PHYSICAL LIMITATIONS OF OMNIDIRECTIONAL ANTENNAS

L J CHU

TECHNICAL REPORT NO. 64

MAY 1, 1948

RESEARCH LABORATORY OF ELECTRONICS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

The research reported in this document was made possible through support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the Army Signal Corps, the Navy Department (Office of Naval Research), and the Air Force (Air Materiel Command), under the Signal Corps Contract No. W-36-039 sc-32037.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Research Laboratory of Electronics

Technical Report No. 64 May 1, 1948

PHYSICAL LIMITATIONS OF OMNIDIRECTIONAL ANTENNAS

L. J. Chu

Abstract

The physical limitations of omnidirectional antennas are considered. With the use of the spherical wave functions to describe the field, the directivity gain G and the Q of an unspecified antenna are calculated under idealized conditions. To obtain the optimum performance, three criteria are used: (1) maximum gain for a given complexity of the antenna structure, (2) minimum Q, (3) maximum ratio of G/Q. It is found that an antenna of which the maximum dimension is 2a has the potentiality of a broad bandwidth provided that the gain is equal to or less than $4a/\lambda$. To obtain a gain higher than this value, the Q of the antenna increases at an astronomical rate. The antenna which has potentially the broadest bandwidth of all omnidirectional antennas is one which has a radiation pattern corresponding to that of an infinitesimally small dipole.

PHYSICAL LIMITATIONS OF OMNIDIRECTIONAL ANTENNAS

1. Introduction

An antenna system, functioning as a transmitter, provides a practical means of transmitting, to a distant point or points in space, a signal which appears in the form of r-f energy at the input terminals of the transmitter. The performance of such an antenna system is judged by the quality of transmission, which is measured by both the efficiency of transmission and the signal distortion. At a single frequency, transmission efficiency is determined by the power gain of the antenna system in a desired direction or directions. The distortion depends not only on the frequency characteristics of the antenna input impedance, but also on variations of phase and of power gain with frequency. It is common practice to describe the performance of an antenna system in terms of its power gain and the bandwidth of its input impedance.

Designers of antennas at VLF range have always been faced with the problems of excessive conduction losses in the antenna structure and a narrow bandwidth. At this frequency range, the physical size of the antenna is necessarily small in terms of the operating wavelength. For a broadcasting antenna, with a specified distribution of radiated power in space, it was found that the antenna towers must be spaced at a sufficient distance apart so as not to have excessive currents on the towers. At microwave frequencies, where a high gain has been made possible with a physically small antenna, there seems to be a close relationship between the maximum gain thus far obtainable and the size of the antenna expressed in terms of the operating wavelength. At optical frequencies where a different language is used, the resolving power of a lens or a reflector is proportional to the ratio of the linear dimension to wavelength. Thus, over the entire frequency range, there seems to be a practical limit to the gain or the directivity of a radiating or focussing system.

From time to time, there arises the question of achieving a higher gain from an antenna of given size than has been obtained conventionally. Among published articles, Schelkunoff[1] has derived a mathematical expression for the current distribution along an array which yields higher directivity gain than that which has been usually obtained. It is mentioned at the end of this article that an array carrying this current distribution would have a narrow bandwidth as well as high conduction loss. In 1943, LaPaz and Miller[2] obtained an optimum current distribution on a vertical antenna of given length which gives the maximum possible field strength on the horizon for a given power output. The result was disputed in a later paper by Bouwkamp and deBruijn[3] who developed a method of realizing an

---

1. S. A. Schelkunoff, B.S.T.J. V. 22, pp. 80-107, Jan. 1943.
2. L. LaPaz and G. A. Miller, Proc. I.R.E. V.31, pp. 214-232, May, 1943.
3. C. J. Bouwkamp and N. G. deBruijn , Philips Research Reports V. 1, pp. 135-158, Jan. 1946.

arbitrarily sharp vertical radiation pattern by a suitable choice of the current distribution. In a later report by Laemmel of the Polytechnic Institute of Brooklyn, a method was presented for finding a source distribution function which results in an arbitrarily large gain relative to an isotropic radiator and which at the same time is contained within an arbitrarily small region of space.

In all the above articles, the authors invariably have computed the source distribution required to obtain a directivity gain higher than that obtained in practice with an antenna of a given size. As a result, it can be said that there is no mathematical limit to the directivity gain of an antenna of given size. The possibility of arranging on paper the current distribution on an antenna at r-f frequencies exists because of the absence of the severe restriction which must be observed, on account of the incoherent nature of the energy, in designing a system at optical frequencies.

In 1941, Stratton demonstrated the impracticality of supergain antennas. In his unpublished notes he derived the source distribution within a sphere of finite radius for any prescribed distribution of the radiation field in terms of a complete set of orthogonal, spherical, vector wave functions[4]. Mathematically, the series representing the source distribution diverges as the directivity gain of the system increases indefinitely. Physically, high current amplitude on the antenna, if it can be realized, implies high energy storage in the system, a large power dissipation, and a low transmission efficiency.

This paper presents an attempt to determine the optimum performance of an antenna in free space and the corresponding relation between its gain and the bandwidth of the input impedance under various criteria. Let the largest linear dimension of the antenna structure be 2a, such that the complete antenna structure including transmission lines and the oscillator can be enclosed inside a geometrical spherical surface of radius a. The field outside the sphere due to an arbitrary current or source distribution inside the sphere can be expressed in terms of a complete set of spherical vector waves[4]. Each of these waves represents a spherical wave propagating radially outward. However, the current or source distribution <u>inside</u> the sphere is not uniquely determined by the field distribution <u>outside</u> the sphere. It is mathematically possible to create a given field distribution outside the sphere with an infinite number of different source distributions. We shall confine our interest to the most favorable source distribution and the corresponding antenna structure. To circumvent the difficult task of determining the latter, the most favorable conditions will be assumed to exist inside the sphere. The current or source distribution inside the sphere necessary to produce the desired field distribution outside will be assumed to require the minimum amount of energy stored inside the sphere so that one has a pure resistive input impedance at a single frequency. Also, to simplify the problem, the conduction loss will be neglected.

Under these conditions it is not possible to calculate the behavior of this

---

4. J. A. Stratton, "Electromagnetic Theory", Ch. 7, p. 392, McGraw-Hill, 1941.

antenna over a finite frequency range since the exact nature of the antenna structure is not known. At one frequency we can determine the radiation characteristics of the system from the expressions for the field, including the directivity gain of the antenna in a given direction. The directivity gain is equal to the power gain in the absence of conduction loss in the antenna structure. We shall utilize the conventional concept of Q, computed from the energy and power at a single frequency, to obtain the frequency characteristics of the input impedance by extrapolation. It is understood that the physical interpretation of Q as so computed becomes rather vague whenever the value of Q is low.

After obtaining the gain and Q of the antenna corresponding to an arbitrary field distribution outside the sphere, we then proceed to determine the optimum distribution of the field outside the sphere under different criteria and the corresponding gains and Q through the process of maximization and minimization.

Antennas can be classified according to their radiation characteristics as follows: (1) omni-directional antennas, (2) pencil-beam antennas, (3) fanned-beam antennas, and (4) shaped-beam antennas. The first type of antenna will be discussed in detail in this article. The physical limitations of pencil-beam antennas will be dealt with in an article to be published later.

The problem has been worked out independently by Ramo and Taylor of Hughes Aircraft with a slightly different procedure. Their results are essentially in agreement with what follows.

