Journal of Research of the National Bureau of Standards—D. Radio Propagation
Vol. 64D, No. 1, January–February 1960

Effect of Antenna Size on Gain, Bandwidth, and Efficiency[1]

Roger F. Harrington[2]

(June 29, 1959)

A theoretical analysis is made of the effect of antenna size on parameters such as gain, bandwidth, and efficiency. Both near-zone and far-zone directive gains are considered. It is found that the maximum gain obtainable from a broad-band antenna is approximately equal to that of the uniformly illuminated aperture. If higher gain is desired, the antenna must necessarily be a narrow-band device. In fact, the input impedance becomes frequency sensitive so rapidly that, for large antennas, no significant increase in gain over that of the uniformly illuminated aperture is possible. Also, if the antenna is lossy, the efficiency falls rapidly as the gain is increased over that of the uniformly illuminated aperture.

1. Introduction

As a practical matter, the maximum directive gain (directivity) of an antenna depends upon its physical size compared to wavelength. The uniformly illuminated aperture type of antenna has been found to give a higher gain in practice than other antennas, at least for large apertures. However, the uniformly illuminated aperture does not represent a theoretical limit to the gain. Higher directive gains appear to be possible, but analyses of projected "supergain" antennas reveal extreme frequency sensitivity at best, excessive losses at worst. This paper gives a theoretical treatment to the general problem, from which quantitative bounds to antenna performance may be obtained. The analysis considers both the near-zone and the far-zone gain of antennas.

Let the spherical coordinate system be defined as in figure 1. The directive gain as a function of distance from an antenna is defined as the ratio of the maximum density of outward-directed power flux to the average density. In equation form this is

$$G(\beta r)=\frac{4\pi r^2 \mathrm{Re}(S_r)_{\max}}{\mathrm{Re}(P)}, \tag{1}$$

where S_r is the radial component of the complex Poynting vector at a distance r, and P is the outward-directed complex power over a sphere of radius r.

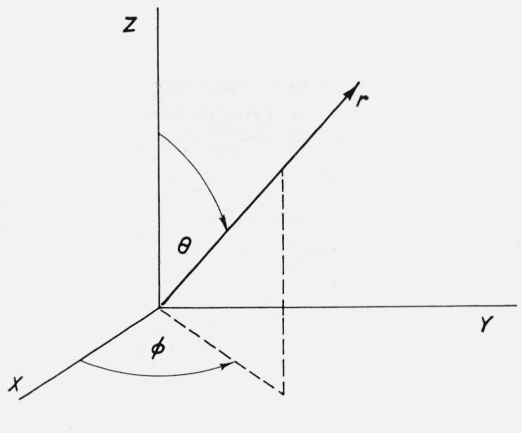

FIGURE 1. *The spherical coordinate system.*

[1] This work has been supported by General Electric Company, subcontract No. 500–7119
[2] Department of Electrical Engineering, Syracuse University, Syracuse, N.Y.

The field external to a sphere containing all sources can be expanded in terms of spherical wave functions. The general form is [1] [3]

$$\boldsymbol{E}=-\nabla\times(\mathbf{r}\psi)+\frac{1}{j\omega\epsilon}\nabla\times\nabla\times(\mathbf{r}\hat{\psi})$$
$$\boldsymbol{H}=\nabla\times(\mathbf{r}\hat{\psi})+\frac{1}{j\omega\mu}\nabla\times\nabla\times(\mathbf{r}\psi) \tag{2}$$

where $\boldsymbol{r}$ is the radius vector from the origin and

$$\psi=\sum_{m,n}A_{mn}h_n^{(2)}(\beta r)P_n^m(\cos\theta)\cos(m\phi+\alpha_{mn})$$
$$\hat{\psi}=\sum_{m,n}B_{mn}h_n^{(2)}(\beta r)P_n^m(\cos\theta)\cos(m\phi+\hat{\alpha}_{mn}) \tag{3}$$

where A_{mn}, B_{mn}, α_{mn}, and $\hat{\alpha}_{mn}$ are coefficients which do not depend on r, θ, or ϕ, and $\beta=2\pi$/wavelength. From the above formulas for the field can be calculated the power over a sphere of radius r

$$\mathrm{Re}(P)=\frac{4\pi}{\beta^2}\sum_{m,n}\frac{n(n+1)(n+m)!}{\epsilon_m(2n+1)(n-m)!}\left(\frac{1}{\eta}|A_{mn}|^2+\eta|B_{mn}|^2\right) \tag{4}$$

where ϵ_m is Neumann's number and $\eta\cong120\pi$. The radial component of the power flux density in the $\theta=0$ direction is

$$S_r=E_xH_y^*-E_yH_x^* \tag{5}$$

where

$$E_x=\sum_n\frac{n(n+1)}{2\beta r}\left[A_{1n}\sin\alpha_{1n}F_n(\beta r)-j\eta B_{1n}\cos\hat{\alpha}_{1n}F_n'(\beta r)\right]$$
$$E_y=\sum_n\frac{n(n+1)}{2\beta r}\left[A_{1n}\cos\alpha_{1n}F_n(\beta r)+j\eta B_{1n}\sin\hat{\alpha}_{1n}F_n'(\beta r)\right]$$
$$H_x=\sum_n\frac{n(n+1)}{2\beta r}\left[-B_{1n}\sin\hat{\alpha}_{1n}F_n(\beta r)+\frac{j}{\eta}A_{1n}\cos\alpha_{1n}F_n'(\beta r)\right] \tag{6}$$
$$H_y=\sum_n\frac{n(n+1)}{2\beta r}\left[-B_{1n}\cos\hat{\alpha}_{1n}F_n(\beta r)-\frac{j}{\eta}A_{1n}\sin\alpha_{1n}F_n'(\beta r)\right]$$

and

$$F_n(\beta r)=\beta r\,h_n^{(2)}(\beta r). \tag{7}$$

The coordinate axes are to be oriented so the maximum radiation is in the $\theta=0$ direction. Then the S_r of (5) is the $(S_r)_{\max}$ of (1).

