Geometrical Theory of Diffraction*

JOSEPH B. KELLER
Institute of Mathematical Sciences, New York University, New York, New York
(Received September 13, 1961)

The geometrical theory of diffraction is an extension of geometrical optics which accounts for diffraction. It introduces diffracted rays in addition to the usual rays of geometrical optics. These rays are produced by incident rays which hit edges, corners, or vertices of boundary surfaces, or which graze such surfaces. Various laws of diffraction, analogous to the laws of reflection and refraction, are employed to characterize the diffracted rays. A modified form of Fermat's principle, equivalent to these laws, can also be used. Diffracted wave fronts are defined, which can be found by a Huygens wavelet construction. There is an associated phase or eikonal function which satisfies the eikonal equation. In addition complex or imaginary rays are introduced. A field is associated with each ray and the total field at a point is the sum of the fields on all rays through the point. The phase of the field on a ray is proportional to the optical length of the ray from some reference point. The amplitude varies in accordance with the principle of conservation of energy in a narrow tube of rays. The initial value of the field on a diffracted ray is determined from the incident field with the aid of an appropriate diffraction coefficient. These diffraction coefficients are determined from certain canonical problems. They all vanish as the wavelength tends to zero. The theory is applied to diffraction by an aperture in a thin screen diffraction by a disk, etc., to illustrate it. Agreement is shown between the predictions of the theory and various other theoretical analyses of some of these problems. Experimental confirmation of the theory is also presented. The mathematical justification of the theory on the basis of electromagnetic theory is described. Finally, the applicability of this theory, or a modification of it, to other branches of physics is explained.

1. INTRODUCTION

GEOMETRICAL optics, the oldest and most widely used theory of light propagation, fails to account for certain optical phenomena called diffraction. We shall describe an extension of geometrical optics which does account for these phenomena. It is called the geometrical theory of diffraction.[1,2] Like geometrical optics, it assumes that light travels along certain straight or curved lines called rays. But it introduces various new ones, called diffracted rays, in addition to the usual rays. Some of them enter the shadow regions and account for the light there while others go into the illuminated regions.

Diffracted rays are produced by incident rays which hit edges, corners, or vertices of boundary surfaces, or which graze such surfaces. Ordinary geometrical optics does not describe what happens in these cases but the new theory does. It does so by means of several laws of diffraction which are analogous to the laws of reflection and refraction. Like them, the new laws are deducible from Fermat's principle, appropriately modified. Away from the diffracting surfaces, diffracted rays behave just like ordinary rays.

In terms of the new rays, diffracted wave fronts can be defined. A Huygens wavelet construction can also be devised to determine them. It is also possible to introduce an eikonal or phase function which is constant on these wave fronts and which satisfies the usual eikonal equation. Thus all the fundamental principles of ordinary geometrical optics can be extended to the geometrical theory of diffraction.

Ordinary geometrical optics is often used to determine the distribution of light intensity, polarization, and phase throughout space. This is accomplished by assigning a field value to each ray and letting the total field at a point be the sum of the fields on all the rays through that point. The phase of the field on a ray is assumed to be proportional to the optical length of the ray from some reference point where the phase is zero. The amplitude is assumed to vary in accordance with the principle of conservation of energy in a narrow tube of rays. The direction of the field, when it is a vector, is given by a unit vector perpendicular to the ray. This vector slides parallel to itself along the ray in a homogeneous medium, and rotates around the ray in a specific way as it slides along it in an inhomogeneous medium.

Exactly the same principles as those just described can be used to assign a field to each diffracted ray. The only difficulty occurs in obtaining the initial value of the field at the point of diffraction. In the case of the ordinary rays, the field on a ray emerging from a source is specified at the source. But on a reflected or transmitted ray, the initial value is obtained by multiplying the field on the incident ray by a reflection or transmission coefficient. By analogy the initial value of the field on a diffracted ray is obtained by multiplying the field on the incident ray by a diffraction coefficient, which is a matrix for a vector field. There are different coefficients for edge diffraction, vertex diffraction, etc.

All the diffraction coefficients vanish as the wavelength λ of the field tends to zero. Then the sum of the fields on all diffracted rays, which we call the diffracted field, also vanishes. The geometrical optics field alone remains in this case, as we should expect because diffraction is usually attributed to the fact that λ is not zero.

* This paper was presented as an invited address at the Seminar on Recent Developments in Optics and Related Fields at the October 13, 1960 meeting of the Optical Society of America, Boston, Massachusetts. The research on which it is based was sponsored in part by the Air Force Cambridge Research Laboratories, Office of Aerospace Research.

[1] J. B. Keller, "The geometrical theory of diffraction," Proceedings of the Symposium on Microwave Optics, Eaton Electronics Research Laboratory, McGill University, Montreal, Canada (June, 1953).

[2] See J. B. Keller, in *Calculus of Variations and its Applications, Proceedings of Symposia in Applied Math*, edited by L. M. Graves (McGraw-Hill Book Company, Inc., New York and American Mathematical Society, Providence, Rhode Island, 1958), Vol. 8.

Dimensional considerations show that edge-diffraction coefficients are proportional to $\lambda^{\frac{1}{2}}$ and tip- or vertex-diffraction coefficients to λ. The field diffracted around a curved surface decreases exponentially with λ, and is consequently weaker than the field diffracted by a tip which is in turn weaker than that diffracted by an edge.

Diffraction coefficients can be characterized by recognizing that only the immediate neighborhood of the point of diffraction can affect their values. Thus the directions of incidence and diffraction, the wavelength, and the geometrical and physical properties of the media at the point of diffraction determine them. This suggests that they are determined by certain simpler problems in which only the local geometrical and physical properties enter. These "canonical" problems must be solved in order to determine the diffraction coefficients mathematically. Alternatively, experimental measurements on these canonical configurations can yield the coefficients.

The theory outlined above suffices for the analysis of a large class of situations involving diffraction. However, there remain phenomena which can be analyzed only by the inclusion of still another type of ray—the complex or imaginary ray. In a homogeneous medium such a ray is a complex straight line, while in an inhomogeneous medium it is a complex-valued solution of the differential equations for rays. Such rays occur as transmitted rays whenever total internal reflection occurs, and also in many other situations. Fields can be associated with these rays just as well as with the other kinds of rays.

A different kind of diffraction effect, not covered by the theory as explained so far, occurs at a caustic or focus of the ordinary or the diffracted rays. At such places neighboring rays intersect and the cross-sectional area of a tube of rays becomes zero. Consequently the principle of energy conservation in a tube of rays leads to an infinite amplitude for the field there. In order to obtain a finite value for the field at such places the present theory introduces a caustic correction factor. When the field on a ray passing through a caustic is multiplied by the appropriate factor, it becomes finite at the caustic. The caustic correction factors are determined by the wavelength and the local-ray geometry near the caustic, and are obtained from canonical problems.

In the subsequent sections this theory will be explained more fully and applied to some typical illustrative examples. Other applications of the theory will also be described. The mathematical interpretation of the theory in terms of asymptotic expansions will be explained in order to relate it to electromagnetic theory. The construction of similar theories in other branches of physics will also be commented upon.

2. EDGE-DIFFRACTED RAYS

The fundamental premise underlying the geometrical theory of diffraction is that light propagation is entirely a local phenomenon because the wavelength of light is so small. By this it is meant that the manner of propagation at a given point is determined solely by the properties of the medium and the structure of the field in an arbitrarily small neighborhood of the point. Furthermore all fields, no matter how they are produced, must have the same local structure, namely, that of a plane wave. Therefore, all fields must propagate in the same way. In particular, then, diffracted fields must travel along rays just like the ordinary geometrical optics field does, and, in fact, these rays must obey the ordinary geometrical-optics laws. The rays along which the diffracted field propagates are the diffracted rays.

Where do diffracted rays come from? It seems clear that they must be produced by some of the ordinary optical rays, but, which ones? The laws of propagation, reflection, and refraction enable us to follow the usual rays from the source outward, and determine where they go and what rays they produce, with some exceptions. The usual laws fail to specify what happens to a ray which hits an edge or a vertex, or grazes a boundary surface. Therefore, such rays must give rise to diffracted rays. We hypothesize that they do. In the case of edges, this hypothesis is related to Thomas Young's idea that diffraction is an edge effect.

This hypothesis can be tested mathematically in those cases where diffraction problems have been solved by other means. One such case is the diffraction of a plane wave by a semi-infinite screen with a straight edge. The solution of this problem obtained by Sommerfeld[3] consists of incident and reflected plane waves plus a third wave which is called a diffracted wave. When the incident wave is propagating in a direction normal to the edge of the screen, the diffracted wave is cylindrical with the edge as its axis. The straight lines orthogonal to the cylindrical wave fronts of the diffracted wave appear to come from the edge and can be identified with our diffracted rays. This example also suggests that an incident ray normal to the edge produces diffracted rays which are also normal to the edge and which leave it in all directions.

When the incident rays in the direction of propagation of the incident wave are oblique to the edge of the screen, the diffracted wave in Sommerfeld's solution is conical. This means that the diffracted wave fronts are parallel cones with the edge as their common axis. The straight lines orthogonal to these cones also appear to come from the edge, and can be identified with our diffracted rays. This example suggests the *law of edge diffraction*. A diffracted ray and the corresponding incident ray make equal angles with the edge at the point of diffraction, provided that they are both in the same medium. They lie on opposite sides of the plane normal to the edge at the point of diffraction. When the two rays lie in different media, the ratio of the sines of the

[3] A. J. W. Sommerfeld, *Optics* (Academic Press, Inc., New York, 1954).

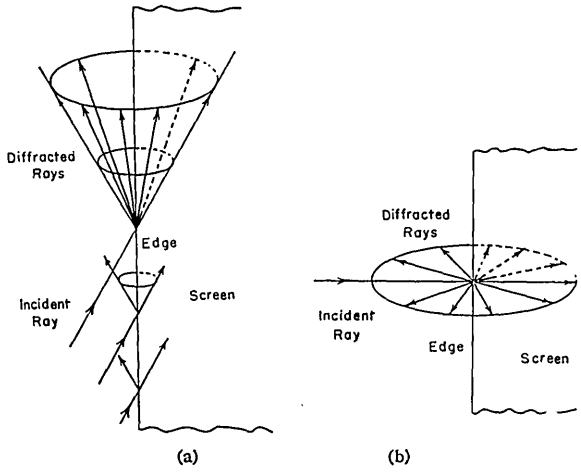

FIG. 1. (a) The cone of diffracted rays produced by an incident ray which hits the edge of a thin screen obliquely. (b) The plane of diffracted rays produced by a ray normally incident on the edge of a thin screen.

angles between the incident and diffracted rays and the normal plane is the reciprocal of the ratio of the indices of refraction of the two media. See Fig. 1.

