Diffraction Theory of Electromagnetic Waves

J. A. STRATTON AND L. J. CHU
Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received May 15, 1939)

It has been shown by Larmor, Kottler and others that the classical method of calculating diffraction from the Kirchhoff formula in terms of a scalar light function cannot be applied directly to the electromagnetic field since it takes into account neither the vector character of the field nor the effect of charges along the contour of the opening. The field equations are integrated directly by means of a vector analog of Green's theorem. The results are applied to the calculation of diffraction of electromagnetic waves from a rectangular slit in a screen of infinite conductivity. The results are compared with an exact solution obtained recently by Morse and Rubenstein.

1. INTRODUCTION

RECENT advances in the technique of generating ultra-high frequency radio waves have stimulated interest in a number of problems of electromagnetic theory which heretofore have had only academic importance. A natural consequence of this trend towards short waves is an application of the methods of physical optics to determine the intensity and distribution of radiation from hollow tubes, horns, or small openings in cavity resonators. Now it is well known that the application of the Kirchhoff diffraction formula to an opening in an opaque screen involves several fundamental errors of principle. Nevertheless a remarkably good agreement between measured and calculated intensities is obtained in the region directly in front of the opening provided the wave-length is small compared to the size of the opening. If, on the other hand, the wave-length is relatively large, the radiation is distributed over a wide angle and the Kirchhoff formula proves to be definitely inaccurate. It must then be extended to account first for the vector character of the wave, and secondly for the discontinuities introduced about the contour of the opening. Both of these factors as sources of error have been recognized by previous writers. In the present paper we shall recall first the restrictions on the Kirchhoff formula when applied to a scalar wave function. A vector analog of Green's theorem will then be derived, from which an integral representation of electromagnetic fields can be obtained in a very simple manner. Next the theory is extended to surfaces over which the field vectors are discontinuous, and the results are finally compared to those obtained from the solution of a boundary value problem by rigorous methods.

2. THE KIRCHHOFF FORMULA FOR SCALAR WAVE FUNCTIONS

Let $\varphi(x, y, z)$ be any solution of

$$\nabla^2\varphi+k^2\varphi=0, \tag{1}$$

which is continuous and has continuous first derivatives throughout a closed domain V, and let φ_S, $(\partial\varphi/\partial n)_S$ denote the values of φ and its normal derivative on the bounding surface S. Then if the direction of the normal is outward from S, the integral

$$u(x', y', z') = \frac{1}{4\pi}\int_S\left[\left(\frac{\partial\varphi}{\partial n}\right)_S \frac{e^{ikr}}{r} - \varphi_S\frac{\partial}{\partial n}\left(\frac{e^{ikr}}{r}\right)\right]da \tag{2}$$

represents a discontinuous function which at all interior points x', y', z' is equal to φ and at all external points is zero. Since the values of φ and its derivatives at all interior points are uniquely determined by the value of φ_S alone on S (Dirichlet Problem), or by $(\partial\varphi/\partial n)_S$ alone (Neumann Problem), the functions φ_S and $(\partial\varphi/\partial n)_S$ cannot be chosen independently if φ and u are to be identical at interior points. The function defined by (2) satisfies (1) and is regular within the domain V whatever the choice of φ_S and $(\partial\varphi/\partial n)_S$, but the values assumed by u and $\partial u/\partial n$ on S will in general differ from those assigned to φ_S and $(\partial\varphi/\partial n)_S$.

A Kirchhoff diffraction problem is formulated usually somewhat as follows. A primary wave is incident upon an opaque screen in which there is an opening S_1. The scalar potential of the wave, or a rectangular component of a field vector, is represented by φ satisfying (1). It is now assumed that on the dark side of the screen φ_S and $(\partial\varphi/\partial n)_S$ are zero except over the opening S_1, where φ_S and $(\partial\varphi/\partial n)_S$ have their undisturbed values. The diffracted wave is then calculated from (2), the integral extending over S_1 alone.

Such a procedure cannot possibly give an exact result. For in the first place the assumption of zero values for the light function and its derivative over the screen S_2 implies a discontinuity about the contour C_1 bounding the opening, while Green's theorem, upon which (2) is based, is valid only for functions which are continuous over a complete bounding surface. This difficulty cannot be obviated by the common expedient of replacing the contour of discontinuity by a thin region of rapid but continuous transition. The vanishing of φ_S and $(\partial\varphi/\partial n)_S$ on any finite part of S would then entail a zero value everywhere. The integral represents in fact a scalar wave function which approximates the true intensity to a degree which must be determined by other means. If the ratio of wave-length to size of S_1 is small, the radiation is thrown largely forward and the Kirchhoff function may differ by a negligible amount from the assumed zero value over the screen. If, on the other hand, the wave-length is large, the wave function calculated on the basis of undisturbed values over S_1 will be found to have values over S_2 which are by no means small. It is sometimes suggested[1] that this distribution over S_2 be applied once more to the Kirchhoff formula (2)—a method of successive approximations. Quite apart from questions of convergence, the difficulties of evaluating the resulting surface integrals in most cases make such a procedure of little practical value.