Note that α_{mn} and $\hat{\alpha}_{mn}$ do not enter into the formula for power, (4). Also, from the symmetry of (5) and (6) it is evident that the density of power flux in the $\theta=0$ direction is independent of α_{1n} and $\hat{\alpha}_{1n}$. Hence, they may be chosen in any convenient manner. In particular, let $\alpha_{1n}=\pi/2$, $\hat{\alpha}_{1n}=\pi$, which give a field linearly polarized in the x direction. With this choice, the gain (1) becomes

$$G=\frac{\mathrm{Re}[(\sum_n a_nF_n+jb_nF_n')(\sum_n b_nF_n+ja_nF_n')^*]}{4\sum_{m,n}\frac{n(n+1)(n+m)!}{\epsilon_m(2n+1)(n-m)!}(|A_{mn}|^2+|\eta B_{mn}|^2)} \tag{8}$$

[3] Figures in brackets indicate the literature references at the end of this paper.

where

$$a_n = n(n+1)A_{1n} \qquad b_n = \eta n(n+1)B_{1n}. \tag{9}$$

Equation (8) is a general formula for directive gain.

2. Maximum Gain

If all orders of spherical wave functions are permitted there is no limit to the gain of an antenna. However, a definite limit to the gain exists if wave functions are restricted to orders $n \leq N$. Only the A_{1n} and B_{1n} contribute to the numerator of (8), so the gain can be increased by setting

$$A_{mn} = B_{mn} = 0 \qquad m \neq 1. \tag{10}$$

Furthermore, the gain formula is symmetrical in a_n and b_n so the maximum gain will exist under the condition

$$a_n = b_n. \tag{11}$$

Equation (8) has now been reduced to

$$G = \frac{\mathrm{Re}\left[\left(\sum_{n=1}^{N} a_n u_n\right)\left(\sum_{n=1}^{N} a_n u_n\right)^*\right]}{4\sum_{n=1}^{N} |a_n|^2 \frac{1}{2n+1}} \tag{12}$$

where

$$u_n(\beta r) = F_n(\beta r) + j F_n'(\beta r). \tag{13}$$

The numerator of (12) can now be increased without changing the denominator by setting

$$\text{phase } (a_n) = -\text{phase } (u_n) \tag{14}$$

in which case (12) becomes

$$G = \frac{\left(\sum_{n=}^{N} |a_n|\,|u_n|\right)^2}{4\sum_{n=1}^{N} |a_n|^2 \frac{1}{2n+1}}. \tag{15}$$

Finally, the $|a_n|$ are adjusted for maximum gain by requiring $\partial G/\partial |a_i| = 0$ for all a_i. The result is

$$G(\beta r)_{\max} = \tfrac{1}{4}\sum_{n=1}^{N}(2n+1)\,|u_n(\beta r)|^2 \tag{16}$$

which is the maximum directive gain obtainable using wave functions of order $n \leq N$. The maximization procedure also results in the relationship

$$\left|\frac{a_n}{a_i}\right| = \frac{(2n+1)|u_n|}{(2i+1)|u_i|}. \tag{17}$$

From (13) it follows that

$$|u_n(\beta r)|^2 = |F_n'(\beta r)|^2 + |F_n(\beta r)|^2 + 2. \tag{18}$$

As $\beta r \rightarrow \infty$, $|u_n|^2 \rightarrow 4$, so in the far zone (16) reduces to

$$G(\infty)_{\max} = \sum_{n=1}^{N}(2n+1) = N^2 + 2N \tag{19}$$

which has been previously published [1].

The $|u_n|^2$ functions, which enter into (16) and also into later formulas, are shown in figure 2. The maximum directive gain for various βr and various N has been calculated. Figure 3 shows the ratio of the near-zone gain to the far-zone gain for several N. Note that the maximum near-zone gain is essentially the same as the maximum far-zone gain unless $\beta r < N$. Keep in mind that the excitation of the antenna is changed as βr is varied so that it is always adjusted for maximum gain at the given radius. One can think of the antenna as being focused at the distance r.