The second part of this law, pertaining to diffraction into a different medium, is not suggested by the above example but is suggested by Snell's law of refraction. Both parts of the law are consequences of the following modified form of Fermat's principle which we call *Fermat's principle for edge diffraction*. An edge-diffracted ray from a point P to a point Q is a curve which has stationary optical length among all curves from P to Q with one point on the edge.

The derivation of the law of edge diffraction from this principle is particularly simple when the edge is straight and both rays lie in the same homogeneous medium. Then it is obvious that the ray consists of two straight-line segments meeting at a point on the edge. Let us rotate the plane containing the edge and the point Q around the edge until it contains P. In doing so the length of the segment from Q to the axis is unchanged, and the angle between the segment and the axis is unchanged. After the rotation, P, Q, and the edge lie in one plane and the stationary optical path is that given by the law of reflection. Thus the two segments must make equal angles with the edge and lie on opposite sides of the plane normal to the edge at the point of diffraction —but this is the law of edge diffraction for rays in the same medium. A similar argument, using Snell's law, yields the law of edge diffraction for rays in different media.

The law of edge diffraction is also confirmed by several approximate solutions of the problem of diffraction by an aperture in a thin screen. The Kirchhoff method, sometimes called the method of physical optics, represents the field diffracted through an aperture as an integral over the aperture. Rubinowicz[4] evaluated this integral asymptotically for short wavelengths when the incident field was a spherical wave, by using the Maggi transformation. He showed that the diffracted field at a point Q consisted of contributions from a small number of points on the edge. If we draw straight lines from these points to Q and call them diffracted rays we find that all of them at smooth parts of the edge satisfy the law of edge diffraction. The integrals of the Kirchhoff and modified Kirchhoff method, which employs Rayleigh's formulas, have been evaluated asymptotically for short wavelengths by van Kampen[5] and by Keller *et al.*[6] for arbitrary incident fields. The latter authors also evaluated Braunbek's[7] improved version of the Kirchhoff integrals. In all cases the points on smooth parts of the edge which contribute to the diffracted field correspond to diffracted rays satisfying the law of edge diffraction.

An indirect experimental verification of the existence of edge diffracted rays and of the law of edge diffraction is contained in the results of Coulson and Becknell.[8] They photographed the cross sections of the shadows of thin opaque disks of various shapes and detected bright lines in the shadows. When the incident field was normally incident on a disk, the bright line was found to be the evolute of the edge of the disk. In the special case of a circle the evolute is just the center, which appears as the well-known bright spot on the axis of the shadow. According to the law of edge diffraction, the diffracted rays lie in planes normal to the edge when the incident rays are normal to the edge. Therefore the caustic of these rays is a cylinder normal to the disk. Its cross section is the envelope of the normals to the edge in the plane of the disk, which is just the evolute of the edge. Thus the cross section of the caustic of the diffracted rays coincides with the bright lines which were observed. A similar interpretation applies to the bright lines observed at oblique incidence. Similar bright lines were observed by Nienhuis[9] in the diffraction patterns of apertures. He found that they were exactly the caustics of diffracted rays which he assumed to emanate normally from the edge.

3. FIELDS DIFFRACTED BY STRAIGHT EDGES

Let us now consider the field u_e on a ray diffracted from an edge.[10] For simplicity let us suppose that the ray is in a homogeneous medium so that it is a straight line. Let us begin with the two-dimensional case in which the edge is a straight line and the incident rays all lie in planes normal to the edge. Then the diffracted rays are also normal to the edge, and emanate from it in all

[4] A. Rubinowicz, Ann. Physik **53**, 257 (1917); **73**, 339 (1924).

[5] N. G. van Kampen, Physica **14**, 575 (1949).

[6] J. B. Keller, R. M. Lewis, and B. D. Seckler, J. Appl. Phys. **28**, 570 (1957).

[7] W. Braunbek, Z. Physik **127**, 381 (1950); **127**, 405 (1950).

[8] J. Coulson and G. G. Becknell, Phys. Rev. **20**, 594 (1922); **20**, 607 (1922).

[9] K. Nienhuis, Thesis, Groningen, 1948.

[10] J. B. Keller, J. Appl. Phys. **28**, 426 (1957).

directions. Thus it suffices to consider only the rays in one plane normal to the edge.

If λ denotes the wavelength of the incident field u_i, it is convenient to define the propagation constant $k=2\pi/\lambda$. We also denote by r the distance from the edge. Then the phase on a diffracted ray is just kr plus the phase ψ_i of the incident ray at the edge. To find the amplitude $A(r)$, which we assume to be a scalar, we consider as a tube of rays two neighboring rays in the same plane normal to the edge. Actually the tube is a cylinder of unit height. The cross-sectional area of this tube is proportional to r and the flux through it is proportional to rA^2. Since the flux must be constant, $A(r)$ is proportional to $r^{-\frac{1}{2}}$. The amplitude is also proportional to the incident amplitude A_i at the edge so we write $A(r)=DA_ir^{-\frac{1}{2}}$, where D is a diffraction coefficient. Thus the diffracted field is

$$u_e=DA_ir^{-\frac{1}{2}}e^{i(kr+\psi_i)}=Du_ir^{-\frac{1}{2}}e^{ikr}. \tag{1}$$

Let us compare our result (1) with Sommerfeld's[3] exact solution for diffraction of a plane scalar wave by a half-plane. When his result is asymptotically expanded for large values of kr, it agrees perfectly with (1) provided that

$$D=-\frac{e^{i\pi/4}}{2(2\pi k)^{\frac{1}{2}}\sin\beta}\left[\sec\tfrac{1}{2}(\theta-\alpha)\pm\csc\tfrac{1}{2}(\theta+\alpha)\right]. \tag{2}$$

Here β is the angle between the incident ray and the edge which is $\pi/2$ in this case. The angles between the incident and diffracted rays and the normal to the screen are θ and α, respectively (see Fig. 2). The upper sign applies when the boundary condition on the half-plane is $u=0$ while the lower sign applies if it is $\partial u/\partial n=0$.

The agreement between (1) and the exact solution confirms our theory and also determines the edge diffraction coefficient D. Similar agreement occurs for oblique incidence on a half-plane when (1) is replaced by the appropriate expression and the denominator $\sin\beta$ is included in (2). In this case θ and α are defined as above after first projecting the rays into the plane normal to the edge. In case the half-plane is replaced by a wedge of angle $(2-n)\pi$, comparison of (1), and its modified form for $\beta\neq\pi/2$, with Sommerfeld's exact solution for a wedge yields agreement when

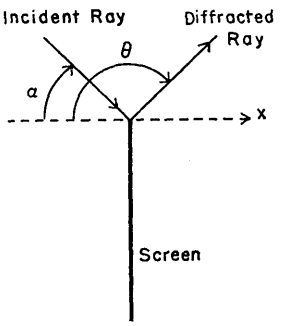

FIG. 2. The projection of incident and diffracted rays into a plane normal to the edge of a screen. The angles α and θ are those between the projections and the normal to the screen, measured as shown in the figure. The edge is normal to the plane of the figure.

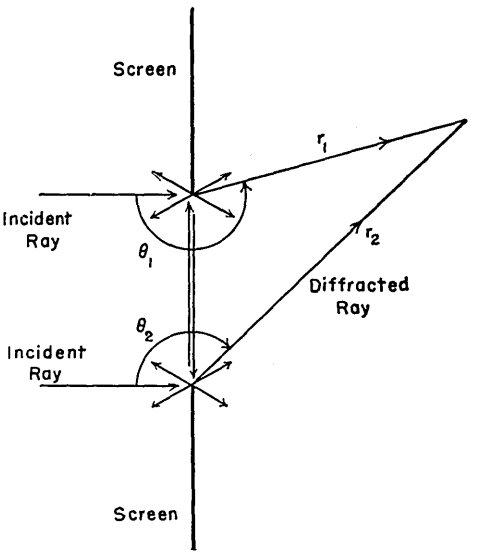

FIG. 3. The diffracted rays produced by a plane wave normally incident on a slit in a thin screen. The two incident rays which hit the slit edges are shown, with some of the singly diffracted rays they produce. One diffracted ray from each edge is shown crossing the slit and hitting the opposite edge, producing doubly diffracted rays.

$$D=\frac{e^{i\pi/4}\sin\frac{\pi}{n}}{n(2\pi k)^{\frac{1}{2}}\sin\beta}\left[\left(\cos\frac{\pi}{n}-\cos\frac{\theta-\alpha}{n}\right)^{-1}\mp\left(\cos\frac{\pi}{n}-\cos\frac{\theta+\alpha+\pi}{n}\right)^{-1}\right]. \tag{3}$$

For $n=2$ the wedge becomes a half-plane and (3) reduces to (2). [Equation (3) is misprinted in Eq. (A10) of reference 10.] In the electromagnetic case D is a matrix which has been determined in reference 10 for a perfectly conducting thin screen.

We shall now apply (1) and (2) to determine the field diffracted through an infinitely long slit of width $2a$ in a thin screen. By Babinet's principle, this will also be the field diffracted by a thin strip of width $2a$. For simplicity we shall assume that the incident field is a plane wave propagating in a direction normal to the edges of the slit. Then the problem is two dimensional and we can confine our attention to a plane normal to the edges. In this plane let the screen lie in the y axis of a rectangular coordinate system with the edges of the slit at $x=0$, $y=\pm a$. Let the incident field be $e^{ik(x\cos\alpha-y\sin\alpha)}$. Two singly diffracted rays, one from each edge, pass through any point P. Thus the singly diffracted field at P, $u_{e1}(P)$, is the sum of two terms of the form (1).

$$u_{e1}(P)=-\frac{e^{ik(r_1-a\sin\alpha)+i\pi/4}}{2(2\pi kr_1)^{\frac{1}{2}}}\left[\sec\tfrac{1}{2}(\theta_1+\alpha)\pm\csc\tfrac{1}{2}(\theta_1-\alpha)\right]$$
$$-\frac{e^{ik(r_2+a\sin\alpha)+i\pi/4}}{2(2\pi kr_2)^{\frac{1}{2}}}+\left[\sec\tfrac{1}{2}(\theta_2-\alpha)\pm\csc\tfrac{1}{2}(\theta_2+\alpha)\right]. \tag{4}$$

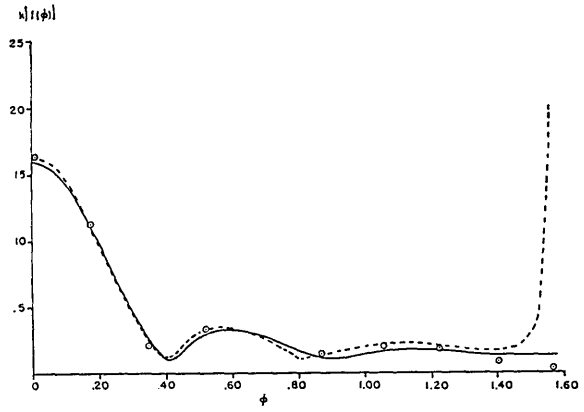

FIG. 4. The far-field diffraction pattern of a slit of width $2a$ hit normally by a plane wave; $ka=8$. The solid curve based upon Eq. (6) results from single diffraction, and applies to a screen on which $u=0$ or $\partial u/\partial n=0$. The dashed curve includes the effects of multiple diffraction for a screen on which $u=0$. The dots are based upon the exact solution of the reduced wave equation for a screen on which $u=0$.[11] The ordinate is $k|f(\varphi)|$ and the abscissa is φ in radians.

