STAT

UNCLASSIFIED

TOPICS IN THE DESIGN OF ANTENNAS FOR SCATTER

John Granlund

Group 33

STAT

Technical Report No. 135

23 November 1956

ABSTRACT

Unlike the signal received over a line-of-sight path, scatter signals arrive at the receiving site from a continuum of directions with intensities that may be described by a directional pattern similar to an antenna pattern. This report shows how the mean signal power available at the terminals of a receiving antenna may be expressed in terms of the antenna pattern and of a pattern of incoming power density. Methods of measuring the power-density pattern at the receiving site are discussed. Designs maximizing the available power are obtained for two rather general types of antenna when they are illuminated by arbitrary power-density patterns. Numerical results, obtained for an assumed Gaussian-shaped power-sensity pattern, suggest that only a very small increase in available power may be obtained by readjusting the pattern of an antenna that was initially adjusted to have maximum plane-wave gain in the direction of maximum power density.

A large array of small antennas may be subject to "gain loss" but, if the connections between array elements are allowed to vary in time at the fading rate, the array will not exhibit "gain loss." Appropriate electronic circuitry for the interconnections is discussed.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY LEXINGTON, MASSACHUSETTS

UNCLASSIFIED

TOPICS IN THE DESIGN OF ANTENNAS FOR SCATTER

I. INTRODUCTION

Practical communication links utilizing the weak, fluctuating scatter signals received from a remote transmitter have been demonstrated.[1] Unlike signals received over a short line-of-sight path, these scatter signals arrive at the receiving site from a continuum of directions with intensities that may be described by a directional pattern similar to an antenna pattern. This spread in the directions of arrival of the signal components gives rise to a phenomenon that has misleadingly been called "gain loss": The strengths of scatter signals from a remote transmitter, received on adjacent, dissimilar antennas, are not in the same proportion as the plane-wave gains of the antennas.

As suggested by Schott,[2] the mean signal power available at the terminals of a receiving antenna is proportional to the sum (or integral) over all directions of the pattern of incoming power density weighted by the antenna pattern. In arriving at this result, Schott makes the plausible assumption that the plane-wave components of the signal arriving from different directions are uncorrelated. This assumption, which is also made here, is implicit in the concept of a pattern of power density; the assumption is discussed in greater detail in the next section.

The aim of this report is to provide a method for arriving at a "best" antenna design for the reception of scatter signals. Certainly the "best" antenna should deliver the "greatest" signal strength to its associated receiver but, since scatter signals are subject to fading, the choice of a "greatest" signal implies a criterion of measurement. Also, in general, the larger the antenna, the greater the signal it will deliver. In the sense of this report, the optimum receiving antenna is that stationary structure of a given size which delivers maximum mean signal power to a matched load.

Although the pattern of incoming power density gives necessary and sufficient signal data for the design of an optimum receiving antenna, the shape of this pattern depends on the shape of the transmitting antenna pattern as well as on the behavior of the scattering mechanism. In other words, a detailed description of the scattering mechanism itself would be necessary for the simultaneous optimization of both transmitting and receiving antennas. Such a description is beyond the scope of this report. Nevertheless, the remainder of this introductory section is given to an elementary discussion of the effects that the transmitting and receiving antenna patterns may have on the received signal power.

Following a simplified version of the Booker-Gordon model of the scattering mechanism,[3] let us suppose that the transmitting antenna emits power uniformly in a circular cone having a small angle Θ_t, as in Fig. 1. Similarly, let the receiving antenna accept power uniformly only in a cone having a small angle Θ_r. Suppose that the region above the surface of the earth is entirely uniform except for some subregion having a volume V, where scattering takes place. Suppose that each volume element, ΔV, of V scatters power isotropically, so that $K\Delta V$ watts per unit solid angle are scattered when the volume element is illuminated with unit power density

from the transmitting antenna, and let K be very small so that V intercepts negligible power, and multiple scattering may be neglected. Let the linear dimensions of V be small compared with the distances from V to the transmitting and receiving sites. Finally, let V be entirely contained within the common volume defined by the intersection of the transmitting- and receiving-antenna cones.

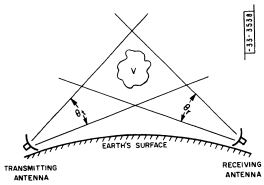

Fig. 1. The scattering region is contained in the common volume.

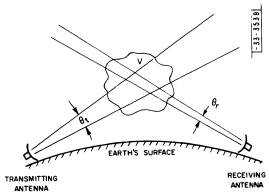

Fig. 2. If beamwidths are reduced, the common volume is contained in the scattering region.

The power gain of the transmitting antenna is inversely proportional to Θ_t^2, and the power gain of the receiving antenna is inversely proportional to Θ_r^2. The signal power available at the terminals of the receiving antenna is proportional to the product of the antenna gains, the scattering coefficient K, and the volume of the region in which scattering takes place, V. As far as the antennas are concerned,

$$(\text{received power}) \propto \frac{1}{\Theta_t^2} \frac{1}{\Theta_r^2} \quad .$$

If the angular width of the receiving antenna pattern, Θ_r, were varied, the received power would be just proportional to the gain of that antenna: In the present elementary circumstance, the "gain loss" phenomenon does not appear provided the scattering region V lies entirely within the common volume. In other words, the dependence of received power on antenna gains is the same as that for a line-of-sight path if the antennas cannot "resolve" the scattering region.

On the other hand, if the antenna beamwidths are reduced to increase the received power, a point is reached where the common volume no longer contains the scattering region, V. Let us next examine the dependence of received power on antenna beamwidths when this transition has been completed; that is, let us assume that the common volume of the antenna cones is entirely contained within the scattering region V, as in Fig. 2. Suppose that the transmitting-antenna cone is the larger, and that it is pierced through its center by the receiving-antenna cone.

The region in which scattering takes place is now just the common volume, which can be described roughly as a cylinder having a height proportional to Θ_t and a cross-sectional area

proportional to Θ_r^2. The volume of this cylinder is proportional to $\Theta_t \Theta_r^2$; the received power is proportional to it and to the product of the antenna gains:

$$(\text{received power}) \propto \Theta_t \Theta_r^2 \frac{1}{\Theta_t^2} \frac{1}{\Theta_r^2} = \frac{1}{\Theta_t} \quad .$$

No further increase in received power can be obtained by reducing the receiving-antenna beamwidth Θ_r, and the received power increases at only half the rate of the previous case when the transmitting-antenna beamwidth Θ_t is reduced. Both antennas seem to exhibit "gain loss."

Suppose that the transmitting-antenna beamwidth is reduced to increase the received power. When the two beamwidths are about the same size, the received power will cease to increase because the assumption about the shape of the common volume will no longer be valid: When the transmitting-antenna beamwidth Θ_t is the smaller, the common volume is more nearly proportional to $\Theta_t^2 \Theta_r$, and

$$(\text{received power}) \propto \frac{1}{\Theta_r} \quad .$$

But at this point, the receiving-antenna beamwidth may be reduced to increase the received power. In any case, it appears that, in general, the most economical design of a practical scatter link would call for equal transmitting- and receiving-antenna beamwidths.

Before closing this discussion of the joint effects of the transmitting- and receiving-antenna patterns on the received signal power, it is appropriate to discuss the information required for the simultaneous optimization of both antennas and to remark on the difficulties of obtaining this information by radio measurements. In general, each volume element in the scattering region can be expected to scatter power according to some pattern, generally not isotropic, and the shape of this pattern may depend on the position of the volume element in the scattering region. We may regard the magnitude of the scattering pattern in the direction of the receiving antenna, produced when one volume element is illuminated by unit power density from the transmitting antenna, as the scattering efficiency of that volume element. Thus, even if questions of polarization are somehow set aside, it would be necessary to know the scattering efficiency throughout the scattering region in order to calculate the received power when given transmitting and receiving antennas are used.

A measurement of the power density arriving at the receiving site gives only two-dimensional information about the scattering region; the power density is a function of the azimuth and elevation angles. The pattern of arriving power density is like a "photograph" of the illuminated scattering region as seen from the receiving site and is not enough information to determine the behavior of each point in the volume. The additional information that would be made available by a power-density measurement at a second receiving site does not appear to be helpful because the scattering pattern of each volume element would contribute differently to the new "photograph."

The remainder of this report will deal with the density of power arriving at a single receiving site when a fixed transmitting antenna is used. As shown in the foregoing discussion, the results will pertain only to the effectiveness of the receiving antenna in extracting power when it is illuminated by this power-density pattern.

II. CROSS-CORRELATION BETWEEN ARRIVING SIGNAL COMPONENTS

In this section, we will consider the assumption implied by the introductory discussion, namely, that the plane-wave components of the signal arriving from different directions are uncorrelated. At the receiving site, the electric field due to the arriving signal may be resolved into an angular spectrum of incoming plane waves. The antenna weights the amplitude of a plane-wave component F_i arriving from a particular direction i, in accordance with the amplitude of the antenna pattern in that direction, a_i, and the open-circuit signal voltage V at the antenna terminals is the vector sum (or integral) of the weighted plane-wave components:

$$V = \sum_i a_i F_i \quad . \tag{1}$$

The index i is associated with the direction of arrival of the i^{th} component.

The amplitudes and phases of the plane-wave components F_i, and hence the amplitude and phase of their weighted sum V, vary in time as a result of motions in the scattering region and possibly also because the transmitted signal is modulated. Temporarily, it will be convenient to assume that the transmitter is not modulated, so that attention is focused on random fluctuations caused by the scattering mechanism. Later we will show that the results also apply to modulated signals.

In this report, the notations Av.$\{\ \}$ and $\overline{(\ )}$ represent a time average, and $(\ )^*$ means that the complex conjugate is to be taken. Using them, the mean-square terminal voltage, a quantity proportional to the time-average available signal power at the receiving-antenna terminals, may be written as

$$\overline{|V|^2} = \text{Av.}\left\{\left(\sum_i a_i F_i\right)\left(\sum_k a_k F_k\right)^*\right\} = \sum_i \sum_k a_i a_k^* \overline{F_i F_k^*} \quad .$$

The cross-correlations $\overline{F_i F_k^*}$ cannot be regarded as power densities because, although their weighted sum over all directions (i and k) is proportional to the total power, it is not possible to associate a single direction of arrival with $\overline{F_i F_k^*}$. (Is the direction i or k?) On the other hand, if

$$\overline{F_i F_k^*} = 0 \quad \text{for} \quad k \neq i \quad ,$$

the mean-square terminal voltage could be written

$$\overline{|V|^2} = \sum_i |a_i|^2 \overline{|F_i|^2} \quad ,$$

and a single direction of arrival i could be ascribed to each term contributing to the total power.

But the above assumption places too severe a restriction on the arriving signal components. It is neither reasonable nor necessary to assume that two signal components, arriving from directions that are arbitrarily close together, are uncorrelated. A physical antenna, having finite aperture dimensions, has a smooth pattern that does not change abruptly with small changes in direction: for radian changes in direction less than the reciprocal of the largest

aperture dimension expressed in wavelengths, the antenna pattern remains essentially constant. Thus, depending on the aperture dimensions, there is a "cone" of directions within which $a_k \approx a_i$. Let it be assumed that

$$\overline{F_i F_k^*} = 0 \quad \text{for all k such that } a_k \neq a_i \quad . \tag{2}$$

The mean-square terminal voltage can now be written

$$\overline{|V|^2} = \sum_i |a_i|^2 \sum_k \overline{F_i F_k^*} \quad , \tag{3}$$

and the second sum may be regarded as the power density, since a single direction of arrival i may be ascribed to it.

The relation, (2), and the resulting simplification, (3), indicate the nature of the assumption of incoherence that was implied by the introductory discussion and that will be made in this report. Essentially, we assume that no antenna we would consider building will have a narrow enough beamwidth to enable the antenna to resolve the correlation angle: the increment in angle-of-arrival over which arriving plane-wave components are appreciably correlated. Note that, although we assume no correlation over angular spreads as large as or larger than the beamwidth of some "biggest" antenna, we make no assumption about the shape of the cross-correlation for angular increments less than the correlation angle; none is necessary. It is quite conceivable, for example, that the cross-correlation would decrease more rapidly with increments in elevation than with increments in azimuth.

Since no experimental evidence tending to invalidate the assumption of incoherence made here has come to the attention of the author, the assumption will be considered in the light of various theories of the scattering mechanism. In terms of the Booker-Gordon theory,[3] we assume that the angle at the receiving site subtended by a length in the scattering region equal to the scale of turbulence is less than the beamwidth of the "biggest" antenna. Airborne refractometer measurements, as interpreted by Gordon,[4] yield values of the scale of turbulence of the order of 50 meters for tropospheric scatter. Even for a short 100-km path, for which the distance from the scattering region to the receiving site is about 50 km, the correlation angle is of the order of

$$\frac{50\ \text{m}}{50\ \text{km}} = 10^{-3} \text{ radian} = 3.4 \text{ minutes of arc.}$$

An antenna capable of resolving such a small correlation angle would have aperture dimensions at least of the order of 10^3 wavelengths – a 30-meter dish receiving a 10,000-Mcps scatter signal, for example. Thus, if the scattering mechanism is accurately described by the Booker-Gordon theory, the assumption of incoherence will be valid in most practical cases. Only in the case of large antennas, extremely high frequencies, and short path lengths, will the assumption become questionable.

Bailey, et al.,[5] have applied the Booker-Gordon theory to E-layer scatter, using a value of 100 meters for the scale of turbulence. Since ionospheric scatter becomes important at much lower frequencies (20 to 100 Mcps) and for greater path lengths (of the order of 10^3 km) than

does tropospheric scatter, it is clear that the assumption of incoherence is in complete agreement with this theory of E-layer scatter.

Villars and Weisskopf,[6] discussing turbulence from a fluid-dynamic point of view, show that kinetic energy in the atmosphere is dissipated by being transferred from a large turbulent eddy to successively smaller eddies. In spite of the resulting assortment of eddy sizes, they attribute the scattering of radio waves only to those eddies having a size of the order of the vertical wavelength,

$$L = \frac{\Lambda}{2 \sin \theta/2} \quad ,$$

where Λ is the wavelength and Θ is the scattering angle through which the radio rays are bent in the scattering process. The correlation angle is approximately the angle subtended by a length L in the scattering region at the receiving site. The angle varies inversely with the frequency, but the aperture dimensions of an antenna big enough to resolve the correlation angle are independent of frequency. They increase very nearly as the square of the path length. For a short 100-km path, the aperture dimensions would have to be of the order of 800 meters if the antenna were to resolve the correlation angle. Thus, in most practical situations, the assumption of incoherence is compatible with the Villars-Weisskopf theory.

Perhaps the most interesting justification of the assumption of incoherence is the following demonstration that, in certain circumstances, the maximum signal power available at the terminals of an antenna receiving coherent radiation is inversely proportional to the antenna gain. Needless to say, such a situation has not been observed with properly pointed antennas.

The coefficient a_i, which has been described as the amplitude of the receiving-antenna pattern in the i-direction, may be regarded as the amplitude of the plane-wave component that would leave the receiving antenna in the i-direction if a current of one ampere were applied to the antenna terminals. The density of power leaving the antenna in the i-direction is then proportional to $|a_i|^2$ and, if the indices are appropriately assigned to the directions,

$$W = \sum_i |a_i|^2 \tag{4}$$

is proportional to the total power radiated from the receiving antenna when it is excited with unit current. The quantity W is also proportional to the radiation resistance R_a of the receiving antenna, since $W = 1/2\ |I|^2\ R_a$, and we have chosen unit current I. Thus, the signal power available at the terminals of the receiving antenna is proportional to

$$W_r = \frac{\overline{|V|^2}}{R_a} = \frac{\text{Av.} \left\{ \left| \sum_i a_i F_i \right|^2 \right\}}{\sum_i |a_i|^2} \tag{5}$$

from (1) and (4). Explicit expressions corresponding to (4) and (5) will be developed in Secs. III and IV.

Now suppose that the antenna pattern is confined to directions for which the arriving plane-wave components F_i are perfectly correlated; the sums in (5) will cover only those directions. If F_i and F_k are "perfectly correlated," they will "fade together" with fluctuations in the scattering region: The temporal variations of F_i will be proportional to the temporal variations of F_k. We may write

$$F_i(t) = g(t)\, f_i \quad \text{with} \quad \overline{|g(t)|^2} = 1 \quad ,$$

where the complex time function g(t) is independent of direction i, and the complex numbers f_i are not time functions. In that case, (5) becomes

$$W_r = \frac{\overline{|g(t)|^2} \left| \sum_i a_i f_i \right|^2}{\sum_i |a_i|^2} = \frac{\left| \sum_i a_i f_i \right|^2}{\sum_i |a_i|^2} \quad ,$$

and, by the Schwarz inequality, the numerator is less than or equal to

$$\sum_i |a_i|^2 \sum_k |f_k|^2 \quad ,$$

which means that the received power, or at least W_r, which is proportional to it, is

$$W_r \leq \sum_i |f_i|^2 = \sum_i \overline{|F_i|^2} \quad .$$

The inequality holds with the equal sign when $a_i = f_i^*$. In that case, the more directions i involved in the sum, i.e., the greater the antenna beamwidth, the greater will be the received power. If, in addition, $\overline{|F_i|^2} = \overline{|F_k|^2}$ in the range of directions covered by the sum, the received power would be inversely proportional to the antenna gain!

In all fairness to the reader, it should be pointed out that the antenna pattern adjustment $a_i = f_i^*$ required for this unusual conclusion will, in general, be practical only if the magnitudes and phases of the f_i do not change abruptly from one direction to the next.

* * *

So far, we have justified the assumption of incoherence, which allows the received signal power to be written in terms of the antenna pattern and of the pattern of incoming power density, as in (3). Temporarily, we assumed that the fluctuations of the plane-wave components F_i are entirely due to fluctuations in the scattering region. But if the shape of the power density pattern were affected by the additional fluctuations caused by modulation of the transmitted signal, the forthcoming antenna design considerations, which are based on the power-density pattern, would also depend on the modulation. Fortunately, it appears that this is not the case, as we shall now see. Wide-band scatter signals may be distorted because of multipath transmission; it should be noted that we will not attempt to disprove the existence of this effect.

We wish to show that the incoming power density is proportional only to the mean transmitted power and is independent of the modulation. It will be convenient to calculate the power received with a very directive antenna, pointed in a typical direction, and to show that this power does not depend on the modulation. If this hypothetical receiving antenna is bigger than the "biggest" antenna we might consider building, and if it has a beamwidth of the order of the correlation angle, the power density is proportional to the mean-square signal voltage at its terminals. We will suppose, then, that our hypothetical receiving antenna delivers an output only when illuminated by plane-wave components arriving from within a cone having an angle of the order of the correlation angle, and that the pattern amplitude is constant within this cone.

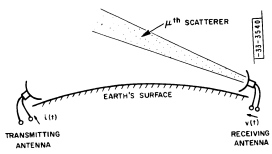

Fig. 3. The received signal is a superposition of signals scattered from the individual scatterers.

In what follows, it will be necessary to work with a model of the scattering mechanism. We will assume that the signal voltage at the receiving-antenna terminals can be calculated as the sum of signals scattered from moving scatterers in the receiving-antenna beam, as shown in Fig. 3. Neglecting the possible dispersion effects of a single scattering, the contribution to the received signal voltage from the μ^{th} scatterer can be written

$$b_\mu\, i(t - \tau_\mu) \quad ,$$

and the total received voltage is

$$v(t) = \sum_\mu b_\mu\, i(t - \tau_\mu) \quad .$$

i(t) is a measure of the instantaneous transmitted signal, say the current at the terminals of the transmitting antenna. The time delay τ_μ is the time-of-flight along the path from the transmitting antenna to the μ^{th} scatterer and thence to the receiving antenna; τ_μ changes as the μ^{th} scatterer moves. The real coefficient b_μ is the ratio of the component of v(t) coming from the μ^{th} scatterer to the transmitted signal measured τ_μ seconds earlier, $i(t - \tau_\mu)$; b_μ depends on the scattering strength of the μ^{th} scatterer, the antenna patterns, and inverse distance.

The usual expression for the instantaneous transmitted signal as the projection of a rotating vector is

$$i(t) = \mathcal{R}e\, [I(t)\, e^{j\omega_o t}] \quad ,$$

where ω_o is the radian carrier frequency. The vector I(t) describes the modulation, a stationary random process; if the transmitter were unmodulated, I(t) = constant. The received voltage can be similarly represented. In terms of the vector expression for i(t),

$$v(t) = \mathcal{R}e \left[e^{j\omega_o t} \sum_\mu b_\mu\, e^{-j\omega_o \tau_\mu}\, I(t - \tau_\mu) \right] \quad .$$

We will call the complex coefficient

$$b_\mu e^{-j\omega_o \tau_\mu} \equiv c_\mu \quad ,$$

so that the time-dependent vector giving the magnitude and phase of the received voltage is

$$V(t) = \sum_\mu c_\mu I(t - \tau_\mu) \quad .$$

According to the foregoing analysis, the magnitude of c_μ does not depend on the modulation, and its phase is the carrier phase delay of the path through the μ^{th} scatterer, again independent of the modulation. Actually, a single scattering may be slightly dispersive so that the magnitude and phase of the scattering strength of a single scatterer may depend weakly on the signal frequency. Also, the magnitudes and phases of the antenna patterns may vary slowly with signal frequency. These effects would make the coefficients c_μ dependent on the modulation. However, we are certainly justified in neglecting these dependences for most communication signals, since they generally occupy a very narrow band of frequencies relative to the carrier frequency. It would be unreasonable to suppose that the modulation depends on the random motions in the scattering region. Thus, we may take c_μ and $I(t - \tau_\mu)$ to be independent random variables, for, although both depend on τ_μ, the specification of $I(t - \tau_\mu)$ provides no information about c_μ, and conversely.

The power received by our hypothetical narrow-beam receiving antenna is proportional to

$$\overline{v^2(t)} = 1/2 \; \overline{|V(t)|^2} = 1/2 \text{ Av.} \left\{ \sum_\mu c_\mu I(t - \tau_\mu) \sum_\nu c_\nu^* I^*(t - \tau_\nu) \right\}$$

$$= 1/2 \sum_\mu \sum_\nu \overline{c_\mu c_\nu^* I(t - \tau_\mu) I^*(t - \tau_\nu)} \quad ,$$

and, in view of the statistical independence of the c's and the I's,

$$\overline{v^2(t)} = 1/2 \sum_\mu \sum_\nu \overline{c_\mu c_\nu^*} \; \overline{I(t - \tau_\mu) I^*(t - \tau_\nu)} \quad .$$

Finally, if the cross-correlation between the gains of paths through individual scatterers, $\overline{c_\mu c_\nu^*}$, decreases essentially to zero for such short differences in time-of-flight, $\tau_\mu - \tau_\nu$, that the delayed modulation is essentially unchanged, that is, if

$$\overline{c_\mu c_\nu^*} = 0 \text{ for all } \nu \quad (6)$$

such that $I(t - \tau_\nu) \neq I(t - \tau_\mu)$,

then the mean-square received voltage becomes

$$\overline{v^2(t)} = 1/2 \; \overline{|I(t)|^2} \sum_\mu \sum_\nu \overline{c_\mu c_\nu^*} \quad , \quad (7)$$

and the received power is seen to be proportional to the mean transmitted power and independent of the modulation, as required.

In arriving at the last result, it has again been necessary to assume a short correlation distance – in this case, in the direction of most rapid change of time-of-flight with distance: essentially in the vertical direction in the scattering region. But because of the usually small scattering angles Θ, through which the radio rays are bent in the scattering process, the time of flight over the μ^{th} path changes relatively slowly with the height of the μ^{th} scatterer. For example, Villars and Weisskopf have chosen a correlation distance (the size of eddies primarily responsible for scattering) in such a way that the difference in path length between paths passing through the top and bottom of such an eddy is just one wavelength. The difference in time-of-flight over these paths is therefore equal to a period of the carrier. Thus, according to the Villars-Weisskopf theory, a scatter signal would have to occupy a bandwidth of the order of the carrier frequency to invalidate the assumption, (6), leading to the result, (7)!

The signal bandwidths required by the Booker-Gordon theory to invalidate (6) are generally less than the carrier frequencies, particularly for the microwave frequencies and for the greater path lengths. Nevertheless, the bandwidths are great enough so that there can be little question about the validity of (6).

In this section, we have given an expression for the signal power available at the terminals of an antenna illuminated by scattered radio waves. We have seen that this expression can be interpreted as the sum of weighted power densities, provided that plane-wave components of the signal arriving from appreciably different directions are uncorrelated, and this assumption of incoherence has been justified. Finally, we have shown that the power density, which is basic to the body of this paper, is proportional to the mean transmitted power and is not dependent on the signal modulation.

Although this section has been concerned specifically with scatter reception, it seems probable that the conclusions reached here, and hence the forthcoming results based on them, are more generally applicable. The assumption of incoherence is certainly justified for a line-of-sight link, since the signal arrives essentially from one direction. Short-wave signals received over long-distance ionospheric paths may very likely also be treated in the manner of this report.

III. DESCRIPTION OF THE ANTENNA

Complete descriptions of the antenna pattern and of the pattern of incoming power density will be required for the design of an optimum antenna. The antenna is considered in this section, and its behavior when receiving scatter signals is considered in Sec. IV. In arriving at explicit descriptions of the antenna pattern and the power-density pattern, the concept of an angular spectrum of plane waves, as discussed by Booker and Clemmow[7] and by Booker, Ratcliffe and Shinn,[8] will be extended to three dimensions and used as a mathematical tool.

The receiving antenna will be characterized either by the electric fields it would produce if a radio-frequency current of one ampere were applied to its terminals or by the pattern of plane waves that would emerge from the antenna with the same excitation. Points in space will be located on a system of rectangular coordinates, with the antenna in the vicinity of the origin. As shown in Fig. 4, it will be convenient to speak of the positive z-direction as being "up," although the coordinate axes may always be oriented to suit a particular situation. It will also be convenient to assume that the region above the xy-plane is free space and that all the radiation from the antenna crosses the xy-plane and leaves the antenna in directions above the horizontal; at least, it will be assumed that any obstacles above the xy-plane are at great enough distances so that their presence does not materially affect the driving-point impedance of the antenna. Thus, we imagine an antenna located more or less at the origin of coordinates, radiating into the upper hemisphere, and having all its active conductors, supporting members, etc., in the lower hemisphere.

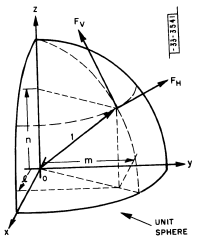

Fig. 4. Geometry of the space surrounding the receiving antenna.

Although the restriction to one hemisphere holds approximately for many scatter antennas that have negligible back radiation, the restriction would not be at all reasonable for simpler antennas, such as a dipole. In these latter cases, it will generally be possible to piece together separate solutions for each hemisphere.

We will measure distances in wavelengths, so that

$$\mathrm{Re}\left\{E\, e^{j2\pi(f_o t - z)}\right\}$$

is the expression for a component of the electric field of a plane wave of carrier frequency f_o, which has left the antenna and is traveling toward the zenith. We shall represent this wave component as

$$E\, e^{-j2\pi z} ,$$

it being understood that this and similar expressions for radio-frequency quantities are to be multiplied by $e^{j2\pi f_o t}$ and the real part taken.

In giving the direction of travel of a plane wave, we shall use the direction cosines ℓ, m and n, which are the projections of a unit vector on the x-, y- and z-axes, respectively, as shown in Fig. 4. We regard the direction cosines ℓ and m as the primary variables; we will use

$$\ell^2 + m^2 + n^2 = 1$$

to write n in terms of ℓ and m. Thus, we might represent a component of a plane wave ascending in the direction (ℓ, m) as

$$E\, e^{-j2\pi(x\ell+ym+z\sqrt{1-\ell^2-m^2})} \quad ,$$

and we would mean that the square root is to be taken with the positive sign, since the wave is traveling into the upper hemisphere. However, we shall often write n in place of the radical for brevity.

If the projection on the xy-plane of the intersection of an incremental solid angle with the unit sphere is $d\ell\, dm$, the increment of solid angle itself is

$$d\Omega = \frac{d\ell\, dm}{n} \quad . \tag{8}$$

We shall often have occasion to integrate plane-wave components and other functions of direction, (ℓ, m), over all directions in the upper hemisphere. These integrals will be written as

$$\int_{u.h.} d\Omega \quad \text{or} \quad \iint_{u.h.} d\ell\, dm \quad .$$

The second integral may also be taken to cover the inside of the unit circle, $\ell^2 + m^2 = 1$.

Now we will use a physical argument to express the fields, produced when the antenna is excited with a current of one ampere, as a superposition of the plane waves that emerge from the antenna. Let the x-, y- and z-components of the electric field at the point (x, y, z) be the complex functions, $E_x(x, y, z)$, $E_y(x, y, z)$ and $E_z(x, y, z)$, respectively. As shown in Fig. 4, let $F_H(\ell, m)\, d\Omega$ be the horizontally polarized component of the electric field of the plane wave that leaves the antenna and progresses into the incremental solid angle $d\Omega$ in the general direction (ℓ, m). The phase of $F_H(\ell, m)$ will be referred to the origin of coordinates. Similarly, let $F_V(\ell, m)\, d\Omega$ be the vertically polarized component. As usual, we have taken the horizontal component to be tangent to a horizontal circle of constant latitude on the unit sphere; the "vertical" component is tangent to a longitude circle.

At the point (x, y, z), the x-components of the plane-wave components add to give

$$E_x(x, y, z) = \int_{u.h.} d\Omega\, e^{-j2\pi(x\ell+ym+zn)} \times \left[-\frac{m}{\sqrt{\ell^2 + m^2}} F_H(\ell, m) - \frac{n\ell}{\sqrt{\ell^2 + m^2}} F_V(\ell, m) \right] .$$

Similarly,

$$E_y(x, y, z) = \int_{u.h.} d\Omega\, e^{-j2\pi(x\ell+ym+zn)} \times \left[\frac{\ell}{\sqrt{\ell^2 + m^2}} F_H(\ell, m) - \frac{nm}{\sqrt{\ell^2 + m^2}} F_V(\ell, m) \right] ,$$

$$E_z(x, y, z) = \int_{u.h.} d\Omega\, e^{-j2\pi(x\ell+ym+zn)} \left[\sqrt{\ell^2 + m^2}\, F_V(\ell, m) \right] .$$

Using the index i to represent one of the subscripts x, y or z, these three equations may all be written as

$$E_i(x, y, z) = \iint_{u.h.} P_i(\ell, m)\, e^{-j2\pi(x\ell+ym+zn)}\, d\ell\, dm \quad , \tag{9}$$

if the P_i's are defined by

$$\left.\begin{aligned} P_x &= -\frac{m}{n\sqrt{\ell^2 + m^2}} F_H - \frac{\ell}{\sqrt{\ell^2 + m^2}} F_V \\ P_y &= +\frac{\ell}{n\sqrt{\ell^2 + m^2}} F_H - \frac{m}{\sqrt{\ell^2 + m^2}} F_V \\ P_z &= +\frac{\sqrt{\ell^2 + m^2}}{n} F_V \end{aligned}\right\} . \tag{10}$$

The P's may be regarded as rectangular plane-wave components as contrasted with the F's, which are horizontal and vertical plane-wave components. It should be noted, however, that the resultant electric vector defined by the set of P's differs in magnitude from the electric vector defined by the F's by the factor $1/n$, which was absorbed into the P's by the change of variable, (8).

Before continuing with the argument, we will pause to consider the relations, (10). If the F's are eliminated, there results

$$\ell\, P_x + m\, P_y + n\, P_z = 0 \quad , \tag{10a}$$

which may be interpreted to mean that the scalar product of a unit vector, pointing in the direction of motion of a plane wave, with the electric vector of that plane wave is zero. In other words, the electric vector is normal to the direction of propagation. This result is a consequence of our having (correctly) taken the F's to be normal to the direction of propagation.

Suppose that we have two plane waves with rectangular components $P_{i1}(\ell, m)$ and $P_{i2}(\ell, m)$ and corresponding horizontal and vertical components $F_{k1}(\ell, m)$ and $F_{k2}(\ell, m)$. (We use the first subscript, in this case, k, on the F's to indicate the polarization, H or V, in the same way that we have used the first subscript on the P's to indicate which rectangular component, x, y or z, is meant.) Using (10), we find

$$P_{x1}\,P_{x2} + P_{y1}\,P_{y2} + P_{z1}\,P_{z2} = \frac{1}{n^2}\,(F_{H1}\,F_{H2} + F_{V1}\,F_{V2}) \quad ,$$

or, more compactly,

$$\sum_i P_{i1}\,P_{i2} = \frac{1}{n^2}\sum_k F_{k1}\,F_{k2} \quad , \tag{10b}$$

so that, if $P_{i1} = P_{i2}^* = P_i$,

$$\sum_i |P_i|^2 = \frac{1}{n^2}\sum_k |F_k|^2 \quad . \tag{10c}$$

Continuing the argument, we turn our attention to (9), which gives the field components in terms of the rectangular plane-wave components. For a fixed value of z, (9) relates a function of (x, y) to a function of (ℓ, m):

$$E_i(x,y,z)\Big|_{z=\text{const.}} = \iint_{u.h.} [P_i(\ell,m)\, e^{-j2\pi z n}]\, e^{-j2\pi(x\ell+ym)}\, d\ell\, dm \quad . \tag{9a}$$

In the plane, z = constant, let E_i have a Fourier transform, and assume that the inverse transform also exists. The inverse transform could be written

$$E_i(x,y,z)\Big|_{z=\text{const.}} = \int_{-\infty}^{\infty}\int [P_i'(\ell,m)\, e^{-j2\pi z n}]\, e^{-j2\pi(x\ell+ym)}\, d\ell\, dm \quad .$$

But in the subregion, $\ell^2 + m^2 \leqslant 1$, we have found that $P_i' = P_i$. It would appear that some "plane-wave components," which leave the antenna in the fictitious directions, $\ell^2 + m^2 > 1$, but which nevertheless may contribute to E_i, are missing from (9a). Indeed, the fields produced by a physical antenna on a plane in front of its aperture (z = constant) may vary so rapidly with position, (x, y), that the Fourier transform of these fields cannot possibly be zero everywhere outside $\ell^2 + m^2 = 1$. Let us examine the nature of the contribution to E_i of these fictitious "plane-wave components." Writing n in terms of ℓ and m,

$$dE_i(x,y,z) = P_i(\ell,m)\, e^{-j2\pi(x\ell+ym+z\sqrt{1-\ell^2-m^2})}\, d\ell\, dm \quad .$$

Outside of $\ell^2 + m^2 = 1$ the radical is imaginary, and we must reconsider its sign. If n were positive imaginary, the contribution $dE_i(x,y,z)$ would increase exponentially with height z, which would not be compatible with the chosen source location. But if n is negative imaginary, the contribution decreases exponentially with height, which seems more reasonable.