There are other difficulties. An electromagnetic field at a point on the closed surface S is characterized by a set of scalar functions which represent the rectangular components of the electric and magnetic vectors. Each one of these scalar functions satisfies (1) and its value at an interior point x', y', z' is therefore expressed by (2) in terms of its values over the boundary S. But these components at an interior point must not only satisfy the wave equation, *they must also be solutions of the Maxwell field equations.* The real problem, therefore, is not the integration of a scalar wave equation, nor even a vector wave equation, but of a simultaneous system of first-order vector equations relating the vectors E and H. There is nothing new in this remark. The integration of the field equations in closed form has been discussed by Love,[2] Larmor,[3] v. Ignatowsky,[4] Tonolo,[5] Macdonald,[6] Tedone,[7] and most completely by Kottler,[8] but their results appear to have been commonly disregarded in treatises on physical optics. Recently, however, the subject has been reviewed by Schelkunoff[9] in connection with equivalence theorems.

3. A Vector Analog of Green's Theorem

The integration of the field equations can be achieved most directly and rigorously by a method which is wholly analogous to the treatment of the scalar wave equation. Let V be a closed region of space bounded by a regular surface S, and let $\mathbf{P}$ and $\mathbf{Q}$ be two functions of position which together with their first and second derivatives are continuous throughout V and on the surface S. The divergence theorem is applied to the vector $\mathbf{P}\times\nabla\times\mathbf{Q}$, giving

$$\int_V \nabla\cdot(\mathbf{P}\times\nabla\times\mathbf{Q})dv=\int_S(\mathbf{P}\times\nabla\times\mathbf{Q})\cdot\mathbf{n}da, \quad (3)$$

where $\mathbf{n}$ is a unit normal vector directed outward from S. Upon expanding the integrand of the volume integral a vector analog of Green's first identity is obtained,

[1] M. Born, *Optik*, p. 152.

[2] A. E. H. Love, Phil. Trans. **A197**, 1 (1901).

[3] J. Larmor, Lond. Math. Soc. Proc. **1**, 1 (1903).

[4] W. v. Ignatowsky, Ann. d. Physik **23**, 875 (1907); **25**, 99 (1908).

[5] A. Tonolo, Annali di Matematica **3**, 17, 29 (1910).

[6] H. M. Macdonald, Proc. Lond. Math. Soc. **10**, 91 (1911); and Phil. Trans. **A212**, 295 (1912).

[7] O. Tedone, Linc. Rendi (5), **1**, 286 (1917).

[8] F. Kottler, Ann. d. Physik **71**, 457 (1923).

[9] S. A. Schelkunoff, Bell Sys. Tech. J. **15**, 92 (1936).

$$\int_V (\nabla\times\mathbf{P}\cdot\nabla\times\mathbf{Q}-\mathbf{P}\cdot\nabla\times\nabla\times\mathbf{Q})dv = \int_S (\mathbf{P}\times\nabla\times\mathbf{Q})\cdot\mathbf{n}da. \quad (4)$$

The vector analog of Green's second identity ("Green's theorem") is obtained by reversing the roles of $\mathbf{P}$ and $\mathbf{Q}$ in (4) and subtracting one expression from the other.

$$\int_V (\mathbf{Q}\cdot\nabla\times\nabla\times\mathbf{P}-\mathbf{P}\cdot\nabla\times\nabla\times\mathbf{Q})dv = \int_S (\mathbf{P}\times\nabla\times\mathbf{Q}-\mathbf{Q}\times\nabla\times\mathbf{P})\cdot\mathbf{n}da. \quad (5)$$

The first identity (4) is the logical basis for uniqueness proofs in connection with vector fields. It will be noted that if one place $\mathbf{P}=\mathbf{Q}=\mathbf{E}$, (4) proves to be identical with Poynting's theorem.

4. Application to the Field Equations

The field equations in a rationalized m.k.s. system of units are

$$\text{(I)}\quad \nabla\times\mathbf{E}-i\omega\mu\mathbf{H}=-\mathbf{J}^*, \qquad \text{(III)}\quad \nabla\cdot\mathbf{H}=\rho^*/\mu,$$

$$\text{(II)}\quad \nabla\times\mathbf{H}+i\omega\epsilon\mathbf{E}=\mathbf{J}, \qquad \text{(IV)}\quad \nabla\cdot\mathbf{E}=\rho/\epsilon.$$

It is assumed here that the fields are harmonic and that all quantities contain the time in the form of a factor $\exp(-i\omega t)$.

The field intensity $\mathbf{E}$ is measured in volts per meter, $\mathbf{H}$ in ampere-turns per meter, current density $\mathbf{J}$ in amperes per square meter. In free space $\mu_0=4\pi\times10^{-7}$ henry per meter, $\epsilon_0=(1/36\pi)\times10^{-9}$ farad per meter. The medium is assumed to be homogeneous and isotropic, and of zero conductivity. The quantities $\mathbf{J}^*$ and ρ^* are fictitious densities of magnetic current and magnetic charge. In normal fields they are zero, but the arbitrary (and physically impossible) assumption of discontinuities in the tangential components of $\mathbf{E}$ about the contour of an opening can only be accounted for by some such assumption. Currents and charges are related by the equations of continuity,

$$\text{(V)}\qquad \nabla\cdot\mathbf{J}-i\omega\rho=0, \qquad \nabla\cdot\mathbf{J}^*-i\omega\rho^*=0.$$

The vectors $\mathbf{E}$ and $\mathbf{H}$ satisfy

$$\nabla\times\nabla\times\mathbf{E}-k^2\mathbf{E}=i\omega\mu\mathbf{J}-\nabla\times\mathbf{J}^*, \quad (6)$$

$$\nabla\times\nabla\times\mathbf{H}-k^2\mathbf{H}=i\omega\epsilon\mathbf{J}^*+\nabla\times\mathbf{J}, \quad (7)$$

where $k^2=\omega^2\epsilon\mu$. Solutions of (6) and (7) are to be found which are finite and single-valued at all interior points of V and at all points on S.