In (4) r_1 and r_2 denote the distances from P to the upper and lower edges, and the angles θ_1 and θ_2 are determined by the corresponding rays, as shown in Fig. 3.

The diffraction pattern of the slit, $f_s(\varphi)$, can be obtained from (4) by introducing the polar coordinates r,φ of P. When P is far from the slit ($r\gg a$) we have $r_1\sim r-a\sin\varphi$, $r_2\sim r+a\sin\varphi$, $\theta_1\sim\pi+\varphi$, and $\theta_2\sim\pi-\varphi$. Then (4) becomes

$$u_{c1}(P)=-(k/2\pi r)^{\frac{1}{2}}e^{ikr+i\pi/4}f_s(\varphi). \tag{5}$$

The diffraction pattern $f_s(\varphi)$ due to single diffraction is found to be

$$f_s(\varphi)=i\frac{\sin[ka(\sin\varphi+\sin\alpha)]}{k\sin\frac{1}{2}(\varphi+\alpha)}\pm\frac{\cos[ka(\sin\varphi+\sin\alpha)]}{k\cos\frac{1}{2}(\varphi-\alpha)}. \tag{6}$$

This equation shows that $|f_s(\varphi)|$ is not zero for any value of φ and that it is the same for the two types of boundary conditions on the screen.

In Figs. 4 and 5, $|kf_s(\varphi)|$ is shown as a function of φ for normal incidence ($\alpha=0$). The range of φ is 0 to $\pi/2$ which is the region behind the slit. In Fig. 4 for which $ka=8$, the exact values of the diffraction pattern are shown. They are based upon numerical evaluation of the exact solution of the reduced wave equation with $u=0$ on the screen, and are given by Karp and Russek.[11] The agreement between our theory and the exact values is seen to be quite good even though the slit is only about 3 wavelengths wide.

The transmission cross section σ of the slit per unit length can be obtained simply by employing the cross-section theorem. According to this theorem it is equal to the imaginary part of the diffraction pattern in the forward direction $\varphi=-\alpha$. By applying the theorem to $f_s(\varphi)$ we obtain $\sigma=2a\cos\alpha$, which is just the geometrical optics result.

[11] S. N. Karp and A. Russek, J. Appl. Phys. 27, 886 (1956).

To obtain a better value of σ, let us consider the two doubly diffracted rays which go in the direction $\varphi=-\alpha$. They are produced by two singly diffracted rays which cross the slit and hit the opposite edges. The field at the upper edge on the singly diffracted ray from the lower edge is given by the second term in (4) with $\theta_2=\pi/2$ and $r_2=2a$. If we choose the upper sign, appropriate to a screen on which $u=0$, we obtain

$$u_{c1}=-\frac{e^{ika(2+\sin\alpha)+i\pi/4}}{2(\pi ka)^{\frac{1}{2}}}\sec\frac{1}{2}\left(\frac{\pi}{2}-\alpha\right). \tag{7}$$

We now use (7) as the incident field in (1) to obtain the field on the doubly diffracted rays. We proceed similarly to find the doubly diffracted field from the lower edge and add the two results to obtain the doubly diffracted field u_{c2}. Far from the slit we can write it in the form (5) with $f_d(\varphi)$ instead of $f_s(\varphi)$. In the direction $\varphi=-\alpha$, f_d has the value

$$f_d(-\alpha)=-\frac{1}{k(\pi ka)^{\frac{1}{2}}}\left[\frac{e^{i2ka(1+\sin\alpha)+i\pi/4}}{1+\sin\alpha}+\frac{e^{i2ka(1-\sin\alpha)+i\pi/4}}{1-\sin\alpha}\right]. \tag{8}$$

To obtain σ we apply the cross-section theorem which requires taking the imaginary part of $f_s(-\alpha)+f_d(-\alpha)$. The result is

$$\sigma=2a\cos\alpha-\frac{1}{k(\pi ka)^{\frac{1}{2}}}\left[\frac{\cos[2ka(1+\sin\alpha)-\pi/4]}{1+\sin\alpha}+\frac{\cos[2ka(1-\sin\alpha)-\pi/4]}{1-\sin\alpha}\right]. \tag{9}$$

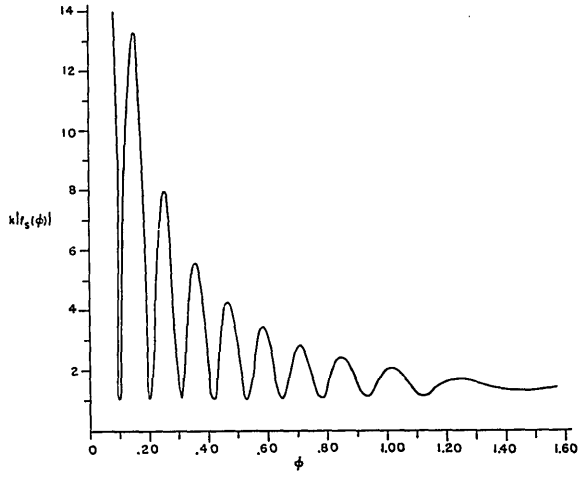

FIG. 5. The far-field diffraction pattern of a slit of width $2a$ hit normally by a plane wave; $ka=10\pi$. The curve, based on Eq. (6), results from single diffraction for a screen on which $u=0$ or $\partial u/\partial n=0$. The ordinate is $k|f_s(\varphi)|$ and the abscissa is φ in radians. The value at $\varphi=0$ is finite but too high to be shown.

This expression has also been obtained by H. Levine by a different method. For normal incidence, $\alpha=0$, and (9) becomes

$$\frac{\sigma}{2a}=1-\frac{\cos(2ka-\pi/4)}{\pi^{\frac{1}{2}}(ka)^{\frac{3}{2}}}. \tag{10}$$

In Fig. 6 $\sigma/2a$ is shown as a function of ka based on (10), and points obtained from the exact solution of the boundary-value problem are also shown. The agreement is quite good even when the wavelength is several times the slit width. When additional multiply diffracted rays are taken into account, even better agreement is obtained, as the dashed curve in the figure shows. The expression for the field obtained by summing all the multiply diffracted fields coincides with that obtained by Karp and Russek[11] in a different way.

The above calculation of the doubly diffracted field was performed for a screen on which $u=0$. For a screen on which $\partial u/\partial n=0$ no diffracted field is produced when a plane wave travels toward the edge in a direction parallel to the screen. As a consequence the diffraction coefficient D, given by (2) with the lower sign, vanishes when $\alpha=\pi/2$. In this case the diffracted field is proportional to $\partial u_i/\partial n$, the normal derivative of the incident field at the edge. Thus (1) must be replaced by

$$u_e=D'\frac{\partial u_i}{\partial n}\cdot r^{-\frac{1}{2}}e^{ikr}. \tag{11}$$

Here D' is a new diffraction coefficient which can be obtained by solving the problem of diffraction of the wave xe^{-iky} by a half-plane lying on the negative y axis. This was done by Karp and Keller[12] with the result

$$D'(\theta)=-\frac{1}{ik}\frac{\partial}{\partial\alpha}D(\theta,\pi/2)$$

$$=-\frac{e^{-i\pi/4}}{2(2\pi)^{\frac{1}{2}}k^{\frac{3}{2}}\sin^2\beta}\frac{\sin\left(\frac{\pi}{4}-\frac{\theta}{2}\right)}{\cos^2\left(\frac{\pi}{4}-\frac{\theta}{2}\right)}. \tag{12}$$

Let us use (11) and (12) to determine the doubly diffracted field for a screen on which $\partial u/\partial n=0$. First we obtain from the second term in (4) at $\theta_2=\pi/2$, $r_2=2a$ the result

$$\frac{\partial u_{e1}}{\partial n}=-\frac{e^{ika(2+\sin\alpha)+i\pi/4}}{8(\pi ka^3)^{\frac{1}{2}}}\cdot\frac{\sin\left(\frac{\pi}{4}-\frac{\alpha}{2}\right)}{\cos^2\left(\frac{\pi}{4}-\frac{\alpha}{2}\right)}. \tag{13}$$

We now use (13) and (12) in (11) to obtain the doubly diffracted field from the upper edge, and do a similar

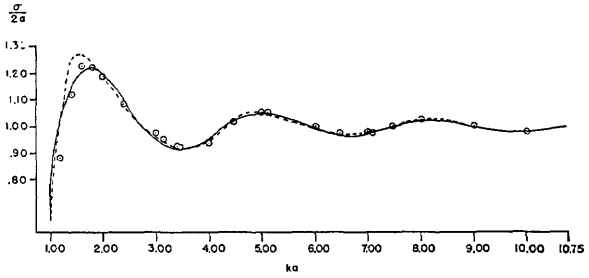

FIG. 6. The transmission cross section of a slit of width $2a$ as a function of ka, for normal incidence with $u=0$ on the screen. The solid curve, based on Eq. (10), results from single and double diffraction; the dashed curve includes single and all multiple diffraction. The dots are based upon the exact solution of the reduced wave equation with $u=0$ on the screen.[11] The ordinate is $\sigma/2a$ and the abscissa is ka.