Since the contribution to the fields from the $P_i(\ell, m)$ for which $\ell^2 + m^2 > 1$ becomes negligible at sufficiently great distances from the antenna, we associate the storage fields with these fictitious "plane-wave components." The radiation fields, as we have already seen, are accounted for by the components of those plane waves which travel in the real directions, $\ell^2 + m^2 \leqslant 1$. Thus, the complete expression for the fields is

$$\left.\begin{aligned} E_i(x,y,z) &= \int_{-\infty}^{\infty}\iint P_i(\ell,m)\, e^{-j2\pi(x\ell+ym+zn)}\, d\ell\, dm \\ &\text{n is positive for } \ell^2 + m^2 \leqslant 1 \\ &\text{n is negative imaginary for } \ell^2 + m^2 > 1 \end{aligned}\right\} \quad , \tag{11}$$

and the direct transform, giving the rectangular plane-wave components in terms of the fields, is

$$P_i(\ell,m) = \int_{-\infty}^{\infty}\iint E_i(x,y,z)\, e^{j2\pi(x\ell+ym+zn)}\, dx\, dy \quad . \tag{12}$$

The height z is a parameter in both of these expressions. Note that the fields at any point in the region above the xy-plane can be calculated from the P's using (11). Only two of these three rectangular plane-wave components are required, since the three are interrelated by (10a). The P's are determined, in turn, through (12), by the fields on one of the planes, z = constant, above the antenna.

The foregoing argument, although not rigorous, has provided a considerable physical interpretation for the results, (11) and (12), and the supporting relations. The results can be made rigorous by showing that the fields given by (11) satisfy Maxwell's equations in the region where (11) is applicable, and that they match the boundary conditions appropriate to that region. The Maxwell equations applicable to the electric field in the free-space upper hemisphere are the wave equation and the divergence relation, usually written

$$\nabla^2 \vec{\mathcal{E}} = \epsilon\mu\, \frac{\partial^2 \vec{\mathcal{E}}}{\partial t^2} \quad , \quad \text{and} \quad \nabla \cdot \vec{\mathcal{E}} = 0 \quad .$$

Since we have measured distances in wavelengths, the wave equation becomes, for the sinusoidal steady state,

$$\frac{\partial^2 E_i}{\partial x^2} + \frac{\partial^2 E_i}{\partial y^2} + \frac{\partial^2 E_i}{\partial z^2} = -4\pi^2 E_i \quad , \quad i = x, y, z \quad ,$$

and for our purposes these expressions are a satisfactory approximation to the wave equation when the E_i are the complex instantaneous amplitudes of the field components of a typical modulated wave. As may be seen by applying them to (11), these expressions are satisfied if $\ell^2 + m^2 + n^2 = 1$, as we intend. The divergence relation may be written

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0 \quad ,$$

and it is satisfied by (11) if

$$\ell P_x + m P_y + n P_z = 0 \quad , \tag{10a}$$

as we discovered previously.

The region of interest is the upper hemisphere, which has the xy-plane as its only boundary near the source. The region will be excited by a physical antenna located in the vicinity of the

origin and somewhat below the xy-plane. Surely, the Fourier transform of the fields set up on the xy-plane by this antenna will exist, and the inverse transform will also exist. Thus, boundary conditions can be matched on the xy-plane. The choice of sign for the direction cosine n, given in (11), will insure that the fields given by (11) will decrease appropriately in intensity at great distances from the origin.

* * *

Next we will study the behavior of the fields at a point a great distance

$$r = \sqrt{x^2 + y^2 + z^2}$$

from the antenna and in the (real) direction (ℓ_o, m_o, n_o) in the upper hemisphere, so that the point of space in question is

$$(x, y, z) = (r\ell_o, rm_o, rn_o) \quad .$$

We expect the fields at that point to behave like

$$E_i(r\ell_o, rm_o, rn_o) \sim \frac{e^{-j2\pi r}}{r} \quad ,$$

so it is appropriate to evaluate the limit,

$$L = \lim_{r\to\infty} r e^{j2\pi r} E_i(r\ell_o, rm_o, rn_o)$$

$$= \lim_{r\to\infty} r e^{j2\pi r} \iint_{-\infty}^{\infty} P_i(\ell, m)\, e^{-j2\pi r(\ell\ell_o + mm_o + nn_o)}\, d\ell\, dm \quad ,$$

from (11). Since we are interested in the far fields, only those plane-wave components for which $\ell^2 + m^2 \leq 1$ will contribute, and the range of integration may be reduced to the upper hemisphere. Using (8),

$$L = \lim_{r\to\infty} r e^{j2\pi r} \int_{u.h.} n P_i(\ell, m)\, e^{-j2\pi r(\ell\ell_o + mm_o + nn_o)}\, d\Omega \quad .$$

It will be convenient to regard the exponent as the scalar product of two unit vectors: one, $\vec{U}_o$, having the direction (ℓ_o, m_o, n_o), and the other, $\vec{U}$, pointed at the incremental solid angle $d\Omega$:

$$L = \lim_{r\to\infty} r \int_{u.h.} n P_i(\ell, m)\, e^{j2\pi r(1 - \vec{U}\cdot\vec{U}_o)}\, d\Omega \quad .$$

The geometry of this integral is shown in Fig. 5. We will integrate first around strips on the unit sphere in which the scalar product $\vec{U} \cdot \vec{U}_o$, and hence the entire exponent, is constant; then we will add up the contributions of the strips. The strips are defined by circles of constant

latitude, formed with the polar axis in the direction of $\vec{U}_o$. Since the integral is taken only over the upper hemisphere, some of the strips are incomplete; we may account for this condition by defining the integrand to be zero below the xy-plane.

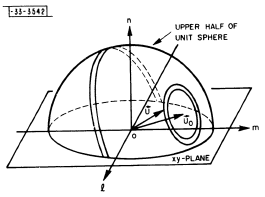

Fig. 5. $\vec{U} \cdot \vec{U}_o$ is constant throughout one of the strips.

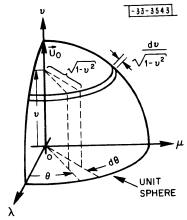

Fig. 6. The strips are parallel to the λμ-plane in the new coordinate system.

The integration is facilitated by taking new axes, (λ, μ, ν), in such a way that the directions of ν and $\vec{U}_o$ coincide, as shown in Fig. 6. The new variables of integration will be the direction cosine ν and the azimuthal angle Θ defined with the polar axis in the ν-direction. We will define

$$Q(\nu, \Theta) = \begin{cases} n\, P_i(\ell, m) & \text{for} \quad n \geq 0 \quad , \\ 0 & \text{for} \quad n < 0 \quad . \end{cases}$$

With these changes, the limit becomes

$$L = \lim_{r\to\infty} r \int_{-1}^{1} e^{j2\pi r(1-\nu)} \frac{d\nu}{\sqrt{1-\nu^2}} \int_0^{2\pi} Q(\nu, \Theta)\, \sqrt{1-\nu^2}\, d\Theta$$

$$= \lim_{r\to\infty} r \int_{-1}^{1} e^{j2\pi r(1-\nu)}\, d\nu \int_0^{2\pi} Q(\nu, \Theta)\, d\Theta$$

$$= \lim_{r\to\infty} r \int_{-1}^{1} q(\nu)\, e^{j2\pi r(1-\nu)}\, d\nu \quad ,$$

where

$$q(\nu) = \int_0^{2\pi} Q(\nu, \Theta)\, d\Theta$$

Although we will not attempt an explicit expression for the function $q(\nu)$, some of its properties are of interest. The rectangular plane-wave components P_i, which are the Fourier transforms of the components of the electric field set up on the xy-plane by a physical antenna, will be continuous, bounded functions of direction in the upper hemisphere. The function $q(\nu)$ is an integral of one of the P_i over a strip having a length that is also (in general) a continuous, bounded

function of ν; thus, $q(\nu)$ will be a continuous, bounded function of ν in the range $-1 \leq \nu \leq 1$. (In the particular case where the direction of $\vec{U}_o$ is the vertical, the strip length will be discontinuous at the point $\nu = n = 0$, corresponding to the horizon, but the integrand $n P_i$ decreases continuously to zero with n at that point, making $q(\nu)$ again continuous and bounded.) Also, since we are interested in directions $\vec{U}_o$ in the upper hemisphere, the value $q(-1)$ would involve only contributions from the lower hemisphere; in other words, $q(-1) = 0$. A representative plot of $|q(\nu)|$ is given in Fig. 7.

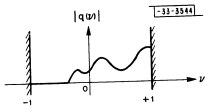

Fig. 7. $q(\nu)$ is continuous and bounded, and $q(-1) = 0$.

Consider the following integration by parts:

$$\int_{-1}^{1} e^{j2\pi r(1-\nu)}\, dq(\nu) = q(1) - e^{j4\pi r} q(-1) + j2\pi r \int_{-1}^{1} q(\nu)\, e^{j2\pi r(1-\nu)}\, d\nu \quad .$$

The second member on the right has been shown to be zero, and as $r \to \infty$, the left-hand side goes to zero according to the Riemann-Lebesque lemma. Thus, in the limit,

$$0 = q(1) - 0 + j2\pi \lim_{r\to\infty} \int_{-1}^{1} q(\nu)\, e^{j2\pi r(1-\nu)}\, d\nu$$

or

$$L = \lim_{r\to\infty} r \int_{-1}^{1} q(\nu)\, e^{j2\pi r(1-\nu)}\, d\nu = \frac{j}{2\pi}\, q(1)$$

$$= \frac{j}{2\pi} \int_{0}^{2\pi} Q(1,\Theta)\, d\Theta = j n_o P_i(\ell_o, m_o) \quad .$$

We have shown that

$$L = \lim_{r\to\infty} r\, e^{j2\pi r}\, E_i(r\ell_o, rm_o, rn_o) = j n_o P_i(\ell_o, m_o) \quad , \tag{13}$$

which means that, at great distances r from the antenna, the electric field components in the direction (ℓ_o, m_o, n_o) are

$$E_i(r\ell_o, rm_o, rn_o) \approx j n_o P_i(\ell_o, m_o)\, \frac{e^{-j2\pi r}}{r} \quad .$$

We note that the far field propagates in the direction away from the antenna, and that, since its components are proportional to the corresponding rectangular plane-wave components, the total electric field is normal to the direction of propagation.

The far field may also be described in terms of the horizontal and vertical plane-wave components, F_H and F_V. At the distant point $(r\ell_o, rm_o, rn_o)$, the horizontal component of the total electric field, which has the same direction as F_H in Fig. 4, is

$$E_H = -\frac{m_o}{\sqrt{\ell_o^2 + m_o^2}}\, E_x + \frac{\ell_o}{\sqrt{\ell_o^2 + m_o^2}}\, E_y \quad ,$$

and the vertical component, having the same direction as F_V, is

$$E_V = -\frac{n_o \ell_o}{\sqrt{\ell_o^2 + m_o^2}}\, E_x - \frac{n_o m_o}{\sqrt{\ell_o^2 + m_o^2}}\, E_y + \sqrt{\ell_o^2 + m_o^2}\, E_z \quad .$$

Using (13) to write the far fields in terms of the rectangular plane-wave components, and (10) to convert to the horizontal and vertical plane-wave components, there results, for large distances r,

$$\left.\begin{aligned} E_H(r\ell_o, rm_o, rn_o) &\to j F_H(\ell_o, m_o)\, \frac{e^{-j2\pi r}}{r} \\ \text{and} \\ E_V(r\ell_o, rm_o, rn_o) &\to j F_V(\ell_o, m_o)\, \frac{e^{-j2\pi r}}{r} \end{aligned}\right\} \quad . \tag{14}$$

* * *

At a great distance, $R = \Lambda r$ meters, Λ being the wavelength, the Poynting vector is directed away from the antenna. Its magnitude,

$$S = \frac{1}{2\eta} \sum_i |E_i|^2 \text{ watts/meter}^2 \quad ,$$

(η = characteristic impedance of free space)

is the power from the antenna crossing a square meter of the surface of a sphere of radius R meters, centered at the origin. The solid angle subtended by that square meter at the origin is

$$\Omega_R = \frac{1}{R^2} \quad ,$$

and the angular density of power from the antenna is

$$w = \frac{S}{\Omega_R} = \frac{R^2}{2\eta} \sum_i |E_i|^2 \text{ watts/solid angle} \quad .$$

The concept of an angular density of power is useful for the purposes of this paper. When we refer to a power density, we will mean an angular density.

Using (13) and (10c), the density of power leaving the antenna and progressing in the direction (ℓ, m) is

$$w(\ell, m) = \frac{(\Lambda r)^2}{2\eta} \left(\frac{n}{r}\right)^2 \sum_i |P_i(\ell, m)|^2$$

$$= \frac{\Lambda^2}{2\eta}\, n^2 \sum_i |P_i(\ell, m)|^2 = \frac{\Lambda^2}{2\eta} \sum_k |F_k(\ell, m)|^2 \quad , \tag{15}$$

and the total power emitted from the antenna is

$$\left.\begin{aligned} W &= \frac{\Lambda^2}{2\eta} \sum_k \int_{u.h.} |F_k(\ell, m)|^2 \, d\Omega \\ &= \frac{\Lambda^2}{2\eta} \sum_i \iint_{u.h.} n \, |P_i(\ell, m)|^2 \, d\ell \, dm \\ &= \frac{\Lambda^2}{2\eta} \sum_i Re\left\{ \iint_{-\infty}^{\infty} n \, |P_i(\ell, m)|^2 \, d\ell \, dm \right\} \end{aligned}\right\} , \qquad (16)$$

since n is imaginary outside of $\ell^2 + m^2 = 1$.

The ratio

$$\frac{w(\ell, m)}{W}$$

may be called the gain of the antenna in the direction (ℓ, m): neglecting antenna losses, the ratio multiplied by 4π is the gain relative to the gain of an "isotropic radiator."

Before closing this section, we will digress to develop expressions for the transmitted power in terms of the near fields of the antenna. Starting with the last form of (16) and using (12),

$$W = \frac{\Lambda^2}{2\eta} \sum_i Re\left\{ \iint_{-\infty}^{\infty} n \, P_i(\ell, m) \, d\ell \, dm \iint_{-\infty}^{\infty} E_i^*(x, y, o) \, e^{-j2\pi(x\ell + ym)} dx \, dy \right\} .$$

Changing the order of integration,

$$W = \frac{\Lambda^2}{2\eta} \sum_i Re\left\{ \iint_{-\infty}^{\infty} E_i^*(x, y, o) \, dx \, dy \iint_{-\infty}^{\infty} n \, P_i(\ell, m) \, e^{-j2\pi(x\ell + ym)} d\ell \, dm \right\} ,$$

and the last integral is recognized from (11) to be

$$\frac{j}{2\pi} \frac{\partial E_i(x, y, o)}{\partial z}$$

so that the transmitted power can be written

$$W = \frac{\Lambda^2}{2\eta} \sum_i Re\left\{ \frac{j}{2\pi} \iint_{-\infty}^{\infty} E_i^*(x, y, o) \frac{\partial E_i(x, y, o)}{\partial z} \, dx \, dy \right\} , \qquad (17)$$

which gives the transmitted power in terms of the fields and their z-derivatives on the xy-plane. The same result could have been obtained by integrating the real part of the vertical component of the Poynting vector over the xy-plane. Next we seek an expression for the transmitted power involving only the fields on the xy-plane. Starting again with (16) and using (12) to express both P_i and P_i^*,

$$W = \frac{\Lambda^2}{2\eta} \sum_i \iint_{u.h.} n \, d\ell \, dm \iint_{-\infty}^{\infty} E_i(x, y, o) \, e^{j2\pi(x\ell + ym)} dx \, dy$$

$$\times \iint_{-\infty}^{\infty} E_i^*(x', y', o) \, e^{-j2\pi(x'\ell + y'm)} \, dx' \, dy' ,$$

or with $x' = x + u$, $y' = y + v$,

$$W = \frac{\Lambda^2}{2\eta} \sum_i \iint_{u.h.} n \, d\ell \, dm \iint_{-\infty}^{\infty} E_i(x, y, o) \, dx \, dy$$

$$\times \iint_{-\infty}^{\infty} E_i^*(x + u, y + v, 0) \, e^{-j2\pi(u\ell + vm)} \, du \, dv ,$$

which becomes

$$W = \frac{\Lambda^2}{2\eta} \sum_i \iint_{-\infty}^{\infty} du \, dv \iint_{u.h.} n \, e^{-j2\pi(u\ell + vm)} d\ell \, dm$$

$$\times \iint_{-\infty}^{\infty} E_i(x, y, o) \, E_i^*(x + u, y + v, 0) \, dx \, dy$$

when the order of integration is changed. The last integral is the spatial auto-correlation of field components on the xy-plane:

$$\phi_i(u, v) = \iint_{-\infty}^{\infty} E_i(x, y, o) \, E_i^*(x + u, y + v, 0) \, dx \, dy . \qquad (18)$$

The second integral, which plays the role of a weighting function, is the Fourier transform of the real part of the direction cosine n:

$$N_r(u, v) = \iint_{u.h.} \sqrt{1 - \ell^2 - m^2} \, e^{\pm j2\pi(u\ell + vm)} d\ell \, dm .$$

This integral is best evaluated in spherical coordinates and with repeated reference to Watson.[9] At first, let $\ell = \rho \cos\Theta$, $m = \rho \sin\Theta$, so that

$$N_r(u, v) = \int_0^1 \rho \, d\rho \int_0^{2\pi} d\Theta \left[\sqrt{1 - \rho^2} \, e^{\pm j2\pi\rho(u \cos\Theta + v \sin\Theta)} \right]$$

$$= 2\pi \int_0^1 \rho \sqrt{1 - \rho^2} \, J_0\left(2\pi\rho\sqrt{u^2 + v^2}\right) d\rho .$$

Then let $\rho = \sin\phi$:

$$N_r(u,v) = 2\pi \int_0^{\pi/2} J_0\left(2\pi\sqrt{u^2+v^2}\,\sin\phi\right) \sin\phi\,\cos^2\phi\,d\phi$$

$$= 2\pi\,\frac{2^{1/2}\Gamma(3/2)}{\left(2\pi\sqrt{u^2+v^2}\right)^{3/2}}\,J_{3/2}\left(2\pi\sqrt{u^2+v^2}\right) \quad .$$

$$N_r(u,v) = \frac{1}{2\pi(u^2+v^2)}\left[\frac{\sin\left(2\pi\sqrt{u^2+v^2}\right)}{2\pi\sqrt{u^2+v^2}} - \cos\left(2\pi\sqrt{u^2+v^2}\right)\right] \quad . \tag{19}$$

In terms of the auto-correlation of field components on the xy-plane, (18), and the weighting function, (19), the transmitted power may be written

$$W = \frac{A^2}{2\eta}\sum_i \int_{-\infty}^{\infty}\int N_r(u,v)\,\Phi_i(u,v)\,du\,dv \quad . \tag{20}$$

IV. DESCRIPTION OF SCATTER

It will be convenient to describe the scattered fields arriving at the receiving site in the same terms that were used in Sec. III to describe the fields that would be produced by the receiving antenna if it were excited with a unit current. The same coordinate system (Fig. 4) and similar notation will be used. The electric field components of the arriving radiation will be called $E_{ia}(x, y, z)$. Similarly, the horizontal and vertical plane-wave components arriving from the direction (ℓ, m) in the upper hemisphere will be designated $F_{ka}(\ell, m)$, and the corresponding rectangular plane-wave components will be $P_{ia}(\ell, m)$. The second subscript, a, will be used to indicate arriving quantities. The phases of the $F_{ka}(\ell, m)$ and of the $P_{ia}(\ell, m)$ (and also the amplitudes of these quantities when $\ell^2 + m^2 > 1$) will again be referred to the origin of coordinates.

We will assume that all the sources of radiation incident on the receiving antenna are located above the xy-plane. Physically, this radiation will be reflected (or scattered) from the receiving antenna and other obstacles below the xy-plane, so that the total field above the xy-plane will be a superposition of the incident fields and the fields reflected from below the xy-plane. The reflected fields will depend in part on the structure of the receiving antenna and its surroundings, but the incident fields will depend only on the transmitter and its antenna, the scattering mechanism and the geographical location of the receiving site. We will use only the incident fields (not the total fields) to describe the arriving scatter radiation. The field components E_{ia}, then, pertain only to the incident fields; they may, however, be regarded as the total field components that would be present if the receiving antenna and other reflecting objects were removed from below the xy-plane, leaving free space. Similarly, the plane-wave components F_{ka} and P_{ia} represent plane waves arriving only from above the xy-plane; if they correspond to the total field, the lower hemisphere must be free space. This restriction to incident fields only will have to be considered again when the density of power arriving from the scattering region is to be measured in terms of the incident fields.

The arguments of the last section leading up to and justifying (11) and (12) could be repeated almost word-for-word for the arriving quantities. The result would also be similar: the expression for the arriving fields is

$$\left.\begin{aligned} E_{ia}(x,y,z) = \int_{-\infty}^{\infty}\int P_{ia}(\ell,m)\,e^{j2\pi(x\ell+ym+zn)}\,d\ell\,dm \\ n \text{ is positive for } \ell^2+m^2 \leqslant 1 \\ n \text{ is negative imaginary for } \ell^2+m^2 > 1 \end{aligned}\right\} , \tag{21}$$

and the direct transform, giving the rectangular plane-wave components in terms of the fields, is

$$P_{ia}(\ell,m) = \int_{-\infty}^{\infty}\int E_{ia}(x,y,z)\,e^{-j2\pi(x\ell+ym+zn)}\,dx\,dy \quad . \tag{22}$$

Note that the exponents in (11) and (21), and also the exponents in (12) and (22), have opposite signs. This is because the transmitted fields, which leave in the direction (ℓ, m) are propagating in the opposite direction from the arriving fields, which arrive from the direction (ℓ, m).

Making the direction cosine n again negative imaginary for $\ell^2 + m^2 > 1$ insures that the storage fields die out at the greater distances (more negative z this time) from the source. In fact, it will usually be reasonable to assume that the storage fields of the source do not contribute to the fields arriving at the receiving site. In these cases, $P_{ia}(\ell, m) = 0$ for $\ell^2 + m^2 > 1$, and the range of integration in (21) may be reduced to the upper hemisphere.

The relations (10) between the horizontal and vertical plane-wave components and the rectangular plane-wave components do not depend on the direction of propagation of the components. They are also valid for the arriving quantities.

Up to this point, we have not mentioned the magnetic fields that must accompany the corresponding electric fields. In general, since the magnetic field can be computed from the electric field with the aid of the appropriate Maxwell equation, it is not necessary to be concerned with the magnetic fields. However, magnetic quantities will be convenient in the forthcoming determination of the open-circuit voltage at the receiving-antenna terminals due to an arriving scatter signal.

Temporarily, then, we introduce the arriving magnetic-field components $H_{ia}(x, y, z)$ and their Fourier transforms $Q_{ia}(\ell, m)$, the (magnetic) rectangular plane-wave components. The magnetic field also obeys the wave equation and a divergence relation. By analogy with (21), (22) and (10a),

$$\left.\begin{aligned} H_{ia}(x, y, z) = \int_{-\infty}^{\infty}\int Q_{ia}(\ell, m)\, e^{j2\pi(x\ell + ym + zn)}\, d\ell\, dm \\ n \text{ is positive for } \ell^2 + m^2 \leqslant 1 \\ n \text{ is negative imaginary for } \ell^2 + m^2 > 1 \end{aligned}\right\}, \tag{23}$$

$$Q_{ia}(\ell, m) = \int_{-\infty}^{\infty}\int H_{ia}(x, y, z)\, e^{-j2\pi(x\ell + ym + zn)}\, dx\, dy \quad , \tag{24}$$

$$\ell\, Q_x + m\, Q_y + n\, Q_z = 0 \quad . \tag{25}$$

The similarity here is so complete that no comment is required.

Next, we relate the arriving magnetic fields to the arriving electric fields, using the Maxwell curl equation which is usually written

$$\nabla \times \vec{\mathcal{E}} = -\frac{\partial \vec{\mathcal{B}}}{\partial t} \quad .$$

In the present circumstances (sinusoidal steady state, free space, and distances measured in wavelengths), this equation may be written

$$\frac{\partial E_{za}}{\partial y} - \frac{\partial E_{ya}}{\partial z} = -j2\pi\eta\, H_{xa} \quad ,$$

$$\frac{\partial E_{xa}}{\partial z} - \frac{\partial E_{za}}{\partial x} = -j2\pi\eta\, H_{ya} \quad ,$$

$$\frac{\partial E_{ya}}{\partial x} - \frac{\partial E_{xa}}{\partial y} = -j2\pi\eta\, H_{za} \quad ,$$

or, in terms of the arriving rectangular plane-wave components, from (21) and (23),

$$\left.\begin{aligned} m\, P_{za} - n\, P_{ya} = -\eta\, Q_{xa} \\ n\, P_{xa} - \ell\, P_{za} = -\eta\, Q_{ya} \\ \ell\, P_{ya} - m\, P_{xa} = -\eta\, Q_{za} \end{aligned}\right\} . \tag{26}$$

It is enlightening to summarize in vector notation. Let $\vec{P}_a(\ell, m)$ be an electric vector having the components $P_{xa}(\ell, m)$, $P_{ya}(\ell, m)$ and $P_{za}(\ell, m)$; let $\vec{Q}_a(\ell, m)$ be the corresponding magnetic vector; and let $\vec{U}(\ell, m)$ be a vector in the direction (ℓ, m). The wave equation requires that

$$\vec{U} \cdot \vec{U} = 1 \quad ,$$

so that $\vec{U}$ is a unit vector. The divergence relations may be written

$$\vec{U} \cdot \vec{P} = 0 \tag{10a'}$$

and

$$\vec{U} \cdot \vec{Q} = 0 \quad , \tag{25'}$$

and the curl equations become

$$\vec{U} \times \vec{P}_a = -\eta\, \vec{Q}_a \tag{26'}$$

and

$$\eta\, \vec{U} \times \vec{Q}_a = \vec{P}_a \quad .$$

Although these vector equations have meaning only for the real directions, their algebraic counterparts, written in terms of the direction cosines, are also valid for the fictitious directions $\ell^2 + m^2 > 1$, as we have seen. In view of the definition of the Poynting vector, it is clear from either of the curl equations that $\vec{P}_a(\ell, m)$ and $\vec{Q}_a(\ell, m)$ represent radiation that is propagating in the opposite direction from that of $\vec{U}$. In any case, the signs of the vector products would have been reversed if we had been describing the transmitted fields. It may also be seen that a number of these vector relations are redundant.

* * *

Now we are ready to develop an expression for the open-circuit voltage V at the receiving antenna terminals due to an arriving scatter signal. As suggested by Fig. 8, we will calculate the open-circuit voltage at the terminals of a small test dipole when the receiving antenna is

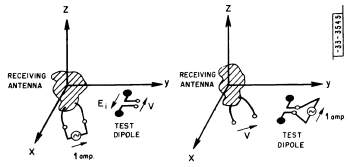

Fig. 8. The current source and voltmeter may be interchanged without affecting the voltmeter reading.

excited with a radio-frequency current of one ampere; then we will use reciprocity to infer that the same open-circuit voltage would appear at the terminals of the receiving antenna if, instead, the test dipole carried unit current. Let the test dipole be located at the point (x, y, z) above the xy-plane, let it be oriented in the i-(x-, y- or z-) direction, and let it have an effective length of a meters, short compared with the wavelength. The open-circuit voltage at its terminals, caused by excitation of the receiving antenna, is then

$$V = a\, E_i(x, y, z) \quad ,$$

and the same voltage would appear at the receiving-antenna terminals if the test dipole were excited with one ampere.

On the other hand, the scatter fields arrive from a variety of directions, producing very complex field patterns at the receiving site. A single test dipole could hardly be expected to reproduce such patterns. Tentatively, we suppose that a current sheet on some plane, z = constant, above the xy-plane is capable of producing field patterns that duplicate the scatter-field patterns in the vicinity of the antenna. The current sheet may be considered to arise from an infinite array of incremental test dipoles, oriented in the x- and y-directions on the plane z = constant and excited with different currents. Since we know the open-circuit voltage at the receiving-antenna terminals due to the excitation of each of these dipoles, these contributions may be superposed to obtain the open-circuit voltage due to the current sheet.

Let the components of the linear current density on the sheet be $J_x(x, y, z)$ and $J_y(x, y, z)$ amperes/meter. Consider an incremental rectangle bounded by (x, x + dx) and (y, y + dy) in the plane z = constant. The total current flowing in the x-direction in this rectangle is $(\Lambda dy)\, J_x(x, y, z)$ amperes, and the effective length of this rectangle of current is Λdx meters. Its contribution to the open-circuit voltage is therefore

$$dV = -(\Lambda dx)\, E_x(x, y, z)\, (\Lambda dy)\, J_x(x, y, z) \quad .$$

(The negative sign arises from the polarity conventions shown in Fig. 8.) Superposing the contributions from all the x-components and all the y-components of current density on the sheet, we obtain the open-circuit voltage due to the current sheet:

$$V = -\Lambda^2 \int_{-\infty}^{\infty}\int (E_x J_x + E_y J_y)\, dx\, dy \quad .$$

The Biot-Savart law states that the magnetomotive force around a closed loop equals the net current enclosed; directions are given by the right-hand rule. If the loop lies in the yz-plane with horizontal sides just above and just below the current sheet, as shown in Fig. 9, the current involved will be related to J_x. If the loop has differential length dy and vanishing height dz << dy, so that displacement currents can be neglected, we have

$$(\Lambda dy)\, J_x(x, y, z) = (\Lambda dy)\, [H_y^-(x, y, z) - H_y^+(x, y, z)] \quad ,$$

$$J_x = H_y^- - H_y^+ \quad ,$$

where, as in Fig. 9, H_y^+ and H_y^- are the y-components of the magnetic field just above and just below the current sheet. In order to compare the fields due to the current sheet with the incident scatter fields, we imagine that the receiving antenna and other reflecting objects are removed from below the xy-plane, leaving free space. In this case, the tangential magnetic fields above and below the current sheet are equal in amplitude and opposite in direction:

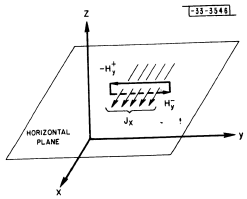

Fig. 9. A loop in a vertical plane encloses current flowing horizontally on the sheet.

$$-H_y^+ = H_y^- \quad \text{so that} \quad J_x = 2H_y^- \quad .$$

The corresponding relation,

$$J_y = -2H_x^- \quad ,$$

for the other components could be obtained by applying the Biot-Savart law to a similar loop in the xz-plane. In terms of the magnetic field tangential to a plane just below the current sheet, the open-circuit voltage of the receiving antenna is

$$V = 2\Lambda^2 \int_{-\infty}^{\infty}\int (E_y H_x^- - E_x H_y^-)\, dx\, dy \quad .$$

At this point in the discussion, the region below the current sheet is free space. Accordingly, H_x^- and H_y^- must be components of a magnetic field that is propagating into the lower

hemisphere. The corresponding plane-wave components are given by (24), and the missing vertical component is then given by (25). The magnetic fields anywhere below the current sheet could be calculated from (23), and the corresponding electric fields could be calculated from (26), (10a) and (21). In short, the fields anywhere below the current sheet are completely determined by the tangential magnetic field components H_x^- and H_y^- on a plane just below the current sheet. Since this tangential magnetic field is proportional, in turn, to the density of current on the sheet, we conclude that the current sheet is capable of reproducing the field patterns of the arriving scatter radiation at the receiving site, and we deduce that

$$V = 2\Lambda^2 \iint_{-\infty}^{\infty} (E_y H_{xa} - E_x H_{ya})\, dx\, dy \tag{27}$$

is a general expression for the open-circuit voltage at the terminals of the receiving antenna in terms of the arriving magnetic field and the electric field produced by the antenna when it is excited with one ampere.