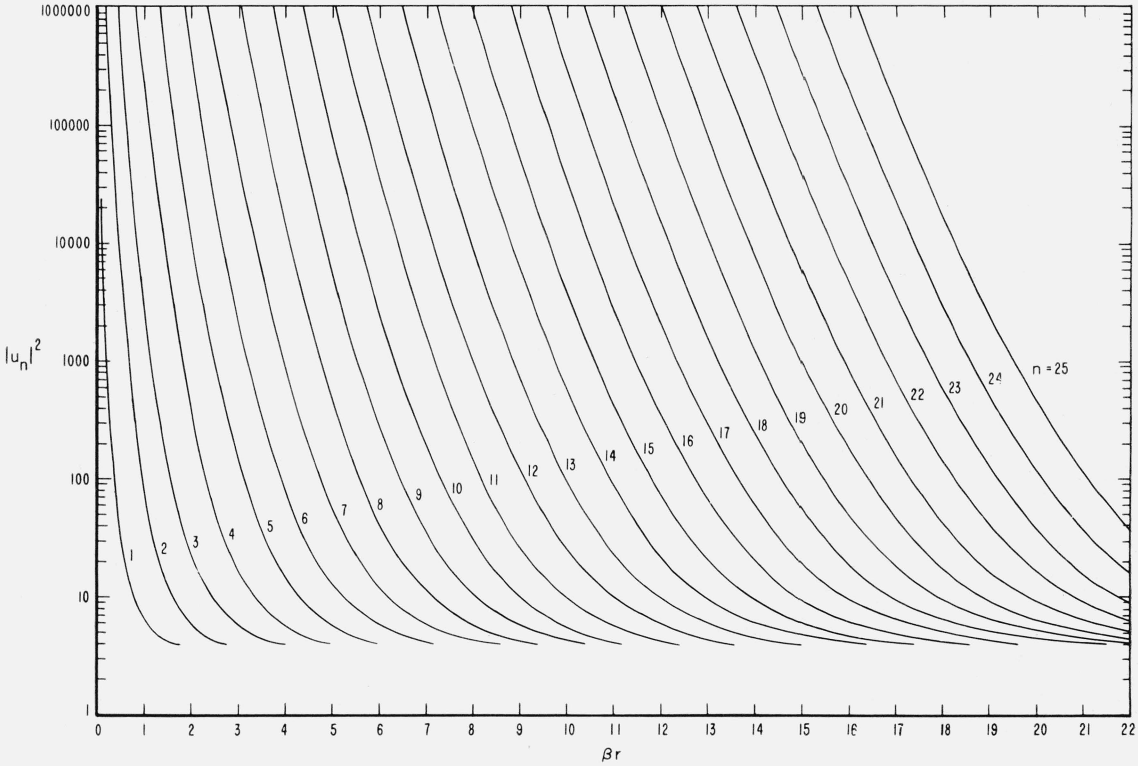

FIGURE 2. *The functions* $|u_n(\beta r)|^2$.

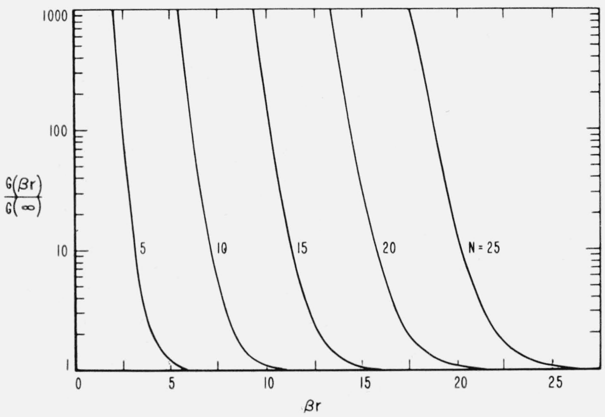

FIGURE 3. *Ratio of the maximum gain at a distance* r *to the maximum gain at infinity, using wave functions of order* $n \leq N$.

3. Quality Factor

So far the antenna structure has not been mentioned. An ideal loss-free antenna of radius R is defined as one having no energy storage $r<R$, which is the same as Chu's definition [2]. The quality factor

$$Q=\begin{cases} \dfrac{2\omega\, W^{\text{elect}}}{\text{Re}(P)} & W^{\text{elect}}>W^{\text{mag}} \\[2ex] \dfrac{2\omega\, W^{\text{mag}}}{\text{Re}(P)} & W^{\text{mag}}>W^{\text{elect}} \end{cases} \tag{20}$$

for this ideal antenna must be less than or equal to that for any other loss-free antenna fitting into the sphere $r=R$, since any field $r<R$ can only add to the energy storage. In (20) the energies W are those obtained by subtracting the radiation field from the total field. If the Q of an antenna is high, it can be interpreted as the reciprocal of the fractional bandwidth of the input impedance. If the Q is low the antenna has broadband potentialities.

Because of the orthogonality of the spherical wave functions, the total electric energy, magnetic energy, and power radiated is the sum of the corresponding quantities associated with each mode. The Q can therefore be found by treating the field of each spherical wave as if it existed on a "spherical waveguide" isolated from all other waves [2,3]. The energy and power formulas of transmission line theory apply to each spherical waveguide if a voltage, current, and characteristic impedance are defined for each TE_{mn} wave as

$$\begin{aligned} V_{mn}^{TE}&=\frac{A_{mn}}{\beta}\sqrt{\frac{4\pi n(n+1)(n+m)!}{\eta\epsilon_m(2n+1)(n-m)!}}\,F_n(\beta r) \\ I_{mn}^{TE}&=\frac{A_{mn}}{\beta}\sqrt{\frac{4\pi n(n+1)(n+m)!}{\eta\epsilon_m(2n+1)(n-m)!}}\,jF_n'(\beta r) \\ Z_{mn}^{TE}&=F_n(\beta r)/jF_n'(\beta r) \end{aligned} \tag{21}$$

and for each TM_{mn} wave as

$$\begin{aligned} V_{mn}^{TM}&=\frac{B_{mn}}{\beta}\sqrt{\frac{\eta 4\pi n(n+1)(n+m)!}{\epsilon_m(2n+1)(n-m)!}}\,jF_n'(\beta r) \\ I_{mn}^{TM}&=\frac{B_{mn}}{\beta}\sqrt{\frac{\eta 4\pi n(n+1)(n+m)!}{\epsilon_m(2n+1)(n-m)!}}\,F_n(\beta r) \\ Z_{mn}^{TM}&=jF_n'(\beta r)/F_n(\beta r). \end{aligned} \tag{22}$$