In (5) let $\mathbf{P}=\mathbf{E}$ and $\mathbf{Q}=\psi\mathbf{a}$, where $\mathbf{a}$ is a unit vector in an arbitrary direction and $\psi=e^{ikr}/r$. The Green's function $\mathbf{Q}$ is essentially the vector potential of a unit current element. Distance is measured from the element at x, y, z to the point of observation at x', y', z'.

$$r=[(x'-x)^2+(y'-y)^2+(z'-z)^2]^{\frac{1}{2}}. \quad (8)$$

The following identities are readily verified.

$$\nabla\times\mathbf{Q}=\nabla\psi\times\mathbf{a}, \qquad \nabla\times\nabla\times\mathbf{Q}=\mathbf{a}k^2\psi+\nabla(\mathbf{a}\cdot\nabla\psi),$$
$$\nabla\times\nabla\times\mathbf{P}=k^2\mathbf{E}+i\omega\mu\mathbf{J}-\nabla\times\mathbf{J}^*.$$

By application of the divergence theorem and further transformation, it is easily shown that $\mathbf{a}\cdot$ is a factor common to all the terms in (5), and since the direction of $\mathbf{a}$ is arbitrary, it follows that

$$\int_V \left[i\omega\mu\mathbf{J}\psi-\nabla\times\mathbf{J}^*\psi+\frac{1}{\epsilon}\rho\nabla\psi\right]dv = \int_S [i\omega\mu(\mathbf{n}\times\mathbf{H})\psi+(\mathbf{n}\times\mathbf{E})\times\nabla\psi + (\mathbf{n}\cdot\mathbf{E})\nabla\psi-\mathbf{n}\times\mathbf{J}^*\psi]da. \quad (9)$$

An application of the identity

$$\int_V \nabla\times\mathbf{J}^*\psi dv=\int_S \mathbf{n}\times\mathbf{J}^*\psi da+\int_V \mathbf{J}^*\times\nabla\psi dv \quad (10)$$

reduces this to

$$\int_V \left[i\omega\mu\mathbf{J}\psi-\mathbf{J}^*\times\nabla\psi+\frac{1}{\epsilon}\rho\nabla\psi\right]dv = \int_S [i\omega\mu(\mathbf{n}\times\mathbf{H})+(\mathbf{n}\times\mathbf{E})\times\nabla\psi + (\mathbf{n}\cdot\mathbf{E})\nabla\psi]da. \quad (11)$$

Because of the singularity of the function ψ at $r=0$, the identity (11) holds only when this point is excluded from V. The procedure is familiar to everyone acquainted with potential theory. The point x', y', z' is taken as the center of a small sphere of radius r_1. The normal over the sphere is directed out of V, and consequently radially *towards* the center.

$$\nabla\psi = (1/r - ik)(e^{ikr}/r)\mathbf{r}_0, \tag{12}$$

and on the sphere $\mathbf{n} = \mathbf{r}_0$. The area of the sphere vanishes with the radius as $4\pi r_1^2$, and since

$$(\mathbf{n}\times\mathbf{E})\times\mathbf{n} + (\mathbf{n}\cdot\mathbf{E})\mathbf{n} = \mathbf{E}, \tag{13}$$

the contribution of the spherical surface to the right-hand side of (11) reduces to $4\pi\mathbf{E}(x', y', z')$. The value $\mathbf{E}$ at any internal point of V is therefore

$$\begin{aligned}\mathbf{E}(x', y', z') &= \frac{1}{4\pi}\int_V \left[i\omega\mu\mathbf{J}\psi - \mathbf{J}^*\times\nabla\psi + \frac{1}{\epsilon}\rho\nabla\psi \right] dv \\ &\quad - \frac{1}{4\pi}\int_S [i\omega\mu(\mathbf{n}\times\mathbf{H})\psi \\ &\qquad + (\mathbf{n}\times\mathbf{E})\times\nabla\psi + (\mathbf{n}\cdot\mathbf{E})\nabla\psi] da. \end{aligned}\tag{14}$$

An appropriate interchange of vectors gives

$$\begin{aligned}\mathbf{H}(x', y', z') &= \frac{1}{4\pi}\int_V \left[i\omega\epsilon\mathbf{J}^*\psi + \mathbf{J}\times\nabla\psi + \frac{1}{\mu}\rho^*\nabla\psi \right] dv \\ &\quad + \frac{1}{4\pi}\int_S [i\omega\epsilon(\mathbf{n}\times\mathbf{E})\psi - (\mathbf{n}\times\mathbf{H})\times\nabla\psi \\ &\qquad - (\mathbf{n}\cdot\mathbf{H})\nabla\psi] da. \end{aligned}\tag{15}$$

It will be shown below under more general circumstances that (14) and (15) satisfy the field equations at all points of V and S.

