tapped delay cable where the taps are made. Note the small volume, 7 x 7 x 3 inches, occupied by this assembly of 14 amplifiers.

Eight-foot lengths of RG-71/U, 93-ohm coaxial cable, are used to connect the 336 tap amplifier outputs to the 14 combiner circuits. Each of the 14 antenna beams is formed by combining appropriate tap signals from the 24 antennas. The 24 tap signals are fed to a common bus in the combiner rack, see figure 13, where a resistive termination is provided. The combiner rack also includes the r-f gate circuits, (one beside each combiner bus), the programmer, and the power supply.

Programmer and Display

The programmer provides the timed signals necessary to control the r-f gate circuits and the oscilloscope sweep for the visual display. These circuits sequentially sample each beam of the ISCAN array and apply the output to a receiver whose i-f output is displayed on the oscilloscope. The display receiver, oscilloscope, and three auxiliary receivers are located in the display console, as shown in figure 14. The three rotary selector switches for the auxiliary R-390 communication receivers are positioned below the display oscilloscope. Toggle switches are also located on the control console to enable the operator to turn ON or OFF any or all of the driver amplifiers in the process of checkout and adjustment of the array.

The programmer circuit consists of a clock multivibrator, four bistable multivibrators, and a diode matrix. The output of the diode matrix furnishes the control signals for the r-f gates which all have a common load resistor. Vertical deflection for the display oscilloscope is provided by the 14 gated signals which appear across this common load. The horizontal sweep voltage, derived from the clock circuit, is delayed in time to coincide with the output of the R-390 receiver.

The display, as observed on the oscilloscope, consists of 14 vertical lines corresponding to the relative signal strength of the beams. The repetition rate of the display can be adjusted over a range of approximately 20 to 60 times per second. It is a simple procedure for the system operator to observe the displayed output of the 14 beams and select whichever one he deems best for operation with one of the auxiliary receivers.

System Performance

Operation

Installation of the ISCAN system was completed on 30 June 1961, and although the system has been operational since that time the evaluation of performance is far from complete. However, the Signal Corps has conducted some operational tests, employing short pulse transmissions from Germany, which prove the ability of the array to separate multipath signals and interference. The signal consisted of 0.5 millisecond pulses at a repetition rate of 25 per second and carrier frequencies of 13.565 mc and 18.208 mc. A comparison was made, on a dual trace oscilloscope, of the outputs from the ISCAN array and a typical broad beam receiving antenna. The display showed that the broadbeam antenna might pick up from one station as many as 4 or 5 multipath signals at certain times, with delays of one to several milliseconds between them. It was possible, by beam selecting with the ISCAN array, to receive any one of these signals and essentially reject the others.

Figure 15 shows a photograph of the ISCAN visual display for a long range signal taken at three different times about 30 seconds apart. A short period of system operation using the oscilloscope display is sufficient to show that multipath occurs frequently and that rapid changes in angle of arrival can occur. The multiple narrow beams of the ISCAN system along with the provision for rapid sequential sampling and visual display offer an advanced capability for the study of ionospheric propagation when combined with data gathered by other research tools. The primary objective is, of course, improving long distance communication.

Reliability

The operational period has not been sufficiently long to fully evaluate the system reliability, noise figure, and intermodulation distortion. Nevertheless, a design study reveals that the system itself is inherently reliable because of the redundancy of signal paths in the equipment. Each of the 24 signals routed to a combiner arrives via a separate antenna, cable, amplifier, and delay line. A malfunction in any one or even several signal paths will not appreciably affect the overall system performance. Should several of the signal paths in the equipment cease to function, the effect is to reduce the array aperture causing a reduction of antenna gain and an increase in beamwidth and sidelobes. But the system will continue to operate with only moderate degradation in performance.

Mention has been made previously that the beam forming system involved no complicated electronic or mechanical phase shifting. The simplicity of the overall ISCAN system promises a rather high degree of system reliability.

Conclusions

Satisfactory performance has been achieved from the ISCAN array although system evaluation has not been completed. Data obtained to date proves that it is possible to produce broadband, multiple antenna beams from an array using tapped delay lines and to achieve inertialess beam steering by electronic sampling of these multiple overlapping beams. The ISCAN system has sufficient directivity to discriminate against multipath and interference and thereby enhance long-distance, h-f communication. The antenna system is inherently very reliable because it makes use of redundant signal paths. It has been experimentally verified that phase and amplitude errors can be controlled within acceptable tolerances. The non-uniformly spaced, unfilled array greatly reduces cost without sacrificing performance.

Other significant accomplishments are the development of a ground-independent antenna element, very accurate measurement of electrical cable length, development of r-f amplifiers (some transistorized) with good noise figure and phase tracking, and design of a visual display which permits the system operator to observe the relative outputs of the multiple beams at all times.

The experimental antenna is only a first step towards development of an operational system. A final version would include automatic beam selection, provision for angular diversity, and possibly polarization diversity or automatically adjusted optimum-polarization.

Acknowledgement

The authors wish to acknowledge the contributions of F. Triolo and G. Vogt of the Signal Corps, K. Meredith, J. Kuecken and W. Munsey of Avco, Electronics and Ordnance Division, as well as those of the Avco Electronics Research Laboratory who were associated with the Study Phase of this program.

References

1. H. T. Fries and C. B . Feldman, "A Multiple Unit Steerable Antenna for Short-Wave Reception", Proc. IRE, Vol. 16, July, 1937.
2. Avco Corp., "Study Report - Inertialess Steerable Communication Antenna", 15 December 1959, Contract DA-36-039-SC 78293.
3. Avco Corp., "Study Report Addendum - Inertialess Steerable Communication Antenna", 31 May 1960, Contract DA-36-039-SC 78293.
4. Avco Corp., "Final Report - Inertialess Steerable Communication Antenna", 1 November 1961, Contract DA-36-039-SC 78293, Astia Document No. AD 265 579.
5. Reference (3) pages 31 - 36.
6. Patent Disclosure, "Delay Line Phasing of Linear Arrays", by H. Brueckmann; Docket No. 11,982, USA Patent Agency, dated 13 April 1960.

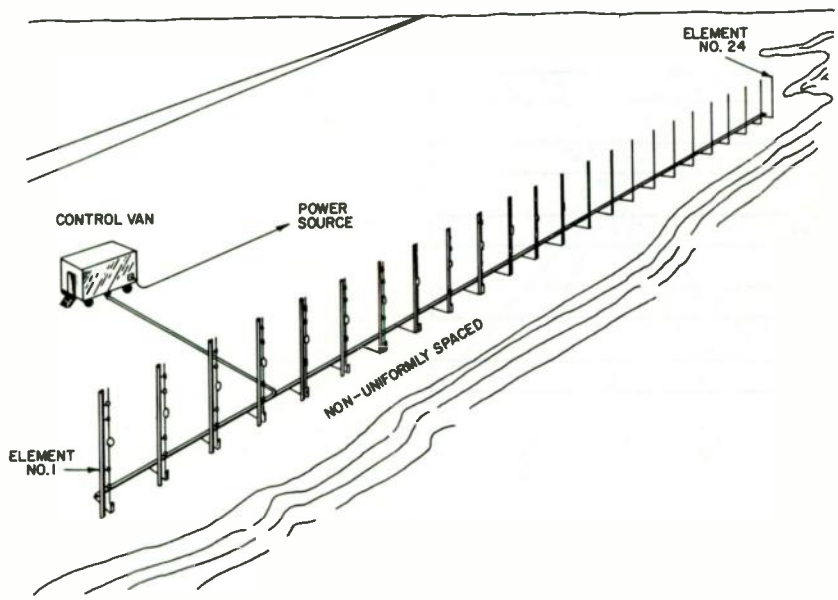


Fig. 1 Sketch of ISCAN System

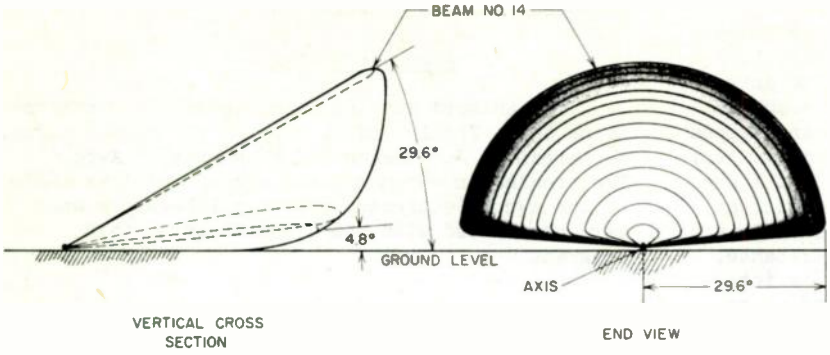


Fig 2 Array Beam Patterns, Pictorial Representation

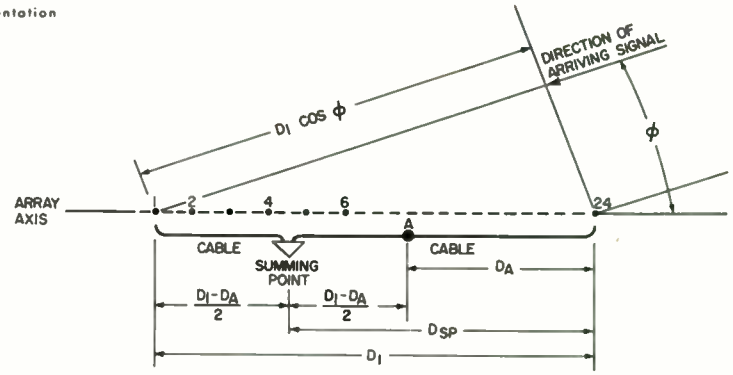


Fig. 3 Geometry of ISCAN Array

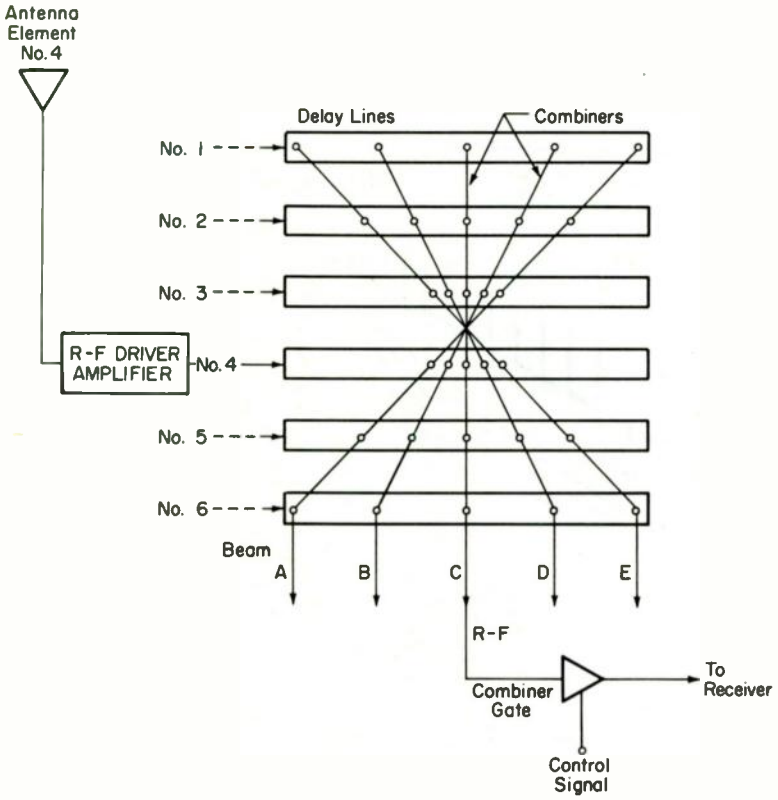


Fig. 4 Beam Forming by Tapped Delay Lines, Simplified Diagram

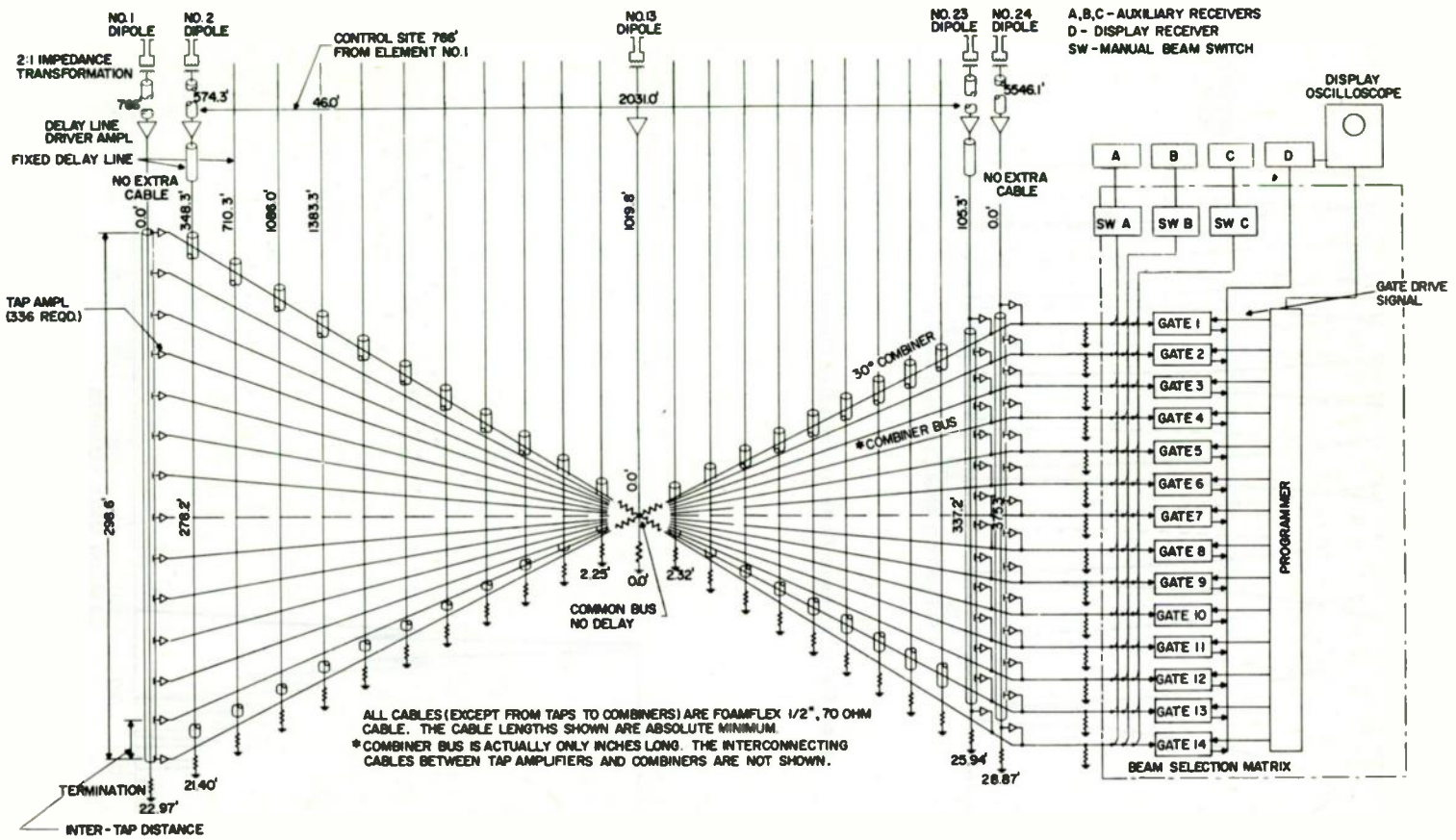


Fig. 5 System Diagram

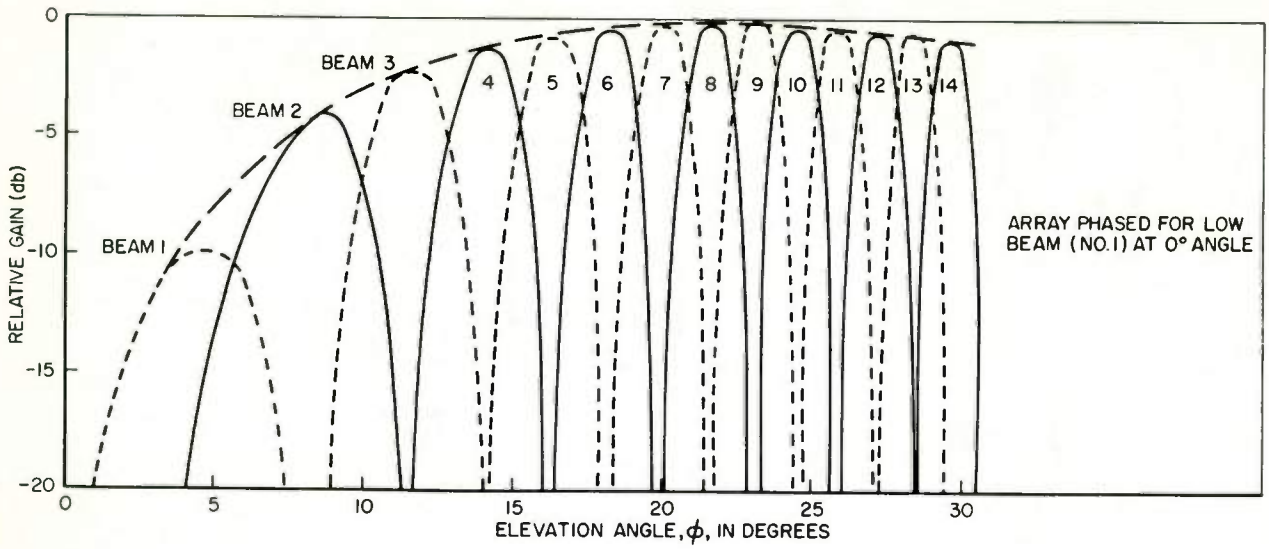
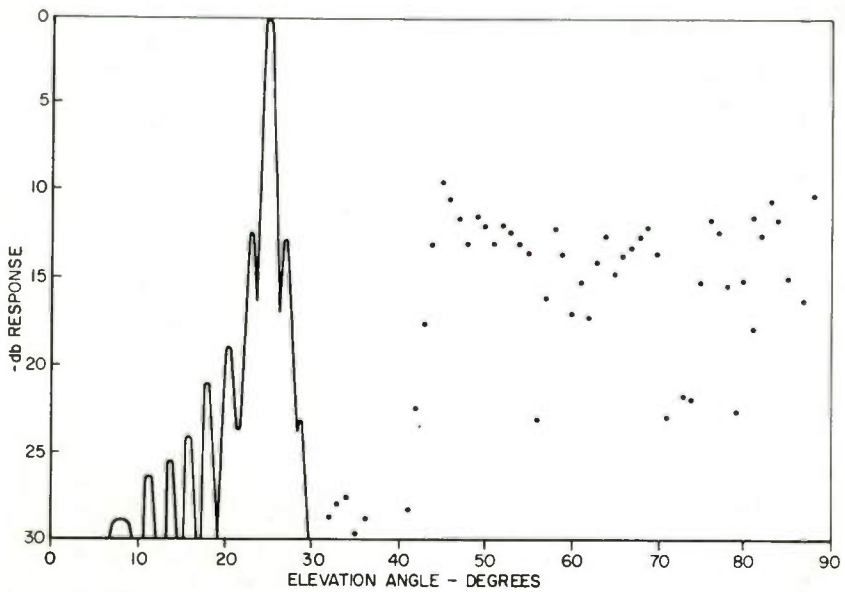


Fig. 6 Elevation Beam Patterns, Including Element Factor



(The Response From 0 to 30 Degrees Has Been Ploted on The Basis of Calculations Every 0.2 Degrees Higher Angles Have Been Calculated And Plotted Only Every Degree.)