An antenna adjusted for maximum gain has equal excitation of TE and TM waves. It is therefore convenient to define modal quality factors

$$Q_n=\frac{2\omega\, W_{mn}^{\text{elect}}}{\text{Re}(P_{mn})}=\frac{2\omega\, W_{mn}^{\text{mag}}}{\text{Re}(P_{mn})}, \tag{23}$$

where the W_{mn} and P_{mn} are the sum of the energies and powers of both the TE_{mn} and TM_{mn} waves. Note that the Q_n's are independent of m since the characteristic impedances, (21) and (22), are independent of m. These are the same Q_n's defined by Chu for circularly polarized omnidirectional antennas [2]. Abstracting from Chu's work, one has

$$Q_n(\beta R)=\tfrac{1}{2}|F_n(\beta R)|^2\beta R X_n'(\beta R) \tag{24}$$

where R is the radius of the spherical antenna and

$$X_n(\beta R) = \text{Im}\,[Z_{mn}^{TM}(\beta R)]. \tag{25}$$

These Q_n are approximately one-half those calculated by Chu for linearly polarized omnidirectional antennas. A plot of some Q_n is given in figure 4.

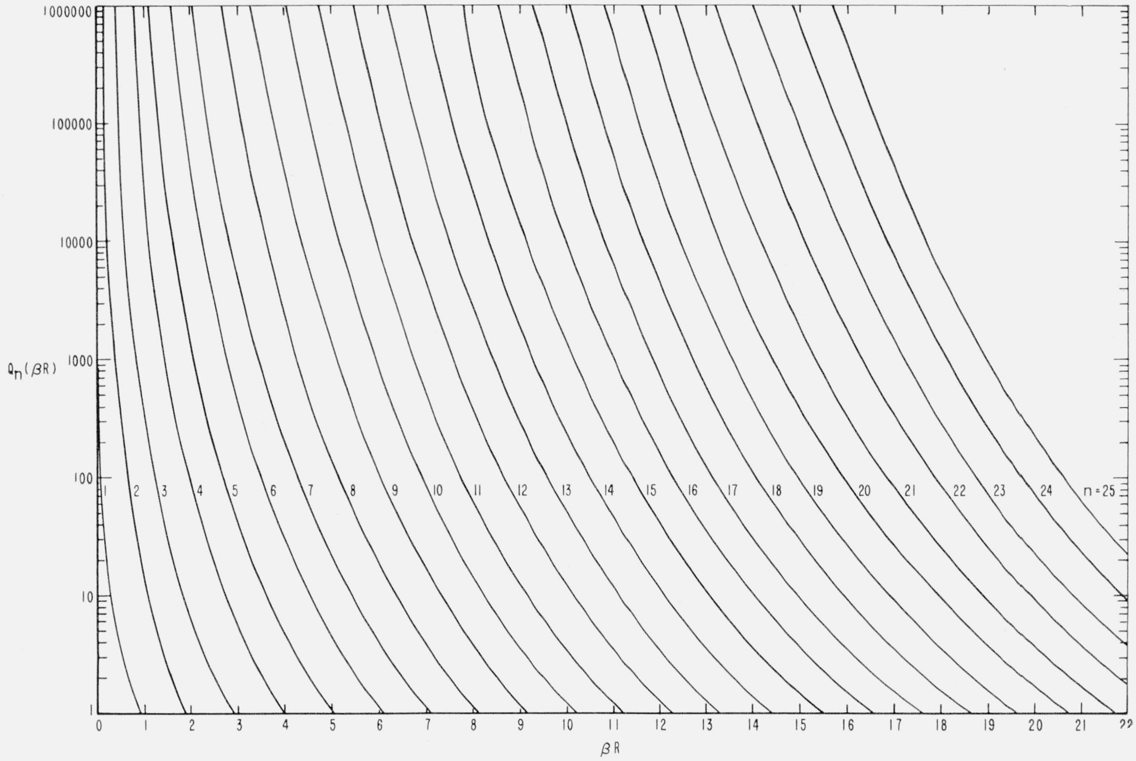

FIGURE 4. *Modal quality factors for wave functions of order* n.

