Journal of Research of the National Bureau of Standards Vol. 59, No. 6, December 1957 Research Paper 2807

Excitation of Surface Waves on Conducting, Stratified, Dielectric-Clad, and Corrugated Surfaces

James R. Wait

An expression for the field of an electric dipole located over a flat surface with a specified surface impedance Z is derived from the formal integral solution by a modified saddle-point method. Using the value of Z appropriate for a homogeneous conducting ground, the general expressions reduce to those given by Norton. In this case the phase of Z lies between 0 and 45°. When the phase exceeds 45°, as it may for a stratified ground, the radiated wave of the dipole becomes partially trapped to the interface. This effect is most pronounced for an inductive surface where the phase of Z is 90°, in which case the energy of the wave is confined within a small distance from the surface. Such inductive surfaces are a metallic plane with a thin dielectric film or a corrugated surface.

This unifying treatment provides a link between the surface waves of Zenneck, Sommerfeld, Norton, and Goubau, and indicates that the phase angle of Z controls the extent to which these waves may exist for a dipole excitation.

1. Introduction

The propagation of radio waves along the surface of the ground has been discussed from a theoretical standpoint for many years. As long ago as 1907, Zenneck [1] [1] showed that a wave, which was a solution of Maxwell's equations, traveled without change of pattern over a flat surface bounding two homogeneous media of different conductivity and dielectric constants. When the upper medium is air and the lower medium is a homogeneous dissipative ground, the wave is characterized by a phase velocity greater than that of light and a small attenuation in the direction along the interface. Furthermore, this Zenneck surface wave, as it has been called, is highly attenuated with height above the surface. In 1909 Sommerfeld [2] solved the problem of a vertical dipole over a homogeneous ground (half-space). In an attempt to explain the physical nature of his solution, he divided the expression for the field into a "space wave" and a "surface wave." Both parts, according to Sommerfeld, are necessary in order to satisfy Maxwell's equations and the appropriate boundary conditions. The surface-wave part varied inversely as the square root of the range, and it was identified as the radial counterpart of the plane Zenneck surface wave.

For many years it was believed that this Sommerfeld surface wave was the predominant component of the field radiated from a vertical antenna over a finitely conducting ground. It was not until 1935 that an error in sign in Sommerfeld's 1909 paper [2] was pointed out by Norton [3], which also partly accounted for the unusual calculated field-strength curves of Rolf [4]. At about this time there was a series of papers by Weyl, Van der Pol, Niessen, Wise, and Norton, deriving more accurate representations for the field of the dipole. A discussion of this later work has been given by Bouwkamp [5]. Norton [6, 7], in particular, has developed his formula for the field components to a stage where numerical results can readily be obtained. It is now generally accepted that the Sommerfeld surface wave (or the radial Zenneck wave) does not bear any similarity to the total field of a vertical dipole over a homogeneous conducting earth. In fact, the field excited by a dipole varies as $1/d$, where d is the distance for low frequencies, and it varies as $1/d^2$ for high frequencies. Norton has suggested that the field in air of a dipole over a homogeneous ground be expressed as a sum of three components: A direct ray (or primary influence), a reflected ray which is to be modified by an appropriate Fresnel reflection coefficient, and a correction term. Norton has described the first and second components as the space wave; the third or correction term, the surface wave. This seems to be a logical step,

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

[2] It has been suggested by B. M. Fannin (Final Report, Pt. III, Investigations of air to air and air to ground experimental data, Elec. Eng. Dept., Cornell University, 1951) that Summerfeld did not actually make an "error in sign" in his 1909 paper. Fannin's contention is that Rolf and other workers simply misinterpreted Sommerfeld's formula for the attenuation factor. This can hardly be so, as Sommerfeld himself in 1910 (The propagation in wireless telegraphy, Jahrb. d. Drahtl. Tel. 4, 157, 1910) used the incorrect formula and its resulting series expansion.

although, taken separately, the space and surface waves of Norton are not solutions of Maxwell's equations. On the other hand, his "space wave" is the contribution that would be derived on the basis of geometrical optics, and his "surface wave" is the correction from wave theory. This latter term will be called the "Norton surface wave" as distinct from the Zenneck and Sommerfeld surface waves, and the trapped surface waves discussed below.

It was pointed out by Sommerfeld [8] nearly 60 years ago that a straight cylindrical conductor of finite conductivity can act as a guide for electromagnetic waves. Some 50 years later Goubau [9, 10] demonstrated that such a cylindrical surface can be launched with reasonable efficiency from a coaxial line whose outer surface is terminated in a conical horn. The improvement in the transmission properties by coating the wire with a thin dielectric film has been discussed in detail by Goubau [10]. The plane counterpart of the Goubau-Sommerfeld cylindrical surface wave is obtained when a flat metallic surface is coated with a dielectric film. Atwood [11] has discussed the nature of the surface waves that can exist in a structure of this type. When the film thickness is small compared to the wavelength, the single propagating mode has a phase velocity slower than that of light and is attenuated rapidly above the surface in the air. Such surfaces have been called inductive because the normal surface impedance, looking into the surface, is almost purely imaginary for a low-loss dielectric on a well-conducting base. A similar type of surface wave can exist over a corrugated surface [12, 13, 14] which is inductive if the periodicity and depth of the corrugation are small compared to the wavelength.

The excitation of surface waves on dielectric-clad plane conducting surfaces has been discussed by Tai [15] for a line current source, and Brick [16] for a dipole source. Corresponding treatments for corrugated surfaces have been given by Cullen [17] for a line magnetic or infinite slot source, and Barlow and Fernando [18] for a vertical electric dipole source. In the case of dipole excitation, the field varies predominantly as the inverse square root of the distance along both the dielectric-clad and the corrugated surfaces. Surface waves of this type, which have a phase velocity less than that of light, can be called "trapped surface waves" because they carry most of their energy within a small distance from the interface [19, 20].