2. Analysis

2.1 Field of a Vertically Polarized Omnidirectional Antenna. The type of antenna under consideration here gives rise to an omnidirectional pattern. It is commonly used as a beacon or broadcasting antenna. The radiated power is distributed uniformly around a vertical axis, which we take as the axis of a spherical coordinate system (R,θ,ϕ). We shall discuss first the case where the electric field lies in planes passing through the axis of symmetry. The antenna is schematically shown in Fig. 1 and lies totally within a spherical surface of radius a. For an arbitrary current distribution and antenna structure, the field outside the sphere can be expressed in terms of a complete set of orthogonal, spherical waves, propagating radially outward. For the type of antenna under consideration only TM_{no} waves are required to describe the circularly symmetrical field with the specified polarization. By ignoring all the other spherical waves we have the expressions of the three non-vanishing field components:

$$\begin{aligned}
H_\phi &= \sum_n A_n P_n^1(\cos\theta)\, h_n(kR) \\
E_R &= -j\sqrt{\frac{\mu}{\epsilon}} \sum_n A_n\, n(n+1)\, P_n(\cos\theta)\, \frac{h_n(kR)}{kR} \qquad (1) \\
E_\theta &= j\sqrt{\frac{\mu}{\epsilon}} \sum_n A_n P_n^1(\cos\theta)\, \frac{1}{kR}\frac{d}{dR}[R\, h_n(kR)]
\end{aligned}$$

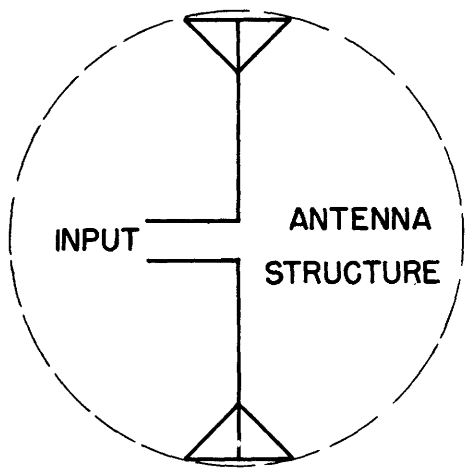

Figure 1. Schematic diagram of a vertically polarized omnidirectional antenna.

where

$P_n(\cos\theta)$ is the Legendre polynomial of order n,

$P_n^1(\cos\theta)$ is the first associated Legendre polynomial,

$h_n(kR)$ is the spherical Hankel function of the second kind.

$k = \omega\sqrt{\epsilon\mu} = 2\pi/\lambda$

$\sqrt{\mu/\epsilon}$ is the wave impedance of a plane wave in free space and

$1/\sqrt{\epsilon\mu}$ is the velocity of a plane wave in free space.

The time factor $e^{j\omega t}$ is omitted throughout the paper and the rationalized MKS unit system is used. The coefficients A_n's are, in general, complex quantities. In synthesis problems, the A_n's are specified by the desired radiation characteristics. When the antenna structure is given, the A_n's are determined from the boundary conditions on the surfaces of the antenna structure. For the time being, the A_n's are a set of unspecified coefficients.

2.2 Radiation Characteristics. At a sufficiently large distance from the sphere, the transverse field components become asymptotically

$$\left.\begin{aligned} E_\theta &= \sqrt{\frac{\mu}{\epsilon}}\,\frac{e^{-jkR}}{kR}\sum_n A_n\,(-1)^{\frac{n+1}{2}}\,P_n^1(\cos\theta) \\ H_\phi &= \sqrt{\frac{\epsilon}{\mu}}\,E_\theta \end{aligned}\right\}. \qquad (2)$$

The angular distribution of the radiation field is given by the series of the associated Legendre polynomials. This series behaves somewhat like a Fourier series in the interval from $\theta = 0$ to $\theta = \pi$. Using the conventional definition for the directivity gain, we have

$$G(\theta) = \frac{4\pi \; |E_\theta|^2}{\int_0^\pi \int_0^{2\pi} |E_\theta|^2 \sin\theta \, d\theta \, d\phi} = \frac{\left|\sum_n A_n (-1)^{\frac{n+1}{2}} \; P_n^1(\cos\theta)\right|^2}{\sum_n |A_n|^2 \frac{n(n+1)}{2n+1}} . \quad (3)$$

The denominator is obtained from the orthogonality of the associated Legendre polynomials:

$$\int_0^\pi \left[P_n^1(\cos\theta)\right]^2 \sin\theta \, d\theta = \frac{2n(n+1)}{2n+1}$$

and

$$\int_0^\pi P_n^1(\cos\theta) \, P_{n'}^1 (\cos\theta) \sin\theta \, d\theta = 0 \quad \text{for } n \neq n'.$$

We shall limit our attention to the gain in the equatorial plane $\theta = \pi/2$. In this plane,

$$P_n^1(0) = 0 \text{ for n even, and}$$

$$P_n^1(0) = (-1)^{\frac{n-1}{2}} \frac{n!}{2^{n-1}(\frac{n-1}{2}!)^2} \quad \text{for n odd} \quad .$$

Thus the terms of even n do not contribute to the radiation field along the equator. In order to have a high directivity gain in the equatorial plane it is necessary to have

$$A_n = 0 \text{ for n even}$$

while all the rest of the A_n's must have the same phase angle. From here on, we shall consider all A_n's to be positive real quantities for odd n and zero for even n. Thus the directivity gain on the equatorial plane becomes

$$G(\frac{\pi}{2}) = \frac{\left[\Sigma' A_n (-1)^{\frac{n+1}{2}} \; P_n^1(0)\right]^2}{\Sigma' A_n^2 \; \frac{n(n+1)}{2n+1}} \quad (4)$$

where Σ' represents the sum over odd n only.

<u>2.3 Power and Energy Outside the Sphere.</u> With the field of the omnidirectional antenna outside the sphere given by Eqs. (1), the total complex power computed at the surface of the sphere is the integral of the complex Poynting vector over the same sphere:

$$P(a) = j2\pi \sqrt{\frac{\mu}{\epsilon}} \quad \Sigma' \left\{\frac{A_n}{k}\right\}^2 \frac{n(n+1)}{2n+1} \; \rho h_n (\rho h_n)' \quad (5)$$

where

$$\rho = ka$$
$$h_n = h_n(\rho)$$
$$(\rho h_n)' = \frac{d}{d\rho}\rho h_n(\rho) \quad .$$

The average power radiated is the real part of (5):

$$P_{av} = 2\pi\sqrt{\frac{\mu}{\epsilon}}\sum{}' \left(\frac{A_p}{k}\right)^2 \frac{n(n+1)}{2n+1} \quad . \tag{6}$$

It is possible to calculate the total electric energy and the magnetic energy stored outside the sphere. On the c-w basis the total stored energy is infinitely large provided any one of the A_n's is finite, since the radiation field which is inversely proportional to the radial distance extends to infinity. As in the case of an infinite transmission line with no dissipation, most of the energy appears in the form of a traveling wave which propagates toward infinity and never returns to the source. The total energy calculated on this basis has no direct bearing upon the performance of the antenna. It is difficult to separate the energy associated with the local field in the neighborhood of the antenna from the remainder. The energy is not linear in the field components and hence the law of linear superposition cannot be applied directly to it. The imaginary part of the integral of the complex Poynting vector is proportional only to the difference of the electrical and magnetic energy stored outside the sphere. In order to separate the energy associated with radiation from that associated with the local field, we shall reduce the field problem to a circuit problem where the radiation loss is replaced by an equivalent conduction loss.