The Q of an antenna having equal excitation of TE and TM waves is given by

$$Q = \frac{\sum_{m,n} \text{Re}(P_{mn}) Q_n}{\sum_{m,n} \text{Re}(P_{mn})} \tag{26}$$

since it is merely necessary to add the energies and powers of the individual waves. For maximum gain all $P_{mn}=0$ except P_{1n}, $n \leq$ N. It follows from (9), (11), (21), and (22) that

$$\begin{aligned} \text{Re}\,(P_{1n}) &= \text{Re}\,(V_{1n}^{TE} I_{1n}^{TE*} + V_{1n}^{TM} I_{1n}^{TM*}) \\ &= \frac{4\pi}{\eta \beta^2} \frac{1}{2n+1} |a_n|^2 \end{aligned} \tag{27}$$

so the formula for Q becomes

$$Q = \frac{\sum_{n=1}^{N} |a_n|^2 \left(\frac{1}{2n+1}\right) Q_n}{\sum_{n=1}^{N} |a_n|^2 \left(\frac{1}{2n+1}\right)}. \tag{28}$$

Finally, from (17) one has

$$Q=\frac{\sum_{n=1}^{N}(2n+1)|u_n(\beta r)|^2 Q_n(\beta R)}{\sum_{n=1}^{N}(2n+1)|u_n(\beta r)|^2} \tag{29}$$

as the quality factor for an ideal antenna of radius R, adjusted for maximum gain at a radius r, using wave functions of order $n \leq N$. For an antenna adjusted for maximum gain at infinity

$$Q \xrightarrow[\beta r\to\infty]{} \frac{\sum_{n=1}^{N}(2n+1)Q_n(\beta R)}{N^2+2N}. \tag{30}$$

Figure 5 shows the Q of antennas focused at infinity for several N. (The dashed lines represent

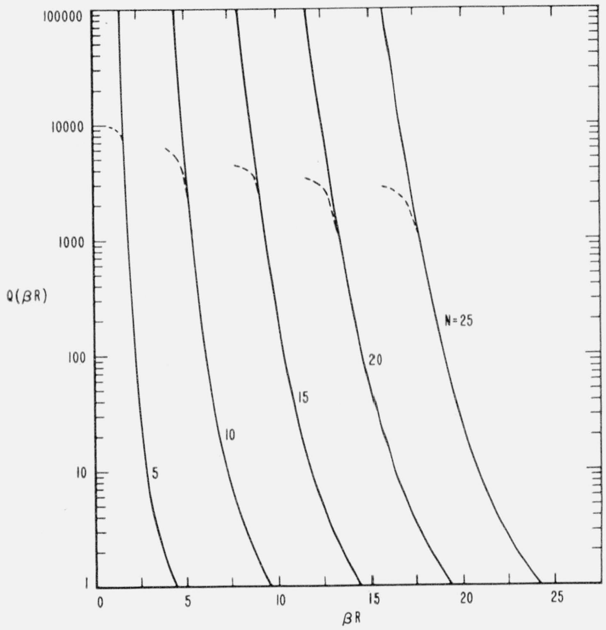

FIGURE 5. *Quality factors for ideal loss-free antennas adjusted for maximum gain at infinity, using wave functions of order* $n \leq N$ *(dashed lines show the effect of losses).*

the effect of antenna losses, considered below.) Calculations for antennas focused at other r show that the Q is substantially independent of the radius of focus. No appreciable change in Q occurs until r is almost equal to R, that is, the field point is almost at the antenna surface. When $r=R$, the Q of the antenna is close to the Q_n of the highest order wave present. When $\beta R>N$, all Q_n are of the order of unity or less, and the quality factor is

$$Q\leq 1 \qquad \beta R>N. \tag{31}$$

In this case the antenna is potentially broad band.

4. Effect on Antenna Losses

To obtain quantitative results for the effect of conduction losses on antenna performance an idealized model is again postulated. Intuitively one would expect the losses on a metal antenna to be smaller the more effectively the sphere is utilized. Therefore, for the ideal lossy antenna is postulated a spherical conductor of radius R excited by the magnetic sources

$$\mathbf{M}=\mathbf{E}\times\mathbf{n} \tag{32}$$

on its surface. So long as the sphere is a good conductor the source of (32) will generate the desired field [4]. If the conductivity of the sphere is poor, (32) can be modified to allow for a field internal to the conductor.

The above postulated ideal lossy antenna is particularly simple to analyze because the wave functions are orthogonal over its surface. The effect of the spherical conductor is that of a discontinuity in the characteristic impedance of each "spherical waveguide" at the radius R. The effect of the source $\mathbf{M}$ is that of a voltage source in series with each waveguide at the radius R. The waveguides are matched in each direction, so the equivalent circuit for each mode is a voltage source in series with the two characteristic impedances, $r<R$ and $r>R$. For $r<R$ the characteristic impedances for the various modes are

$$
\begin{aligned}
Z_{mn}^{TE} &= \frac{\eta_c}{\eta}[F_n(kr)/jF_n'(kr)]^* \approx \eta_c/\eta \\
Z_{mn}^{TM} &= \frac{\eta_c}{\eta}[jF_n'(kr)/F_n(kr)]^* \approx \eta_c/\eta
\end{aligned}
\tag{33}
$$

where k and η_c are the wave number and intrinsic impedance in the conductor,

$$
k \approx (1-j)\sqrt{\frac{\omega\mu\sigma}{2}} \qquad \eta_c \approx (1+j)\sqrt{\frac{\omega\mu}{2\sigma}}. \tag{34}
$$