calculation for the lower edge. Then we add the two doubly diffracted fields to obtain

$$u_{e2}=-\frac{e^{ika(2+\sin\alpha)+ikr_1}}{16\pi(ka)^{\frac{3}{2}}(2kr_1)^{\frac{1}{2}}}\cdot\frac{\sin\left(\frac{\pi}{4}-\frac{\alpha}{2}\right)}{\cos^2\left(\frac{\pi}{4}-\frac{\alpha}{2}\right)}\cdot\frac{\sin\left(\frac{\pi}{4}-\frac{\theta_1}{2}\right)}{\cos^2\left(\frac{\pi}{4}-\frac{\theta_1}{2}\right)},$$

$$-\frac{e^{ika(2-\sin\alpha)+ikr_2}}{16\pi(ka)^{\frac{3}{2}}(2kr_2)^{\frac{1}{2}}}\cdot\frac{\sin\left(\frac{\pi}{4}+\frac{\alpha}{2}\right)}{\cos^2\left(\frac{\pi}{4}+\frac{\alpha}{2}\right)}\cdot\frac{\sin\left(\frac{\pi}{4}-\frac{\theta_2}{2}\right)}{\cos^2\left(\frac{\pi}{4}-\frac{\theta_2}{2}\right)}. \tag{14}$$

Far from the slit we write u_{e2} in the form (5) and obtain $f_d(\varphi)$. Then we apply the cross-section theorem to $f_s(-\alpha)+f_d(-\alpha)$ and find

$$\sigma=2a\cos\alpha-\frac{2a}{32\pi^{\frac{1}{2}}(ka)^{\frac{5}{2}}}$$

$$\times\left\{\frac{\sin^2\left(\frac{\pi}{4}-\frac{\alpha}{2}\right)}{\cos^4\left(\frac{\pi}{4}-\frac{\alpha}{2}\right)}\sin[2ka(1+\sin\alpha)-\pi/4]\right.$$

$$\left.+\frac{\sin^2\left(\frac{\pi}{4}+\frac{\alpha}{2}\right)}{\cos^4\left(\frac{\pi}{4}+\frac{\alpha}{2}\right)}\sin[2ka(1-\sin\alpha)-\pi/4]\right\}. \tag{15}$$

For normal incidence $\alpha=0$ and (15) becomes

$$\frac{\sigma}{2a}=1-\frac{\sin(2ka-\pi/4)}{8\pi^{\frac{1}{2}}(ka)^{\frac{3}{2}}}. \tag{16}$$

[12] S. N. Karp and J. B. Keller, Optica Acta (Paris) 8, 61 (1961).

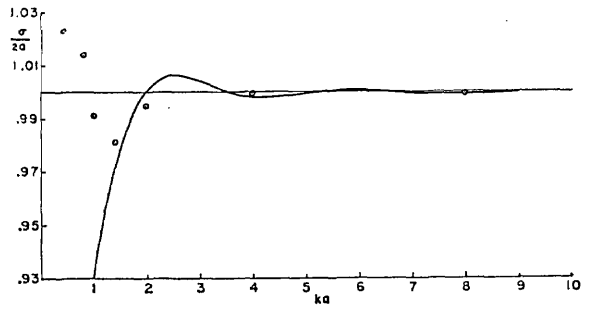

FIG. 7. The transmission cross section of a slit of width $2a$ as a function of ka, for normal incidence with $\partial u/\partial n=0$ on the screen. The curve, based on Eq. (16), results from single and double diffraction. The dots are based upon the exact solution of the reduced wave equation.[13] The ordinate is $\sigma/2a$ and the abscissa is ka.

A graph of $\sigma/2a$ versus ka based on (16) is shown in Fig. 7 together with some values of $\sigma/2a$ computed by Skavlem[13] from the exact solution of the boundary value problem.

Let us now consider a diffraction grating consisting of $2N+1$ parallel slits each of width $2a$ with centers a distance b apart. Let us again consider the two-dimensional case in which a plane wave is incident at angle α with its wavefronts parallel to the edges. Then one singly diffracted ray from each edge will pass through every point and the total field is the sum of $2(2N+1)$ terms of the form (1). Let us number the slits from $t=-N$ to $t=N$ with the center of slit t at $y=tb$ and let r_t,φ_t be polar coordinates with their origin at this center. We now add together the fields on the two rays from the edges of slit t. If $r_t\gg a$ we obtain the result (5) for that slit, with r_t,φ_t instead of r,φ and multiplied by the factor $e^{-ikt\sin\alpha}$ to account for the phase of the incident wave. Thus far from the grating we have

$$u_{01}=-\sum_{t=-N}^{t=N}\left(\frac{k}{2\pi r_t}\right)^{\frac{1}{2}}e^{ik(r_t-tb\sin\alpha)+i\pi/4}f_s(\varphi). \quad (17)$$

To simplify (17), let r,φ be polar coordinates with their origin at the center of the middle slit. At points far from the grating compared to its length $2Nb+2a$ we have $\varphi_t\sim\varphi$ and $r_t\sim r-tb\sin\varphi$ and (17) becomes

$$u_{01}=-\left(\frac{k}{2\pi r}\right)^{\frac{1}{2}}e^{ikr+i\pi/4}f_s(\varphi)\sum_{t=-N}^{t=N}e^{-iktb(\sin\varphi+\sin\alpha)}$$

$$=-\left(\frac{k}{2\pi r}\right)^{\frac{1}{2}}e^{ikr+i\pi/4}f_s(\varphi)$$

$$\times\frac{\sin[(N+\frac{1}{2})kb(\sin\varphi+\sin\alpha)]}{\sin[\frac{1}{2}kb(\sin\varphi+\sin\alpha)]}. \quad (18)$$

The same result applies with N instead of $N+\frac{1}{2}$ if there are $2N$ slits and the origin is at the midpoint of the grating. The result expresses the diffraction pattern as the product of the pattern of a single slit and a grating factor which is the quotient of sines.

The method of this section has been applied to diffraction by a semi-infinite thick screen with a flat end and by a truncated wedge by Burke and Keller.[14] In computing the doubly diffracted field for a screen upon which $u=0$ it was found that $D=0$. This is related to the fact that the singly diffracted field vanishes at the edge. Here again the doubly diffracted field is proportional to $\partial u_i/\partial n$ at the edge. The corresponding diffraction coefficient $D'(\theta,n)$ was found by solving a canonical problem, as was done to obtain (12). If $D(\alpha,\theta,n)$ denotes the coefficient in (3), the result is

$$D'(\theta,n)=\frac{1}{ik}\frac{\partial}{\partial\alpha}D(-\pi/2,\theta,n)=\frac{2e^{-i\pi/4}\sin(\pi/n)}{n^2k^{\frac{1}{2}}(2\pi)^{\frac{1}{2}}}$$

$$\times\frac{\sin\left[\left(\theta+\frac{\pi}{2}\right)/n\right]}{\left\{\cos\frac{\pi}{n}-\cos\left[\left(\theta+\frac{\pi}{2}\right)/n\right]\right\}}. \quad (19)$$

In the case of grazing incidence additional considerations which we shall not explain are required because the incident ray and a diffracted ray both continue along the surface together.

4. FIELDS DIFFRACTED BY CURVED EDGES

The field on an edge-diffracted ray is given by (1) only in the special case when the diffracted wave is cylindrical. In general (1) must be replaced by

$$u_e=Du_i[r(1+\rho_1^{-1}r)]^{-\frac{1}{2}}e^{ikr}. \quad (20)$$

Here ρ_1 denotes the distance from the edge to the caustic of the diffracted rays, measured negatively in the direction of propagation. (See Fig. 8.) We obtain (20) by employing the energy principle according to which A varies inversely as the square root of the cross-sectional area of a tube of rays. If ρ_1 and ρ_2 denote the radii of curvature of the wave front which is orthogonal to the tube at one cross section, then ρ_1+r and ρ_2+r are the corresponding radii a distance r further along the tube. (See Fig. 9.) Thus the cross-sectional area is proportional to $(\rho_1+r)(\rho_2+r)$ and A is proportional to $[(\rho_1+r)(\rho_2+r)]^{-\frac{1}{2}}$. When $\rho_2=0$, A varies as $[r(1+\rho_1^{-1}r)]^{-\frac{1}{2}}$. Since the diffracted rays form a caustic at the edge, $\rho_2=0$ there. If r denotes distance along a ray from the edge, the last expression for A applies. By using it together with the phase factor and the factors Du_i previously explained, we obtain (20).

The distance ρ_1 in (20) can be found by geometrical considerations. When the edge is a plane curve, it is

[13] C. J. Bouwkamp, Repts. Progr. in Phys. 17, 35 (1954).

[14] J. E. Burke and J. B. Keller, Research Rept. EDL-E48, Electronic Defense Laboratories, Sylvania Electronic Systems, Mountain View, California (March, 1960).

given by the following formula, which is analogous to the lens and mirror laws

$$\frac{1}{\rho_1}=-\frac{\dot\beta}{\sin\beta}-\frac{\cos\delta}{\rho\sin^2\beta}. \tag{21}$$

Here $\rho\gtrless 0$ denotes the radius of curvature of the edge, β is the angle between the incident ray and the (positive) tangent to the edge, $\dot\beta$ is the derivative of β with respect to arclength s along the edge, and δ is the angle between the diffracted ray and the normal to the edge. Since both β and $\partial/\partial s$ change sign when the direction of increasing arclength is reversed, this direction can be chosen arbitrarily without effecting the value of ρ_1. The normal, which lies in the plane of the edge, is assumed to point towards its center of curvature.

As an application of (21) let us consider a spherical wave diffracted by a straight edge. In this case $\rho=\infty$ so the focal length of the edge is infinite and the last term in (21) vanishes. A simple calculation shows that $\dot\beta=-R^{-1}\sin\beta$ at a point on the edge a distance R from the source. Thus (21) yields $\rho_1=R$ as we should have expected. If $u_i=e^{ikR}/R$ then (20) yields

$$u_e=-\frac{e^{ik(R+r)+i\pi/4}}{2\sin\beta(2\pi k)^{\frac12}[rR(r+R)]^{\frac12}}\times\left[\sec\tfrac12(\theta-\alpha)\pm\csc\tfrac12(\theta+\alpha)\right]. \tag{22}$$

This result coincides with the asympotic expansion of the diffracted field given by the exact solution of the reduced wave equation for this problem, providing another confirmation of our theory.

When a plane wave is normally incident upon a screen containing an aperture of any shape, $\beta=\pi/2$ for every ray and (21) yields $\rho_1=-\rho/\cos\delta$. If $\beta=\pi/2$ it follows from the definitions of θ and δ that $\delta=\theta-\pi/2$ so $\rho_1=-\rho/\sin\theta$. For a circular aperture of radius a, $\rho=a$ and $\rho_1=-a/\sin\theta$. In this case two singly diffracted rays pass through each point P not on the axis. They come from the nearest and farthest points on the edge. If the incident field is $u_i=e^{ikx}$ and the screen is in the plane $x=0$, we obtain by adding two terms of the form (20) the result

$$u_{el}(P)=-\frac{e^{ikr+i\pi/4}}{2(2\pi k)^{\frac12}}\left[\sec\frac{\theta_1}{2}\pm\csc\frac{\theta_1}{2}\right][r_1(1-a^{-1}r_1\sin\theta_1)]^{-\frac12}$$
$$-\frac{e^{ikr_2+i\pi/4}}{2(2\pi k)^{\frac12}}\left[\sec\frac{\theta_2}{2}\pm\csc\frac{\theta_2}{2}\right]\times[r_2(1-a^{-1}r_2\sin\theta_2)]^{-\frac12}. \tag{23}$$

The angles and distances in (23) are as shown in Fig. 3.