It is the purpose of the present paper to study the general problem of a vertical dipole over a flat surface with a specified surface impedance, Z. A solution is obtained for the field in terms of the error function of complex argument. It is shown that, using an appropriate value for Z, the expressions for the Norton space and surface wave for a homogeneous conducting ground are obtained as a special case [7]. By allowing Z to be purely imaginary and using asymptotic expansions for the error function, the field is shown to be of a trapped-surface-wave type [14, 19]. Another special case is for a two-layer conducting ground [21, 22] which can exhibit some of the features of trapped surface waves. Thus, a general connection between these various forms of seemingly unrelated surface waves is developed.

2. Basic Formulation

A vertical electric dipole of length ds and current I is located at a height h over a flat ground plane which exhibits the property of surface impedance. Here the ratio of the tangential electric and magnetic fields is specified on the surface of the ground plane and is denoted by Z. Choosing cylindrical coordinates, the dipole is located at $(0,0,h)$ and the ground plane at $z=0$. As will be seen, the Hertz vector at (ρ,ϕ,z) need have only a z component, Π, to satisfy the boundary condition; consequently, the fields are given by [3]

$$E_\rho=\frac{\partial^2\Pi}{\partial\rho\partial z},\qquad E_\phi=0,\qquad E_z=\left(\beta^2+\frac{\partial^2}{\partial z^2}\right)\Pi \tag{1}$$

$$H_\rho=0,\qquad H_\phi=\frac{\beta^2}{i\mu\omega}\frac{\partial\Pi}{\partial\rho},\qquad H_z=0,$$

where $\mu=4\pi\times10^{-7}$ and $\beta=2\pi$/free space wavelength. The resultant Hertz vector, Π, can now

[3] A time factor exp $(i\omega t)$ is suppressed throughout.

be represented as a sum of a primary and secondary part in the manner

$$\Pi = \Pi_p + \Pi_r, \tag{2}$$

where

$$\Pi_p = \bar{c} e^{-i\beta r_1}/r_1 \qquad \text{with } \bar{c} = \frac{Ids}{4\pi i\omega\epsilon} \qquad \text{and} \qquad r_1 = [(z-h)^2+\rho^2]^{1/2} \tag{3}$$

and Π_r is to be regarded as the influence of the ground plane. In the most elementary case of a perfectly conducting ground plane (i. e., $Z=0$), Π_r would be given by the simple relation

$$\Pi_r = \bar{c} e^{-i\beta r_2}/r_2 \qquad \text{with} \qquad r_2 = [(z+h)^2+\rho^2]^{1/2}. \tag{4}$$

To find Π_r (or Π) for the more general case, it is necessary to express the primary field in terms of an integral representation [2] (sometimes called Sommerfeld's integral), as follows:

$$\frac{e^{-i\beta r_1}}{r_1} = \int_0^\infty \frac{J_0(\lambda\rho) e^{-u|z-h|} \lambda d\lambda}{u}, \tag{5}$$

where $u = (\lambda^2 - \beta^2)^{1/2}$ and $J_0(\lambda\rho)$ is a Bessel function of order zero and argument $\lambda\rho$. Due to the presence of the branch point of u, the contour of the integral is indented *upward* slightly at $\lambda=\beta$. Physically, this integral can be regarded as a means of representing the primary field as a spectrum of plane waves whose angles of incidence, θ, on the ground plane are related to the integration variable by $\lambda = \beta \sin\theta$. Consequently, the angles include both real and complex values, ranging from 0 to $\pi/2$ along the real axis and from $\pi/2$ to $\pi/2 + i\infty$ along a line parallel to the imaginary axis.

The form of eq (5) suggests writing Π_r in the following form:

$$\Pi_r = \bar{c} \int_0^\infty \frac{J_0(\lambda\rho) e^{-u(z+h)} R(\lambda) \lambda d\lambda}{u}, \tag{6}$$

where $R(\lambda)$ is now a reflection coefficient for plane wave incident on the ground plane at angle of incidence $\theta = \sin^{-1}\lambda/\beta$. Consequently,

$$R(\lambda) = \frac{u - i\beta\Delta}{u + i\beta\Delta}, \tag{7}$$

where

$$\Delta = \frac{Z\beta}{\mu\omega} = \frac{Z}{120\pi}.$$

In the limiting case of a perfect conductor ($Z=0$) it can be seen that $R(\lambda)=1$, and the integral is of the form of a Sommerfeld integral and Π_r is again given by eq (4).

The Hertzian potential Π can then be written in the form

$$\Pi = \bar{c} \left[\frac{e^{-i\beta r_1}}{r_1} + \frac{e^{-i\beta r_2}}{r_2} - 2P \right] \tag{8}$$

where $2P$ is the effect of the imperfection of the ground plane which results from a finite value of the surface impedance Z.