Fig. 7 Calculated Antenna Pattern for Twenty-Four Dipole Elements, Non-Uniformly Spaced, with Illumination, Steered to 25°

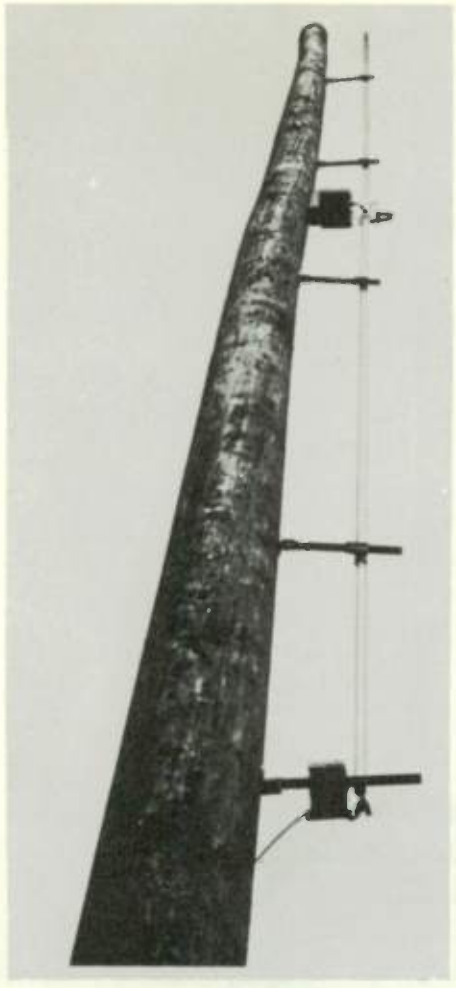


Fig. 8 Dipole Antenna Element

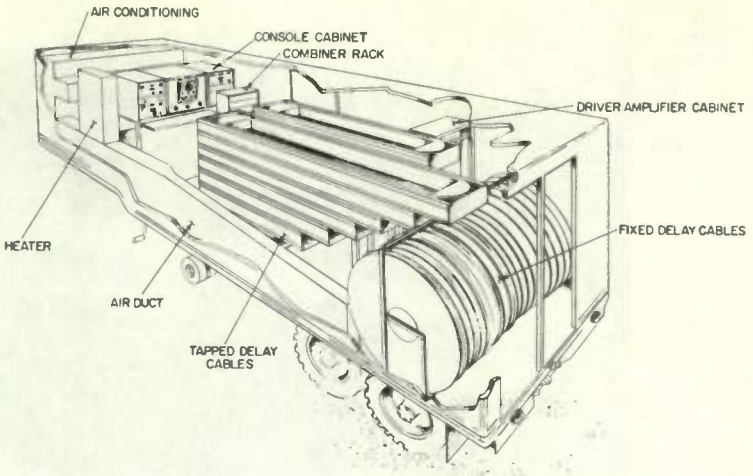


Fig. 10 General Layout of Control Van

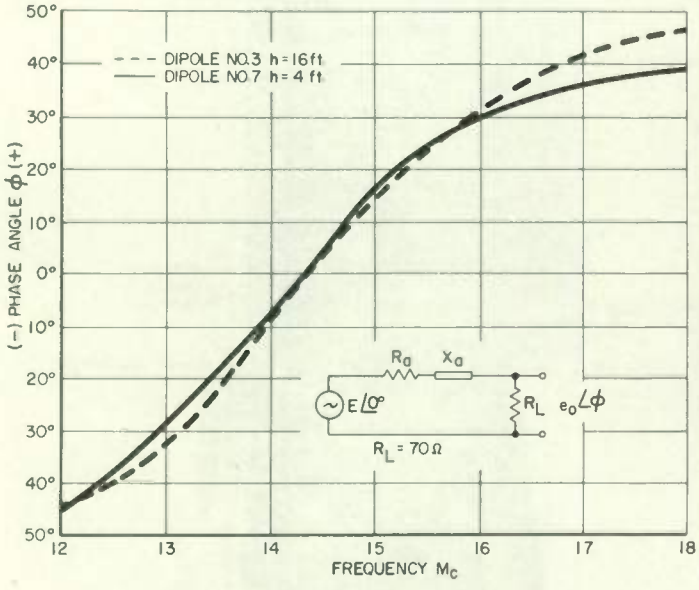


Fig. 9 Phase Shift versus Frequency for Two Antenna Elements at Different Heights

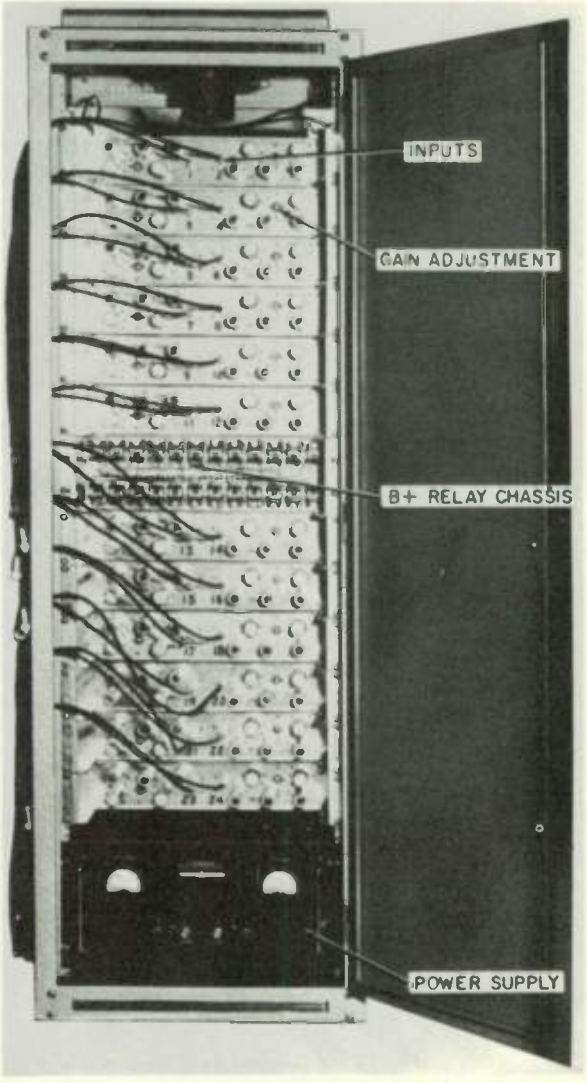


Fig. 11 Front of Driver Amplifier Rack

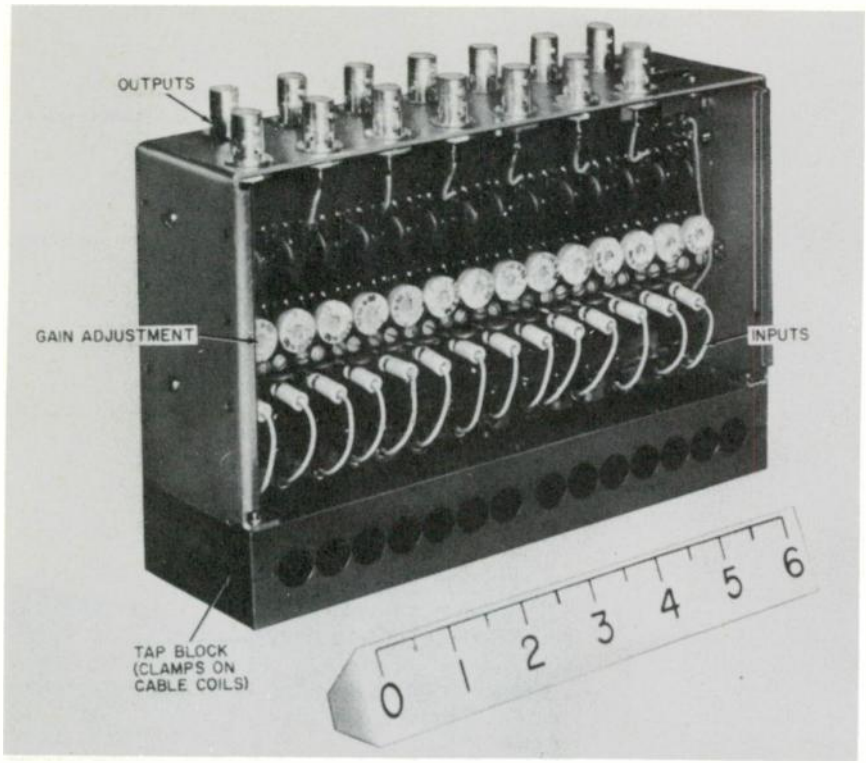


Fig. 12 Rear View of Tap Amplifier Chassis

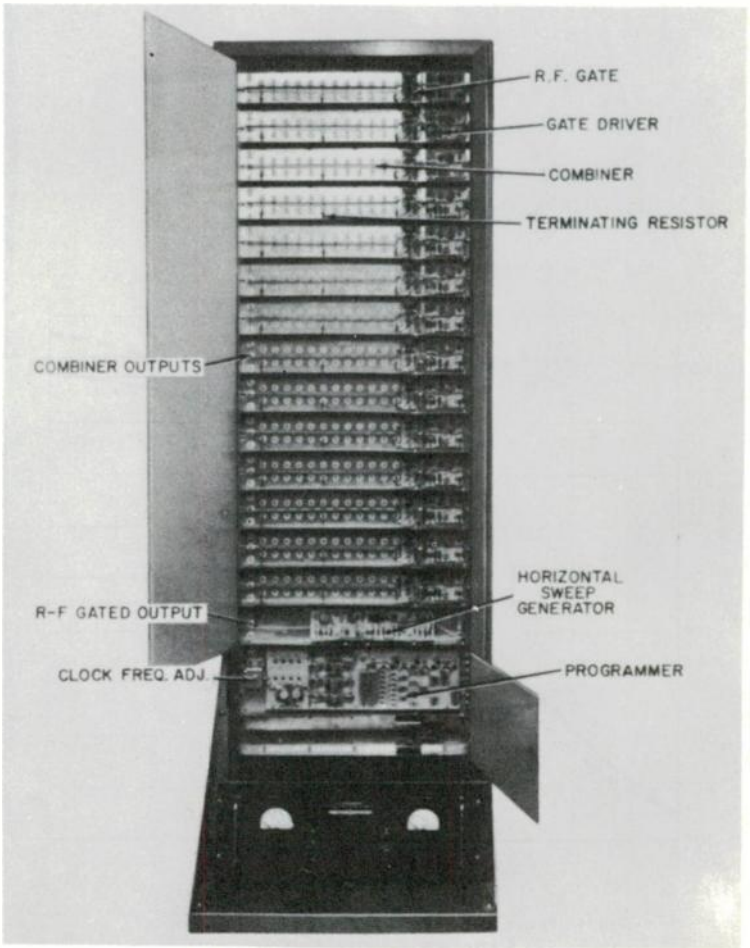


Fig. 13 Combiner Rack, Rear View

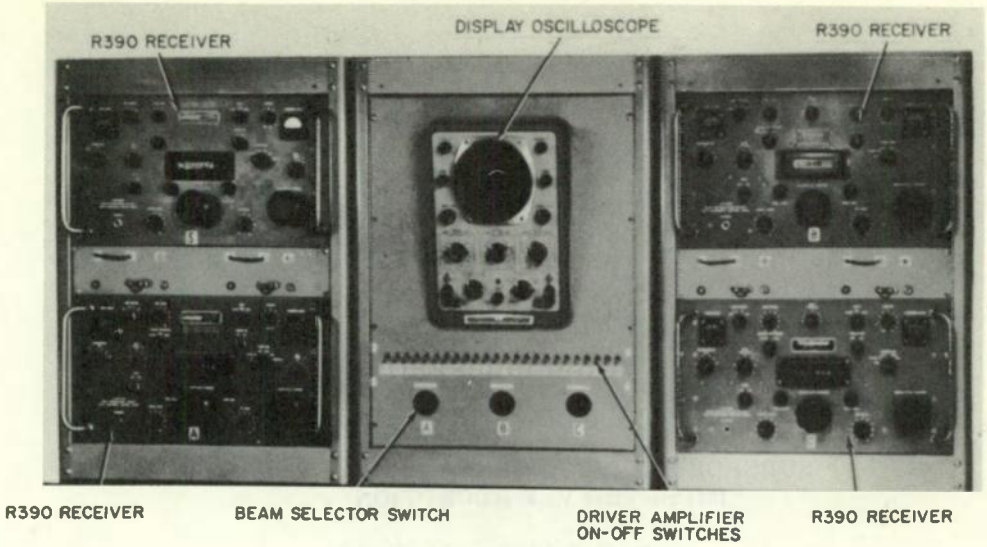


Fig. 14 Display and Operation Console

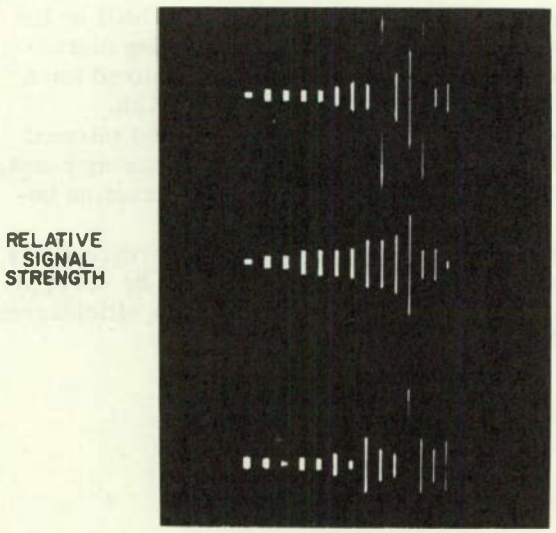


Fig. 15 Visual Display Patterns

SUPERDIRECTIVE ANTENNA ARRAYS FOR IMPROVED VLF RECEPTION

R. O. Schildknecht, Sichak Associates, Nutley, N. J.

Abstract

Superdirective antenna arrays offer a practical and significant improvement in the reception of VLF radio signals. Based on supergain techniques, such arrays enhance the SNR by the use of a highly directional pattern without occupying more than a few per cent of the space that would be required for a conventional array, and do not degrade the bandwidth.

Supergain techniques are not usually of practical interest because of unavoidable losses in efficiency and in many cases, drastic reduction in bandwidth resulting from interaction between elements.

It is shown that most of the problems with superdirectivity are peculiar to the case of efficient antennas, and do not apply to VLF receiving antennas in which extremely low efficiencies are easily tolerated.

THE COUPLING AND MUTUAL IMPEDANCE BETWEEN CONICAL LOG-SPIRAL ANTENNAS IN SIMPLE ARRAYS

John D. Dyson
Electrical Engineering Laboratory
University of Illinois
Urbana, Illinois

Summary

The radiation characteristics of the conical logarithmic spiral antenna make it attractive as an element for circularly polarized arrays. The design of such arrays requires a knowledge of the radiation coupling and mutual impedance between these log-spiral elements as a function of rotation as well as element spacing, and for some array configurations a knowledge of the location of the phase center of the elements.

A general consideration of the conical log-spiral antenna as a locally periodic, slow wave structure is outlined, and in so-doing a quasi-empirical formula for the location of the antenna phase center is developed.

An investigation of the mutual impedance between the log-spiral antennas has shown it to be very low. Information on this impedance and on the coefficient of coupling between antennas is supplied for two different geometrical array configurations and for various antenna parameters. Pattern modification due to the presence of parasitic elements has been investigated. For element to element spacing of at least one half wavelength, this consists primarily of an increase in beamwidth. Radiation patterns for several arrays, including one 4 element conical array are shown. This latter array should be of interest as a primary antenna or as a feed for a large parabolic reflector in telemetry applications.

1. Introduction

There are many applications which require a circularly polarized field with narrower beamwidths and higher gains than can be secured with the individual planar and conical logarithmic spiral antennas which have been developed to date[1]. Many of these applications have been filled by constructing arrays of helical antennas. For wide band applications, it would appear that a considerable advantage could be obtained from the extremely wide bandwidth of the conical log-spiral antenna as an array element. Thus, although the array factor of conventional arrays would be a limiting factor, the element factor could be made essentially independent of frequency over a considerable band of frequencies. In addition to bandwidth, the conical log-spiral antenna is unique in that the radiated fields are essentially circularly polarized over considerable angles off the axis of the antenna.

There are various geometrical configurations that can be used when arraying the conical antennas.[2] Four of these are shown in Figure 1. The parallel array and the collinear array would have a frequency dependent array pattern; however, the conical array and coaxial arrays could theoretically be designed to be independent of frequency. The conical array has the same geometry as that used by DuHamel and Berry when arraying log-periodic elements.[3] This geometry limits the array to a few log-spiral elements; therefore, to obtain very narrow beamwidths from an array the conventional parallel configuration may offer the only solution. In addition to the use of arrays as primary radiators, consideration has recently been given to using an array of two contrawound conical log-spiral antennas as the feed for large parabolic tracking antennas.[4-6]

The design of arrays of conical antennas requires a knowledge of the radiation coupling between antennas to determine the effect on the element pattern, and the mutual impedance between antennas to compute the element terminal impedance. Since a simple rotation of these antennas about their axes allows phasing of the array, a knowledge of the radiation coupling and mutual impedance as a function of rotation as well as a function of element spacing is desirable. The determination of the element to element separation requires a knowledge of the location of the element phase center. The purpose of this paper is to supply this information for parallel and conical linear arrays over a restricted, although representative, range of antenna parameters.

2. Coupled Antennas

Consider a two element array in which currents I_1 and I_2 flow as a result of voltages V_1 and V_2 applied across the feed terminals of the two elements. As far as these terminal voltages and currents are concerned, the two antennas may be represented by the general four-terminal network in Figure 2a.