Using (23) to express the magnetic field components in (27), we have

$$V = 2\Lambda^2 \iint_{-\infty}^{\infty} dx\, dy \iint_{-\infty}^{\infty} (E_y Q_{xa} - E_x Q_{ya})\, e^{j2\pi(x\ell+ym+zn)}\, d\ell\, dm \quad ,$$

and, after reversing the order of integration, we recognize the second integrals as (12). Accordingly,

$$V = 2\Lambda^2 \iint_{-\infty}^{\infty} (P_y Q_{xa} - P_x Q_{ya})\, d\ell\, dm \quad .$$

Using (26) to write the magnetic plane-wave components in terms of the electric plane-wave components, and using (10a), the open-circuit voltage can be written more compactly in terms of electric plane-wave components only:

$$V = \frac{2\Lambda^2}{\eta} \sum_i \iint_{-\infty}^{\infty} n\, P_i(\ell, m)\, P_{ia}(\ell, m)\, d\ell\, dm \quad . \tag{28}$$

A word of caution about (28) is in order. This relation is valid even if the sources of the arriving radiation are near the receiving antenna; it is valid for the current sheet, for example. But when the current in the current sheet is turned off so that the pattern of the receiving antenna can be measured, nothing but free space is left where the sheet had been; the current sheet does not disturb the field pattern of the receiving antenna. A physical source near the receiving antenna, on the other hand, would generally reflect some of the fields from the antenna, which would invalidate (28). In that case, (27) would be valid if E_x and E_y were interpreted as components of the <u>total</u> field on the plane z = constant when the receiving antenna is excited with unit current and the nearby source is turned off (voltage sources short-circuited and current sources open-circuited).

However, we shall assume that the sources of the arriving scatter fields are at great enough distances so that they do not affect the fields produced when the receiving antenna is excited. We shall also assume that the storage-field region of these sources does not extend to the receiving site. The scatter fields, then, will arrive only from the real directions, so that the range of integration in (28) may be reduced to cover only the upper hemisphere:

$$\left.\begin{aligned} V &= \frac{2\Lambda^2}{\eta} \sum_i \iint_{\text{u.h.}} n\, P_i(\ell, m)\, P_{ia}(\ell, m)\, d\ell\, dm \\ &= \frac{2\Lambda^2}{\eta} \sum_k \int_{\text{u.h.}} F_k(\ell, m)\, F_{ka}(\ell, m)\, d\Omega \end{aligned}\right\} \tag{29}$$

from (10b).

* * *

Having developed expressions for the open-circuit voltage at the receiving-antenna terminals due to an arriving scatter signal, (29), and for the total power radiated when the antenna is driven with unit current, (16), we can now write an expression for the time-average signal power available at the antenna terminals. In terms of the unit current and the radiation resistance R_a of the antenna, the radiated power is $W = 1/2\, |I|^2 R_a$, and, neglecting antenna losses, the mean available signal power is

$$\begin{aligned} W_a &= \overline{\frac{|V|^2}{8R_a}} = \overline{\frac{|V|^2}{16W}} = \frac{\Lambda^2}{2\eta} \frac{N}{D} \\ &= \frac{\Lambda^2}{2\eta} \frac{\text{Av.} \left\{ \left| \sum_i \int_{\text{u.h.}} F_i(\ell, m)\, F_{ia}(\ell, m)\, d\Omega \right|^2 \right\}}{\sum_i \int_{\text{u.h.}} |F_i(\ell, m)|^2\, d\Omega} \quad . \end{aligned}$$

Next we manipulate the numerator N of this expression for mean power, making use of the assumption of incoherence discussed in Sec. II. It should be noted that only the arriving quantities F_{Ha} and F_{Va} are time-dependent; the time average will apply to them. Using a second, similar set of directional variables such that

$$d\Omega' = \frac{d\ell'\, dm'}{n'} \quad ,$$

we have

$$N = \sum_i \sum_k \int_{\text{u.h.}} F_i(\ell, m)\, d\Omega \int_{\text{u.h.}} F_k^*(\ell', m')\, \overline{F_{ia}(\ell, m)\, F_{ka}^*(\ell', m')}\, d\Omega' \quad ,$$

or with $\ell' = \ell + \lambda$, $m' = m + \mu$,

$$N = \sum_i \sum_k \int_{u.h.} F_i(t, m)\, d\Omega$$

$$\times \iint F_k^*(t + \lambda,\ m + \mu)\, \overline{F_{ia}(t, m)\, F_{ka}^*(t + \lambda,\ m + \mu)}\, \frac{d\lambda\, d\mu}{n'} \quad .$$

The second (double) integral involves the cross-correlation between arriving plane-wave components. The angular separation between the directions of arrival of the two components is given by the variables of integration λ and μ, which are the differences in the direction cosines. As in Sec. II, we assume that the cross-correlation

$$\overline{F_{ia}(t, m)\, F_{ka}^*(t + \lambda,\ m + \mu)}$$

will be non-zero only for such a small range of directions of arrival (λ, μ) that the antenna pattern will be constant over that range:

$$F_k^*(t + \lambda,\ m + \mu) \approx F_k^*(t, m) \quad .$$

The direction cosine n', which generally varies more slowly with direction than the antenna pattern, will also remain constant over the range of appreciable contribution to the second integral. Finally, the range of integration, which strictly speaking is a displaced unit circle, may be extended without changing the value of the integral. With these changes, the numerator becomes

$$N = \sum_i \sum_k \int_{u.h.} F_i(t, m)\, F_k^*(t, m)\, d\Omega$$

$$\times \frac{1}{n} \iint_{-\infty}^{\infty} \overline{F_{ia}(t, m)\, F_{ka}^*(t + \lambda,\ m + \mu)}\, d\lambda\, d\mu \quad ,$$

and the available power may be written

$$W_a = \frac{\Lambda^2}{2\eta} \frac{N}{D} = \frac{\Lambda^2}{2\eta} \frac{\sum_i \sum_k \int_{u.h.} F_i(t, m)\, F_k^*(t, m)\, \Phi_{ik}(t, m)\, d\Omega}{\sum_i \int_{u.h.} |F_i(t, m)|^2\, d\Omega} \qquad (30)$$

in terms of the quantities

$$\Phi_{ik}(t, m) = \frac{1}{n} \iint_{-\infty}^{\infty} \overline{F_{ia}(t, m)\, F_{ka}^*(t + \lambda,\ m + \mu)}\, d\lambda\, d\mu \quad , \qquad (31)$$

which we recognize from (30) to be proportional to the densities of power arriving from the direction (t, m). The power densities themselves are the quantities,

$$\frac{\Lambda^2}{2\eta}\, \Phi_{ik}(t, m) \quad ,$$

but it will sometimes be convenient to refer to the Φ's as power densities.

Suppose that the antenna emits only horizontally-polarized waves when it is excited, so that the available signal power is

$$W_a = \frac{\Lambda^2}{2\eta} \frac{\int_{u.h.} |F_H(t, m)|^2\, \Phi_{HH}(t, m)\, d\Omega}{\int_{u.h.} |F_H(t, m)|^2\, d\Omega} \quad .$$

Let the power-density pattern have a maximum in the direction $\vec{U}_m$, as shown in Fig. 10. It is then clear that, for maximum power, the antenna pattern should be a thin spike in the direction $\vec{U}_m$, because a change to any other pattern of the same "area" would reduce the numerator integral, leaving the value of the denominator integral unchanged. Of course, such an optimization would require unreasonably large antenna-aperture dimensions. We will discover that, in general, if the aperture is specified, the optimum antenna has almost, but not quite, maximum plane-wave gain; its pattern will be clustered about the direction of maximum power density.

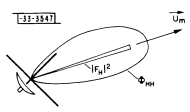

Fig. 10. The available power is maximized by the antenna having a pattern that is sharply peaked in the direction of maximum power density.

The case of antenna noise is also of interest. If "hot spots" in the sky are neglected, antenna noise may be considered to arise from a uniform distribution over the upper hemisphere of incoherent sources of all polarizations, such that the power densities are

$$\Phi_{HH}(t, m) = \Phi_{VV}(t, m) = \Phi_o \quad , \quad \text{a constant}$$

and

$$\Phi_{HV}(t, m) = \Phi_{VH}(t, m) = 0 \quad .$$

The available noise power is, from (30),

$$W_a = \frac{\Lambda^2}{2\eta} \frac{\int_{u.h.} (|F_H|^2 + |F_V|^2)\, \Phi_o\, d\Omega}{\int_{u.h.} (|F_H|^2 + |F_V|^2)\, d\Omega} = \frac{\Lambda^2}{2\eta}\, \Phi_o \quad ,$$

which is independent of the antenna pattern. Apparently, the antenna pattern that maximizes the mean signal power will also maximize the signal-to-noise ratio, since the noise power will not be affected by the pattern adjustment.

We have finally arrived at the core of the problem. From the point of view of the antenna designer, the power densities are necessary and sufficient data for the design of an optimum antenna. His job is to produce an antenna having a pattern, determined by the $F_i(t, m)$, which maximizes the available power, (30), subject to appropriate restrictions. The proprietor of the receiving site, on the other hand, will describe the power densities to the designer and specify some of the restrictions. Section V will be given to a discussion of the power densities and suggestions for their measurement; the following sections will deal with some of the antenna designer's problems.

V. DETERMINATION OF POWER DENSITY

It should be emphasized that four power densities, given by (31) in terms of the arriving plane-wave components, were required to express the mean signal power available at the receiving-antenna terminals, (30). Since the numerator of (30), being a squared magnitude, is a real, non-negative number for any antenna pattern (for any set of F_i's), it must be that

$$\Phi_{HV}(\ell, m) = \Phi^*_{VH}(\ell, m) \quad ,$$

that the self-densities Φ_{HH} and Φ_{VV} are real non-negative quantities, and that the matrix of power densities

$$[\Phi(\ell, m)] = \begin{bmatrix} \Phi_{HH}(\ell, m) & \Phi_{HV}(\ell, m) \\ \Phi_{VH}(\ell, m) & \Phi_{VV}(\ell, m) \end{bmatrix} \tag{32}$$

is Hermitian, with a non-negative determinant. For a particular direction, (ℓ, m), the determinant becomes zero if and only if the polarization of the plane wave arriving from (ℓ, m) remains fixed in time as the signal fades. In this special case, the incoming radiation could be described in terms of two power densities but, in general, three are required; it will be convenient to retain all four to preserve symmetry. If the available power had been written in terms of the rectangular plane-wave components, nine rectangular power densities would have resulted; these could, however, have been reduced to the above four with the aid of the relations (10).

It is the purpose of this section, then, to suggest means for determining the four power-density patterns: the four average quantities that describe the arriving scatter radiation. At the outset, it should be noted that, since the recorded strengths of received scatter signals generally show seasonal as well as diurnal trends, certainly the amplitudes and very possibly also the shapes of the power-density patterns may show corresponding long-term variations. If the patterns are to be used in the design of an optimum fixed receiving antenna that is expected to deliver maximum mean signal power for its useful life, the power densities should be averages that are taken over periods of time long compared with the periods of variation of the received signal strength. Averaging periods at least of the order of several years are indicated!

On the other hand, the proprietor of the receiving site might want instead an antenna that delivers maximum mean power during periods of weak received signal. In that case, he should average the power densities only during such periods. Of course, the power-density patterns characteristic of weak-signal periods might show more statistical regularity, so that shorter averaging times could be employed. Nevertheless, a program of measurement of the power densities is likely to be a very long one. It may be desirable instead to estimate the power-density patterns from one of the theories of the scattering mechanism and a knowledge of the transmitting-antenna pattern.

* * *

Throughout the remainder of this section we will suppose that the power densities are to be measured. Questions as to when and for how long to average are left to the judgment of the proprietor of the receiving site; our object will be to suggest methods of measurement.

One of the rectangular power densities, which we will call

$$\psi_{aik}(\ell, m) = \int_{-\infty}^{\infty}\int \overline{P_{ia}(\ell, m)\, P_{ka}^{*}(\ell + \lambda,\ m + \mu)}\, d\lambda\, d\mu \quad , \tag{33}$$

can be expressed in terms of the arriving fields as follows:

$$\psi_{aik}(\ell, m) = \overline{P_{ia}(\ell, m)\, E_{ka}^{*}(0, 0, 0)} \quad ,$$

$$= \int_{-\infty}^{\infty}\int \phi_{aik}(x, y)\, e^{-j2\pi(x\ell + ym)}\, dx\, dy \quad , \tag{34}$$

in which

$$\phi_{aik}(x, y) = \overline{E_{ia}(x, y, 0)\, E_{ka}^{*}(0, 0, 0)} \quad . \tag{35}$$

[We have used (21) and (22).] A method of measurement suggests itself immediately: Voltages proportional to the fields at two points on the xy-plane would be available at the terminals of a pair of appropriately placed dipoles. The Fourier transform of (35), the temporal cross-correlation of the fields, is the rectangular power density. The cross-correlation should be measured with one dipole located at the origin and the other located, in turn, at each of a sufficient number of points $(x, y, 0)$ on the xy-plane to determine the behavior of $\phi_{aik}(x, y)$ throughout the plane.

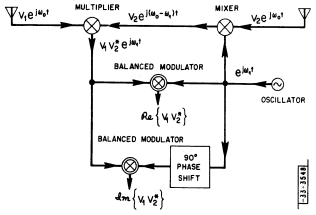

Fig. 11. The radio-frequency portion of a cross-correlator.

Figure 11 is a block diagram showing the essential parts of the circuitry that might be used to cross-correlate the dipole voltages V_1 and V_2; the simpler circuit of Fig. 12 could be used to obtain $\phi_{aii}(0, 0)$, the "cross-correlation" for identically oriented dipoles at the same point in space. A radian carrier frequency ω_0 is assumed for the arriving signal, and a local oscillator of radian frequency $\omega_1 < \omega_0$ is shown in Fig. 11. It is assumed that the output circuits of the

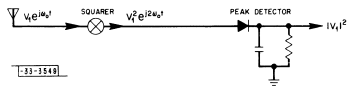

Fig. 12. The auto-correlator is just a watt-hour meter.

four multiplying elements shown in Fig. 11 include filters that reject all but the difference-frequency component of the inputs. The balanced-modulator outputs are the direct ("d-c") voltages,

$$\mathcal{R}e\{V_1 V_2^{*}\} \quad \text{and} \quad \mathcal{I}m\{V_1 V_2^{*}\} \quad ,$$

which could be averaged, for example, by Miller integrators[10] and recorded at appropriate intervals. In terms of these averages, the cross-correlation of the fields is

$$\phi_{aik}(x, y) \propto \mathcal{R}e\{\overline{V_1 V_2^{*}}\} + j\, \mathcal{I}m\{\overline{V_1 V_2^{*}}\} \quad ,$$

the proportionality constant being determined by the gain of the system. Again in Fig. 12 it is assumed that the output circuit of the squarer includes a filter to reject all but the second harmonic. The peak detector output $|V_1|^2$ should be similarly averaged to obtain a quantity proportional to $\phi_{aii}(0, 0)$.

Only four rectangular power densities need be measured. For purposes of illustration, we will work with the four rectangular densities that would be measured with x- and y-oriented dipoles, although this choice is quite arbitrary. The "horizontal- and vertical-power densities," (32), can then be obtained from the measured ψ_{aik}'s with the aid of the relations (10). The necessary algebra, however, is somewhat tedious. At the risk of clouding the main issues, we will introduce matrix notation to simplify the algebraic manipulations. Suppose then, that we have the elements of the matrix

$$[\psi_a(\ell, m)] = \begin{bmatrix} \psi_{axx}(\ell, m) & \psi_{axy}(\ell, m) \\ \psi_{ayx}(\ell, m) & \psi_{ayy}(\ell, m) \end{bmatrix} \tag{36}$$

of rectangular power densities, measured with x- and y-oriented dipoles only. Define the following column matrices:

$$P_a(\ell, m)] = \begin{bmatrix} P_{xa}(\ell, m) \\ P_{xa}(\ell, m) \end{bmatrix} \quad ; \quad F_a(\ell, m)] = \begin{bmatrix} F_{Ha}(\ell, m) \\ F_{Va}(\ell, m) \end{bmatrix} \quad . \tag{37}$$

The relation between P] and F] is given by the transformation matrix that has the coefficients of the first two of equations (10) as elements:

$$[T(\ell, m)] = \begin{bmatrix} \dfrac{-m}{n\sqrt{\ell^2+m^2}} & \dfrac{-\ell}{\sqrt{\ell^2+m^2}} \\ \dfrac{\ell}{n\sqrt{\ell^2+m^2}} & \dfrac{-m}{\sqrt{\ell^2+m^2}} \end{bmatrix} . \tag{38}$$

Note that the determinant of the T-matrix is

$$|\,[T(\ell, m)]\,| = \frac{1}{n} ,$$

so that [T] may be inverted freely except possibly for the directions $\ell^2 + m^2 = 1$, corresponding to the horizon. According to (10), the relation between P] and F] is

$$P] = [T] \times F] \quad ; \quad F] = [T]^{-1} \times P] . \tag{39}$$

We will write the transpose of F] as

$$\underline{F} = \underline{P} \times [T]_t^{-1} ,$$

reserving the star to mean that the complex conjugate is to be taken.

The matrix of power densities, (32), may now be written

$$[\Phi(\ell, m)] = \frac{1}{n} \int_{-\infty}^{\infty}\int \overline{F_a(\ell, m)] \times \underline{F_a^*(\ell + \lambda,\ m + \mu)}}\ d\lambda\, d\mu , \tag{31a}$$

$$= \frac{1}{n} \int_{-\infty}^{\infty}\int [T(\ell, m)]^{-1} \times \overline{P_a(\ell, m)] \times \underline{P_a^*(\ell + \lambda,\ m + \mu)}} \times [T(\ell + \lambda,\ m + \mu)]_t^{-1}\, d\lambda\, d\mu .$$

The elements of [T] will remain essentially constant over the range of appreciable contribution to the integral, so the T-matrices may be removed from the integral. We recognize the remaining integral as $[\psi_a(\ell, m)]$, giving the result

$$[\Phi(\ell, m)] = \frac{1}{n}\,[T(\ell, m)]^{-1} \times [\psi_a(\ell, m)] \times [T(\ell, m)]_t^{-1} . \tag{40}$$

To recapitulate, the cross-correlations of x- and y-fields on the xy-plane (35) are measured, these four quantities are transformed (34), giving the elements of the ψ_a-matrix (36), which, when operated on by the T-matrix (38), gives the matrix of power densities (40).

But there is a fundamental difficulty with this procedure for measuring the power densities: We have tacitly assumed that the total field at each of the measuring dipoles is the incident field arriving from the scattering region; no fields reflected from below the xy-plane are assumed

present. In other words, we have assumed that the region below the xy-plane is free space. In some cases, this assumption may be reasonable, but usually an antenna will be erected over a more or less level ground, which must be taken into consideration in measuring the arriving power densities.

* * *

If the ground at the receiving site is smooth enough, if irregularities in the terrain are much smaller than a wavelength, the reflected fields are easily calculated. In this case, the foregoing measurement procedure may be re-interpreted to account for the presence of the ground. We will suppose that the xy-plane is parallel to the ground and at a distance h wavelengths above the ground. Cross-correlations may be measured as before on the xy-plane or on some other plane that is tilted with respect to the ground. We will start by interpreting cross-correlation measurements on the xy-plane over a lossy, but otherwise smooth, homogeneous ground.

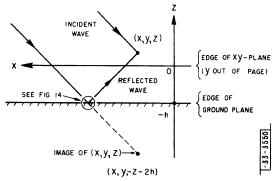

Fig. 13. The reflected wave travels a greater distance than the incident wave.

Receiving-site geometry is shown in Fig. 13. It may be seen that the reflected waves arriving at the point (x, y, z) above the ground travel a greater distance than the corresponding incident waves. This extra distance is just accounted for by calculating the phase of a reflected wave as though it were not reflected, but continued instead, in the absence of the ground, to the image of the point (x, y, z).

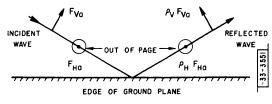

Fig. 14. The incident and reflected waves very close to the ground.

On being reflected, the wave will suffer an additional phase shift and attenuation, which will depend on the angle of incidence and the polarization. This effect may be accounted for by assigning reflection coefficients $\rho_H(\ell, m)$ and $\rho_V(\ell, m)$ to the reflected horizontal and vertical plane-wave components. Sign conventions for the reflection coefficients are shown in Fig. 14. Note that the reflected horizontal component $\rho_H F_{Ha}$ contributes to the x- and y-fields with the same sign as the incident horizontal component; the reflected vertical component $\rho_V F_{Va}$, however, contributes to the tangential fields with the opposite sign from the incident vertical component. If the ground were a perfect conductor, $\rho_H = -1$ and $\rho_V = +1$, so that both reflected components would contribute to the tangential fields with opposite signs from the incident components.

In view of the foregoing discussion, and using (21) and (10), the tangential components of the total arriving field in the presence of the ground may be written

$$E_{-g}(x, y, z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_{-g}(\ell, m, z)\, e^{j2\pi(x\ell+ym-zn)}\, d\ell\, dm$$

with the definitions

$$\left.\begin{aligned} P_{xg}(\ell, m, z) &= -\frac{m}{n\sqrt{\ell^2+m^2}}\left\{1+\rho_H(\ell, m)\, e^{-j4\pi(z+h)n}\right\} F_{Ha}(\ell, m) \\ &\quad -\frac{\ell}{\sqrt{\ell^2+m^2}}\left\{1-\rho_V(\ell, m)\, e^{-j4\pi(z+h)n}\right\} F_{Va}(\ell, m) \\ P_{yg}(\ell, m, z) &= \frac{\ell}{n\sqrt{\ell^2+m^2}}\left\{1+\rho_H(\ell, m)\, e^{-j4\pi(z+h)n}\right\} F_{Ha}(\ell, m) \\ &\quad -\frac{m}{\sqrt{\ell^2+m^2}}\left\{1-\rho_V(\ell, m)\, e^{-j4\pi(z+h)n}\right\} F_{Va}(\ell, m) \end{aligned}\right\} \quad . \tag{41}$$

On the xy-plane, which is parallel to the ground, (41) may be written in matrix notation as

$$E_p(x, y)] = \begin{bmatrix} E_{xg}(x, y, 0) \\ E_{yg}(x, y, 0) \end{bmatrix} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [T_p(\ell, m)] \times F_a(\ell, m)]\, e^{j2\pi(x\ell+ym)}\, d\ell\, dm \tag{42}$$

in which

$$[T_p(\ell, m)] = [T(\ell, m)] \times \{[1] + e^{-j4\pi hn}\, [\rho(\ell, m)]\}$$

$$= \begin{bmatrix} -\frac{m}{n\sqrt{\ell^2+m^2}}\left\{1+\rho_H\, e^{-j4\pi hn}\right\} & -\frac{\ell}{\sqrt{\ell^2+m^2}}\left\{1-\rho_V\, e^{-j4\pi hn}\right\} \\ \frac{\ell}{n\sqrt{\ell^2+m^2}}\left\{1+\rho_H\, e^{-j4\pi hn}\right\} & -\frac{m}{\sqrt{\ell^2+m^2}}\left\{1-\rho_V\, e^{-j4\pi hn}\right\} \end{bmatrix} \tag{43}$$

and

$$[1] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad ; \quad [\rho(\ell, m)] = \begin{bmatrix} \rho_H(\ell, m) & 0 \\ 0 & -\rho_V(\ell, m) \end{bmatrix} \quad . \tag{44}$$

We suppose that the elements of the matrix,

$$[\phi_p(x, y)] = \overline{E_p(x, y)] \times \underbrace{E_p^*(0, 0)}} \quad , \tag{45}$$

of cross-correlations of x- and y-fields on the xy-plane have been measured, and we wish to determine the power densities from them. From (42),

$$[\phi_p(x, y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [T_p(\ell, m)]\, e^{j2\pi(x\ell+ym)}\, d\ell\, dm$$

$$\times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \overline{F_a(\ell, m)] \times \underbrace{F_a^*(\ell', m')}} \times [T_p^*(\ell', m')]_t\, d\ell'\, dm' \quad .$$

If the height h of the measurement plane is not too great, $[T_p^*]_t$ will remain essentially constant over the range of appreciable contribution to the second integral; it may be removed from the integral, which is then recognized as $n[\Phi(\ell, m)]$. Consolidating,

$$[\phi_p(x, y)] \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\psi_p(\ell, m)]\, e^{j2\pi(x\ell+ym)}\, d\ell\, dm \quad ,$$

from which

$$[\psi_p(\ell, m)] = n[T_p(\ell, m)] \times [\Phi(\ell, m)] \times [T_p^*(\ell, m)]_t$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\phi_p(x, y)]\, e^{-j2\pi(x\ell+ym)}\, dx\, dy \tag{46}$$

and

$$[\Phi(\ell, m)] = \frac{1}{n}\, [T_p(\ell, m)]^{-1} \times [\psi_p(\ell, m)] \times [T_p^*(\ell, m)]_t^{-1} \quad . \tag{47}$$

The result has the same form as (40), in which the ground was assumed absent: The cross-correlation matrix, (45), is measured, transformed (46), and operated on by the T_p-matrix, (43), to give the result, (47). It may be noted that, if the ground is a perfect conductor, (47) reduces to

$$[\Phi(\ell, m)] = \frac{1}{n\left|1 - e^{-j4\pi hn}\right|^2} [T(\ell, m)]^{-1} \times [\psi_p(\ell, m)] \times [T(\ell, m)]_t^{-1} \quad . \tag{47a}$$

Given cross-correlation measurements on the xy-plane but no information about the surroundings of the plane or its orientation with respect to the scattering region, we are able to determine the direction cosines ℓ and m of the power density impinging on the plane, but we are not able to determine the sign of the direction cosine n of the arriving power density. In other words, we are not able to distinguish between power density arriving from a point above the measurement plane and power density arriving from the image of that point below the measurement plane. We were able to interpret measurements on a plane parallel to the ground because we knew that the scattering region must be above the plane and that plane waves arriving from below the plane must be reflected copies of the incident plane waves. If the measurement plane were tilted with respect to the ground, however, it would be possible for incident radiation to arrive from both sides of the plane. Nevertheless, it will generally be possible to orient a tilted measurement plane with respect to the scattering region in such a way that we are reasonably sure that the incident plane waves arrive only from one side of the plane.

Next we will interpret cross-correlation measurements on a vertical plane above the ground: the xz-plane of Fig. 13. We suppose that the positive y-axis is oriented in the general direction of the transmitter, so that all incident radiation arrives with a positive direction cosine m. We will obtain a two-lobed power-density pattern: One lobe will extend above the ground in front of the xz-plane, and the other lobe, below the ground, will be the somewhat attenuated image of the first because of ground reflection. At first glance, it would appear that these two lobes, taken together, could be delivered to the antenna designer as the power-density pattern for the design of an optimum antenna to be erected at the origin of coordinates in Fig. 13. This, however, would be a mistake! A plane wave arriving from above the xy-plane is perfectly correlated with the corresponding plane wave that is reflected from the ground and arrives from the image direction below the xy-plane, so that the assumption of incoherence, discussed in Sec. II, is not valid for the superposition of incident and reflected plane waves. As we shall see, the two lobes will be scalloped by alternate constructive and destructive interference between incident and reflected waves. The angular locations of the scallops will depend critically on the height h of the reference dipole used in the measurement. In contrast, if any two arriving plane-wave components were uncorrelated unless they arrived from nearly coincident directions, as required by the assumption of incoherence, the power densities would be, for all practical purposes, independent of the location of the reference dipole used in measuring them.

The proprietor of the receiving site should measure the densities of incident power and deliver these, together with the ground-reflection coefficients, to the antenna designer for his consideration.

Returning to the problem at hand, we may write the column matrix of x- and y-field components on the xz-plane, which is normal to the ground as

$$[E_n(x, z)] = \begin{bmatrix} E_{xg}(x, 0, z) \\ E_{yg}(x, 0, z) \end{bmatrix}$$

$$= \iint_{-\infty}^{\infty} [T(\ell, m)] \times \{[1] + e^{-j4\pi(z+h)n} [\rho(\ell, m)]\} \times F_a(\ell, m)] \, e^{j2\pi(x\ell+zn)} \, d\ell \, dm \tag{48}$$

from (38), (41) and (44). The cross-correlation matrix to be measured is

$$[\phi_n(x, z)] = \overline{E_n(x, z)] \times \underline{E_n^*(0, 0)}} \quad . \tag{49}$$

In writing an expression for this matrix in terms of the power densities, it will be convenient to convert one of the ingredients (48) into an integral over the direction cosines ℓ and $\underline{n}$, m being determined as the positive square root of $1 - \ell^2 - n^2$:

$$d\ell \, dm = n d\Omega = \frac{n}{m} \, d\ell \, dn \quad .$$

In this way, (49) becomes

$$[\phi_n(x, z)] = \iint_{-\infty}^{\infty} [T(\ell, m)] \times \{[1] + e^{-j4\pi(z+h)n} [\rho(\ell, m)]\} \, e^{j2\pi(x\ell+zn)} \, \frac{n}{m} \, d\ell \, dn$$

$$\times \iint_{-\infty}^{\infty} \overline{F_a(\ell, m)] \times \underline{F_a^*(\ell', m')}} \times \{[1] + e^{j4\pi hn'} [\rho^*(\ell', m')]\} \times [T(\ell', m')]_t \, d\ell' \, dm' \quad .$$

As before, if the height h of the reference dipole is not too great, the last two factors may be removed from the second integral, which is then recognized as $n[\Phi]$. The result,

$$[\phi_n(x, z)] = \iint_{-\infty}^{\infty} \frac{n^2}{m} \, e^{j2\pi(x\ell+zn)} \, d\ell \, dn$$

$$\times [T] \times \{[1] + e^{-j4\pi(z+h)n} [\rho]\} \times [\Phi] \times \{[1] + e^{j4\pi hn} [\rho^*]\} \times [T]_t \quad ,$$

however, is not in the desirable form of a Fourier transform because of the function of z, $e^{-j4\pi(z+h)n}$, appearing in the integrand. This difficulty may be remedied by splitting $[\phi_n]$ into the sum of two integrals:

$$[\phi_n(x, z)] = \iint_{-\infty}^{\infty} \frac{n^2}{m} \, e^{j2\pi(x\ell+zn)} \, d\ell \, dn \times [T] \times [\Phi] \times \{[1] + e^{j4\pi hn} [\rho^*]\} \times [T]_t$$

$$+ \iint_{-\infty}^{\infty} \frac{n^2}{m} \, e^{j2\pi(x\ell-zn)} \, d\ell \, dn \times e^{-j4\pi hn} [T] \times [\rho] \times [\Phi] \times \{[1] + e^{j4\pi hn} [\rho^*]\} \times [T]_t$$

and changing the sign of the variable n in the second integral. The effect of this change will be emphasized by writing [T], [Φ] and [ρ] as functions of n, which, of course, is only part of the truth, since [T] and [Φ] depend on two direction cosines. With the change of sign, $[\phi_n]$ can be written in terms of its Fourier transform:

$$[\phi_n(x,z)] \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{[\psi^+(n)] + [\psi^-(n)]\}\, e^{j2\pi(x\ell+zn)}\, d\ell\, dn \quad ,$$

in which

$$[\psi^+(n)] = \frac{n^2}{m}\,[T(n)] \times [\Phi(n)] \times \{[1] + e^{j4\pi hn}\,[\rho^*(n)]\} \times [T(n)]_t$$

and

$$[\psi^-(n)] = \frac{n^2}{m}\, e^{j4\pi hn}\,[T(-n)] \times [\rho(-n)] \times [\Phi(-n)] \times \{[1] + e^{-j4\pi hn}\,[\rho^*(-n)]\} \times [T(-n)]_t \qquad (50)$$

These two ψ-matrices describe the two lobes mentioned in the preliminary discussion. The presence of the matrix,

$$\{[1] + e^{j4\pi hn}\,[\rho^*(n)]\} \quad ,$$

causes the scalloping of the upper lobe and the corresponding matrix in the expression for $[\psi^-]$ causes the scalloping of the lower lobe.

It is important to note that, because of the presence of the ground, the matrix of arriving power densities, [Φ(n)], will be zero for directions $n < 0$, below the horizon. Similarly, [Φ(−n)] will be zero for $n > 0$. In other words, the upper lobe is zero below the horizon, and the lower lobe is zero above the horizon. These facts allow the two lobes to be separated. Thus, from the measured matrix of cross-correlations of x- and y-fields on the xz-plane, (49), we may obtain the upper lobe:

$$[\psi^+] = \begin{cases} \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\phi_n(x,z)]\, e^{-j2\pi(x\ell+zn)}\, dx\, dz \quad , & n > 0 \\ 0 \quad , & n < 0 \quad . \end{cases} \qquad (51)$$

According to (50), the desired matrix of arriving power densities may then be calculated from

$$[\Phi] = \frac{m}{n^2}\,[T]^{-1} \times [\psi^+] \times [T]_t^{-1} \times \{[1] + e^{j4\pi hn}\,[\rho^*]\}^{-1} \quad . \qquad (52)$$

* * *

If the distance between transmitting and receiving sites is very great, the signals received on the measurement dipoles may be so weak and corrupted by noise that it may be desirable to make the measurements instead with more directive antennas to improve the signal-to-noise ratio. In this connection, it should be noted that the long averaging times contemplated for the cross-correlation of antenna voltages will not be effective in removing that part of the "noise" in the measured cross-correlation which is due to coherent noises received on the two measuring antennas. Suppose that the input to the circuit of Fig. 11 from the left-hand antenna is

$$V_1 + (N_1 + k_1 N_o)$$

and the input from the right-hand antenna is

$$V_2 + (N_2 + k_2 N_o) \quad .$$

V_1 and V_2 are the two signal voltages, $(N_1 + k_1 N_o)$ and $(N_2 + k_2 N_o)$ are the two noise voltages, and k_1 and k_2 are constants. With the circuit of Fig. 11 and associated integrators we will measure

$$\overline{\{V_1 + (N_1 + k_1 N_o)\}\,\{V_2^* + (N_2^* + k_2^* N_o^*)\}} = \overline{V_1 V_2^*} + \overline{(N_1 + k_1 N_o)\,(N_2^* + k_2^* N_o^*)} \quad ,$$

since the signals and noises will be uncorrelated. The noise components, N_o, N_1 and N_2 are assumed uncorrelated; N_o is that part of the noise which is common to both antennas. Note that the measured cross-correlation, which reduces to

$$\overline{V_1 V_2^*} + k_1 k_2^*\, \overline{|N_o|^2} \quad ,$$

includes a term due to this common noise.