3. Evaluation of the Integral P

The integral

$$P = \int_0^\infty \frac{(i\beta\lambda\Delta) e^{-u(z+h)}}{(u+i\beta\Delta)u} J_0(\lambda\rho) d\lambda \tag{9}$$

must now be evaluated. Before proceeding, certain conditions and restrictions must be clearly stated. At a later stage, these may be relaxed somewhat. For the moment

$$|\Delta|<1, \qquad 0<\arg \Delta<\frac{\pi}{4}, \qquad \text{and } \beta\rho>>1.$$

Since

$$J_0(x)=\frac{1}{2}\left[H_0^{(1)}(x)+H_0^{(2)}(x)\right]$$

and

$$H_0^{(2)}(x)=-H_0^{(1)}(-x)$$

where $H_0^{(1)}$ and $H_0^{(2)}$ are Hankel functions of the first and second kind, respectively, it follows that

$$P=\frac{i\beta\Delta}{2}\int_{-\infty}^{\infty}\frac{\lambda e^{-u(z+h)}}{(u+i\beta\Delta)u}H_0^{(1)}(\lambda\rho)d\lambda \tag{10}$$

Introducing the substitutions $\lambda=\beta \cos \alpha$ and $\Delta=\sin \alpha$, it is seen that

$$P=\frac{i\beta\Delta}{4}\int_{-i\infty}^{\pi+i\infty}\frac{\cos \alpha e^{-i\beta(z+h)\sin\alpha}\,H_0^{(1)}(\beta\rho \cos\alpha)}{\sin\frac{\alpha+\alpha_0}{2}\cos\frac{\alpha-\alpha_0}{2}}\,d\alpha. \tag{11}$$

Essentially P is now a spectrum of plane waves traveling away from the ground plane so that the vertical component of the propagation constant is $\beta \sin \alpha$. Again, complex angles are included in the spectrum. When $|\beta\rho \cos \alpha|>>1$, the first term of the asymptotic expansion of the Hankel function can be employed. That is,

$$H_0^{(1)}(\beta\rho \cos \alpha)\cong\left[\frac{2}{\pi\beta\rho \cos\alpha}\right]^{1/2}e^{i\beta\rho\cos\alpha}e^{-i\pi/4}. \tag{12}$$

When this is inserted into eq (11), the equation for P becomes

$$P=-e^{-i\pi/4}\frac{\Delta}{2}\left(\frac{\beta}{2\pi\rho}\right)^{1/2}\int_{-i\infty}^{\pi+i\infty}\frac{(\cos\alpha)^{1/2}e^{-i\beta R\cos(\theta-\alpha)}}{\sin\frac{\alpha+\alpha_0}{2}\cos\frac{\alpha-\alpha_0}{2}}\,d\alpha. \tag{13}$$

where R and θ are defined by $z+h=R\sin\theta$ and $\rho=-R\cos\theta$. In the usual case, the separation ρ between transmitter and receiver is large compared to their respective heights, h and z, and consequently, θ is slightly less than π. It can be seen that the exponential factor in the integrand of P is rapidly varying except for a region near $\alpha=\theta\cong\pi$. This is, of course, the saddle point of the integrand or the point of stationary phase. The important part of the integrand is in the region near the saddle point. In fact, this is the justification for employing the asymptotic expansion for the Hankel functions $H_0^{(1)}$; the argument $\beta\rho \cos \alpha$ is always large in the region near the saddle point if $\beta\rho$ itself is large and $z+h$ is less than ρ.

The integral for P is now in a form where the saddle-point method of integration can be applied. The usual technique [23] is to deform the contour (which in the present case is along the negative imaginary axis, the real axis from 0 to π, and then along a line parallel to the positive imaginary axis) to a path of steepest descent. For example, if we let

$$\cos(\alpha-\theta)=1-i\tau^2$$

and let τ range from $-\infty$ to $+\infty$ through real values, an integral of the type

$$\int_{-i\infty}^{\pi+i\infty} G(\cos\alpha)e^{-i\beta R\cos(\alpha-\theta)}\,d\alpha \tag{14}$$

where $G(\cos\alpha)$ is slowly varying at $\alpha=\theta$, is transformed to

$$(2i)^{1/2}e^{-i\beta R}\int_{-\infty}^{+\infty} \frac{G(\cos\alpha)}{\sqrt{1-i\tau^2/2}}\,e^{-i\beta R\tau^2}d\tau. \tag{15}$$

In this deformation of the contour, account must be made for the singularities of the integrand that are crossed. In the case of poles, $2\pi i$ times the sum of the residues is added to the new integral. The final step in this classical procedure is to expand $G(\cos\alpha)/\sqrt{1-i\tau^2/2}$ in a power series in τ enabling the integration to be carried out term by term. The leading term of this resulting asymptotic expansion is

$$(2\pi i)^{1/2}G(\cos\theta)e^{-i\beta R}(\beta R)^{-1/2} \tag{16}$$

and succeeding terms contain $(\beta R)^{-3/2}$, $(\beta R)^{-5/2}$, and so on.

As is often the case, the contribution from the saddle point (which is the branch point in the original λ plane) cannot be separated from the pole(s) of the integrand. For example, in the present problem,

$$G(\cos\alpha)=\frac{(\cos\alpha)^{1/2}e^{-i\beta R\cos(\theta-\alpha)}}{\sin\dfrac{\alpha+\alpha_0}{2}\cos\dfrac{\alpha-\alpha_0}{2}} \tag{17}$$

has a pole at $\alpha=\pi+\alpha_0$ which is near the saddle point $\alpha=\theta$ since α_0 is small and θ is near π. In other words, the integrand is not slowly varying near the saddle point on account of the factor $\cos(\alpha-\alpha_0)/2$. The integral is then of the form

$$I=\int_{-i\infty}^{\pi+i\infty}\frac{e^{-i\beta R\cos(\theta-\alpha)}}{\cos\dfrac{\alpha-\alpha_0}{2}}\,d\alpha \tag{18}$$

after the slowly varying factors have been separated out. The necessary modification of the saddle-point method to treat integrals of this type was devised by Van der Waerden [24] and Clemmow [25]. On making the usual deformation of the contour via the substitution $\cos(\theta-\alpha)=1-i\tau^2$, the integral now becomes

$$I=-2e^{-i\pi/4}e^{-i\beta R}\sqrt{2}\cos\left(\frac{\theta-\alpha_0}{2}\right)\int_{-\infty}^{+\infty}\frac{e^{-\beta R\tau_2}d\tau}{\tau^2+i2\left[\cos\left(\dfrac{\theta-\alpha_0}{2}\right)\right]^2}. \tag{19}$$