Far from the aperture (23) simplifies to

$$u_{el}=-\frac{ke^{ikr}}{2\pi r}f_s(\varphi). \tag{24}$$

The diffraction pattern $f_s(\varphi)$ due to single diffraction is from (23)

$$f_s(\varphi)=k^{-\frac12}(2\pi a/\sin\varphi)^{\frac12}\left\{i\frac{\sin[ka\sin\varphi-(\pi/4)]}{\sin(\varphi/2)}\pm\frac{\cos[ka\sin\varphi-(\pi/4)]}{\cos(\varphi/2)}\right\}. \tag{25}$$

A graph of $|kf_s(\varphi)|$ based on (25) is given in Fig. 10 for $ka=3\pi$. It shows that $f_s(\varphi)$ is infinite at $\varphi=0$, which is a consequence of the fact that the axis, a caustic of the diffracted rays, and the shadow boundary $r=a$ both extend in the direction $\varphi=0$.

Let us examine (23) near the axis by introducing ρ and x, the distances from P to the axis and to the plane of the screen, and $\delta=\tan^{-1}(x/a)$. When $\rho\ll a$, (23) can be simplified to

$$u_{el}=-\frac{a^{\frac12}e^{ik(x\sin\delta+a\cos\delta)}}{(2\pi k\rho)^{\frac12}(x^2+a^2)^{\frac14}}\left[\sec\left(\frac{\delta}{2}+\frac{\pi}{4}\right)\pm\csc\left(\frac{\delta}{2}+\frac{\pi}{4}\right)\right]\times\cos\left(k\rho\cos\delta-\frac{\pi}{4}\right). \tag{26}$$

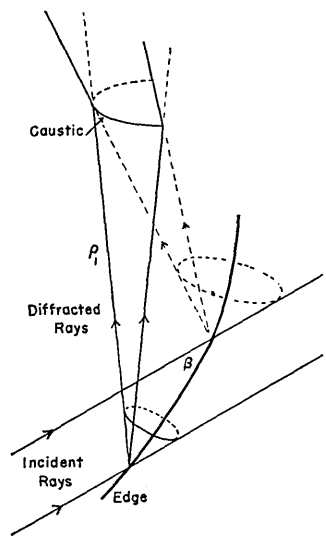

FIG. 8. A pair of neighboring incident rays hitting a curved edge, and some of the resulting diffracted rays. The two cones of diffracted rays intersect at the caustic, which is at the distance ρ_1 from the edge along the rays.

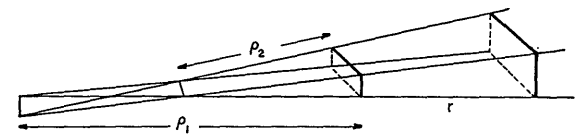

FIG. 9. A tube of rays and two small portions of wave fronts normal to them a distance r apart. The neighboring rays intersect at the two centers of curvature of the wave fronts. They are at the distances ρ_1 and ρ_2 from one wave front and at ρ_1+r and ρ_2+r from the other. The ratio of the areas of the wavefront sections is seen to be $\rho_1\rho_2/(\rho_1+r)(\rho_2+r)$.

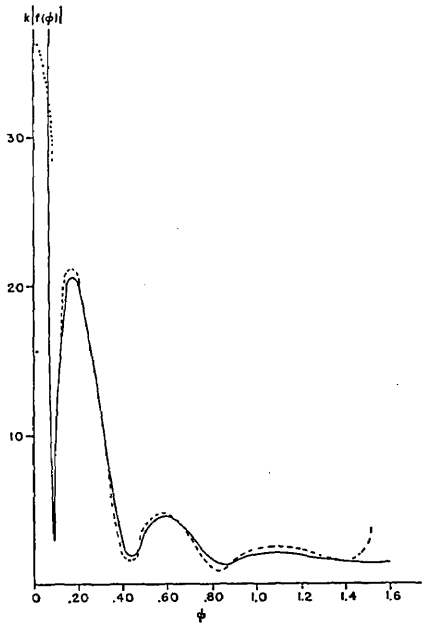

FIG. 10. The diffraction pattern of a circular hole of radius a hit normally by a plane wave; $ka=3\pi$. The solid curve, based on Eq. (25) results from single diffraction with either $u=0$ or $\partial u/\partial n=0$ on the screen. The dashed curve, for the case $u=0$ on the screen, also includes multiple diffraction. The dotted curve near $\varphi=0$ includes the caustic and shadow-boundary corrections given by (30).

This equation shows that u_{el} is singular like $\rho^{-\frac{1}{2}}$ on the axial caustic $\rho=0$. To modify it so that u_{el} is finite there, we consider the following exact solution of the reduced wave equation, in which δ is a constant and J_0 is the zero-order Bessel function

$$e^{ikz\sin\delta}J_0(k\rho\cos\delta)\sim\frac{2e^{ikz\sin\delta}}{(2\pi k\rho\cos\delta)^{\frac{1}{2}}}\cos(k\rho\cos\delta-\pi/4). \quad (27)$$

The right side of (27) is the asymptotic expansion of the solution on the left side when $k\rho$ is large. It shows that the solution corresponds to rays converging on an axial caustic, making the angle $\pi/2-\delta$ with the axis, just as is the case in (26). Near the axis the right side of (27) is not applicable and the same is true of (26). Therefore we assume that the right side of (26) can be converted to its correct value on and near the axis by multiplying it by the ratio of the left side of (27) to the right side, and we call this ratio the axial caustic correction factor F

$$F=\tfrac{1}{2}(2\pi k\rho\cos\delta)^{\frac{1}{2}}\sec(k\rho\cos\delta-\pi/4)J_0(k\rho\cos\delta). \quad (28)$$

[The factor $\frac{1}{2}$ was omitted in (A16) of reference 10]. When we multiply the right side of (26) by F, we obtain on and near the axis

$$u_{el}=-\frac{(a\cos\delta)^{\frac{1}{2}}}{2(x^2+a^2)^{\frac{1}{2}}}\left[\sec\left(\frac{\delta}{2}+\frac{\pi}{4}\right)\pm\csc\left(\frac{\delta}{2}+\frac{\pi}{4}\right)\right]\times J_0(k\rho\cos\delta)\exp[ik(x^2+a^2)^{\frac{1}{2}}]. \quad (29)$$

Far from the screen δ approaches $\pi/2$ and $\sec[(\delta/2)+(\pi/4)]\approx 2x/a$, so u_{el} does not decrease with x along the axis. This is because the shadow boundary $r=a$ extends in the direction parallel to the boundary. When we combine u_{el} with the geometrical optics field $u_g=e^{ikx}$ in the region $r<a$, the plane wave terms cancel and we obtain a spherical wave

$$u_g+u_{el}\approx e^{ikx}-\frac{a}{2x}e^{ikx}\left(1+\frac{ika^2}{2x}\right)\frac{2x}{a}J_0(ka\sin\varphi)=-\frac{ke^{ikx}}{2\pi x}i\pi a^2J_0(ka\sin\varphi). \quad (30)$$

Thus $f_s(\varphi)$ has the finite value $i\pi a^2J_0(ka\sin\varphi)$ near $\varphi=0$, rather than the singular value given by (25). The transmission cross section σ is the imaginary part of $f(0)$ according to the cross-section theorem, so (30) yields $\sigma=\pi a^2$, the geometrical optics value.

The result (30) applies only in the immediate vicinity of $\varphi=0$ since its derivation employed the geometrical optics field in the region $r<a$. To obtain an improved form of the diffraction pattern (25) which is valid for a larger range of φ, we shall introduce a caustic and shadow boundary correction directly into (25). We do this by observing that $(\sin\varphi)^{-\frac{1}{2}}\sin(ka\sin\varphi-\pi/4)$ and $(\sin\varphi)^{-\frac{1}{2}}\cos(ka\sin\varphi-\pi/4)$ are proportional to the asymptotic forms of $J_1(ka\sin\varphi)$ and $J_0(ka\sin\varphi)$, respectively. Upon using these functions in (25) in place of their asymptotic forms, we obtain

$$f_s(\varphi)=\frac{\pi a}{k}\left[\frac{iJ_1(ka\sin\varphi)}{\sin(\varphi/2)}\pm\frac{J_0(ka\sin\varphi)}{\cos(\varphi/2)}\right]. \quad (31)$$

Let us now consider the doubly diffracted field in order to obtain a more accurate value of σ. For a screen on which $u=0$ the first term in (23) becomes at $r_1=2a$, $\theta=\pi/2$

$$u_{el}=-\frac{e^{2ika+i\pi/4}}{(2\pi ka)^{\frac{1}{2}}}. \quad (32)$$

This is the field on a ray which has crossed the aperture to the opposite edge. By treating it as the incident field in (20) we obtain for the doubly diffracted field

$$u_{e2}=\frac{e^{ik(r_1+2a)}}{2\pi ka^{\frac{1}{2}}}[r_1(1-a^{-1}r_1\cos\delta_1)]^{-\frac{1}{2}}\sec\frac{\delta_1}{2}+\frac{e^{ik(r_2+2a)}}{2\pi ka^{\frac{1}{2}}}[r_2(1-a^{-1}r_2\cos\delta_2)]^{-\frac{1}{2}}\sec\frac{\delta_2}{2}. \quad (33)$$

Near the axis we find by comparing (33) with (26) that

$$u_{e2}=-\frac{2e^{2ika+\pi/4}}{(2\pi ka)^{\frac{1}{2}}}\sec\frac{\delta}{2}\left[\sec\left(\frac{\delta}{2}+\frac{\pi}{4}\right)+\csc\left(\frac{\delta}{2}+\frac{\pi}{4}\right)\right]^{-1}u_{el}. \quad (34)$$

From (34) we see that u_{e2} becomes infinite on the axis like u_{e1} does and it can be made regular by the same caustic correction. The result can be obtained by using (29) in (34). In this way we obtain at points far behind the screen and near the axis

$$u_{e2}=-\frac{ke^{ikx}}{2\pi x}\cdot\left[-\frac{2}{k}\left(\frac{\pi a}{k}\right)^{\frac{1}{2}}e^{2ika-i\pi/4}J_0(ka\sin\varphi)\right]. \quad (35)$$

The bracketed quantity in (35) is $f_d(\varphi)$, the diffraction-pattern contribution from double diffraction. Upon applying the cross-section theorem to it, and adding the result to πa^2, which was obtained from (30), we find

$$\frac{\sigma}{\pi a^2}=1-\frac{2\sin[2ka-\pi/4]}{\pi^{\frac{1}{2}}(ka)^{\frac{3}{2}}}. \quad (36)$$

This result, which was also obtained in a different way by Levine,[15] is shown in Fig. 11. Values of $\sigma/\pi a^2$ obtained from the exact solution of the corresponding boundary value problem for the reduced wave equation are also shown. The agreement can be seen to be quite good, even when the wavelength is twice the diameter.