2.4 Equivalent Circuits for Spherical Waves. Because of the orthogonal properties of the spherical wave functions employed, the total energy, electric or magnetic, stored outside the sphere is equal to the sum of the corresponding energies associated with each spherical wave, and the complex power transmitted across a closed spherical surface is equal to the sum of the complex powers associated with each spherical wave. Insofar as the total energies and power are concerned, there is no coupling between any two of the spherical waves outside the sphere. Consequently, we can replace the space outside the sphere by a number of independent equivalent circuits, each with a pair of terminals connected to a box which represents the inside of the sphere. From this box, we can pull out a pair of terminals representing the input to the antenna as shown in Fig. 2. The total number of pairs of terminals is equal to the number of spherical waves used in describing the field outside the sphere, plus one. We have now managed to transform a space problem to the problem of its equivalent circuit.

When the field outside the sphere has been specified by Eq.(1), the current, voltage, and impedance of the equivalent circuit for each spherical wave are uniquely

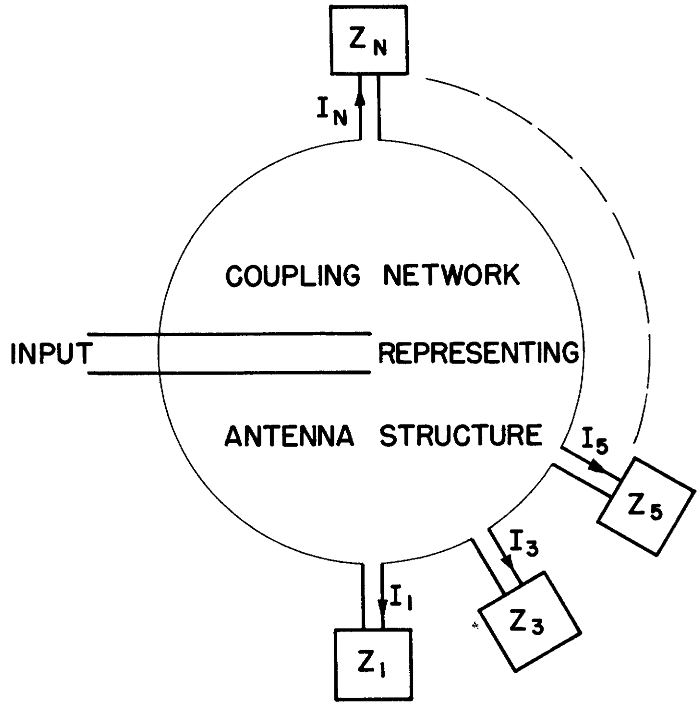

Figure 2. Equivalent circuit of a vertically polarized omnidirectional antenna.

determined except for an arbitrary real transformer ratio by equating the complex power associated with the spherical waves to that of the circuits. We shall define the voltage, current, and impedance of the equivalent circuit for the TM_n wave as:

$$
\begin{aligned}
V_n &= \sqrt[4]{\frac{\mu}{\epsilon}}\ \frac{A_n}{k} \sqrt{\frac{4\pi n(n+1)}{2n+1}}\ j(\rho h_n)' \\
I_n &= \sqrt[4]{\frac{\mu}{\epsilon}}\ \frac{A_n}{k} \sqrt{\frac{4\pi n(n+1)}{2n+1}}\ \rho h_n
\end{aligned}
\qquad (7)
$$

and

$$Z_n = j(\rho h_n)' / \rho h_n \ .$$

The voltage is proportional to the θ component of electric field and the current is proportional to the magnetic field H_ϕ of the TM_n wave on the surface of the sphere. The normalized impedance Z_n is equal to the normalized radial wave impedance on the surface. It can be shown that not only is the complex power which is fed into the equivalent circuit equal to the complex power associated with the TM_n wave but the instantaneous powers are also equal to each other. The impedance Z_n of the equivalent circuit is a physically realizable impedance and Eq. (7) is valid at all frequencies.

Using the recurrence formulas of the spherical Bessel functions, one can write the impedance Z_n as a continued fraction:

$$Z_n = \frac{n}{j\rho} + \cfrac{1}{\cfrac{2n-1}{j\rho} + \cfrac{1}{\cfrac{2n-3}{j\rho} + \ddots \, \cfrac{1}{\cfrac{3}{j\rho} + \cfrac{1}{\cfrac{1}{j\rho} + 1}}}} \quad . \tag{8}$$

This can be interpreted as a cascade of series capacitances and shunt inductances terminated with a unit resistance. For n = 1, the impedance consists of the three elements shown in Fig. 3. This is the equivalent circuit representing a wave which could be generated by an infinitesimally small dipole. At low frequencies, most of the voltage applied to the terminal appears across the capacitance, and the unit resistance is practically short-circuited by the inductance. At high frequencies,

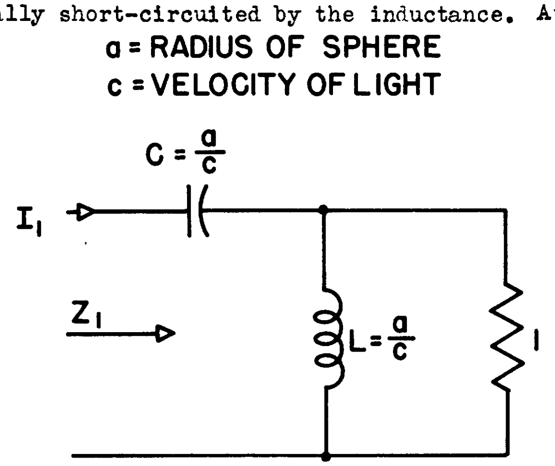

Figure 3. Equivalent circuit of electric dipole.

the impedance Z_1 is practically a pure resistance of amplitude unity. At intermediate frequencies, the reactance of Z_1 remains capacitive. The equivalent circuit of Z_n is schematically shown in Fig. 4. The circuit, for all values of n,

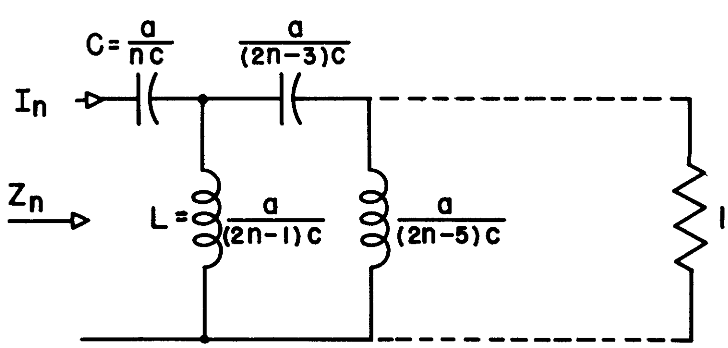

Figure 4. Equivalent circuit of TM_n spherical wave.

acts as a high-pass filter. With the dissipative element hidden at the very end of the cascade, the difficulty of feeding average power into the dissipative element at a single frequency increases with the order of the wave. The dissipation in the resistance is equal to the radiation loss in the space problem. The capacitances and inductances are proportional to the ratio of the radius of the sphere to the velocity of light. The increase of the radius of the sphere has the same effect on the behavior of the equivalent circuit as the increase of frequency.