These expressions are essentially equivalent to those obtained by v. Ignatowsky for a closed surface. If all sources can be enclosed within a sphere of finite radius, the field is regular at infinity and either side of S may be chosen as its "interior," or S may be closed at infinity. It may be remarked that (14) and (15) are convenient expressions for the calculation of a field directly from a given distribution of current without the intervention of vector and scalar potentials or of a Hertzian vector. The surface S recedes to infinity and the fictitious magnetic sources are placed equal to zero. Then

$$\begin{aligned}\mathbf{E}(x', y', z') &= \frac{1}{4\pi}\int_V \left[i\omega\mu\mathbf{J}\psi + \frac{1}{\epsilon}\rho\nabla\psi \right] dv, \\ \mathbf{H}(x', y', z') &= \frac{1}{4\pi}\int_V \mathbf{J}\times\nabla\psi dv. \end{aligned}\tag{16}$$

Since the current distribution is given, the charge density can be determined from the equation of continuity.

It is known from the uniqueness theorem of electromagnetic theory that a field within a bounded domain is completely determined by the specification of the tangential components of $\mathbf{E}$ and $\mathbf{H}$ on the surface. It follows in (14) and (15) that when $\mathbf{n}\times\mathbf{E}$ and $\mathbf{n}\times\mathbf{H}$ have been fixed, the choice of $\mathbf{n}\cdot\mathbf{E}$ and $\mathbf{n}\cdot\mathbf{H}$ is no longer arbitrary. The selection must be consistent with the conditions on a field satisfying Maxwell's equations. The same limitations on the choice of φ_S and $(\partial\varphi/\partial n)_S$ were pointed out in §2. The dependence of the normal component of $\mathbf{E}$ upon the tangential component of $\mathbf{H}$ is equivalent to that of ρ upon $\mathbf{J}$.

Let us suppose for the moment that the charge and current distributions in (14) are confined to a thin layer at the surface S. As the depth of the layer diminishes the densities may be increased so that in the limit the volume densities are replaced in the usual way by surface densities. If the region V contains no charge or current within its interior or on its boundary S, the field at an interior point is

$$\begin{aligned}\mathbf{E}(x', y', z') &= \frac{1}{4\pi}\int_S [i\omega\mu(\mathbf{n}\times\mathbf{H})\psi \\ &\qquad + (\mathbf{n}\times\mathbf{E})\times\nabla\psi + (\mathbf{n}\cdot\mathbf{E})\nabla\psi] da. \end{aligned}\tag{17}$$

It is now clear that this is exactly the field that would be produced by a distribution of electric current over S with surface density $\mathbf{K}$, a distribution of magnetic current of density $\mathbf{K}^*$, and a surface electric charge of density η, where

$$\mathbf{K} = -\mathbf{n}\times\mathbf{H}, \qquad \mathbf{K}^* = \mathbf{n}\times\mathbf{E}, \qquad \eta = -\epsilon\mathbf{n}\cdot\mathbf{E}. \tag{18}$$

The values of $\mathbf{E}$ and $\mathbf{H}$ in (18) are those just *inside* the surface S. The function $\mathbf{E}(x', y', z')$ defined by (17) is discontinuous across S. It is well known[10] that the integral

$$\mathbf{E}_3(x', y', z') = \frac{1}{4\pi\epsilon}\int_S \eta \nabla\psi da \tag{19}$$

suffers a discontinuity on transition through S given by $\mathbf{n}\cdot\Delta\mathbf{E}_3 = \eta/\epsilon$, where $\Delta\mathbf{E}_3$ is the difference of the values outside and inside. The third term of (17) does not affect the transition of the tangential component but reduces the normal component to zero. Likewise the discontinuity of

$$\mathbf{E}_2(x', y', z') = \frac{1}{4\pi}\int_S \mathbf{K}^*\times\nabla\psi da \tag{20}$$

is specified by $\mathbf{n}\times\mathbf{E}_2 = \mathbf{K}^*$, so that the second term in (17) reduces the tangential component of $\mathbf{E}$ to zero without affecting the normal component. The first term in (17) is continuous across S, but has discontinuous derivatives.

$$\nabla'\times\mathbf{E}(x', y', z') = -\frac{i\omega\mu}{4\pi}\int_S (\mathbf{n}\times\mathbf{H})\times\nabla\psi da. \tag{21}$$

The vector $\mathbf{E}$ and the tangential component of its curl is zero on the positive side of S; it is therefore zero at all external points. The same analysis applies to $\mathbf{H}$.

5. Extension to Discontinuous Surface Distributions

The results of the preceding section hold only if the vectors $\mathbf{E}$ and $\mathbf{H}$ are continuous and have continuous first derivatives at all points of S. They cannot, therefore, be applied directly to the problem of diffraction at a slit. To obtain the required extension of (17) to such cases, consider the closed surface S (surfaces closed at infinity are included) to be divided into two zones S_1 and S_2 by a closed contour C lying on S, as in Fig. 1. The vectors $\mathbf{E}$ and $\mathbf{H}$ and their first derivatives are continuous over S_1 and satisfy the field equations. The same is true over S_2. However the components of $\mathbf{E}$ and $\mathbf{H}$ which are tangential to the surface are now subject to a discontinuous change in passing across C from one zone to the other. The occurrence of such discontinuities can be reconciled with the field equations only by the further assumption of a line distribution of charges or currents about the contour C. This line distribution of sources contributes to the field and only when it is taken into account do the resultant expressions for $\mathbf{E}$ and $\mathbf{H}$ satisfy Maxwell's equations. The calculation of a diffraction pattern from an integral extended over a portion of a surface only, as has been the customary practice, must necessarily lead to erroneous results.