For an aperture in a screen on which $\partial u/\partial n=0$ we must use (11) with a factor $(1+\rho_1^{-1}r)$ on the right side, to find the doubly diffracted field. We begin by obtaining from the first term in (23) at $r_1=2a$, $\theta_2=\pi/2$ the result

$$\frac{\partial u_{e1}}{\partial n}=-\frac{e^{2ika-i\pi/4}}{4(2\pi ka^3)^{\frac{1}{2}}}. \quad (37)$$

This is the normal derivative of the singly diffracted field on a ray which has crossed the aperture to the opposite edge. We now use (37) in (11) with the extra

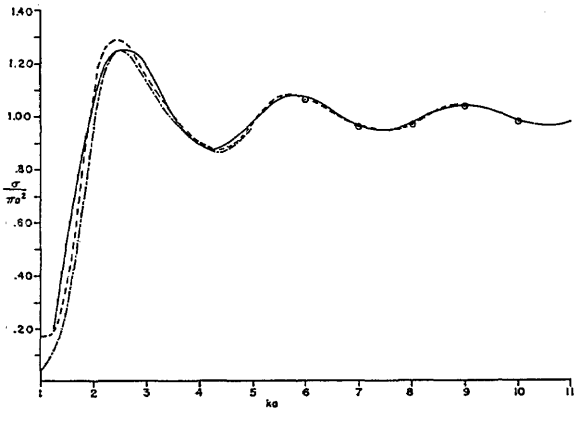

FIG. 11. The transmission cross section σ of a circular aperture of radius a in a thin screen on which $u=0$. The wave is normally incident. The solid curve, based on Eq. (36), results from single and double diffraction; the dashed curve also includes all multiple diffractions. The dots and the broken curve up to $ka=5$ are based upon the exact solution of the reduced wave equation.[13] The ordinate is $\sigma/\pi a^2$ and the abscissa ka.

[15] H. Levine, Institute of Mathematical Sciences, New York University, New York, Research Rept. EM-84 (1955).

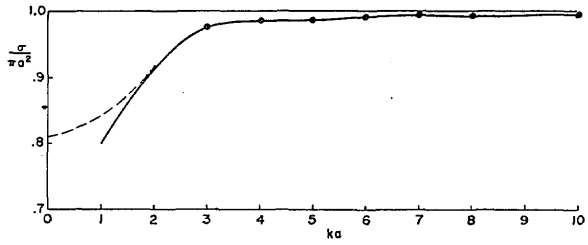

FIG. 12. The transmission cross section σ for normal incidence on a circular hole in a screen on which $\partial u/\partial n=0$. The ordinate is $\partial/\pi a^2$ and the abscissa is ka. The solid curve is based upon Eq. (40) which results from single and double diffraction. The encircled points and the dashed curve are the exact values computed by Bouwkamp.[13]

factor to obtain the doubly diffracted field which is

$$u_{e2}=-\frac{ie^{ik(r_1+2a)}}{16\pi(k^4a^3)^{\frac{1}{2}}}[r_1(1-a^{-1}r_1\cos\delta_1)]^{-\frac{1}{2}}\frac{\sin(\delta_1/2)}{\cos^2(\delta_1/2)}$$
$$-\frac{ie^{ik(r_2+2a)}}{16\pi(k^4a^3)^{\frac{1}{2}}}[r_2(1-a^{-1}r_2\cos\delta_2)]^{-\frac{1}{2}}\frac{\sin(\delta_2/2)}{\cos^2(\delta_2/2)}. \quad (38)$$

Proceeding as before, we evaluate (38) at points near the axis, then apply the caustic correction factor (28), and finally evaluate the result far from the screen. In this way we obtain

$$u_{e2}=-\frac{ke^{ikx}}{2\pi x}\left[\frac{\pi^{\frac{1}{2}}J_0(ka\sin\varphi)}{4k^{\frac{5}{2}}a^{\frac{1}{2}}}e^{2ika+i\pi/4}\right]. \quad (39)$$

To compute σ we apply the cross-section theorem which requires us to take the imaginary part of the bracketed expression in (39) at $\varphi=0$. We must add this contribution to σ from double diffraction to πa^2, obtained from (30). We also add the term $-\pi/4ka^2$ which was shown by Buchal and Keller[16] to come from the second term in the singly diffracted field u_{e1}. In this way we obtain

$$\frac{\sigma}{\pi a^2}=1-\frac{1}{4(ka)^2}+\frac{\sin(2ka+\pi/4)}{4\pi^{\frac{1}{2}}(ka)^{\frac{5}{2}}}. \quad (40)$$

This result was also obtained by Levine and Wu.[17] It is shown in Fig. 12 together with values based upon the exact solution of the reduced wave equation with $\partial u/\partial n=0$ on the screen. The agreement appears to be quite good for $ka\geqq 2$.

De Vore and Kouyoumjian[18] have used the method of this section to calculate the singly and doubly diffracted fields produced by a plane electromagnetic wave inci-

[16] R. N. Buchal and J. B. Keller, Commun. Pure Appl. Math. 8, 85 (1960).

[17] H. Levine and T. T. Wu, Tech. Rept. 71, Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California, (July, 1957).

[18] R. DeVore and R. Kouyoumjian, "The back scattering from a circular disk," URSI-IRE Spring meeting, Washington D. C. (May, 1961).

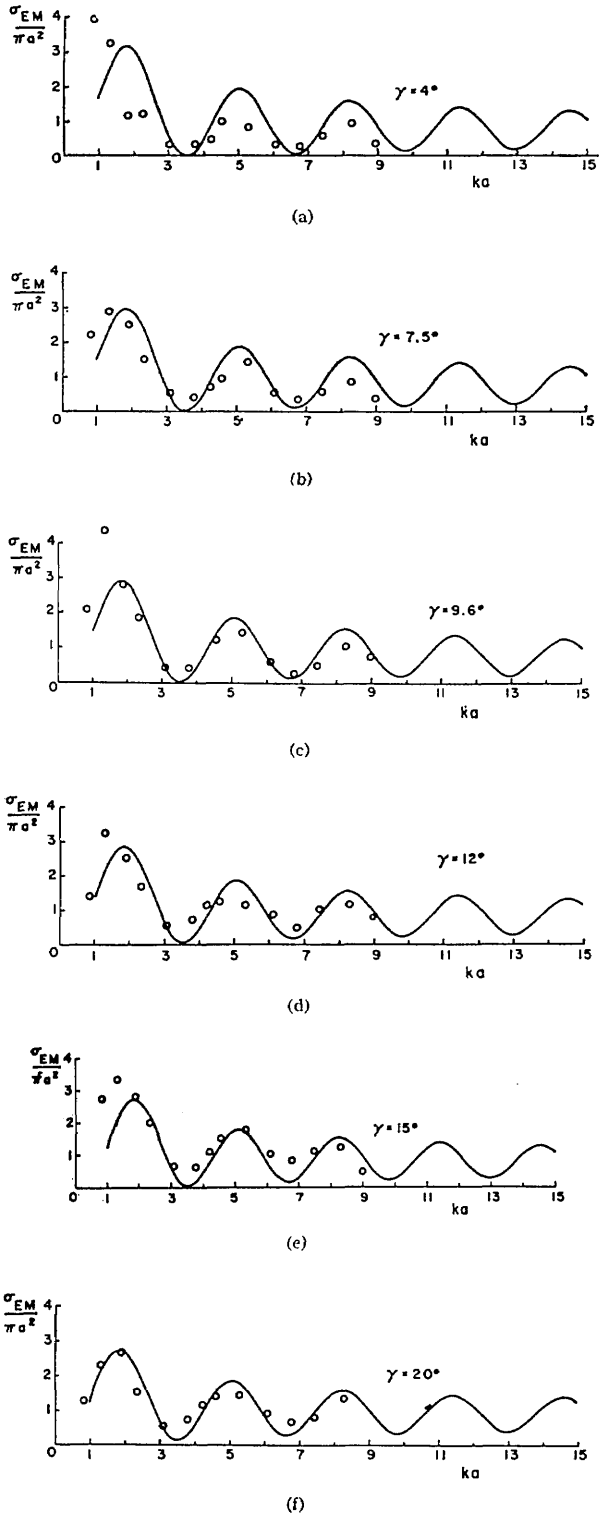

FIG. 13. Experimental and theoretical values of the electromagnetic back scattering cross section σ_{EM} of a finite circular metal cone for axial incidence. The ordinate is $\sigma_{EM}/\pi a^2$ and the abscissa is ka where a is the radius of the base of the cone. The circles are measured values and the curves are computed using the singly and doubly diffracted fields from the edge. The half-angle γ of the cone is indicated on each curve.

