BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 84, Number 5, September 1978

RAYS, WAVES AND ASYMPTOTICS[1]

BY JOSEPH B. KELLER

1. Introduction. In 1929 the American Mathematical Society established an annual lectureship named after Josiah Willard Gibbs (1839–1903), Professor of Mathematical Physics at Yale University from 1871 to 1903. Gibbs contributed essentially to the development of statistical mechanics and physical chemistry, and invented vector analysis. Therefore, it is appropriate that these lectures concern "mathematics or its applications" and "the contribution mathematics is making to present-day thinking and to modern civilization."

In this fiftieth Gibbs lecture, I will try to fulfill these objectives by describing some developments in the field of wave propagation. I hope that they will show also how mathematics itself is enriched by interaction with scientific and technical problems. In keeping with the intention that the lectures be "of a semipopular nature," I will omit as much technical detail as possible.

At first I was especially pleased that this is the fiftieth Gibbs lecture, because 50 is so special in our number system. This is because it is the product of the number of fingers on one hand multiplied by the number of fingers on two hands. But from this point of view, 50 is not a dimensionless number, since it has the dimensions of (fingers per hand) squared. Therefore, its numerical value depends upon the choice of units, so it has no intrinsic significance. This is a reminder that it is only dimensionless numbers which we can regard as large or small, as in the asymptotic analysis I am going to discuss later.

My plan is to begin with light rays and to describe their theory and use in optics. Then I will demonstrate some of their properties with the aid of a laser, kindly lent to me by Arthur Schawlow of Stanford University. Next, I will explain how rays were displaced by waves, which were introduced to provide a more accurate description of observed phenomena. The wave theory required solving certain partial differential equations, and numerous methods were devised to do this in special cases. However, in all other cases

The fiftieth Josiah Willard Gibbs Lecture presented at the 83rd Annual Meeting of the American Mathematical Society, St. Louis, Missouri, January 27, 1977; received by the editors October 3, 1977.

AMS (MOS) subject classifications (1970). Primary 35-02, 35A35, 35B40, 78-02, 78A05, 78A25, 78A45.

Key words and phrases. Diffraction, geometrical theory of diffraction, wave propagation, asymptotic expansions, ray theory, partial differential equations, optics, electromagnetic theory.

[1]Supported by the Office of Naval Research, Contract N00014-76-C0010, the Air Force Office of Scientific Research, Contract AFOSR-76-2884 and the National Science Foundation Grants MCS75-09837 and MCS75-23332-A02.

this was a major difficulty. Ultimately rays, generalized in various ways, provided a method for solving such equations asymptotically. Thus, asymptotics provided the reconciliation of the ray and wave theories. Finally, I shall indicate the broader implications of this story for mathematics and for science in general.

2. Rays. Rays were originally defined in Optics, the science of light, as the paths along which light travels. They were of three kinds, direct, reflected and refracted, characterized as follows (see Figure 1):

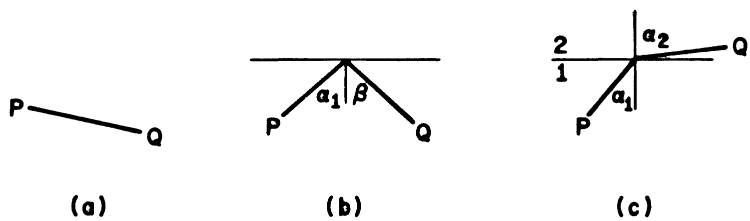


FIGURE 1. a. The direct ray from P to Q in a homogeneous medium is a straight line.
b. A cross-section of the shadow showing the predicted bright spot on the axis. When this bright spot was observed by Aarago, Poisson became an ardent proponent of the wave theory.
c. A ray from P in medium 1 is shown hitting the interface between media 1 and 2. The refracted ray in medium 2 is shown. The angle of refraction α_2 is related to the angle of incidence α_1 by Snell's law.

1. A direct ray in a homogeneous medium is a straight line.
2. A reflected ray is produced when a ray is incident upon a smooth surface. In a homogeneous medium it is also a straight line. The angle β between the reflected ray and the normal to the surface is equal to the angle α_1 between the incident ray and the normal. Furthermore, the reflected ray lies in the plane determined by the normal and the incident ray, and is on the opposite side of the normal from the incident ray.
3. A refracted ray is produced when a ray in one medium, say medium 1, is incident upon the interface between medium 1 and another medium, say medium 2. It lies in medium 2 and is a straight line if medium 2 is homogeneous. The refracted ray lies on the opposite side of the normal from the incident ray, in the plane containing the incident ray and the normal. The angle α_2 between the refracted ray and the normal is related to α_1 by Snell's law:

$$\sin \alpha_2 / \sin \alpha_1 = n_1 / n_2.$$

Here n_i, the index of refraction of medium i, is a characteristic property of medium i ($i = 1, 2$). It is a constant if medium i is homogeneous.

These characterizations of the rays may be called the laws of propagation, reflection and refraction respectively. Euclid stated the first two of them, but omitted the coplanarity part of the law of reflection. This was first added by the Arabic scientist Alhazen in the tenth century. Snell's law was not discovered until 1626. Much earlier Ptolemy had stated that $\alpha_2/\alpha_1 = n_1/n_2$, which is the form that Snell's law takes when α_1 and α_2 are small. In 1637 Descartes also published Snell's law, but it is not clear how he obtained it–perhaps by reading Snell's paper.

These three laws are the basis of Geometrical Optics, the science of the propagation of light, in piecewise homogeneous media. They suffice for the design and analysis of mirrors, lenses and other optical instruments. They also provide the foundation for the applications of Geometrical Optics, such as those published by Gauss in 1846. (At this point, rays, lenses, and light pipes were demonstrated.)

3. The Principle of Least Time. Euclid knew that a straight line is the shortest path between two points. Therefore, he also knew that the direct ray between two points P and Q in a homogeneous medium is the shortest path from P to Q. Later Heron, an Alexandrian, showed that the shortest path from P to a plane boundary and then to Q consists of an incident ray from P to the boundary plus a reflected ray satisfying the law of reflection, from the boundary to Q, provided that P and Q lie on the same side of the boundary.

These two results were generalized by Fermat in 1661. He considered the line integral L of the index of refraction along any path from P to Q:

$$L = \int_P^Q n\, ds.$$

Here ds denotes the element of arclength. Fermat stated that light travels from P to Q along that path which minimizes L. Now n is inversely proportional to the velocity of light so $n\, ds$ is proportional to the time required for light to travel the distance ds. Therefore, L is proportional to the time required for light to travel along the path from P to Q. Consequently, Fermat called his characterization of rays the Principle of Least Time.

This principle implies that in a homogeneous medium, the light path or direct ray from P to Q is a straight line. It also implies that in a homogeneous medium with a plane boundary, light follows a straight line from P to the boundary and another straight line from the boundary to Q, with the law of reflection obeyed. To obtain this conclusion, we must consider only paths with a point on the boundary, and then use Heron's result. Fermat also showed that when P and Q lie on opposite sides of a plane interface between two media, his principle yields an incident and refracted ray satisfying the law of refraction.

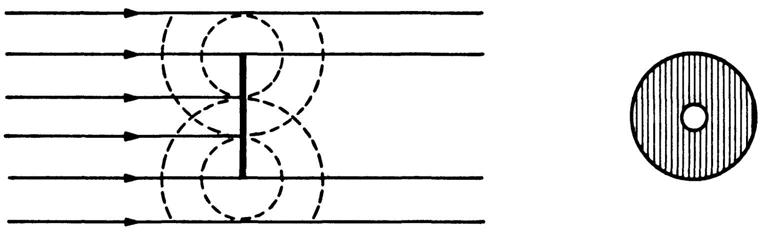
Thus, Fermat's principle implies the three laws of geometrical optics in piecewise homogeneous media. It also provides a characterization of rays in more general inhomogeneous media, for which $n(x)$ is not piecewise constant. Therefore, it provides a succinct basis for geometrical optics in arbitrary media.

When curved surfaces occur, the ray path determined by the laws of geometrical optics sometimes makes L a maximum rather than a minimum. In other cases it just makes L stationary. Consequently, the correct characterization of the rays, which includes all these cases, is that the rays are the paths which make L stationary. Therefore, Fermat's principle should be the Principle of Stationary Time. Fermat knew this, but ignored it–perhaps because "least" is shorter than "stationary", or because he wanted to show that God or Nature is economical, as we should all be. Only in the twentieth century did Carathéodory partially justify Fermat's use of "least" by showing that every sufficiently short portion of a light ray does minimize L among all paths between its endpoints.

In 1833 Hamilton provided still other ways of describing rays. He introduced Hamilton's equations, various characteristic functions, the eiconal equation, etc.