It can readily be verified that no poles of the integrand are crossed in this deformation for the argument of α_0 (or Δ) less than 45°. The integral to consider is now of the type

$$A=\int_{-\infty}^{+\infty}\frac{e^{-x\tau^2}}{\tau^2+c}\,d\tau \tag{20}$$

where, for real τ, $x=\beta R$ and

$$c=i2\left[\cos\left(\frac{\theta-\alpha_0}{2}\right)\right]^2.$$

Since $\int_{-\infty}^{+\infty}e^{-x\tau^2}d\tau=\left(\frac{\pi}{x}\right)^{1/2}$, it follows that

$$A=\frac{1}{c}\left(\frac{\pi}{x}\right)^{1/2}-\frac{2}{c}\int_0^{\infty}\frac{\tau^2}{\tau^2+c}\,e^{-x\tau^2}d\tau. \tag{21}$$

Now it can be seen readily that

$$\frac{d}{dx} e^{-xc}A = -\left(\frac{\pi}{x}\right)^{1/2} e^{-xc} \tag{22}$$

and an integration with respect to x from x to ∞ leads to

$$e^{-xc}A = \int_x^\infty \left(\frac{\pi}{x}\right)^{1/2} e^{-xc}dx, \tag{23}$$

or, after a change of variable,

$$A = 2(\pi/c)^{1/2}e^{xc}\int_{\sqrt{xc}}^\infty e^{-z^2}dz, \tag{24}$$

which holds for *all* c in the c plane with a cut along the negative real axis. This can be written

$$A = \frac{\pi}{\sqrt{c}} e^{xc} erfc\sqrt{xc}, \tag{25}$$

where *erfc* is the complement of the error integral given by

$$erfc z_0 = \frac{2}{\sqrt{\pi}}\int_{z_0}^\infty e^{-z^2}dz, \tag{26}$$

where the integration in the z plane is directed from z_0 towards the right to $+\infty$. Therefore

$$I = -i2\pi e^{-i\beta R}e^{-w}erfc(iw^{1/2}), \tag{27}$$

where

$$w = -\beta Rc = -2i\beta R\left[\cos\left(\frac{\theta-\alpha_0}{2}\right)\right]^2,$$

and finally, since $R = r_2$,

$$P = \int_0^\infty \frac{(i\beta\lambda\Delta)}{(u_0+i\beta\Delta)} e^{-u(z+h)}J_0(\lambda\rho)d\lambda \tag{28}$$

$$\simeq i(\pi p)^{1/2}e^{-w}erfc(iw^{1/2})e^{-i\beta r_2}/r_2,$$

where

$$w \simeq p\left(1+\frac{(z+h)}{\Delta r_2}\right)^2 \tag{29}$$

and

$$p = -\frac{i\beta r_2}{2}\Delta^2 = |p|e^{ib}.$$

While the derivation was carried out for the argument of Δ lying in the range 0 to $\pi/4$ or b in the range 0 to $-\pi/2$, the formula is actually valid for all values of b if the above definition of the error function complement is used. The justification for this step is based on the principle of analytical continuation. If, in the initial derivation, the argument of Δ was allowed to exceed $\pi/4$, we would find that a pole would be crossed in the deformation of the contour in the α plane. There would be a corresponding change in the integral A but, after a change of the sign of the variable τ, the expression yielded for I would be unchanged in form if we are consistent in the definition of the complex error function.

4. General Properties of the Solution

The general behavior of the result will be considered now. For the moment the heights of the transmitter and receiver are assumed to be very small. That is $(z+h) <<< \rho$ and, consequently, $w \simeq p$ and $r_1 \simeq r_2 \simeq \rho$. The Hertz potential then has the form

$$\Pi = 2\bar{c}e^{-i\beta\rho}F(p), \tag{30}$$

where

$$F(p)=1-i(\pi p)^{1/2}e^{-p}erfc(ip^{1/2}). \tag{31}$$

The vertical electric field is then given by

$$E_z=\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\frac{\partial\Pi}{\partial\rho} \tag{32}$$

and, since $\beta\rho>>1$ and $F(p)$ is a slowly varying function, it follows that

$$E_z=+2\bar{c}\beta^2e^{-i\beta\rho}\rho^{-1}F(p). \tag{33}$$

For a highly conducting flat ground plane, $[|p|<<1]$, the electric field denoted by E_z^∞ is given by

$$E_z^\infty=2\bar{c}\beta^2e^{-i\beta\rho}\rho^{-1}. \tag{34}$$

This result can be obtained by elementary methods, being just the free space radiation field of an electric dipole of moment $2\bar{c}$. The factor 2 arises from the imaging of the primary dipole in the perfectly reflecting surface. The ratio

$$F(p)=E_z/E_z^\infty \tag{35}$$

is by definition the complex attenuation factor and is a measure of the efficiency of transmission of waves along the interface for a vertical-electric-dipole source.