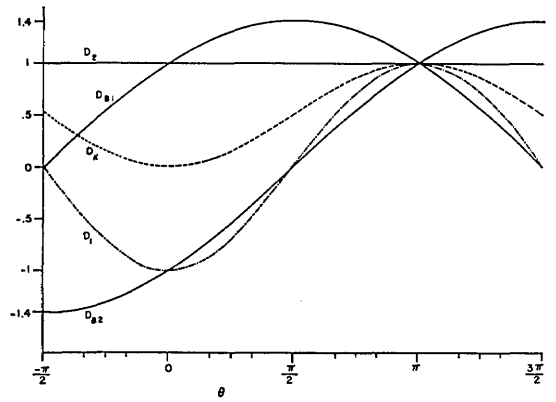

FIG. 14. Comparison of the Kirchhoff edge-diffraction coefficients D_1, D_2, and D_k with the Braunbek coefficients D_{B1} and D_{B2} for normal incidence ($\alpha=0$). The latter coincide with our D, given by Eq. (2) with the upper and lower signs, respectively. D_1 and D_2 result from the two usual modifications of the Kirchhoff theory and $D_k=\frac{1}{2}(D_1+D_2)$ comes from the usual form of it. The ordinate is $-2(2\pi k)^{\frac{1}{2}}e^{-i\pi/4}\sin\theta D$ and the abscissa is θ. All the coefficients are singular on the shadow boundary $\theta=\pi$ but the singular factor $\sin\theta$ has been removed. The dark side of the screen is at $\theta=3\pi/2$ and the illuminated side at $\theta=-\pi/2$.

dent from any direction on a perfectly conducting thin disk. They also performed accurate measurements of the diffracted field and found excellent agreement with the theory. Keller[19] has also used it to calculate the singly and doubly diffracted fields scattered by a perfectly conducting finite circular cone with a flat base hit by an axially incident plane wave. The back-scattered field comes primarily from the sharp edge at the rear of the cone and may be calculated by using the diffraction coefficient D given by (3) with the sign depending upon the orientation of the incident electric field. The results have been compared[20] with the measured values of Keys and Primich.[21] The comparison, shown in Fig. 13, indicates that the theory is fairly accurate even when ka is as small as unity. The theory has also been applied to objects of other shapes by Burke and Keller.[22]

It has been shown[6] by asymptotic evaluation, that the Kirchhoff theory, its two usual modifications and Braunbek's modification all lead to expressions for the field diffracted by an aperture in a thin screen, which can be interpreted as sums of fields on diffracted rays. The expression for the field on each ray was found to be exactly of the form (21), which is an additional corroboration of our theory. The Braunbek modification yielded exactly the diffraction coefficients (2) given by our theory, but the various other Kirchhoff methods gave different coefficients. The various coefficients are compared in Fig. 14 from which it can be seen that the

[19] J. B. Keller, IRE Trans. Antennas and Prop. AP-8, 175 (1960).
[20] J. B. Keller, IRE Trans. Antennas and Prop. AP-9, 411 (1961).
[21] J. E. Keys and R. I. Primich, Defense Research Telecommunications Establishment, Ottawa, Canada, Rept. 1010 (May, 1959).
[22] J. E. Burke and J. B. Keller, Research Rept. EDL-E49, Electronic Defense Laboratories, Sylvania Electronic Systems, Mountain View, California (April, 1960).

Kirchhoff results agree with ours in the forward direction, but differ considerably at other angles. Thus we may expect the Kirchhoff theory to lead to incorrect results at large diffraction angles.

The present method has been used by Keller[23] to determine the force and torque exerted on a rigid immobile thin disk or strip by an incident plane acoustic wave. In this case, if u denotes the acoustic pressure, the force on the disk is proportional to its total scattering cross section. By Babinet's principle this is found to be twice the transmission cross section of the complementary aperture. Similarly the torque on the disk is proportional to the angular derivative of the diffraction pattern of the disk, evaluated in the forward direction, and this pattern is the negative of that of the complementary aperture. For strips the same statements apply to the force and torque per unit length. Thus the results obtained above for σ yield the force directly. By differentiating the pattern $f(\varphi)$ with respect to φ at $\varphi=-\alpha$, we can also obtain the torque. Graphs of some of the results for the torque are shown in Figs. 15–17.

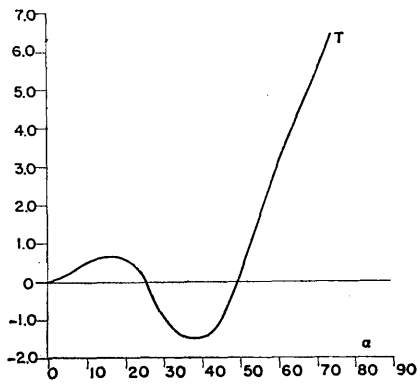

FIG. 15. The acoustic torque T per unit length on a strip of width $2a$ as a function of the angle of incidence α for $ka=5$. The vertical scale is the value of $T|a^2P^2/\rho_0c^2\pi^{\frac{1}{2}}(ka)^{\frac{1}{2}}|^{-1}$, where P is the amplitude of the incident pressure wave, ρ_0 is the density of the medium and c is its sound speed. The values of α at which $T=0$ are equilibrium angles which are stable if $T_\alpha<0$ and unstable if $T_\alpha>0$. The curve does not apply near $\alpha=\pi/2$, which corresponds to grazing incidence, at which $T=0$. T is an odd function of α.

[23] J. B. Keller, J. Acoust. Soc. Am. 29, 1085 (1957).

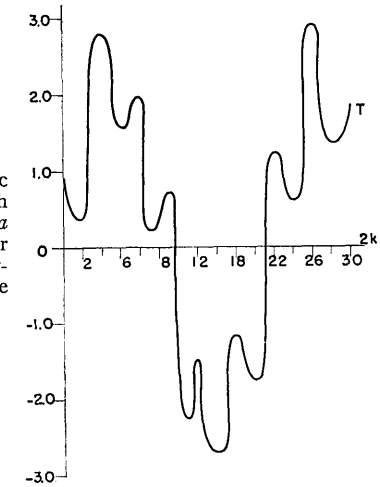

FIG. 16. The acoustic torque T per unit length on a strip of width $2a$ as a function of ka for $\alpha=\pi/4$ (45°). The vertical scale is the same as in Fig. 15.

5. CORNER OR TIP DIFFRACTION

Let us now consider diffraction by special points such as the corner of an edge of a boundary surface or the tip of a conical boundary surface. We call such points vertices and assume that an incident ray which hits one produces infinitely many diffracted rays which satisfy the *Law of vertex diffraction:* A diffracted ray and the corresponding incident ray may meet at any angle at a vertex of a boundary surface. This law follows at once from *Fermat's principle for vertex diffraction:* A vertex-diffracted ray from a point P to a point Q is a curve which has stationary optical length among all curves from P to Q passing through the vertex.

Both of these formulations show that a ray incident on a vertex produces a two-parameter family of diffracted rays which leave the vertex in all directions. In a homogeneous medium they are straight lines and the corresponding diffracted wave fronts are spheres with the vertex as their center. Thus the diffracted wave is spherical, as we might expect. This expectation is verified by the exact solution of the boundary-value problem for the reduced wave equation, corresponding to diffraction by a cone. This confirmation lends support to the law of vertex diffraction. Additional support is provided by the asymptotic evaluation[6] for short wavelengths of the various forms of the Kirchhoff approximation for diffraction through an aperture in a thin screen. This evaluation shows that the field at any point contains a sum of terms, one for each vertex on the edge of the aperture. Each of these terms corresponds to a vertex-diffracted ray, in agreement with our theory.

To determine the field on a vertex-diffracted ray let us still consider a homogeneous medium. Then the phase on such a ray at the distance r from the vertex is $kr+\Psi_i$ where Ψ_i is the phase of the incident field at the vertex. The amplitude varies as r^{-1} since the cross-sectional area of a tube of rays is proportional to r^2. Therefore we write the field on the diffracted ray as

$$u=Cu_i(e^{ikr}/r). \tag{41}$$

Here C is the appropriate vertex or corner diffraction coefficient. It depends upon the directions of the incident and diffracted rays, the local geometry of the vertex and the local properties of the media at it. Dimensional considerations show that C is proportional to a length, so it must be proportional to k^{-1}.

The field diffracted by any cone when a plane wave is incident is of the form (41), as we find by analyzing the boundary value problem for the reduced wave equation. This not only confirms our theory but enables us to determine C when the boundary-value problem can be solved. It has been solved for elliptic cones, including the plane angular sector, by Kraus and Levine,[24] but C

[24] L. Kraus and L. Levine, Commun. Pure Appl. Math. 14, 49 (1961).

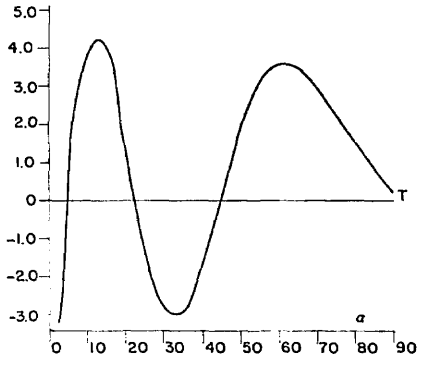

FIG. 17. The acoustic torque T on a circular disk of radius a as a function of the angle of incidence α for $ka=5$. The vertical scale is the value of $T|P^2a^3/\rho_0 c(ka)^2|^{-1}$. P, ρ_0, and c are defined in the caption of Fig. 15. The values of x at which $T=0$ are equilibrium values, stable if $T_\alpha<0$ and unstable if $T_\alpha>0$. The curve does not apply near $\alpha=0$ (normal incidence) nor $\alpha=\pi/2$ (grazing incidence).

has not been evaluated. For circular cones, it has been evaluated by Felsen[25] and by Siegel *et al.*[26] The various forms of the Kirchhoff integral for diffraction through an aperture, when evaluated asymptotically,[6] yield fields of the form (41) for each corner on the aperture edge. Here again the resulting expressions for C are different for the different versions of the Kirchhoff method, and none can be expected to coincide with the exact value contained in the solution of Kraus and Levine. We shall not consider any applications of vertex diffraction.

6. SURFACE-DIFFRACTED RAYS

Let us now consider an incident ray which grazes a boundary surface, i.e., is tangent to the surface. We assume that such a ray gives rise to a surface-diffracted ray in accordance with the *law of surface diffraction:* An incident ray and the resulting surface diffracted ray in the same medium are parallel to each other at the point of diffraction and lie on opposite sides of the plane normal to the ray at this point. When the two rays lie in different media, they obey the law of refraction.

The surface ray travels along the surface in a manner determined by the usual differential equations for rays on a surface. Therefore in a homogeneous medium it is an arc of a geodesic or shortest path on the surface. At every point the surface ray sheds a diffracted ray satisfying the law of surface diffraction. All of these properties of surface-diffracted rays follow from *Fermat's principle for surface diffraction:* A surface-diffracted ray from a point P to a point Q is a curve which makes stationary the optical length among all curves from P to Q having an arc on the boundary surface.

To illustrate these ideas, let us consider a boundary surface which is a cylinder parallel to the z axis. Let the intersection of the cylinder with the xy plane be a smooth closed convex curve C and suppose that the medium surrounding the cylinder is homogeneous. To find a surface-diffracted ray from P to Q, both of which lie outside C, we observe that the optical length of a curve is proportional to its geometric length. Therefore to utilize Fermat's principle we imagine a string from P to Q, and consider it to be pulled taut. Then it will consist of two straight-line segments from P and Q to the cylinder and a geodesic arc along the cylinder. When P and Q are both in the xy plane the arc is just an arc of C; otherwise it is a helical arc along the cylinder. It may wrap around the cylinder any number of times in either direction, so there are countably many surface diffracted rays from P to Q.

