AFCRL-62-34

NEW YORK UNIVERSITY

Courant Institute of Mathematical Sciences

Division of Electromagnetic Research

RESEARCH REPORT No. EM-171

Electromagnetic Theory and Geometrical Optics

MORRIS KLINE

Contract No. AF 19(604)5238

FEBRUARY, 1962

AFCRL-62-34

NEW YORK UNIVERSITY

Courant Institute of Mathematical Sciences
Division of Electromagnetic Research

Research Report No. EM-171

ELECTROMAGNETIC THEORY AND GEOMETRICAL OPTICS

Morris Kline

AF 19(604)5238

Project 5635

Task 56350

February, 1962

Prepared for
Electronics Research Directorate
Air Office Cambridge Research Laboratories
Office of Aerospace Research
United States Air Force
Bedford, Mass.

Requests for additional copies by Agencies of the Department of Defense, their contractors, and other Government agencies should be directed to the:

ARMED SERVICES TECHNICAL INFORMATION AGENCY
DOCUMENTS SERVICE CENTER
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

Department of Defense contractors must be established for ASTIA service or have their 'need-to-know' certified by the cognizant military agency of their project or contract.

All other persons and organizations should apply to the:

U.S. DEPARTMENT OF COMMERCE
OFFICE OF TECHNICAL SERVICES
WASHINGTON 25, D.C.

I. INTRODUCTION. It may seem unnecessary at this late date to discuss the relationship of electromagnetic theory to geometrical optics. The content of both fields is well known and everyone knows also that geometrical optics is the limit for vanishing wave length of electromagnetic theory. Moreover, since Maxwell's theory supersedes the older geometrical optics, presumably, then, geometrical optics could be discarded. The optical industry continues to use it but perhaps that is because it is behind the times.

There are, however, at least three major reasons for pursuing and clarifying the relationship in question. The first is the purely theoretical or academic problem of building a mathematical bridge between the two domains, electromagnetic theory and geometrical optics. The older bases for asserting that geometrical optics is a limiting case of electromagnetic theory are vague and from a mathematical standpoint highly unsatisfactory.

The second major reason for the investigation is a practical one. To solve problems of electromagnetic theory, whether in the range of radio frequencies or visible light frequencies, one should solve Maxwell's equations with the appropriate initial and boundary conditions. However, as is well known, Maxwell's equations can be solved exactly in only a few problems. Hence physicists and engineers, especially those concerned with ultra-high frequency problems, have resorted to the simpler methods of geometrical optics. Although these methods have proved remarkably efficacious in the optical domain, they are intrinsically limited; they do not furnish information about some of the most important phenomena such as diffraction, polarization, and interference, to say nothing about the numerical accuracy of what geometrical optics does yield. Hence the practical question becomes whether the establishment of a better link between Maxwell's theory and geometrical optics will provide more accurate approximate methods of solving electromagnetic problems. Insofar as ultra-high frequency problems are concerned, the answer, based on work of the last ten years, can already be given affirmatively. It is also a fact that optical people are now looking more and more into diffraction effects and one might venture that the practice of optics is on the verge of entering into an electromagnetic treatment of optical problems.

The investigation serves a third purpose. In principal it is concerned with the relationship between a wave theory and a non-periodic phenomenon with the latter in some sense a limiting case of the wave theory as a parameter, the wave length in the case of electromagnetic phenomena, goes to zero. However there are many branches of physics, acoustics, hydrodynamics, magnetohydrodynamics and quantum mechanics, which also treat wave theories. Hence in each case there should be a corresponding "optical" theory or if one exists, as in the case of quantum mechanics, the present theory should shed light on the two complementary domains. We shall in fact see that the electromagnetic investigations to be surveyed here do indeed lead to new creations or new insights into other branches of physics.

2. SOME RELEVANT HISTORY. To appreciate just what the problem of reconciling geometrical optics and electromagnetic theory amounts to we shall examine briefly the historical background.