The behavior of $F(p)$ is very dependent on the phase of p, and since

$$P=|p|e^{ib}=-\frac{i\beta\rho}{2}\Delta^2=-\frac{i\beta\rho}{2}\left(\frac{Z}{120\pi}\right)^2, \tag{36}$$

the surface impedance, Z, plays an important role. For example, in the case of a homogeneous ground of conductivity σ and dielectric constant ϵ,

$$\Delta=\frac{Z}{120\pi}=\frac{1}{N}\left(1-\frac{1}{N^2}\right)^{1/2} \tag{37}$$

with

$$N=\left[\frac{\epsilon-i\sigma/\omega}{\epsilon_0}\right]^{1/2}.$$

The refractive index, N, is usually large compared to unity with a phase near, but always less than $-45°$. Correspondingly, for a homogeneous half-space, the phase b lies between 0° and $-90°$. This is precisely the case that has been treated in detail by Norton, who gives numerical results of $F(p)$ as a function of $|p|$ for b varying from 0 to $-180°$.[4] In all cases for $|p|>0$, $|F(p)|$ was less than unity. In other words, the field E_z was never greater than E_z^∞ for the perfectly conducting ground plane. For purposes of numerical work, the error-function complement can be expanded in an ascending power series, as is well known. This leads to the following series formula for $F(p)$:

$$F(p)=1-i(\pi p)^{1/2}e^{-p}-2p+\frac{(2p)^2}{1\cdot3}-\frac{(2p)^3}{1\cdot3\cdot5}+\ldots, \tag{38}$$

which is absolutely convergent.

The second term of this series, the quantity $-i(\pi p)^{1/2}e^{-p}$, is rather interesting if it is considered by itself. The vertical electric field, E_z^s, which would result by ignoring the other terms, is

$$E_z^s=-2\bar{c}\beta^2\frac{e^{-i\beta\rho}}{\rho}i(\pi p^{1/2})e^{-p} \tag{39}$$

$$=\bar{c}e^{-3\pi i/4}\beta^2\frac{\sqrt{2\pi}}{\sqrt{\beta\rho}}\Delta e^{-i\beta\left[1-\frac{\Delta^2}{2}\right]\rho}. \tag{40}$$

[4] Norton used a time factor $e^{-i\omega t}$, whereas $e^{i\omega t}$ is employed in the present paper.

This field has the characteristics of a Zenneck surface wave; it varies inversely with the distance along with an exponential attenuation, and it has a phase velocity greater than that of light. By retracing the steps in the derivation it is not difficult to show that E_z^s is the residue at the pole α_0 in the complex α plane. Taken by itself, E_z^s could easily exceed E_z^∞ by a great amount if b was near $-90°$ corresponding to the phase of Z (or Δ) being near zero. This corresponds to a small conductivity in the lower half space.

The above argument would indicate that the Zenneck surface wave is excited by a vertical dipole source on a homogeneous half-space. Of course, when the other terms in the series for $F(p)$ are considered, the total field does not have the attenuation characteristics of a Zenneck surface wave. For example, if the asymptotic expansion for the complement of the error function is employed, the following series for $F(p)$ results:

$$F(p) = -\frac{1}{2p} - \frac{1\cdot 3}{(2p)^2} - \frac{1\cdot 3\cdot 5}{(2p)^3} - \frac{1\cdot 3\cdot 5\cdot 7}{(2p)^4} - \ldots, \tag{41}$$

which are also asymptotic and valid for large values of p and $-2\pi < b < 0$. This would indicate that the field E_z at large ranges varies inversely as the square of distance, whereas the Zenneck surface wave varies inversely as the square root of the distance. This viewpoint is confirmed by the work of Wise [26] and Rice [27], who develop rigorous expansions for the Hertz potential on a homogeneous earth.

The absence of any features associated with a trapped surface wave on a homogeneous earth is due to the restricted range of b (namely between 0 and $-\pi/2$) for a homogeneous earth. It is of interest to note that Sommerfeld in his 1909 paper had an error in sign which was finally pointed out by Norton [3]. This error is equivalent to the interchanging of the b and $-b$. In other words, the complex conjugate of p is used in place of the proper quantity. Rolf [4] employed Sommerfeld's formulas along with the incorrect sign. Rolf's series expansions for $F(p)$ would correspond to eq (38) with b replaced by $-b$. His calculations therefore correspond to values of $F(p)$ for b from 0 to $+90°$. The appearance of his curves is indeed weird. As will be indicated below, the function $F(p)$ can exceed unity in certain instances when b is positive. This is indicated in Rolf's curves. The irregularities and dips to zero, however, must be attributed to errors in computation. The series formula for $F(p)$, although mathematically convergent, takes a great number of terms when p is greater than 5 or so. And, furthermore, the asymptotic expansion of $F(p)$ is not accurate for the real part of p less than about 15. The anomalous features in Rolf's curves occur just where the accurate computation of $F(p)$ would require tens or hundreds of terms in the power series formula. This viewpoint is confirmed when more elaborate calculations are carried out [21].

The nature of the attenuation function $F(p)$ when b is positive can be best seen from an asymptotic development. This is obtained by noting that

$$erfc(ip^{1/2}) = \frac{2}{\sqrt{\pi}}\int_{ip^{1/2}}^{\infty} e^{-x^2}dx = 2 - \frac{2}{\sqrt{\pi}}\int_{ip^{1/2}}^{\infty} e^{-x^2}dx. \tag{42}$$

The latter integral asymptotically vanishes for $p\to\infty$ if the real part of $ip^{1/2}$ is less than zero. Therefore

$$erfc(ip^{1/2}) \simeq 2 - \frac{ie^{+p}}{\sqrt{\pi p}}\left(1 + \frac{1}{2p} + \frac{1\cdot 3}{(2p)^2} + \frac{1\cdot 3\cdot 5}{(2p)^3} + \ldots\right) \tag{43}$$

for $2\pi > b > 0$ and $|p| >> 1$. Utilizing this result, it then follows that

$$\begin{aligned} F(p) &= 1 - i\sqrt{\pi p}\, e^{-p} erfc(ip^{1/2}) \\ &\simeq -2i\sqrt{\pi p}\, e^{-p} - \frac{1}{2p} - \frac{1\cdot 3}{(2p)^2} - \frac{1\cdot 3\cdot 5}{(2p)^3} - \ldots \end{aligned} \tag{44}$$

in the asymptotic sense. The term $-2i\sqrt{\pi p}\; e^{-p}$ has all the characteristics of a surface wave.