4. Waves. Because geometrical optics was so successful in describing the propagation of light, some scientists sought to explain why light obeyed its laws. Furthermore, new optical phenomena were discovered which were not accounted for by geometrical optics. First there was diffraction, which is the occurrence of light where there should be none according to geometrical optics, observed by Grimaldi in 1665. Then there were polarization phenomena found by Bartholinus in 1670, the finite speed of light measured by Römer in 1675 and the colors which appeared when thin plates were illuminated by white light, described by Newton in 1704. To account for these things, the theory that light was a wave phenomenon was proposed by Huygens (1690) and developed by Newton (1704), Young (1802) and Fresnel (1819).

In 1818 Poisson devised a reductio ad absurdum argument to show that the wave theory was untenable. He considered an opaque circular disk, such as a coin, illuminated from one side by a distant source of light. See Figure 2.



(a) (b)

FIGURE 2. a. The experiment proposed by Poisson to refute the wave theory of light consists of an opaque circular disk illuminated by a distant light source. Some rays from the source are shown, together with the shadow of the disk, which is devoid of rays. In addition waves spreading out from the edge of the disk are shown. These waves will arrive in phase at points on the axis of the disk, making the axis bright. Poisson considered this unreasonable conclusion a death blow to the wave theory.

b. A cross-section of the shadow the predicted bright spot on the axis. When this bright spot was observed by Aarago, Poisson became an ardent proponent of the wave theory.

According to the wave theory, waves would emanate from all points on the rim of the disk and they would spread into the shadow of the disk. They would all arrive at the axis in the same phase, so the axis would be very bright. Therefore, a cross-section of the shadow would contain a bright spot at its center. Since this is ridiculous, the wave theory must be wrong. When Arago performed this experiment in 1818 and found the bright spot, Poisson was immediately converted and began working enthusiastically on the wave theory.

In the search for the correct equations to describe light waves, various wave equations were proposed, and the equations of elasticity were discovered. The culmination of the search was the realization by Maxwell that light was an electromagnetic phenomenon. Therefore light waves were governed by Maxwell's equations (1864), a set of partial differential equations which, he correctly concluded, were satisfied by all electromagnetic fields. The property that distinguishes light from other electromagnetic waves is their short wavelength λ, which ranges from about $\lambda = 4 \times 10^{-5}$ centimeters for violet light to about $\lambda = 6 \times 10^{-5}$ centimeters for red light. The corresponding vibration frequencies ν range from about $\nu = 5 \times 10^{16}$ cycles/second for red to about $\nu = 7.5 \times 10^{16}$ cycles/second for violet.

5. The task of solving partial differential equations. Once the equations governing a phenomenon such as light have been found, there arises the task of solving those equations in particular cases. If there are many solutions, as there are for partial differential equations, the solution which is appropriate to a particular case must be found. This solution is characterized by auxiliary conditions, such as initial conditions, boundary conditions, regularity conditions, radiation conditions, etc. These conditions themselves are determined by the circumstances of the particular case–the location of the boundaries, the physical properties of the boundaries, the properties of the light sources in optical problems, etc. The combination of one or more partial differential equations and a set of auxiliary conditions is called a problem.

From the beginning of the nineteenth century to the present time, various methods have been devised to solve such problems. The basis for many of them is the technique of separation of variables. This technique applies to partial differential equations which possess special solutions which are products of functions of one variable each. When this is the case, each of the factors satisfies a separated ordinary differential equation. Most of the special functions of analysis arose as solutions of such separated equations, e.g., those of Bessel, Hankel, Legendre, Hermite, Lamé, Laguerre, Whittaker, etc.

If the equation has "enough" product solutions, it may be possible to represent the solution of a particular problem as an integral or series of them with suitable coefficients. The theory of transforms–Fourier, Laplace, Hankel, Mellin, Lebedev, etc.–and the theory of orthogonal functions, were developed to determine these coefficients.

When these methods are applicable, they yield an integral representation or a series representation of the solution. For certain values of the parameters in the problem, it may be easy to evaluate the integral or series. In wave propagation problems this is usually the case when the wavelength λ is large compared to other lengths in the problem. However, when λ is small compared to them, as is the case in optics, it is usually very difficult to evaluate the integral or series. To overcome this difficulty, methods were devised for the asymptotic evaluation of integrals–Kelvin's method of stationary phase, the method of steepest descent, etc.–and of series–Poisson's summation formula, the Watson transformation, the theory of alternative representations, etc.

The foregoing methods, based upon separation of variables, are applicable only when the boundary conditions are imposed upon a complete coordinate surface in a coordinate system in which the partial differential equation is separable. In the late nineteenth century a systematic search was begun by

Michel (1890) and Bocher (1894) for all the coordinate systems in which Laplace's equation is separable. This work was continued and extended to other equations by Robertson (1927), Eisenhart (1934), Redheffer (1948) and others. The conclusion was that each equation is separable in only a limited set of coordinate systems. For example, in three dimensions Laplace's equation $\Delta u = 0$ is separable only when the coordinate surfaces are confocal cyclides and limiting forms of them. These are certain fourth degree surfaces which include quadratics, and are most simply described by quadratic equations in pentaspheric coordinates. The reduced wave equation

$$(\Delta + k^2)u = 0$$

is separable only when the coordinate surfaces are confocal quadrics and their degenerate forms. This yields just 11 coordinate systems.

If the boundary surfaces are not complete coordinate surfaces of a separable coordinate system, the methods described so far do not apply. Then in some cases eigenfunction expansions may be used, but some means must be found to construct the eigenfunctions and to find the eigenvalues. For this purpose variational methods were devised, such as that of Rayleigh and Ritz. Another method is the conversion of the problem into an integral equation by means of Green's functions. But then it is necessary to find some way to solve the integral equation. In very special cases the Wiener-Hopf method can be used for this purpose, but not in general.

6. Approximation methods and existence theory. Because of these difficulties in constructing explicit solutions of problems, numerous methods of approximating solutions have been employed. Some of the most successful of them are based upon physical intuition, such as the Kirchhoff method for treating wave propagation problems involving short waves, and the static method for corresponding problems involving long waves. Gradually various systematic approximation methods have been developed. These include the regular perturbation method, singular perturbation theory, boundary layer theory, the use of asymptotic expansions, the method of matched asymptotic expansions, series truncation methods, etc. With the advent of high speed computers, a host of numerical methods have been implemented, such as the method of finite differences, the finite element method, the method of lines, the collocation method, etc.

In those special cases in which explicit solutions can be found, there is no question about the existence of a solution. Furthermore, the method of construction of the solution often shows that it is unique. In addition, the expression for the solution usually shows that it depends continuously upon the data of the problem. When a problem has these three properties–that a solution exists, is unique, and depends continuously on the data in a suitable norm–it is said to be well posed, properly posed or properly formulated, a concept introduced by Hadamard in 1921. For the great majority of problems, which cannot be solved explicitly, there arises the question of whether they are well posed.

To answer this question, the existence and uniqueness theory of partial differential equations was developed, together with the tool of functional analysis. Sometimes engineers and scientists, especially when studying mathematics, criticize this theory. They point out that the physical phenomenon which the problem describes is properly posed, so that it is unnecessary to prove it. What they fail to realize is that the purpose of the theory is to determine whether the mathematical problem does describe the physical phenomenon. It can do so only if it is well posed, because the physical phenomenon is well posed. A more appropriate criticism is that the theory stops with the demonstration of well-posedness, rather than proceeding to the determination of specific properties of the solution. Regularity theory, which concerns the degree of smoothness of the solution, is a first step in this direction.

Another question arises about approximations to solutions constructed by an approximation method. The question is "How close is the approximation to the solution?" Usually this question is answered by giving an "order estimate" of the error. Rarely the constant in this estimate can be found also. In other cases the approximate solution can be compared with explicit solutions of special problems and with experimental results, as we shall see.

7. The Geometrical Theory of Diffraction. In 1953 I introduced a new method, the Geometrical Theory of Diffraction, for solving approximately problems of wave propagation. The method is intended to apply to problems in which the wavelength λ is small compared to any other length a in the problem, so that $\lambda/a \ll 1$. In terms of the propagation constant or wavenumber $k = 2\pi/\lambda$, this condition is $ka \gg 1$. If a is used as the unit of length, the condition for validity of this method is just $\lambda \ll 1$ or $k \gg 1$. As we shall see, the method is useful not only in this range of λ, but also far outside it. Furthermore, it can be applied to any linear partial differential equation or system of such equations.