Except for the equivalent circuit of the electric dipole, it would be tedious to calculate the total electric energy stored in all the capacitances of the equivalent circuit for Z_n. We shall therefore approximate the equivalent circuit for Z_n by a simple series RLC circuit which has essentially the same frequency behavior in the neighborhood of the operating frequency. We compute R_n, L_n, and C_n of the simplified equivalent circuit by equating the resistance, reactance, and the frequency derivative of the reactance of Z_n to those of the series RLC circuit. The series resistance R_n is of course equal to the real part of Z_n at the operating frequency. The results are the following:

$$
\begin{aligned}
R_n &= \left|\rho h_n\right|^{-2} \\
C_n &= \frac{2}{\omega^2}\left[\frac{dX_n}{d\omega} - \frac{X_n}{\omega}\right]^{-1} \\
L_n &= \frac{1}{2}\left[\frac{dX_n}{d\omega} + \frac{X_n}{\omega}\right]
\end{aligned}
\tag{9}
$$

where

$$X_n = \left[\rho j_n(\rho j_n)' + \rho n_n(\rho n_n)'\right] \left|\rho h_n\right|^{-2}$$

and j_n and n_n are the spherical Bessel functions of the first and second kind. Except for the frequency variation of the resistance R_n in the original equivalent circuit, the simplified circuit is accurate enough to describe Z_n in the immediate neighborhood of the operating frequency.

Based upon this simplified equivalent circuit, the average power dissipation in Z_n is

$$P_n = \sqrt{\frac{\mu}{\epsilon}}\ \frac{2\pi n(n+1)}{2n+1}\left(\frac{A_n}{k}\right)^2 \tag{10}$$

and the average electric energy stored in Z_n is

$$W_n = \sqrt{\frac{\mu}{\epsilon}}\ \frac{\pi n(n+1)}{2(2n+1)}\left(\frac{A_n}{k}\right)^2 \left|\rho h_n\right|^2 \left[\frac{dX_n}{d\omega} - \frac{X_n}{\omega}\right] \tag{11}$$

which is larger than the average magnetic energy stored in Z_n. We shall define Q_n for the TM_n wave as

$$Q_n = \frac{2\omega W_n}{P_n} = \frac{1}{2}\left|\rho h_n\right|^2 \left[\rho \frac{dX_n}{d\rho} - X_n\right]. \tag{12}$$

The bandwidth of the equivalent circuit of the TM_n wave is equal to the reciprocal of Q_n, when it is matched externally with a proper amount of stored magnetic energy. When Q_n is low, the above interpretation is not precise, but it does indicate qualitatively the frequency sensitivity of the circuit.

In Fig. 5, Q_n of the TM_n waves is plotted against $2\pi a/\lambda$. We observe that Q_n is of the order of unity or less whenever the abscissa $2\pi a/\lambda$ is greater than the order n of the wave. Here the stored electric energy in the equivalent circuit of the wave is insignificant and the circuit behaves essentially as a pure resistance. When $2\pi a/\lambda$ is less than n, the circuit behaves essentially as a pure capacitance. Q_n increases at an astronomical rate as the abscissa decreases. In terms of wave propagation, the TM_n wave will propagate from the surface of the sphere without an excessive amount of energy stored in the neighborhood of the sphere only when the radius of the sphere is greater than $n\lambda/2\pi$.

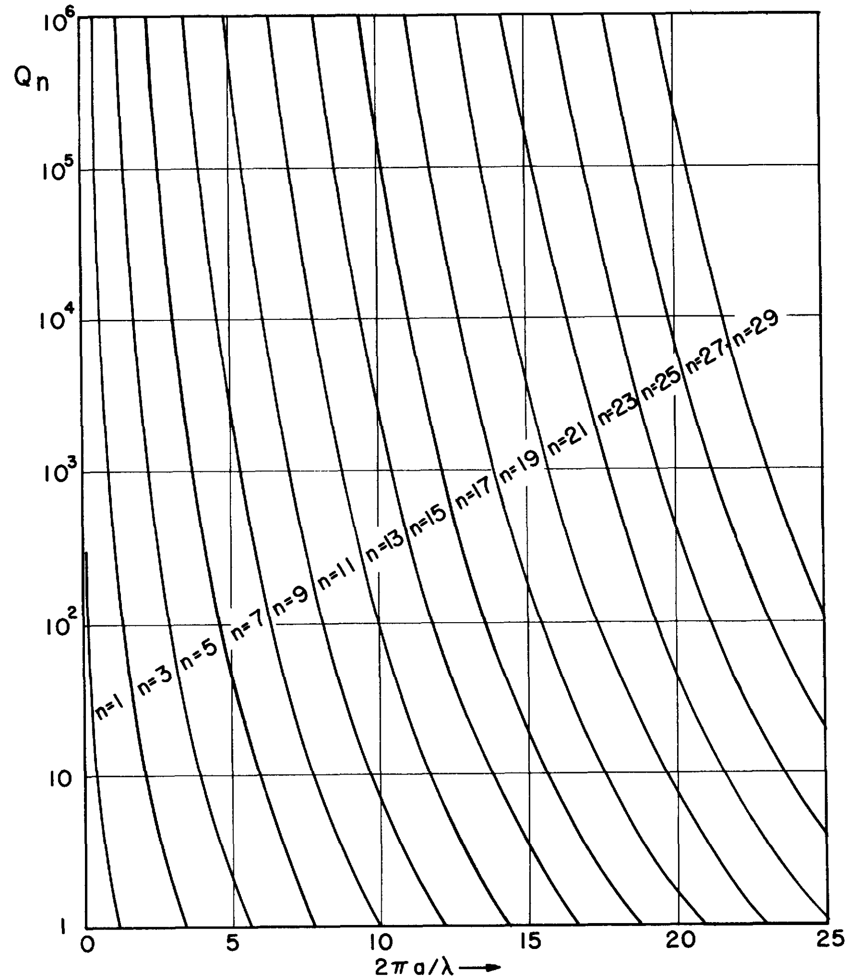

Figure 5. Q_n of the equivalent circuit of TM_n or TE_n wave.

<u>2.5 Equivalent Circuit of the Antenna.</u> The complete equivalent circuit of the antenna system is shown in Fig. 2. The circular box is a coupling network representing the space inside the geometrical sphere shown in Fig. 1. It couples the system feeding the input terminals to the various equivalent circuits of the TM waves connected externally to the box. The voltage V_n and current I_n are those given by Eq. (7).

In Sec. 2.2 on radiation characteristics, it was pointed out that for each term to contribute positively to the gain of the antenna in the equatorial plane, it is necessary for all A_n's to have the same phase angle. The spherical Hankel function $h_n(\rho)$ is essentially a positive imaginary quantity when its argument is less than the order n. Thus, the currents of the equivalent circuits Z_n for n greater than the argument, are essentially in phase, and the instantaneous electric energies stored in all the equivalent circuits oscillate in phase.

We have calculated the power dissipation as well as the average energy stored in Z_n for the simplified circuit of the TM_n wave. The total electric energy stored and the power dissipated in all the circuits connected to the coupling network is equal to the summation of W_n and P_n, respectively. The total power dissipated $\sum_n' P_n$ is, of course, equal to the total power radiated into space, while the total electric energy stored $\sum_n' W_n$ is that associated with the local field outside of the sphere.

Theoretically, there is no unique antenna structure or source distribution inside the sphere which generates the field distribution given by Eq. (1). Consequently, the coupling network representing the space inside the sphere is not unique. The process of determining the optimum antenna structure for a given field distribution outside or the optimum coupling network is by no means a simple matter. To simplify the problem and to give the best antenna structure the benefit of doubt, the following most favorable conditions for energy storage and power dissipation inside the sphere will be assumed:

1. There will be no dissipation in the antenna structure in the form of conduction loss.
2. There will be no electrical energy stored except in the form of a traveling wave.
3. The magnetic energy stored will be such that the total average electric energy stored beyond the input terminals of the antenna is equal to the average magnetic energy stored beyond the terminals at the operating frequency.