Another way of looking at the noise problem is that we contemplate using a circuit like that of Fig. 11 to measure the density of power arriving at the receiving site. In addition to the signal power, we expect a more or less uniform distribution of noise power density, as discussed at the end of Sec. IV. Since signal and noise are incoherent, their power densities add, and we should expect to measure the sum of these two power densities with equipment appropriate for the measurement of power density. The advantage in using directive antennas for the measurement is that they would capture less of the total noise power, but essentially all of the signal power. To a first and fairly good approximation, a constant power density Φ_o may be subtracted from the measured self-densities, $\Phi_{HH}(\ell, m)$ and $\Phi_{VV}(\ell, m)$ to account for the noise.

The measurement may be improved considerably by transmitting an unmodulated carrier and using sharply tuned circuits to remove all but a narrow band of noise. The tuned circuits should precede the multiplying elements in Figs. 11 and 12.

It may also be desirable to make measurements with directive antennas for another reason: The ground at the receiving site may be so irregular that a calculation of the fields reflected from it would be impractical. This difficulty may be overcome by measuring the power densities with antennas that are directive enough to reject, substantially, the radiation reflected from the ground. As before, a measurement of cross-correlations of antenna terminal voltages will be required; one antenna will be located at the origin of coordinates and another should be

located in turn at a number of points on the xy-plane. Since we will assume that neither antenna is sensitive to fields reflected from the ground, considerable freedom in the choice of orientation of the measurement plane (the xy-plane) is left to the proprietor of the receiving site. This freedom will be restricted only by the requirement that all the incident radiation should arrive from above the xy-plane: with positive direction cosine n.

In Sec. IV, we developed an expression, (29), for the open-circuit voltage at the terminals of a receiving antenna due to an arriving scatter signal. The reference point on the antenna to which the phase of its pattern is referred was assumed to be located at the origin of coordinates. If we had assumed, instead, that the antenna is located at the point (x, y, z), i.e., the reference point on the antenna is at (x, y, z), we would have discovered that its terminal voltage can be expressed as

$$V(x,y,z) = \frac{2\Lambda^2}{\eta} \sum_i \int_{u.h.} F_i(\ell,m)\, F_{ia}(\ell,m)\, e^{j2\pi(x\ell+ym+zn)}\, d\Omega \quad . \tag{53}$$

The phases of the arriving plane-wave components are shifted by the same factor as in (21), the expression for the arriving fields. Note that, in moving this antenna from the origin to its new location, we keep it pointed in the same direction, and we do not rotate it, so that its polarization remains the same.

In the dipole measurements we used x- and y-oriented dipoles; in the present measurements with directive antennas we will also need two kinds of antennas, which will have terminal voltages $V_1(x,y,z)$ and $V_2(x,y,z)$, given by (53). Antenna 1 might be a small dish with horizontally polarized feed, and antenna 2 might be the same dish with a vertically polarized feed, for example. In place of the T-matrix, we will need the square matrix,

$$[F(\ell,m)] = \begin{bmatrix} F_{H1}(\ell,m) & F_{H2}(\ell,m) \\ F_{V1}(\ell,m) & F_{V2}(\ell,m) \end{bmatrix} , \tag{54}$$

of horizontal and vertical plane-wave components emitted by the two antennas when they are excited with unit current. When the antennas are located on the xy-plane, the column matrix of their open-circuit voltages is

$$V(x,y)] = \begin{bmatrix} V_1(x,y,0) \\ V_2(x,y,0) \end{bmatrix} = \frac{2\Lambda^2}{\eta} \int_{u.h.} [F(\ell,m)]_t \times F_a(\ell,m)]\, e^{j2\pi(x\ell+ym)}\, d\Omega \quad , \tag{55}$$

and the cross-correlation matrix to be measured is

$$[\phi(x,y)] = \overline{V(x,y)] \times \underline{V^*(0,0)}}$$

$$= \left(\frac{2\Lambda^2}{\eta}\right)^2 \int_{u.h.} [F(\ell,m)]_t\, e^{j2\pi(x\ell+ym)}\, d\Omega$$

$$\times \int_{u.h.} \overline{F_a(\ell,m)] \times \underline{F_a^*(\ell',m')}} \times [F^*(\ell',m')]\, d\Omega' \quad .$$

The range of both integrals may be extended, and the matrix of antenna patterns, $[F^*]$, may be removed from the second integral, which is then recognized as $[\Phi(\ell,m)]$. The result of these changes is

$$[\phi(x,y)] \equiv \int_{-\infty}^{\infty}\int [\psi(\ell,m)]\, e^{j2\pi(x\ell+ym)}\, d\ell\, dm \quad ,$$

in which

$$[\psi(\ell,m)] = \left(\frac{2\Lambda^2}{\eta}\right)^2 \frac{1}{n}\, [F(\ell,m)]_t \times [\Phi(\ell,m)] \times [F^*(\ell,m)]$$

$$= \int_{-\infty}^{\infty}\int [\phi(x,y)]\, e^{-j2\pi(x\ell+ym)}\, dx\, dy \quad . \tag{57}$$

The matrix of measured cross-correlations, (56), is transformed, (57), and operated on by the matrix of antenna patterns, (54), to give the matrix of power densities:

$$[\Phi(\ell,m)] = n\left(\frac{\eta}{2\Lambda^2}\right)^2 [F(\ell,m)]_t^{-1} \times [\psi(\ell,m)] \times [F^*(\ell,m)]^{-1} \quad . \tag{58}$$

* * *

We have suggested a number of methods for measuring the power densities, all of which involve measuring the cross-correlations of antenna voltages on some plane, transforming, and performing algebraic operations on the transforms. It is time to take stock of the results to see which parts of these procedures will present real difficulties and, if possible, to suggest short cuts. The algebraic operations, being relatively simple and straightforward, may be dismissed immediately.

The measurement of the cross-correlations will be the most difficult operation. Making a measurement at each point on the measurement plane, leaving the movable antenna fixed at each point for enough time to take an average, would be physically impossible. Even measuring at a sufficient number of points to determine the behavior of the cross-correlation throughout the plane will be a time-consuming operation. It will be advantageous to reduce the number of measurement points to the bare minimum.

With the exception of the measurement on a plane perpendicular to the ground, which will be discussed separately, all the results, (40), (47) and (58), have been of the form

$$[\Phi(\ell, m)] = [A(\ell, m)] \times [\psi(\ell, m)] \times [A^*(\ell, m)]_t \quad ,$$

in which $[\Phi]$ is the matrix of arriving power densities, $[\psi]$ is the matrix of the transforms of the cross-correlations – a matrix of rectangular power densities – and $[A]$ is the appropriate transformation matrix. $[\Phi]$ is the matrix of a non-negative definite Hermitian form, as we discovered at the beginning of this section. Because of the symmetry of its expression in terms of $[\psi]$, it follows that $[\psi]$ is also the matrix of a non-negative definite Hermitian form. In particular,

$$\psi_{ik}(\ell, m) = \psi^*_{ki}(\ell, m) \quad ,$$

so that the ψ_{ii}'s are real. The cross-correlations, which are the Fourier transforms of the ψ's, must therefore obey

$$\phi_{ik}(x, y) = \phi^*_{ki}(-x, -y) \quad ,$$

so that the self terms are conjugate symmetric:

$$\phi_{ii}(x, y) = \phi^*_{ii}(-x, -y) \quad .$$

By making use of these symmetries, it will be possible to obtain the four power densities from a measurement of the four cross-correlations over only half of the xy-plane – an appreciable saving in effort.

This saving is not possible for the measurement on a plane normal to the ground as we have developed it, because the matrix of transforms of the cross-correlations, $[\psi^+(n)] + [\psi^-(n)]$, is not Hermitian, as can be seen from (50). The extra penalty attached to this measurement may be attributed to the fact that it is also a measurement of the ground-reflection coefficients: If $[\Phi]$ is eliminated from the equations (50), there results

$$[\rho(n)] = e^{j4\pi hn} [T(n)]^{-1} \times [\psi^-(-n)] \times [\psi^+(n)]^{-1} \times [T(n)] \quad .$$

Although this feature may be desirable, one could certainly devise simpler experiments for the measurement of the reflection coefficients.

In some cases, it may be reasonable to assume another kind of symmetry among the ψ's, due to symmetry about the plane containing the great circle path from transmitter to receiving site. If the y-axis lies in this great-circle plane, the cross-correlations will show some kind of symmetry about the y-axis, depending on the polarizations of both the transmitting antenna and the measuring antennas. In these cases, which will not be enumerated, it will be possible to obtain the power densities from a measurement of the cross-correlations over a quarter of the measurement plane except where the measurement is on a plane perpendicular to the ground. In that case, the yz-plane should coincide with the great-circle plane, the symmetries will be about the z-axis, and a measurement over half of the xz-plane will suffice.

It is a fundamental property of Fourier transforms that the fine structure of a function is determined by the gross behavior – the extent – of its transform, and conversely. Thus, the

required extent on the measurement plane of the cross-correlation measurements would depend on the amount of fine structure in the power densities if the object of the experiment were to determine the power densities as accurately as possible. But if the measured power densities are to be used in the design of an antenna, we should like to know how much of the fine structure is necessary for the design; perhaps some could be neglected with a resulting saving in measurement effort. As will be shown in the next sections, this question is best answered directly in terms of the necessary extent of the cross-correlation measurements on the measuring plane: It will be necessary to know the cross-correlation between arriving fields at any two points on and possibly slightly beyond the aperture of the antenna; if the antenna is to have a large aperture, the measurements should extend over a correspondingly large area.

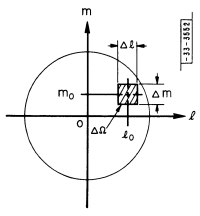

Fig. 15. Measurement effort is reduced if the power densities arrive from within a narrow range of directions, $\Delta\Omega$.

On the other hand, the fine structure of the cross-correlations, which will dictate the allowable separation of the measurement points, is determined by the angular extent of the power densities. If it may be assumed that all the plane-wave components of the signal arrive from within the rectangular range of directions $\Delta\Omega$, shown in Fig. 15, the familiar Sampling Theorem will suggest an appropriate spacing of measurement points and perform the Fourier transform of the cross-correlations measured at these points.

One of the rectangular power densities, which will be zero outside of $\Delta\Omega$, may be represented by a Fourier series inside $\Delta\Omega$. In fact, the product of this rectangular density with an arbitrary function of (ℓ, m) may be represented by a Fourier series inside $\Delta\Omega$:

$$\psi_{ik}(\ell, m)\, e^{j2\pi[x_o(\ell-\ell_o)+y_o(m-m_o)]} = \sum_{r'} \sum_{s'} C_{r's'}\, e^{-j2\pi[r'(\ell-\ell_o)/\Delta\ell + s'(m-m_o)/\Delta m]}$$

so that the rectangular power density itself is

$$\psi_{ik}(\ell, m) = \sum_{r', s'} C_{r's'}\, e^{-j2\pi[(x_o+r'/\Delta\ell)(\ell-\ell_o)+(y_o+s'/\Delta m)(m-m_o)]} \quad .$$

The coefficients $C_{r's'}$ and the constants x_o and y_o are as yet undetermined. The cross-correlation $\phi_{ik}(x, y)$ is the transform of $\psi_{ik}(\ell, m)$:

$$\phi_{ik}(x, y) = \iint_{\Delta\Omega} \psi_{ik}(\ell, m)\, e^{j2\pi(x\ell+ym)}\, d\ell\, dm$$

$$= \Delta\ell\, \Delta m\, e^{j2\pi(x\ell_o+ym_o)} \sum_{r', s'} C_{r's'} \frac{\sin \pi[(x-x_o)\,\Delta\ell - r']}{\pi[(x-x_o)\,\Delta\ell - r']} \frac{\sin \pi[(y-y_o)\,\Delta m - s']}{\pi[(y-y_o)\,\Delta m - s']} \quad .$$

Note that when $x = x_o + r/\Delta\ell$ and $y = y_o + s/\Delta m$, all the terms in the double sum are zero except the one for which $r' = r$ and $s' = s$. Accordingly,

$$\phi_{ik}(x_o + r/\Delta\ell,\ y_o + s/\Delta m) = \Delta\ell\,\Delta m\, C_{rs}\, e^{j2\pi[(x_o + r/\Delta\ell)\ell_o + (y_o + s/\Delta m)m_o]} ,$$

which evaluates the coefficients C_{rs}. The cross-correlation becomes

$$\phi_{ik}(x, y) = e^{j2\pi[(x-x_o)\ell_o + (y-y_o)m_o]} \sum_{r,s} \phi_{ik}(x_o + r/\Delta\ell,\ y_o + s/\Delta m)$$
$$\times\, e^{-j2\pi(r\ell_o/\Delta\ell + sm_o/\Delta m)}\, \frac{\sin\pi[(x - x_o)\,\Delta\ell - r]}{\pi[(x - x_o)\,\Delta\ell - r]}\, \frac{\sin\pi[(y - y_o)\,\Delta m - s]}{\pi[(y - y_o)\,\Delta m - s]} \quad . \qquad (59)$$

The assumption that the power densities arrive from within the rectangular range of directions $\Delta\Omega$ has made possible this expression for the cross-correlation at any point on the xy-plane in terms of samples of the cross-correlation,

$$\phi_{ik}(x_o + r/\Delta\ell,\ y_o + s/\Delta m) \quad ,$$

taken at a rectangular array of discrete points on the xy-plane. Rows of points are separated by $1/\Delta\ell$ in the x-direction and $1/\Delta m$ in the y-direction. The smaller the range of directions $\Delta\Omega$ of the arriving power densities, the greater is the allowable spacing of the measurement points. Even if we could only state with certainty that none of the storage fields of the source reach the receiving site, so that the power densities might arrive from any direction within $\ell^2 + m^2 = 1$, we could still inscribe that range in a square of sides $\Delta\ell = \Delta m = 2$, and we could determine the cross-correlations everywhere on the xy-plane from measurements taken at a rectangular array of points spaced by a half wavelength.

The constants x_o and y_o are still quite arbitrary (they should be real numbers); they may be chosen at the discretion of the experimenter to displace the measurement grid with respect to the origin, where the reference antenna is located, so that none of the measurement points coincides with the origin.

The rectangular power density, which is the Fourier transform of the cross-correlation, is just the Fourier series postulated at the start, which, in terms of the coefficients we have determined, becomes

$$\psi_{ik}(\ell, m) = \begin{cases} \dfrac{1}{\Delta\ell\,\Delta m} \displaystyle\sum_{r,s} \phi_{ik}(x_o + r/\Delta\ell,\ y_o + s/\Delta m)\, e^{-j2\pi[(x_o + r/\Delta\ell)\ell + (y_o + s/\Delta m)m]} \\ \qquad\qquad \text{for} \quad (\ell, m) \text{ inside } \Delta\Omega \\ 0 \text{ otherwise} \quad . \end{cases} \qquad (60)$$

Unfortunately, this application of the Sampling Theorem only applies directly to the measurement on a plane perpendicular to the ground if the range of directions $\Delta\Omega$ is large enough to include the directions of arrival of both incident and reflected plane waves. This would tend to reduce the permissible vertical separation of measurement points on the xz-plane.

* * *

The methods of power-density measurement discussed so far have had the common advantage that relatively modest antennas may be used. As one of these is shifted around on the measurement plane, the field pattern that would appear over the aperture of a larger antenna is explored. But the most intuitive device for measuring power density – a very directive (and hence very large) steerable antenna – has not yet been discussed.

It should be clear that if a single, large antenna is steered across the sky and the received power measured at appropriate directions, only a single power density can result. With two such antennas, used separately, it might be possible to measure the self-densities Φ_{HH} and Φ_{VV}, but for the other two, cross-correlation measurements will again be needed. One attack is suggested by the first part of (34), which reads

$$\psi_{aik}(\ell, m) = \overline{P_{ia}(\ell, m)\, E^*_{ka}(0, 0, 0)} \quad .$$

A very directive "i-polarized" antenna pointed in the direction (ℓ, m) would have a terminal voltage proportional to $P_{ia}(\ell, m)$, and a k-oriented dipole located at the origin would have a terminal voltage proportional to $E_{ka}(0, 0, 0)$. Cross-correlating these two, we would obtain one of the rectangular power densities. Unfortunately, there are a number of difficulties with an experiment of this type. One of these is that, when a practical steerable antenna of finite aperture dimensions is used, it is advantageous to mount the steerable antenna close to the reference dipole – so close, in fact, that the structure of the steerable antenna is just in front of or just behind the reference dipole, interfering with its reception.

A more convenient measurement can be made with a single, large, steerable dish having both horizontal and vertical feeds at the focal point. The feeds should be arranged so that the mutual impedance between the horizontal terminals and the vertical terminals is zero. This allows the horizontal terminals and vertical terminals to be regarded as the terminals of two separate antennas, each steered in the same direction and located at the same point in space (the origin of coordinates). These antennas may be described by the square matrix,

$$[F_s(\ell, m, \varphi, \Theta)] = \begin{bmatrix} F_{Hh}(\ell, m, \varphi, \Theta) & F_{Hv}(\ell, m, \varphi, \Theta) \\ F_{Vh}(\ell, m, \varphi, \Theta) & F_{Vv}(\ell, m, \varphi, \Theta) \end{bmatrix} ,$$

of horizontal and vertical plane-wave components emitted when the antennas are excited with unit current. F_{Hh} and F_{Vh} are the horizontal and vertical patterns resulting from excitation of the horizontal feed; F_{Hv} and F_{Vv} are the horizontal and vertical patterns resulting from

excitation of the vertical feed. To a good approximation, F_{Vh} and F_{Hv} are zero, but we shall retain them for completeness. As has been indicated, the elements of the pattern matrix will depend on the angles φ and Θ through which the antenna structure has been elevated and rotated in azimuth.

The open-circuit voltages of the two sections of the antenna, according to (55), are

$$\left.\begin{matrix} V_h(\varphi,\Theta) \\ V_v(\varphi,\Theta) \end{matrix}\right] = \frac{2\Lambda^2}{\eta} \int_{u.h.} [F_s(t,m,\varphi,\Theta)]_t \times F_a(t,m)]\, d\Omega \quad .$$

This time the correlation matrix to be measured is

$$[\psi_s(\varphi,\Theta)] = \overline{V_s(\varphi,\Theta)] \times \underline{V_s^*(\varphi,\Theta)}}$$

$$= \left(\frac{2\Lambda^2}{\eta}\right)^2 \int_{u.h.} [F_s(t,m,\varphi,\Theta)]_t \, d\Omega$$

$$\times \int_{u.h.} \overline{F_a(t,m)] \times \underline{F_a^*(t',m')}} \times [F_s^*(t',m',\varphi,\Theta)]\, d\Omega' \quad ,$$

and, as before, the pattern matrix may be removed from the second integral, which is then recognized as $[\Phi(t,m)]$. The remaining expression,

$$[\psi_s(\varphi,\Theta)] = \left(\frac{2\Lambda^2}{\eta}\right)^2 \int_{u.h.} [F_s(t,m,\varphi,\Theta)]_t \times [\Phi(t,m)] \times [F_s^*(t,m,\varphi,\Theta)]\, d\Omega \quad ,$$

will have to be treated somewhat differently this time, however. It will be assumed that, when the antenna structure is elevated by an angle φ and rotated by an angle Θ, the patterns have sharp maxima in the direction (t_o, m_o) – so sharp that the power-density patterns remain essentially constant over the range of directions for which there is an appreciable contribution to the integral. The integral becomes

$$[\psi_s(\varphi,\Theta)] = \left(\frac{2\Lambda^2}{\eta}\right)^2 \int_{u.h.} [F_s(t,m,\varphi,\Theta)]_t \times [\Phi(t_o,m_o)] \times [F_s^*(t,m,\varphi,\Theta)]\, d\Omega \quad . \qquad (62)$$

In this form, the Φ-matrix is constant so far as the integration is concerned. Unfortunately, it cannot be removed from the integrand because of the variable matrices that pre- and post-multiply it. Nevertheless, the 16 cross-products that multiply the various elements of the Φ-matrix may be integrated, and from these and the measured elements of the correlation matrix $[\psi_s]$ the four power densities may be obtained. If F_{Vh} and F_{Hv} may be neglected, only four integrals, one of which is the conjugate of another, need be evaluated; each power density is then proportional, through one of the four integrals, to the corresponding correlation.

From (62), the density of power arriving from (t_o, m_o) may be obtained. The measurement and calculations should be repeated for other directions until there are enough data to determine the power-density patterns in their entirety. Strictly speaking, the pattern integrals in (62) should be re-evaluated each time the elevation angle φ is changed. However, if the patterns are extremely peaked and if no measurements are to be made in the general direction of the zenith, this may not be necessary.

The same kind of long-time average measurement must be made with the antenna pointed in one direction as is necessary when the movable antenna is located at one point in the alternate technique of making cross-correlation measurements on a plane. And, in general, the number of directions in which measurements are made should agree with the number of measurement points chosen for the alternate scheme. Thus, although the use of the very directive, steerable antenna for power-density measurement is conceptually more straightforward than the cross-correlation measurements on a plane, more elaborate equipment is required, and the effort involved will be about the same.

VI. OPTIMIZATION OF AN ARRAY

The design of two specific types of receiving antenna will be considered – one in this section and another in Sec. VII. In each case, the required information about the arriving scatter signal, which is given by the power-density patterns, is presumed known. The two antennas will differ because different restrictions will be placed on their construction. It is apparent that some restriction should be placed on the shape, and particularly on the size, of the aperture: In general, the larger the aperture dimensions, the more power may be extracted from the arriving scatter fields, although, as we saw in Sec. I, the point of diminishing returns is reached when the aperture of the receiving antenna is as big as the transmitting-antenna aperture.

The antenna to be considered in this section is an array of smaller, not necessarily identical, antennas. The number, construction, location and orientation of the array elements is presumed to be included in the list of restrictions on the construction of the antenna; although we will not specify these choices, they are important and should be given due consideration. The remainder of the design, which will be considered here, will amount to the design of the interconnections between the terminals of the array elements and the common load to which the received power is to be supplied – a network problem.

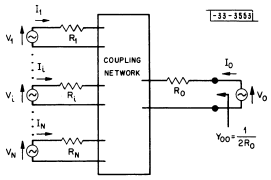

Fig. 16. The coupling network for a group of disconnected sources.

As shown in Fig. 16, we will start by designing a linear coupling network that will extract the maximum mean power from a group of disconnected sources and supply it to a single load. N sources, each having an open-circuit voltage V_i, an internal resistance R_i, and supplying a current I_i, are indicated on the left of the coupling network. The single load to which the power is to be supplied is R_o on the right; I_o is the load current. The extra voltage source V_o on the right does not belong in the description of a receiving antenna; it will be used temporarily to help demonstrate a restriction on the elements of the coupling network. As usual, the V_i and the I_i are the time-varying complex amplitudes of the corresponding voltages and currents. In practice, the coupling network might consist of a collection of transmission lines that pass all the frequency components of a typical communication signal without appreciable relative attenuation or phase shift: We will assume that the input-output relations of the coupling network, calculated for steady-state sinusoidal inputs, will hold when the inputs are fading, modulated carriers.

The circuit of Fig. 16 is a fairly accurate representation of the interconnections between the elements of an array and their common load. V_i may be regarded as the open-circuit voltage of the i^{th} element and R_i its radiation resistance, so that the i^{th} source on the left is the Thévenin equivalent of the i^{th} element as seen from its terminals. (It is assumed that the driving-point reactance of each element has been resonated.) But if two antennas are located close together as they might be in an array, and if one is driven with a generator, a fraction of the generator

voltage will appear at the terminals of the other, indicating a mutual impedance between the two antennas. These mutual impedances, which are neglected in the present representation, will be considered later.

The currents in Fig. 16 are related to the voltages by the set of N + 1 linear equations,

$$I_i = \sum_{k=0}^{N} Y_{ik} V_k \quad , \quad i = 0, 1, 2, \ldots N \quad . \tag{63}$$

where the Y_{ik}'s are the short-circuit driving-point and transfer admittances of the coupling network and source and load resistances, i.e., the network to which the voltage sources are connected.

Regardless of the construction of the coupling network, maximum power will be delivered to the load R_o if it is matched to the output of the coupling network. This means that the impedance seen looking in at the output terminals of the coupling network (with the sources dead) will also be R_o, and the output self-admittance, shown in Fig. 16, will be

$$Y_{oo} = \frac{1}{2R_o} \quad . \tag{64}$$

We may now develop a restriction on the admittances Y_{ik} that describe the coupling network. The restriction we will need is a direct consequence of the obvious fact that the power from generator V_o dissipated in the source and load resistances $R_o \ldots R_N$ is less than or equal to the power supplied by generator V_o, the other generator voltages, $V_1 \ldots V_N$, being zero:

$$\tfrac{1}{2} \sum_{i=0}^{N} |I_i|^2 R_i \leq \tfrac{1}{2} \mathrm{Re}\{V_o I_o^*\} \quad .$$

Using (63) to express the currents in terms of V_o, the voltage of the only live source,

$$\tfrac{1}{2} \sum_{i=0}^{N} |Y_{io}|^2 R_i |V_o|^2 \leq \tfrac{1}{2} \mathrm{Re}\{Y_{oo}^* |V_o|^2\} \quad ,$$

$$\sum_{i=0}^{N} |Y_{io}|^2 R_i \leq \mathrm{Re}\{Y_{oo}^*\} = \frac{1}{2R_o} \quad .$$

The first term of the sum is $|Y_{oo}|^2 R_o = \frac{1}{4R_o}$, so that the remaining terms are

$$\sum_{i=1}^{N} |Y_{io}|^2 R_i \leq \frac{1}{4R_o} \quad ,$$

which may be written

$$\sum_{i=1}^{N} \left| 2\sqrt{R_o R_i}\, Y_{io} \right|^2 = \sum_{i=1}^{N} \left| 2\sqrt{R_o R_i}\, Y_{oi} \right|^2 \leq 1 \quad ,$$

since $Y_{ik} = Y_{ki}$ by reciprocity. The quantity

$$t_i = 2\sqrt{R_o R_i}\, Y_{oi} \tag{65}$$

is the transmission coefficient from the i^{th} source on the left in Fig. 16 to the load (or from the right-hand source to the i^{th} source resistance on the left). We have just demonstrated that

$$\sum_{i=1}^{N} |t_i|^2 \leq 1 \quad , \tag{66}$$

which is equivalent to the statement that the sum of the squared magnitudes of the elements in the first row of the scattering matrix[11] is less than or equal to one when (in the present case) the load reflection coefficient – the diagonal element – is zero.

Evidently the inequality (66) becomes an equality if the coupling network is lossless. It will also be an equality if the coupling network contains resistors that are not coupled to the load, but that may, however, be coupled to one or more of the left-hand sources.

We will have no further use for the right-hand voltage source, which may be considered to be removed ($V_o = 0$). The mean power in the load from the left-hand sources is

$$W_a = 1/2 |I_o|^2 R_o$$

$$= 1/2 R_o \text{ Av.} \left\{ \left| \sum_{i=1}^{N} Y_{oi} V_i \right|^2 \right\}$$

$$= 1/8 \text{ Av.} \left\{ \left| \sum_{i=1}^{N} \left(2\sqrt{R_o R_i}\, Y_{oi} \right) \frac{V_i}{\sqrt{R_i}} \right|^2 \right\}$$

from (63). Or, using the transmission coefficients, (65),

$$W_a = 1/8 \text{ Av.} \left\{ \left| \sum_{i=1}^{N} t_i \frac{V_i}{\sqrt{R_i}} \right|^2 \right\} \quad . \tag{67}$$

It has become clear that, for maximum power, the restriction (66) should be satisfied with the equal sign, for if it read

$$\sum_{i=1}^{N} |t_i|^2 = a^2 \quad , \quad 0 < a^2 < 1 \quad ,$$

so that

$$\sum_{i=1}^{N} \left| \frac{t_i}{a} \right|^2 = 1 \quad ,$$

the mean load power, which could be written

$$W_a = 1/8\, a^2 \text{ Av.} \left\{ \left| \sum_{i=1}^{N} \frac{t_i}{a} \frac{V_i}{\sqrt{R_i}} \right|^2 \right\} \quad ,$$

would only be the fraction a^2 of that possible if (66) were satisfied with the equal sign. Thus, we should adjust the transmission coefficients in such a way as to maximize (67) subject to the constraint,

$$\sum_{i=1}^{N} |t_i|^2 = 1 \quad . \tag{68}$$

We have shown that the constraint (68) is a necessary condition on the realizability of the optimum coupling network, but as yet we have not shown that it is sufficient. To show that the constraint is sufficient, it will be simplest to produce a realization of the coupling network, which, of course, is one of the desired results. Such a realization is shown in Fig. 17. The

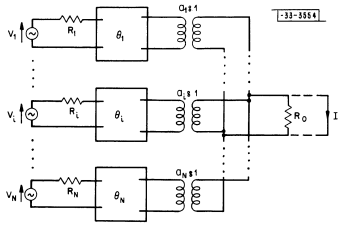

Fig. 17. A realization of the optimum coupling network.

boxes labeled $\theta_1 \cdots \theta_N$ are phase shifters that are coupled by ideal transformers to the load R_o. Physically, the i^{th} phase shifter might be a transmission line of characteristic impedance R_i and electrical length θ_i radians. The i^{th} ideal transformer, which has a turns ratio of a_i to 1, might also be approximated by appropriate connections to a resonant length of transmission line. The parallel rather than the series connection of transformer secondaries has been chosen to simplify the construction of the coupling network: A single autotransformer (a single resonant line) with various primary taps will serve the purposes of all N ideal transformers shown in Fig. 17.

With the load short-circuited, the current flowing in the short is

$$I = \sum_{i=1}^{N} a_i e^{-j\theta_i} \frac{V_i}{R_i} \quad ,$$

and if the load is matched, the mean load power is

$$W_a = 1/8 \, \overline{|I|^2} \, R_o = 1/8 \, R_o \, \text{Av.} \left\{ \left| \sum_{i=1}^{N} a_i e^{-j\theta_i} \frac{V_i}{R_i} \right|^2 \right\} \quad ,$$

which may be written

$$W_a = 1/8 \, \text{Av.} \left\{ \left| \sum_{i=1}^{N} \left(a_i \sqrt{\frac{R_o}{R_i}} e^{-j\theta_i} \right) \frac{V_i}{\sqrt{R_i}} \right|^2 \right\} \quad . \tag{69}$$

The load will be matched if the admittances add up to

$$\sum_{i=1}^{N} \frac{a_i^2}{R_i} = \frac{1}{R_o} \quad ,$$

which may be written

$$\sum_{i=1}^{N} \left| a_i \sqrt{\frac{R_o}{R_i}} e^{-j\theta_i} \right|^2 = 1 \quad . \tag{70}$$

Comparing (69) with (67) and (70) with (68), we recognize that the transmission coefficients for the network of Fig. 17 may be taken to be

$$t_i = a_i \sqrt{\frac{R_o}{R_i}} e^{-j\theta_i} \quad . \tag{71}$$

If the turns ratios and phase shifts are appropriately chosen, the network of Fig. 17 will indeed be a realization of the optimum coupling network.

It remains to maximize the mean load power by adjusting the transformers and phase shifters. We will work with the transmission coefficients, from which the turns ratios and phase shifts may be determined by (71). Manipulating (67), the mean load power becomes

$$W_a = 1/8 \sum_{i=1}^{N} \sum_{k=1}^{N} t_i t_k^* \frac{\overline{V_i V_k^*}}{\sqrt{R_i R_k}} = \sum_{i=1}^{N} \sum_{k=1}^{N} t_i t_k^* \phi_{ik} \quad . \tag{72}$$

And we are not surprised to discover that the data required for the maximization consists of the set

$$\phi_{ik} = \frac{1}{8\sqrt{R_i R_k}} \overline{V_i V_k^*} \tag{73}$$

of quantities proportional to the temporal cross-correlations of open-circuit source voltages. We presume that these quantities, which may be determined from the power densities by the methods of Sec. V (or which may be measured directly!), are known. In any case, we note from (73) that $\phi_{ik} = \phi_{ki}^*$, so that the auto-correlations are real and non-negative.