It is not present in the asymptotic development for $F(p)$ when b is negative. Of course, for dissipative media such as a stratified conductor, the real part of p is large and the surface-wave term almost vanishes compared to the term $-1/2p$.

The most interesting case is when the surface impedance is purely inductive. For example, if a flat perfectly conducting ground plane is coated with a thin layer (of thickness h) of pure dielectric (with refractive index N), then the surface impedance, Z, is given to a good approximation by [19]

$$Z \cong i\mu\omega h\left[1-\frac{1}{N^2}\right]. \tag{45}$$

Another example is when a flat metal surface is corrugated in the form of periodic rectangular slots of width d and depth D with a separation w between slot centers. The surface impedance for the magnetic vector parallel to the directions of the corrugations is then

$$Z = i\frac{d}{w}\eta_0 \tan \beta D, \tag{46}$$

provided that there are at least five corrugations per wavelength along the surface [13]. For $D < \lambda/4$, the surface impedance is purely inductive, and since

$$p = -\frac{i\beta\rho}{2}(Z/\eta_0)^2, \tag{47}$$

$p+i|p|$ and $b=\pi/2$. The attenuation function then has the asymptotic development,

$$F(p) \cong -2e^{i3\pi/4}\sqrt{\pi|p|}\, e^{-i|p|} + \frac{i}{2|p|} + \frac{3}{4|p|^2} - \frac{15i}{8|p|^3} \cdots, \tag{48}$$

where

$$|p| = \frac{\beta\rho}{2}\frac{X^2}{\eta_0^2} \text{ and } X = |Z|.$$

The vertical electric field at $z=0$ is then given by

$$E_z \cong 2c\beta^2 e^{-i\beta\rho}\rho^{-1}F(p). \tag{49}$$

The surface wave term now varies as

$$\frac{1}{\sqrt{\beta\rho}} e^{-i\beta\left(1+\frac{X^2}{2\eta_0^2}\right)\rho},$$

which *is not a Zenneck surface wave* since the phase velocity, v, given by

$$v = \frac{1}{1+X^2/(2\eta_0^2)} \frac{1}{\sqrt{\mu_0\epsilon_0}}, \tag{50}$$

is less than the velocity of light, $1/\sqrt{\mu_0\epsilon_0}$.

If the surface is purely capacitive such that $Z=-iX'$, the argument of the error function is

$$ip^{1/2} = e^{-i\pi/4}\sqrt{\beta\rho/2}(X'/\eta_0)^2, \tag{51}$$

so that $erfc(ip^{1/2})$ can be expanded directly to lead to the asymptotic representation

$$F(p) \cong -\frac{i}{2|p|} + \frac{3}{2|p|^2} + \frac{15i}{8|p|^3} \cdots, \tag{52}$$

which contains no surface-wave term.

5. Height-Gain Functions

It is desirable now to return to the general formula for $F(p)$ which is a (modified) saddle point approximation for the field of a vertical dipole over the impedive surface. For $h=0$, it is

$$W=1-i\sqrt{\pi p}e^{-w}erfc(iw^{1/2}), \tag{53}$$

where

$$w=p\left(1+\frac{z}{\Delta r_2}\right)^2 \simeq p\left(1+\frac{z}{\Delta p}\right)^2.$$

For small heights and receiver it is of interest to develop W in a power series in z. For example,

$$W(z)=W(0)+z\left(\frac{dW}{dz}\right)_{z=0}+\frac{z^2}{2}\left(\frac{d^2W}{dz^2}\right)^2_{z=0}+\cdot\cdot\cdot\cdot \tag{54}$$

Now

$$\frac{dW}{dz}=\left[-\frac{p}{w}+i\sqrt{\pi p}\,e^{-w}erfc(iw^{1/2})\right]\frac{dw}{dz} \tag{55}$$

and

$$\frac{dw}{dz}=2p\left(1+\frac{z}{\Delta\rho}\right)\frac{1}{\Delta\rho}.$$

Therefore

$$\left[\frac{dW}{dz}\right]_{z=0}=ik\Delta W(0)=ik\Delta F(p)$$

and, to a first order in z,

$$W(z)\simeq F(p)\ [1+ik\Delta z]. \tag{56}$$

The generation of higher-order terms using this method is cumbersome, and it is really better to return to the original integral for the Hertz potential $\Pi(z)$ given by

$$\Pi(z)=2\bar{c}\int_0^\infty \frac{\lambda J_0(\lambda\rho)}{u+i\beta\Delta}e^{-uz}d\lambda \tag{57}$$

for $h=0$. Expanding the exponential, this is rewritten

$$\Pi(z)=2\int_0^\infty [1-uz+u^2z^2/2-\ \ldots\ ]\frac{\lambda J_0(\lambda\rho)}{u+i\beta\Delta}d\lambda \tag{58}$$

$$=2\bar{c}\left[\Pi(0)+i\beta\Delta\,\Pi(0)+\frac{z^2}{2}\left(\frac{\partial^2\Pi(z)}{\partial z^2}\right)_{z=0}+\ \ldots\ \right].$$