A method of determining a contour distribution consistent with the requirements of the problem was proposed by Kottler.[8] It was shown in §4 that the field at an interior point in (17) is identical with that produced by the surface currents and charges specified in (18). A discontinuity in the tangential components of $\mathbf{E}$ and $\mathbf{H}$ in passing on the surface from zone S_1 to zone S_2 implies therefore an abrupt change in the surface current density. The termination of a line of current, in turn, can be accounted for on the basis of the equation of continuity by an accumulation of charge on the contour. Let $d\mathbf{s}$ be an element of length along the contour in the positive direction as determined by the positive normal $\mathbf{n}$, Fig. 1. Let $\mathbf{n}_1$ be a unit vector lying in the surface, normal to both $\mathbf{n}$ and the contour element $d\mathbf{s}$, and directed into zone (1). The line densities of electric and magnetic charge will be designated by σ and σ^*. Then Eqs. (V), when applied to surface currents, become

$$\mathbf{n}_1\cdot(\mathbf{K}_1-\mathbf{K}_2) = i\omega\sigma, \quad \mathbf{n}_1\cdot(\mathbf{K}_1^*-\mathbf{K}_2^*) = i\omega\sigma^*, \tag{22}$$

and hence by (18),

$$\begin{aligned} i\omega\sigma &= \mathbf{n}_1\cdot(\mathbf{n}\times\mathbf{H}_2 - \mathbf{n}\times\mathbf{H}_1) \\ &= (\mathbf{H}_2-\mathbf{H}_1)\cdot(\mathbf{n}_1\times\mathbf{n}), \\ i\omega\sigma^* &= \mathbf{n}_1\cdot(\mathbf{n}\times\mathbf{E}_1 - \mathbf{n}\times\mathbf{E}_2) \\ &= -(\mathbf{E}_2-\mathbf{E}_1)\cdot(\mathbf{n}_1\times\mathbf{n}). \end{aligned} \tag{23}$$

The vector $\mathbf{n}_1\times\mathbf{n}$ is in the direction of $d\mathbf{s}$. If S_2 represents an opaque screen over which $\mathbf{E}_2 = \mathbf{H}_2 = 0$, the field at any point on the shadow side is

$$\begin{aligned} \mathbf{E}(x', y', z') = &-\frac{1}{i\omega\epsilon}\frac{1}{4\pi}\oint_C \nabla\psi\mathbf{H}_1\cdot d\mathbf{s} \\ &-\frac{1}{4\pi}\int_{S1}[i\omega\mu(\mathbf{n}\times\mathbf{H}_1)\psi + (\mathbf{n}\times\mathbf{E}_1)\times\nabla\psi \\ &+(\mathbf{n}\cdot\mathbf{E}_1)\nabla\psi]da, \end{aligned} \tag{24}$$

[10] H. B. Phillips, *Vector Analysis* (Wiley and Sons, 1933), p. 206. In greater detail, H.Poincaré, *Théorie Mathématique de la Lumière*, II, Ch. VII.

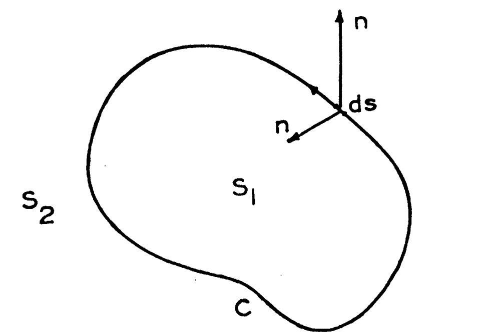

FIG. 1. The contour on which the fields and their derivatives are discontinuous. The lower **n** lying within the zone S_1 should be $\mathbf{n}_1$.

which can be shown to be identical with Kottler's result,

$$4\pi\mathbf{E}(x', y', z') = -\frac{1}{i\omega\epsilon}\oint_C \nabla\psi\mathbf{H}_1\cdot d\mathbf{s} + \oint_C \psi\mathbf{E}_1\times d\mathbf{s} - \int_{S_1}\left(\mathbf{E}_1\frac{\partial\psi}{\partial n} - \psi\frac{\partial\mathbf{E}_1}{\partial n}\right)da. \quad (25)$$

For the magnetic field one obtains

$$\begin{aligned}4\pi\mathbf{H}(x', y', z') &= \frac{1}{i\omega\mu}\oint_C \nabla\psi\mathbf{E}_1\cdot d\mathbf{s} + \int_{S_1}[i\omega\epsilon(\mathbf{n}\times\mathbf{E}_1)\psi \\ &\quad - (\mathbf{n}\times\mathbf{H}_1)\times\nabla\psi - (\mathbf{n}\cdot\mathbf{H}_1)\nabla\psi]da \\ &= \frac{1}{i\omega\mu}\oint_C \nabla\psi\mathbf{E}_1\cdot d\mathbf{s} + \oint_C \psi\mathbf{H}_1\times d\mathbf{s} \\ &\quad - \int_{S_1}\left(\mathbf{H}_1\frac{\partial\psi}{\partial n} - \psi\frac{\partial\mathbf{H}_1}{\partial n}\right)da. \quad (26)\end{aligned}$$

It remains to be shown that the fields expressed by these integrals are in fact divergenceless and satisfy (I) and (II). Consider first the divergence of (24) at a point x', y', z'.