The science of geometrical optics was founded in the seventeenth century. To the law of relection, known since Euclid's day, René Descartes and Willebrord Snell added the law of refraction; Robert Boyle and Robert Hooke discovered interference; Olaf Römer established the finiteness of the velocity of light; F. M. Grimaldi and Hooke discovered diffraction; Erasmus Bartholinus discovered double refraction in Iceland spar; and Newton discovered dispersion.

Two physical theories of light were created in the seventeenth century. Christiaan Huygens formulated the "wave" theory of light[1] and Newton formulated a theory of propagation of particles[2]. Huygens thought of light as a longitudinal motion of ether and as spreading out at a finite velocity from a point source. The farthermost position reached by the light in space filled out a surface which he called the front of the wave. In homogeneous media this surface is a sphere. To explain further how light propagates, Huygens supposed that when the disturbance reached any point in the ether this point imparted its motion to all neighboring points. Thus if the wave front at time t_1 should be the surface S_1 and if P is a typical point on S_1, the point P communicated its motion to all points in its neighborhood and from P the light spread out in all directions. Its velocity in these various directions depended upon the nature of the medium. Thus in some small interval of time (and in an isotropic medium) the front of the light emanating from a point would be a sphere with P as a center. The same would be true at any other point of the surface S_1, except that the radii of the spheres might differ as the medium differs along S_1. The new position of the front at some time t_2 greater than t_1 is the envelope in the mathematical sense of the family of spheres attached one to each point of S_1. (There is according to this theory also a backward wave. This backward wave troubled scientists until

Kirchhoff showed under his formulation that it does not exist. We shall not pursue this historical point.) To explain reflection and refraction Huygens supposed that the same phenomenon takes place at each point on the reflecting or refracting interface when the front reaches it, except, of course, that no waves penetrate the reflecting surface.

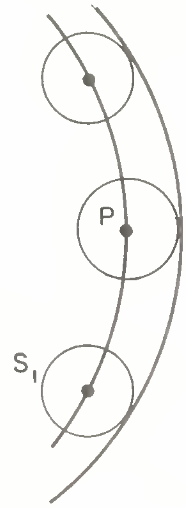

Figure 1

There are many more details to Huygens' theory which explain the phenomena of geometrical optics including double refraction. However, more relevant for us is the fact that Huygens considered light as a series of successive impulses each travelling as already described and he did not explain the relationship of the impulses to each other. Thus the periodicity of light is not contained in Huygens' theory. Also, though the phenomenon of diffraction had already been observed by Hooke and Grimaldi, Huygens apparently did not know it and he did not consider it though his theory could have covered at least a crude theory of diffraction.

The second major theory of light was Newton's. He suggested in opposition to Huygens' "wave" theory, that a source of light emits a stream of particles in all directions in which the light propagates. These particles are distinct from the ether in which the particles move. In homogeneous space these particles travel in straight lines unless deflected by foreign bodies such as reflecting and refracting bodies. Newton did introduce a kind of periodicity, "fits", which he used to explain bright and dark rings appearing in certain phenomena of refraction. However, the nature of the periodicity was vague. His theory was on the whole crude for the variety of phenomena he tried to embrace and he made many ad hoc assumptions. Nevertheless, Newton developed this mechanical theory so thoroughly that its completeness—it included diffraction—and Newton's own great reputation caused others, aside from Euler, to accept it for 100 years. Huygens' work was, on the whole, ignored. Both men, incidentally, obtained some inkling of polarization through reasoning about double refraction in Iceland spar.

Despite the recognition in the seventeenth century of phenomena such as diffraction, a limited theory of light called geometrical optics was erected on the basis of four principles. In homogeneous media light travels in straight lines. The light rays from a source travel out independently of one another. That light rays obey the law of reflection was the third principle, and that they obey the law of refraction for abrupt or continuous changes in the medium was the fourth. (The phenomenon of double refraction in crystals was embraced by supposing that the medium has two indices of refraction which depend upon position and the direction of the propagation.)