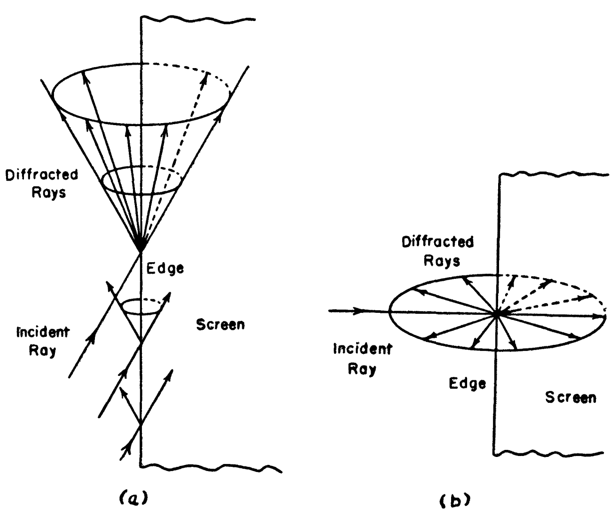


FIGURE 3. a. The cone of diffracted rays produced by an incident ray which hits the edge of a thin screen obliquely.

b. When the incident ray is normal to the edge, the cone of diffracted rays opens up to become a plane of diffracted rays.

The basic idea is that short waves propagate along rays, as is the case in geometrical optics. However, in addition to the three kinds of rays considered in geometrical optics, there are other kinds of rays which I called diffracted rays and complex rays. They are characterized as follows:

4. An *edge diffracted ray* is produced when a ray hits an edge of a boundary or interface. In a homogeneous medium it is a straight line. The angle between the diffracted ray and the edge is equal to the angle between the incident ray and the edge, when both rays lie in the same medium. Otherwise, the angles between the two rays and the plane normal to the edge are related by Snell's law. Furthermore, the diffracted ray lies on the opposite side of the normal plane from the incident ray. From this law of edge diffraction it follows that one incident ray produces a cone of edge diffracted rays. See Figure 3.

5. A *vertex diffracted ray* is produced when a ray hits a vertex of a boundary or interface. In a homogeneous medium it is a straight line. The vertex diffracted rays leave the vertex in all directions, which is the law of vertex diffraction. It's not much of a law, but at least it's democratic. It follows that in three dimensions a single incident ray produces a two parameter family of vertex diffracted rays.

6. A *surface diffracted ray* is produced when a ray is incident tangentially on a smooth boundary or interface. It is a geodesic on the surface in the metric $n\,ds$, where n is the refractive index of the medium on the side of the surface containing the incident ray. The surface ray is tangent to the incident ray. At every point it sheds a diffracted ray along its tangent. See Figure 4. A surface diffracted ray is also produced on side 2 of an interface by a ray incident from side 1 at the critical angle $\sin^{-1}(n_1/n_2)$, which is real if $n_1/n_2 < 1$. In this case at every point it sheds rays back toward side 1 at the critcial angle.

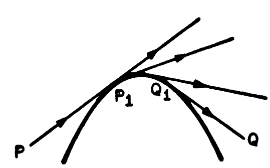


FIGURE 4. An incident ray from P hits a smooth surface tangentially at P_1 and produces a surface diffracted ray. This ray is a geodesic on the surface. At each point it sheds a diffracted ray along its tangent.

7. A *complex ray* is a complex curve which satisfies the equations characterizing a ray. Such a ray can be defined only if $n(x)$ is analytic or piecewise analytic. In a homogeneous medium a complex ray is a complex straight line. There are also complex reflected, refracted, edge diffracted, vertex diffracted and surface diffracted rays, provided that the boundaries, interfaces and edges are analytic or piecewise analytic.

These seven kinds of rays are the building blocks for more complicated ray paths which are composed of them. Thus, for example, an incident ray may hit a surface to produce a reflected ray which in turn hits another surface to produce a doubly reflected ray, or hits an edge to produce an edge diffracted ray, etc. All the direct, singly, and multiply reflected, refracted, and diffracted real and complex rays are taken into account in the Geometrical Theory of Diffraction.

Before explaining how the rays are used, we shall present another characterization of them.

8. Variational characterization of rays. The principle of stationary time, described in §3, can be extended to yield the diffracted and complex rays as well as the ordinary rays. To extend it we shall introduce seven classes of curves $\mathcal{C}_j, j = 1, \ldots, 7$, one class corresponding to each of the seven types of rays. Then a ray of type j is defined to be a curve which makes the optical length $L = \int_P^Q n\, ds$ stationary in class $\mathcal{C}_j, j = 1, \ldots, 7$.

The direct rays from P to Q are the stationary curves in the class $\mathcal{C}_1$ of all piecewise smooth real curves from P to Q which have no points on the boundaries or interfaces. The incident plus reflected ray paths from P to Q are stationary in $\mathcal{C}_2$, which contains curves with one point on a boundary or interface, but which do not cross an interface. The incident plus refracted ray paths are stationary in $\mathcal{C}_3$, which also contains curves with one point on an interface but which do cross the interface.

The incident plus edge diffracted ray paths are stationary in $\mathcal{C}_4$, containing curves with one point on an edge. Incident plus vertex diffracted paths lie in the class $\mathcal{C}_5$ of curves with one point on a vertex. Incident plus surface diffracted ray paths lie in the class $\mathcal{C}_6$ of curves with an arc on a boundary or interface. Complex direct rays lie in the class $\mathcal{C}_7$ of complex curves from P to Q with no points on boundaries or interfaces. Other complex rays can be defined similarly.

The rays defined by this variational principle are exactly the same as those defined explicitly before. This principle makes precise the definition of a ray in an inhomogeneous medium. We did not make that explicit before because it involves certain ordinary differential equations–either Lagrange's or Hamilton's equations–which are just the Euler equations corresponding to the optical length integral.

The variational principle can also be extended to include rays which are multiply reflected, refracted and/or diffracted. For this purpose it is merely necessary to introduce other classes of curves, such as that with r points on edges, s points on vertices, and t points or arcs on boundaries or interfaces. The stationary curves in this class will yield rays which have been diffracted r times by edges, s times by vertices and have had a combination of t reflections, refractions or diffractions by surfaces. By considering the rays in all such classes of curves, we obtain all the rays from P to Q.

9. Asymptotic construction of solutions. We shall now show how to use the rays to construct approximate solutions of boundary value problems. The approximations will be asymptotic to the exact solutions as λ tends to zero or k tends to infinity. For definiteness we shall consider the reduced wave equation for a scalar function $u(x)$:

$$\Delta u + k^2 n^2(x) u = \delta(x - x_0), \quad x \text{ in} D. \tag{9.1}$$

The delta function represents a point source of unit strength at x_0, and the equation holds in an exterior domain D. On the boundary ∂D of D some boundary condition must be imposed, and we shall choose

$$u(x) = 0, \qquad x \text{ in } \partial D. \tag{9.2}$$

As $|x| = r$ tends to infinity, we want u to be an outgoing wave. Therefore, we require that u satisfy the radiation condition. If D is a two dimensional domain as we assume, then the radiation condition is

$$\lim_{r\to\infty} r^{1/2}(u_r - iknu) = 0. \tag{9.3}$$

Physically, this problem has the following interpretation. A periodic point source is located at x_0 in the exterior of an opaque object. On the boundary ∂D of the object the field produced by the source must vanish. The frequency of the source is proportional to k, and the amplitude of the field is proportional to u. The field is in a steady state, emanating from the source, being scattered by the object, and radiating outward to infinity.

To construct an asymptotic approximation to $u(x)$ we proceed as follows:

(a) First, we determine all the rays starting at the source point x_0 and passing through the point x. It is convenient to label each such ray with an integer $j = 1, 2, \ldots$.

(b) Next, we associate with the jth ray a field $u_j(x) \sim A_j(x)e^{ikS_j(x)}$. Here $S_j(x)$ is called the phase function and $A_j(x)$ is called the amplitude function. The phase function S_j at x is just the optical length of the jth ray from x_0 to x. Thus,

$$S_j(x) = \int_{x_0}^{x} n\, ds. \tag{9.4}$$

The amplitude $A_j(x)$ is determined from the law of energy conservation in a narrow tube of rays, which can be expressed in the form

$$nA_j^2 d\sigma|_x = nA_j^2 d\sigma|_{x'}. \tag{9.5}$$

Here x and x' are two points on the jth ray while $d\sigma(x)$ and $d\sigma(x')$ are the corresponding cross-sectional areas of a narrow tube of rays containing this ray. From this relation we obtain

$$A_j(x) = \left[\frac{n(x')d\sigma(x')}{n(x)d\sigma(x)}\right]^{1/2} A_j(x'). \tag{9.6}$$

(c) Finally, we add together the fields on all the rays through x to obtain the result

$$u(x) = \sum_j u_j(x) \sim \sum_j A_j(x)e^{ikS_j(x)}. \tag{9.7}$$

Here S_j is given by (9.4) and A_j by (9.6).

To illustrate the use of these formulas, we shall consider the case of a two dimensional homogeneous medium with $n(x) = 1$. Then (9.4) shows that $S_j(x)$ is just the length of the jth ray from x_0 to x. Furthermore all rays, except surface rays, are straight lines. In two dimensions a tube of rays is just a strip bounded by two rays. When these rays are straight, the cross-sectional "area" of the strip at x is $d\sigma(x) = r d\theta$. Here $d\theta$ is the angle between the two bounding rays and r is the distance from their point of intersection to x.