We will use the method of Lagrange to maximize (72) subject to the constraint (68): We will maximize

$$U = \sum_{i=1}^{N} t_i \sum_{k=1}^{N} t_k^* \phi_{ik} - \alpha \sum_{i=1}^{N} |t_i|^2 \quad ,$$

where α is the Lagrangian multiplier. There are 2 N relations; N state that U is stationary with the N real parts of the t's, and N state that U is stationary with the imaginary parts. Temporarily, we let $t_i = \tau_i + j\sigma_i$, where the τ_i and σ_i are real. Then

$$U = \sum_{i=1}^{N} (\tau_i + j\sigma_i) \sum_{k=1}^{N} (\tau_k - j\sigma_k)\, \phi_{ik} - \alpha \sum_{i=1}^{N} (\tau_i^2 + \sigma_i^2) \quad ,$$

and the first N relations are

$$\frac{\partial U}{\partial \tau_n} = \sum_{k=1}^{N} (\tau_k - j\sigma_k)\, \phi_{nk} + \sum_{i=1}^{N} (\tau_i + j\sigma_i)\, \phi_{in} - 2\alpha\, \tau_n = 0$$

$$\text{for} \quad n = 1, 2, \ldots, N \quad .$$

The second N relations read

$$\frac{\partial U}{\partial \sigma_n} = j \sum_{k=1}^{N} (\tau_k - j\sigma_k)\, \phi_{nk} - j \sum_{i=1}^{N} (\tau_i + j\sigma_i)\, \phi_{in} - 2\alpha\, \sigma_n = 0$$

$$\text{for } n = 1, 2, \ldots, N \quad .$$

The sum

$$1/2 \left(\frac{\partial U}{\partial \tau_n} + j \frac{\partial U}{\partial \sigma_n} \right) = \sum_{i=1}^{N} (\tau_i + j\sigma_i)\, \phi_{in} - \alpha\, (\tau_n + j\sigma_n) = 0$$

utilizes all 2 N relations in N complex relations:

$$\sum_{i=1}^{N} t_i\, \phi_{in} - \alpha\, t_n = 0 \quad , \qquad n = 1, 2, \ldots, N \quad . \tag{74}$$

At this point, it will be convenient to convert to matrix notation. Let the column matrix of transmission coefficients and the column matrix of normalized open-circuit source voltages be

$$\underline{t} = \begin{bmatrix} t_1 \\ t_2 \\ \cdot \\ \cdot \\ \cdot \\ t_N \end{bmatrix} \quad \text{and} \quad \left[\frac{V}{\sqrt{8R}} \right] = \begin{bmatrix} \frac{V_1}{\sqrt{8R_1}} \\ \frac{V_2}{\sqrt{8R_2}} \\ \cdot \\ \cdot \\ \cdot \\ \frac{V_N}{\sqrt{8R_N}} \end{bmatrix} \tag{75}$$

so that the square matrix of normalized cross-correlations is

$$[\phi] = \overline{\left[\frac{V}{\sqrt{8R}} \right]} \times \underline{\left[\frac{V^*}{\sqrt{8R}} \right]} = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2N} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \phi_{N1} & \phi_{N2} & \cdots & \phi_{NN} \end{bmatrix} \quad . \tag{76}$$

Evidently, the constraint may be written

$$\underline{t} \times t^*] = 1 \quad ; \tag{68a}$$

the mean load power is

$$W_a = \underline{t} \times [\phi] \times t^*] \quad ; \tag{72a}$$

and the result of the maximization reads

$$\underline{t} \times [\phi] = \alpha\, \underline{t} \quad . \tag{74a}$$

Post-multiplying both sides of (74a) by t*], we have

$$\underline{t} \times [\phi] \times t^*] = \alpha\, \underline{t} \times t^*] = \alpha \quad ,$$

so that the Lagrange multiplier α is the maximum mean load power we seek.

The solution of (74a) for the maximum multiplier and corresponding t-matrix is the well-known characteristic value problem.[12] Only enough will be said about it to indicate the nature of its solution and to show the physical significance of certain special cases.

The result, (74a), may be written

$$\underline{t} \times \{ [\phi] - \alpha\, [1] \} = 0 \quad ,$$

in which [1] is the unit matrix of order N. This equation has non-trivial solutions only if the determinant

$$| \{ [\phi] - \alpha\, [1] \} | = 0 \quad ,$$

that is,

$$\begin{vmatrix} \phi_{11} - \alpha & \phi_{12} & \cdots & \phi_{1N} \\ \phi_{21} & \phi_{22} - \alpha & \cdots & \phi_{2N} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \phi_{N1} & \phi_{N2} & \cdots & \phi_{NN} - \alpha \end{vmatrix} = 0 \quad , \tag{77}$$

which is the characteristic equation of the matrix $[\phi]$. Evidently, this equation has N α-roots or characteristic values. In spite of the fact that the off-diagonal elements ϕ_{ik} are, in general, complex, (72a) is a non-negative definite Hermitian form, so the characteristic values must be real and non-negative, and the coefficients of the characteristic equation must be real, and, if none of them is zero, they must alternate in sign. We are interested in obtaining the maximum power; the greatest characteristic value α_m is that maximum power.

If the array elements have been properly placed, α_m will be a distinct (not multiple) root of the characteristic equation. In that case, the rank of the system determinant

$$|[\phi] - \alpha_m [1]|$$

will be equal to $(N-1)$, and at least one of the cofactors C_{ik} of its elements will be different from zero. (74a) will then be satisfied by the characteristic vector (the set of transmission coefficients)

$$t_i \propto C_{ik}$$

and the column k may be chosen so that at least one of the C_{ik} is different from zero. In other words, the elements of t] are proportional to the elements of one column of the matrix of cofactors. The constraint, (68), will also be satisfied if the constant of proportionality is chosen so that

$$t_i = \frac{C_{ik}}{\sqrt{\sum_{i=1}^{N} |C_{ik}|^2}} e^{j\Theta} \quad , \tag{78}$$

which determines the transmission coefficients to within an arbitrary phase shift Θ common to all of them.

On the other hand, if α_m is a multiple root of the characteristic equation, the rank of the system determinant will be less than $(N-1)$, and all its cofactors will be zero, so that (78) will be indeterminate. In this case, (74a) is satisfied with $\alpha = \alpha_m$ by at least two independent sets $t_a]$ and $t_b]$ or by any linear combination of these:

$$t] = a\, t_a] + b\, t_b] \quad .$$

As a result, any one of the transmission coefficients may be made zero by properly choosing the constants a and b. Of course, making $t_i = 0$ amounts to disconnecting the i^{th} source entirely from the load. For this to be possible without reducing the maximum available load power, the array elements would have to be particularly poorly placed.

As an example of the foregoing considerations, suppose that a coupling network is desired for a three-element array having the matrix of normalized cross-correlations,

$$[\phi] = \begin{bmatrix} 4 & 1+j & -j2 \\ 1-j & 5 & 1+j \\ j2 & 1-j & 4 \end{bmatrix} \quad .$$

Note from the diagonal elements that, by itself, source 1 would supply 4 watts to a matched load; similarly, sources 2 and 3 would supply 5 watts and 4 watts, respectively. The characteristic equation is

$$\begin{vmatrix} 4-\alpha & 1+j & -j2 \\ 1-j & 5-\alpha & 1+j \\ j2 & 1-j & 4-\alpha \end{vmatrix} = 36 - 48\alpha + 13\alpha^2 - \alpha^3 \quad ,$$

$$= -(\alpha-1)(\alpha-6)^2 = 0 \quad ,$$

so that the maximum of 6 watts may be drawn from the three sources in a number of ways. The system determinant for a 6-watt connection is

$$\begin{vmatrix} -2 & 1+j & -j2 \\ 1-j & -1 & 1+j \\ j2 & 1-j & -2 \end{vmatrix} = 0 \quad ,$$

and, in addition, it may be seen by inspection that all 9 cofactors are zero. We will obtain the 6 watts in such a way that the third source is entirely disconnected from the load ($t_3 = 0$). Deleting the third row and column, the system determinant becomes

$$\begin{vmatrix} -2 & 1+j \\ 1-j & -1 \end{vmatrix} = 0 \quad ,$$

and the matrix of its cofactors is

$$\begin{bmatrix} -1 & -1+j \\ -1-j & -2 \end{bmatrix} \quad .$$

The reduced t-matrix will be taken from the first column of this matrix of cofactors, so that the entire t-matrix is

$$\underline{t}] \propto \begin{bmatrix} -1 \\ -1-j \\ 0 \end{bmatrix} .$$

It may also be seen by inspection that

$$\underline{-1, -1-j, 0} \times \begin{bmatrix} 4 & 1+j & -j2 \\ 1-j & 5 & 1+j \\ j2 & 1-j & 4 \end{bmatrix} = 6\,\underline{-1, -1-j, 0} ,$$

so that 6 watts are still available after the elimination of the third source. In normalizing the characteristic vector (78), the phase shift $\Theta = 3\pi/4$ will lead to the simplest network. The t-matrix is then

$$\underline{t}] = \begin{bmatrix} \frac{1}{\sqrt{3}} e^{-j3\pi/4} \\ \sqrt{\frac{2}{3}} \\ 0 \end{bmatrix} ,$$

and the corresponding network, for assumed equal source and load resistances, is shown in Fig. 18.

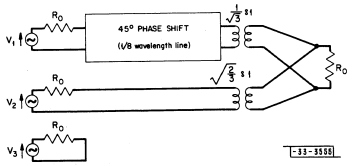

Fig. 18. In this example, one source may be entirely disconnected from the load.

As the number of sources (the order of the associated correlation matrix) increases, the labor involved in the method of solution just demonstrated soon becomes prohibitive. Fortunately, an iterative procedure, which is particularly well-suited to the problem of finding the greatest characteristic value and corresponding vector, is available.[13] Automatic computing machines may be programmed to follow this procedure.

The case of perfect correlation between source voltages and the case of complete incoherence of these voltages are two extremes worth mentioning. As discussed in Sec. II, if V_i and V_k are "perfectly correlated," they will "fade together" with fluctuations in the scattering region. We may write the column matrix of normalized open-circuit source voltages (75) as

$$\frac{\underline{V}}{\sqrt{8R}}] = g(t) \cdot \underline{v}] = g(t) \cdot \begin{bmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ \cdot \\ v_N \end{bmatrix} \quad \text{with} \quad \overline{|g(t)|^2} = 1 ,$$

where the v_i are not time functions. The matrix of normalized cross-correlations (76) becomes

$$[\phi] = \overline{|g(t)|^2} \cdot \underline{v}] \times \underline{v}^* = \underline{v}] \times \underline{v}^* .$$

In this case, the characteristic value problem is particularly simple. (74a) reads

$$\underline{t} \times [\phi] = \underline{t} \times \underline{v}] \times \underline{v}^* = \alpha\, \underline{t} ,$$

but this time, the product $\underline{t} \times \underline{v}]$ is a scalar constant, so that the result of the maximization is $\underline{t} \propto \underline{v}^*$. As before, we may choose the phase of the proportionality constant at our convenience, and its magnitude should be chosen to satisfy the constraint, (68a). Thus, if

$$\underline{v} \times \underline{v}^*] = W_c , \tag{79}$$

then $\underline{t}$ may be chosen as

$$\underline{t} = \frac{1}{\sqrt{W_c}}\, \underline{v}^* . \tag{80}$$

The maximum available power, (72a), is

$$W_a = \underline{t} \times \underline{v}] \times \underline{v}^* \times \underline{t}^*]$$
$$= \frac{1}{W_c}\, \underline{v}^* \times \underline{v}] \times \underline{v}^* \times \underline{v}] = W_c .$$

The maximum available power, (79), is seen to be the sum of the powers available from each source separately. As may be seen by expanding the characteristic equation, (77), the coefficient of $(-\alpha)^{(N-1)}$, which is also the sum of the characteristic values, is just the sum of the diagonal

elements: the sum of the powers available from each source separately. Since one characteristic value is equal to the latter sum, and since the characteristic values are non-negative, all the other characteristic values must be zero. The case of perfect correlation just considered might arise if a small array were illuminated by a point source or by a relatively concentrated source having a power-density pattern that is sharply peaked in one direction. It might also arise if the object of the design were to produce a coupling network that is optimum only for a period of time short compared with the fading period of the arriving signal. The latter possibility, which leads to a time-varying coupling network, will be considered in Sec. VIII.

The other extreme case is that of complete incoherence of the source voltages. In this case, the cross-correlations ϕ_{ik} between the source voltages are zero, and the correlation matrix reduces to the diagonal form:

$$[\phi] = \begin{bmatrix} \phi_{11} & 0 & \cdots & 0 \\ 0 & \phi_{22} & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & \phi_{NN} \end{bmatrix} .$$

The characteristic values of this matrix are just its diagonal elements. If the greatest of these is ϕ_{ii}, the maximum power of ϕ_{ii} watts may evidently be obtained by matching the i^{th} source directly to the load and disconnecting the other sources entirely. Almost complete incoherence would be expected for an array of widely separated elements that is illuminated by a diffuse source having a broad power-density pattern.

For example, if the array is receiving only noise, and if that noise arrives uniformly from all directions in space, we would expect the array element voltages to be completely incoherent. We discovered in Sec. IV that the available noise power is independent of the antenna pattern; the noise power must be independent of the connections in the coupling network for the array. Since the characteristic values of the correlation matrix represent powers available with various connections, all the characteristic values must be equal. As may be seen by imagining that the sources are coupled to the load two at a time, this implies that the off-diagonal elements ϕ_{ik} of the correlation matrix are zero, and that the diagonal elements are equal.

On the other hand, suppose that two of the elements of the array are short dipoles. In the last section, it was shown that the cross-correlation of their open-circuit voltages is proportional to the Fourier transform of a rectangular power density. Since noise power arrives only from the real directions, $\ell^2 + m^2 \ll 1$, this rectangular power density cannot be a constant: It must be zero outside of $\ell^2 + m^2 = 1$ and different from zero inside. Its transform, the cross-correlation, therefore, cannot be an impulse in space that is zero unless the locations of the two dipoles coincide. In other words, the open-circuit noise voltages of the two dipoles will generally

be partially correlated! This apparent paradox has resulted from our neglecting the mutual impedances between the array elements. When these are properly accounted for, the noise power is truly independent of the connections in the coupling network between the array elements and the load.

* * *

Instead of representing the array, as seen at the terminals of its elements, by the set of N disconnected sources shown at the left in Fig. 16, we are led to the more general representation of Fig. 19. The self and mutual impedances between array elements are the elements of the

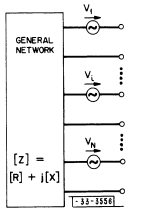

Fig. 19. A set of connected sources.

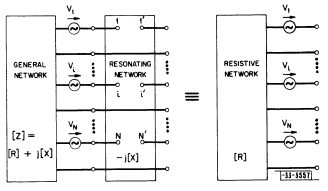

Fig. 20. The resonating network is connected in series with the connected sources.

matrix [Z] of open-circuit driving-point and transfer impedances measured at the terminals of the array elements before these have been coupled to the load. The open-circuit voltages of the array elements are the elements of the column matrix

$$V] = \begin{bmatrix} V_1 \\ V_2 \\ \cdot \\ \cdot \\ \cdot \\ V_N \end{bmatrix} . \qquad (81)$$

The open-circuit impedances and voltages have been chosen for convenience; in some cases, the open-circuit voltage of a dipole-array element can be calculated under the assumption that the other dipole elements, which are open-circuited, are removed entirely.

Again we are to design a coupling network that will extract maximum power from the array of coupled sources and deliver it to a single load. We will attack this problem by reducing it to the one we have just solved: the design of an optimum coupling network for a set of disconnected sources. The first step is to resonate the self and mutual reactances. If the resonating network is connected in "series" with the array elements as shown in Fig. 20, and if the self and mutual reactances of the resonating network are of opposite sign from the corresponding array reactances, the self and mutual impedances measured at the new terminal pairs are the self and mutual resistances of the array. We are left with the resistive network on the right in Fig. 20.

It should be emphasized that we are dealing strictly with terminal pair behavior: There would be no connection, for example, between terminal 1 and terminal i in the resonating network; current can flow only to terminal i' from terminal i. It may be imagined that terminals i – i' are the ends of the primary winding of an ideal transformer, and that all other connections in the resonating network are made to the secondary winding. In any case, it should be clear that the open-circuit voltages of the resistive network we have obtained are the same as the corresponding open-circuit voltages of the array.

Note that if we started with the resistive network on the right in Fig. 20 and added a "series" reactive network described by the matrix +j[X], we would reproduce the representation of the array (Fig. 19). Thus, if we can find a coupling network that extracts maximum power from the resistive network and delivers it to a single load, the resonating network and that coupling network are an arrangement for extracting maximum power from the array, for, if more power were available from the array alone, the coupling network could include the "series" reactive network +j[X] to obtain that greater power.

The logic of the preceding argument hinges on the fact that the process of adding the resonating network is reversible. We shall have to use the same type of argument again, because we have not discovered how to extract maximum power from the resistive network.

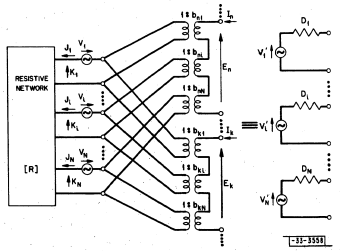

Fig. 21. The ideal transformers can create N disconnected sources from the N connected sources.

The next step is to diagonalize the resistance matrix of the resistive network. For this purpose, we consider the ideal transformer connection shown in Fig. 21. From the N terminal pairs of the resistive network we connect N^2 transformers with primaries in parallel and secondaries in series in such a way that N new terminal pairs are created. (Only the n^{th} and k^{th} output terminal pairs are shown in Fig. 21.) The transformer connecting the i^{th} resistive-network terminal pair to the n^{th} output circuit has a turns ratio $1 : b_{ni}$; all N^2 transformers are described by the matrix of turns ratios

$$[b] = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1N} \\ b_{21} & b_{22} & \cdots & b_{2N} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ b_{N1} & b_{N2} & \cdots & b_{NN} \end{bmatrix} . \qquad (82)$$

The transformers have been connected in such a way that, when the output terminal pairs are open-circuited, no current flows in the resistive network, and the open-circuit voltage at the n^{th} output terminal pair, which is equal to the open-circuit voltage of the n^{th} pair in the equivalent circuit on the right in Fig. 21, is

$$E_n = V'_n = \sum_{i=1}^{N} b_{ni} V_i \quad .$$

In matrix notation, from (82), the column matrix, (81), of resistive-network open-circuit voltages, and the corresponding column matrix of open-circuit voltages of the equivalent right-hand circuit,

$$V'] = [b] \times V] \quad . \qquad (83)$$

The self and mutual output resistances may be calculated with the sources dead (V] = 0). In that case, the currents J_i are related to the currents I_n by

$$J_i = \sum_{n=1}^{N} b_{ni} I_n \quad ; \quad J] = [b]_t \times I] \quad .$$

The voltages K] are

$$K] = [R] \times J] = [R] \times [b]_t \times I] \quad ,$$

and the voltages E] are

$$E] = [b] \times K] = [b] \times [R] \times [b]_t \times I] \quad .$$

We will call the matrix of output resistances [D] in anticipation of its being diagonal:

$$E] = [D] \times I] \quad ; \quad [D] = [b] \times [R] \times [b]_t \quad . \tag{84}$$

Evidently, a set of transformer connections described by any real matrix [b] can be realized; if $b_{ni} = 0$, the corresponding transformer should be removed, leaving its primary circuit open and its secondary short. Because of reciprocity, the resistance matrix [R] is symmetric, and it may indeed be diagonalized by the operation (84) with considerable freedom in the choice of the elements of [b].

Guillemin[14] discusses a method for obtaining the diagonalizing matrix, [b], as the product of $M = N(N-1)/2$ elementary matrices:

$$[b] = [b_M] \times [b_{M-1}] \times \ldots \times [b_2] \times [b_1] \quad .$$

The elementary matrix $[b_1]$ is chosen so that $[b_1] \times [R] = [R']$ has a zero in the second row, first column:

$$\begin{bmatrix} 1 & 0 & 0 \ldots \\ -\frac{R_{21}}{R_{11}} & 1 & 0 \ldots \\ 0 & 0 & 1 \ldots \\ . & . & . \\ . & . & . \\ . & . & . \end{bmatrix} \times \begin{bmatrix} R_{11} & R_{12} & R_{13} \ldots \\ R_{21} & R_{22} & R_{23} \ldots \\ R_{31} & R_{32} & R_{33} \ldots \\ . & . & . \\ . & . & . \\ . & . & . \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \ldots \\ 0 & R'_{22} & R'_{23} \ldots \\ R_{31}, & R_{32} & R_{33} \ldots \\ . & . & . \\ . & . & . \\ . & . & . \end{bmatrix} ,$$

in which $R'_{2i} = R_{2i} - \frac{R_{21}}{R_{11}} R_{1i}$. Next, $[b_2]$ is chosen so that $[b_2] \times [R'] = [R'']$ has a zero in the third row, first column:

$$\begin{bmatrix} 1 & 0 & 0 \ldots \\ 0 & 1 & 0 \ldots \\ -\frac{R_{31}}{R_{11}} & 0 & 1 \ldots \\ . & . & . \\ . & . & . \\ . & . & . \end{bmatrix} \times \begin{bmatrix} R_{11} & R_{12} & R_{13} \ldots \\ 0 & R'_{22} & R'_{23} \ldots \\ R_{31} & R_{32} & R_{33} \ldots \\ . & . & . \\ . & . & . \\ . & . & . \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \ldots \\ 0 & R'_{22} & R'_{23} \ldots \\ 0 & R''_{32} & R''_{33} \ldots \\ . & . & . \\ . & . & . \\ . & . & . \end{bmatrix} ,$$

in which $R''_{3i} = R_{3i} - \frac{R_{31}}{R_{11}} R_{1i}$. Next, $[b_3]$ is chosen so that $[b_3] \times [R''] = [R'']$ has a zero in the third row, second column:

$$\begin{bmatrix} 1 & 0 & 0 \ldots \\ 0 & 1 & 0 \ldots \\ 0 & -\frac{R''_{32}}{R'_{22}} & 1 \ldots \\ . & . & . \\ . & . & . \\ . & . & . \end{bmatrix} \times \begin{bmatrix} R_{11} & R_{12} & R_{13} \ldots \\ 0 & R'_{22} & R'_{23} \ldots \\ 0 & R''_{32} & R''_{33} \ldots \\ . & . & . \\ . & . & . \\ . & . & . \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \ldots \\ 0 & R'_{22} & R'_{23} \ldots \\ 0 & 0 & R'''_{33} \ldots \\ . & . & . \\ . & . & . \\ . & . & . \end{bmatrix} ,$$

in which $R'''_{3i} = R''_{3i} - \frac{R''_{32}}{R'_{22}} R'_{2i}$. The process continues into the fourth and lower rows until the remaining matrix, $[b] \times [R]$, is triangular. The diagonal elements of this matrix (R_{11}, R'_{22}, $R'''_{33} \ldots$) are the elements of the diagonal matrix [D], the source resistances on the right in Fig. 21:

$$[D] = [b] \times [R] \times [b]_t = \begin{bmatrix} R_{11} & 0 & 0 \ldots \\ 0 & R'_{22} & 0 \ldots \\ 0 & 0 & R'''_{33} \ldots \\ . & . & . \\ . & . & . \\ . & . & . \end{bmatrix} .$$

When this diagonalization procedure is employed, the determinant of each elementary matrix, and hence the determinant of [b], is unity. [b] has the triangular form

$$[b] = \begin{bmatrix} 1 & 0 & 0 \ldots \\ b_{21} & 1 & 0 \ldots \\ b_{31} & b_{32} & 1 \ldots \\ . & . & . \\ . & . & . \\ . & . & . \end{bmatrix} ,$$

so that no more than $N(N-1)/2$ transformers and exactly N direct connections will complete the set of transformer connections described by [b]. This saving of somewhat over half of the possible N^2 transformers is a decided advantage to the method of diagonalization under discussion.

Since it represents a physically realizable network, the matrix [D] of the resistive network as seen at the output terminals of the transformers must be non-negative definite. In addition, since it would be unreasonable to expect infinite power to be available from the array, each source resistance D_i must be greater than zero, or if $D_i = 0$, then $V_i' = 0$. The case $D_i = 0$ and $V_i' \neq 0$ would not be expected in a practical situation; it could arise, for example, if some of the array elements were interconnected through a lossless network in such a way as to create an extra pair of terminals. We may expect, then, that [D] and [R] are positive definite, and that the diagonal elements D_i are all greater than zero. It will be noted that the condition $D_i \neq 0$ for $i = 1, 2, \ldots k$ is necessary and sufficient to allow the diagonalization procedure to be carried through the $(k+1)^{st}$ row.

Since the determinant $|[b]| = 1$, $[b]^{-1}$ must exist. (It has the same triangular form as [b].) (83) and (84) may then be rearranged to read

$$V_] = [b]^{-1} \times V'] \tag{83a}$$

and

$$[R] = [b]^{-1} \times [D] \times [b]_t^{-1} \quad . \tag{84a}$$

These equations show that if the disconnected sources on the right in Fig. 21 are followed by the set of ideal transformer connections described by $[b]^{-1}$, the terminal-pair behavior of the resistive network and its open-circuit voltages are reproduced. In other words, the process brought about by the set of transformer connections is reversible. Thus, the set of transformer connections, followed by a coupling network that extracts maximum power from the disconnected sources and delivers it to a single load, is an optimum arrangement for extracting power from the resistive network, for if more power were available from the resistive network, the coupling network could include the set of transformer connections described by $[b]^{-1}$ to obtain that greater power.

We have shown that if the array is followed by a resonating network (Fig. 20), then by a set of transformer connections (Fig. 21) and, finally, by a coupling network (Fig. 17), the combined circuit will deliver the maximum power from the array to a single load. If the signal data consist of the elements of the matrix of cross-correlations of array-element voltages, $\overline{V] \times \underline{V^*}}$, the matrix of cross-correlations of the open-circuit voltages of the disconnected sources, shown on the right in Fig. 21, is

$$\overline{V'] \times \underline{V'^*}} = [b] \times \overline{V] \times \underline{V^*}} \times [b]_t$$

from (83). And the matrix of normalized cross-correlations to be used in the characteristic value problem associated with the disconnected sources is

$$[\phi] = 1/8 \left[\frac{1}{\sqrt{D}}\right] \times [b] \times \overline{V] \times \underline{V^*}} \times [b]_t \times \left[\frac{1}{\sqrt{D}}\right] \quad , \tag{85}$$

where the diagonal matrix

$$\left[\frac{1}{\sqrt{D}}\right] = \begin{bmatrix} \frac{1}{\sqrt{D_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{D_2}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & \frac{1}{\sqrt{D_N}} \end{bmatrix} \quad . \tag{86}$$

The ϕ-matrix has again been normalized in such a way that its characteristic values are extreme available powers. As usual, the form of the final coupling network is dictated by the solution vector corresponding to the greatest characteristic value.

It is somewhat disheartening to contemplate the hardware required by the foregoing method of extracting maximum power from a large array. Consider a 100-element array consisting of 10 rows and 10 columns of antennas. The number N^2 is 10,000; about 5000 transformers and a similar number of elements in the resonating network would be required. Any such design would clearly be uneconomical. Of course, the mutual impedance between widely separated array elements would be small, so that many of the mutual impedances could probably be neglected altogether. This would effect a considerable saving in both the number of transformers and the number of elements in the resonating network. The mutual impedances would be particularly small for an array of moderately directive elements that were arranged as far as possible not to have overlapping storage fields. Nevertheless, a number of circuit elements greatly in excess of the number of array elements would probably be required in any case.

The advantage of the foregoing development has been that we were able to prove that it led to the maximum available power. Fortunately, the circuits we developed are not the only ones that can collect maximum power from an array; there are many simpler ones. In fact, any circuit that, when connected to the array and the load, produces the same load current, will serve just as well. The load current, from (63), is

$$I_o = \sum_{k=1}^{N} Y_{ok} V_k' \quad .$$

The admittances and voltages pertain to the disconnected sources and their coupling network. In terms of the transmission coefficients, (65),

$$I_o = \frac{1}{2\sqrt{R_o}} \sum_{k=1}^{N} t_k \frac{1}{\sqrt{D_k}} V_k' \quad ,$$

or in matrix notation,

$$I_o = \frac{1}{2\sqrt{R_o}} \underline{t} \times \left[\frac{1}{\sqrt{D}}\right] \times V'] \quad .$$

Finally, using (83),

$$I_o = \frac{1}{2\sqrt{R_o}}\ \underline{t} \times \left[\frac{1}{\sqrt{D}}\right] \times [b] \times V] \quad . \tag{87}$$

The load current has been expressed in terms of $\underline{t}$ (the normalized solution vector of the characteristic value problem), the source resistances of the separate sources, the matrix of turns ratios that diagonalizes the resistance matrix of the array, and the open-circuit array element voltages. Any circuit that delivers a load current that is the same weighted sum of open-circuit array element voltages will deliver the same maximum power as the circuits we have developed. This new circuit must provide the same weighting coefficients (transfer admittances) as appear in (87).

The calculation of the transfer admittances for the new circuit can also be simplified; it is not necessary to go through the intermediate step of calculating the matrix [b] that diagonalizes the resistance matrix, as we shall now see. The characteristic value problem for the array consists of determining the set of transmission coefficients $\underline{t}$ that satisfies $\underline{t} \times [\phi] = \alpha \underline{t}$ with the greatest characteristic value α_m. With $[\phi]$ given by (85), the equation reads

$$\underline{t} \times \left[\frac{1}{\sqrt{D}}\right] \times [b] \times \{\tfrac{1}{8}\ \overline{V] \times \underline{V^*}}\} \times [b]_t \times \left[\frac{1}{\sqrt{D}}\right] = \alpha \underline{t} \quad ,$$

and, after post-multiplying by $\left[\frac{1}{\sqrt{D}}\right]^{-1} \times [b]_t^{-1}$,

$$\left\{\underline{t} \times \left[\frac{1}{\sqrt{D}}\right] \times [b]\right\} \times \{\tfrac{1}{8}\ \overline{V] \times \underline{V^*}}\} = \alpha \underline{t} \times \left[\frac{1}{\sqrt{D}}\right]^{-1} \times [b]_t^{-1} \quad .$$

If the right-hand side is post-multiplied by

$$[1] = [R]^{-1} \times [R] = [b]_t \times [D]^{-1} \times [b] \times [R] \quad ,$$

from (84a), we have

$$\left\{\underline{t} \times \left[\frac{1}{\sqrt{D}}\right] \times [b]\right\} \times \{\tfrac{1}{8}\ \overline{V] \times \underline{V^*}}\} = \alpha \left\{\underline{t} \times \left[\frac{1}{\sqrt{D}}\right] \times [b]\right\} \times [R] \quad ,$$

which may be written

$$\underline{y} \times \{\tfrac{1}{8}\ \overline{V] \times \underline{V^*}}\} = \alpha \underline{y} \times [R] \quad , \tag{88}$$

in which

$$\underline{y} = \underline{t} \times \left[\frac{1}{\sqrt{D}}\right] \times [b] \quad ; \quad \underline{t} = \underline{y} \times [b]^{-1} \times \left[\frac{1}{\sqrt{D}}\right]^{-1} \quad .$$

The constraint on the transmission coefficients, $\underline{t} \times t^*] = 1$, becomes

$$\underline{y} \times [b]^{-1} \times \left[\frac{1}{\sqrt{D}}\right]^{-1} \times \left[\frac{1}{\sqrt{D}}\right]^{-1} \times [b]_t^{-1} \times y^*] = 1 \quad ,$$

and, with the aid of (84a), it may be written

$$\underline{y} \times [R] \times y^*] = 1 \quad . \tag{89}$$

The load current, (87), written in terms of the transfer admittances, is simply

$$I_o = \frac{1}{2\sqrt{R_o}}\ \underline{y} \times V] \quad . \tag{90}$$

In designing an optimum coupling network for an array of coupled sources, therefore, the characteristic value problem, (88), should be solved for the characteristic vector y], corresponding to the maximum characteristic value α_m, which is the maximum available power. The characteristic vector should be normalized in such a way that the constraint, (89), is satisfied; as before, this operation will fix the magnitudes of the y's, but leave a common phase shift undetermined. Finally, a coupling network must be found that will deliver the load current, (90), for any set V] of open-circuit array element voltages; the transfer admittances are the set

$$\frac{1}{2\sqrt{R_o}}\ \underline{y} \quad .$$

The characteristic value problem, (88), is more general than the one we were originally led to. Nevertheless, if (88) is post-multiplied by $[R]^{-1}$:

$$\underline{y} \times \{\tfrac{1}{8}\ \overline{V] \times \underline{V^*}} \times [R]^{-1}\} = \alpha \underline{y} \quad ,$$

it becomes formally identical with (74a) and may be solved by the same methods. The matrix $\{\frac{1}{8}\ \overline{V] \times \underline{V^*}} \times [R]^{-1}\}$, however, is generally not Hermitian, and it may be more convenient to solve (88) directly for its maximum characteristic value and corresponding vector. The fact that $\overline{V] \times \underline{V^*}}$ is non-negative definite Hermitian and [R] is positive definite symmetric is sufficient to insure that the characteristic values are real, finite, and non-negative. This more general characteristic value problem is also well known and may be solved by iteration.[13]

One of the simplest forms the coupling network may take is shown in Fig. 22. The box labeled θ_i is a transmission line of electrical length θ_i radians and having any convenient

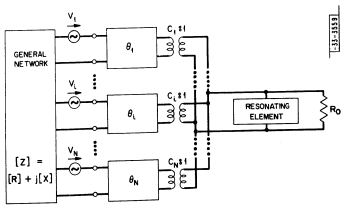

Fig. 22. A simple realization of the optimum coupling network for a set of connected sources.

characteristic impedance; it may be simply an extension of the feeder from the i^{th} array element. The ideal transformers may be realized approximately by appropriate connections to a single length of resonant transmission line. And if this resonant line is cut long (or short, as required), it will also perform the function of the resonating element. The design of this circuit is straightforward when the characteristic value problem has been solved, making the coefficients in (90) available.

* * *

Before closing this section, we will choose a simple array and show that the power available from it, when it is subjected to noise that arrives uniformly from all directions, is independent of the connections in the coupling network that connects the array to its load. For this to be the case, the cross-correlations between the open-circuit voltages $\overline{V_i\, V_k^*}$ of the disconnected sources on the right in Fig. 21 must be zero, and the same noise power must be available from each source. In other words,

$$[\phi] \propto [1] \quad .$$

Alternatively, as may be seen from (88), it must be that

$$\overline{[V] \times \underline{V}^*} \propto [R] \quad ,$$

which we will demonstrate.