In this network, Z_{11} is the mesh impedance of mesh 1 (less the generator impedance) and is the impedance of antenna 1 with antenna 2 present and open circuited (not equal to the self-impedance of antenna 1 when isolated). Thus, we have

$$V_1 = Z_{11} I_1 + Z_{12} I_2$$
$$V_2 = Z_{21} I_1 + Z_{22} I_2 \qquad (1)$$

where Z_{12} is the mutual impedance between the antennas. In the usual case where the medium is

reciprocal, $Z_{12} = Z_{21}$. If $V_2 = 0$, the input presented to the generator at reference plane A_1 by antenna 1 with antenna 2 present is

$$\frac{V_1}{I_1} = Z_{11} - \frac{Z_{12} Z_{21}}{Z_{22}} \tag{2}$$

Rearranging; we can express this ratio as

$$\frac{V_1}{I_1} = Z_{11} (1 - k^2) \tag{3}$$

where

$$k = \frac{Z_{12}}{\sqrt{Z_{11} Z_{22}}} \tag{4}$$

The complex number k is defined to be the coefficient of coupling between antennas one and two[7-8]. For identical antennas $Z_{11} = Z_{22}$ and

$$Z_{12} = k Z_{11} \tag{5}$$

If antenna 1 is driven and antenna 2 open circuited, $I_2 = 0$, and it can be shown that

$$k = \frac{V_2}{V_1} \tag{6}$$

Thus the measurement procedure is reduced to the determination of one impedance and the complex ratio of two voltages.[9] This method has advantages of accuracy and ease of measurement over the short circuit-open circuit method or the symmetric-antisymmetric method when measuring low values of mutual impedance. These advantages have been pointed out by Stratoti and Wilkinson who measured the coupling between short helices.[10]

The impedance of the antenna and its mutual impedance with the other antenna, referred to the feed terminals, are the basic quantities of interest. However for the conical log-spiral, which radiates a broad lobe off the apex of the cone, it is not practical to attempt to measure the magnitude and phase of the voltage across the feed terminals which are situated at this apex. Thus it is necessary to measure the coupling and mutual impedance at some length of a transmission line and, taking into account line loss, transform these quantities down this line to the antenna terminals. Although it is not convenient to establish and measure the open circuit voltage at reference plane A_2 (apex terminals of antenna 2), it is still desirable to use the impedance-voltage-ratio method, hence we force I_2', the current at the desired reference plane B_2 in Figure 2b to be zero. For $I_2' = 0$, and $Z_{11} = Z_{22}$, the terminal impedance Z_{11} and the mutual impedance Z_{12} can be expressed in terms of the measurable quantities Z'_{11}, Z'_{22}, and Z'_{12} at reference planes B_1 and B_2, and the characteristic impedance and length of the transmission lines involved.[11]

The coupled energy measured at the antenna terminals may be only a partial indication of the interaction of antennas operated in an array. This is true in particular for traveling wave antennas as has been recognized by several authors. Recently Rupp[12] has proposed that this measured characteristic be referred to as the "terminal coupling factor" to differentiate it from a "pattern modification factor". He aptly points out that since traveling wave antennas may couple to each other in a directional manner a pair of such antennas may be considered to be a radiating directional coupler. Under certain conditions it would thus be possible for more coupled energy to be radiated than transmitted to the terminals. For these reasons a study was made of the modification of the radiation pattern due to the presence of parasitic elements.

3. The Phase Center

In the parallel array the effective spacing of the antennas is simply the separation between the axes of the cones. In the conical array the effective separation is the spacing in wavelengths between the phase centers of the antennas. Thus a knowledge of the relative position of the phase center is required to calculate conical array patterns. In this section we shall outline a general application of the "backward-wave concept" to the conical log-spiral antenna and in so doing develop a quasi-empirical formula for the location of the phase center of the radiated field.

Mayes, Deschamps, and Patton[13] introduced the basic idea that the logarithmic periodic antenna could be considered to be a locally periodic structure whose period varies slowly, increasing linearly with the distance to the point of excitation. It was pointed out that for normal operation there is a radiation region with a phase of excitation such that the structure will radiate toward the feedpoint. Among such structures the backward wave zig-zag antenna is the counterpart of the log-periodic zig-zag, and the backward wave bifilar helix the periodic counterpart of the conical log-spiral antenna. To apply this concept to the conical log-spiral antenna, it is instructive to consider some of the techniques which have been useful in the study of helices.

Consider the truncated portion of the wire version of the conical log-spiral antenna shown in Figure 3. This structure is of the same basic geometry as the helix, and a local period of the structure can be defined in terms of a helix with parameters equivalent to the corresponding log-spiral parameters averaged over that period. The study of periodic structures such as the helix has been facilitated for a number of years by the use of the "k-β" diagrams to display the frequency variation of the propagation constant along the structure. Recently this approach has been applied to radiating structures.[14-15]

The solution of the determinantal equation for the real phase constants on the helix, as a function of ka, requires that the fields go to zero amplitude as the distance extends to infinity. As a result we require

$$|\beta_n a| > ka \qquad (7)$$

where β_n is the phase constant of the n-th space harmonic on the structure, k is the phase constant of wave propagation in free space and "a" is the radius of the helix. Thus there are regions on the k-β diagram where the roots of the transcendental equation are real, corresponding to slow waves. Similarly this condition produces so-called "forbidden regions" in which solutions are complex.

The boundaries of these regions can be expressed by the inequalities[16]

$$|m| + \frac{ka}{\cot \xi} \leq \frac{|\beta_n| a}{\cot \xi} \leq |m| - \frac{ka}{\cot \xi}$$

and (8)

$$\frac{ka}{\cot \xi} \leq \frac{N}{2}$$

where for multi-wire helices in the higher modes

$$m = -1 + n N \qquad (9)$$

ξ is the pitch angle, N is the number of wires or arms, and n is an integer. For a two wire symmetrical helix with the arms excited with a 180^o phase reversal, the k-β diagram takes the form of Figure 4.

The parameters normally plotted for the helix are the pitch normalized to the wavelength in free space and to the wavelength of the surface wave. On the conical spiral the pitch distance on the surface is p', and

$$\frac{p'}{\lambda_o} = \frac{ka}{\cot \xi} = \frac{ka}{\tan \alpha}$$

(10)

$$\frac{p'}{\lambda_s} = \frac{\beta a}{\tan \alpha}$$

where a is the average radius of the period with pitch p', λ_o the wavelength in free space and λ_s the wavelength of the surface wave. These are the parameters used in Figure 4.

Sensiper has shown that a good approximation to the locus of phase constants on the helix for ka in the range of interest here, is a straight line drawn through the origin with slope

$$\frac{ka}{\beta a} = \sin \xi \qquad (11)$$

and the boundaries of the m = 1 forbidden region. Such an approximation assumes propagation of energy with the speed of light along the arms. Although the determination of the propagation constant along the surface of the cylindrical helix applies to the infinite structure, it has been shown that the solution is useful for interpretation of the characteristics of the finite monofilar endfire helix[17] and the finite backward wave bifilar helix[18] with thin wire arms.

On the conical spiral the ratio of the velocity of the slow wave to that in free space is given by

$$\frac{ka}{\beta a} = \frac{p'}{\ell} = \cos \alpha \qquad (12)$$

The radius "a" of the cylindrical helix is a constant, therefore as the frequency of operation is increased the parameter ka increases. Under the above approximation, as ka is increased the propagation constant of the wave propagating along the surface of the structure away from the point of excitation increases and as it approaches that of the 1st backward space wave there will be strong coupling between these waves. When coupling begins the propagation constant will depart from the straight line approximation. The slope of this curve at any point on the curve, is the ratio of the group velocity of the surface wave, v_g, to that in free space and hence when

$$\frac{d(ka)}{d(\beta a)} = \frac{v_g}{v_o} = 0 \qquad (13)$$

propagation in the surface wave stops. This point corresponds to one end point of a stop band. For ka above this point the propagation constant is complex.[19] In the open structure this implies a transfer of energy to the backward space wave, and experimental evidence indicates that the effective phase center of the antenna is located in this region in normal backward wave radiation. As ka is increased further it is possible to force the beam to scan through the visible region of the spectrum.

In the above discussion we have considered operating conditions with a change in ka. On a periodic structure such as the helix, with constant radius and pitch angle, a change in ka implies a change in the frequency of operation. On the conical log-spiral antenna the spiral angle α is constant but the radius "a" is a linearly varying function of the distance from the apex along the conical surface. Thus a change of ka can result from either a change in position on the cone or from a change in frequency. Further, on the structure of infinite length any particular value of ka can be found for any frequency of operation. Thus, to consider the behavior of the structure one of these two variables must be held fixed. Again, we will consider only those waves predicted by the straightline approximation above, since there is experimental evidence indicating that only the lowest order mode consistent with the excitation and winding of the arm contributes significantly to the radiation pattern of the two arm balanced conical antenna when operated such that there is negligible "end effect".[20]

If we assume operation at a fixed frequency, a wave propagating along the arms away from the apex implies an increasing ka. Hence at some point on the structure, the propagation constant of the surface wave associated with the flow of energy along the arms increases until there is

coupling of energy to the backward space wave and rapid radiation. Experimental evidence indicates that the position of the phase center, expressed in wavelengths to the apex, remains essentially fixed.[2] Therefore since a/λ_o at the phase center remains constant, ka remains constant for all frequencies of operation, excluding those frequencies where normal operation is distorted by the finite truncations of the antenna, i.e. end-effect. The operating point of the frequency independent antenna does not scan along the line $ka = \beta a \cos \alpha$ with a change in frequency; instead the parameters ka and βa scale with frequency.

In Figure 4 if we calculate an expression for the intersection of the asymptote

$$ka = \beta a \cos \alpha$$

and the edge of the m = 1 forbidden region, we obtain

$$ka = \frac{\sin \alpha}{1 + \cos \alpha} \tag{14}$$

which is the circumference such that the antenna is phased to radiate backward along the surface of the cone. However, a study of the symmetry of the conical antenna indicates that all radiation from that along the surface of the cone out to that at an angle of θ_o from the surface will radiate in the backward direction. Thus it can be shown that the average circumference, of the period such that the following relationship is satisfied is the approximate region on the conical log-spiral structure that is phased for backward radiation.

$$\frac{\sin \alpha}{1 + \cos \alpha} \leq ka \leq \frac{\sin \alpha}{1 + \cos \alpha \cos \theta_o} \tag{15}$$

If the efficiency of excitation of the first backward space wave is high enough it can be postulated in the light of the above discussion that the radiating region is confined to approximately that circumference which satisfies equation (14). However, Sensiper has shown that the approximation due to the use of the asymptote and the region boundaries results in a ka that is slightly high since the coupling between these waves forces the curve down. Thus, this intersection point can be considered as only an upper bound for the start of the backward wave radiating region.

The measured phase centers of several antennas with narrow constant width arms and with cone and spiral angles varying from $15^o \leq \theta_o \leq 30^o$ and $60^o \leq \alpha \leq 83^o$ are plotted in Figure 5. These phase centers occur at a ka considerably below the intersection point and their location on the k-β diagram may be approximated by a straight line.

The orientation of this line is interesting. The distance between the intersections of the asymptote for that particular antenna and this line, and the edge of the m = 1 forbidden region is directly related to the beamwidth of the antenna. As this distance decreases more of the active region is moved into the visible region of the spectrum and more energy is radiated at an angle from the surface. Wire antennas wound with an α of 60^o have a wide beamwidth (on the order of 120^o) while those with $\alpha = 83^o$ have patterns with beamwidths on the order of 70 degrees. The line also implies that at a spiral angle α of 45^o the antenna at the phase center (and hence we could expect the major portion of the active region) is phased to radiate at an angle from the axis of the antenna. In agreement with this, it has been determined that thin wire antennas wound with an α of 45^o have multiple side-lobes and a major portion of the radiation directed in an endfire direction.

The data plotted in Figure 5 is for wire antennas. From pattern information, and a previous study of the near fields on the planar antenna, we could expect the slope of the approximate locus of the phase centers for antennas with wider exponentially expanding arms, to become more negative and thus more nearly parallel to the region boundary, with a minimum change for large α.

If we calculate an expression for the intersection of the phase center line and the asymptote taking into account the cone angle θ_o, we get

$$ka \simeq \frac{1.2 \sin \alpha}{1.4 + \cos \alpha \cos \theta_o} \tag{16}$$

This expression is plotted in Figure 6. A family of curves, one for each value of θ_o could be plotted, but for antennas that are good unidirectional radiators the cone angle varies as $10^o \leq 2\theta_o \leq 45^o$ and therefore $.996 \geq \cos \theta_o \geq .924$. Thus $\cos \theta_o \simeq 1$, and it is felt that the single curve is sufficient.

A recent study of the near fields on the conical log-spiral antenna by McClelland indicates good agreement between the measured characteristics of the near fields and those that might be predicted in terms of the applications of the above concepts.[21]

4. Experimental Considerations

The conical log-spiral antenna with controlling parameters and an associated coordinate system is shown in Figure 7. For the present investigation a series of antennas with parameters $2\theta_o = 15^o$, D = 17.5 cm, d = 4.5 cm, h = 47.3 cm and α = 60, 73, and 83^o were used. The antennas were constructed from 1/4 inch copper tubing supported by 1/2 × 1 inch wood struts as shown in Figure 8. The feed cable RG-141/U was carried through one of the arms to the apex of the cone.

All measurements were made at a frequency of 610 mc which, for these antennas, was near the low end of the range of frequencies where the antennas could be expected to operate in a frequency independent manner. Thus the active region

was at the lower portion of the antenna and in the case of parallel antennas the coupling could be expected to be the tightest.

5. Measured Coefficient of Coupling

The magnitude of the coupling coefficient for parallel antennas is plotted in Figure 9. The measured coupling as a function of separation shown in Figure 9a is low, approximately -27 db for α of 60°, and -41 db for α of 73° and 83° at S = $\lambda/2$. The coupling rapidly decreases as the spacing increases, with antennas having a wider beamwidth (α = 60°), being more closely coupled than those with a narrower beamwidth. There is some scatter of the measured points for larger separations. This was attributable to back scatter at the site. Although the antennas were outdoors, it was determined that there is approximately a 5 db possible error due to the site for separations of a wavelength or more where the coupling level was -50 db or greater. These spurious reflections did not prove bothersome when the separation was fixed and the parasitic antenna simply rotated.

The variation in terminal coupling as a function of relative rotation of the parasitic antenna is plotted in Figure 9c. The antennas are more closely coupled when oriented in a like manner, and loosely coupled when rotated 90° with respect to each other. Thus, although a rotation of the antennas to phase an array would cause as much as 10 db variation in the coupling, the general level of coupling is so low that, as indicated later, the change in mutual impedance is small compared to the self impedance of the antennas.

Terminal coupling between right and left hand wound antennas shown in Figure 9b is on the same order of magnitude as that between like antennas, with the coupling between the contrawound antennas actually a little higher for close spacing. The term contrawound is used in this paper to mean an antenna with a winding of the opposite sense to that of the antenna to which it is coupled. It is not used to mean one antenna with arms wound in both a right and left hand direction. The character of the curves is undoubtedly a function of the reflections at the site which return with the opposite sense of polarization. The dashed lines may be a more valid estimate of the true coupling. The data as a function of rotation does not exhibit a well-defined character and again the variations are probably due in part to reflected energy.

Figure 9e indicates that the coupling from the n-th to the n + 1 and at least to the n + 2 element must be considered. It is interesting to note that there is almost a 90° shift between curves B and C. This 90° shift was observed for 83° antennas but only a 20° shift for α = 60° antennas. The period of curve C shifts slightly along the horizontal axis with rotation of the center antenna, indicating that there must be some coupling through or reradiation from the center antenna as well as direct coupling between the outside antennas.

The coupling between conical log-spiral antennas in a conical array is considerably greater, as shown in Figure 10. Data was taken for a range of ψ, the array angle. For minimum ψ, approximately 17°, the struts of the antennas were just touching. In this position the separation angle between the surface of the cones was 2 degrees. Unlike the parallel case, the surfaces of the cones are now closely coupled over the portion of the antenna that supports a strong surface wave. In addition the tip of the parasitic antenna is in strong fields radiated back toward the tip of the driven element. Measured values of terminal coupling were approximately -18, -26, and -35 db for α of 60, 73, and 83. As the radiation patterns indicate in a later section, in this case more of the coupled energy appears to be reradiated than reflected back to the feed.

As the array angle was increased, coupling decreased to a minimum at an angle of approximately 35 to 50° depending upon the spiral angle α. A further increase in ψ increased the coupling due to the fact that the antennas were tending to radiate more energy toward each other. In the limit when ψ = 180° there is maximum coupling.

This curve should be interpreted in conjunction with the radiation patterns which indicate that the array angle will normally be limited to around 60° or less to maintain reasonably low side lobes. Since S is a function of ψ, the dual scales on Figure 10(a,b,c) may be used as a nomogram to convert ψ to spacing in wavelengths.

Figure 10b indicates less variation with rotation. The increase in variation with increasing α is due in part to the fact that these curves are taken with $\lambda/2$ between phase centers and hence the array angle (and of more significance the separation angle) is smaller for the α = 83° antennas.

For small array angles the coupling between contrawound antennas in a conical array shown in Figure 10b, is on the same general level as that for identical antennas. It however does not increase with an increase in the array angle since the two antennas are of opposite sense of polarization. Figure 10e indicates that in this array as well, the coupling to the second antenna in an array is on the same order of magnitude as the coupling to the nearest neighbor.

6. Mutual Impedance

The real and reactive parts of Z_{11} and of the mutual impedance, Z_{12}, for parallel antennas is shown in Figure 11. As the coefficient of coupling indicates, the mutual impedance is low. For an α equal to or greater than 73° the real and reactive parts for antennas with the same sense of winding are less than 1 ohm. For antennas with opposite sense of winding they are less than 2 ohms. The real part of the self

impedance of the antennas for these cases lies between 110 and 160 ohms.

The mutual impedance between elements in a conical array is larger than it is in the parallel array. At first inspection Figures 12b and c may be misleading, however pattern considerations will limit the normal range of α to approximately 60° or less and in this range the mutual impedance is quite low for the 73° and 83° antennas. At spacings up to $S = \lambda/2$ there is little difference in the general level of the mutual impedance between the same or oppositely sensed antennas.

7. Radiation Patterns -- Parasitic Arrays

The radiation patterns of the individual antennas are shown in Figure 13. Also shown is the distortion in the individual element pattern due to the presence of a parasitic antenna in parallel and conical arrays. For all patterns the feed line to the parasitic element was terminated in its characteristic impedance. There was however some mismatch between the line and the antenna. For parallel antennas the principal change is the introduction of some squint in the E_{ϕ} pattern taken in the plane of the array and a broadening of the beamwidths in this plane at a spacing of one half wavelength. As the separation is increased this pattern modification diminishes until at one wavelength the patterns are very similar to the single element pattern. Pattern modification is less when the parasitic antenna is one with an opposite sense of polarization.

As previously pointed out, when the antennas are placed in a conical array the separation between the surface of the elements is reduced and there is consequently considerably stronger coupling. For the smallest array angle used, $\Psi = 17^{\circ}$, i.e., an angular separation between the conical surfaces of 2°, pattern modification is quite severe for antennas with the same sense of polarization. Modification is much less where the parasitic element is of opposite sense although for this close spacing the patterns taken in the plane perpendicular to the array are slightly distorted. In addition the axial ratio on axis may rise sharply as indicated for $\alpha = 83^{\circ}$. Patterns taken in the plane perpendicular to the array were essentially unchanged for the same sense antennas.

The presence of a second parasitic element restores symmetry to the pattern but it again increases the beamwidth. Although the terminal coupling varies considerably with rotation of the parasitic element, the effect of such rotation on the radiation pattern is usually small compared to the effect due to the physical presence of the element.

As the array angle is increased the pattern distortion rapidly diminishes and for a separation angle of only 10° the element pattern is modified only slightly by an oppositely sensed parasitic element.