Similarly $d\sigma(x') = r'd\theta$. Thus, (9.6) becomes

$$A_j(x) = (r'/r)^{1/2} A_j(x'). \tag{9.8}$$

From (9.8) we see that $A_j(x)$ decreases like $r^{-1/2}$ as r increases.

In order to use (9.8) it is necessary to determine $A_j(x')$ at some point x' on the ray. It is natural to try to do so at the point where the ray originates. Thus, $A_j(x')$ should be determined at the source on a ray which emanates from the source, at the point of reflection or refraction on a reflected or refracted ray, and at the point of diffraction on a diffracted ray. We shall now explain how to do this.

10. Diffraction coefficients and canonical problems. To find the reflected amplitude $A_{\text{ref}}(x_r)$ at the point of reflection x_r, we require the sum of the incident and reflected fields to satisfy the boundary condition at x_r asymptotically. In general this yields the result

$$A_{\text{ref}}(x_r) = RA_{\text{inc}}(x_r). \tag{10.1}$$

Here $A_{\text{inc}}(x_r)$ is the amplitude on the incident ray at x_r and R is a factor called the reflection coefficient of the boundary or interface at x_r. If the boundary condition is $u = 0$ we find that $R = -1$, while if instead it is the normal derivative of u which must equal zero then $R = +1$. Similar considerations at an interface, in which the transmitted amplitude $A_{\text{trans}}(x_r)$ is involved and suitable interface conditions are imposed, yield (10.1) and

$$A_{\text{trans}}(x_r) = TA_{\text{inc}}(x_r). \tag{10.2}$$

Both R and the transmission coefficient T depend upon the angle of incidence and the two values of n at x_r.

In order to determine $A_{\text{diff}}(x')$, the amplitude on an edge diffracted ray, we note that the diffracted rays all emanate from the edge. For simplicity we shall assume that $n(x) = 1$. Therefore, (9.8) holds with the origin at the edge. But this implies that $A_{\text{diff}}(x')$ becomes infinite as r' tends to zero, in such a way that $(r')^{1/2}A_{\text{diff}}(x')$ remains constant. In analogy with (10.1) and (10.2) we assume that this constant is proportional to the incident amplitude $A_{\text{inc}}(x_d)$ at the point of diffraction x_d:

$$(r')^{1/2} A_{\text{diff}}(x') = DA_{\text{inc}}(x_d). \tag{10.3}$$

We call the factor D introduced in (10.3) a diffraction coefficient. It depends upon the angles of incidence and diffraction, the boundary conditions on the surfaces meeting at the edge, the angle between these surfaces, etc.

To determine D we must solve a canonical problem. This is a simpler problem which has the same local geometry and other local properties as does the actual problem near x_d. By examining the solution of the local problem at a distance of many wavelengths from the edge, we find that the diffracted amplitude does have the form (9.8). Furthermore, $(r')^{1/2}A_{\text{diff}}(x')$ is of the form (10.3) and this relation yields D. This procedure for finding D is equivalent to the formal application of boundary layer analysis, or of the method of matched asymptotic expansions, to the present problem.

From (10.3) we see that D has the dimensions of (length)$^{1/2}$ so we can write

$$D = k^{-1/2}\tilde{D} \tag{10.4}$$

where $\tilde{D}$ is the dimensionless edge diffraction coefficient. This expression shows that an edge diffracted field is weaker than the incident field by the factor $k^{-1/2}$, which tends to zero as k becomes infinite. Upon using (10.4) and (10.3) in (9.8) we obtain for the amplitude on an edge diffracted ray the result

$$A_{\text{diff}}(x) = \tilde{D}A_{\text{inc}}(x_d)/(kr)^{1/2}. \tag{10.5}$$

To find $S_{\text{diff}}(x)$ on a diffracted ray we use (9.4) to obtain

$$S_{\text{diff}}(x) = \int_{x_0}^{x_d} ds + \int_{x_d}^{x} ds = S_{\text{inc}}(x_d) + r. \tag{10.6}$$

Here we have denoted the first integral in (10.6) by $S_{\text{inc}}(x_d)$, the phase of the incident wave at the point of diffraction x_d. Now from (10.5) and (10.6) we have

$$\begin{aligned} u_{\text{diff}}(x) &\sim A_{\text{diff}}(x)e^{ikS_{\text{diff}}(x)} \\ &= \frac{\tilde{D}A_{\text{inc}}(x_d)}{(kr)^{1/2}} e^{ikS_{\text{inc}}(x_d)+ikr} \sim \frac{\tilde{D}e^{ikr}}{(kr)^{1/2}} u_{\text{inc}}(x_d). \end{aligned} \tag{10.7}$$

A completely similar analysis of diffraction by a vertex in three dimensions yields the result

$$A_{\text{diff}}(r) = \tilde{C}A_{\text{inc}}(x_d)/kr. \tag{10.8}$$

Here $\tilde{C}$ is the dimensionless vertex diffraction coefficient. From (10.8) we see that a vertex diffracted field is weaker by the factor k^{-1} than the incident field. It is also weaker than an edge diffracted field, and is in fact as weak as a field doubly diffracted by edges.

For the amplitude on a surface diffracted ray there are corresponding formulas. They involve additional diffraction coefficients as well as decay exponents. However, we shall not present them because we shall not use them.

To find the amplitude on a ray which has been multiply reflected, refracted, edge diffracted, vertex diffracted, or surface diffracted, we just apply the preceding formulas repeatedly. In this way we can find the amplitudes on all the rays through the point x, and thus evaluate the asymptotic approximation (9.7) for $u(x)$. The amplitude is decreased by a factor $k^{-1/2}$ at each edge diffraction, k^{-1} at each vertex diffraction, and by an exponential factor at each surface diffraction. Therefore, to obtain an approximation which is asymptotic to $u(x)$ to a given order in k^{-1}, it suffices to consider only rays which have undergone a limited number of diffractions.

The equation (9.6) yields an infinite value for the amplitude at a point where the cross-sectional area $d\sigma$ of a ray tube vanishes. The locus of such points is called a caustic surface of the ray family, and on a caustic (9.6) is not valid. Instead the asymptotic form of the solution is different on the caustic and in a "boundary layer" near it, from what it is elsewhere. This different asymptotic form is given by multiplying the expression for u by a factor called a caustic correction factor. This factor is also obtained from a canonical problem, or by using boundary layer theory or the method of matched expansions. The result is that the field on a caustic exceeds that off the caustic by a factor which is a positive power of k, typically $k^{1/6}$. Thus, the field on a caustic is much larger than that elsewhere when k is large.

11. Applications. Now we shall apply the preceding construction to several problems. In doing so we shall assume that the medium is uniform and that the source is very far away from the scatterer. Then we can treat the incident wave as a plane wave with negligible error. To show this, we consider the spherical wave $Ae^{ik|x-x_0|}/|x - x_0|$ of amplitude A produced by a source at x_0. We suppose that the source location x_0 recedes to infinite distance in the direction of the unit vector $-\hat{k}$. In order that the wave have a finite limit, we assume that both the phase and modulus of A increase linearly with $|x_0|$, so that $A = |x_0|^{-ik|x_0|}$. Then the spherical wave $e^{ik(|x-x_0|-|x_0|)}|x_0|/|x - x_0|$ has the limit $e^{ik\hat{k}\cdot x}$, which is a plane wave. This justifies the use of a plane wave for a sufficiently distant source.

As the first application, we consider the canonical two dimensional problem of diffraction of a plane wave by a half-plane or semi-infinite thin screen. The asymptotic expansion of Sommerfeld's exact solution of this problem for kr large contains a diffracted wave of the form (10.7). It agrees precisely with (10.7) if $\tilde{D}$ is given by

$$\tilde{D} = -\frac{e^{i\pi/4}}{2(2\pi)^{1/2}}\left[\sec\tfrac{1}{2}(\theta-\alpha) \pm \csc\tfrac{1}{2}(\theta+\alpha)\right]. \qquad (11.1)$$

Here θ and α are respectively the angles between the incident and diffracted rays and the normal to the screen. The upper sign holds when the boundary condition is $u = 0$ on the half-plane and the lower sign when it is $\partial u/\partial n = 0$. This comparison determines $\tilde{D}$ for the edge of a thin screen on which either of these boundary conditions holds.

Now we can use (11.1) and (10.7) to construct asymptotic solutions of other diffraction problems. Let us do so for the two dimensional problem of diffraction of a plane wave by a slit in a thin screen. We suppose that the screen lies on the parts $y \geqslant a$ and $y \leqslant -a$ of the y-axis of a rectangular coordinate system, with its edges at $x = 0, y = \pm a$. Thus the slit of width $2a$ is the interval $-a < y < a$, of the y-axis. Let the incident field be

$$u_{\text{inc}}(x,y) = e^{ik(x\cos\alpha - y\sin\alpha)}.$$

Then one incident ray hits each edge normally and produces diffracted rays as in Figure 3b. This is shown in Figure 5 for the particular case in which $\alpha = 0$ so that the incident rays are also normal to the screen.