As an array, we will choose a pair of short vertical half-dipoles just above a perfectly conducting horizontal plane and driven between their lower ends and the plane. The dipoles will be illuminated with noise which arrives uniformly from all directions in the upper hemisphere. If the dipoles have effective heights of a_1 and a_2 meters, and if they are separated by r wavelengths, the matrix of their open-circuit driving-point and transfer resistances is[15]

$$[R] = \frac{2\pi\eta}{\Lambda^2} \begin{bmatrix} a_1^2\, f(0) & a_1 a_2\, f(r) \\ a_1 a_2\, f(r) & a_2^2\, f(0) \end{bmatrix} ,$$

where

$$f(r) = \frac{\sin 2\pi r}{2\pi r} + \frac{\cos 2\pi r}{(2\pi r)^2} - \frac{\sin 2\pi r}{(2\pi r)^3} \quad ,$$

$$f(0) = \lim_{r \to 0} f(r) = 2/3 \quad .$$

The open-circuit voltage of one of the dipoles is proportional to the z-component of the electric field at the dipole. The cross-correlation of open-circuit voltages is therefore proportional to the cross-correlation of z-fields, which, as is suggested by (34), may be calculated as the Fourier transform of the rectangular power density ψ_{azz}. From (33) and (10),

$$\psi_{azz}(\ell, m) = \frac{\ell^2 + m^2}{n} \int\int_{-\infty}^{\infty} \frac{1}{n}\, \overline{F_{Va}(\ell, m)\, F_{Va}^*(\ell + \lambda,\ m + \mu)}\, d\lambda\, d\mu \quad .$$

The integral, as may be seen by comparing it with (31), is just $\Phi_{VV}(\ell, m)$, the density of arriving vertically polarized noise power. As we discovered in Sec. IV, this quantity is a constant for the real directions in the case of uniform noise illumination; we shall again call it Φ_o:

$$\psi_{azz}(\ell, m) = \begin{cases} \dfrac{\ell^2 + m^2}{n}\, \Phi_o \quad , & \ell^2 + m^2 \leqslant 1 \quad , \\ 0 \quad , & \ell^2 + m^2 > 1 \quad . \end{cases}$$

The cross-correlation of z-fields is

$$\phi_{azz}(x, y) = \int\int_{-\infty}^{\infty} \psi_{azz}(\ell, m)\, e^{j2\pi(x\ell + ym)}\, d\ell\, dm$$

$$= \Phi_o \int\int_{u.h.} \frac{\ell^2 + m^2}{\sqrt{1 - \ell^2 - m^2}}\, e^{j2\pi(x\ell + ym)}\, d\ell\, dm \quad .$$

The integral is again best evaluated in spherical coordinates with the aid of Watson.[9] With $\ell = \rho \cos\Theta$, $m = \rho \sin\Theta$, it becomes

$$\phi_{azz}(x, y) = \Phi_o \int_o^1 \rho\, d\rho \int_o^{2\pi} d\Theta\, \frac{\rho^2}{\sqrt{1 - \rho^2}}\, e^{j2\pi\rho(x \cos\Theta + y \sin\Theta)}$$

$$= 2\pi\Phi_o \int_o^1 \frac{\rho^3}{\sqrt{1 - \rho^2}}\, J_o\left(2\pi\rho\sqrt{x^2 + y^2}\right) d\rho \quad .$$

The quantity $\sqrt{x^2 + y^2}$ is the separation r of the two dipoles. With $\rho = \sin\phi$, the integral becomes

$$\phi_{azz}(r) = 2\pi\Phi_o \int_o^{\pi/2} J_o(2\pi r \sin\phi) \sin^3\phi\, d\phi \quad ,$$

$$= 2\pi\Phi_o \int_o^{\pi/2} [J_o(2\pi r \sin\phi) \sin\phi - J_o(2\pi r \sin\phi) \sin\phi \cos^2\phi]\, d\phi$$

$$= 2\pi\Phi_o \left[\frac{2^{-1/2}\Gamma(1/2)}{(2\pi r)^{1/2}}\, J_{1/2}(2\pi r) - \frac{2^{1/2}\Gamma(3/2)}{(2\pi r)^{3/2}}\, J_{3/2}(2\pi r) \right]$$

$$= 2\pi\Phi_o\, f(r) \quad .$$

Allowing for reflections from the conducting plane, the matrix of cross-correlations of open-circuit voltages of the dipoles is

$$\overline{\underline{V}\,|\times|\,\underline{V}^*} = 8\pi\Phi_o \begin{bmatrix} a_1^2\, f(0) & a_1 a_2\, f(r) \\ a_1 a_2\, f(r) & a_2^2\, f(0) \end{bmatrix},$$

which is indeed proportional to their resistance matrix, as expected. We have shown that the noise power available from this particular array is independent of the connections in the (lossless) coupling network connecting it to the load.

Incidentally, the foregoing argument suggests an unusual way of calculating the real part of the mutual impedance between two antennas.

VII. OPTIMIZATION OF A CONTINUOUS ANTENNA

In this section, we will discover how to adjust a receiving antenna of the type described in Sec. III in such a way that maximum signal power is available at its terminals. Instead of adjusting a discrete set of transmission coefficients, as was done in Sec. VI, we will allow the adjustment of the electric field produced at each point on the aperture plane of the antenna when a unit current is applied to its terminals. In contrast to the array considered in Sec. VI, the antenna considered here is described by a continuous field distribution; it may be called a continuous antenna.

There is a pitfall in the path of the present investigation that we will dispense with at the start. Even if the region of space containing the active conductors of the antenna were specified – in fact, even if this region had maximum dimensions small compared with a wavelength – it would be generally possible to fit into this region an antenna having an arbitrarily sharp pattern! An antenna having a radian beamwidth small compared with the reciprocal of its largest aperture dimension, expressed in wavelengths, (or having a small beamwidth if the largest aperture dimension is small) is known as a superdirective antenna.[16] Such an antenna is undesirable because its adjustment is critical and its losses are relatively high. We will neglect antenna losses, and we will assume that the optimum configuration we arrive at analytically can be duplicated exactly in practice. Since, as we discovered in Sec. IV, the greatest signal power is available from the antenna having the greatest gain in the direction of maximum power density, we might expect our maximization of available signal power to lead to a superdirective antenna having infinite gain. This undesirable result will be prevented by appropriately restricting the construction of the antenna.

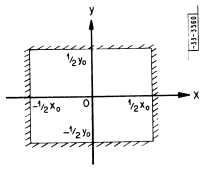

Fig. 23. The antenna is a rectangular hole in an infinite, perfectly conducting plane.

The antenna to be considered in this section consists of a rectangular hole in an infinite, perfectly conducting plane, which will be taken as the xy-plane. The origin of coordinates will be centered at the mouth of the hole, as shown in Fig. 23. The aperture dimensions, which are the length and width of the hole, are x_o wavelengths in the x-direction and y_o wavelengths in the y-direction. In some manner, which we will not investigate, electric fields will be produced on the xy-plane at the mouth of the hole when a generator is connected to the antenna terminals. It will be assumed that all the power from this generator is radiated into the upper hemisphere.

The gain of the antenna, the ratio of (15) to (16), is

$$\frac{n^2\left[|P_x(\ell,m)|^2 + |P_y(\ell,m)|^2 + |P_z(\ell,m)|^2\right]}{\iint_{u.h.} n\left[|P_x(\ell,m)|^2 + |P_y(\ell,m)|^2 + |P_z(\ell,m)|^2\right] d\ell\, dm}.$$

The conducting plane imposes the boundary condition that the tangential electric fields $E_x(x,y,o)$ and $E_y(x,y,o)$ are zero except for (x,y) inside the aperture. The rectangular plane-wave

components $P_x(\ell, m)$ and $P_y(\ell, m)$, which are the Fourier transforms of the tangential fields, must therefore be well-behaved functions of direction. Although we would expect $P_x(\ell, m)$ to have a peak in the direction of maximum power density, the sharpness of that peak is limited by the aperture size. (It is easily shown, for example, that

$$\frac{|P_x(\ell, m)|^2}{\iint_{-\infty}^{\infty} |P_x(\ell, m)|^2 \, d\ell \, dm}$$

is less than or equal to the aperture area expressed in square wavelengths.) The third rectangular plane-wave component, $P_z(\ell, m)$, is linked to the other two by the divergence relation,

$$\ell P_x + m P_y + n P_z = 0 \quad . \tag{10a}$$

Except possibly in the directions, n = 0, corresponding to the horizon, $P_z(\ell, m)$ will also be a well-behaved function of direction; its peak will again be limited in sharpness by the aperture dimensions. In other words, regardless of how the antenna is adjusted, its gain in any direction, except possibly toward the horizon, will remain finite.

Because of the divergence relation (10a), only two of the three field components may be adjusted independently. It will be convenient to work with $E_x(x, y, o)$ and $E_y(x, y, o)$ or their transforms, $P_x(\ell, m)$ and $P_y(\ell, m)$. The rectangular shape of the aperture has been chosen to simplify the calculations. Since either tangential field, $E_i(x, y, o)$, is zero outside the rectangular aperture, its transform, $P_i(\ell, m)$, according to the Sampling Theorem, is completely determined by a set of samples taken on a grid of points in the ℓm-plane. Rows of sampling points are spaced by $1/x_o$ in the ℓ-direction and by $1/y_o$ in the m-direction:

$$P_i(\ell, m) = \sum_{r,\,s} P_i(\ell_o + r/x_o, m_o + s/y_o) \frac{\sin \pi[x_o(\ell - \ell_o) - r]}{\pi[x_o(\ell - \ell_o) - r]} \frac{\sin \pi[y_o(m - m_o) - s]}{\pi[y_o(m - m_o) - s]} \quad , \tag{91}$$

and the Fourier series representing the corresponding tangential field is

$$E_i(x, y, o) = \begin{cases} \dfrac{1}{x_o y_o} \displaystyle\sum_{r,\,s} P_i(\ell_o + r/x_o, m_o + s/y_o) \, e^{-j2\pi[x(\ell_o + r/x_o) + y(m_o + s/y_o)]} \\ \qquad \text{for} \quad -1/2\, x_o < x < 1/2\, x_o \quad \text{and} \quad -1/2\, y_o < y < 1/2\, y_o \\ 0 \text{ otherwise} \quad . \end{cases} \tag{92}$$

These relations may be derived in the same way as (59) and (60) were derived; in fact, they may be inferred from (59) and (60). The constants ℓ_o and m_o play the same role as the constants x_o and y_o did in (59) and (60); they may be chosen arbitrarily to displace the sampling grid with respect to the origin in the ℓm-plane (the zenith). It will often be convenient to pick the direction (ℓ_o, m_o) to coincide with the direction of maximum power density.

In maximizing the available signal power, we will adjust the coefficients, $P_x(\ell_o + r/x_o, m_o + s/y_o)$ and $P_y(\ell_o + r/x_o, m_o + s/y_o)$. Strictly speaking, there is a doubly infinite set of each but, as a practical matter, only a finite number of points $(\ell_o + r/x_o, m_o + s/y_o)$ on the

sampling grid fall within $\ell^2 + m^2 = 1$: Only a finite number of points correspond to real directions. Changing the coefficients corresponding to samples taken very far outside of the unit circle will not have much effect on the antenna pattern in real directions, as can be seen from (91). Practically then, we may be content to adjust a finite number of coefficients.

We will start with the expression (30) for the available power, which we will write in matrix notation as

$$W_a = \frac{\Lambda^2}{2\eta} \frac{\int_{\text{u.h.}} d\Omega \{\underline{F} \times [\Phi] \times F^*\}}{\int_{\text{u.h.}} d\Omega \{\underline{F} \times F^*\}} \quad ,$$

in which $[\Phi] = [\Phi(\ell, m)]$ is the matrix of power densities (32), and

$$F] = F(\ell, m)] = \begin{bmatrix} F_H(\ell, m) \\ F_V(\ell, m) \end{bmatrix} \quad .$$

The next step is to write the available power in terms of the x- and y-plane-wave components:

$$P] = P(\ell, m)] = \begin{bmatrix} P_x(\ell, m) \\ P_y(\ell, m) \end{bmatrix} \quad . \tag{93}$$

As suggested by (39), this may be accomplished with the aid of the transformation matrix [T], given by (38). The result is

$$W_a = \frac{\Lambda^2}{2\eta} \frac{\int_{\text{u.h.}} d\Omega \{\underline{P} \times [T]_t^{-1} \times [\Phi] \times [T]^{-1} \times P^*\}}{\int_{\text{u.h.}} d\Omega \{\underline{P} \times [T]_t^{-1} \times [T]^{-1} \times P^*\}} \quad .$$

If the square matrix in the numerator is called

$$[N] = [T]_t^{-1} \times [\Phi] \times [T]^{-1} \quad , \tag{94}$$

and the corresponding denominator matrix is called† .

†It may be seen from (40) that the numerator matrix is not the matrix of rectangular power densities used in Sec. V, but may be expressed in terms of that matrix as

$$[N] = \frac{1}{n} [D] \times [\theta_a] \times [D] \quad .$$

$$[D] = [T]_t^{-1} \times [T]^{-1} = \begin{bmatrix} 1-m^2 & \ell m \\ \ell m & 1-\ell^2 \end{bmatrix} , \tag{95}$$

the available power may be written

$$W_a = \frac{A^2}{2\eta} \frac{\int_{u.h.} d\Omega \{\underline{P} \times [N] \times P^*\}}{\int_{u.h.} d\Omega \{\underline{P} \times [D] \times P^*\}} . \tag{96}$$

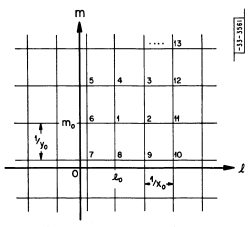

Fig. 24. Renumbering the sampling grid.

In order to express the available power in terms of the samples of the antenna pattern, it will be convenient to renumber the points on the sampling grid consecutively, using a scheme such as is shown in Fig. 24. In that figure, the point $\ell = \ell_o$, $m = m_o$, corresponding to $(r, s) = (0, 0)$, has been given the new number 1; the new number 2 has been given to the point $\ell = \ell_o + 1/x_o$, $m = m_o$, which was originally numbered $(r, s) = (1, 0)$. The new numbering spirals outward until all points on the grid have been assigned a number. If the new number k corresponds to the original numbering (r, s), the sum (94) may be rewritten

$$P_i(\ell, m) = \sum_{k=1}^{\infty} p_{ik} f_k , \tag{94a}$$

in which

$$p_{ik} = P_i(\ell_o + r/x_o,\ m_o + s/y_o) \tag{94b}$$

and

$$f_k = \frac{\sin \pi[x_o(\ell - \ell_o) - r]}{\pi[x_o(\ell - \ell_o) - r]} \frac{\sin \pi[y_o(m - m_o) - s]}{\pi[y_o(m - m_o) - s]} . \tag{94c}$$

The integrand in the numerator of (96) presently reads

$$\underline{P_x, P_y} \times \begin{bmatrix} N_{xx} & N_{xy} \\ N_{yx} & N_{yy} \end{bmatrix} \times \begin{bmatrix} P_x^* \\ P_y^* \end{bmatrix} ,$$

all quantities being functions of (ℓ, m). If the matrices involved are extended, the integrand may be written in terms of the pattern samples, (91b), as

$$\underline{p} \times [N] \times p^*$$

in which p is the column matrix of pattern samples,

$$p = \begin{bmatrix} p_{x1} \\ p_{x2} \\ \vdots \\ p_{y1} \\ p_{y2} \\ \vdots \end{bmatrix} , \tag{97}$$

and

$$[N] = \left[\begin{array}{cc|cc} N_{xx}f_1f_1 & N_{xx}f_1f_2 \cdots & N_{xy}f_1f_1 & N_{xy}f_1f_2 \cdots \\ N_{xx}f_2f_1 & N_{xx}f_2f_2 \cdots & N_{xy}f_2f_1 & N_{xy}f_2f_2 \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \hline N_{yx}f_1f_1 & N_{yx}f_1f_2 \cdots & N_{yy}f_1f_1 & N_{yy}f_1f_2 \cdots \\ N_{yx}f_2f_1 & N_{yx}f_2f_2 \cdots & N_{yy}f_2f_1 & N_{yy}f_2f_2 \cdots \\ \vdots & \vdots & \vdots & \vdots \end{array}\right] . \tag{98}$$

The extra marks on the matrix brackets indicate the partitioning shown in (97) and (98). If the denominator integrand is extended in the same way, the available power may be written

$$W_a = \frac{\Lambda^2}{2\eta} \frac{\int_{u.h.} d\Omega \{ \underset{\sim}{p} \times [\underset{\sim}{N}_i] \times p^* \}}{\int_{u.h.} d\Omega \{ \underset{\sim}{p} \times [\underset{\sim}{D}_i] \times p^* \}} \quad .$$

In this form, only the square matrices $[\underset{\sim}{N}_i]$ and $[\underset{\sim}{D}_i]$ are functions of direction. In terms of

$$[\underset{\sim}{G}_i] = \int_{u.h.} [\underset{\sim}{N}_i] \, d\Omega \quad \text{and} \quad [\underset{\sim}{H}_i] = \int_{u.h.} [\underset{\sim}{D}_i] \, d\Omega \quad , \tag{99}$$

the available power may finally be written as

$$W_a = \frac{\Lambda^2}{2\eta} \frac{\underset{\sim}{p} \times [\underset{\sim}{G}_i] \times p^*}{\underset{\sim}{p} \times [\underset{\sim}{H}_i] \times p^*} \quad . \tag{100}$$

It will be recalled from Sec. IV that the denominator of the expression for the available power is proportional to the radiation resistance of the antenna, and the numerator is proportional to the mean-square open-circuit voltage. The same is true of (100). This means that any matrix $[\underset{\sim}{H}_i]$ of finite order is positive definite Hermitian, so that its inverse exists, and the matrix $[\underset{\sim}{G}_i]$ of like order is at least non-negative definite Hermitian. [The denominator matrix $[\underset{\sim}{H}_i]$ is actually symmetric, as may be seen from (95).] The expression (100) may be maximized by holding the denominator constant and maximizing the numerator. (This amounts to maximizing the mean-square open-circuit voltage while holding the radiation resistance constant.) If a procedure similar to the one leading to (74) is used, the condition for maximum available power is found to be

$$\underset{\sim}{p} \times [\underset{\sim}{G}_i] = \alpha \, \underset{\sim}{p} \times [\underset{\sim}{H}_i] \quad , \tag{101}$$

where α is again the Lagrangian multiplier. Post-multiplying both sides of (101) by p^*, we have

$$\underset{\sim}{p} \times [\underset{\sim}{G}_i] \times p^* = \alpha \, \underset{\sim}{p} \times [\underset{\sim}{H}_i] \times p^* \quad ,$$

so that the maximum available power is just $(\Lambda^2/2\eta)\,\alpha$. The problem of solving (101) for its maximum characteristic value and corresponding vector (the set of pattern samples) is essentially identical with the general characteristic value problem (88) discussed in Sec. VI. The main difference is that we have not shown a constraint corresponding to (89). It will be noted that if p is multiplied by a constant in (100), that expression is not changed, so no constraint is necessary. On the other hand, if we had stated the available power to be

$$W_a = \frac{\Lambda^2}{2\eta} \, \underset{\sim}{p} \times [\underset{\sim}{G}_i] \times p^*$$

if

$$\underset{\sim}{p} \times [\underset{\sim}{H}_i] \times p^* = 1 \quad ,$$

the solution vector resulting from (101) would have to be normalized to satisfy the above constraint. These are just two equivalent ways of stating the problem.

It is difficult to make a precise statement about the order of the G- and H-matrices (the number of pattern samples) that need be considered. In general, if the pattern samples included in the maximization are taken at all points on the sample grid from which appreciable power density arrives, the indicated maximum available power cannot be increased appreciably by including more points in the maximization. The fact that the antenna gain is bounded insures that, except in the case of a point source located on the horizon, there will be a limit to the available power regardless of the number of pattern samples included in the maximization.

* * *

The foregoing maximization is complicated by the fact that both the optimum pattern shape and the optimum polarization of the receiving antenna are determined simultaneously. As we discovered, it is really necessary to proceed in this manner if a general power-density matrix is under consideration. On the other hand, if the transmitting antenna is horizontally polarized, for example, it would appear to be obvious that the receiving antenna should also be horizontally polarized. We will next investigate the validity of such conjecture by seeking the conditions under which the optimum antenna, when driven at its terminals, produces tangential fields $E_x(x, y, o)$ and $E_y(x, y, o)$ that are in a fixed ratio, independent of the location of the point (x, y, o) on the aperture plane. Equivalently, we ask under what conditions the optimum antenna has the pattern $P_y(t, m) = a\, P_x(t, m)$, with a some complex constant.

If the power densities are sufficiently well-behaved, as they will be in practice, the pattern

$$P] = P_x(t, m) \begin{bmatrix} 1 \\ a \end{bmatrix} \tag{102}$$

must maximize the ratio of the numerator integrand to the denominator integrand of (96) in each direction (t, m), for if it did not, more power would be available if $P_y(t, m) \neq a\, P_x(t, m)$. In other words, if the optimum antenna is to have the pattern $P_y(t, m) = a\, P_x(t, m)$, we must have

$$\{P_x(t, m) \cdot \underline{1, a}\} \times [N(t, m)] = \beta(t, m) \{P_x(t, m) \cdot \underline{1, a}\} \times [D(t, m)] \quad , \tag{103}$$

or simply

$$\underline{1, a} \times [N(t, m)] = \beta(t, m) \, \underline{1, a} \times [D(t, m)] \quad .$$

We have written the condition under which (102) maximizes the ratio of the numerator integrand to the denominator integrand of (96). Note that the Lagrange multiplier β depends on the direction (t, m). Post-multiplying by the transformation matrix $[T]$, given by (38), and using (94) and (95), we have

$$\{\underline{1, a} \times [T]_t^{-1}\} \times [\Phi] = \beta \{\underline{1, a} \times [T]_t^{-1}\} \quad ,$$

so that the Lagrange multiplier is a characteristic value of the power-density matrix in each

direction of space. The corresponding characteristic vector (consisting of a pair of horizontal and vertical plane-wave components, incidentally) is proportional to

$$\underline{1, a} \times [T]_t^{-1} = \frac{-1}{\sqrt{\ell^2 + m^2}} \underline{n(m - a\ell),\ (\ell + am)}$$

and may be taken to be

$$\underline{n(m - a\ell),\ (\ell + am)} \quad .$$

We seek the conditions under which

$$\underline{n(m - a\ell),\ (\ell + am)} \times \begin{bmatrix} \Phi_{HH} - \beta & \Phi_{HV} \\ \Phi_{VH} & \Phi_{VV} - \beta \end{bmatrix} = 0 \quad ,$$

with β an arbitrary (real, positive) function of (ℓ, m) and $\underline{a}$ an arbitrary (complex) constant. If β is eliminated from the two equations, there results, after some algebra,

$$(n^2 |m - a\ell|^2 - |\ell + am|^2)\ \Phi_{HV} = n(m - a^*\ell)\ (\ell + am)\ (\Phi_{HH} - \Phi_{VV}) \quad , \qquad (104)$$

the desired relation among the power densities that allows the polarization of the receiving antenna to be fixed in such a way that $P_y(\ell, m) = a\, P_x(\ell, m)$ before the pattern shape is determined.

The condition (104) does not impose a particularly severe restriction on the power-density matrix. It is obviously satisfied by uniform noise, for which $\Phi_{HV} = 0$ and $\Phi_{HH} = \Phi_{VV}$: This was, of course, to be expected, since the available noise power does not depend on the adjustment of the antenna. (104) is also satisfied by the power-density matrix

$$[\Phi(\ell, m)] = \Phi(\ell, m) \begin{bmatrix} n^2 |m - a\ell|^2 & n(m - a^*\ell)\ (\ell + am) \\ n(m - a\ell)\ (\ell + am^*) & |\ell + am|^2 \end{bmatrix} ,$$

in which $\Phi(\ell, m)$ is an arbitrary real positive function of direction and $\underline{a}$ is an arbitrary complex constant. It will be noted that this matrix is singular, so that the polarization of the plane wave arriving from each direction (ℓ, m) is also fixed in time. We have given examples of extreme cases in which (104) is satisfied.

If (104) is satisfied, so that the antenna pattern is of the form (102), we have, after post-multiplying (103) by P*],

$$\begin{aligned} \underline{P} \times [N] \times P^*] &= \beta(\ell, m)\ \underline{P} \times [D] \times P^*] \\ &= \beta(\ell, m)\ |P_x(\ell, m)|^2\ \underline{1, a} \times [D] \times \begin{bmatrix} 1 \\ a^* \end{bmatrix} \\ &= \beta(\ell, m)\ |P_x(\ell, m)|^2\ A(\ell, m) \quad , \end{aligned}$$

in which

$$\begin{aligned} A(\ell, m) &= \underline{1, a} \times [D] \times \begin{bmatrix} 1 \\ a^* \end{bmatrix} \\ &= (1 - m^2) + 2\ell m\, Re\{a\} + |a|^2\ (1 - \ell^2) \quad , \end{aligned} \qquad (105)$$

so that the power available with optimum polarization, as expressed by (96), becomes

$$W_a = \frac{\Lambda^2}{2\eta} \frac{\int_{u.h.} \beta(\ell, m)\ |P_x(\ell, m)|^2\ A(\ell, m)\ d\Omega}{\int_{u.h.} |P_x(\ell, m)|^2\ A(\ell, m)\ d\Omega} \quad , \qquad (106)$$

with $\beta(\ell, m)$ the greater of the two characteristic values of $[\Phi(\ell, m)]$ in each direction of space. At this point, it should be clear that we are justified in regarding $\beta(\ell, m)$, or more properly $(\Lambda^2/2\eta)\ \beta(\ell, m)$, as the power density.

Having determined the polarization of the antenna, it remains to maximize the available power by adjusting the shape of its pattern. We need work only with the x-samples, $P_x(\ell_o + r/x_o, m_o + s/y_o)$. If these are numbered consecutively as before to form the set

$$p] = \begin{bmatrix} p_{x1} \\ p_{x2} \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \quad , \qquad (107)$$

and if we use the column matrix of corresponding f's given by (91c),

$$f] = \begin{bmatrix} f_1 \\ f_2 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \quad , \qquad (108)$$

then, from (91a), we can write

$$|P_x(\ell, m)|^2 = \underline{p} \times f] \times \underline{f} \times p^*] \quad .$$

In terms of

$$\left.\begin{aligned} [G] &= \int_{u.h.} \beta(\ell, m)\ f] \times \underline{f}\ A(\ell, m)\ d\Omega \\ \text{and} \qquad [H] &= \int_{u.h.} f] \times \underline{f}\ A(\ell, m)\ d\Omega \end{aligned}\right\} \quad , \qquad (109)$$

the power available with optimum polarization (106) can finally be written as

$$W_a = \frac{\lambda^2}{2\eta} \frac{\underline{p} \times [G] \times \underline{p}^*]}{\underline{p} \times [H] \times \underline{p}^*]} \quad . \tag{110}$$

As before, the condition for maximum available power is

$$\underline{p} \times [G] = \alpha \, \underline{p} \times [H] \quad , \tag{111}$$

and that maximum power is just $(\lambda^2/2\eta)\ \alpha$. It may be noted that both integrands (109) are real, so that both matrices [G] and [H] are now real. This means that the characteristic vector p] will also be real; the conjugate signs may be removed from (110).

* * *

A considerable simplification of the problem resulted when we were able to determine the optimum polarization of the antenna before finding its pattern shape; the order of the G- and H-matrices was cut in half. Further simplification is possible if the aperture dimensions of the antenna are large compared with the wavelength. In that case, the $\frac{\sin x}{x}$ -functions f_k given by (91c) vary rapidly with direction and tend to bunch in one direction. Let us examine the denominator integral (109), giving the element of [H] in the i^{th} row and k^{th} column. Let the original numbering (r, s) correspond to i and the original numbering (p, q) correspond to k:

$$H_{ik} = \iint_{u.h.} \frac{\sin \pi[x_o(\ell - \ell_o) - r]}{\pi[x_o(\ell - \ell_o) - r]} \frac{\sin \pi[y_o(m - m_o) - s]}{\pi[y_o(m - m_o) - s]}$$

$$\times \frac{\sin \pi[x_o(\ell - \ell_o) - p]}{\pi[x_o(\ell - \ell_o) - p]} \frac{\sin \pi[y_o(m - m_o) - q]}{\pi[y_o(m - m_o) - q]} A(\ell, m) \frac{d\ell\, dm}{n} \quad .$$

For large x_o and y_o, the behavior of the integrand is largely determined by the four $\frac{\sin x}{x}$ -functions, so that the remaining factor, $A(\ell, m)/n$, may be removed from the integral after being set equal to its value at the point of maximum contribution of the $\frac{\sin x}{x}$ -functions. (The fact that $1/n \to \infty$ at the horizon does not cause any difficulty, since

$$\iint_{u.h.} \frac{d\ell\, dm}{n} = \int_{u.h.} d\Omega = 2\pi$$

exists.) In addition, the range of integration may be extended to cover the entire ℓm-plane. The double integral may then be split into the product of two single integrals of the form

$$\int_{-\infty}^{\infty} \frac{\sin \pi[x_o(\ell - \ell_o) - r]}{\pi[x_o(\ell - \ell_o) - r]} \frac{\sin \pi[x_o(\ell - \ell_o) - p]}{\pi[x_o(\ell - \ell_o) - p]} d\ell = \begin{cases} \frac{1}{x_o} & \text{if} \quad p = r \\ 0 \text{ otherwise} & \end{cases} .$$

The denominator integral under consideration is therefore approximated by

$$H_{ik} \approx \begin{cases} \frac{1}{x_o y_o} \frac{A(\ell_o + r/x_o,\ m_o + s/y_o)}{\sqrt{1 - (\ell_o + r/x_o)^2 - (m_o + s/y_o)^2}} \\ \qquad \text{if} \quad k = i \quad \text{and} \quad (\ell_o + r/x_o)^2 + (m_o + s/y_o)^2 < 1 \\ 0 \text{ otherwise} \quad , \end{cases} \tag{112}$$

so that the matrix [H] of (109) becomes diagonal and of finite order. The same approximations, when applied to the more general denominator integral (99), result in a matrix of finite order having three non-zero diagonals:

$$\{H\} = \frac{1}{x_o y_o} \left[\begin{array}{cccc|cccc} B_1 & 0 & \cdots & & C_1 & 0 & \cdots & \\ 0 & B_2 & \cdots & & 0 & C_2 & \cdots & \\ \cdot & \cdot & & & \cdot & \cdot & & \\ \cdot & \cdot & & & \cdot & \cdot & & \\ \cdot & \cdot & & & \cdot & \cdot & & \\ \hline C_1 & 0 & \cdots & & D_1 & 0 & \cdots & \\ 0 & C_2 & \cdots & & 0 & D_2 & \cdots & \\ \cdot & \cdot & & & \cdot & \cdot & & \\ \cdot & \cdot & & & \cdot & \cdot & & \\ \cdot & \cdot & & & \cdot & \cdot & & \end{array}\right] . \tag{113}$$

It is interesting to note that these same approximations can be deduced from a physical argument, which, although more difficult to justify than the preceding mathematical argument, provides a valuable insight into the meaning of the approximations. The denominator of the expression for available power, we recall, is proportional to the radiation resistance of the antenna, or, originally, the power emitted by the antenna when a unit current is applied to its terminals. The approach leading to the approximations consists of calculating the transmitted power from (17), using the Fourier series (92) for the fields.

Strictly speaking, (92) applies only to the tangential fields, $E_x(x, y, o)$ and $E_y(x, y, o)$, and only on the aperture plane of the antenna. On the other hand, any one of the terms of the sum behaves like a component of a plane wave over the aperture. It would seem reasonable that the normal field, $E_z(x, y, o)$, should behave like the third component of a plane wave, the tangential components of which are given by (92), at least we would expect this behavior of the normal field at points within the aperture that are well removed from the edge. In other words, if the aperture is large enough, (92) is approximately valid for all three field components.

In order to make use of (17), it is necessary to calculate the z-derivatives of the field components, which requires a knowledge of the field components as a function of z in the immediate vicinity of the aperture plane. Again it would seem reasonable to extend (92) as though each term represented the component of a plane wave progressing into the upper hemisphere. For $z \approx 0$, then,

$$E_i(x,y,z) \approx \begin{cases} \dfrac{1}{x_o y_o} \sum\limits_{r,s} P_i(\ell_o + r/x_o,\ m_o + s/y_o)\, e^{-j2\pi[x(\ell_o + r/x_o) + y(m_o + s/y_o) + zn]} \\ \qquad \text{for} \quad (x,y) \text{ inside the aperture} \\ 0 \text{ otherwise} \quad . \end{cases}$$

$$n = \sqrt{1 - (\ell_o + r/x_o)^2 - (m_o + s/y_o)^2} \quad .$$

If this approximation is used in calculating (17), the transmitted power becomes, after several steps,

$$W = \frac{\Lambda^2}{2\eta} \frac{1}{x_o y_o} \sum_i Re\left\{ \sum_{r,s} \sqrt{1 - (\ell_o + r/x_o)^2 - (m_o + s/y_o)^2}\ |P_i(\ell_o + r/x_o,\ m_o + s/y_o)|^2 \right\} \quad .$$

Since the radical is pure imaginary outside of the unit circle, only a finite number of terms are involved in the double sum. If the divergence relation (10a) is used to express P_z in terms of P_x and P_y, this expression for the transmitted power becomes identical with the expression that would be obtained with the approximate H-matrix, (113):

$$W = \frac{\Lambda^2}{2\eta} \underline{p} \times \{\dot{H}\} \times p^* \} \quad .$$

In other words, the physical argument leads to the approximation (113) for the H-matrix; it would also produce the diagonal matrix with coefficients given by (112) if the polarization of the antenna were predetermined.