All of these laws follow from Fermat's Principle of Least Time. This principle presupposes that any medium is characterized by a function $n(x, y, z)$ called the index of refraction (the absolute index or index to a vacuum). The optical distance between two points P_1 with coordinates (x_1, y_1, z_1) and P_2 with coordinates (x_2, y_2, z_2) over any given path is defined to be the line integral

$$\int_{P_1}^{P_2} n(x, y, z)\, ds$$

taken over that path. Fermat's principle as stated by him and others following him, says that the optical path, the path which light actually takes, between P_1 and P_2, is that curve of all those joining P_1 and P_2 which makes the value of the integral least. This formulation is physically incorrect, as can be shown by examples, and the correct statement is that the first variation of this integral, in the sense of the calculus of variations, must be zero. This principle could be and was applied to the design of numerous optical instruments. It is to be noted that this principle or any other formulation of geometrical optics says nothing about the nature of light.

The mathematical theory of geometrical optics received its definitive formulation in the work of William R. Hamilton during the years 1824 to 1844.[3] Though Hamilton was aware of Fresnel's work, which we shall mention shortly, he was indifferent to the physical interpretation, that is Huygens' or Newtons's, and to a possible extension to include interference. He was concerned to build a deductive, mathematical science of optics. Though his work is described as geometrical optics, he did include doubly refracting media (which are sometimes regarded as outside the pale of strict geometrical optics) and dispersion.

Hamilton's chief idea was a characteristic function, of which he gave several types. The basic one of these expressed the optical length of the ray which joined a point in the object space to a point in the image space as a function of the positions of these two points. The partial derivatives of this function give the direction of the light ray at the point in question. Hamilton also introduced three other types of characteristic functions. He shows that from a knowledge of any one of these, all problems in optics involving, for example, lenses, mirrors, crystals, and propagation in the atmosphere, can be solved. From Hamilton's work the equivalence of Fermat's principle and Huygens' principle is clear.

As we have already observed, geometrical optics cannot be regarded as an adequate theory of light because it does not take into account interference, diffraction, polarization, or even a measure of the intensity of light. In the early part of the nineteenth century new experimental work by Thomas Young, Augustin Fresnel, E. L. Malnus, D. F. J. Arago, J. B. Biot, D. Brewster, W. H. Wollaston and others made it clear that a wave theory of light was needed to account for all these phenomena. Fresnel extended Huygens' theory by adding periodicity in space and time to Huygens' wave fronts. Thereby interference was incorporated and Fresnel used the extended theory to explain diffraction as the mutual interference of the secondary waves emitted by those portions of the original wave front which have not been obstructed by the diffracting obstacle.

Up to this time (1818) thinking on the wave theory of light (and for that matter even the corpuscular theory) had been guided by the analogy with sound. Young in 1817 suggested transverse rather than longitudinal wave motion. Young's suggestion caused Fresnel to think about waves in solids and to suggest that rigidity should give rise to transverse waves. This idea was important for the yet to be developed theory of waves in elastic solids and also for the ether. He sought then to base the theory of light on the dynamical properties of ether.

However, Fresnel's theoretical foundations were incomplete and even inconsistent. He tried to explain the physical nature of light propagating through isotropic and anisotropic media by regarding the ether as a quasi-elastic medium and the light as a displacement of the ether particles. When an ether particle was displaced, the other particles exerted a restoring force proportional to displacement. But the phenomena of interference, the intensity in reflection and refraction, and particularly polarization, led to the conclusion that the vibrations of the ether particles must be transverse, whereas an elastic medium can support transverse and longitudinal waves. Nor could the ether be a rare gas because there only longitudinal waves are transmitted and there is no elastic resistance. Hence Fresnel assumed his ether was infinitely compressible. It was like a gas but with elasticity in place of viscosity. The theory of waves in elastic media was not well developed in Fresnel's time so that his approach was over-simple, and he could not readily eliminate the longitudinal waves which an elastic medium can support.