The characteristic impedance $r<R$ is extremely small for good conductors, so $V_{mn}\approx 0$, $r<R$. The current I_{mn} must be continuous at $r=R$. Hence for any mode the ratio of power dissipated to power radiated is given by

$$
\frac{P_{diss}}{P_{rad}} = \frac{|I_{mn}|^2\,\mathrm{Re}(Z_{mn}^-)}{|I_{mn}|^2\,\mathrm{Re}(Z_{mn}^+)} = \frac{\mathrm{Re}(\eta_c)}{\eta\,\mathrm{Re}(Z_{mn}^+)} \tag{35}
$$

where the superscripts $+$ and $-$ refer to $r>R$ and $r<R$, respectively. Dissipation factors D_n are defined for the case of equal TE_{mn} and TM_{mn} excitation as

$$
\begin{aligned}
D_n &= \frac{P_{diss}^{TE}+P_{diss}^{TM}}{P_{rad}^{TE}+P_{rad}^{TM}} = \frac{P_{diss}^{TE}}{2P_{rad}^{TE}} + \frac{P_{diss}^{TM}}{2P_{rad}^{TM}} \\
&= \frac{\mathrm{Re}(\eta_c)}{2\eta}\left[\frac{1}{\mathrm{Re}(Z_{mn}^{TE})} + \frac{1}{\mathrm{Re}(\substack{TM \\ Zn})}\right].
\end{aligned}
\tag{36}
$$

The D_n are independent of m because the Z_{mn} are independent of m. Using (21) and (22) one has

$$
\begin{aligned}
D_n(\beta R) &= \frac{\mathrm{Re}(\eta_c)}{2\eta}[|F_n'(\beta R)|^2 + |F_n(\beta R)|^2] \\
&= \frac{\mathrm{Re}(\eta_c)}{2\eta}[|u_n(\beta R)|^2 - 2]
\end{aligned}
\tag{37}
$$

where the $|u_n|^2$ are plotted in figure 1. Note that the D_n are essentially proportional to the $|u_n|^2$ when $\beta R<n$.

The dissipation factor for an antenna having equal excitation of TE and TM waves is defined as

$$
D = \frac{P_{\mathrm{diss}}}{P_{\mathrm{rad}}} = \frac{\sum_{m,n}\mathrm{Re}(P_{mn})D_n}{\sum_{m,n}\mathrm{Re}(P_{mn})} \tag{38}
$$

where P_{mn} is $P_{mn}^{TE}+P_{mn}^{TM}$. The second equality of (38) follows from the orthogonality property of the modes. For maximum gain all $P_{mn}=0$ except P_{1n}, $n\leq N$, which are given by (27). Using this and (17) one has

$$D=\frac{\sum_{n=1}^{N}(2n+1)|u_n(\beta r)|^2 D_n(\beta R)}{\sum_{n=1}^{N}(2n+1)|u_n(\beta r)|^2} \tag{39}$$

as the dissipation factor for an ideal lossy antenna of radius R, adjusted for maximum directive gain at a radius r, using wave functions of order $n\leq N$. For an antenna adjusted for maximum gain at infinity

$$D\xrightarrow[\beta r\to\infty]{}\frac{\sum_{n=1}^{N}(2n+1)D_n(\beta R)}{N^2+2N}. \tag{40}$$

Figure 6 shows the dissipation factor of antennas focused at infinity for several N. Calculations for antennas focused at other values of r show that D is essentially independent of the radius

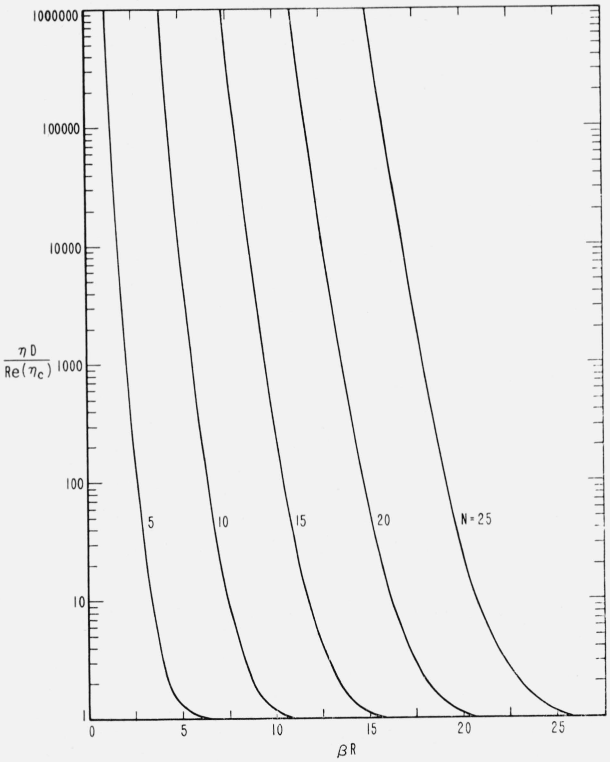

FIGURE 6. *Dissipation factors* (D) *for ideal lossy antennas adjusted for maximum gain at infinity, using wave functions of order* n≤N.

of focus, except when $r\approx R$. When $r=R$, the D of the antenna is approximately the D_n of the highest order wave present. If $\beta R>N$, all D_n are approximately $\mathrm{Re}(\eta_c)/\eta$, and the dissipation factor is

$$D\approx \mathrm{Re}\,(\eta_c)/\eta \tag{41}$$

and the antenna has very small loss. The efficiency of the antenna is

$$\%\text{ efficiency}=\frac{100\,P_{\mathrm{rad}}}{P_{\mathrm{rad}}+P_{\mathrm{diss}}}=\frac{100}{1+D}. \tag{42}$$