$$\begin{aligned}\nabla'\cdot E(x', y', z') &= \frac{1}{i\omega\epsilon}\frac{1}{4\pi}\oint_C \nabla^2\psi\mathbf{H}_1\cdot d\mathbf{s} + \frac{1}{4\pi}\int_S [i\omega\mu(\mathbf{n}\times\mathbf{H}_1)\cdot\nabla\psi \\ &\quad + (\mathbf{n}\cdot\mathbf{E}_1)\nabla^2\psi]da \\ &= -\frac{k^2}{i\omega\epsilon}\frac{1}{4\pi}\oint_C \psi\mathbf{H}_1\cdot d\mathbf{s} - \frac{k^2}{4\pi}\int_{S_1}(\mathbf{n}\cdot\mathbf{E}_1)\psi da \\ &\quad + \frac{i\omega\mu}{4\pi}\int_{S_1}(\mathbf{n}\times\mathbf{H}_1)\cdot\nabla\psi da, \quad (27)\end{aligned}$$

taking into account the relation $\nabla' = -\nabla$ when applied to ψ or its derivatives. Now

$$\int_S (\mathbf{n}\times\mathbf{H})\cdot\nabla\psi da = \int_S \psi\nabla\times\mathbf{H}\cdot\mathbf{n}da - \oint_C \psi\mathbf{H}\cdot d\mathbf{s}. \quad (28)$$

The line integral resulting from this transformation is zero when S is closed, but otherwise just cancels the contour integral in (27). Inversely, only the presence of the contour integral leads to a zero divergence and transverse waves at great distances from the opening S_1. From (II) and the relation $k^2/i\omega\epsilon = -i\omega\mu$ follows immediately the result

$$\nabla'\cdot\mathbf{E}(x', y', z') = 0. \quad (29)$$

An identical proof holds for $\mathbf{H}(x', y', z')$.

Finally it will be shown that (24) and (26) satisfy (I) and (II).

$$\begin{aligned}\nabla'\times\mathbf{H}(x', y', z') &= -\frac{1}{4\pi}\int_{S_1}[(\mathbf{n}\times\mathbf{H}_1)\cdot\nabla\nabla\psi \\ &\quad + k^2\psi(\mathbf{n}\times\mathbf{H}_1) - i\omega\epsilon(\mathbf{n}\times\mathbf{E}_1)\times\nabla\psi]da, \quad (30)\end{aligned}$$

since the curl of the gradient is identically zero. Furthermore

$$\begin{aligned}\int_{S_1}(\mathbf{n}\times\mathbf{H}_1)\cdot\nabla\nabla\psi da &= -\int_{S_1}(\mathbf{n}\cdot\nabla_{\nabla\psi}\times\mathbf{H}_1)\nabla\psi da \\ &= \int_{S_1}(\mathbf{n}\cdot\nabla\times\mathbf{H}_1)\nabla\psi da - \int_{S_1}\mathbf{n}\times\nabla\cdot(\mathbf{H}_1\nabla\psi)da \\ &= -\oint_C \nabla\psi\mathbf{H}_1\cdot d\mathbf{s} - i\omega\epsilon\int_{S_1}(\mathbf{n}\cdot\mathbf{E}_1)\nabla\psi da, \quad (31)\end{aligned}$$

where the operator $\nabla_{\nabla\psi}$ acts on $\nabla\psi$ only. The last integral takes account of the fact that the field equations are by hypothesis satisfied on S_1. Then

$$\begin{aligned}\nabla'\times\mathbf{H}(x', y', z') &= \frac{1}{4\pi}\oint_C \nabla\psi\mathbf{H}_1\cdot d\mathbf{s} + \frac{i\omega\epsilon}{4\pi}\int_{S_1}[i\omega\mu(\mathbf{n}\times\mathbf{H}_1)\psi \\ &\quad + (\mathbf{n}\times\mathbf{E}_1)\times\nabla\psi + (\mathbf{n}\cdot\mathbf{E}_1)\nabla\psi]da \\ &= -i\omega\epsilon\mathbf{E}(x', y', z'). \quad (32)\end{aligned}$$

The validity of (I) is established in the same manner.

6. Application to a Rectangular Slit

In Fig. 2 a rectangular slit is shown in an infinite, perfectly conducting screen coinciding with the xy plane. Plane waves coming from the left are incident on this slit at various angles. The n rmal to the slit surface is drawn to the left and the intensity of the field is to be calculated at any point whose coordinates are R', θ', φ' lying to the right of the screen. Then at sufficiently large distances

$$r \simeq R' - x\cos\varphi'\sin\theta' - y\sin\varphi'\sin\theta'. \quad (33)$$

Consider first the case of a normal incidence, with the electric vector polarized along the x axis and of unit intensity. The field at great distances from the slit obtained by evaluating (25) is

$$\begin{aligned} 4\pi E_{\theta'}' &= -ik(1+\cos\theta')\cos\varphi' A, \\ 4\pi E_{\varphi'}' &= ik(1+\cos\theta')\sin\varphi' A, \\ E_{R'}' &= 0, \end{aligned} \quad (34)$$

where

$$A = 4\frac{\sin(\tfrac{1}{2}ka\cos\varphi'\sin\theta')}{k\cos\varphi'\sin\theta'} \times \frac{\sin(\tfrac{1}{2}kb\sin\varphi'\sin\theta')}{k\sin\varphi'\sin\theta'}\frac{e^{ikR'}}{R'}. \quad (35)$$