By Poynting's theorem it can be shown that the input impedance of the antenna is a pure resistance at the operating frequency under these conditions.

With this particular antenna structure, and its corresponding equivalent circuit, we can proceed to define a quantity Q at the input terminals:

$$Q = \frac{2\omega \text{ times the mean electric energy stored beyond the input terminals}}{\text{power dissipated in radiation}}$$

If this Q is high, it can be interpreted as the reciprocal of the fractional frequency bandwidth of the antenna. If it is low, the input impedance of the antenna varies slowly with frequency and the antenna has potentially a broad bandwidth. The ratio Q can therefore be used in the latter case as a crude indication for a broadband.

Upon summing up the mean electric energy stored in all the simplified equivalent circuits representing the spherical waves outside the sphere, and the total power radiated, the Q of the idealized antenna is

$$Q = \frac{\sum' A_n^2 \frac{n(n+1)}{2n+1} Q_n(\rho)}{\sum' A_n^2 \frac{n(n+1)}{2n+1}} \qquad (13)$$

where Q_n is given by Eq. (12).

We have defined and calculated two fundamental quantities G and Q for this somewhat idealized antenna. We have imposed a number of conditions on the coefficients A_n as well as on the energy and power inside the sphere. Otherwise, the set of coefficients A_n is yet unspecified. Additional conditions must be imposed on G and Q to determine the ultimate limits of antenna performance under various criteria.

2.6 Criterion I: Maximum Gain. Whenever the antenna structure must be confined within a small volume, and high gain is required, the logical criterion would be to demand maximum gain with an antenna structure of given complexity. The series of Legendre polynomials representing the field distribution behaves angularly in the same fashion as a Fourier series. The complexity of the source distribution required to generate the n-th term increases with the order n. To specify the number of terms in Eq. (1) to be used is therefore equivalent to specifying the complexity of the antenna structure. We shall therefore exclude all the terms for $n > N$, where N is an odd integer, and proceed to calculate the maximum gain as a function of N.

Differentiating the gain in the equatorial plane, [Eq. (4)], with respect to the coefficient A_n and setting the derivative to zero, we have

$$A_n = (-1)^{\frac{n+1}{2}} \frac{2n+1}{n(n+1)} P_n^1(0) \frac{\sum' A_n^2 \frac{n(n+1)}{2n+1}}{\sum' A_n (-1)^{\frac{n+1}{2}} P_n^1(0)} .$$

There are as many equations of this form as the number of terms in the series. We can therefore solve for A_n in terms of the first coefficient A_1:

$$A_n = (-1)^{\frac{n-1}{2}} \frac{2(2n+1)}{3n(n+1)} P_n^1(0) A_1 . \qquad (14)$$

The corresponding gain and Q of the antenna are

$$G(\tfrac{\pi}{2}) = \sum_1^{N}{}' a_n \qquad (15)$$

$$Q = \sum_1^N{}' a_n Q_n \Big/ \sum_1^N{}' a_n \qquad (16)$$

where $a_n = \frac{2n+1}{n(n+1)} \left[P_n^1(0) \right]^2$. (17)

Except for the first few terms,

$$a_n \cong 4/\pi . \qquad (18)$$

The formula (15) for the maximum gain was previously obtained by W. W. Hansen[5].

TABLE I

The Maximum Gain Versus N

N	1	3	5 - - - - - - - - - - - - - - - -	N
Gain	1.5	3.81	4.10	$2N/\pi$

The value of the maximum gain for different values of N is given in Table I. For N = 1, the gain is that of an electric dipole. For large values of N, the gain is proportional to N. Under the present criterion, the gain is independent of the size of the antenna. It indicates that an arbitrarily high gain can be obtained with an arbitrarily small antenna, provided the source distribution can be physically arranged.

Figure 6 shows the Q of an antenna designed to obtain the maximum gain with a given number of terms, as a function of $2\pi a/\lambda$. While the terms in the denominator

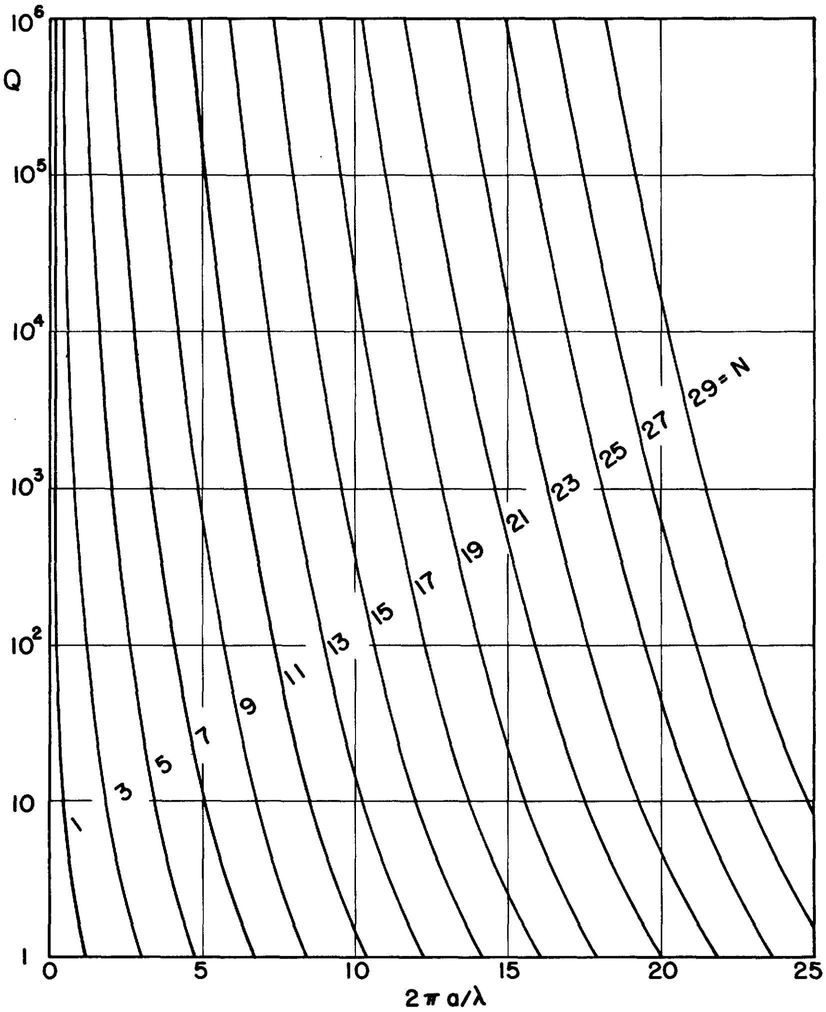

Figure 6. Q of omnidirectional antenna. Criterion: Max. gain with fixed number of terms.