We define a surface-diffracted wavefront to be a surface orthogonal to a family (i.e., normal congruence) of surface-diffracted rays. In the example just described let us suppose that P is a line source parallel to the z axis. Then the surface-diffracted wavefronts which it produces are cylinders with generators parallel to the z axis. Their intersections with the xy plane are the involutes of the curve C.

Two examples of surface-diffracted rays are shown in Figs. 18 and 19. In both cases the grazing ray is incident horizontally from the left. One of the shed surface-diffracted rays is shown in Fig. 18 and two are shown in Fig. 19. In both cases the figures show cross sections of opaque screens surrounded by homogeneous media.

To construct the field on a surface diffracted ray we assume that the phase of the field increases in proportion to optical length along the ray. The amplitude is assumed to be proportional to the amplitude of the incident field at the point of diffraction. The coefficient of proportionality involves a new surface-diffraction coefficient B which depends upon local properties of the boundary surface and the media at the point of diffraction. Along the portion of the diffracted ray on the surface, the amplitude is assumed to vary in accordance with the principle of energy conservation in a narrow strip of rays on the surface. Thus it varies inversely as the square root of the width of this strip because of

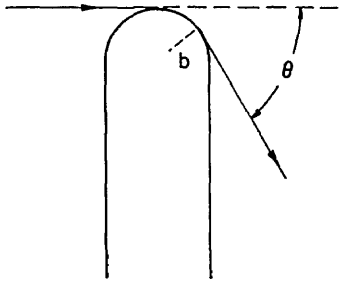

FIG. 18. Cross section of a screen of width $2b$ with a rounded end of radius b. A plane wave is normally incident upon it from the left. The dashed line is the shadow boundary. An incident ray which grazes the end of the screen is shown together with one of the surface diffracted rays which it produces in the shadow region. The angle between this ray and the shadow boundary is θ.

[25] L. B. Felsen, J. Appl. Phys. **26**, 138 (1955).
[26] K. M. Siegel, J. W. Crispin, and C. E. Schensted, J. Appl. Phys. **26**, 309 (1955).

convergence or divergence. However, it also decays because of radiation from the surface along the shed diffracted rays. The decay rate is assumed to be determined by a decay exponent α which depends upon local properties of the surface and media. The amplitude on a shed diffracted ray varies in the usual manner and is proportional to the field on the surface diffracted ray at the point of shedding. The proportionality factor involves another diffraction coefficient which can be shown to be the same as the coefficient B introduced above as a consequence of the principle of reciprocity. The details of this construction of the field are given for cylinders by Keller[27] and for three-dimensional surfaces by Levy and Keller.[28]

The theory has been tested by applying it to diffraction of a cylindrical wave by a circular cylinder and comparing the predicted field with the asymptotic expansion of the exact solution. The two results agreed exactly when appropriate expressions were used for the diffraction coefficient B and the decay rate α. Thus the theory was confirmed and B and α were determined. The theory was further confirmed by making similar comparisons for diffraction by a parabolic cylinder, an elliptic cylinder (Levy[29]), a spheroid (Levy and Keller[30]) and the screen of Fig. 19 (Magiros and Keller[31]). Of course, in all cases the same values of B and α were employed.

A numerical test of the accuracy of the theory was made by Keller[32] for the screen of Fig. 19. The field u far from the end of the screen in the shadow region can be written as $u=f(\theta,kb)e^{ikr+i\pi/4}(kr)^{-\frac{1}{2}}$. The function $f(\theta,kb)$ was evaluated for $\theta=\pi/4$ as a function of kb from the exact solution of the boundary-value problem and from the formula given by our theory. The results are shown in Fig. 20. The upper curves and points apply to a screen on which $\partial u/\partial n=0$; the lower ones to a screen on which $u=0$. The solid curves are obtained when a simple

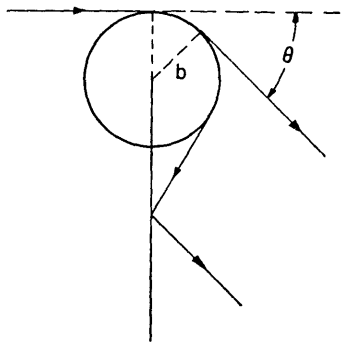

FIG. 19. Cross section of an infinitely thin screen with a cylindrical tip of radius b. A plane wave is normally incident upon it from the left. The dotted line is the shadow boundary. An incident ray which grazes the tip is shown together with two of the surface diffracted rays which it produces. One of them is reflected from the screen, and both ultimately make the angle θ with the shadow boundary.

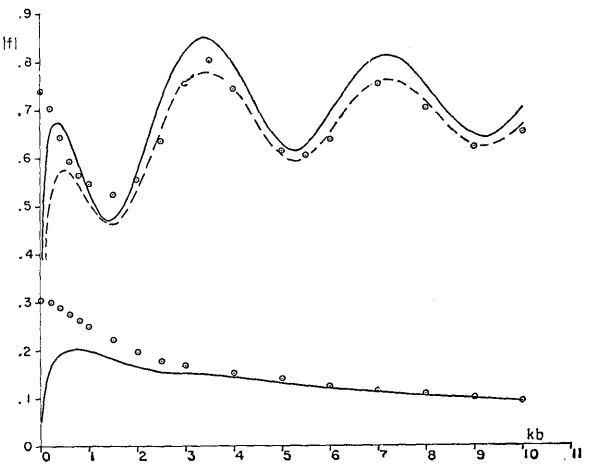

FIG. 20. The far-field amplitude $|f(\pi/4,kb)|$ in the shadow of the screen of Fig. 19 in the direction $\theta=\pi/4$ as a function of kb. The upper curves and points apply to a screen on which $\partial u/\partial n=0$ while the lower ones pertain to a screen on which $u=0$. The encircled points were obtained by laborious numerical computation of the series solution of the boundary-value problem; the curves were obtained from the formulas given by the geometrical theory of diffraction. The dashed curve was obtained from a formula in which an improved expression for α was used.

expression is used for the decay exponent α and the dashed curve when a more accurate expression is used for α. The agreement between the curves given by our theory and the points from the exact solution appears to be quite good for $kb>2$. Similar curves for the screen of Fig. 18 are shown in Fig. 21, but there is no exact solution with which to compare them.

Diffracted rays have also been introduced by Franz and Depperman[33] who called the associated waves "creeping waves." They applied them to explain oscillations in the measurements of the radar back-scattering cross section of metallic circular cylinders. A refinement of the method described above for determining B and α was given by Levy and Keller.[34] They showed how variation of the curvature of the diffracting surface modifies the values of B and α determined from a circular cylinder. The determination of the field near the diffracting surface requires special considerations, described in references 27 and 28, because the surface is a caustic of the diffracted rays. Another special treatment described by Buchal and Keller[16] is required near the shadow boundary. Near the point of diffraction, where the shadow boundary meets the diffracting surface, a still different special analysis due to V. Fock and to C. L. Pekeris is required. A uniform expression for the field in these various regions has been obtained, in two-dimensional cases, by Logan and Yee[35] by combining

[27] J. B. Keller, IRE Trans. Antennas and Propagation AP-4, 243 (1956).
[28] B. R. Levy and J. B. Keller, Commun. Pure Appl. Math. 12, 159 (1959).
[29] B. R. Levy, J. Math. and Mech. 9, 147 (1960).
[30] B. R. Levy and J. B. Keller, Can. J. Phys. 38, 128 (1960).
[31] D. Magiros and J. B. Keller, Commun. Pure. Appl. Math. 14, 457 (1961).
[32] J. B. Keller, J. Appl. Phys. 30, 1452 (1952).
[33] W. Franz and K. Depperman, Ann. Physik 10, 361 (1952).
[34] B. R. Levy and J. B. Keller, IRE Trans. Antennas and Propagation, AP-7, 552 (1959).
[35] N. A. Logan and K. S. Yee, Symposium on Electromagnetic Theory, U. S. Army Mathematics Research Center, University of Wisconsin, Madison, Wisconsin (April, 1961).

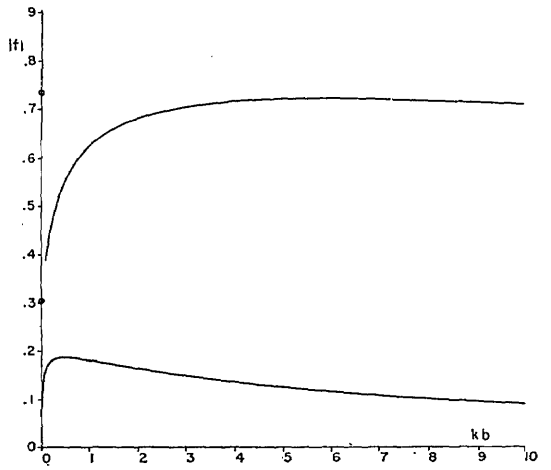

FIG. 21. The far-field amplitude $|f(\pi/4,kb)|$ in the shadow of the screen of Fig. 18 in the direction $\theta=\pi/4$ as a function of kb. The curves are based upon the geometrical theory of diffraction while the points at $kb=0$ are determined from the solution for diffraction by a half-plane. The upper curve and point apply to a screen on which $\partial u/\partial n=0$; the lower ones to a screen on which $u=0$. Probably only the decreasing portions of the curves are correct, and most likely the amplitude decreases monotonically for all values of kb.

the method of Fock and Pekeris with the geometrical considerations we have described.

7. FURTHER DEVELOPMENTS

The geometrical theory of diffraction which we have described has been applied to inhomogeneous media by Friedrichs and Keller[36] and by Seckler and Keller.[37] It has also been extended by the introduction of complex or imaginary rays.[2]

A similar theory can be constructed to describe any kind of wave propagation and this has been done to some extent for water waves, elastic waves, quantum-mechanical waves, surface waves, etc. When anisotropic media are present, or when more than one propagation velocity exist, there are more kinds of rays and the theory is correspondingly more complicated. However the principles are essentially unchanged.

From a mathematical point of view, the field constructed by means of the present theory is the leading part of the asymptotic expansion of the exact field for small values of λ or large values of k. The full asymptotic expansion consists of additional terms in the amplitude of the field on each ray. These terms are smaller than the first term by factors of k^{-n}, $n=1, 2, \cdots$. These statements have been proved in special cases, but not in general.

[36] K. O. Friedrichs and J. B. Keller, J. Appl. Phys. 26, 961 (1955).

[37] B. D. Seckler and J. B. Keller, J. Acoust. Soc. Am. 31, 192 (1958).