A number of great mathematical physicists, C. L. Navier, S. D. Poisson, A. L. Cauchy, G. Green, F. Neumann, G. Lamé and J. W. Strutt (Lord Rayleigh) worked on the theory of waves in elastic media and the application of this theory to light[4]. In all this work the ether was an elastic medium which existed in isotropic and anisotropic media. Some of the theories supposed that the ether particles interacted with the particles of ponderable matter through which the light passed. This approach to light was pursued even after Maxwell's time but was never quite satisfactory. One of the principal difficulties was to explain away longitudinal waves. Another was the lack of a consistent explanation of the phenomena of reflection and refraction at the boundaries of isotropic and anisotropic media. A third was that dispersion was not explained.

Of additional efforts preceding Maxwell's work, we shall mention the work of James MacCullagh. MacCullagh in 1839 (published 1848) changed the nature of the elastic solid which represented ether. Instead of a solid which resists compression and distortion, he introduced one whose potential energy depends only on the rotation of the volume elements. Waves in MacCullagh's ether could be only transverse and the vector $\underline{e}$ which represented a wave motion satisfied the equation

$$\mu \Delta \underline{e} = \rho \frac{\partial^2 \underline{e}}{\partial t^2} .$$

Moreover, div $\underline{e} = 0$. MacCullagh did have to introduce independent boundary conditions. (Whittaker, following Heaviside, points out that this $\underline{e}$ amounts to the magnetic field intensity of Maxwell.)

This solid ether of MacCullagh placed difficulties in the way of representing the relationship between ether and ordinary matter (when light travels through matter) and obliged him to postulate a particular force (later called Kirchhoff's force) in order to explain the differing elasticity of the ether on the two sides of a surface which separates diversely refracting media. What is significant about MacCullagh's work is that his differential equations are closely related to Maxwell's though physically the former's theory bore no relation to electromagnetism.

The most satisfactory theory of light which we have today came about not through the study of light per se but through the development of electricity and magnetism by Clerk Maxwell. We shall not pursue here the history of the researches in electricity and magnetism of Gauss, Oersted, Ampère, Faraday, Riemann and others because their contributions are still taught as a basis for Maxwell's electromagnetic theory and so are largely familiar. It is well known that one of Maxwell's great discoveries was the realization that light must be an electromagnetic phenomenon. Maxwell wrote to a friend in January of 1865 "I have a paper afloat, with an electromagnetic theory of light, which 'till I am convinced to the contrary, I hold to be great guns."

Though Maxwell did try unsuccessfully to obtain a mechanical theory of electromagnetic phenomena in terms of pressures and tensions in an elastic medium and after Maxwell, H. Hertz, W. Thomson, C. A. Bjerknes and H. Poincaré tried to improve mechanical models but equally unsuccessfully, the acceptance of Maxwell's theory marked the end of elastic theories of light. The adoption of Maxwell's theory means also the adoption of a purely mathematical view, for the knowledge that light consists of a conjoined electric and magnetic field travelling through space hardly explains the physical nature of light. It merely reduces the number of mysteries in science by compounding one of them.

We might mention that the possibility of linking light and electromagnetism was considered by several predecessors of Maxwell. Euler, Young and Faraday had suggested this possibility on different grounds. Riemann had observed the identity of the velocity of light with the ratio of the electrostatic to the electromagnetic units of charge and so produced an ad hoc theory by extending the electrostatic potential equation

$$\Delta \phi = 4\pi\rho$$

to

$$\Delta\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = 4\pi\rho \ .$$

Thus he had a wave motion which for the proper value of c moved with the velocity of light. However light was still a scalar in this theory nor was there any physical justification for adding $\partial^2\phi/\partial t^2$.

Maxwell's assertion that light is an electromagnetic wave had other arguments to recommend it than the wave equation to which his equations reduce and the fact that the ratio of the electrostatic to the electromagnetic unit of charge is the velocity of light. It is well known that from the first two equations when expressed in rectangular coordinates, for example, and in a non-conducting medium one can obtain for any component of E or H precisely the same mathematical equation which Navier and Poisson had derived for waves in an elastic medium[5] and these latter waves did explain many of the phenomena of light. Moreover, Maxwell's equations possessed a superior feature. Navier, Poisson and other workers in the elastic theory of light had to make the arbitrary assumption that the dilatation (divergence) of the medium is 0 to eliminate longitudinal waves. In Maxwell's equations this condition is automatically present, that is, $\text{div}\,\underline{D} = 0$ and $\text{div}\,\underline{B} = 0$. One could also derive from his equations, as Helmholtz did, the proper boundary conditions at an interface between two media without additional assumptions. Of course Hertz's experimental confirmations, principally the existence of travelling electromagnetic fields, at least showed that radio waves behave like light waves. One must remember, however, that Maxwell's assertion about light was bold and even questionable in his day. The sources of light available then and even up to the present day are not monochromatic and so no fine experimental confirmation could be expected. We are just at the point today, in the development of lasers, of producing coherent monochromatic light.