5. W. W. Hansen, "Notes on Microwaves", M. I. T. Rad. Lab. Report T-2.

of Eq. (16) have approximately equal amplitudes, the numerator is an ascending series of $(N+1)/2$ terms. For any given value of $2\pi a/\lambda$, Q_n increases with n at a rapid rate as shown in Fig. 5. The numerator is essentially determined by the last few terms of the ascending series. For $2\pi a/\lambda$ greater than N, Q is of the order of unity or less, indicating the potentiality of a broad-band system. For $2\pi a/\lambda$ less than N, the value of Q rises astronomically as $2\pi a/\lambda$ decreases. The transition occurs for

$$2\pi a/\lambda \cong N \tag{19}$$

corresponding to a gain

$$G \cong \frac{2}{\pi} \cdot \frac{2\pi a}{\lambda} = \frac{4a}{\lambda} \;. \tag{20}$$

The gain of an omnidirectional antenna as given by Eq. (20) will be called the normal gain. It is equal to the gain obtained from a current distribution of uniform amplitude and phase along a line of length 2a. In Fig. 7, curve I shows the Q of an antenna of normal gain. To increase the gain by a factor of two, we have to use twice as many

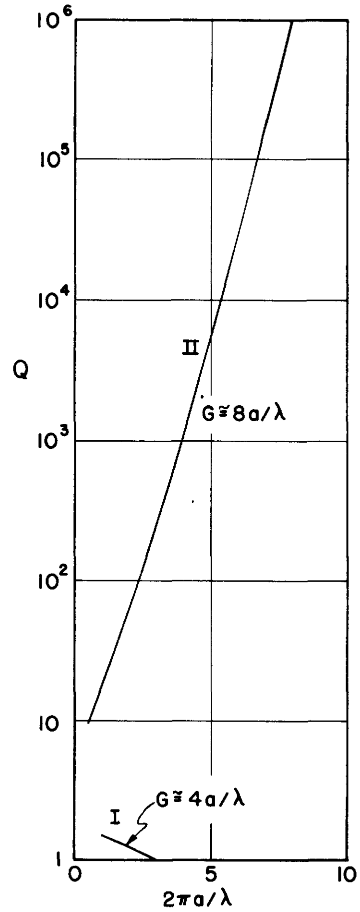

Figure 7. Q for omnidirectional antenna.
Criterion: Max. G with fixed number of terms.
I normal gain.
II twice the normal gain.

terms, and pay dearly in Q as shown by Curve II. The slope of Curve II indicates the increasing difficulty of obtaining additional gain as the normal gain increases.

Under the present criterion, no special conditions have been imposed on Q. It can be shown that the Q obtained here is by no means the minimum for a given gain and antenna size. Since the gain is maximized [Eq. (15)] with respect to A_n, a small variation of A_N will not affect the gain. Instead, the Q will vary rapidly as indicated by Eq. (13) when $N > 2\pi a/\lambda$.

2.6 Criterion II: Minimum Q. In this section, we shall proceed to find a combination of A_n's to give the minimum Q with no separate conditions imposed on the gain of the antenna. Differentiating Q with respect to A_n, we have the following equation:

$$Q_n \Sigma' A_n^2 \frac{n(n+1)}{2n+1} = \Sigma' A_n^2 \frac{n(n+1)}{2n+1} Q_n \ .$$

For any given value of $2\pi a/\lambda$, all the Q_n's have different values. Hence, the above equation can be satisfied when there is only one term under the summation sign. The corresponding Q of the antenna is equal to the Q_n of the term used. Since Q_1 has the lowest amplitude, we conclude that the antenna which generates a field outside the sphere corresponding to that of an infinitesimally small dipole has potentially the broadest bandwidth of all antennas. The gain of this antenna is 1.5.

2.7 Criterion III: Maximum G/Q. As a compromise between the two criteria just mentioned, we shall now maximize the ratio of the gain to Q. The process can be interpreted as the condition for the minimum Q to achieve a certain gain or as the condition for the maximum gain at a given Q. The problem is that of finding the proper combination of A_n's for maximum G/Q. From Eqs. (4) and (13), we have

$$\frac{G}{Q} = \frac{\left[\Sigma' A_n (-1)^{\frac{n+1}{2}} P_n^1(0)\right]^2}{\Sigma' A_n^2 \frac{n(n+1)}{2n+1} Q_n} \ . \tag{21}$$

With the same method used before, we obtain

$$A_n = (-1)^{\frac{n-1}{2}} \frac{2(2n+1)}{3n(n+1)} P_n^1(0) \frac{Q_1}{Q_n} A_1 \ . \tag{22}$$

The corresponding values of G, Q, and the ratio G/Q are

$$G = \frac{\left[\Sigma' a_n/Q_n\right]^2}{\Sigma' a_n/Q_n^2} \tag{23}$$

$$Q = \frac{\Sigma' a_n/Q_n}{\Sigma' a_n/Q_n^2} \tag{24}$$

$$G/Q = \Sigma' a_n/Q_n \tag{25}$$

where a_n is given in Eq. (17). The gain and G/Q are plotted against $2\pi a/\lambda$ with N as a parameter in Figs. 8 and 9, respectively. In using the above formulas, Q_n is arbitrarily considered to be unity whenever its actual value is equal to or less than unity. The Q for all the points on the curve is about unity. Since the two series involved in Eqs. (23) and (24) converge rapidly as N is increased indefinitely, the gain approaches asymptotically the approximate value of $4a/\lambda$ which is the normal gain derived under criterion I. There is a definite limit to the gain if the Q of the antenna is required to be low. It is this physical limitation, among others, which limits the gain of all the practical antennas to the approximate value $4a/\lambda$.

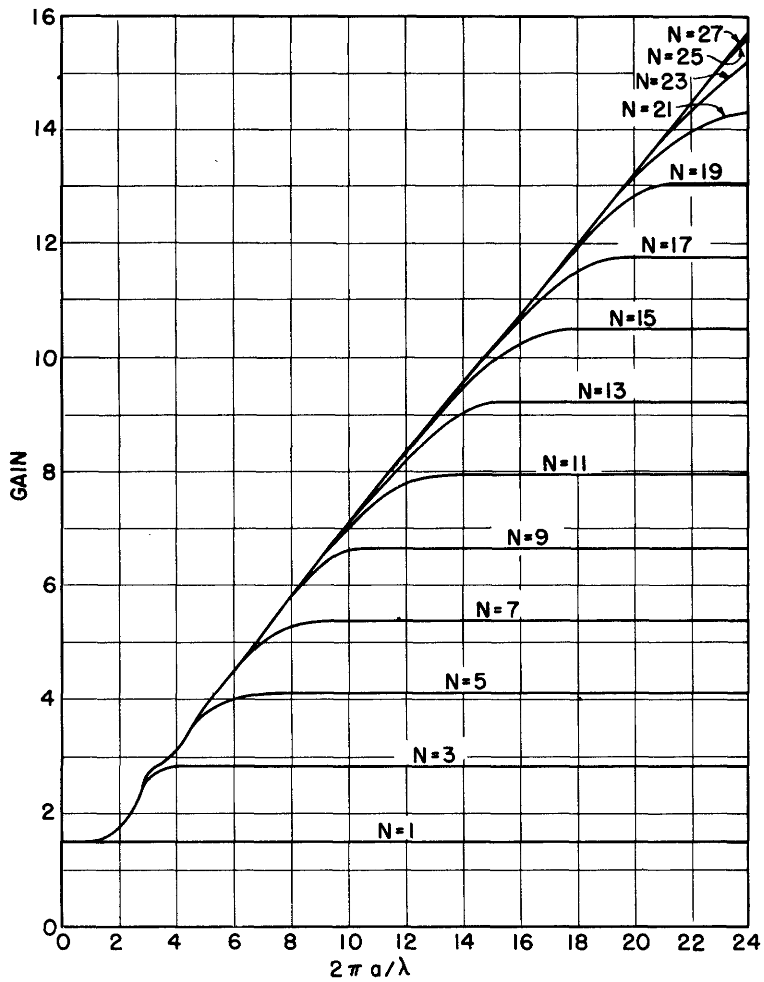

Figure 8. Gain of omnidirectional antenna. Criterion: Max. G/Q. When $Q_n < 1$, it is considered to be unity.