The modification in the element beamwidth of an $\alpha = 83^{\circ}$ antenna due to one and two parasitic elements is indicated in Table No. 1. Changes for 73° antennas are of the same order of magnitude as those shown for 83°, those for 60° antennas are considerably less.

Table 1

Approximate half-power beamwidth of isolated element and with parasitic and contrawound parasitic antennas present. $S = \lambda/2$, $\alpha = 83^{\circ}$, $2\theta_o = 15^{\circ}$

Antennas	Array	$\phi = 0^{\circ}$		$\phi = 90^{\circ}$	
		E_{ϕ}	E_{θ}	E_{ϕ}	E_{θ}
Element	-	68	64	72	63
With CW Par	Parallel	66	60	68	64
With Par.	Parallel	98	74	69	60
Par. each side	Parallel	119	100	65	54
With CW Par	Conical	67	54	72	63
With Par.	Conical	65	55	73	60

Most of the pattern modification can be explained in terms of array theory. A phase delay of less than $\pi/2$ in the excitation of one element of a two element array scans the beam of the array toward this unit. In Figure 13e, the separation between the element phase centers is approximately $\lambda/4$ for $\Psi = 17^{\circ}$. The radiation patterns taken in the plane of the array in this case are approximately those of a two element array with $S = \lambda/4$ and a phase delay of $\pi/4$ to one element, this would indicate that a significant portion of the energy is coupled to the parasitic element and reradiated with a phase delay. The excitation of an array with an equivalent phase advance to one element scans the beam in the opposite direction. In the parallel parasitic array in Figure 13b, the energy coupled to the parasitic antenna is delayed 180° by the separation. A further delay in reradiating the energy will appear as a phase advance to that element. The result is a squint away from the parasitic antenna.

The fact that a rotation of the parasitic element has only a secondary effect would indicate that the net phase delay, excluding element separation, in the excitation of and reradiation from the parasitic element is approximately the same for any orientation.

8. Radiation Patterns -- Driven Arrays

The electric field radiation pattern of the isolated, $\alpha = 73^{\circ}$, elements used are shown in Figure 14. These patterns may be approximated very well by $\cos^{1.7}\theta$ and $\cos^{1.3}\theta$ functions. The radiation pattern of a parallel two element in-phase array with half-wavelength spacing was calculated using an assumed element function of $\cos^{1.5}\theta$ as a compromise between the isolated

element functions of Figure 14. As shown in Figure 15 this calculated pattern is in good agreement with the measured pattern except that it is approximately one degree narrower at the half-power points. Since Figure 13 indicates that the element pattern is broadened when the element is placed in a linear array a new element function was assumed by fitting a cosine curve to these patterns. A calculated array pattern using $\cos^{.8}(\theta+4^{\circ}) + .04$ for E_{ϕ}, and $\cos^{.9}\theta$ for E_{θ}, was approximately 1 degree wider than the measured patterns.

The geometry of a conical array is indicated in Figure 16 where the electric field at a distant point is given approximately by

$$E = \cos^{n}(\theta + \frac{\psi}{2}) \exp j(\beta d \sin\frac{\psi}{2} \sin\theta) + \cos^{n}(\theta - \frac{\psi}{2}) \exp j(-\beta d \sin\frac{\psi}{2} \sin\theta) \quad (17)$$

where $\cos^{n}(\theta + \frac{\psi}{2})$ is the assumed element function, d is the separation between element phase centers, ψ is the array angle, and θ the angle from the axis of the array in the plane of the array.

The radiation pattern of a two element in-phase conical array was calculated from (19) using $n = 1.5$ and $\psi = 34^{\circ}$ ($S = .5\lambda$). This pattern is compared to the measured patterns in Figure 17. The half-power beamwidth of the measured E_{ϕ} pattern is within one degree of the calculated pattern and there is good agreement between the patterns for $\theta \pm 40^{\circ}$ degrees. Beyond $\theta = 40^{\circ}$ the measured patterns are somewhat narrower than calculated. The E_{θ} array pattern is somewhat narrower than calculated; better agreement could be secured by using a slightly modified element function.

Measured radiation patterns of several two and three element arrays are shown in Figure 18. The arrays, when excited with the elements in phase, have predictable patterns for a given element spacing. A total array angle of 60 to 70° appears to be the maximum useable angle for reasonable side lobes for both two and three element conical arrays. A three element, $\alpha = 73^{\circ}$, conical array of 15° cones with 33° between axes of adjacent elements has a pattern half power beamwidth of approximately 40°. Beyond this the beamwidth rises.

The array may be phased by a rotation of one element.[17] The similarity between the patterns in Figure 18c and Figure 13e may be noted. Rotation of the one element 180° in an excited array places a null on axis. Patterns for this case and those for the in-phase condition are essentially the sum and difference patterns. Suitable hybrid circuitry[22] permits the use of these modes simultaneously.

When interpreting the conical array patterns it should be noted that a change in the array angle ψ, which brings about a change in the spacing between the phase centers of the elements is not equivalent to a change in the element spacing in a parallel array. In a parallel array the change in radiation characteristics with frequency can be simulated by maintaining the frequency constant and varying the element spacing, provided the element pattern does not change with frequency. In a conical array this is not true since a change in ψ changes the orientation of the element pattern with respect to the array geometry. The element pattern can be approximated by a function such as $\cos^{n}(\theta + \frac{\psi}{2})$. Thus a change in S, and hence ψ effectively changes this function as well as changing the element spacing.

That the conical array geometry does preserve the beamwidth of the array with changing frequency of operation is shown in Figure 19. A "conical quad-spiral array" with representative radiation patterns is shown. The particular array shown has a nominal gain of 12 db over a circularly polarized isotropic source. The bandwidth may readily be extended to cover the 220 through 2400 mc telemetry bands and the capability of covering this complete range of frequencies with a single antenna should make it attractive for telemetering and acquisition functions. In its present form, or with two of the elements replaced by contrawound units to obtain selectable polarization, it should make an excellent feed for a large parabolic reflector.

9. Conclusions

A consideration of the conical log-spiral antenna as a locally periodic slow wave structure has led to somewhat better understanding of its characteristics and operation. Under this concept an approximate expression for the location of the phase center of the antenna has been obtained.

An experimental investigation of the coupling between conical log-spiral antennas indicates (1) that coupling is low; on the order of -30 db or greater for element to element spacing of $\lambda/2$ or more in a parallel array. (2) This coupling varies with rotation, being a minimum for a 90° rotation between elements. (3) Coupling is on the order of 10 db greater in the conical array. Minimum coupling for 15° cones occurs at an array angle of approximately 35 to 50° depending upon the spiral angle α. (4) Changes in the basic element pattern caused by the presence of other elements are minor for an element to element spacing of at least one-half wavelength, and consist mainly of a broadening of the element pattern beamwidth in the plane of the array. (5) A good approximation to the array pattern of small arrays may be gained from the use of the isolated element pattern function.

A four element conical array that should prove useful as a primary radiator or as a primary feed for a large parabolic reflector has been constructed to operate satisfactorily over an 8 to one range of frequencies. This range of

frequencies may be readily extended to cover the 220 through 2400 telemetering bands with one antenna.

10. Acknowledgement

The author would like to acknowledge many helpful discussions with members of the University of Illinois Antenna Laboratory and in particular Professors G. A. Deschamps and P. E. Mayes, and Mr. O. L. McClelland. In addition he is pleased to acknowledge the assistance of the members of the laboratory and in particular that of Mr. R. E. Griswold who made the extensive coupling and mutual impedance measurements.

This work was conducted under the sponsorship of the Aeronautical Systems Division, Wright-Patterson Air Force Base under Contract AF33(616)-6079.

References

1. Dyson, J.D., "A Survey of the Very Wide Band and Frequency Independent Antennas - 1945 to the Present" Journal of Research of the N.B.S. Vol. 66D, No. 1, pp. 1-6, Jan. - Feb., 1962.

2. Dyson, J.D., "The Conical Log-Spiral Antenna in Simple Arrays", Abstracts of the 11th Annual Symposium on the USAF Antenna Research and Dev. Program, sponsored by Aeronautical Systems Div., Air Force Systems Command, Oct. 16, 1961.

3. DuHamel, R.H. and Berry, D. G., "Logarithmically Periodic Antenna Arrays", IRE WESCON Convention Record, Part I, 1958, pp. 161-177.

4. Webster, R. and Lyles, J.F., "Application of Frequency Independent Feeds to Automatic Tracking Antennas", Abstracts of the 11th Annual Symposium on the USAF Antenna Res. and Dev. Program, sponsored by Aeronautical Systems Div., Air Force Systems Command, Oct. 16, 1961.

5. Bales, B. W. and Voklenburg, E., "Antenna Feed Has Ten-to-One Bandwidth" Electronics, Vol. 34, No. 48, pp. 56-58, Dec. 1, 1961.

6. Bresler, A.D., Jasik, H. and Kampensky, A. "A Wide Band Conical Scan Antenna Feed System", The Microwave Journal, Vol. 4, pp. 97-101, Dec. 1961.

7. King, R.W.P., "Coupled Antennas and Transmission Lines", Proc. IRE, vol. 31, November, 1943, pp. 626-640.

8. Blasi, E. A., "The Theory and Application of the Radiation Mutual-Coupling Factor", Proc. IRE, Vol. 42, July, 1954, pp. 1179-1183.

9. King, R.W.P., "Theory of Linear Antennas," Harvard University Press, Cambridge, Mass., 1956, p. 349.

10. Stratoti, A.R. and Wilkinson, E.J., "An Investigation of the Complex Mutual Impedance between Short Helical Array Elements", Trans. IRE, Vol. AP-7, July, 1959, pp. 279-280.

11. Dyson, J.D., "The Radiation Coupling and Mutual Impedance between Balanced Wire-Arm Conical Log-Spiral Antennas" Technical Report No. 54, Antenna Laboratory, Elec. Engr. Dept., University of Illinois, April 1962. (The content of the present paper is treated in greater detail).

12. Rupp, W. E., "Coupled Energy as a Controlling Factor in the Radiation Patterns of Broadside Arrays", Abstracts of the 11th Annual Symposium on the USAF Antenna Research and Dev. Program, Sponsored by Aeronautical Systems Div., AFSC, October 1961.

13. Mayes, P. E., Deschamps, G.A., and Patton, W. T., "Backward-Wave Radiation from Periodic Structures and Applications to the Design of Frequency-Independent Antennas", Proc. IRE, Vol. 49, No. 5, pp. 962-963, May 1961.

14. Oliner, A.A., and Hessel, A., "Guided Waves on Sinusoidally-Modulated Reactance Surfaces", IRE Trans., Vol. AP-7, Dec. 1959, pp. 8201-8208.

15. Ishimaru, A., and Tuan, H., "Frequency Scanning Antennas", 1961 IRE International Convention Record, Part I, pp. 101-109.

16. Sensiper, S., "Electromagnetic Wave Propagation on Helical Conductors", Tech. Report No. 194, Res. Lab. of Elec., MIT, May 16, 1951.

17. Marston, A.E. and Adcock, M.D., "Radiation from Helices", Report 3634, Naval Research Laboratory, March 8, 1950.

18. Patton, W.T., unpublished work.

19. Mayes, P. E., "Coupled Mode Analysis of Stop-Band Characteristics of Modulated Reactance Surfaces", Internal TM 62-2, Antenna Lab., Univ. of Ill., Feb. 1962 (Unpublished).

20. Dyson, J.D. and Mayes, P.E., "New Circularly Polarized Frequency Independent Antennas with Conical Beam or Omnidirectional Patterns", IRE Trans., Vol. AP-9, July 1961, pp. 334-352.

21. McClelland, O.L., "An Investigation of the Near Fields on the Conical Equiangular Spiral Antenna", M.S. Thesis, Department of Electrical Engineering, Univ. of Ill.,1962.

22. Glass, J.E., Minton, T.M., Jr., and Peeler, G.D., "A Dual Mode Spiral Antenna" Proc. of the 1961 Nat'l Aero. Elec. Conf., May 1961.

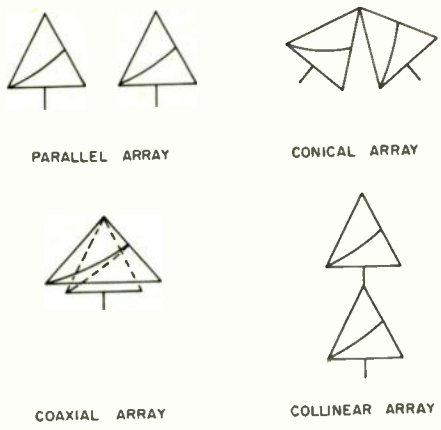


Fig. 1. Possible configurations for arraying the conical log-spiral antenna

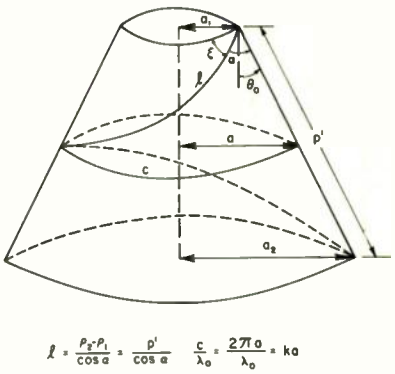


Fig. 3. Truncated portion of conical log-spiral antenna

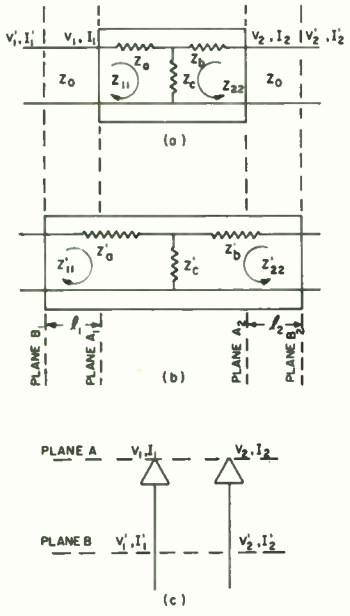


Fig. 2. Equivalent network for coupled antennas

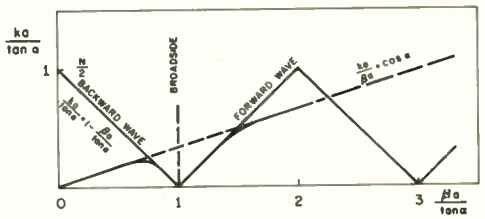


Fig. 4. "k-β" diagram for bifilar helix

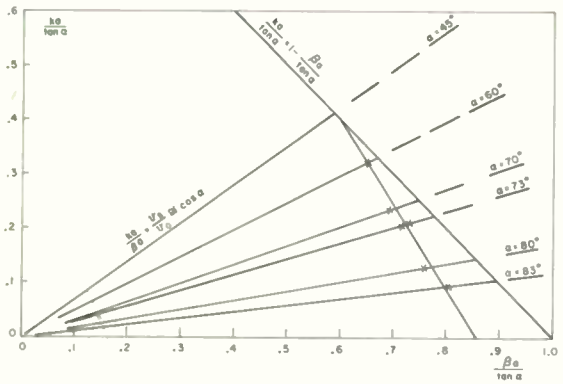


Fig. 5. Measured phase centers for several wire arm conical log-spiral antennas

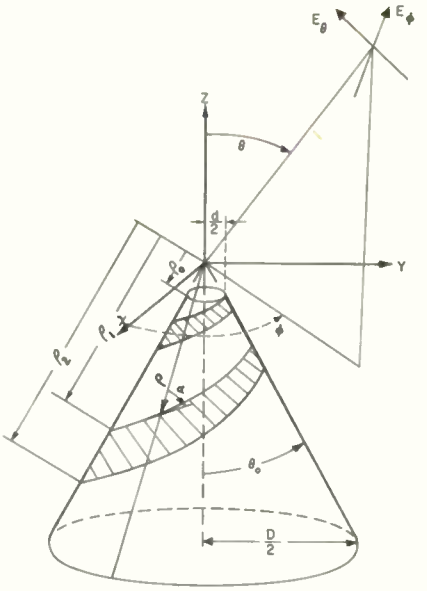


Fig. 7. A conical antenna with associated coordinate system

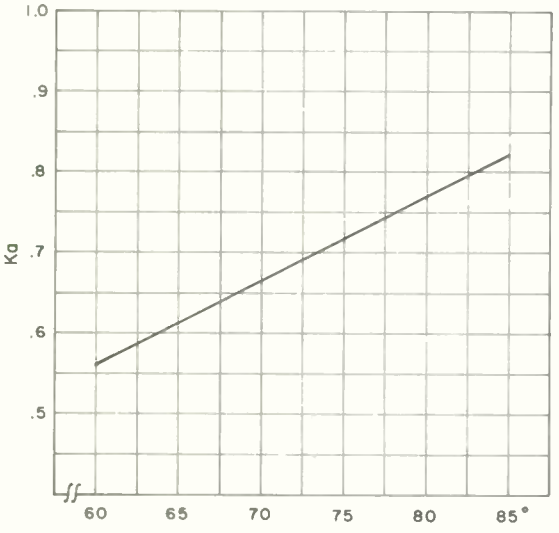


Fig. 6 Approximate circumference in wavelengths at the phase center of balanced wire arm conical log-spiral antennas $10^{o} \leq 2\theta_{o} \leq 45^{o}$

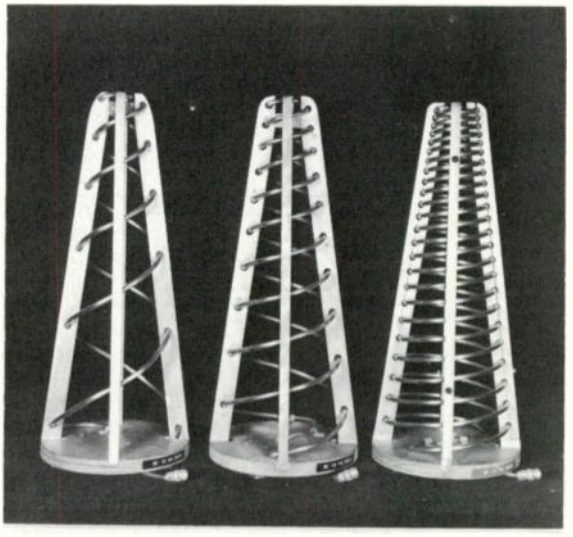


Fig. 8. Conical log-spiral antennas. α = 60,73, and 83°; $2\theta_{o}$ = 15°; d = 4.5 cm; D = 17.5 cm.