The antenna remains reasonably efficient until D is of the order of unity. The dissipation also changes the effective Q of the antenna, defined as

$$Q_{\mathrm{eff}}=\begin{cases}\dfrac{2\omega W^{\mathrm{elect}}}{P_{\mathrm{rad}}+P_{\mathrm{diss}}} & W^{\mathrm{elect}}>W^{\mathrm{mag}}\\[2ex] \dfrac{2\omega W^{\mathrm{mag}}}{P_{\mathrm{rad}}+P_{\mathrm{diss}}} & W^{\mathrm{mag}}>W^{\mathrm{elect}}\end{cases} \tag{43}$$

Because of the orthogonality of energy and power, $P_{\mathrm{rad}}+P_{\mathrm{diss}}$ is simply the sum of the corresponding quantities for each mode, that is,

$$P_{\mathrm{rad}}+P_{\mathrm{diss}}=\sum_{m,n}(P_{mn})_{\mathrm{rad}}(1+D_n). \tag{44}$$

Thus, for an antenna containing only the 1, n modes, $n\leq N$, one has instead of (28)

$$Q_{\mathrm{eff}}=\frac{\sum_{n=1}^{N}|a_n|^2\left(\frac{1}{2n+1}\right)Q_n}{\sum_{n=1}^{N}|a_n|^2\left(\frac{1+D_n}{2n+1}\right)}. \tag{45}$$

Finally, if the antenna is adjusted for maximum directive gain, the a_n are given by (17) and

$$Q_{\mathrm{eff}}=\frac{\sum_{n=1}^{N}(2n+1)|u_n(\beta r)|^2Q_n(\beta R)}{\sum_{n=1}^{N}(2n+1)|u_n(\beta r)|^2[1+D_n(\beta R)]}. \tag{46}$$

If the antenna is adjusted for maximum gain at infinity this becomes

$$Q_{\mathrm{eff}}\xrightarrow[\beta r\to\infty]{}\frac{\sum_{n=1}^{N}(2n+1)Q_n(\beta R)}{\sum_{n=1}^{N}(2n+1)[1+D_n(\beta R)]}. \tag{47}$$

The dashed lines in figure 5 show the effective Q for various N assuming $\mathrm{Re}(\eta_c)/2\eta=10^{-4}$, which corresponds to good conductors in the vicinity of 10,000 Mc. Further calculations show that the effective Q is essentially independent of the radius of focus, just as in the loss-free case. For large antennas, the maximum effective Q is of the order of that for a good spherical resonator constructed of the same metal.

While the directive gain of an antenna is unaffected by dissipation (assuming that the current distribution is unchanged), the overall gain

$$g(\beta r)=\frac{4\pi r^2\ \mathrm{Re}\ (S_r)_{\mathrm{max}}}{P_{\mathrm{rad}}+P_{\mathrm{diss}}} \tag{48}$$

is affected. This is the gain usually of primary interest for antenna evaluation. One can quite simply go back and maximize g, since the P_{diss} of each mode is related to the P_{rad} of each mode by (35). The only difference in the equations for G and those for g is that the factor $1/(2n+1)$ is replaced by $(1+D_n)/(2n+1)$. Hence, (16) becomes [4]

$$g_{\mathrm{max}}=\frac{1}{4}\sum_{n=1}^{N}\frac{(2n+1)}{1+D_n(\beta R)}|u_n(\beta r)|^2 \tag{49}$$

[4] This procedure is slightly in error since dissipation factors for TE and TM modes alone are not quite equal. The correction is small, however, until the dissipation factors become quite large.

and the wave amplitudes, related previously by (17), are related by

$$\frac{|a_n|}{|a_i|}=\frac{(1+D_i)(2n+1)|u_n|}{(1+D_n)(2i+1)|u_i|} \tag{50}$$

Note that the maximum gain is now a function of R (antenna size) as well as r (field point). If the antenna is focused at infinity (49) reduces to

$$g_{max}\xrightarrow[\beta r\to\infty]{}\sum_{n=1}^{N}\frac{2n+1}{1+D_n(\beta R)}. \tag{51}$$

So long as $D_N<1$ the overall gain g is substantially equal to the directive gain G. If $D_N>1$ the maximum gain depends upon the surface resistance of the metal, $\mathrm{Re}(\eta_c)$. However, the $|u_n|^2$ functions, which enter into D_n according to (37), rise very rapidly. For good conductors the "cut-off" of the summation of (46) occurs at approximately the same value of n as that for which the Q_n level off.

When the overall gain instead of the directive gain is maximized it affects the dissipation and quality factors. The dissipation factor becomes

$$D=\frac{\sum_{n=1}^{N}(2n+1)|u_n(\beta r)|^2D_n(\beta R)/[1+D_n(\beta R)]^2}{\sum_{n=1}^{N}(2n+1)|u_n(\beta r)|^2/[1+D_n(\beta R)]^2} \tag{52}$$

instead of (39). The principal difference between (39) and (52) is that in the latter case D levels off when the D_n becomes greater than unity. This can be thought of as due to the non-utilization of modes which are highly dissipative. The quality factor becomes

$$Q_{eff}=\frac{\sum_{n=1}^{N}(2n+1)|u_n(\beta r)|^2Q_n(\beta R)/[1+D_n(\beta R)]^2}{\sum_{n=1}^{N}(2n+1)|u_n(\beta r)|^2/[1+D_n(\beta R)]} \tag{53}$$

instead of (46). A plot of (53) would again give curves similar to the dashed lines of figure 5.