<u>2.8 Horizontally Polarized Omnidirectional Antenna.</u> By interchanging E and H in Eq. (1), and replacing $\sqrt{\epsilon/\mu}$ by $-\sqrt{\mu/\epsilon}$, we have the field outside the geometric sphere, for a horizontally polarized omnidirectional antenna, expressed as a summation of circularly symmetrical TE_n waves:

$$E_\phi = \sqrt{\frac{\mu}{\epsilon}}\ \Sigma' B_n\ P_n^1(\cos\theta)\ h_n(kR)$$

$$H_r = j\ \Sigma' B_n\ n(n+1)\ P_n(\cos\theta)\ \frac{h_n(kR)}{kR} \qquad (26)$$

$$H_\theta = -j\ \Sigma' B_n\ P_n^1(\cos\theta)\ \frac{1}{kR}\ \frac{d}{dR}\left[\ R\ h_n(kR)\right]$$

where the B_n's are arbitrary constants. As before, each TE_n wave at the surface of the sphere can be replaced by a two-terminal equivalent circuit defined on the same basis as that of the corresponding TM_n wave. The voltage, current, and admittance at the input of the circuit are the following:

$$V_n = \sqrt[4]{\frac{\mu}{\epsilon}}\ \frac{B_n}{k}\ \sqrt{\frac{4\pi n(n+1)}{2n+1}}\ \rho\, h_n \qquad (27)$$

$$I_n = \sqrt[4]{\frac{\mu}{\epsilon}}\ \frac{B_n}{k}\ \sqrt{\frac{4\pi n(n+1)}{2n+1}}\ \ j(\rho\, h_n)' \qquad (28)$$

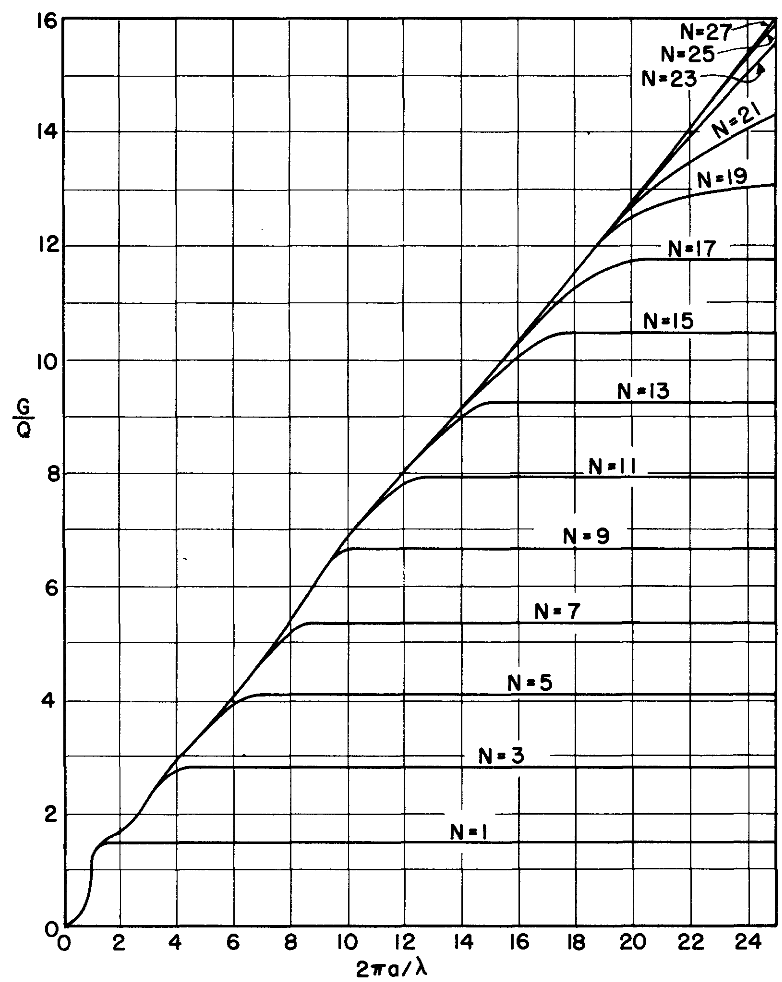

Figure 9. G/Q of omnidirectional antenna. Criterion: Max. G/Q. When $Q_n < 1$, it is considered to be unity.

$$Y_n = j(\rho h_n)' / \rho h_n \quad . \tag{29}$$

The admittance is equal to the normalized wave admittance of the TE_n wave on the surface of the sphere, and is also equal to the impedance Z_n of the equivalent circuit for the TM_n wave. This circuit is a cascade of shunt inductances and series capacitances terminated with a unit conductance as shown in Fig. 10. At low

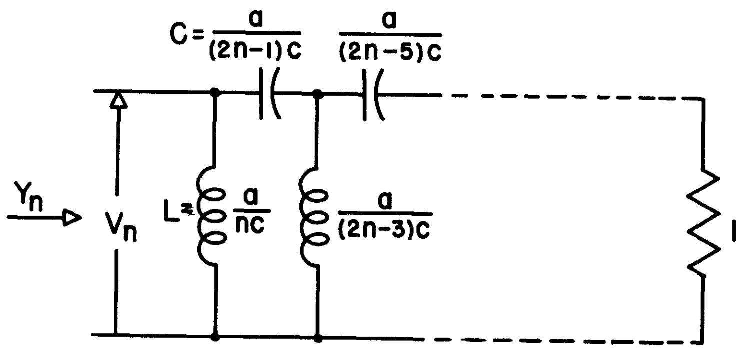

Figure 10. Equivalent circuit of TE_n spherical wave.

frequencies, the admittance is practically that of the first inductance. The admittance remains inductive at all frequencies and approaches a pure conductance of unit amplitude as the frequency increases.

The analysis of the horizontally polarized omnidirectional antenna follows exactly that of the vertically polarized one. The formulas for G, P_n, W_n, Q_n, and Q remain unchanged if we replace all the A_n's by B_n's. The quantity W_n is now to be interpreted as the mean magnetic energy stored in the simplified equivalent circuit of the TE_n wave (a parallel RLC circuit). Results obtained previously apply to the present problem without further modification.

2.9 Circularly Polarized Omnidirectional Antenna. The field of an elliptically polarized omnidirectional antenna can be expressed as a sum of TM_n waves [Eq. (1)] and TE_n waves [Eq. (26)]. To obtain circular polarization everywhere, we must have

$$B_n = \pm j A_n \quad . \tag{30}$$

Under this condition, the gain of the circularly polarized antenna is again given by Eq. (3).

The equivalent circuit of the circularly polarized omnidirectional antenna consists of 2N+1 pairs of terminals where N is the highest order of the spherical waves used. If we are only interested in the gain along the equator, the even terms of the series can be excluded. The number of pairs are reduced to N+2 including the input pair. It is interesting to observe that the instantaneous total energy density at any point outside the sphere is independent of time when Eq. (30) is satisfied. The difference between the mean electric energy density and the mean magnetic energy density is zero at any point outside the sphere enclosing the antenna. Furthermore, the instantaneous Poynting vector is independent of time. This implies that the power flow from the surface of the sphere enclosing a truly circularly polarized omnidirectional antenna is a d-c flow, and the instantaneous power is equal to the radiated power. These relationships are due to the dual nature of TE waves and TM waves as well as the 90° difference in time phase between the two sets of waves.