Now if this solution is extended analytically towards the screen, it is evident that it does not vanish on the plane $z=0$, as required by the boundary conditions. In fact nothing has been stated as to the location of the surface S_2 which closes the slit surface S_1, and one is free to choose it in the way which is least liable to violate the actual conditions. Thus, in the present case S_1 may be closed by a surface S_2 which lies just behind S_1 and over which it is assumed that the field vectors vanish. This is equivalent to saying that the field functions (34) can be continued analytically into the region for which $\theta' > \pi/2$. The effect of the screen or baffle is now taken into account by assuming it to act as a perfect reflector. The phase relations after reflection are determined by the condition that the resultant tangential component of $\mathbf{E}$ must be zero at $\theta' = \pi/2$, while the magnitude of the initial normal component is doubled. In the present instance these conditions are expressed by

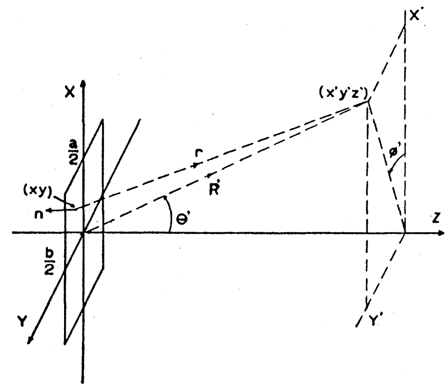

Fig. 2. Coordinate system for calculating the diffraction by a rectangular slit.

$$\begin{aligned} E_{\theta'} &= E_{\theta'}'(\theta') + E_{\theta'}'(\pi-\theta'), \\ E_{\varphi'} &= E_{\varphi'}'(\theta') - E_{\varphi'}'(\pi-\theta'), \end{aligned} \quad (36)$$

where $E_{\theta'}$ and $E_{\varphi'}$ are normal and tangential components of the *resultant* field. Applied to (34) this gives

$$\begin{aligned} 4\pi E_{\theta'} &= -2ik\cos\varphi' A, \\ 4\pi E_{\varphi'} &= 2ik\cos\theta'\sin\varphi' A. \end{aligned} \quad (37)$$

The energy flow or intensity of the diffracted wave is proportional to the sum of the squares of these quantities. The solid curves of Fig. 3 are plots of the intensity in the vertical plane $\varphi'=0$ for several ratios of slit breadth to wavelength. Similar curves are drawn in Fig. 6 for the horizontal plane $\varphi'=\pi/2$.

In the more general case the direction of propagation of the incident wave makes an arbitrary angle α with the z axis. If the polarization is parallel to the x axis, the components of the incident wave are

$$E_x = e^{ik(y\sin\alpha + z\cos\alpha)}, \quad H_y = (\epsilon/\mu)^{\frac{1}{2}}\cos\alpha E_x, \quad (38)$$
$$H_z = -(\epsilon/\mu)^{\frac{1}{2}}\sin\alpha E_x.$$

The diffracted field is found from (25) to be

$$\begin{aligned} 4\pi E_{\theta'}' &= -ik\cos\varphi'(1+\cos\theta'\cos\alpha)A, \\ & \qquad E_{R'}' = 0 \\ 4\pi E_{\varphi'}' &= ik\sin\varphi'(\cos\theta'+\cos\alpha)A, \end{aligned} \quad (39)$$

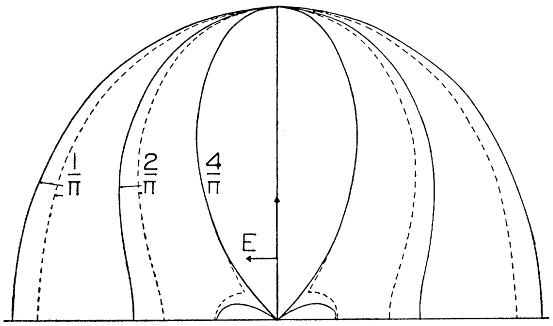

FIG. 3.

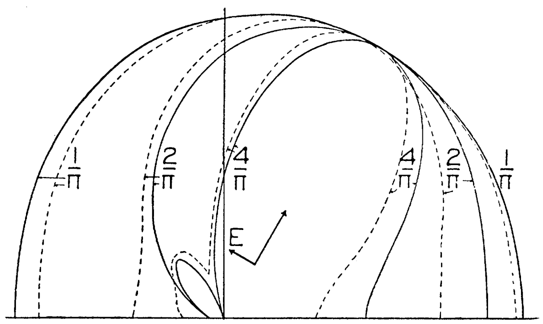

FIG. 4.

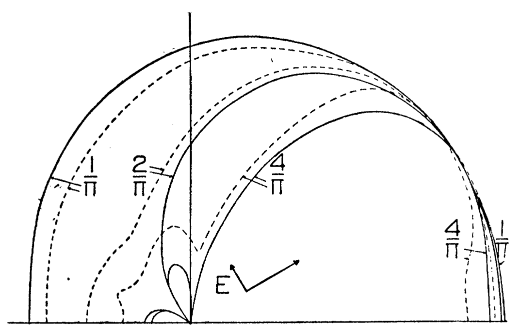

FIG. 5.