Now since $\Pi(z)$ satisfies the wave equation, $\partial^2/\partial z^2$ can be replaced by

$$-\beta^2-\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho}$$

so that

$$\Pi(z)=\left[1+i\beta\Delta z-\frac{z^2}{2}\left(\beta^2+\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho}\right)+\ \ldots\ \right]\Pi(0). \tag{59}$$

Since

$$\Pi(0)=2\bar{c}e^{-i\beta\rho}\rho^{-1}F(p), \tag{60}$$

the contribution from the second-order term is

$$\frac{z^2}{2}\left[\beta^2+\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho}\right]\Pi(0)=\frac{2\overline{c}e^{-i\beta\rho}}{2\rho}F(p)\left\{\frac{z^2}{\rho^2}+\frac{i(\beta z)^2}{(\beta\rho)}-\frac{z^2}{2\rho}\frac{F'(p)}{F(p)}-i\beta z^2\frac{F'(p)}{F(p)}+\frac{z^2}{2}\frac{F''(p)}{F(p)}\right\}, \quad (61)$$

where the prime(s) indicate a derivative(s) with respect to ρ. It is not difficult to see that for $2\pi<b<0$, $F(p)$ is slowly varying compared to other factors and the term in curly brackets is much less than unity if $\beta\rho>>1$ and $\rho^2>>z^2$. In the case of the inductive surface (i. e., $2\pi>b>0$), it is convenient to write

$$F(p)=F_s(p)+F_a(p), \quad (62)$$

where asymptotically

$$F_s(p)\simeq-2i\sqrt{\pi p}\,e^{-p} \quad (63)$$

is the surface wave term and

$$F_a(p)\simeq-\frac{1}{2p}-\frac{3}{4p^2}-\frac{15}{8p^3}-\ \cdots \quad (64)$$

is the remaining (nonsurface) contribution. [For the case $-2\pi<b<0$, $F(p)\simeq F_a(p)$].

Now

$$\frac{F'_a(p)}{F_a(p)}\sim-\frac{1}{\rho}+0\left(\frac{1}{p^2}\right)$$

and

$$\frac{F''_a(p)}{F'_a(p)}\sim-\frac{2}{\rho}+0\left(\frac{1}{p^2}\right)$$

but

$$\frac{F'_s(p)}{F_s(p)}\sim\frac{i\beta\Delta^2}{2}$$

and

$$\frac{F''_s(p)}{F_s(p)}\sim\left(-\frac{i\beta\Delta^2}{2}\right)^2.$$

Therefore, the Hertz potential $\Pi_s(z)$, corresponding to the surface-wave field, is given by

$$\Pi_s(z)\simeq\Pi_s(0)\left[1+i\beta\Delta z+\frac{(i\beta\Delta z)^2}{2}+\cdots\right] \quad (65)$$

$$\simeq\Pi_s(0)e^{i\beta\Delta z}$$

whereas

$$\Pi_a(z)\simeq\Pi_a(0)[1+i\beta\Delta z]. \quad (66)$$

In the case of a purely inductive surface ($Z=iX$), the exponential factor for the surface term is given by

$$e^{i\beta\Delta z}=e^{-(X\beta/\eta_0)z}=e^{-\epsilon\omega Xz}, \quad (67)$$

indicating a rapid decay of the field with height.

By a slight extension of the above, the case of both transmitter and receiver raised yields

$$\Pi_s(z,h)\simeq\Pi_s(0,0)e^{i\beta\Delta(z+h)} \quad (68)$$

and

$$\Pi_a(z,h)\simeq\Pi_a(0,0)(1+i\beta\Delta z)(1+i\beta\Delta h). \quad (69)$$

Still another approach to the effect of raised antennas, which is particularly appropriate for large heights, is now considered. In general,

$$\frac{\Pi}{c}=\frac{e^{-i\beta r_1}}{r_1}+\frac{e^{-i\beta r_2}}{r_2}-2P, \tag{70}$$

where

$$P=\frac{p}{w}\,[1-F(w)]\,\frac{e^{-i\beta r_2}}{r_2} \tag{71}$$

with $F(w)=F_a(w)+\epsilon F_s(w)$, where $\epsilon=0$ for $-2\pi<b<0$, $\epsilon=1$ for $2\pi>b>0$, and $w=p(1+C/\Delta)^2$, $C=(z+h)/r_2$. Asymptotically,

$$F_s(w)=-2i\sqrt{\pi w}\,e^{-w} \tag{72}$$

and

$$F_a(w)=-\frac{1}{2w}-\frac{3}{4w^2}-\frac{15}{8w^3}-\ldots. \tag{73}$$

After some algebraic manipulations,

$$\Pi=\Pi_a+\epsilon\Pi_s$$

$$\frac{\Pi_a}{c}\cong\frac{e^{-i\beta r_1}}{r_1}+\frac{C-\Delta}{C+\Delta}\,\frac{e^{-i\beta r_2}}{r_2}+\left\{-\frac{1}{p(1+C/\Delta)^3}-\frac{1\cdot 3}{2p^2(1+C/\Delta)^5}-\frac{1\cdot 3\cdot 5}{4p^3(1+C/\Delta)^7}\cdots\right\}\frac{e^{-i\beta r_2}}{r_2} \tag{74}$$

and

$$\frac{\Pi_s}{c}\cong\frac{-2\Delta}{\Delta+C}\,[2i\sqrt{\pi w}\,e^{-w}]\,\frac{e^{-i\beta r_2}}{r_2}. \tag{75}$$

The first term in the expression for Π_a is, of course, the direct or primary influence. The second term is the secondary influence modified by a Fresnel reflection coefficient. The terms containing inverse powers of p can be considered as corrections to geometrical optics. The surface-wave term is absent when b is negative. When b is positive, it is a finite contribution to the total field Π, and for a highly inductive surface (i. e., $b=90°$) it is predominant.