One singly diffracted ray from each edge passes through each point P not on the edge. The field on each such ray is of the form (10.7), and the sum of these two fields is the singly diffracted field $u^s(P)$. It is given by

$$u^s(P) \sim -\frac{e^{ik(r_1 - a\sin\alpha) + i\pi/4}}{2(2\pi k r_1)^{1/2}}\left[\sec\tfrac{1}{2}(\theta_1+\alpha) \pm \csc\tfrac{1}{2}(\theta_1-\alpha)\right]$$

$$-\frac{e^{ik(r_2 + a\sin\alpha) + i\pi/4}}{2(2\pi k r_2)^{1/2}} + \left[\sec\tfrac{1}{2}(\theta_2-\alpha) \pm \csc\tfrac{1}{2}(\theta_2+\alpha)\right]. \quad (11.2)$$

Here r_1 and r_2 are respectively the distances from the edges at $y = +a$ and $y = -a$ to P, while the angles θ_1 and θ_2 are determined by the corresponding rays, as shown in Figure 5.

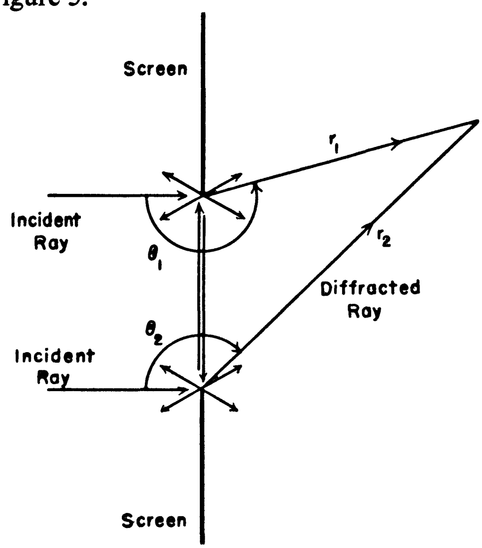


FIGURE 5. The diffracted rays produced by a plane wave normally incident on a slit in a thin screen. The two incident rays which hit the slit edges are shown, with some of the singly diffracted rays they produce. One diffracted ray from each edge is shown crossing the slit and hitting the opposite edge, producing doubly diffracted rays.

From (11.2) we can obtain the far field diffraction pattern of the slit due to single diffraction. To do so we let r, φ be polar coordinates of P and consider points P which are far from the slit, so that $r \gg a$. Then we have $r_1 \sim r - a\sin\varphi$, $r_2 \sim r + a\sin\varphi$, $\theta_1 \sim \pi + \varphi$ and $\theta_2 \sim \pi - \varphi$. By using these relations in (11.2) we can write $u^s(P)$ in the form

$$u^s(r,\varphi) \sim -(k/2\pi r)^{1/2} e^{ikr + i\pi/4} f_s(\varphi). \quad (11.3)$$

Here the far field amplitude $f_s(\varphi)$ due to single diffraction is found to be

$$f_s(\varphi) = i\frac{\sin[ka(\sin\varphi + \sin\alpha)]}{k\sin\frac{1}{2}(\varphi+\alpha)} \pm \frac{\cos[ka(\sin\varphi+\sin\alpha)]}{k\cos\frac{1}{2}(\varphi-\alpha)}. \quad (11.4)$$

We see that the far field diffraction pattern $|f_s(\varphi)|$ is the same for the two different boundary conditions on the screen.

In Figure 6 the solid line shows $|kf_s(\varphi)|$ as a function of φ for normal incidence ($\alpha = 0$), based upon (11.4) with $ka = 8$. For comparison the exact values of the far field diffraction pattern are also shown, as dots, for the boundary condition $u = 0$. They are based upon numerical evaluation of the exact solution of the appropriate boundary value problem for the reduced wave equation. These results are given by S. N. Karp and A. Russek, J. Appl. Phys. 27 (1956), 886. The agreement between the two results is quite good even though the slit is only about 3 wavelengths wide.

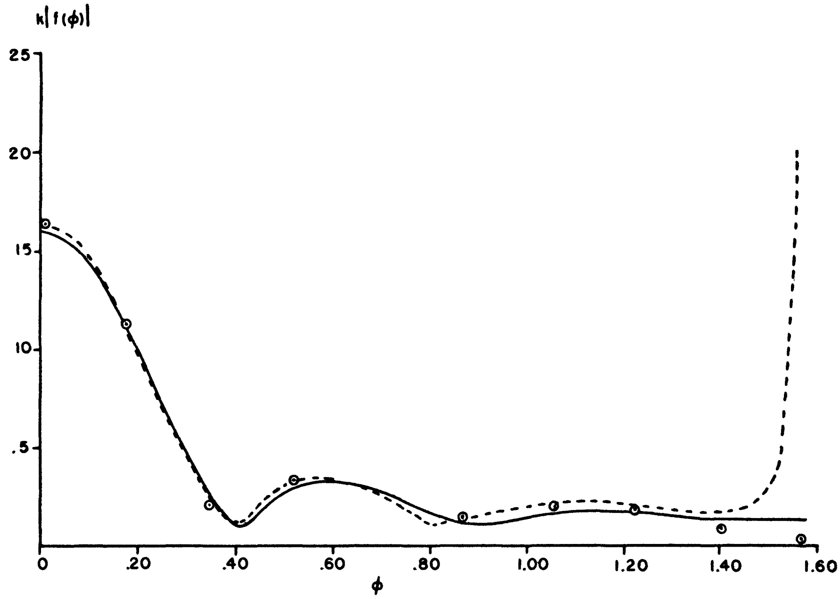


FIGURE 6. The far-field diffraction pattern of a slit of width $2a$ hit normally by a plane wave; $ka = 8$. The solid curve based upon (11.4) results from single diffraction, and applies to a screen on which $u = 0$ or $\partial u/\partial n = 0$. The dashed curve includes the effects of multiple diffraction 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 $u = 0$. The ordinate is $k|f(\varphi)|$ and the abscissa is φ in radians.

The transmission cross-section σ of the slit (per unit length) is a measure of the flux of energy through the slit. According to the cross-section theorem, $\sigma = \operatorname{Im} f(-\alpha)$ where $f(\varphi)$ is the far field amplitude. If we approximate f by the singly diffracted amplitude f_s given by (11.4), we get $\sigma \sim \operatorname{Im} f_s(-\alpha) = 2a \cos \alpha$. This is just the result of geometrical optics. To obtain a more accurate result we shall consider the doubly diffracted rays. They are produced by the two singly diffracted rays which cross the slit and hit the opposite edges.

To find the field u^s_{inc} incident upon the upper edge on the ray singly diffracted from the lower edge, we use the second term in (11.2). We shall choose the upper sign, appropriate to a screen on which $u = 0$. Then we set $\theta_2 = \pi/2$ and $r_2 = 2a$ to obtain

$$u^s_{\text{inc}} \sim -\frac{e^{ika(2+\sin\alpha)+i\pi/4}}{2(2\pi ka)^{1/2}} \sec \tfrac{1}{2}\left(\frac{\pi}{2} - \alpha\right). \tag{11.5}$$

Substitution of this field into (10.7) yields the field on the doubly diffracted rays from the upper edge. A similar calculation yields the doubly diffracted field from the lower edge. The sum of these two doubly diffracted fields is the total doubly diffracted field u^d.

Far from the slit u^d can be written in the form (11.3) with $f_s(\varphi)$ replaced by another function $f_d(\varphi)$. In the forward direction $\varphi = -\alpha$, f_d has the value

$$f_d(-\alpha) = -\frac{1}{k(\pi ka)^{1/2}}\left[\frac{e^{i2ka(1+\sin\alpha)+i\pi/4}}{1+\sin\alpha} + \frac{e^{i2ka(1-\sin\alpha)+i\pi/4}}{1-\sin\alpha}\right]. \quad (11.6)$$

Now we calculate $\sigma \sim \operatorname{Im}[f_s(-\alpha) + f_d(-\alpha)]$ and obtain

$$\sigma = 2a\cos\alpha - \frac{1}{k(\pi ka)^{1/2}}\left[\frac{\cos[2ka(1+\sin\alpha) - \pi/4]}{1+\sin\alpha} + \frac{\cos[2ka(1-\sin\alpha) - \pi/4]}{1-\sin\alpha}\right]. \quad (11.7)$$

For normal incidence, $\alpha = 0$ and (11.7) becomes

$$\frac{\sigma}{2a} = 1 - \frac{\cos(2ka - \pi/4)}{\pi^{1/2}(ka)^{3/2}}. \quad (11.8)$$

In Figure 7 the solid curve shows $\sigma/2a$ as a function of ka, based upon (11.8). The points shown in the figure are obtained from the exact solution of the boundary value problem. They are given by Karp and Russek, loc. cit. The agreement is very good for ka large, as it should be, but it is also surprisingly close even when the wavelength is several times the width of the slit. Even better agreement is obtained when additional multiply diffracted rays are taken into account, as is shown by the dashed curve in the figure.