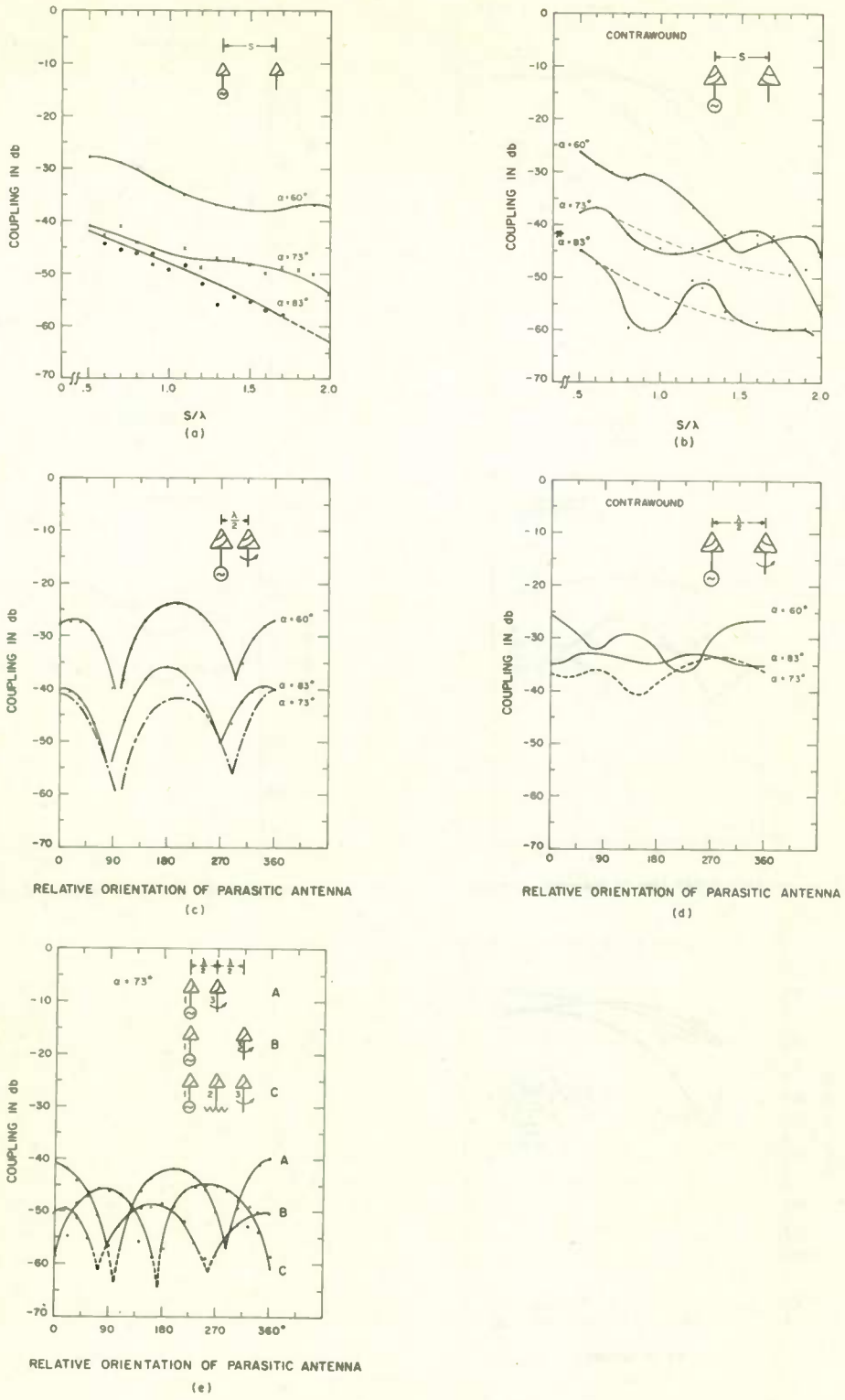


Fig. 9. Magnitude of coupling coefficient measured between parallel antennas (*Erratum: In b above add +10db to indicated coupling Vs S for $\alpha = 83^{o}$)

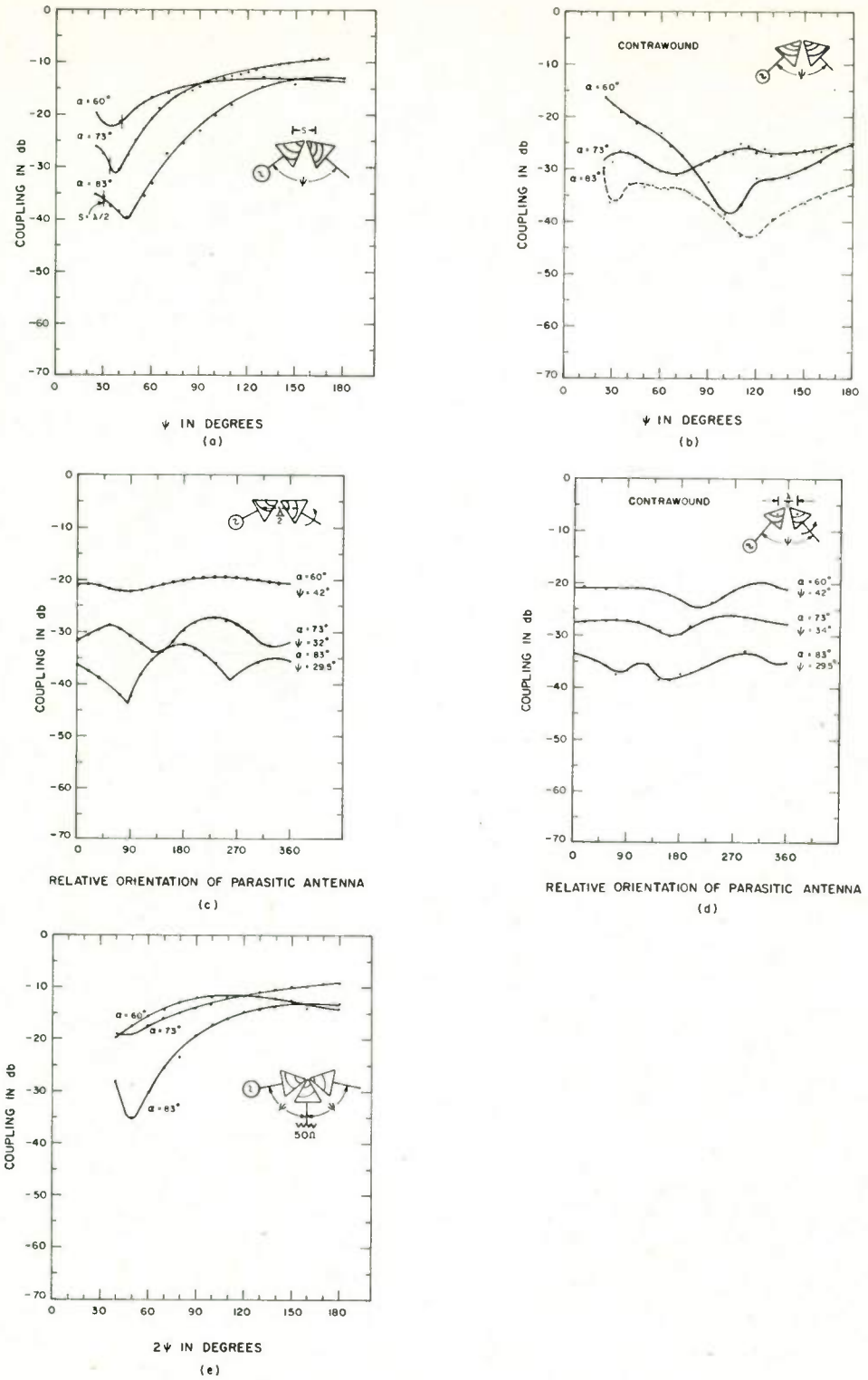


Fig. 10. Magnitude of coupling coefficient measured between antennas in a conical array

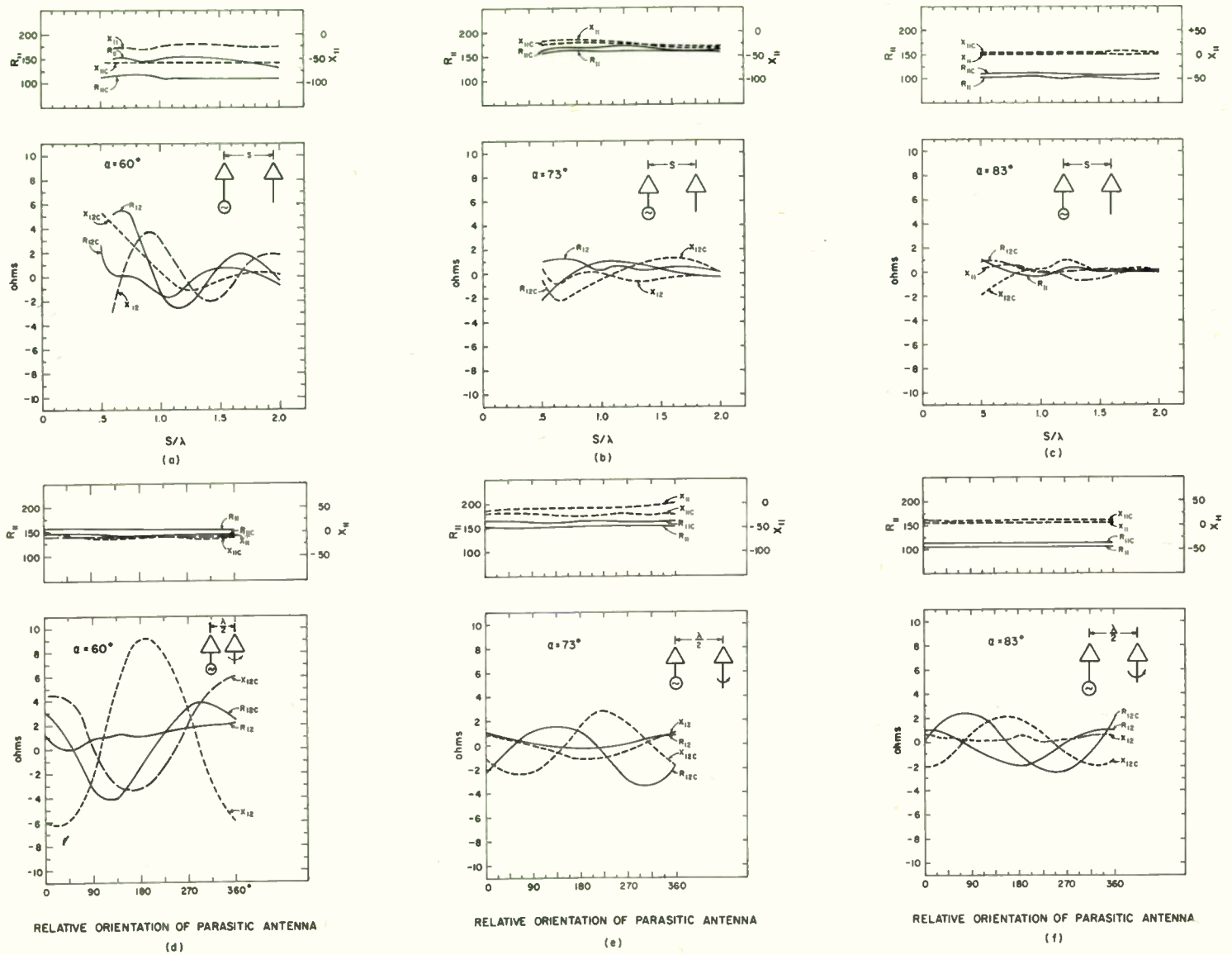


Fig. 11. Mutual impedance measured between parallel antennas

(Subscript C indicates parasitic antenna is contrawound)

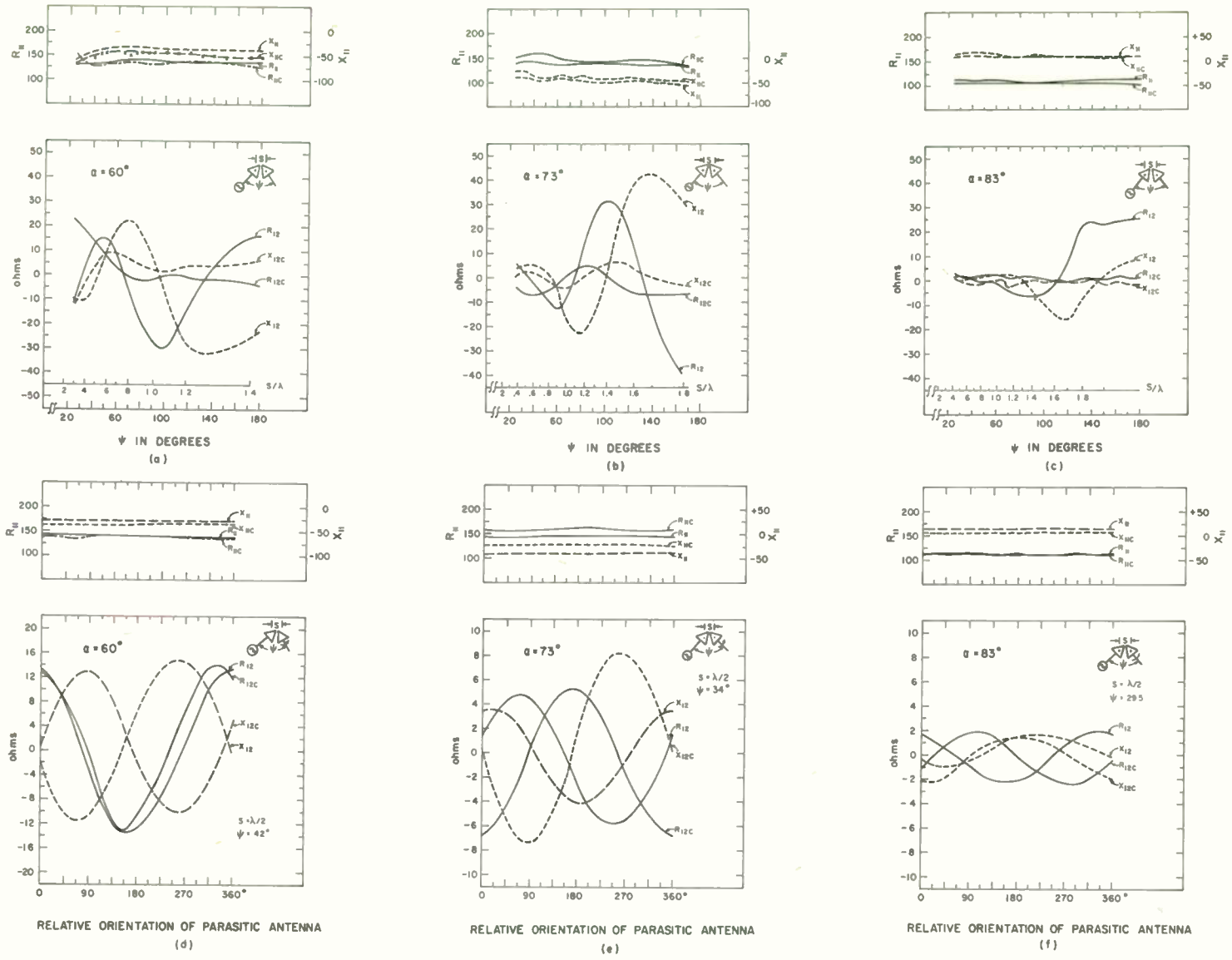


Fig. 12. Mutual impedance measured between antennas in a conical array
(Subscript C indicates parasitic antenna is contrawound)

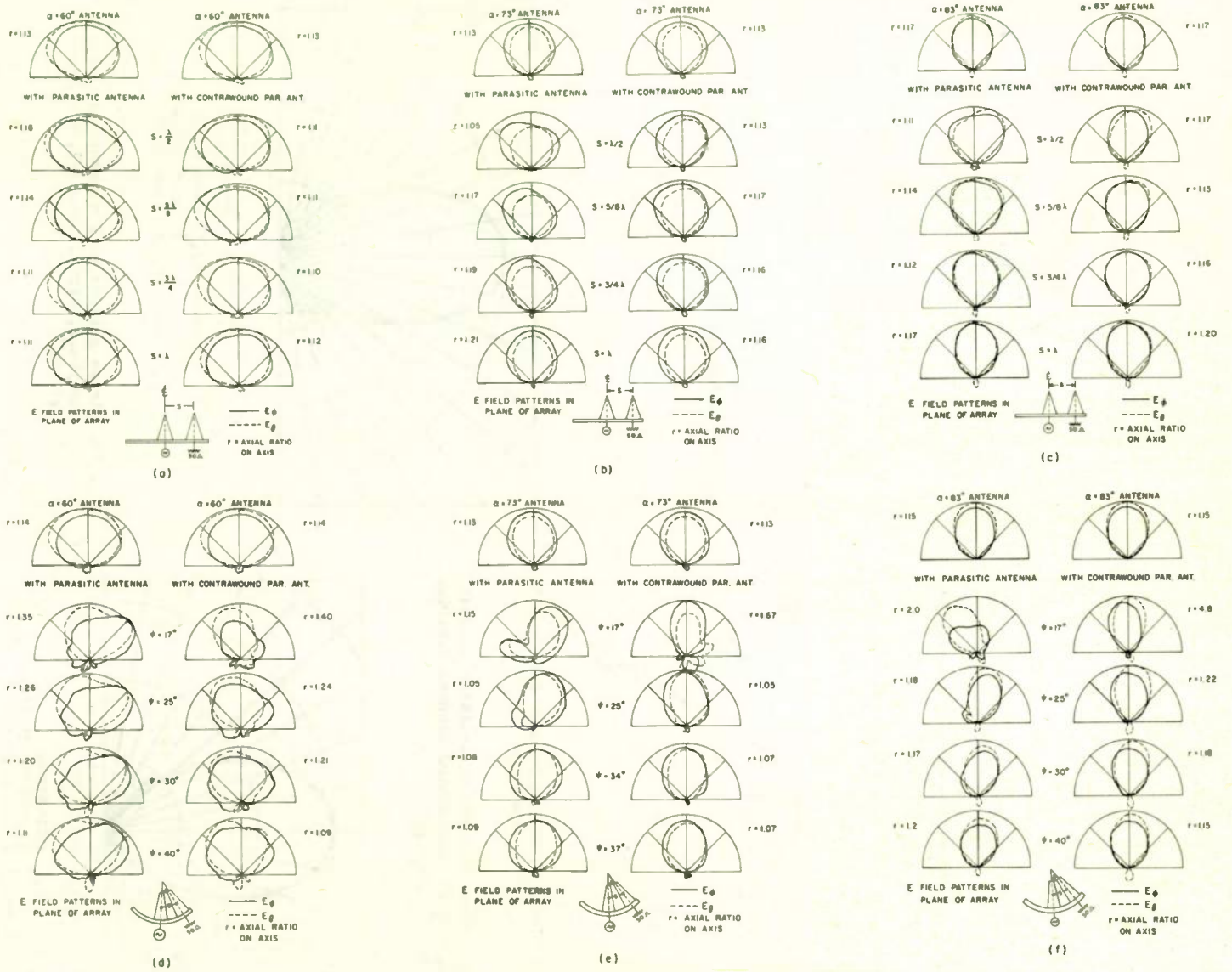


Fig. 13. Electric field radiation patterns of conical log-spiral antennas with and without a parasitic element present. All patterns taken in plane of array. $2\theta_o = 15^o$