5. Discussion

To relate the analysis to practical antenna systems, define the radius R of an antenna system to be the radius of the smallest sphere that can contain it. The Q of the ideal loss-free antenna must then be less than or equal to the Q of any other loss-free antenna of radius R, since fields $r<R$ can only add to energy storage. In other words, the Q of figure 5 is a lower bound to the Q of an arbitrary loss-free antenna of radius R. It would be nice if one could also prove that the dissipation factor D of the ideal lossy antenna were a lower bound to the D of an arbitrary antenna of the same material and radius. The author has not been able to prove this. However, it will be assumed that the D of the ideal lossy antenna is of the same order of magnitude as for other antennas of the same material and radius. Calculations of the D for some practical antennas support this assumption.

It is evident from the foregoing analysis that a marked change in the behavior of an antenna of radius R occurs when wave functions of order $n>\beta R$ are present in its field. In the loss-free case the Q is large. In addition to this, in the lossy case the dissipation is large. In both cases the near-field is characterized by extremely large field intensities. The *normal gain* of an antenna is defined to be the maximum gain obtainable using wave functions of order $n\leq N=\beta R$. Hence, from (19), the normal gain of an antenna of radius R is

$$G_{\mathrm{norm}}=(\beta R)^2+2(\beta R) \tag{54}$$

in the radiation zone. Systems having larger gain than this are called *supergain antennas*. For large βR the gain of a uniformly illuminated aperture of radius R is equal to the above defined normal gain [5]. Therefore, one cannot obtain a gain higher than that of the uniformly illuminated aperture without resorting to a supergain antenna.

It is evident from figure 3 that the maximum near-zone gain of an antenna using wave functions $n \leq N = \beta R$ is essentially the same as the maximum far-zone gain. Hence (54) also defines the normal near-zone gain for all practical purposes. A uniformly illuminated and "focused" aperture (phase adjusted so that all elements contribute in-phase at some distance r) has a near-zone gain approximately equal to the far-zone gain of the "unfocused" aperture (uniform phase). Thus, a near-zone gain greater than that obtainable from a focused uniformly illuminated aperture cannot be obtained without resorting to a supergain antenna.

Having precisely defined the term "supergain," one can now consider the question of how much supergaining is possible. If a particular Q is chosen, the possible increase in gain over normal gain can be readily calculated. For example, if $Q=10^6$ is taken, the decibel increase in gain over normal gain is as shown in figure 7. Note that for small R, substantial increases in gain can be achieved, but for large R the increase becomes insignificant. The curve of figure 7 is relatively insensitive to the particular choice of Q, so long as it is high. This is evident from the rapid rise of the curves of figure 5. The choice $Q=10^6$ represents sort of an upper limit to practically significant Q's, since the bandwidth becomes absurdly narrow for higher Q's. Also, $Q=10^6$ represents the approximate upper limit for antennas constructed of metal, due to dissipation.

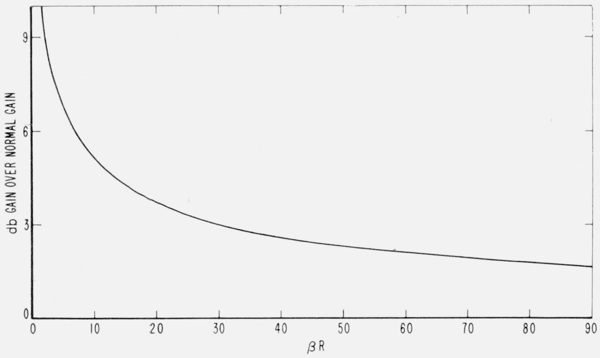

FIGURE 7. *Maximum increase in gain over normal gain obtainable by supergaining if a Q of 10^6 is allowed.*

It is evident from figure 5 that the amount of supergaining possible in large antennas is very small. Hence, for practical purposes, the uniformly illuminated aperture gives optimum gain. For small antennas, however, a significant increase over normal gain is possible. Perhaps the most common example of a small supergain antenna is the short dipole. The problems of narrow bandwidth and high losses associated with this antenna have been thoroughly treated, since it is one of the few antennas that can be used at very low frequencies.

6. References

[1] R. F. Harrington, On the gain and beamwidth of directional antennas, IRE Trans. **AP-6,** 219 (1958).

[2] L. J. Chu, Physical limitations of omnidirectional antennas, J. Appl. Phys. **19,** 1163 (1948).

[3] N. Marcuvitz, Waveguide handbook, pp. 96–100 (McGraw-Hill Book Co., New York, N.Y., 1951).

[4] R. F. Harrington, Introduction to electromagnetic engineering, pp. 167–177 (McGraw-Hill Book Co., New York, N.Y., 1958).

[5] S. Silver, Microwave antenna theory and design, p. 177 (McGraw-Hill Book Co., New York, N.Y., 1949).

BOULDER, COLO. (Paper 64D1–31)