The vertical electric field is simply given by

$$E_z\cong\beta^2S^2\Pi=\beta^2(1-C^2)\Pi \qquad \text{for} \qquad \beta r_1 \text{ and } \beta r_2>>1,$$

and furthermore, $r_2\cong r_1-2hC$.

6. Conclusion

It has been shown that the nature of the field radiated from a dipole over a flat surface depends markedly on the complex value of the surface impedance Z. For example, when the phase of Z is between 0 and 45°, the derived expressions for the field correspond to those of Norton and others for a dipole over a homogeneous flat ground. When the phase of Z exceeds 45° the radiated field becomes trapped by the surface. This effect is most pronounced when Z is purely imaginary, corresponding to an inductive surface. A similar unifying treatment for a dipole over a curved spherical surface will be given in a sequel to this paper.

The author thanks Kenneth A. Norton for his very helpful suggestions and Howard Bussey for his constructive criticism.

7. References

[1] J. Zenneck, Uber die Fortpfanzung ebener electromagnetischer wellen einer ebenen Leiterflache und ihre beziehung zur drahtlosen telegraphie, Ann. Physik [4] **23,** 846 (1907).

[2] A. Sommerfeld, Uber die Ausbreitung der Wellen in der Drahtlosen Telegraphie, Ann. Physik [4] **28,** 665 (1909); **62,** 95 (1920); **81,** 1135 (1926).

[3] K. A. Norton, Propagation of radio waves over a plane earth, Nature **135,** 954 (1935).

[4] B. Rolf, Graphs to Prof. Sommerfeld's attenuation formula for radio waves, Proc. Inst. Radio Engrs. **18,** 391 (1930).

[5] C. J. Bouwkamp, On Sommerfeld's surface wave, Phys. Rev. **80,** 294 (1950).

[6] K. A. Norton, The calculation of ground-wave field intensity over a finitely conducting spherical earth, Proc. Inst. Radio Engrs. **29,** 623 (1941).

[7] K. A. Norton, The propagation of radio waves over the surface of the earth and in the upper atmosphere, I, Proc. Inst. Radio Engrs. **24,** 1367 (1936); II, **25,** 1203 (1937).

[8] A. N. Sommerfeld, Uber di Fortpfanzung electrodynamischer Wellen langs eines Drahtes, Ann. Physik u. Chem. **67,** 233 (1899).

[9] G. Goubau, Surface waves and their application to transmission lines, J. Appl. Phys. **21,** 1119 (1950).

[10] G. Goubau, On the excitation of surface waves, Proc. Inst. Radio Engrs. **40,** 865 (1952).

[11] S. Atwood, Surface wave propagation over a coated plane conductor, J. Appl. Phys. **22,** 504 (1951).

[12] W. Rotman, A study of single surface corrugated guides, Proc. Inst. Radio Engrs. **39,** 952 (1951).

[13] R. S. Elliott, On the theory of corrugated plane surfaces, Trans. Inst. Radio Engrs. AP–**2,** 71 (1954).

[14] F. J. Zucker, The guiding and radiation of surface waves, Proc. Symposium on Modern Advances in Microwave Techniques, p. 403 (Polytechnic Inst. Brooklyn, Nov. 8–10, 1954) (many references are given in this comprehensive paper).

[15] C. T. Tai, The effect of a grounded slab on radiation from a line source, J. Appl. Phys. **22,** 405 (1951).

[16] D. B. Brick, The radiation of a Hertzian dipole over a coated conductor, Monograph 113, Inst. Elec. Engrs. (London), Dec. 15, 1954.

[17] A. L. Cullen, The excitation of plane surface waves, Proc. Inst. Elec. Engrs. **101,** Pt. IV, 225 (1954).

[18] H. E. M. Barlow and W. M. G. Fernando, An investigation of the properties of radial surface waves launched over flat reactive surfaces, Proc. Inst. Elec. Engrs. **103,** 307 (1956).

[19] H. E. M. Barlow and A. L. Cullen, Surface waves, Proc. Inst. Elec. Engrs. **100,** Pt. III, 321 (1953).

[20] N. Marcuvitz, Guided wave concept in electromagnetic theory, Research Rept. R–269–52 (P. I. B.–208, ONR, May 14, 1952, Polytechnic Inst. of Brooklyn, Microwave Research Inst.).

[21] J. R. Wait, Radiation from a vertical dipole over a stratified ground, I, Trans. Inst. Radio Engrs. AP–**1,** 9 (1953); II, AP–**2,** 144 (1954).

[22] J. R. Wait, Radiation from a vertical antenna over a curved stratified ground, J. Research NBS **56,** 232 (1956) RP2671.

[23] A. N. Sommerfeld, Partial differential equations (Academic Press, Inc., New York, N. Y., 1949).

[24] B. L. Van der Waerden, On the method of saddle points, Appl. Sci. Research B–**2,** 33 (1950).

[25] P. C. Clemmow, Some extensions to the method of integration by steepest descents, Quart. J. Mech. Appl. Math. **3,** 241 (1950).

[26] W. H. Wise, The physical reality of the Zenneck surface wave, Bell System Tech. J. **16,** 35 (1937).

[27] S. O. Rice, Series for the wave function of a radiating dipole at the earth's surface, Bell System Tech. J. **16,** 101 (1937).

Boulder, Colo., April 19, 1957.