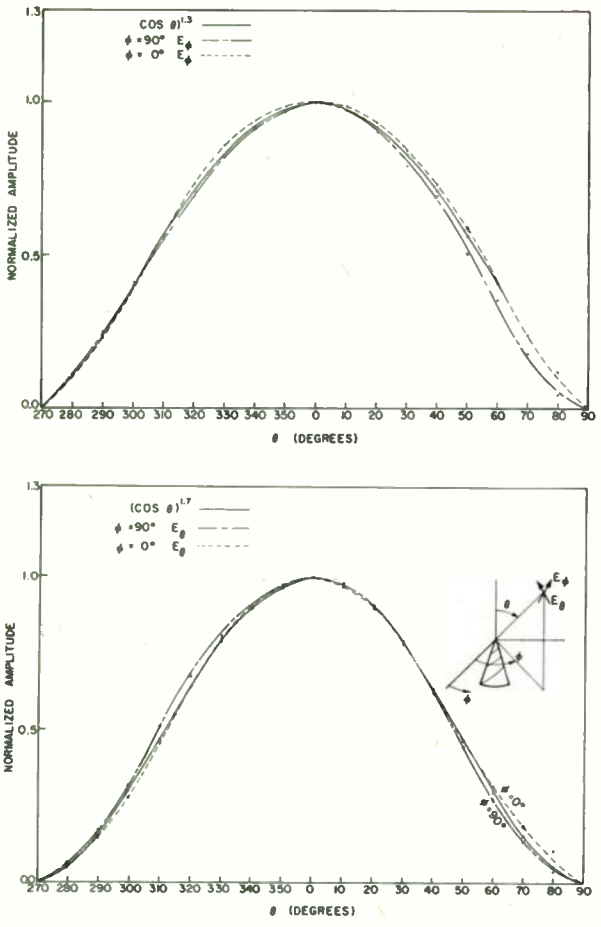


Fig. 14. Radiation patterns of isolated conical log-spiral and assumed element function. $\alpha = 73^{o}$, $2\theta_{o} = 15^{o}$

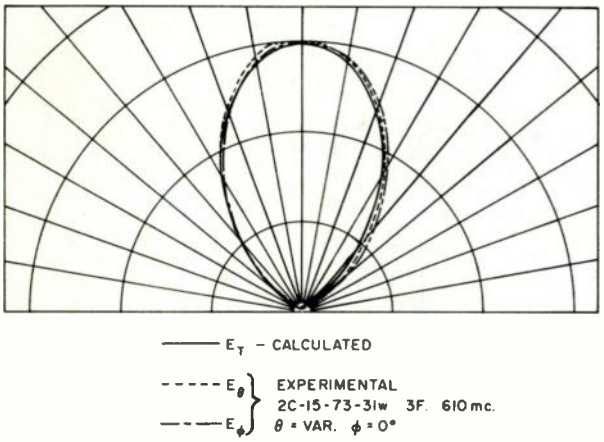


Fig. 15. Array pattern of 2 element parallel, in-phase, array. $S = \lambda/2$, $\alpha = 73^{o}$

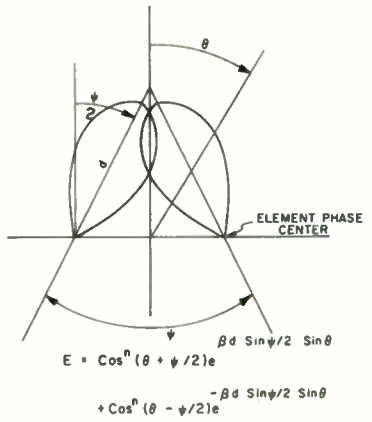


Fig. 16. Geometry of 2 element conical array

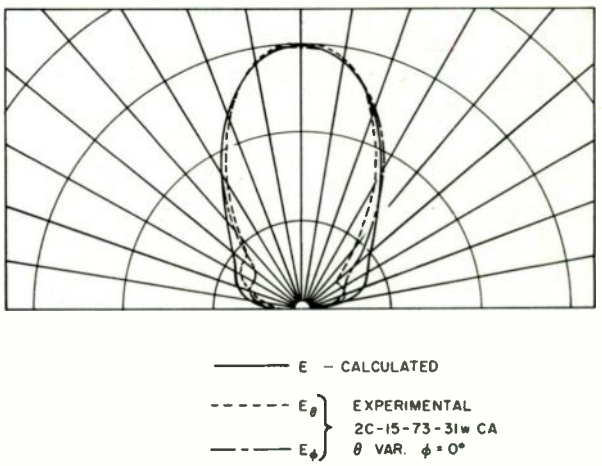


Fig. 17. Array pattern of 2 element conical, in-phase, array. $S = \lambda/2$, $\Psi = 34$, $\alpha = 73^{o}$

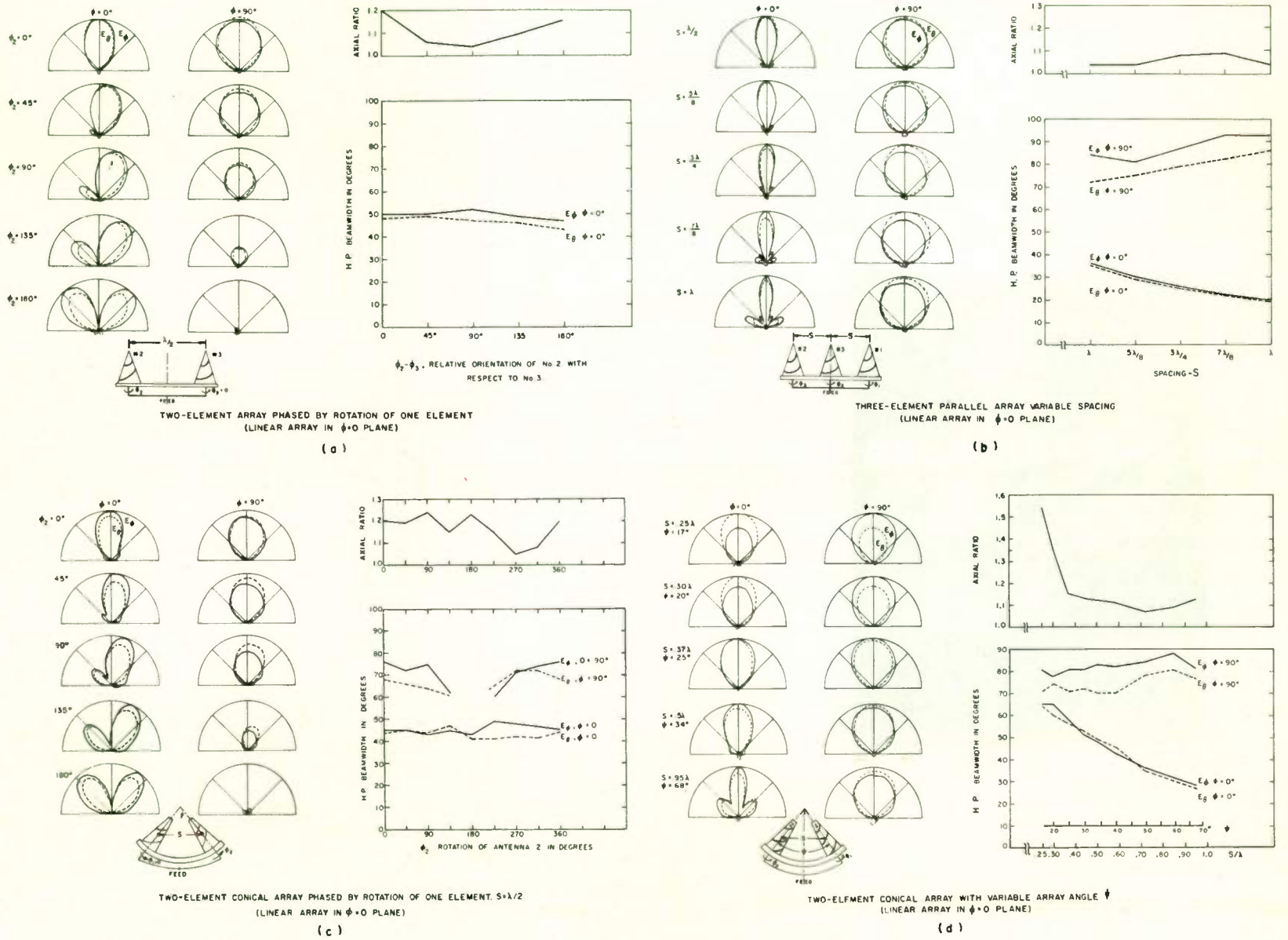
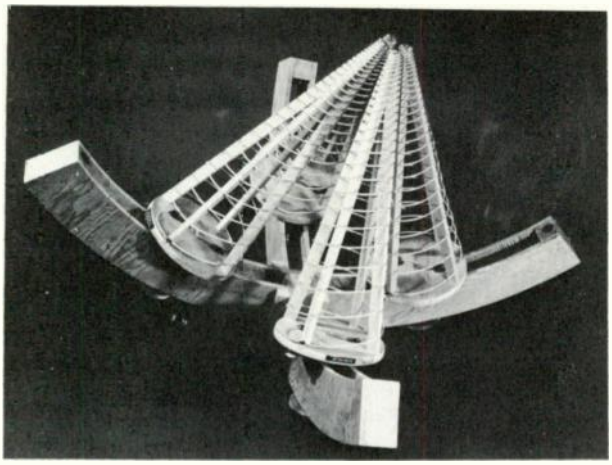


Fig. 18. Electric field patterns of several arrays. Axial ratio recorded on axis of array. $\alpha = 73^{\circ}$, $2\theta_o = 15^{\circ}$



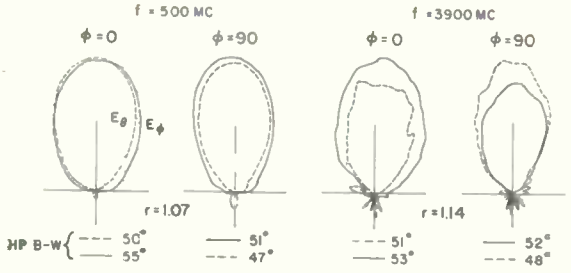


Fig. 19. A "Conical Quad-Spiral Array" with representative E field radiation patterns, r = axial ratio on axis $2\theta_o = 20^o$, $\alpha = 80^o$, $\psi = 50^o$

$$\langle [E(\psi_1) - \langle E(\psi_1) \rangle] [E(\psi_2) - \langle E(\psi_2) \rangle] \rangle$$

$$= 4N \left\{ \langle [A(X_m)\cos X_m \psi_k] [A(X_m) \cos X_m \psi_2] \rangle \right.$$

$$\left. - \langle A(X_m) \cos X_m \psi_1 \rangle \langle A(X_m) \cos X_m \psi_2 \rangle \right\}$$

$$= 4N \left\{ \frac{1}{2N_o} \sum_{k=1}^{N_o} A_k^2 \, [\cos k(\psi_1+\psi_2) + \cos k(\psi_1-\psi_2)] \right.$$

$$\left. - \frac{1}{4N_o^2} \, [E_o(\psi_1) - 1] \, [E_o(\psi_2) - 1] \right\} \qquad (13)$$

so that the correlation coefficient is found to be

$$\rho_a = \frac{N}{N_o \sigma_a(\psi_1) \sigma_a(\psi_2)} \left\{ S(\psi_2 + \psi_1) + S(\psi_2 - \psi_1) \right.$$

$$\left. - 2 - \frac{1}{N_o} [S(\psi_1) - 1] [S(\psi_2) - 1] \right\} \qquad (18)$$

(b) Cosine-Squared Amplitude Distribution:

$A_k = \cos^2 \left(\frac{k\pi}{3N_o}\right)$.

This corresponds to a 4 to 1 amplitude taper.

$$\rho = \frac{N}{N_o \, \sigma(\psi_1) \, \sigma(\psi_2)} \left\{ 2 \sum_{k=1}^{N_o} A_k^2 [\cos k(\psi_1 + \psi_2) + \cos k(\psi_1 - \psi_2)] \right.$$

$$\left. - \frac{1}{N_o} [E_o(\psi_1 - 1] [E_o(\psi_2) - 1] \right\} \qquad (14)$$

III. Specialization to Uniform and Cosine-Squared Amplitude Distributions

The expressions in the preceding section will now be applied to two special cases of practical interest: the uniform and cosine-squared amplitude distributions. Closed-form expressions for μ, σ^2, and ρ are obtained for both of these two cases.

(a) Uniform Amplitude Distribution: $A_k = 1$.

$$E_{oa} = 1 + 2 \sum_{k=1}^{N_o} \cos k\psi$$

$$= \sin \left(\frac{2N_o + 1}{2} \psi \right) / \sin(\psi/2) = S(\psi) \qquad (15)$$

$$\mu_a = \left(1 - \frac{N}{N_o}\right) S(\psi) + \frac{N}{N_o} \qquad (16)$$

$$\sigma_a^2 = \frac{N}{N_o} \left\{ 2N_o + [S(2\psi) - 1] - \frac{1}{N_o} [S(\psi)-1]^2 \right\} \qquad (17)$$

$$E_{ob} = 1 + 2 \sum_{k=1}^{N_o} \cos^2\left(\frac{k\pi}{3N_o}\right) \cos k\psi$$

$$= \frac{1}{4} \left\{ 2S(\psi) + S(\psi + 2\pi/3N_o) + S(\psi - 2\pi/3N_o) \right\} \qquad (19)$$

$$\mu_b = \left(1 - \frac{N}{N_o}\right) E_{ob}(\psi) + \frac{N}{N_o} \qquad (20)$$

$$\sigma_b^2 = \frac{N}{N_o} \left\{ \frac{3}{4} N_o \right.$$

$$+ \frac{1}{2}[S(2\pi/3N_o) + \frac{1}{4} S(4\pi/3N_o) + \frac{3}{4} S(2\psi)$$

$$+ \frac{1}{2} S(2\psi + 2\pi/3N_o) + \frac{1}{2} S(2\psi - 2\pi/3N_o)$$

$$+ \frac{1}{8} S(2\psi+4\pi/3N_o) + \frac{1}{8} S(2\psi-4\pi/3N_o) - \frac{13}{4}]$$

$$\left. - \frac{1}{N_o} [E_{ob}(\psi) - 1]^2 \right\} \qquad (21)$$

$$\rho_b = \frac{N}{N_o \sigma_b(\psi_1)\sigma_b(\psi_2)}$$

$$\left\{ \frac{1}{4} \left[\frac{3}{2} S(\psi_1+\psi_2) + \frac{3}{2} S(\psi_1-\psi_2) \right.\right.$$

$$+ S(\psi_1+\psi_2+2\pi/3N_o) + S(\psi_1-\psi_2+2\pi/3N_o)$$

$$+ S(\psi_1+\psi_2-2\pi/3N_o) + S(\psi_1-\psi_2-2\pi/3N_o)$$

$$+ \frac{1}{4} S(\psi_1+\psi_2+4\pi/3N_o) + \frac{1}{4} S(\psi_1-\psi_2+4\pi/3N_o)$$

$$\left. + \frac{1}{4} S(\psi_1+\psi_2-4\pi/3N_o) + \frac{1}{4} S(\psi_1-\psi_2-4\pi/3N_o) - 8 \right]$$

$$\left. - \frac{1}{N_o} [E_{ob}(\psi_1) - 1] [E_{ob}(\psi_2) - 1] \right\} \qquad (22)$$

IV. Limiting Bounds of Radiation Pattern

Consider now an array of 201 elements ($N_o = 100$). In order to be specific, several restrictions will be imposed on the relative values of the progressive phase shift α and the element spacing d:

$$\alpha \geq -2\pi d/\lambda \qquad (23)$$

$$\alpha \geq (2\pi d/\lambda) - 2\pi + \psi_o \qquad (24)$$

$$\alpha \leq (2\pi d/\lambda) - \pi \qquad (25)$$

Condition (23) is to require that the main beam appears at $\psi = 0$. The results to follow will therefore not pertain to arrays phased for increased directive gain in the endfire direction[4,5]. Condition (24) is to insure that grating lobes will not appear in the visible range. ψ_o may be taken roughly as the location of the first null. For $(2\pi d/\lambda) = \pi$, conditions (23) and (24) are very nearly the same for large N_o; but the latter is more restrictive as $(2\pi d/\lambda)$ becomes greater than π. Condition (25) is imposed to require that the lobes at $\psi = \pm\pi$ remain in the visible range. It is to be noted that conditions (23) - (25) do not represent a weakness in the general development. The validity of the expressions in the preceding section do not depend on these restrictions.

It is interesting to plot the "normalized variances"

$$\sigma_{na}^2 = \sigma_a^2/N \qquad (26)$$

$$\sigma_{nb}^2 = \sigma_b^2/N \qquad (27)$$

for the uniform and cosine-squared amplitude distributions. Normalization renders σ_{na}^2 and σ_{nb}^2 independent of N, the number of deleted pairs of elements. These curves are plotted in Fig. 2.

Examination of Fig. 2 reveals that the normalized varianges for both cases assume essentially constant values (2.0 for the uniform case and 1.1 for the cosine-squared case) except when ψ is close to 0^o and 180^o. This is consistent with the consideration of adding random phasors. The constancy of σ_{na}^2 and σ_{nb}^2 in the side-lobe region results from the fact that the phase of the random phasors spreads over a 2π range, largely independent of ψ. Near $\psi = 0$, the phase distribution clutters around small values, and normalized variances tend to increase with ψ. In the vicinity of $\psi = \pi$, the phase approaches the two alternate values 0 and π, causing a rapid increase in normalized variances.

With the variances known, limiting bounds of the radiation patterns can be computed. The limiting bounds of $E(\psi)$ are taken as $< E(\psi) > \pm 3\sigma(\psi)$. For a normally distributed $\Delta E(\psi)$, the probability that $E(\psi)$ falls within these bounds is 0.9987. The computed results for the uniform amplitude and the cosine-squared amplitude cases are plotted for N = 25 in Figs. 3 and 4 respectively. In each case the pattern E_o for the full 201-element array is also shown. These curves are useful in that they set the limits for what will happen to the radiation pattern. For example, Fig. 3 shows that for an array with uniform amplitude distribution the first (highest) side-lobe level will not be higher than -9.7 db when 25 pairs (25%) of elements are removed at random from a 201-element array (as compared with -13.2 db.) The bounds in the intermediate range of ψ are particularly easy to interpret because they do not vary there.

Bounding curves for $E(\psi)$ with other probability limits can, of course, be similarly computed and plotted.

V. Effects on Side-Lobe Levels and Beamwidth

The computations leading to the bounding curves for E in Figs. 3 and 4 did not require the consideration of the correlation, or mutual effect, of E values at different values of ψ. However, when one considers such questions as the probability that the half-power beamwidth is not increased by more than a certain percentage, the probability that the first side lobe does not deteriorate by a specified amount, or the probability that all side lobes are below a specified level, joint events are involved, and one must examine the correlation coefficients between these events.