Another way of stating the result of the physical argument is that the transmitted power is calculated as though a collection of true plane waves were crossing the xy-plane and progressing into the upper hemisphere; the perfectly conducting sheet, however, is removed from the xy-plane, and only that power passing through the rectangular outline of the aperture is counted. In any case, it appears that the effects of the conducting edge of the aperture may be neglected if the aperture is large enough. Presumably, the edge effects of other large antennas can also be neglected. A dish, for example, which does not have an aperture bounded by the edge of a hole in an infinite conducting sheet, might be handled using the same approximations. Regardless of the structure of the antenna, its optimization will not make it superdirective if these approximations are used.

The same approximations, when applied to the numerator integrals (99) and (109), lead to numerator matrices $\{\dot{Q}\}$ and [G] of the same order as the corresponding denominator matrices, but the simplification of the integrals is not so spectacular in this case. The numerator matrices do not become diagonal because the power densities, as well as the $\frac{\sin x}{x}$-functions, vary rapidly with direction, preventing the integrands from exhibiting the orthogonality possessed by the $\frac{\sin x}{x}$-functions by themselves.

* * *

Before closing this section, we will choose a representative power-density pattern and optimize the patterns of a number of antennas of different size for this power-density pattern. We are primarily interested in the shapes of the optimum patterns and in how much "gain loss" may be recovered by using the optimum antennas; we are not particularly interested in polarization. Accordingly, we will suppose that (104) is satisfied, so that the polarization may be fixed in such a way that $P_y(\ell, m) = a\, P_x(\ell, m)$. We will choose a power-density pattern centered on (ℓ_o, m_o) and having the Gaussian shape

$$\frac{\Lambda^2}{2\eta} \beta(\ell, m) = \frac{x_1 y_1 n}{\pi} e^{-[x_1^2(\ell-\ell_o)^2 + y_1^2(m-m_o)^2]} \quad . \tag{114}$$

The parameters x_1 and y_1 determine the sharpness of the power-density pattern; we will suppose that they are both so large that most of the power arrives from within a very small cone of directions centered on the real direction (ℓ_o, m_o). The power-density pattern has been normalized in such a way that the total power is

$$W = \int_{u.h.} \frac{\Lambda^2}{2\eta} \beta(\ell, m)\, d\Omega \approx \int_{-\infty}^{\infty}\int \frac{\Lambda^2}{2\eta} \beta(\ell, m)\, \frac{d\ell\, dm}{n}$$

$$= \frac{x_1 y_1}{\pi} \int_{-\infty}^{\infty}\int e^{-[x_1^2(\ell-\ell_o)^2 + y_1^2(m-m_o)^2]}\, d\ell\, dm = 1 \text{ watt} \quad . \tag{115}$$

As yet, we have not given a physical interpretation for this "total power," but one will be forthcoming.

We will choose large antennas having aperture dimensions comparable to x_1 and y_1. The grid of points at which pattern samples are taken will be so fine and the number of pattern samples we will require will be so small that the factor $A(\ell, m)/n$ can not only be removed from the matrix integrals, (109), but it can be set equal to its value in the direction of maximum power density. Under these circumstances, there can be no doubt about the pattern samples all being taken in real directions. The denominator matrix, (109), will be

$$[H] \approx \frac{A(\ell_o, m_o)}{n_o} \int_{-\infty}^{\infty}\int \underline{f}^* \times \underline{f}\, d\ell\, dm = \frac{A(\ell_o, m_o)}{n_o x_o y_o} [1] \quad ,$$

and the numerator matrix becomes

$$[G] \approx \frac{2\eta}{\Lambda^2} A(\ell_o, m_o) \int_{-\infty}^{\infty}\int \left\{\frac{\Lambda^2}{2\eta}\frac{\beta(\ell, m)}{n}\right\} f] \times \underline{f}\, d\ell\, dm \quad .$$

The available power, (110), may be stated to be

$$W_a = n_o x_o y_o \, \underline{p} \times [g] \times p] \tag{116}$$

subject to the constraint

$$\underline{p} \times p] = 1 \tag{117}$$

with

$$[g] = \int_{-\infty}^{\infty}\int \left\{\frac{\Lambda^2}{2\eta}\frac{\beta(\ell, m)}{n}\right\} f] \times \underline{f}\, d\ell\, dm \quad . \tag{118}$$

The maximum available power is therefore just $n_o x_o y_o \alpha_m$, where α_m is the maximum characteristic value of [g].

Before calculating the elements of [g] for the intermediate cases, $x_o \approx x_1$ and $y_o \approx y_1$, we will examine the extremes. First, if the aperture is very large ($x_o >> x_1$ and $y_o >> y_1$), the antenna can resolve considerable detail in the power-density pattern; the maximum available power will be greatest in this case. We will suppose that the pattern samples are taken in directions reasonably close to (ℓ_o, m_o). The behavior of the integrand of (118) is then determined primarily by the $\frac{\sin x}{x}$ -functions, so that

$$[g] \approx \left\{\frac{\Lambda^2}{2\eta}\frac{\beta(\ell_o, m_o)}{n_o}\right\} \int_{-\infty}^{\infty}\int f] \times \underline{f}\, d\ell\, dm$$

$$= \frac{x_1 y_1}{\pi x_o y_o} [1] \quad .$$

In this case, regardless of the antenna pattern [within reason: the pattern should still be concentrated in the vicinity of (ℓ_o, m_o)], the maximum available power will be $n_o x_1 y_1 / \pi$.

On the other hand, if the aperture is small, but still has dimensions of many wavelengths so that the approximations for the denominator matrix are valid, if $1 << x_o << x_1$ and $1 << y_o << y_1$, the antenna cannot resolve the solid angle from within which most of the power density arrives. As far as we could tell by adjusting the antenna pattern, the antenna is being illuminated by a point source. The integrand of (118) behaves like the power-density pattern.

Throughout the present example we will use the renumbering scheme shown in Fig. 24 for the pattern samples, so that the first sample is taken in the direction of maximum power density. The other elemental patterns, f_k, are all zero in that direction, as can be seen from (91c). In the present extreme case, then,

$$[g] \approx \int_{-\infty}^{\infty}\int \left\{\frac{\Lambda^2}{2\eta}\frac{\beta(\ell, m)}{n}\right\} d\ell\, dm \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{bmatrix} \quad .$$

The integral is recognized as the "total power," $\mathcal{W} = 1$ watt, (115), and the maximum available power is $n_o x_o y_o$. The quantity $n_o x_o y_o$ is the area in square wavelengths of the projection of the aperture onto a plane normal to the direction (ℓ_o, m_o). This makes the physical significance of the "total power" clear: It is the power per square wavelength of aperture area that would be available at the terminals of a large antenna if all the power density were concentrated in a single direction, and if that direction were normal to the aperture plane.

The characteristic vector making the power $n_o x_o y_o$ available from a point source is

$$p] = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} ,$$

which corresponds to the pattern,

$$P_x(\ell, m) = \frac{\sin \pi\, x_o(\ell - \ell_o)}{\pi\, x_o(\ell - \ell_o)} \frac{\sin \pi\, y_o(m - m_o)}{\pi\, y_o(m - m_o)} \quad .$$

The field component that gives rise to this pattern is of constant amplitude and linearly varying phase over the aperture:

$$E_x(x, y, o) = \begin{cases} \frac{1}{x_o y_o} e^{-j2\pi(x\ell_o + y m_o)} \\ \qquad \text{if} \quad -1/2\, x_o < x < 1/2\, x_o \quad \text{and} \quad -1/2\, y_o < y < 1/2\, y_o \\ 0 \text{ otherwise} \end{cases} .$$

The normalization (117) will allow direct comparison of the amplitude as well as the shape of the pattern that gives maximum plane-wave gain with the pattern giving maximum available power from the diffuse source.

If the pattern giving maximum plane-wave gain is used, the power available from the diffuse source is evidently $n_o x_o y_o g_{11}$, whereas the power available from a point source of the same total power would be $n_o x_o y_o$. The "gain loss" is therefore the ratio of 1 to g_{11}. And since the maximum power available from the diffuse source is $n_o x_o y_o \alpha_m$, the ratio of g_{11} to α_m is the fraction of the "gain loss" that can be recovered by optimum pattern adjustment.

It will be convenient to plot a normalized available power,

$$w_a = \frac{W_a}{n_o x_1 y_1} \quad , \tag{119}$$

as a function of a normalized aperture dimension,

$$c = \frac{x_o}{x_1} \quad . \tag{120}$$

When the aperture dimensions are small compared with x_1 and y_1, or if all the power density is concentrated in a single direction, this normalized available power may be obtained with the pattern giving maximum plane-wave gain and is equal to the normalized aperture area: $w_a = (x_o/x_1)(y_o/y_1)$; when the aperture dimensions are of the order of x_1 and y_1, the normalized available power is $w_a = (x_o/x_1)(y_o/y_1)\, g_{11}$ if the pattern giving maximum plane-wave gain is used, and it is $w_a = (x_o/x_1)(y_o/y_1)\, \alpha_m$ if the optimum pattern is used; when the aperture dimensions are large compared with x_1 and y_1, the normalized available power reaches the upper limit, $w_a = 1/\pi$.

Next, we discuss the calculation of the elements of the g-matrix, (118). Let the original numbering (r, s) correspond to the new number i (Fig. 24), and let (p, q) correspond to k. The element of [g] in the i^{th} row and k^{th} column is then

$$g_{ik} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{x_1 y_1}{\pi}\, e^{-[x_1^2(\ell-\ell_o)^2 + y_1^2(m-m_o)^2]}\, d\ell\, dm$$

$$\times \frac{\sin\pi[x_o(\ell-\ell_o)-r]}{\pi[x_o(\ell-\ell_o)-r]} \frac{\sin\pi[y_o(m-m_o)-s]}{\pi[y_o(m-m_o)-s]} \frac{\sin\pi[x_o(\ell-\ell_o)-p]}{\pi[x_o(\ell-\ell_o)-p]} \frac{\sin\pi[y_o(m-m_o)-q]}{\pi[y_o(m-m_o)-q]}$$

The integral may be broken down into the product of two integrals of the same form:

$$g_{ik} \equiv g_{rs,pq} = J_{rp} K_{sq} = J_{rp}(x_o, x_1)\, J_{sq}(y_o, y_1) \quad . \tag{121}$$

With $x_1(\ell - \ell_o)$ replaced by λ, the first of these becomes

$$J_{rp} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\lambda^2} \frac{\sin\pi(c\lambda - r)}{\pi(c\lambda - r)} \frac{\sin\pi(c\lambda - p)}{\pi(c\lambda - p)}\, d\lambda \quad , \tag{122}$$

with c the normalized aperture dimension (120). The evaluation of this integral is considered in the appendix.

With reference to (121), it should be noted that the new number i corresponds to the original numbering (r, s), not (r, p); it is not possible to express the matrix elements as the product $g_{ik} = a_i b_k$, which, as we discovered in Sec. VI, would lead to a particularly simple result. However, the fact that each matrix element can be expressed as a product of a factor depending on x-dimensions and a factor depending on y-dimensions, (121), will simplify the calculations considerably. (121) is a direct consequence of our having chosen a power-density pattern that can also be written as a product of two functions: In the narrow range of directions where (114) has appreciable magnitude, it may be written as the product of a function of the direction cosine ℓ and a function of the direction cosine m. As a result of (121), the optimization of the antenna pattern may be split into two simpler one-dimensional problems, as we shall now see.

The available power, (116), will be maximized subject to the constraint, (117), by the set of pattern samples satisfying

$$\underline{p} \times [g] = \alpha\, \underline{p} \quad ,$$

which may be written

$$\sum_i p_i\, g_{ik} = \alpha p_k \quad , \text{ all } k$$

in terms of the new sampling grid numbering, or

$$\sum_{r,s} p_{rs}\, g_{rs,pq} = \alpha p_{pq} \quad , \quad \text{all } p \text{ and } q$$

in terms of the original numbering. (The unfortunate double use of the symbol p should not cause any confusion.) Using (121), we have

$$\sum_r J_{rp} \sum_s p_{rs} K_{sq} = \alpha p_{pq} \quad , \quad \text{all } p \text{ and } q,$$

which may be rewritten in matrix notation as

$$[J]_t \times [p] \times [K] = \alpha[p] \quad , \tag{123}$$

in which $[p] = [p_{rs}]$ is the matrix of pattern samples, $[P_x(\ell_o + r/x_o,\ m_o + s/y_o)]$, arranged in a rectangular array corresponding to the array of points on the sampling grid. [J] and [K] are square, symmetric matrices, as may be seen from (122). If the sampling grid has R rows (different values of r) and S columns, [J] will be of order R and [K] will be of order S.

Now suppose that the characteristic value problems,

$$\left.\begin{aligned} \underline{u} \times [J] &= \lambda \underline{u} \\ \text{and} \qquad \underline{v} \times [K] &= \mu \underline{v} \end{aligned}\right\} \quad , \tag{124}$$

have been solved. If the transpose of the first of these equations is post-multiplied by the second, we have

$$[J]_t \times u] \times \underline{v} \times [K] = \lambda\mu\, u] \times \underline{v} \quad ,$$

so that

$$[p] = u] \times \underline{v} \quad , \quad \alpha = \lambda\mu \tag{125}$$

is a solution to (123). In fact, there are R solutions to the first of equations (124) and S solutions to the second equation and hence RS solutions to (123) of the form (125). But, since there are RS points on the sampling grid and hence RS pattern samples to be determined, all solutions to (123) are of the form (125), with $\underline{u}$, $\underline{v}$, λ, and μ determined from (124).

If the constraints,

$$\underline{u} \times u] = 1 \quad \text{and} \quad \underline{v} \times v] = 1 \tag{126}$$

are satisfied by the characteristic vectors of (124), the constraint (117), which may be written variously as

$$1 = \underline{p} \times p] = \sum_i p_i^2 = \sum_{r,s} p_{rs}^2$$

$$= \sum_{r,s} u_r^2 v_s^2 = \sum_r u_r^2 \sum_s v_s^2$$

$$= \underline{u} \times u] \times \underline{v} \times v] \quad ,$$

is also satisfied.

Physically, the characteristic vector $\underline{u}$ determines the behavior of the antenna pattern as a function of the direction cosine ℓ, and $\underline{v}$ determines its behavior with m; the optimum pattern may be expressed as the product of a function of ℓ and a function of m, as may be seen from (91). The tangential fields over the aperture may be written in a corresponding product form, as may be seen from (92). The solutions of (124) corresponding to the greatest characteristic values, λ and μ, will result in the maximum available power.

Since we have chosen a power-density pattern that is symmetric about $\ell = \ell_o$ (and about $m = m_o$), we expect an optimum pattern that is also symmetric about $\ell = \ell_o$ (and about $m = m_o$): We expect the elements of the characteristic vectors of (124) to satisfy $u_{-r} = u_r$ (and $v_{-s} = v_s$). If the sampling grid extends symmetrically about the point $(r,s) = (0,0)$, the elements of the characteristic vectors of (124) will actually show either even symmetry, $u_{-r} = u_r$, or odd symmetry, $u_{-r} = -u_r$ and $u_o = 0$, as we will next demonstrate.

We need the relations $J_{pr} = J_{rp}$ and $J_{-r-p} = J_{rp}$, which may be seen from (122). When these are used, the J-matrix becomes

$$[J] = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdots J_{22} & J_{12} & J_{02} & J_{-12} & J_{-22} & \cdots \\ \cdots J_{12} & J_{11} & J_{01} & J_{-11} & J_{-12} & \cdots \\ \cdots J_{02} & J_{01} & J_{00} & J_{01} & J_{02} & \cdots \\ \cdots J_{-12} & J_{-11} & J_{01} & J_{11} & J_{12} & \cdots \\ \cdots J_{-22} & J_{-12} & J_{02} & J_{12} & J_{22} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \end{bmatrix} .$$

This matrix can be simplified by being transformed by the orthogonal matrix,

$$[M] = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdots \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \cdots \\ \cdots 0 & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 & \cdots \\ \cdots 0 & 0 & 1 & 0 & 0 & \cdots \\ \cdots 0 & -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 & \cdots \\ \cdots -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \end{bmatrix} .$$

The simplification is

$$[M] \times [J] \times [M]_t = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & (J_{22}+J_{-22}) & (J_{12}+J_{-12}) & \sqrt{2}\,J_{02} & 0 & 0 & \cdots \\ \cdots & (J_{12}+J_{-12}) & (J_{11}+J_{-11}) & \sqrt{2}\,J_{01} & 0 & 0 & \cdots \\ \cdots & \sqrt{2}\,J_{02} & \sqrt{2}\,J_{01} & J_{00} & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & (J_{11}-J_{-11}) & (J_{12}-J_{-12}) & \cdots \\ \cdots & 0 & 0 & 0 & (J_{12}-J_{-12}) & (J_{22}-J_{-22}) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

The original characteristic value problem, (124), may be rewritten in terms of the transformed vector set $\underline{u}'$, given by

$$\underline{u} = \underline{u}' \times [M] \quad .$$

The transformed characteristic value problem reads

$$\underline{u}' \times [M] \times [J] = \lambda \underline{u}' \times [M] \quad ,$$

or, after post-multiplication by $[M]_t$,

$$\underline{u}' \times \{[M] \times [J] \times [M]_t\} = \lambda \underline{u}' \times [M] \times [M]_t$$

$$= \lambda \underline{u}'$$

since the transformation matrix is orthogonal, i.e. $[M]_t = [M]^{-1}$. The corresponding constraint, (126), reads

$$\underline{u} \times u] = \underline{u}' \times [M] \times [M]_t \times u'] = \underline{u}' \times u'] = 1 \quad .$$

The characteristic values of the simplified matrix, $[M] \times [J] \times [M]_t$, are characteristic values of the sub-matrices located on its principal diagonal. The corresponding characteristic vectors, $\underline{u}'$, have zero elements in either the last $(R-1)/2$ positions or the first $(R+1)/2$ positions, depending on whether the associated characteristic value is a characteristic value of the upper left-hand sub-matrix or of the lower right-hand sub-matrix. If the characteristic value is from the upper left-hand sub-matrix, only the first $(R+1)/2$ rows of [M] are needed to calculate the characteristic vector $\underline{u}$ of the original problem (124), and it will be noted from [M] that in this case, $u_{-r} = u_r$. On the other hand, if the characteristic value is from the lower right-hand

sub-matrix, only the last $(R-1)/2$ rows of [M] are needed, and $u_{-r} = -u_r$ and $u_o = 0$. Since $u_o = 0$ implies a null in the antenna pattern in the direction of maximum power density it is to be expected that the optimum pattern will show even symmetry, so that the upper left-hand sub-matrix will have the greatest characteristic value. This was found to be the case in all the calculations made.

Although the discussion up to this point has presumed a rectangular aperture and a power-density pattern with unequal parameters, x_1 and y_1, the calculations were further simplified by making $y_o/y_1 = x_o/x_1 = c$, so that the normalized aperture dimensions are equal. It is convenient to regard the aperture as being a square with side x_o and the power density pattern as being circularly symmetric about (t_o, m_o), so that $y_1 = x_1$. A square sampling grid with an odd number of rows and columns has been used. These simplifications make [J] = [K], so that [p] is square; the maximum characteristic values and vectors of (124) are also equal: $\lambda_m = \mu_m$ and $\underline{u} = \underline{v}$.

At the start it was not clear how many samples would be necessary, so the greatest characteristic values of a 25 × 25 g-matrix and a 49 × 49 g-matrix were obtained. These correspond to sampling grids with 5 rows and columns and 7 rows and columns, or to J-matrices of order 5 and 7, respectively. In none of these calculations did the greatest characteristic value of the 25 × 25 g-matrix differ from the greatest characteristic value of the 49 × 49 g-matrix by more than 0.0047 decibels. We may easily conclude that the sampling grid with 5 rows and columns is sufficient for the present example. The difference in characteristic vectors is just barely discernible in the patterns in directions well removed from (t_o, m_o).

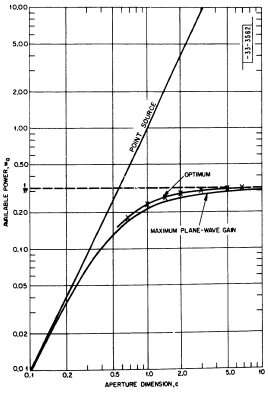

Fig. 25. The available power increases with the aperture area.

The normalized available power (119) is plotted against the normalized aperture dimension (120) in Fig. 25. The bottom curve is the power available from an antenna having maximum plane-wave gain, and the sloped line is the power available from a point source of the same total power, W, when the same antenna is used. The decibel difference between these curves is the "gain loss." The middle curve is the power available with optimum pattern adjustment. Since a characteristic value problem must be solved to obtain each point on the curve, only the seven marked points have been obtained. The decibel difference between this curve and the lower curve is the recoverable "gain loss."

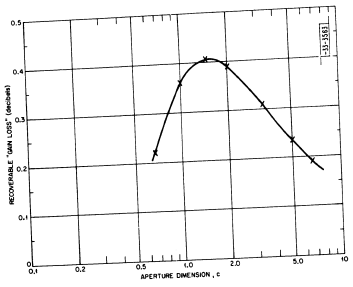

Fig. 26. The power advantage of an antenna having an optimum pattern is not very great.

The recoverable "gain loss" is disappointingly small; it is plotted in Fig. 26.

Figures 27 through 30 are a set of antenna patterns plotted from the characteristic vectors of the 49 × 49 matrices. The optimum patterns are for alternate values of the normalized aperture dimension for which calculations were made: c = 1/1.5, 1/0.7, 1/0.3, 1/0.15. Although the power-density pattern has circular symmetry about the direction of maximum power density, the antenna patterns do not, because the aperture is not circular. Figures 27 through 30 are cross sections of the patterns in the plane $m = m_o$. Also, the patterns really get narrower as the aperture dimension increases, but plotting them against $x_o(\ell - \ell_o)$ makes the antenna patterns appear to remain about the same width, while the power-density pattern appears to get broader. The $\frac{\sin x}{x}$ -pattern giving maximum plane-wave gain and the power-density pattern are plotted for comparison.

The required distribution of tangential electric field on the aperture plane, resulting from excitation of the antenna at its terminals, is plotted in Fig. 31, which shows contours of constant magnitude of electric field across the aperture. (In the present example, the phase varies linearly over the aperture at rates proportional to the direction cosines ℓ_o and m_o.) Figure 31 has been plotted for the normalized aperture dimension, c = 1/0.7, the value at which the recoverable "gain loss" is the greatest. The normalization is such that

$$\int_{-\infty}^{\infty}\int |E_x(x, y, 0)|^2 \, dx \, dy = x_o y_o \quad .$$

We may conclude from the foregoing example that, if the power density pattern has a Gaussian shape, the effort required to design and construct an optimum antenna will generally not be sufficiently compensated by increased available power. In many cases, an antenna designed to have low side lobes will deliver more power than an antenna of the same size having maximum plane-wave gain. It seems reasonable to extrapolate these conclusions to any power-density pattern having a single, fairly well defined maximum.

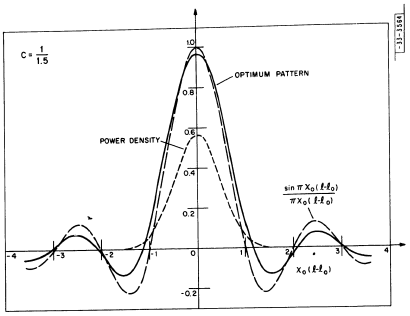

Fig. 27. When the aperture area is small, the optimum pattern is shaped like the $\frac{\sin x}{x}$ -pattern giving maximum plane-wave gain.

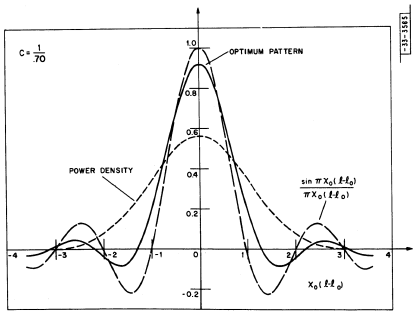

Fig. 28. The aperture area for this figure is larger than the aperture area for the preceding figure.

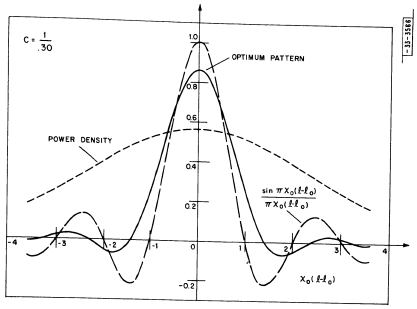

Fig. 29. As the aperture area increases, the side-lobe level of the optimum pattern decreases.

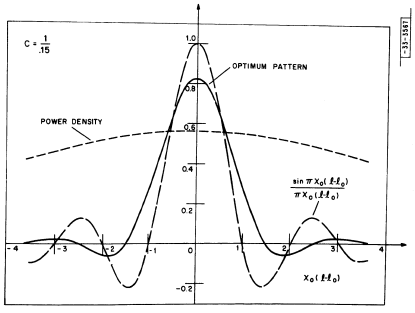

Fig. 30. The aperture area for this figure is large.

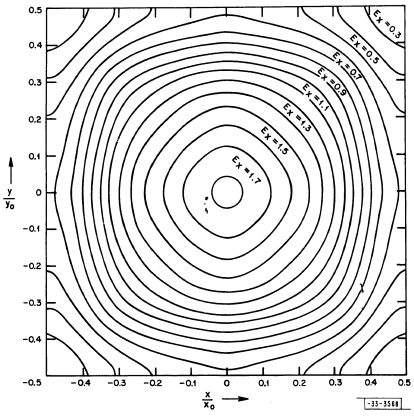

Fig. 31. Contours of electric field amplitude on the aperture plane for an antenna having an optimum pattern.

VIII. OPTIMUM DIVERSITY

In discussing the optimum coupling network for the set of electrically separate sources of Fig. 16, Sec. VI, we found that in the extreme case when the open-circuit source voltages are perfectly correlated, the coupling network can deliver a power to a single load equal to the sum of the powers available from each source separately. But if the source voltages are completely uncorrelated, the other extreme, the maximum power available from the coupling network, is only the greatest of the powers available from each source separately. We also remarked that, if the sources represent elements of a receiving array, the desirable perfect coherence between source voltages may be expected to persist for a period of time short compared with the fading period of the arriving signal, but that a time-varying coupling network would then be required to take advantage of this coherence. Although it will be convenient, initially, to regard the time-varying coupling network as an extension of the work of Sec. VI, the resulting apparatus is really a form of spaced-antenna diversity which results in the maximum signal-to-noise (power) ratio of the resultant signal.

The signal-to-noise ratio available from a number of coherent but noisy sources has been maximized by Kahn,[18] this maximization has been established in a more general form by Brennan,[19] and implementations of the mathematical result have been described by Kahn and by Mack,[20] who calls his equipment "combiner diversity." But whereas the equipment discussed by Kahn and Mack combines the signals after detection, we will arrange to combine signals before detection: The equipment suggested by this section will continue to deliver the maximum signal-to-noise ratio of the resultant signal even if the signal power available from each antenna is less than the noise power.

We will again represent the terminal-pair behavior of an array of N antennas by the set of disconnected sources on the left in Fig. 16. This representation entails no loss in generality, for, if there is appreciable mutual impedance between the terminal pairs of an actual array, the array elements may be interconnected as in Figs. 20 and 21 to produce an equivalent set of N disconnected sources, capable of supplying the same power. The analysis of this section will then apply to the equivalent set of sources.

In terms of the matrix of transmission coefficients, $\underline{t}$, and the matrix of normalized open-circuit signal voltages of the sources in Fig. 16, $V/\sqrt{8R}$] (Eq. 75), the instantaneous signal power in the load, R_o, is

$$W_i = \left[\underline{t} \times \frac{V}{\sqrt{8R}}\right] \times \frac{V^*}{\sqrt{8R}} \times \underline{t}^*] \quad .$$

By "instantaneous" power we mean the power averaged over an integer number of radio-frequency cycles, but not averaged for such a long time that the modulating waveform and particularly the fading waveform has a chance to change. As we discovered in Sec. VI, this instantaneous power is maximized subject to the constraint,

$$\underline{t} \times \underline{t}^*] = 1 \quad , \tag{68a}$$

by the set of transmission coefficients satisfying

$$\underline{t} = \frac{1}{\sqrt{W_{ci}}} \frac{\underline{V}^*}{\sqrt{8R}} e^{j\Theta} \quad , \tag{80a}$$

with Θ an arbitrary phase shift, and the maximum instantaneous power is

$$W_{ci} = \left[\frac{\underline{V}}{\sqrt{8R}} \times \frac{\underline{V}^*}{\sqrt{8R}}\right] \quad , \tag{79a}$$

which is the sum of the instantaneous signal powers available from each source separately. If the transmission coefficients are continuously adjusted according to (80a) to maintain the maximum instantaneous power (79a), the mean signal power is again the sum of the mean signal powers available from each source separately. The mean noise power, as discussed in Sec. VI, is independent of the adjustments of the transmission coefficients if, as we shall assume, the noise arrives uniformly from all directions in space.

It is enlightening to regard this continuously adjusted array as a single composite antenna. We note that the maximum power available from an array (of electrically separate elements) illuminated by a point source is also the sum of the powers available from each element. Thus, if the individual elements of our composite antenna are not subject to "gain loss" when they are illuminated by a diffuse source, the composite antenna itself will not suffer "gain loss" when illuminated by the same diffuse source!

To within the approximations discussed in the last section, which are valid when aperture dimensions are large compared with the wavelength, the power available from an element of the composite antenna is proportional to its aperture area if the elemental antenna is too small to "resolve" the power-density pattern, but approaches an upper limit as the aperture area is increased (see Fig. 25). Apparently, if its total aperture area is fixed, the composite antenna will deliver the greatest power if its elemental antennas are reduced in size (and increased in number) until they can no longer "resolve" the power-density pattern. From the point of view taken in the introductory section, a scatter link employing this continuously adjusted array will be most economical of the aperture areas of transmitting and receiving antennas if the transmitting antenna and each of the elemental antennas of the receiving array are not able to resolve the scattering region.

We have seen that a continuously adjusted array can deliver more signal power to a matched load than a single large antenna having the same aperture area. But if each transmission coefficient of the time-varying coupling network is to be made proportional to the conjugate of the signal voltage at the terminals of the corresponding array element, as required by (80a), it would appear that the signal at each array element would have to be known exactly in order for the composite signal to be received with maximum signal-to-noise ratio! In addition, we should be concerned about whether the time variations of the transmission coefficients will perturb the composite signal. Fortunately, for one choice of the "arbitrary phase shift," Θ of (80a), a complete knowledge of the signals is not required, and the time variations of the transmission coefficients do not perturb the composite signal, as we shall now see.

It will be necessary to differentiate between signal variations caused by modulation of the transmitter and signal variations caused by fading. As shown in Fig. 32, we will suppose that the transmitted signal is modified differently in following each of the paths from transmitting

antenna to the various receiving array elements. $M e^{j\omega_o t}$ is the vector transmitted signal having the radian carrier frequency ω_o and modulation described by the complex time function M. (The instantaneous current at the transmitting antenna terminals would be proportional to $\mathcal{R}e\{M e^{j\omega_o t}\}$.) In traversing the path to the i^{th} array element, the vector signal is subjected to fading; it varies in proportion to the complex time function K_i, which may be regarded as the gain of the i^{th} path. Before the signal appears as the voltage at the terminals of the i^{th} array element, noise is added; the various path noises will be assumed to be uncorrelated and equal in power.

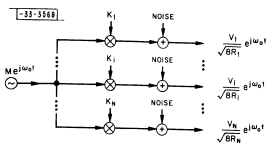

Fig. 32. Model of the radio paths from the transmitting antenna to the various elemental antennas of the receiving array.

Note that the model of Fig. 32 precludes the effects of multipath transmission: If the transmitter is pulsed, a single replica of that pulse is received by each of the elemental antennas. Although this assumption is more restrictive than the assumptions discussed in Sec. II, it is often met in practice and is necessary for the proper functioning of the apparatus we will devise. Since the path gains K_i are assumed to be independent of frequency (over the band of frequencies occupied by the transmitter), no provision is made for the time-of-flight of the signal from transmitter to receiving site; this is of no consequence here.

From Fig. 32, the matrix of normalized open-circuit signal voltages of the array elements may be written

$$\frac{\underline{V}}{\sqrt{8R}} = M\underline{K} \quad , \tag{127}$$

in which $\underline{K}$ is the matrix of path gains. The maximum instantaneous power, (79a), becomes

$$W_{ci} = |M|^2 \left[\underline{K} \times \underline{K}^*\right] \quad , \tag{128}$$

and it may be obtained if the transmission coefficients, from (80a), are

$$\underline{t} = \frac{M^*}{\sqrt{W_{ci}}} \underline{K}^* e^{j\Theta} = \left(\frac{M^*}{|M|} e^{j\Theta}\right) \frac{\underline{K}^*}{\sqrt{\left[\underline{K} \times \underline{K}^*\right]}} \quad . \tag{129}$$

The corresponding normalized open-circuit signal voltage at the output of the coupling network is then

$$\left[\frac{\underline{V}}{\sqrt{8R_o}} \left(= I_o \sqrt{\frac{R_o}{2}} \text{ in the notation of Fig. 16}\right) = \underline{t} \times \frac{\underline{V}}{\sqrt{8R}}\right] = |M| e^{j\Theta} \sqrt{\left[\underline{K} \times \underline{K}^*\right]} \quad . \tag{130}$$

From (130), we see that the signal emerging from the coupling network is proportional, among other factors, to the root-sum-square path gain – a real, slowly varying function of time.