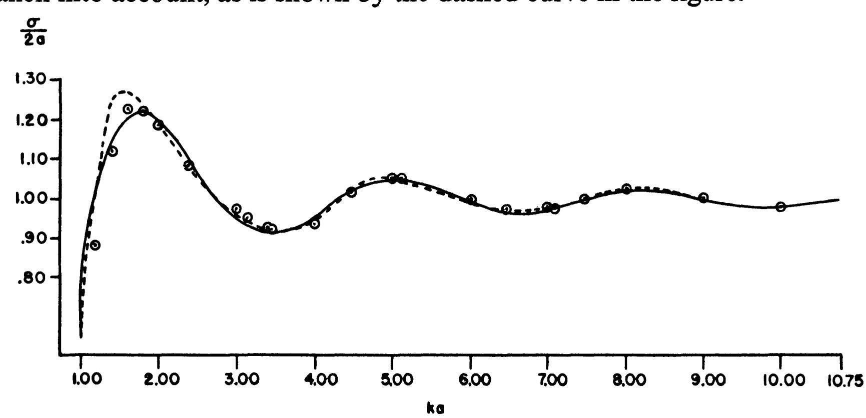


FIGURE 7. The transmission cross-section of a slit of width $2a$ as a function of ka, for normal incidence with $u = 0$ on the screen. The solid curve, based on (11.8) results from single and double diffraction; the dashed curve includes single and all multiple diffraction. The dots are based upon the exact solution of the reduced wave equation with $u = 0$ on the screen. The ordinate is $\sigma/2a$ and the abscissa is ka.

A similar calculation can be made for the three dimensional problem of diffraction of a plane wave by a circular hole of radius a in a thin screen. There are two differences, however. First of all, in calculating the diffracted amplitude from (9.5) we must find the cross-sectional area $d\sigma$ of a tube of diffracted rays in three dimensions. Secondly, we must take account of the fact that the symmetry axis of the circular hole is a caustic of the diffracted rays. When these differences are taken into account, we can calculate the singly and doubly diffracted fields u^s and u^d as before. Then by using the cross-section theorem appropriately, we can find σ. For a wave normally incident on a screen with the boundary condition $u = 0$ we obtain in this way

$$\frac{\sigma}{\pi a^2} = 1 - \frac{2\sin[2ka - \pi/4]}{\pi^{1/2}(ka)^{3/2}}. \tag{11.9}$$

In Figure 8 the solid curve shows $\sigma/\pi a^2$ based upon (11.9) as a function of ka. Results from the exact solution of the boundary value problem are shown for comparison. They are given by C. J. Bouwkamp, Repts. Prog. in Phys. **17** (1954), 35. The agreement is very good for ka large, as we expect it to be. It is also quite good when ka is not very large.

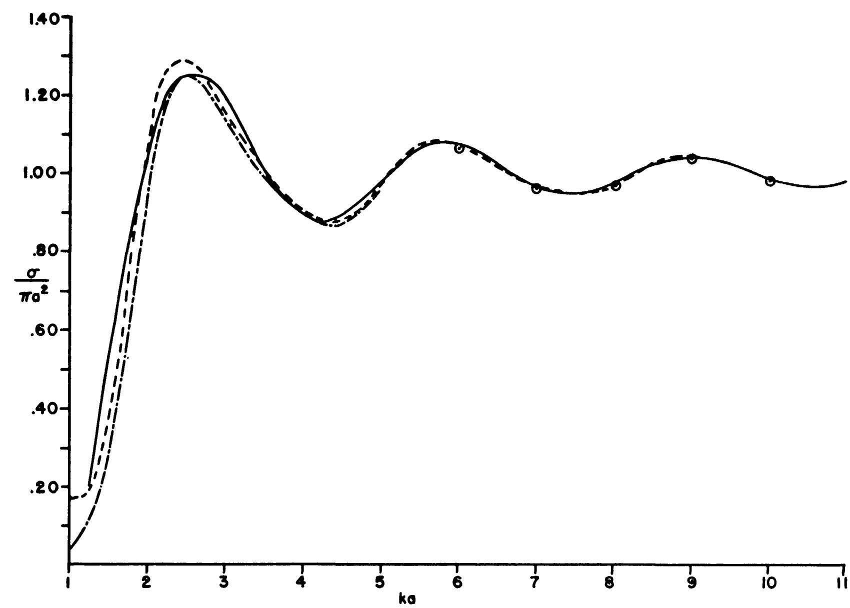


FIGURE 8. The transmission cross-section σ of a circular aperture of radius a in a thin screen on which $u = 0$. The wave is normally incident. The solid curve, based on (11.9), results from single and double diffraction; the dashed curve also includes all multiple diffractions. The dots and the broken curve up to $ka = 5$ are based upon the exact solution of the reduced wave equation. The ordinate is $\sigma/\pi a^2$ and the abscissa ka.

As a final example we shall consider the scattering of a plane electromagnetic wave from the perfectly conducting frustum of a right circular cone shown in Figure 9. In this case u is the electric field, which is a vector field. Consequently the amplitude A_j is a vector. It can be shown that A_j is normal to the jth ray and that its direction remains constant along the ray when the medium is homogeneous. Otherwise A_j undergoes parallel transport with respect to the metric $n(x)\,ds$. Diffracted rays are produced at the two circular edge of the frustum, as in the scalar case. Furthermore the length of A_j is still determined by (9.5). However the diffraction coefficient is a matrix with entries like (11.1). They take account of the finite angle between the surfaces at the edges of the frustum. In fact they are determined from Sommerfeld's solution of diffraction by a wedge on which either $u = 0$ or $\partial u/\partial n = 0$.

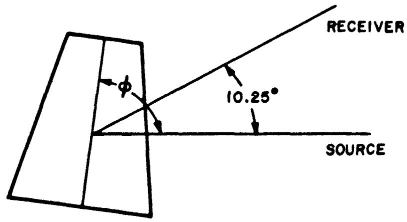


FIGURE 9. Section through the axis of the frustum of a right circular cone. The directions of the radar source and receiver are shown. The angle between these directions, called the bistatic angle, is 10.25°. The frustum can be rotated about an axis normal to the plane of the figure, and its angular position is ϕ.

By using these considerations the scattered field has been calculated, and the results are shown in Figures 10, 11 and 12 along with the corresponding values measured experimentally. These curves are from G. W. Gruver, *Investigation of scattering principles*, Vol. 1, General Dynamics, Fort Worth 1969, pp. 72–74. For this case the exact solution is not known. Instead the comparison with experiment indicates the accuracy of the asymptotic construction.

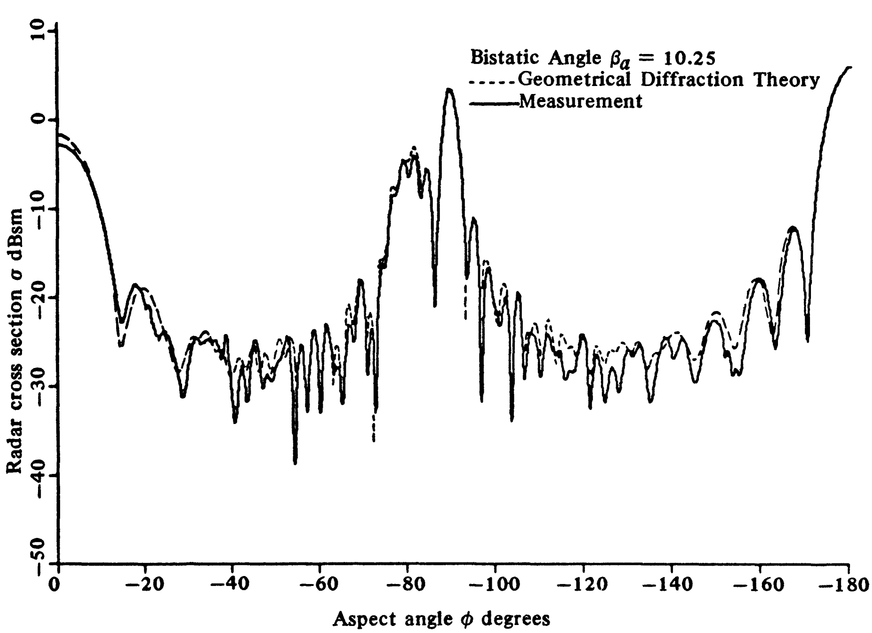


FIGURE 10. The radar cross-section of the frustum in Figure 9 as a function of the angle ϕ. The measured values are represented by the solid curve, and the values computed using the geometrical theory of diffraction are given by the dashed curve. The ordinate is the radar cross-section σ in decibels relative to one square meter and the abscissa is ϕ in degrees. These results are for *VV* polarization, in which the incident field and the measured field both lie in the plane of Figure 9.