The uniform amplitude distribution is a

special case in this respect because the variance for E_a, σ_a^2, is zero at $\psi = 0$. As a consequence, $E_a(0) = \langle E_a(0) \rangle$ is not a random variable and joint probabilities are not involved in the calculations for half-power beamwidth and first side-lobe level for the uniform amplitude case. In any event, both ρ_a and ρ_b have been calculated for values of ψ corresponding to side-lobe locations. Their absolute values are always less than 0.06, and the probability of joint events regarding the side loves is approximately equal to the product of the probabilities of the individual events. For the cosine-squared amplitude distribution, $\sigma_b^2(0) \neq 0$ and $E_b(0)$ is itself a random variable. While it is always possible to make conservative calculations by letting $E_b(0)$ assume its minimum value $\langle E_b(0) \rangle - 3\sigma$, intuitive considerations* of symmetry properties seem to justify the use of $\langle E_b(0) \rangle$. This approximation makes possible the computation of results which otherwise would be unobtainable due to inherent difficulties in evaluating double integrals involving conditional probabilities. Due to the smallness of ρ_b, it is also felt that the probability that all side lobes are below a specified level would be very nearly equal to the product of the probabilities that each side lobe is below that level.

There now remains the question of where the new locations of the side lobes will be after N pairs of elements have been removed. Note from Eq. (9) that $\langle E \rangle$ and E_o have the same lobe locations. Let ψ_{sk}^o denote the location of the kth side love of the original, undisturbed array, and ψ_{sk} that of the array after N pairs of elements have been removed at random. The probability that the kth side lobe is less than, say, -L db. can be written approximately as

$$P\left\{\text{kth S.L. level} < -L\right\} \doteq P\left\{\frac{|E(\psi_{sk})|}{\langle E(0) \rangle} < 10^{-L/20}\right\} \tag{28}$$

in view of the discussion in the preceding paragraph. The right-hand side of Eq. (28) can be rewritten as

$$P\left\{-B < E(\psi_{sk}) < B\right\} = \Phi(z_+) - \Phi(z_-) \tag{29}$$

where B stands for $\langle E(0) \rangle 10^{-L/20}$, $\Phi(z)$ is the standardized Gaussian cumulative distribution function

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-y^2/2}\, dy \tag{30}$$

*Quantitative results obtained by a procedure involving numerical integrations for a particular case bore out the validity of this approximation.

and

$$z_+ = [B - \langle E(\psi_{sk}) \rangle]/\sigma(\psi_{sk}) \tag{31}$$

$$z_- = [-B - \langle E(\psi_{sk}) \rangle]/\sigma(\psi_{sk}) . \tag{32}$$

If the original side-lobe location ψ_{sk}^o, instead of ψ_{sk}, is used, one would have

$$P\left\{-B < E(\psi_{sk}^o) < B\right\} = \Phi(z_+^o) - \Phi(z_-^o) \tag{33}$$

with

$$z_+^o = [B - \langle E(\psi_{sk}^o) \rangle] / \sigma(\psi_{sk}^o) \tag{34}$$

and

$$z_-^o = [-B - \langle E(\psi_{sk}^o) \rangle] / \sigma(\psi_{sk}^o) \tag{35}$$

The task at hand is then to compare (33) with (29). This amounts to a comparison of z_+^o with z_+, and z_-^o with z_-.

In the intermediate range of ψ, the curves in Fig. 2 show that the variance σ^2 is essentially constant. This constancy of variance there can be readily justified for any arbitrary amplitude distribution, as discussed in the previous section. Because $\langle E \rangle$ and E_o have the same lobe locations ψ_{sk}^o, one concludes that

$$|\langle E(\psi_{sk}) \rangle| < |\langle E(\psi_{sk}^o) \rangle| \tag{36}$$

Relation (36), together with the constancy of σ, leads to the following inequalities:

$$z_+^o < z_+ \tag{37}$$

when $\langle E(\psi_{sk}^o) \rangle$ is positive, and

$$z_-^o > z_- \tag{38}$$

when $\langle E(\psi_{sk}^o) \rangle$ is negative. For the values of L considered and $\langle E(\psi_{sk}^o) \rangle$ negative, the following relation holds:

$$3 < z_+ < z_+^o \tag{39}$$

which makes the first term in Eq. (29), and in Eq. (33) practically unity. When $\langle E(\psi_{sk}^o) \rangle$ is positive

$$z_-^o < z_- < -3 \qquad (40)$$

which makes the second term in Eq. (29) and in Eq. (33) practically zero. Hence,

$$P\left\{-B < E(\psi_{sk}^o) < B\right\} < P\left\{-B < E(\psi_{sk}) < B\right\} \qquad (41)$$

The probability that the kth side lobe is below a specified level calculated with the assumption that element removal did not shift the lobe location is therefore less than the actual value, and consequently is on the conservative side.

For cosine-squared amplitude distribution, Fig. 4 shows that the first side lobe occurs at $\psi_{s1} = 3^o$ which is already in the range of constant σ_b. For uniform amplitude distribution, the first few side lobes occur somewhat before the range of ψ where σ_a can be considered constant. Detailed calculations at these side lobes have been made. It was found that in the neighborhood of ψ_{sk}^o the change in $< E >$ overcompensates the corresponding change in σ_a and that inequality (41) still holds.

Figure 2 shows rather violent changes in variance in the neighborhood of $\psi = \pi$. However, when symmetrical pairs of radiators are removed from an array with an odd number of elements, $\psi = \pi$ would remain to be a side-lobe location, and (41) becomes an equality.

The probabilities that the first side lobe lies below -L db when N pairs of elements are removed at random from a 201-element array have been calculated for both the uniform and the cosine-squared distributions, and are plotted versus N in Fig. 5 with L as a parameter. The first side-lobe levels for the original array are -13.2 db and -22.5 db for the uniform and the cosine-squared distributions respectively. It is clear that, as N decreases toward zero, the probabilities approach 1 when L < 13.2 or 22.5 as the case may be, and approach 0 when L > 13.2 or 22.5. Apart from the difference in L values, the curves for these two cases are quite similar. Much interesting information can be derived from Fig. 5. For instance, by removing 25 pairs of elements from a 201-element array at random, there is a probability of 31% that the first side lobe will be equal to or lower than -26 db for the cosine-squared case, an improvement of 3.5 db from that of the original array!

In Fig. 6 are shown curves for the probabilities that all side lobes are below specified levels. The consideration of the other side lobes has significantly different effects on the probability curves for the uniform and cosine-squared cases. For the cosine-squared case where there is a good deal of side-lobe suppression (-22.5 db for the undisturbed array), all lobes are affected significantly by the removal of radiator pairs. This accounts for the rapid drop in probability curves as N increases. For the uniform case, the first side lobe is predominant; the other side lobes are not as significantly affected by the removal of radiator pairs. In fact, comparison of the dashed curves in Figs. 5 and 6 shows that changes occur only when N exceeds 15. Examination of the dashed curves in Fig. 6 also reveals the existence of broad maxima in the neighborhood of N = 25 for probability curves indicating improvement in overall side-lobe structure. These results are in general agreement with what is known about the side-lobe behavior of arrays with uniform and tapered amplitude distributions[6,7]. Tapering yields suppressed side lobes which are more susceptible to change than the one high side lobe in the uniform case, when either elements are removed or excitation coefficients are varied.

Probabilities have also been calculated for changes in half-power beamwidth for both the uniform and the cosine-squared cases. Let P_x represent the probability that the half-power beam angle ψ_{bw} is not increased by x per cent. The results in Table 1 have been obtained. It should

Table 1

$N_o = 100$, $N = 25$

P_x \ A_k	$A_k = 1$	$A_k = \cos^2(k\pi/2N_o)$
5%	0.91	0.80
10%	0.99	0.94

be noted that ψ_{bw} does not specify a spatial angle. Since $\psi = (2\pi d/\lambda)\cos\theta + \alpha$, the actual θ_{bw}^* corresponding to a ψ_{bw} depends on both (d/λ) and α. Table 2 will be helpful in interpreting the results in terms of the spatial angle θ_{bw} for $d = \lambda/4$.

---

*θ_{bw} is the spatial angle measured from the direction of maximum radiation to the half-power point.

Table 2

$N_o = 100$, $N = 25$

		Original θ_{bw}	5% in ψ_{bw}	10% in ψ_{bw}
A_{k_1}	Broadside, $\alpha = 0$	0.51^o	5.9% in θ_{bw}	11.8% in θ_{bw}
	Endfire, $\alpha = -\pi/2$	7.6^o	2.6% in θ_{bw}	5.2% in θ_{bw}
A_{k_2}	Broadside, $\alpha = 0$	0.64^o	4.7% in θ_{bw}	9.4% in θ_{bw}
	Endfire, $\alpha = -\pi/2$	8.6^o	2.3% in θ_{bw}	4.6% in θ_{bw}

$A_{k_1} = 1$; $A_{k_2} = \cos^2(k\pi/3N_o)$

VI. Conclusions

The effects on the radiation pattern when N pairs of symmetrically located radiators are removed from a large $(2N_o + 1)$-element linear array have been studied as a statistical problem. Under the condition $1 \ll N \ll N_o$, the probability distribution for the change in field intensity, ΔE, is found to be approximately Gaussian by invoking the Central Limit Theorem, and expressions for the expected value and the variance of the field intensity are found as functions of ψ. Two specific amplitude distributions, described by $A_k = 1$ and $A_k = \cos^2(k\pi/3N_o)$, have been considered.

It has been possible to determine the limiting bounds of the radiation pattern, the probability that a certain side lobe does not deteriorate by a specified amount, the probability that all side lobes are below a specified level, and the probability that the main-lobe half-power beamwidth is not widened by more than a given percentage. Calculations have been made for the two amplitude distributions for a 201-element array, and the results have been plotted. The curves presented herein should be useful to serve as guides in estimating the probabilities of success, or of failure, when pairs of radiators are removed from a large, symmetrical, linear array.

References

1. H. Unz, "Linear Arrays with Arbitrarily Distributed Elements," IRE Trans. on Antennas and Propagation, vol. AP-8, pp. 222-223; March, 1960.

2. H. Unz, "Multi-Dimensional Lattice Arrays with Arbitrarily Distributed Elements," University of California (Berkeley) Report No. 57 on Contract N7-ONR-29529 (Series No. 60, Issue No. 172); December 19, 1956.

3. D. D. King, R. F. Packard, and R. K. Thomas, "Unequally Spaced, Broad-Band Antenna Arrays," IRE Transactions on Antennas and Propagation, vol. AP-8, pp. 380-385; July, 1960.

4. T. M. Maher and D. K. Cheng, "Optimum Progressive Phase Shifts for Discrete Endfire Arrays," Syracuse University Research Institute Report No. EE492-6002T8, Contract No. AF 30(602)-1640, RADC-TN-60-44; February, 1960.

5. W. W. Hansen and J. R. Woodyard, "A New Principle in Directional Antenna Design," Proc. IRE, vol. 26, pp. 333-345; March, 1938.

6. M. F. Ma and D. K. Cheng, "A Critical Study of Linear Arrays with Equal Side Lobes," 1961 IRE International Convention Record, vol. 9, part 1, pp. 110-122; March, 1961.

7. J. Ruze, "The Effect of Aperture Errors on the Antenna Radiation Pattern," Nuovo Cimento, vol. 9, Suppl. No. 3, pp. 364-380; 1952.

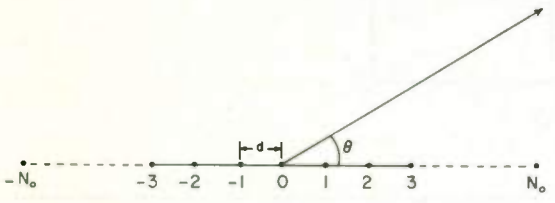


Fig. 1. Linear array of $2N_o+1$ omnidirectinal radiators

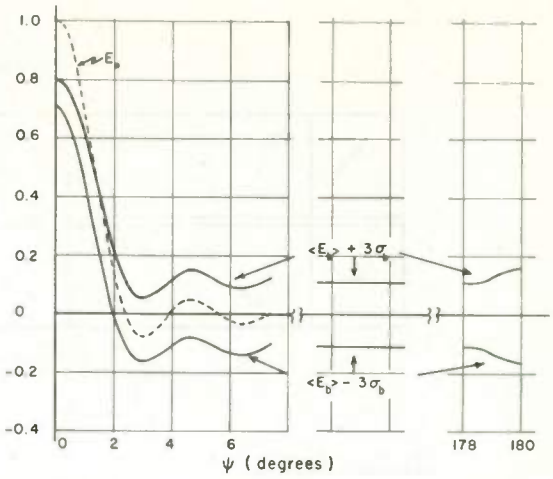


Fig. 4 – Radiation pattern and $\langle E_b \rangle \pm 3\sigma_b$ bounds: $A_k = \cos^2(\frac{k\pi}{3N_o})$

$N_o = 100$, $N = 25$

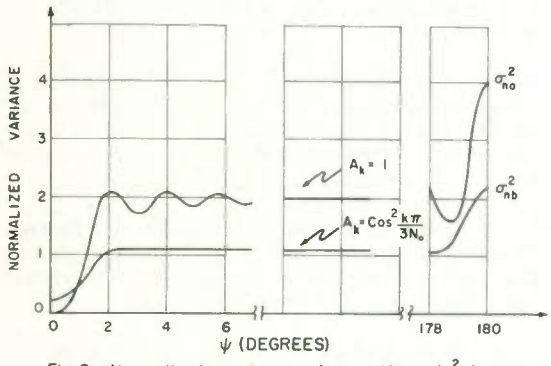


Fig. 2 – Normalized variances for uniform (σ^2_{no}) and cosine-squared (σ^2_{nb}) amplitude distributions

$N_o = 100$

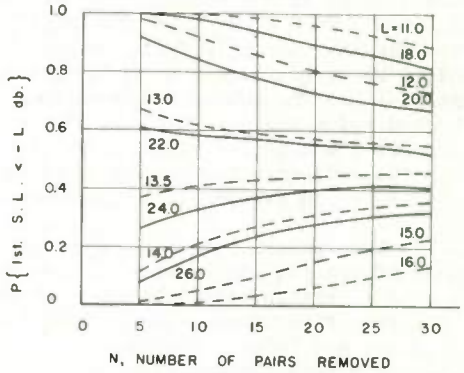


FIG. 5 PROBABILITY DISTRIBUTION FOR FIRST SIDELOBE LEVEL

——— $A_k = \cos^2(\frac{k\pi}{3N_o})$

– – – – – $A_k = 1$

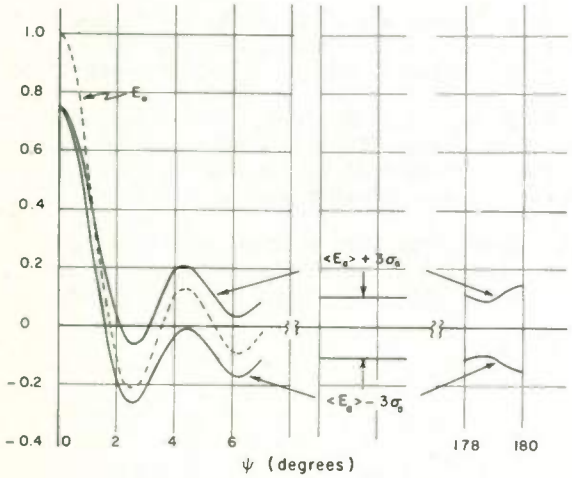


Fig. 3 – Radiation pattern and $\langle E_a \rangle \mp 3\sigma_a$ bounds: $A_k = 1$

$N_o = 100$, $N = 25$

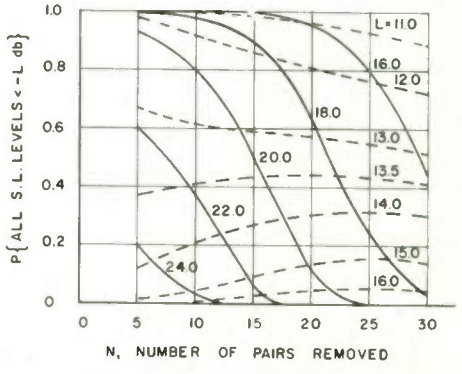


FIG. 6 – PROBABILITY DISTRIBUTION FOR SIDELOBE LEVELS

——— $A_k = \cos^2 \frac{k\pi}{3N_o}$

– – – – $A_k = 1$

A SPACING WEIGHTED ANTENNA ARRAY

Y. T. Lo
Department of Electrical Engineering
University of Illinois
Urbana, Illinois

Summary

This paper considers a non-uniformaly spaced array. By starting with a continuous aperture distribution, the pattern function is formulated from a Lesbegue-Stieltjes integral point of view. A spacing weighting function can thus be generated. Based upon the methods of mechanical quadrature, the integral is reduced to a summation which represents the pattern function of a corresponding array.

As an example, a symmetric linear array of 74 isotropic sources at an average spacing of 1.82 wavelengths is considered. The side lobe level is found to be less than -14 db for $|u| < 0.98$ and below -10 db for $|u| \leq 2$. If these elements were uniformly spaced, secondary beams of 0-db level would occur at $u = \pm 0.552, \pm 1.10$, and ± 1.66.

A systematic optimization in a certain sense by a high speed computer has also been carried out. It indicates that a slight improvement in performance is possible.

Based upon Legendre-Gaussian quadrature, two other linear arrays, both spacing and amplitude weighted, are studied. For relatively small u, their patterns are practically identical to that due to the corresponding continuous distribution.

Introduction

The general method used in designing an antenna array usually follows a uniformly spaced arrangement. However it may be emphasized that this method is originated perhaps from a mathematical convenience rather than some other considerations such as cost, etc. For a small array it may not be a serious problem in considering the over-all cost of a system, but it seems to be in quite a different situation for many extremely large arrays nowadays used for radio astronomy research and high resolution long range radar system. Therefore, it is somewhat doubtful that the application of conventional methods with uniform spacing is justified in these cases. Unfortunately very little is known about the non-uniformly spaced array except a few investigations made in the last few years.[1]

During the fall of 1959, this author made an investigation on a non-uniformly spaced array which was proposed to serve as a feed for the University of Illinois radio telescope. Even in such a case with only a few hundred elements, it has indicated the possibility of reduction in total number of elements by 33% without much sacrifice in the antenna performance. However, in confronting the difficulty in justifying the proposed method of generating the spacing weighting function to be optimum and also general (in fact both are probably not true) we have devised a numerical optimization procedure (in a certain narrow sense) by using a high speed computer. These methods and the results pertaining to the telescope have been reported previously[2] and subsequently in a paper.[2] Recently it is very fortunate to have learned that Dr. Maffet[3] independently proposed the same method in generating the spacing weighting function and also Dr. Andreasen[4] seemed to have used a similar technique in numerical optimization. Although as far as this author knows nothing has been achieved in proving that this spacing weighting function is optimum, (in the sense that the peak side-lobe level is minimized) nor has an algorithm been found for the optimization in a very general sense, yet their findings seem to have added more confidence in these perhaps temporarily content techniques and also prompt this author to present some previously unreported results.

Theory

Uniformly Spaced Array

For simplicity we restrict our discussion to a linear array. The starting point of this approach is to regard a uniformly spaced array as a numerical integration approximation to an integral for its corresponding continuous distribution by the so-called trapezoidal rules. Let the desired radiation pattern be

$$P_o(u) = \int_{-\infty}^{\infty} f_o(x) e^{j2\pi xu}\, dx \qquad (1)$$

where

$u = \sin\theta - \sin\theta_o$,

θ_o = the scan angle of the beam with respect to the normal to the array,

θ = the observation angle also referred to the same normal,

$f_o(x)$ = the magnitude of the aperture distribution normalized such that $P_o(0) = 1$, and it equals identically to zero for $|x| > a/2$

a = the total aperture dimension in λ.