If we chose the phase shift Θ to be a constant, the phase of the output signal would also be constant; only the amplitude of the output signal would vary, in proportion to the magnitude of the modulation. A frequency-modulated signal would emerge from the coupling network as a carrier, fading slowly in amplitude only! Under the same conditions, from (129), we see that the transmission coefficients would have to be adjusted in accordance with the phase of the modulation! This choice of Θ leads to both of the difficulties mentioned previously. Clearly, the appropriate choice is

$$\Theta = \text{Arg } M \quad ; \quad e^{j\Theta} = \frac{M}{|M|} \quad . \tag{131}$$

For this choice, the transmission coefficients need only be adjusted in accordance with the slowly varying path gains:

$$\underline{t} = \frac{\underline{K}^*}{\sqrt{\underline{K} \times \underline{K}^*}} \quad , \tag{129a}$$

the output signal is directly proportional to the modulation:

$$\frac{V}{\sqrt{8R_o}} = M \sqrt{\underline{K} \times \underline{K}^*} \quad , \tag{130a}$$

and neither of the difficulties appears.

If the transmitter is unmodulated (M = 1), the instantaneous power available from the time-varying coupling network, (128), is

$$\underline{K} \times \underline{K}^* = \sum_{i=1}^{N} |K_i|^2 \quad .$$

Unless the magnitudes of the path gains are perfectly correlated, the relative spread in the distribution of power output due to fading will be less (in a mean-square sense) than the relative spread in the distribution of power available from one of the array elements. In other words, the decibel fading range is reduced by the continuously adjusted array — another of its advantages. Staras[24] discusses the case in which the $|K_i|$ are independent and Rayleigh-distributed. These conditions would be expected to hold for short periods of time, say of the order of an hour, and if the array elements were appreciably separated in space. He points out that the power output follows a Chi-square distribution (having 2 N degrees of freedom), which he plots for N up to 10.

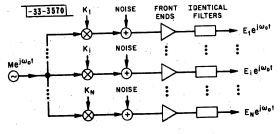

Fig. 33. The model has been modified slightly to include appropriate amplifiers and filters.

It remains for us to find an appropriate measure of the path gains and to show how these may be obtained and used in an electronic realization of the time-varying coupling network. First, since the instrumentation is to be electronic, we will need workable signal levels, and it will be expedient to filter out the noise except within the band of frequencies occupied by the transmitter. Suitable arrangements are indicated in the block diagram of Fig. 33, which

will be used as a model of the problem at hand. The outputs of this model, the E's, are the amplified and filtered array element voltages. It will be assumed that the front-end gains have been set so that the noise levels of these outputs are the same, and, as discussed previously, the noises in the different outputs are uncorrelated. These considerations are important, as we shall see. If the front ends include mixers, as will generally be convenient, it is also important that a common local oscillator be used. The signal-to-noise ratios of the outputs will all be degraded by the noise figures of the front ends, but this unavoidable loss would also appear if a passive coupling network and single front end were used.

After amplification, the question of obtaining maximum power does not arise; it is the signal-to-noise ratio that must be maximized. The latter problem, however, is not new; it has already been solved: We found that, in addition to providing maximum power in a matched load, the coupling network output,

$$\frac{V}{\sqrt{8R_o}} = \underline{t} \times \frac{V}{\sqrt{8R}} \Bigg] \quad ,$$

also results in a maximum signal-to-noise ratio when the transmission coefficients are adjusted according to (129a), since the output noise power is independent of the adjustment of the transmission coefficients (when the noises from the separate sources are uncorrelated and equal in power). By setting the front-end gains in Fig. 33 to equalize the output noise levels, we have brought the output signal voltages, the E's, into proportion with the normalized open-circuit signal voltages of the array elements, i.e.,

$$E] \propto \frac{V}{\sqrt{8R}} \Bigg] \quad .$$

Thus, the signal-to-noise ratio of the combined output,

$$E_o = \underline{t} \times E] \quad , \tag{132}$$

or a multiple thereof, will also be a maximum.

It will be convenient to lump the fixed, equal front-end gains,

$$\frac{E_i}{\frac{V_i}{\sqrt{8R_i}}} \quad ,$$

with the path gains, K_i, in Fig. 33, so that the output signal voltages may be written

$$\underline{E} = M \underline{K} \quad . \tag{133}$$

Defined in this way, the path gains of Fig. 33 are no longer equal to, but are in direct proportion to, the path gains of Fig. 32. This will make no difference in the "transmission coefficients," as may be seen from (129a), and, of course, the absolute level of the combined output (132) is immaterial: We will continue to use the same symbols, K_i.

The combined output,

$$E_o = \underline{\perp} \times E] = \frac{\underline{K^* \times E}]}{\sqrt{\underline{K} \times K^*]}} \quad , \tag{132a}$$

then, would have the maximum signal-to-noise ratio, and it would have a fixed noise level. The root-mean-square signal level would be proportional to $\sqrt{\underline{K} \times K^*]}$, as may be seen from (133) and (132a). The combined output,

$$F_o = \underline{K^* \times E}] = \sum_{i=1}^{N} K_i^* E_i \quad , \tag{134}$$

would also have the maximum signal-to-noise ratio, but the root-mean-square noise level would vary in proportion to $\sqrt{\underline{K} \times K^*]}$, and the root-mean-square signal level would be proportional to $\underline{K} \times K^*]$. (The "mean" in these statements is interpreted as an appropriate short-time average.) In the design of a receiver, it is often convenient to use automatic volume control to hold the root-mean-square signal level fixed. This may be accomplished after the outputs of Fig. 33 have been combined. It will be simplest to realize a circuit having the combined output (134), for which the combining coefficients are simply the conjugates of the path gains.

We return to the problem of finding an appropriate measure of the path gains, which are defined by (133). If it were not for the noise, we might measure M at the transmitter and E_i at the receiver, giving $K_i = E_i/M$. This possibility is not too far-fetched, for we hope to recover M at the receiver as well. It is the noise that causes the real difficulty; this estimate, at any instant of time, would be intolerably noisy. We recall that the path gains, which are the signals received when the transmitter is unmodulated (M = 1), vary only slowly; they describe the fading. Thus a short-time average, taken, at each instant, over a period of the past that is small compared with the fading period, would provide an accurate estimate of a path gain and at the same time would reduce the noise power of an instantaneous estimate by a factor equal to the product of the averaging time and the bandwidth of one of the filters shown in Fig. 33. For example, if the system (and filter) bandwidth were 100 kcps and the fading were slow enough so that we could average over 0.1 second of the past, the averaging process would reduce the noise power by a factor of 10,000. Using this averaging process, we could obtain a continuous good estimate of a path gain for output signal-to-noise ratios of that path greater than about −40 decibels. This short-time average will be indicated by the notation $\underset{f}{\text{Av.}}\{\ \}$.

If the short-time average is performed by a linear passive network, as we intend, contributions to the average will be weighted in accordance with how far back in the past they occurred. Another kind of weighting may be considered: A reasonable measure of a path gain would be

$$K_i = \underset{f}{\text{Av.}}\left\{g \frac{E_i}{M}\right\} \quad ,$$

with the arbitrary time function, g, real and non-negative. To be consistant, the same weighting function, g, would have to be used in measuring the gain of each path.

It would be sensible to pick g in such a way that it increases with the magnitude of the modulation: g a monotonic increasing function of $|M|$. This kind of choice gives more weight to the instantaneous estimate, E_i/M, when the signal amplitude, and hence the signal-to-noise ratio,

is momentarily high. On the other hand, it should be emphasized that the choice of weighting function is relatively unimportant compared with the maximization of signal-to-noise ratio obtained when the outputs of Fig. 33 are properly combined, (134), using the measured path gains. First, the "direct noise" in the combined output, due to a noisy estimate, K_i, but neglecting the noise that has been added to each output, will be almost indistinguishable from the fading, since the rate of variation of the estimated path gain will be limited by the network performing the short-time average. We may therefore disregard the variations in the estimated path gains due to noise, and we are left with the prospect of having a reduced signal-to-noise ratio resulting from improperly determined, but momentarily fixed "transmission coefficients." Except when the estimates are violently in error, this effect is again small. As we discovered in Sec. VI, the correct estimates result in a stationary point of the signal-to-noise ratio. Thus, a small fractional error,

$$\epsilon = \frac{\Delta K_i}{\sqrt{\underline{K} \times K^*]}} \quad ,$$

in one of the estimates would result in a smaller fractional error of the order of $|\epsilon|^2$, a reduction of course, in the signal-to-noise ratio.

A reasonable choice of weighting function is $g = |M|$. We will accept this choice not because it is optimum in some sense, but because it leads to particularly simple circuitry. Our estimate of path gain will be

$$K_i = \underset{f}{\text{Av.}}\left\{|M| \frac{E_i}{M}\right\} = \underset{f}{\text{Av.}}\{E_i e^{-j \operatorname{Arg} M}\} \quad .$$

Not having a copy of the transmitted signal at the receiver, we will use the phase of the combined output (134) as an estimate of the phase of the transmitted signal. This estimate will be quite satisfactory as long as the signal-to-noise ratio of the combined output is only slightly greater than unity, since errors in this estimate affect only the measured path gains. Thus, our measurement of path gain will give

$$K_i = \underset{f}{\text{Av.}}\left\{E_i e^{-j \operatorname{Arg} F_o}\right\} \quad , \tag{135}$$

and we will obtain the combined output

$$F_o = \sum_{i=1}^{N} E_i \underset{f}{\text{Av.}}\left\{E_i^* e^{j \operatorname{Arg} F_o}\right\} \quad . \tag{134a}$$

Containing, as it does, the output F_o on both sides, (134a) is reminiscent of the equation of an oscillator. In fact, the circuit for obtaining this output, shown symbolically in Fig. 34, is an oscillator of sorts. The inputs to this circuit are the outputs of the front ends and filters shown in Fig. 33; the output appears at the upper right of Fig. 34. The tuned circuits marked ω_1 perform the short-time averages. They are all tuned to some (radian) frequency $\omega_1 < \omega_o$ and have bandwidths of the order of the reciprocal of the averaging time (10 cps in the numerical example previously mentioned). The limiter delivers a constant-amplitude output sinusoid (chosen to be

one volt in Fig. 34), and the phase of its output is the same as the phase of the input. The limiter should have sufficient dynamic range to limit properly on noise.

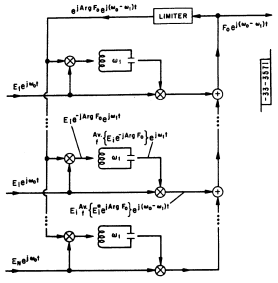

Fig. 34. An electronic realization of the time-varying coupling network.

Noting that the instantaneous frequencies of the signals emerging from the tuned circuits must be close to ω_1, we see that the instantaneous frequency of the signal leaving the second multiplier in each channel must be about $\omega_o - \omega_1$. (It is assumed that product components at the sum frequency are suppressed.) The output, which is the sum of the outputs of the second multipliers, may then be represented as having the carrier frequency $\omega_o - \omega_1$, as has been done in Fig. 34. It may be seen, by following an assumed output through the processes of limiting, first multiplication, averaging, second multiplication, and summing, that the circuit of Fig. 34 does indeed produce the output given by (134a).

The properties of this combining circuit may be deduced either from the figure or from (134a). Suppose, first, that the phases of all the inputs are advanced suddenly by the same amount, simulating a step increase in the phase of the modulation. Nothing will happen immediately to the tuned-circuit voltages (the short-time averages will remain fixed), but the phases of all the second-multiplier outputs and hence the phase of the output will advance by the same amount. The phases of the first-multiplier outputs (the quantities being averaged), however, will not change. Hence there will be no additional, long-term change in the output phase. In other words, the phase of the modulation is reproduced in the output, perturbed, possibly, by noise and fading. The output behaves differently with amplitude changes, however. For rapid changes in the amplitude of the modulation, the output follows the input because the tuned circuits cannot follow the input changes. But if the modulation amplitude changes slowly, at a rate less than or comparable to the fading rate, the output amplitude varies as the square of the modulation amplitude: The tuned circuits can follow these slow changes, and the magnitude of the loop gain of this "oscillator" is twice multiplied by the amplitude of the input signals. From another point of view, the decibel fading range in the output is twice the range that would be obtained from a circuit using normalized transmission coefficients and producing the combined output (132a). As we have already noted, this difficulty may be mitigated by an automatic volume control operating on the combined output F_o.

The behavior of the circuit of Fig. 34 in the absence of signal and under the influence of noise alone has not been studied, either experimentally or analytically. Nevertheless, we may conjecture about its behavior when the inputs are filtered Gaussian noise. First, if the noise inputs are uncorrelated and equal in power, we would expect the tuned-circuit outputs to vary

slowly in amplitude and phase, behaving very nearly like independent filtered Gaussian noises having the same power. The application of even a very weak signal should cause the tuned-circuit outputs to tend to fall in a fixed proportion in such a way as to maximize the signal-to-noise ratio at the output.

On the other hand, if one of the noise powers were greater than the others, we would expect the corresponding tuned-circuit output also to be large. In fact, we would expect the large tuned-circuit output to suppress the others in such a way that the combined output follows the greatest input noise rather closely. Similarly, if the input noises were partially correlated, we would expect the tuned-circuit outputs to tend to adjust themselves so as to maximize the output noise. In either of these conditions, the combining circuit may be regarded as biased; its adjustment favors one set of input signals. A fairly strong signal would then be required to overcome the bias and cause the tuned-circuit outputs to readjust themselves properly for that signal.

The circuit of Fig. 34 bears a resemblance to the circuits devised by Price[22] for another application. The resemblance would be stronger if the limited output were replaced by a stored-reference signal, having some of the characteristics of the signal expected from the transmitter. If a stored-reference signal is available at the receiver, such an arrangement is preferable to the circuit of Fig. 34 when the output signal-to-noise ratio is small.

ACKNOWLEDGMENT

It is a pleasure to acknowledge the many helpful suggestions and criticisms offered by my colleagues at the Massachusetts Institute of Technology and at the Lincoln Laboratory. In particular, the course of this paper has been strongly influenced by discussions with Prof. F.M.H. Villars, Prof. W.M. Siebert, Mr. J.H. Chisholm, Dr. S.C. Wang, Dr. W.L. Root, Dr. S.J. Fricker and Mr. W.C. Mason. The numerical calculations in Sec. VII were performed by and under the supervision of Mr. P.A. Duffy. My wife has been a constant inspiration to me in the preparation of this paper, and she has spent many long hours helping to correct the manuscript. The research reported here was conceived and suggested to me by Dr. J.T. deBettencourt.

REFERENCES

1. Perhaps the best single reference that can be given is the Scatter Propagation Issue of Proc. I.R.E. (Vol. 43, No. 10, October 1955) which contains a series of well-documented articles representing various experimental and theoretical views of the subject.

2. F.W. Schott, "On the Response of a Directive Antenna to Incoherent Radiation," Proc. I.R.E. 39, 677 (1951).

3. H.G. Booker and W.E. Gordon, "A Theory of Radio Scattering in the Troposphere," Proc. I.R.E. 38, 401 (1950).

4. W.E. Gordon, "Radio Scattering in the Troposphere," Proc. I.R.E. 43, 23 (1955).

5. D.K. Bailey, R. Bateman, L.V. Berkner, H.G. Booker, G.F. Montgomery, E.M. Purcell, W.W. Salisbury and J.B. Wiesner, "A New Kind of Radio Propagation at Very High Frequencies Observable over Long Distances," Phys. Rev. 86, 141 (1952).

6. F. Villars and V.F. Weisskopf, "On the Scattering of Radio Waves by Turbulent Fluctuations of the Atmosphere," Proc. I.R.E. 43, 1232 (1955).

7. H.G. Booker and P.C. Clemmow, "The Concept of an Angular Spectrum of Plane Waves, and Its Relation to That of Polar Diagram and Aperture Distribution," Proc. I.E.E. 97, 11 (1950).

8. H.G. Booker, J.A. Ratcliffe and D.H. Shinn, "Diffraction from an Irregular Screen with Application to Ionospheric Problems," Phil. Trans. 242, 579 (1950).

9. G.N. Watson, A Treatise on the Theory of Bessel Functions (Cambridge University Press, 1948), pp. 20, 373 and 54.

10. B. Chance, V. Hughes, E.F. McNichol, D. Sayre and F.C. Williams, Waveforms, Radiation Laboratory Series, Vol. 19 (McGraw-Hill Book Company, New York, 1945), pp. 37, 323 and 664.

11. V. Belevitch, "Synthèse des Réseaux Électriques Passifs à n Paires de Bornes de Matrice de Répartition Prédéterminée," Annales des Télécommunications 6, 302 (1951).

12. F.B. Hildebrand, Methods of Applied Mathematics (Prentice-Hall, New York, 1952), Ch. 1.

13. Ibid., pp. 68-80. The Discussion is generally applicable to Hermitian matrices if the scalar products, e.g. (230), and quadratic forms, e.g. (239), are interpreted as Hermitian scalar products and Hermitian forms. See also pp. 42-46.

14. E.A. Guillemin, The Mathematics of Circuit Analysis (John Wiley & Sons, New York, 1949), pp. 59-62.

15. S.A. Schelkunoff and H.T. Friis, Antennas: Theory and Practice (John Wiley & Sons, New York, 1952), pp. 159 and 162.

16. Ibid., pp. 195-198.

17. J.B. Rosser, Theory and Applications of $\int_o^z e^{-x^2} dx$ and $\int_o^z e^{-p^2y^2} dy \int_o^y e^{-x^2} dx$ (Mapleton House, New York, 1948), pp. 190-191.

18. L.R. Kahn, "Ratio Squarer," Letter to the Editor, Proc. I.R.E. 42, 1704 (1954).

19. D.G. Brennan, "On the Maximum Signal-to-Noise Ratio Realizable from Several Noisy Signals," Letter to the Editor, Proc. I.R.E. 43, 1530 (1955).

20. C.L. Mack, "Diversity Reception in UHF Long-Range Communications," Proc. I.R.E. 43, 1281 (1955).

21. H. Staras, "The Statistics of Combiner Diversity," Letter to the Editor, Proc. I.R.E. 44, 1057 (1956).

22. R. Price, "Optimum Detection of Bandpass Gaussian Signals in White Gaussian Noise, with Application to Receiver Design for Scatter-Path Communication – I," to be submitted to Trans. I.R.E., Professional Group on Information Theory, for publication in December 1956 issue.

APPENDIX

Our purpose here is to evaluate the integral

$$J_{rp} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\lambda^2} \frac{\sin \pi (c\lambda - r)}{\pi (c\lambda - r)} \frac{\sin \pi (c\lambda - p)}{\pi (c\lambda - p)} \, d\lambda \quad ,$$

with r and p integers.

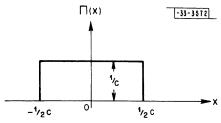

Fig. A-1. The pulse, Π(x).

If Π(x) is defined as in Fig. A-1, we have

$$\int_{-\infty}^{\infty} \Pi(x) \, e^{\pm j2\pi x(\lambda - r/c)} \, dx = \frac{\sin \pi (c\lambda - r)}{\pi (c\lambda - r)} \quad .$$

When these integral representations are used for the $\frac{\sin x}{x}$-functions, we have

$$J_{rp} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\lambda^2} \, d\lambda \int_{-\infty}^{\infty} \Pi(x) \, e^{-j2\pi x(\lambda - r/c)} \, dx \int_{-\infty}^{\infty} \Pi(y) \, e^{j2\pi y(\lambda - p/c)} \, dy \quad ,$$

or, after inverting the order of integration,

$$J_{rp} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \Pi(x) \, e^{j2\pi xr/c} \, dx \int_{-\infty}^{\infty} \Pi(y) \, e^{-j2\pi yp/c} \, dy \int_{-\infty}^{\infty} e^{-\lambda^2} e^{j2\pi(y-x)\lambda} \, d\lambda \quad .$$

The last integral, which we will call I_1, may be evaluated by completing the square:

$$I_1 = \int_{-\infty}^{\infty} e^{-\lambda^2} e^{j2\pi(y-x)\lambda} \, d\lambda$$

$$= e^{-\pi^2(y-x)^2} \int_{-\infty}^{\infty} e^{-[\lambda^2 - j2\pi(y-x)\lambda - \pi^2(y-x)^2]} \, d\lambda$$

$$= e^{-\pi^2(y-x)^2} \int_{-\infty}^{\infty} e^{-[\lambda - j\pi(y-x)]^2} \, d\lambda \quad ,$$

or, with $\mu = \lambda - j\pi (y - x)$,

$$I_1 = e^{-\pi^2(y-x)^2} \int_{-\infty}^{\infty} e^{-\mu^2} \, d\mu = \sqrt{\pi} \, e^{-\pi^2(y-x)^2} \quad ,$$

and J_{rp} becomes

$$J_{rp} = \int_{-\infty}^{\infty} \sqcap(x)\, e^{j2\pi xr/c}\, dx \int_{-\infty}^{\infty} \sqcap(y)\, e^{-j2\pi yp/c}\, e^{-\pi^2(y-x)^2}\, dy \quad .$$

Next we make two changes of variable. First we replace y by (x + u) in the second integral and change the order of integration, giving

$$J_{rp} = \int_{-\infty}^{\infty} e^{-(\pi u)^2}\, e^{-j2\pi up/c}\, du \int_{-\infty}^{\infty} \sqcap(x) \sqcap(x+u)\, e^{j2\pi x(r-p)/c}\, dx \quad .$$

Then we replace x by $v - \frac{u}{2}$ in the second integral:

$$J_{rp} = \int_{-\infty}^{\infty} e^{-(\pi u)^2}\, e^{-j\pi u(r+p)/c}\, du \int_{-\infty}^{\infty} \sqcap\left(v-\tfrac{u}{2}\right) \sqcap\left(v+\tfrac{u}{2}\right) e^{j2\pi v(r-p)/c}\, dv \quad .$$

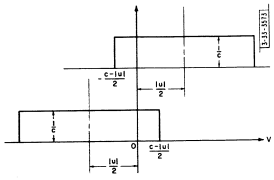

Fig. A-2. The two Π-functions are symmetrically displaced.

We will call the second integral I_2. Apparently it is an even function of u; the magnitude of its integrand is the product of the two Π-functions shown in Fig. A-2. Note that if $|u| > c$, $I_2 = 0$, so we need only be concerned with the range $-c < u < c$. If $p = r$,

$$I_2 = \frac{c - |u|}{c^2} \quad ,$$

and if $p \neq r$,

$$I_2 = \frac{1}{c^2} \int_{-(c-|u|)/2}^{(c-|u|)/2} e^{j2\pi v(r-p)/c}\, dv = \frac{\sin\pi\left(1 - \frac{|u|}{c}\right)(r-p)}{\pi c\,(r-p)}$$

$$= \frac{\sin\pi(r-p)\cos\frac{\pi|u|(r-p)}{c} - \cos\pi(r-p)\sin\frac{\pi|u|(r-p)}{c}}{\pi c\,(r-p)}$$

$$= -(-1)^{(r-p)}\, \frac{\sin\frac{\pi|u|(r-p)}{c}}{\pi c\,(r-p)} \quad ,$$

since r and p are integers.

Taking the special case p = r first,

$$J_{rr} = \frac{1}{c^2} \int_{-c}^{c} (c - |u|)\, e^{-(\pi u)^2}\, e^{-j2\pi ur/c}\, du \quad .$$

If this integral is split into the sum of two integrals having the ranges $-c < u < 0$ and $0 < u < c$, and if u is replaced by $-u$ in the first, the sum may be written

$$J_{rr} = \frac{2}{c^2} \int_0^c (c-u)\, e^{-(\pi u)^2} \cos 2\pi u\, \frac{r}{c}\, du \quad ,$$

or, with u = cx,

$$J_{rr} = 2 \int_0^1 (1-x)\, e^{-(\pi cx)^2} \cos 2\pi rx\, dx \quad . \qquad \text{(A-1)}$$

In the more general case when $p \neq r$,

$$J_{rp} = -\frac{(-1)^{(r-p)}}{\pi c\,(r-p)} \int_{-c}^{c} e^{-(\pi u)^2} \sin\frac{\pi|u|(r-p)}{c}\, e^{-j\pi u(r+p)/c}\, du \quad .$$

The range of this integral may also be cut in half as before:

$$J_{rp} = -\frac{2(-1)^{(r-p)}}{\pi c\,(r-p)} \int_0^c e^{-(\pi u)^2} \sin\frac{\pi u(r-p)}{c} \cos\frac{\pi u(r+p)}{c}\, du \quad ,$$

and, with the aid of the appropriate trigonometric identity,

$$J_{rp} = -\frac{(-1)^{(r-p)}}{\pi(r-p)} (I_r - I_p) \quad , \qquad \text{(A-2)}$$

in which

$$I_r = \frac{1}{c} \int_0^c e^{-(\pi u)^2} \sin 2\pi u\, \frac{r}{c}\, du \quad ,$$

or, with u = cx,

$$I_r = \int_0^1 e^{-(\pi cx)^2} \sin 2\pi rx\, dx \quad . \qquad \text{(A-3)}$$

The integrals (A-1) and (A-3) cannot be performed by elementary methods. However, if the upper limits are arbitrarily extended to $+\infty$, they can both be evaluated in terms of tabulated functions, and if the parameter c (the normalized aperture dimension) is large, the error involved will be small. Even with c as small as unity, the exponential has dropped to about 50×10^{-6} at the present upper limit! Fortunately, we will be interested in values of c of the order of unity or greater, so the approximation appears to be reasonable. We will calculate approximations for the integrals and give an upper bound for the errors involved. Starting with the second integral, (A-3),

$$I_r \approx \int_0^{\infty} e^{-(\pi cx)^2} \sin 2\pi rx\, dx = \mathcal{I}m \left\{ \int_0^{\infty} e^{-(\pi cx)^2}\, e^{j2\pi rx}\, dx \right\} \quad .$$

Completing the square,

$$I_r \approx \mathcal{Im}\left\{e^{-(r/c)^2}\int_0^\infty e^{-[(\pi cx)^2 - j2\pi rx - (r/c)^2]}\,dx\right\}$$

$$= e^{-(r/c)^2}\,\mathcal{Im}\left\{\int_0^\infty e^{-(\pi cx - jr/c)^2}\,dx\right\} ,$$

or, with the complex variable, $z = (\pi cx - j\,r/c)$,

$$I_r \approx \frac{1}{\pi c}\,e^{-(r/c)^2}\,\mathcal{Im}\left\{\int_{-jr/c}^{\infty-jr/c} e^{-z^2}\,dz\right\} .$$

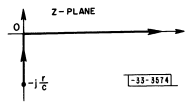

Fig. A-3. Path of integration in the complex plane.

The path of integration may be taken to follow the imaginary axis to the origin and then along the real axis to infinity, as shown in Fig. A-3. The integral along the real axis is pure real, so we are left with the integral along the imaginary axis. With $z = -jy$, the latter integral becomes

$$I_r \approx \frac{1}{\pi c}\,e^{-(r/c)^2}\,\mathcal{Im}\left\{-j\int_{r/c}^{0} e^{y^2}\,dy\right\}$$

$$= \frac{1}{\pi c}\,e^{-(r/c)^2}\int_0^{r/c} e^{y^2}\,dy .$$

The function

$$D(x) = e^{-x^2}\int_0^x e^{y^2}\,dy \qquad \text{(A-4)}$$

is tabulated.[17] In terms of this function, the approximation to I_r reads

$$I_r \approx \frac{1}{\pi c}\,D\left(\frac{r}{c}\right) . \qquad \text{(A-5)}$$

The approximation to J_{rr} is

$$J_{rr} \approx 2\int_0^\infty (1-x)\,e^{-(\pi cx)^2}\cos 2\pi\,rx\,dx$$

$$= 2\int_0^\infty e^{-(\pi cx)^2}\cos 2\pi\,rx\,dx - 2\int_0^\infty x\,e^{-(\pi cx)^2}\cos 2\pi\,rx\,dx$$

$$\equiv I_3 + I_4 .$$

The first integral may be written

$$I_3 = \int_{-\infty}^{\infty} e^{-(\pi cx)^2}\,e^{j2\pi rx}\,dx ,$$

and it may be performed by completing the square. The result is

$$I_3 = \frac{1}{\sqrt{\pi}\,c}\,e^{-(r/c)^2} .$$

The second integral may be integrated by parts:

$$I_4 = \frac{1}{\pi^2 c^2}\int_0^\infty \cos 2\pi\,rx\,d\left[e^{-(\pi cx)^2}\right]$$

$$= \frac{1}{\pi^2 c^2}\cos 2\pi\,rx\,e^{-(\pi cx)^2}\Big|_0^\infty + \frac{2r}{\pi c^2}\int_0^\infty e^{-(\pi cx)^2}\sin 2\pi\,rx\,dx .$$

The integral is recognized as the approximation to I_r, hence

$$I_4 = -\frac{1}{\pi^2 c^2} + \frac{2r}{\pi^2 c^3}\,D\left(\frac{r}{c}\right) ,$$

and the approximation to J_{rr} is

$$J_{rr} \approx \frac{1}{\sqrt{\pi}\,c}\,e^{-(r/c)^2} - \frac{1}{\pi^2 c^2} + \frac{2r}{\pi^2 c^3}\,D\left(\frac{r}{c}\right) . \qquad \text{(A-6)}$$

The approximation to J_{rp}, from (A-2) and (A-5), is

$$J_{rp} \approx -\frac{(-1)^{(r-p)}}{\pi^2 (r-p)\,c}\left[D\left(\frac{r}{c}\right) - D\left(\frac{p}{c}\right)\right] . \qquad \text{(A-7)}$$

It may be noted from (A-1) that the central integral, J_{oo}, can be expressed exactly in terms of tabulated functions:

$$J_{oo} = 2\int_0^1 e^{-(\pi cx)^2}\,dx - \int_0^1 e^{-(\pi cx)^2}\,2x\,dx$$

$$= \frac{1}{\sqrt{\pi}\,c}\left\{\frac{2}{\sqrt{\pi}}\int_0^{\pi c} e^{-y^2}\,dy\right\} - \frac{1}{\pi^2 c^2}\left[1 - e^{-(\pi c)^2}\right] . \qquad \text{(A-8)}$$

This central integral is of primary importance in determining the power available at the antenna terminals; the power available from an antenna with a square aperture having maximum plane-wave gain is proportional to $g_{11} = J_{oo}^2$. The exact value (A-8) is used in the numerical calculations described in Sec. VII.

We will not attempt to relate the error in the maximum characteristic value of the g-matrix to the errors involved in the approximations (A-6) and (A-7). In giving an upper bound for the latter, it will be appropriate to compare the errors with the central integral:

$$\epsilon_{rp} = \frac{|J_{rp}\,(\text{exact}) - J_{rp}\,(\text{approx.})|}{J_{oo}\,(\text{approx.})} .$$

From (A-6),

$$J_{oo} \approx \frac{1}{\sqrt{\pi}\,c} - \frac{1}{\pi^2 c^2} ,$$

and it will be convenient to use

$$J_{oo} \approx \frac{1}{\sqrt{\pi}\, c} \quad ;$$

for c greater than 1/2, this error in the error calculation will reduce the true error by a factor lying between 0.6 and 1.0.

From (A-1), the fractional error in J_{rr} is

$$\epsilon_{rr} = 2\sqrt{\pi}\, c \left| \int_1^\infty (1 - x)\, e^{-(\pi c x)^2} \cos 2\pi\, rx\, dx \right|$$

$$\leq 2\sqrt{\pi}\, c \int_1^\infty (x - 1)\, e^{-(\pi c x)^2}\, dx$$

$$= \sqrt{\pi}\, c \int_1^\infty e^{-(\pi c x)^2}\, 2x\, dx - 2\sqrt{\pi}\, c \int_1^\infty e^{-(\pi c x)^2}\, dx$$

$$= \frac{1}{\pi^{3/2}\, c}\, e^{-(\pi c)^2} - \left[1 - \frac{2}{\sqrt{\pi}} \int_0^{\pi c} e^{-y^2}\, dy \right] \quad . \tag{A-9}$$

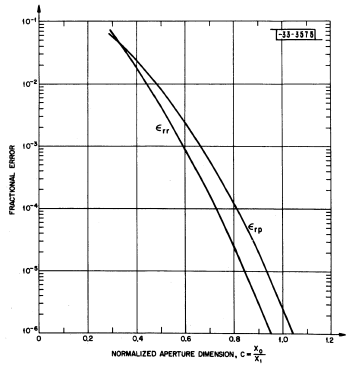

Fig. A-4. The fractional errors soon become insignificant as the normalized aperture dimension increases.

And from (A-2) and (A-3), the fractional error in J_{rp} is

$$\epsilon_{rp} = \sqrt{\pi}\, c \left| -\frac{(-1)^{(r-p)}}{\pi(r-p)} \int_1^\infty e^{-(\pi c x)^2}\, [\sin 2\pi\, rx - \sin 2\pi\, px]\, dx \right|$$

$$< \frac{2c}{\sqrt{\pi}} \int_1^\infty e^{-(\pi c x)^2}\, dx$$

$$= \frac{1}{\pi} \left[1 - \frac{2}{\sqrt{\pi}} \int_0^{\pi c} e^{-y^2}\, dy \right] \quad . \tag{A-10}$$

The upper bounds on the fractional errors, (A-9) and (A-10), are plotted in Fig. A-4. It will be noted that the fractional errors are less than 1 part in 10^3 if c is greater than 0.66. The approximations (A-6) and (A-7) are only used in cases where $c \geq 0.66$.