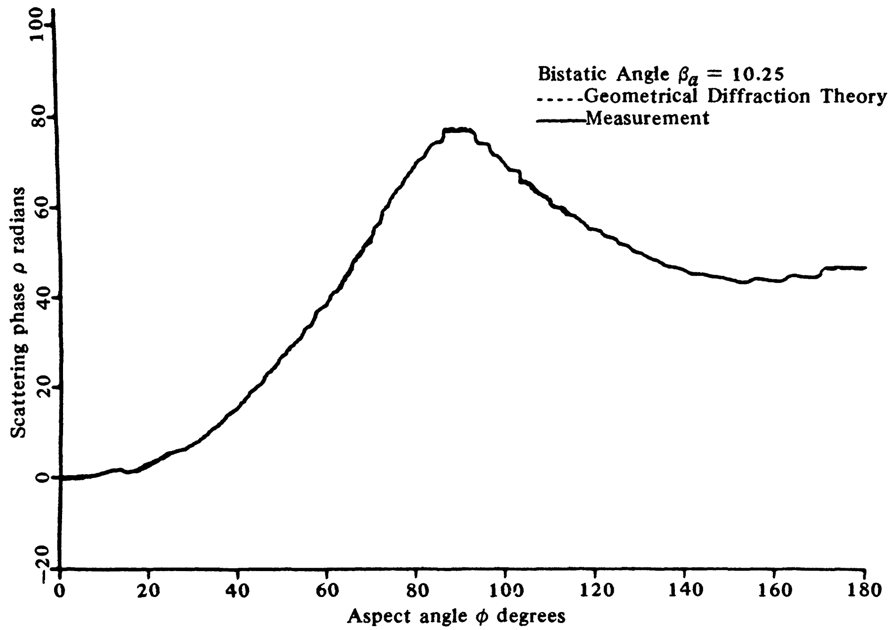


FIGURE 11. The measured and calculated phase of the scattered field for the same case as Figure 10. The ordinate is the scattered phase in radians.

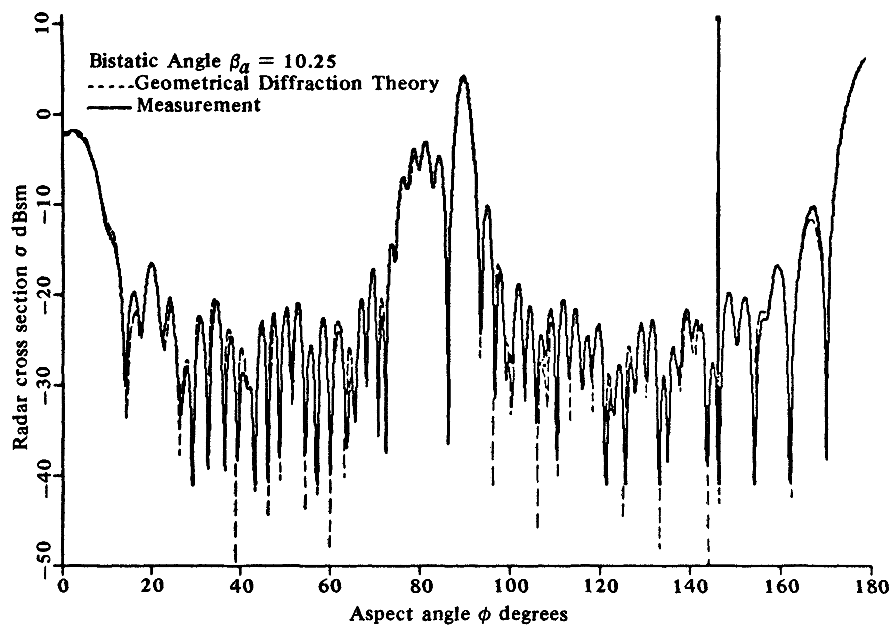


FIGURE 12. Same as Figure 10 for *HH* polarization, in which the incident and measured fields are both normal to the plane of Figure 9.

12. Asymptotics. The Geometrical Theory of Diffraction has been presented as a synthetic procedure for constructing asymptotic approximations to the solutions of a class of boundary value problems. We shall now explain where this procedure comes from, and how to improve the approximation which it yields.

In the nineteenth century many authors tried to solve certain linear ordinary differential equations containing a large parameter k by writing the solutions in the form

$$u(x, k) = e^{ikS(x)} \sum_{n=0}^{\infty} \frac{1}{(ik)^n} A_n(x). \tag{12.1}$$

We shall call the series (12.1) a wave. They substituted the wave (12.1) into some particular equation and equated to zero the coefficient of each power of k. In this way they obtained a recursive system of ordinary differential equations for the successive determination of the phase $S(x)$ and the amplitudes $A_n(x)$. However, they were unable to prove that the resulting series converged.

In 1885 Poincaré proved that in general the series diverges. Instead, he suggested that the series is asymptotic to the solution in the sense that for each N

$$u(x, k) - e^{ikS(x)} \sum_{n=0}^{N} \frac{1}{(ik)^n} A_n(x) = o\left(\frac{1}{k^N}\right) \quad \text{as } k \to \infty. \tag{12.2}$$

In the same year Stieltjes also introduced asymptotic series in another context. Since then the asymptotic character of such series has been proved in great generality by Korn, Birkhoff, Langer, Turrittin, Wasow, Sibuya and many others. The use of such series is often called the WKBJ method after Wentzel, Kramers and Brillouin who used them in quantum mechanics, and Jeffreys who used them in other problems.

The investigations referred to above, and others of the same period, concerned the waves associated with the rays of geometrical optics. None of them considered diffracted waves. However, diffracted waves were present in the asymptotic expansions of the exact and approximate solutions of various problems. For example, edge diffracted waves occurred in Sommerfeld's solutions of diffraction by a half plane and by a wedge. They also appeared in Rubinowicz' approximate solution of diffraction by an aperture. Vertex diffracted waves occurred in diffraction by a semi-infinite circular cone, analyzed by Felsen, by Hansen and Schiff, and by others. Surface diffracted waves were found by Watson in his study of diffraction by a sphere, by Franz in the solution of diffraction by a circular cylinder, and by Friedlander in the analysis of the diffraction of pulses by cylinders with smooth convex cross-sections. Complex waves were present in refraction by a plane interface when the angle of incidence exceeded the critical angle. They also occurred in the plane wave representation of a spherical wave.

The foregoing results and others suggested to me that the asymptotic theory should be extended to include diffracted waves. The method presented in the preceding sections shows how they can be constructed. However it yields only the leading amplitude $A_0(x)$ on each ray. To obtain the full asymptotic expansion, an entire series of the form (12.1) must be used to represent the field on each ray. This requires the introduction of additional diffraction coefficients, and the analysis of further canonical problems to determine them. For diffraction by the edge of a thin screen, some of these coefficients have been found by Karp and Keller (1961) and all of them have been determined by Ahluwalia, Boersma and Lewis (1965).

The preceding considerations indicate how this theory can be applied to any linear partial differential equation or system of such equations, with boundaries of any shape. Many such applications as well as extensions of the theory, have been made by numerous investigators, including R. M. Lewis, D. Ludwig, B. R. Levy, B. D. Seckler, R. N. Buchal, N. Bleistein, J. Cohen, B. Granoff, R. A. Handelsman, B. Matkowsky, F. C. Karal, E. Zauderer, D. S. Ahluwalia, L. Felsen and his co-workers, D. G. Magiros, S. I. Rubinow, B. Morse, Y. M. Chen, G. S. S. Avila, E. B. Hansen, H. Y. Yee, M. C. Shen, R. E. Meyer, E. Resende, V. C. Mow, L. Kaminetzky, W. Streifer, F. Hagin, R. Jarvis, E. Larsen, G. Rosenfeld, R. Voronka, C. Tier, V. Babitch and his co-workers in Leningrad, P. Ufimtsev, etc.

In some cases it has been possible to compare the results of the present theory with the asymptotic expansion of the exact solution, obtained by some other method. In all cases, the two expansions have agreed after computational errors were corrected. This agreement proves the correctness of the theory in those cases, but of course it does not show it in general.

It remains to be proved in general that the series expansion, constructed by the theory described above, is indeed asymptotic to the solution of the corresponding boundary value problem. However, this result has been proved in many cases. Two methods of analysis have been employed. One method deals directly with the constructed expansion, or with some more uniform approximation from which the expansion can be obtained, and attempts to estimate its error. F. Ursell, D. Ludwig, C. S. Morawetz, M. Taylor, and A. Majda have used this method successfully.

The other method, employed by R. K. Luneburg and M. Kline, relates the asymptotic expansion of the solution of an elliptic equation to the singularities of a solution of a corresponding hyperbolic equation. This method requires that the solution of the hyperbolic equation decay sufficiently rapidly as time increases. Great progress in proving that this decay does occur for various problems has been made by C. S. Morawetz, P. D. Lax, R. S. Phillips, C. Bloom, N. D. Kazarinoff, W. Straus, J. Ralston, R. M. Lewis, W. Littman and others.