Now it is to be approximated by a discrete array, namely one with the following aperture distribution

$$f(x) = \frac{a}{2N+1} f_o(x) \sum_{m=-N}^{N} \delta(x-x_m) \qquad (2)$$

where

$2N+1$ = the total number of elements. The assumption of odd numbers here is immaterial to the theory,

δ = the Dirac delta function,

x_m = the location of the m-th element.

Since $f_o(x) = 0$ for $|x| > a/2$, the above equation can be rewritten as follows

$$f(x) = \frac{a}{2N+1} f_o(x) \sum_{m=-\infty}^{\infty} \delta(x-x_m) \qquad (3)$$

For a uniformly spaced array with a common spacing $d = a/2N$

$$x_m = m\, a/2N \qquad (4)$$

Substituting the above into (3), and using the convolution theorem in Fourier transforms, one immediately obtains

$$P(u) = P_o(u) * \sum_{m=-\infty}^{\infty} \delta(u-m\,\frac{2N}{a})$$

$$= \sum_{m=-\infty}^{\infty} P_o(u-m\,\frac{2N}{a}) = \mathrm{Rep}_{2N/a}\, P_o(u) \qquad (5)$$

where * denotes the convolution operator and $\mathrm{Rep}_{2N/a}\, P_o(u)$, following Woodward[6], denotes a function which is the sum of infinitely many P_o's each displaced by $2N/a$ along the u-axis. Equation (5) indicates precisely the difference between $P(u)$ and $P_o(u)$. Since $P(u)$ is a periodic function with a period $2N/a$, the so-called grating lobes will appear at $u = m(2N/a)$, $m = 0, \pm 1, \pm 2, \ldots$. To avoid the appearance of any grating lobes other than the one with $m = 0$ in the visible range, one is forced to choose a sufficiently large N. For a two-dimensional array similar results can be obtained with P periodic in both u and v where $u = \sin\theta\cos\varphi$, $v = \sin\theta\sin\varphi$.

Spacing Weighted Array

It is seen from the above that the existence of large grating lobes is a consequence of the periodicity in element spacing as given by (4). To avoid this one has to abandon any periodicity in spacing. It is well-known that there are available various quadrature methods in numerical integration which do not require uniformly spaced sampling points. To follow our previous approach (1) may be rewritten as

$$P_o(u) = \int_{-\infty}^{\infty} e^{j2\pi xu}\, dF(x) \qquad (6)$$

where

$$F(x) = \int_{-a/2}^{x} f(y)\, dy \qquad (7)$$

This is generally called the Stieltjes integral with a weight or measure F(x). If one divides F from 0 to F(a/2) into 2N equal intervals each equal to ΔF, then the approximation to (6) becomes

$$P_o(u) \approx P(u) = K \sum_{n=-N}^{N} e^{j2\pi x_n u} \qquad (8)$$

where K is a normalizing factor such that $P(0) = P_o(0)$ and x_n satisfies

$$F(x_n) = nF(a/2)/2N \qquad (9)$$

This method of division is motivated by the Lesbegue integral, roughly in a sense that f(x) is divided into intervals such that each interval on x contributing the same amount of power to the beam maximum (for more detail see Ref. 2). This method has been applied to the design of the University of Illinois radio telescope. However, in that particular design, such a spacing weighted array has been used incorporated with a second array amplitude-weighted in order to meet the requirement of very low side-lobe level. The performance of this spacing weighted array has been previously reported only very briefly. In Figure 1, it shows the pattern for u = 0 to 2. It is 132λ long, with 37 pairs of elements symmetrically located. The difference between this array with the one previously reported is that (a) one pair of elements is added due to a frequency change for the telescope; (b) the central portion namely the amplitude-weighted array is absent and the gap so created is closed (in order to show more precisely the behavior of the spacing weighted array); (c) due to (b) the function f(x) in (7) used to generate $\{x_n\}$ is thus a truncated cosine-square function, namely

$$f(x) = 2\cos^2\frac{\pi}{2}\left(\frac{|x|}{a} + \frac{1}{2}\right), \text{ if } |x| \le a/2,$$

$$= 0, \qquad \text{otherwise;}$$

and x_n is given by the solution to

$$\frac{a}{\pi}\sin\frac{\pi}{a}(x_n + \frac{a}{2}) + (x_n + \frac{a}{2}) = \frac{2n}{3} \qquad (10)$$

The set of spacings so obtained is finally subject to a systematic optimization by a computer as discussed above (also Ref. 2) and the final results as shown in Figure 1 indicate that the peak value of the side-lobe is lowered by only about 1 db and the spacings as given by (10) are only slightly perturbed. It should be noted that in Figure 1, except for u = 0 to 0.05 only the envelope of the side-lobes is shown. Since the average spacing is about 1.82λ, if they were uniformly spaced secondary beams of 0-db level would appear at $|u| = 0.552, 1.10, 1.66$ etc. It is also interesting to note that for an array uniform both in magnitude and spacing the side-lobe level near the main beam is always -13.2 db which is extremely difficult to reduce; On the other hand that of this spacing weighted array is at about -26 db.

Legendre-Gauss Array

It is well-known that in numerical integration the Gaussian quadrature in general provides the highest accuracy or for the same accuracy it requires the least number of points. Obviously the Legendre-Gaussian quadrature becomes one of the most attractive methods.

Suppose f(x) is an even function, then (1) becomes

$$P_o(u) = \frac{1}{b}\int_0^{a/2} f(x)\cos 2\pi ux\,dx$$

$$= \frac{1}{b}\int_0^{b} \cos(2\pi x)\,dy(x) \qquad (11)$$

where

$$y(x) = \int_0^x f(x)\,dx$$

and

$$y(a/2) = b$$

Let

$$y = (z+1)b/2$$

Then

$$P_o(u) = \frac{1}{2}\int_{-1}^{1} \cos[2\pi x]dz(x) \quad \frac{1}{2}\sum_{i=1}^{N} H_i \cos 2\pi u x_i \qquad (12)$$

where

$2N$ = total number of elements,

H_i = the weighting coefficient corresponding to the i-th root z_i of the Legendre polynomial of N degree ,

x_i is a solution to

$$\int_0^{x_i} f(x)\,dx = b(z_i+1)/2, \qquad (13)$$

and

$$x_{-i} = -x_i.$$

Due to the fact that if z_i is a root of a Legendre polynomial so is $-z_i$, one can formulate (1) in another manner, instead of (11), as follows

$$P_o(u) = \frac{1}{b}\int_{-a/2}^{a/2} f(x)e^{j2\pi ux}dx = \frac{1}{b}\int_0^{b} e^{j2\pi ux}dy(x) \qquad (14)$$

where

$$y(x) = \int_{-a/2}^{x} f(x)\,dx,$$

Now making the same substitution as in (12), one obtains

$$P_o(u) = \frac{1}{2}\int_{-1}^{1} e^{j2\pi ux}dz(x) \approx \frac{1}{2}\sum_{i=-N}^{N} H_i\, e^{j2\pi ux_i}$$

$$= \sum_{i=1}^{N} H_i \cos(2\pi ux_i) \qquad (15)$$

where x_i is a solution to

$$\int_{-a/2}^{x_i} f(x)\,dx = b(z_i+1)/2$$

and z_i = i-th root of the Legendre polynomial of degree 2N instead of N as given by (12), (13).

There are formulas for the upper bounds of the errors in these quadrature methods; unfortunately, they are not very useful in the present application unless u is very small. It is due to the fact that first the exponential or trigonometrical function in the integrand of (6), (11), and (15) must be retained in the summation in order that each term can be physically interpreted as an isotropic source. Secondly these upper bounds, being proportional to the 2Nth or 4Nth derivative, could become extremely large valued by Berstein's theorem. Nevertheless, this does not necessarily imply that the actual error involved is that large as well illustrated by an example given by Rosser.[7]

As an example to these methods as shown by (12) and (15) the following function is again assumed

$$f(x) = \cos^2 \pi x/a \qquad \text{for } |x| \le a/2$$

$$= 0 \qquad \text{otherwise.}$$

For $a = 132\lambda$, $2N = 80$, the patterns are computed and shown in Figure 2 where for $|u| = 0$ to 0.2 an expanded scale has been drawn to show the detail structure of the main beam and its neighboring lobes; while for $|u| = 0.2$ to 2 only the envelopes of the side-lobes are shown to avoid the confusion of many fast oscillating curves. It is interesting to notice that the approximate array given by (12) gives a pattern almost exactly the same as that of a continuous distribution for $|u|$ up to 0.2 as expected. The first side lobe is -32 db and the rest are even much lower. For $|u| = 0.2$ to 0.35 it is still below -30db; however, after that it begins to build up to -13 db at $|u| = 0.625$ then drops to -17 db at u = 0.75. However, it is somewhat unexpected to find that at this point the pattern breaks up so suddenly into a random behavior and remains so thereafter. The second approximate array as given by (15) gives a pattern almost identical to the one above except that the well behaved portion is extended to $|u| = 0.4$; even for $|u| = 0.5$ it is still below -30 db. Unfortunately after that it builds up at a very

fast rate until reaching about -5 db at $|u| = 0.77$. Then it drops sharply to -40 db, and behaves well again for a range until it reaches -6 db at $|u| = 1.54$. Although no calculation has been made beyond $|u| = 2$, it is reasonable to believe that it will eventually behave in a random manner as in the first approximation. On the other hand calculation has been made for a portion of $|u|$ beyond 2 for the first approximate array, and it shows a randomness very similar to that in interval of $|u|$ between 1 and 2. In fact it is probably true that in most cases of non-uniform spacings the pattern will behave in a random manner for large $|u|$. Investigation along this line of thought on a randomly spaced array has been conducted with some interesting results which will be reported elsewhere.

The spacing of these arrays runs from .06λ to 4λ. Since in Legendre-Gaussian quadrature the end points are not included in the samples, the actual aperture is about 12λ shorter; nevertheless the beam width is even slightly sharper than that of $P_o(u)$ according to calculation. It is also worth noting that the weights H_i and roots z_i are universal constants depending on the degree of the polynomial used; tables for degrees up to 96 are available.[5]

Finally it may be noted that no optimization has been applied to either of these arrays. It is believed that a substantial improvement probably can be achieved for the second approximate array since there are only a few high level peaks and the rest are at a much lower level. Also it is important to note that the amplitude weighting for these arrays has a range of 24 dbs instead of 80 dbs if they were uniformly spaced cosine-square weighted in magnitude.

Conclusions

(1) A non-uniformly spaced array has the potential to provide wide bandwidth, large scan angle, and to reduce the number of elements required. These advantages become even greater for large antennas, and higher dimension arrays.

(2) A spacing weighted array as discussed here has a predictable behavior for relatively small $|u|$, while for large $|u|$, the random behavior prevails and may be estimated in a probabilistic sense.

(3) In general non-uniformly spaced arrays particularly those presented here, have low side-lobe level for small $|u|$ and increasing side-lobe level for large $|u|$ (eventually reaching a more or less saturation level) in contrast to that of a continuous array. This may not be a serious problem in consideration of the yet unaccounted for directivity of most antenna elements.

(4) The range of amplitude weights can be cut down to a much more realistic value by using a spacing weighting.

(5) In view of (2) and (3) above if extremely low side-lobe level is required for an antenna with a moderate number of elements, a combination of an amplitude weighted array and a spacing weighted array (or one with both spacing and amplitude weighting as discussed above) is probably the most promising approach to furnish a solution because at small $|u|$ both their predictable behavior will guarantee a good resultant performance in that region while for large $|u|$ the side-lobe level is essentially that of the non-uniformly spaced array alone. For an example refer to Ref. 2.

Acknowledgment

This work is an extension of the radio astronomy project at the University of Illinois sponsored by NONR 1834(22). This author would like to acknowledge their partial support at the initial stage. He would also like to acknowledge the assistance from an NSF grant for making the presentation of this report possible.

References

(1) D.D. King, R.F. Packard, and R.K. Thomas, Trans. IRE AP-8, 380-384, July, 1960.

S.S. Sandler, Trans. IRE, AP-8, 496-500, Sept. 1960, Other relevant references are listed below.

(2) URSI Spring Meeting, May, 1960, Washington, D.C., G. W. Swenson, Jr. and Y.T. Lo, Trans. IRE, AP-9, 9-16, January 1961.

(3) URSI Spring Meeting, May, 1961, Washington, D.C. Also a paper to be published in IRE Trans. on Antennas and Propagation.

(4) A paper by M.G. Andreasen to be published in IRE Trans. on Antennas and Propagation.

(5) P. Davis and P. Rabinowitz, J. Res. N.B.S. Vol. 56, 35 (1956) and Vol. 60, 613 (1958). H. J. Gaulick, Armament Research and Development Establishment Memo. (B), 77, 1958.

(6) P.M. Woodward, Probability and Information Theory with Application to Radar, Pergamon Press Ltd., London, 1955.

(7) J. B. Rosser, Proc. Am. Math. Soc. 1, 388-389 (1950).

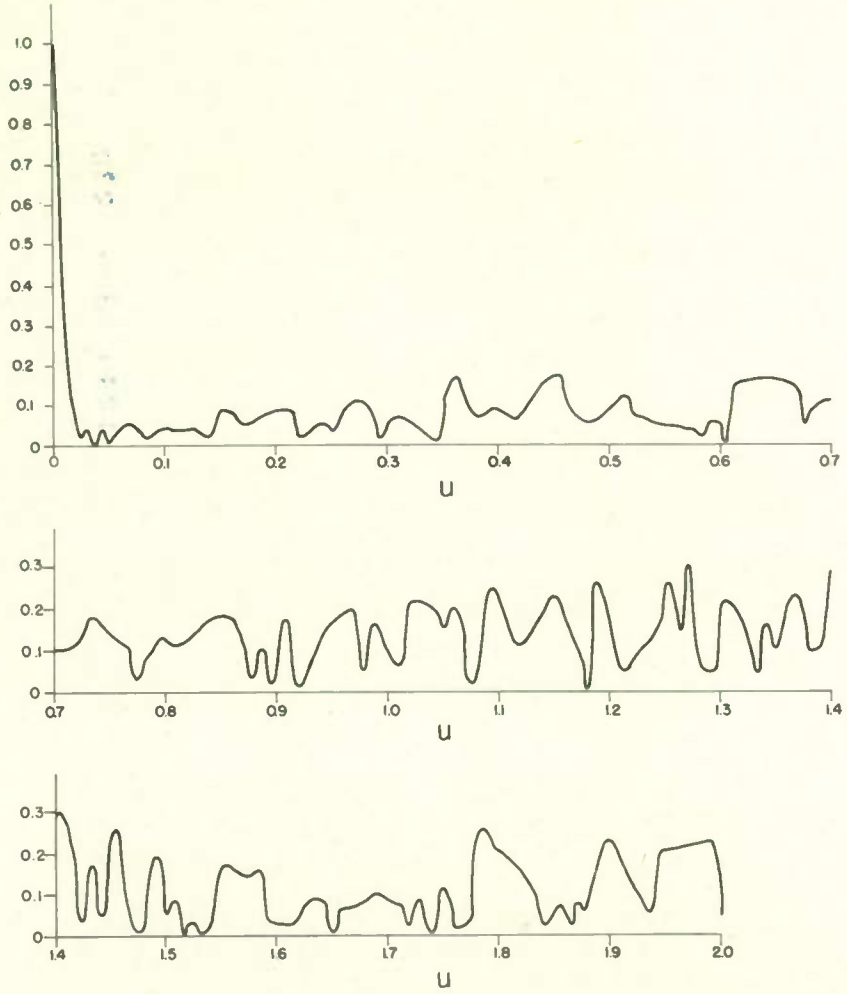


Fig. 1. The field intensity pattern for a linear spacing weighted array of 132 λ in length and 74 elements. For $u \geq 0.05$ only the envelope of the side-lobes is shown. The half-power beamwidth is approximately equal to 0.6°.

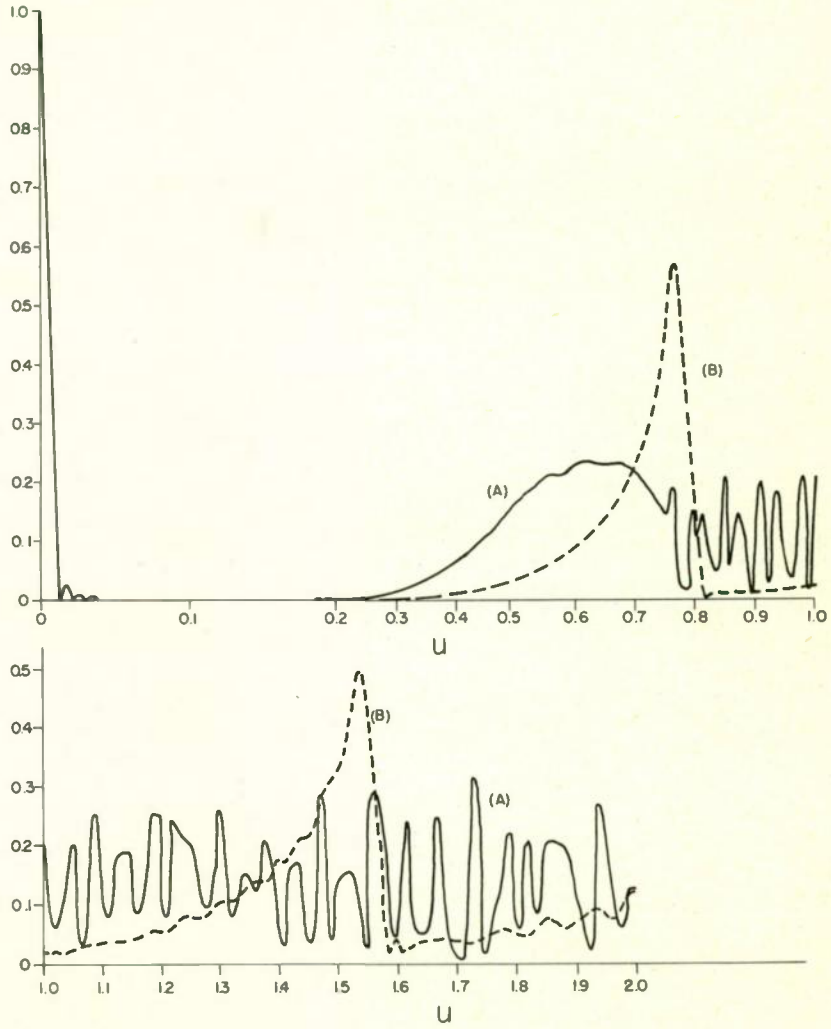


Fig. 2. The field intensity patterns for two linear arrays both amplitude and spacing weighted according to the Legendre-Gaussian quadrature. Both arrays have a nominal length of 132λ, and a total of 80 elements. For ≥ 0.2 only the envelopes of the side lobes are shown. (A) is derived from the polynomial of 40 degree, (B) from that of 80 degree. The half power beamwidth is approximately equal to 0.57°.