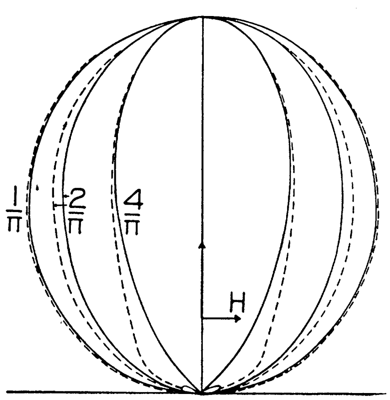

FIG. 6.

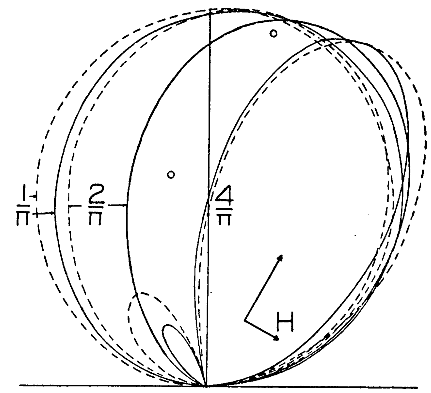

FIG. 7.

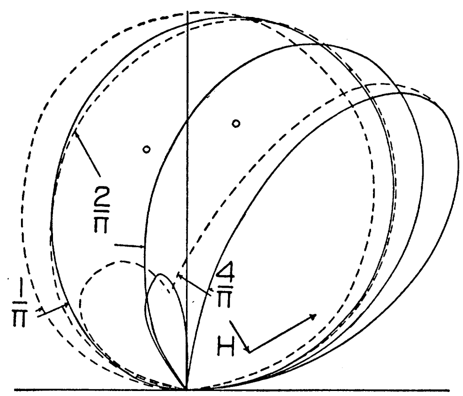

FIG. 8.

FIGS. 3–8. Polar diagrams of the distribution of field intensity, showing the diffraction of a plane wave by a slit in the plane of incidence. The long arrow indicates the direction of incidence and the short arrow indicates the polarization of the field at the slit. The numerals on the curves indicate the width of the slit in wave-lengths (a/λ or b/λ).

and the resultant field to the right of the plane screen obtained by adding the reflected wave is identical with (37), but with α now entering into A in the form

$$A=4\frac{\sin\left(\frac{1}{2}ka\cos\varphi'\sin\theta'\right)}{k\cos\varphi'\sin\theta'}\times\frac{\sin\frac{1}{2}kb(\sin\varphi'\sin\theta'-\sin\alpha)}{k(\sin\varphi'\sin\theta'-\sin\alpha)}\frac{e^{ikR'}}{R'}. \quad (40)$$

In Figs. 7 and 8 the solid curves represent plots of intensity in the horizontal yz plane for $\alpha=30°$ and $\alpha=60°$. The distribution in the vertical plane is identical with Fig. 3.

In the alternate case the magnetic vector is polarized vertically along the x axis and the incident wave is defined by

$$H_x = e^{ik(y \sin \alpha + z \cos \alpha)}, \quad E_y = -(\mu/\epsilon)^{\frac{1}{2}} \cos \alpha H_x, \qquad E_z = (\mu/\epsilon)^{\frac{1}{2}} \sin \alpha H_x. \tag{41}$$

The diffracted field calculated from (25) is then

$$4\pi E_{\theta'}' = i\omega\mu \sin \varphi'(\cos \theta' + \cos \alpha)A, \qquad 4\pi E_{\varphi'}' = i\omega\mu \cos \varphi'(1 + \cos \theta \cos \alpha)A, \tag{42}$$

and the resultant field after reflection is

$$4\pi E_{\theta'} = 2i\omega\mu \sin \varphi' \cos \alpha A, \qquad 4\pi E_{\varphi'} = 2i\omega\mu \cos \varphi' \cos \theta' \cos \alpha\, A, \tag{43}$$

where A is defined by (40). The solid curves of Figs. 4 and 5 represent plots of intensity in the horizontal plane. The distribution in the vertical plane is identical with Fig. 6.

It is exceedingly interesting to compare these results with some calculations made recently by Morse and Rubenstein,[11] who have carried through the two-dimensional problem of diffraction of an electromagnetic plane wave by an infinite slit. The two methods should lead to approximately the same distribution in an equatorial plane. The intensity plots obtained from this rigorous solution are shown as dotted curves. The correspondence on the whole is remarkably good. It will be noted that the most marked deviation occurs in the immediate neighborhood of the screen and probably arises from the errors which are fundamental to the present method: the assumption of unperturbed distributions over the slit and the manner in which the reflection problem has been handled. The authors have discarded two points in Figs. 7 and 8 which fall on the curves published by Morse and Rubenstein[11] and which lead to lobes for which it is difficult to account. This neglect may or may not be justified. Professor Morse has kindly put the original data at the disposition of the authors but insufficient calculations were made to settle the matter one way or the other. It is hard to reconcile the anomalous occurrence of these lobes with the close correspondence of all other points. Whatever the answer to this question, the results give strong support to the belief that Eqs. (25) and (26) can be applied to the calculation of diffracted radiation with assurance of reasonable accuracy. In the case of radiation from hollow tubes and horns, an extension to take account of the internal reflected wave is no doubt possible based on the methods employed in acoustics under similar circumstances.

[11] P. M. Morse and Pearl J. Rubenstein, Phys. Rev. **54**, 895 (1938).