To obtain the Q of the antenna, it is convenient to combine the energies and dissipation in Z_n of the TM_n wave with that in Y_n of the TE_n wave and define a new Q_n as $2\omega W_n/P_n$ where W_n is the mean electric or magnetic energy stored in Z_n and Y_n, and P_n is the total power dissipated in both. Then

$$Q_n = \frac{1}{2} \left|\rho h_n\right|^2 \rho \frac{dX_n}{d\rho} \tag{31}$$

where X_n is the imaginary part of Z_n. For $\rho = 2\pi a/\lambda > n$, this Q_n is approximately equal to one half of the previous Q_n defined for Z_n or Y_n alone, [Eq. (12)]. If no conduction loss and no stored energy inside the sphere are assumed, the expression for Q of a circularly polarized omnidirectional antenna turns out to be identical with that given by Eq. (13), except that Q_n is given by Eq. (31) instead of Eq. (12).

With expressions for G and Q identical with what were obtained previously, we expect similar numerical results for the present case under the various criteria, and the same physical limitation applies.

3. Further Considerations

The conclusion of this paper is self evident. To achieve a gain higher than normal, we must sacrifice the bandwidth under the most favorable conditions. The rest of the paper will be devoted to other considerations not covered in the analysis.

3.1 Practical Limitations. The above analysis does not take into consideration many practical aspects of antenna design. In the following, a qualitative discussion will be given of some of the practical limitations.

It is assumed in the analysis that the antenna under consideration is located in free space. The results, with a minor modification are applicable to the problem of a vertically polarized antenna above a perfectly conducting ground plane. In practice, this condition can seldom be fulfilled. The performance of an antenna designed on the free-space basis will be modified by the presence of physical objects in the neighborhood. Currents will be induced on the objects. They will give rise not only to an additional scattered radiation field but also to a modification of the original current distribution on the antenna structure. Both the gain and Q of the antenna will be changed from their unperturbed values. The currents set up on the objects vary as the unperturbed field intensity at the locations of the objects. For the same power radiated, the r.m.s. amplitude of the unperturbed field intensity in the neighborhood of the antenna is approximately proportional to the square root of Q. In view of the rapid increase of Q as the gain of an antenna is increased above the normal value shown in Fig. 6, the disturbance of the field distribution in space by physical objects in the neighborhood of the antenna becomes increasingly serious.

It is tacitly assumed in the analysis that physically it is possible to design an antenna to achieve an arbitrary current distribution which satisfies the condition for minimum energy storage as discussed in Sec. 2.5. To obtain a gain above normal, additional higher-order spherical waves must be generated outside the sphere with a proper control of amplitudes and relative phases. The corresponding current distribution will have rapid amplitude and phase variation inside the sphere. The practical difficulty of achieving this current distribution will increase with the gain.

We have avoided the question of conduction losses on the antenna structure. In practice, the antenna structure will have conductivities differing from zero or infinity. Neglecting the losses on the transmission line, it can be shown that the minimum conduction loss of the antenna under consideration varies approximately as the mean square of the electric or magnetic field on the surface of the sphere. For a high-Q antenna, the ratio of the minimum conduction loss to the power radiated is therefore approximately proportional to the Q of the antenna computed in the absence of losses. Although this conduction loss is helpful in reducing the Q at the input terminals, it reduces the efficiency and the power gain of the antenna.

The condition of minimum energy storage within the sphere is not always realizable. On account of the unavoidable frequency sensitivities of the elements of the antenna structure or the matching networks, the Q of a practical antenna computed on the no conduction-loss basis will be usually higher than the one derived in this paper.

3.2 Bandwidth and Ideal Matching Network. We have computed the Q of an antenna from the energy stored in the equivalent circuit and the power radiated, and interpreted it freely as the reciprocal of the fractional bandwidth. To be more accurate, one must define the bandwidth in terms of allowable impedance variation or the tolerable reflection coefficient over the band. For a given antenna, the bandwidth can be increased by choosing a proper matching network. The theoretical aspect of this problem has been dealt with by R. M. Fano[6]. Figure 11 given here through his courtesy illustrates the relations among the fractional bandwidth, absolute amplitude of the reflection coefficient, and the parameter $2\pi a/\lambda$ of an antenna which has only the TM_1 wave outside the sphere. As shown in Sec. 2.6, this antenna has the lowest Q of all antennas and its equivalent circuit is shown in Fig. 3. The curve of Fig. 11 is computed on the assumption that the input impedance of the antenna is equal to Z_1, and an ideal matching network is used to obtain a constant amplitude of the reflection coefficent over the band. The phase of the reflection coefficient,

---

6. R. M. Fano, "A Note on the Solution of Certain Problems in Network Synthesis", April 16, 1948, RLE Technical Report No. 62.

however, varies rapidly near the ends of the band.

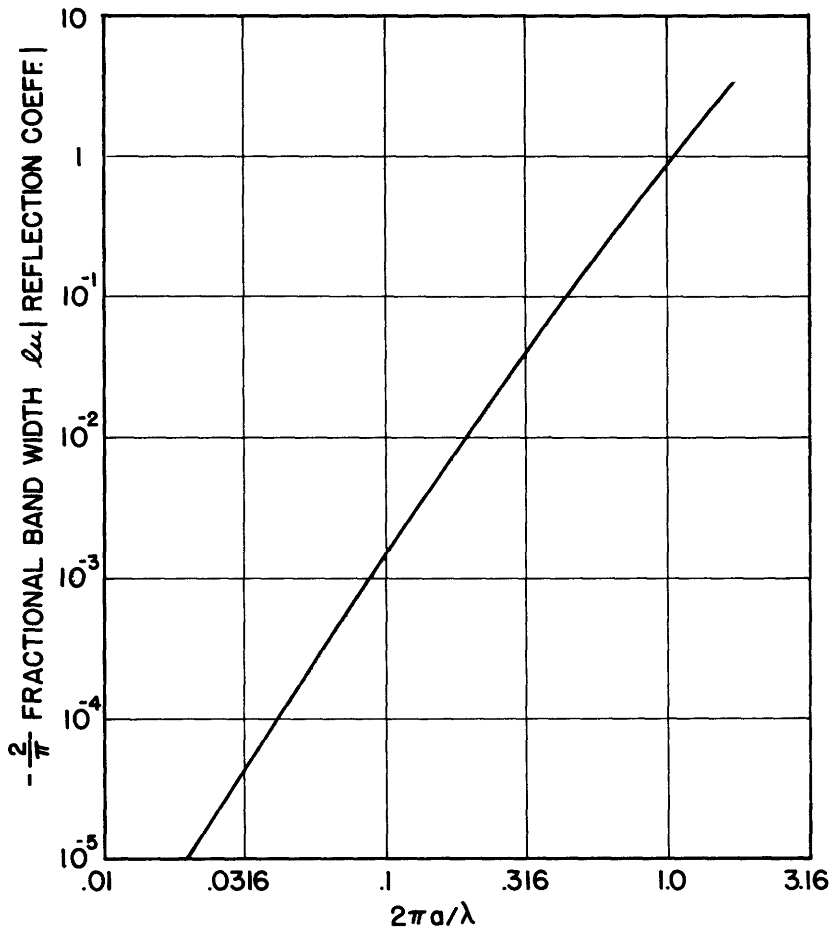

Figure 11. Bandwidth of an ideal dipole with ideal matching network.