It has been pointed out that the geometrical theory requires a special treatment of the solution near a caustic surface. D. Ludwig and Y. Kravtsov devised uniform representations of solutions in regions containing the simplest caustics, which avoid this special treatment. J. Arnold, Duistermaat and others have classified all the generic caustics in any number of dimensions, using the results of R. Thom on the singularities of mappings. Then Duistermaat, using L. Hörmander's Fourier integral operators, extended the uniform asymptotic representation of Ludwig and Kravtsov to these caustics. Ludwig has also obtained uniform asymptotic solutions of various other kinds of problems.

13. Conclusion. We have considered the approximate solution of problems of wave propagation, as exemplified by light propagation, and of linear partial differential equations in general. One feature of the method we employed is that it yields asymptotic expansions of solutions, not convergent expansions. These results are very accurate, even far outside the range where they should be good. The reason why this is so is not known, and explaining it is an outstanding problem. It is related to the fact that an asymptotic expansion of a function is an expansion around a singular point, in this case the point $k = \infty$.

A second desirable feature of the method is that it replaces the problem of solving a partial differential equation by that of solving ordinary differential equations. These are the ray equations and the transport equations. Despite the fact that the partial differential equation is linear, the ray equations are nonlinear. Nevertheless, they enable us to construct approximate solutions.

A third feature is the consideration of canonical problems. It provides a general use for special solutions. This enhances the significance of those solutions. It also enables us to identify other special problems which should be solved.

Finally, we note that an older, displaced theory–the ray theory–was used to solve asymptotically the problems of a new theory which had displaced it–the wave theory. It must be true in general that any outmoded theory which is superseded by a new one is asymptotically correct in some limiting case. Otherwise, it would not have been accepted as a satisfactory theory in the first place. Therefore, it should provide a basis for solving asymptotically the problems of the new theory. This methodological principle, which can be illustrated by other cases as well as the present one, may be helpful in guiding us to the solution of other problems.

References

A general reference for §§2, 3 and 4 is

1. M. Born and E. Wolf, *Principles of optics*, Pergamon Press, Oxford, 1975.

Many of the exact and approximation methods mentioned in §§5 and 6, as well as the theory of separable coordinate systems, are described in detail in

2. P. M. Morse and H. Feshbach, *Methods of theoretical physics*, McGraw-Hill, New York, 1953.

See also

3. D. S. Jones, *The theory of electromagnetism*, Pergamon Press, Oxford, 1964.

The existence and uniqueness theory mentioned in §6 is presented in

4. R. Courant and D. Hilbert, *Methods of mathematical physics*, Vol. 2, *Partial Differential Equations*, Interscience, New York, 1962.

The contents of §§7 through 10 were first presented in

5. J. B. Keller, *The geometrical theory of diffraction*, Proc. Sympos. Microwave Optics, Eaton Electronics Research Lab., McGill Univ., Montreal, Canada, June, 1953, 4 pp.

This was republished, with the title altered by the editors, as

6. ______, *The geometric optics theory of diffraction*, The McGill Symposium on Microwave Optics, Vol. 2, B. S. Karasik and F. J. Zucker, eds., Air Force Cambridge Research Center, Bedford, Mass., 1959, pp. 207–210.

The variational characterization of diffracted rays in §8, and examples of the use of complex rays, were elaborated upon in

7. ______, *A geometrical theory of diffraction*, in Calculus of Variations and Its Applications, L. M. Graves, ed., Proc. Sympos. Appl. Math., Vol. 8, Amer. Math. Soc., Providence, R. I., 1958, pp. 27–52. MR **20** #640.

The asymptotic construction described in §9 was presented in the three publications listed above. It was applied to various problems in the following papers:

8. ______, *Diffraction by a convex cylinder*, Inst. Radio Eng.-Trans. Antennas and Propagation, **AP-4** (1956), 312–321. MR **20** #641.

9. ______, *Diffraction by an aperture*, J. Appl. Phys. **28** (1957), 426–444. MR **20** #5833; **21** #571.

10. B. D. Seckler and J. B. Keller, *The geometrical theory of diffraction in inhomogeneous media*, J. Acous. Soc. Amer. **31** (1959), 192–205. MR **20** #5035, #6926.

11. B. R. Levy and J. B. Keller, *Diffraction by a smooth object*, Comm. Pure Appl. Math. **12** (1959), 159–209. MR **21** #1130.

12. J. B. Keller, *Back scattering from a finite cone*, Inst. Radio Eng.-Trans. Antennas and Propagation, **AP-8** (1960), 175–182.

A survey of the geometrical theory of diffraction and some of its applications, up to 1961 is given in the following paper. It contains the analysis of §10 and the applications in §11, as well as several other applications.

13. ______, *Geometrical theory of diffraction*, J. Opt. Soc. Amer. **52** (1962), 116–130. MR **24** #B1115.

The geometrical theory was adapted to eigenvalue problems in

14. ______, *Corrected Bohr-Sommerfeld quantum conditions for nonseparable systems*, Ann. Physics **4** (1958), 180–188. MR **20** #5650.

It was applied to some specific eigenvalue problems in the two following papers:

15. J. B. Keller and S. I. Rubinow, *Asymptotic solution of eigenvalue problems*, Ann. Physics **9** (1960), 24–75.

16. M. C. Shen, R. E. Meyer and J. B. Keller, *Spectra of water waves in channels and around islands*, Phys. Fluids **11** (1968), 2289–2304.

A sample of other applications of the geometrical theory of diffraction is contained in the three following papers:

17. J. B. Keller and F. C. Karal, *Surface wave excitation and propagation*, J. Appl. Phys. **31** (1960), 1039–1046. MR **25** #2753.

18. H. Y. Yee, L. B. Felsen and J. B. Keller, *Ray theory of reflection from the open end of a waveguide*, SIAM J. Appl. Math. **16** (1968), 268–300. MR **39** #641.

19. W. Streifer and J. B. Keller, *Complex rays with an application to Gaussian beams*, J. Opt. Soc. Amer. **61** (1971), 40–43.

The asymptotic method of §12 has been used every often. Some examples of its use are contained in the following:

20. J. B. Keller, R. M. Lewis and B. D. Seckler, *Asymptotic solution of some diffraction problems*, Comm. Pure Appl. Math **9** (1956), 207–265. MR **17**, 41; **18**, 43.

21. F. G. Friedlander, *Sound pulses*, Cambridge Univ. Press, New York, 1958. MR **20** #3703.

22. F. C. Karal and J. B. Keller, *Elastic wave propagation in homogeneous and inhomogeneous media*, J. Acoust. Soc. Amer. **31** (1959), 694–705. MR **21** #2416.

23. S. I. Rubinow and J. B. Keller, *Asymptotic solution of the Dirac equation*, Phys. Rev. (2) **131** (1963), 2789–2796. MR **28** #2792.

The modifications necessary to treat caustics, edges, and shadow boundaries in a nonuniform way are given in

24. R. N. Buchal and J. B. Keller, *Boundary layer problems in diffraction theory*, Comm. Pure Appl. Math. **13** (1960), 85–114. MR **22** #1797.

The uniform asymptotic theory of caustics was devised in the following papers:

25. Ju. A. Kravcov, a) *A modification of the geometrical optics method*, Izv. Vysš. Učebn. Zaved. Radiofizika **7** (1964), 664–673. (Russian) b) *Asymptotic solutions of Maxwell's equations near a caustic*, Izv. Vysš. Učebn. Zaved. Radiofizika **7** (1964), 1049–1056. (Russian)

26. D. Ludwig, *Wave propagation near a smooth caustic*, Bull. Amer. Math. Soc. **71** (1965), 776–779. MR **32** #753.

27. ______, *Uniform asymptotic expansions at a caustic*, Comm. Pure Appl. Math. **19** (1966), 215–250. MR **33** #4446.

The uniform asymptotic theory of edge diffraction was presented in

28. D. S. Ahluwalia, J. Boersma and R. M. Lewis, *Uniform asymptotic theory of diffraction by a plane screen*, SIAM J. Appl. Math. **16** (1968), 783–807. MR **39** #2392.

The extension of the uniform asymptotic theory to more general equations and more general caustics is given in

29. J. J. Duistermaat, *Oscillatory integrals, Lagrange immersions and unfolding of singularities*, Comm. Pure Appl. Math. **27** (1974), 207–281.

Two recent publications bearing on the proof of the asymptotic character of the formal construction are

30. C. S. Morawetz, *Decay for solutions of the exterior problem for the wave equation*, Comm. Pure Appl. Math **28** (1975), 229–264.

31. C. O. Bloom and N. D. Kazarinoff, *Short wave radiation problems in inhomogeneous media: asymptotic solutions*, Springer-Verlag, New York, 1976.

Courant Institute of Mathematical Sciences, New York University, New York, New York 10012