Though there are unresolved difficulties in Maxwell's theory, chiefly in connection with the interaction of electromagnetic waves with matter (these problems are, of course, being investigated in quantum electrodynamics), we must accept as our best theory that light is an electromagnetic phenomenon subject to Maxwell's equations. Geometrical optics then can be only an approximate representation in several respects. First, wave length considerations do not enter, and so interference is not taken into account. The vector character of the field, that is, polarization, and diffraction, that is, the penetration of the field behind obstacles, are not incorporated. Finally, since wave length considerations do not enter, neither does dispersion.

3. EARLY EFFORTS TO LINK ELECTROMAGNETIC THEORY AND GEOMETRICAL OPTICS. The first significant effort to derive geometrical optics from the electromagnetic theory of light is due to Kirchhoff. Kirchhoff sought a strong mathematical foundation for light and

introduced a modification of Huygen's principle which incorporated the interference in space and time. (The physical interpretation was for him irrelevant.) Since light was represented as a scalar function, in this respect Kirchhoff's representation of light is not directly relevant. Moreover, as is well known, there are difficulties in the use of the Kirchhoff-Huygens principle which he tried to overcome by the assumption of rather arbitrary boundary conditions on the diffracting obstacle and these lead to mathematical inconsistencies.

Nevertheless, in 1882 Kirchhoff did show[6] that when the wave length of the source approaches 0 the wave field given by the Kirchhoff integral approaches the field given by geometrical optics; specifically the diffracted field vanishes and there is sharp transition between the illuminated field and the dark region. That is, the waves behave like straight lines. Hence the idea was generally accepted by the end of the nineteenth century that geometrical optics must be some sort of limit of electromagnetic theory as the wave length goes to 0.

The most widely accepted argument for the connection between electromagnetic theory and geometrical optics is that given by Sommerfeld and Runge who followed a suggestion of P. Debye.[7] In this argument a function u, which may represent some component of $\underline{E}$ or a component of a Hertz vector, is assumed to satisfy the scalar reduced wave equation

$$\Delta u + k^2 u = 0 , \tag{1}$$

wherein $k = \sqrt{\epsilon\mu}\,\omega = 2\pi/\lambda$. Here ϵ and μ may be functions of position and λ is the variable wave length in the inhomogeneous medium. The field is generated by a source, whose frequency is ω and whose wave length in a constant medium ϵ_0, μ_0 is λ_0 so that $k_0 = \sqrt{\epsilon_0\mu_0}\,\omega = 2\pi/\lambda_0$.

Sommerfeld and Runge now make the assumption that

$$u(x,y,z) = A(x,y,z)\,e^{ik_0 S(x,y,z)} , \tag{2}$$

that is, that u is determined by an amplitude function A and a phase function S. The latter, incidentally, is called the eiconal function (because, as we shall see in a moment, it satisfies the eiconal differential equation). While u will vary rapidly as λ_0 approaches 0 or k_0 approaches ∞, it is assumed that A and S do not vary rapidly in x, y and z (relative to the wave length) and that they remain bounded as k_0 approaches ∞. The form of (2) is a generalization of the form of plane waves which exhibit some of the properties of geometrical optics.

By direct differentiation of (2) and substitution in equation (1), one obtains

$$-k_0^2 u\left[\left(\frac{\partial S}{\partial x}\right)^2 + \left(\frac{\partial S}{\partial y}\right)^2 + \left(\frac{\partial S}{\partial z}\right)^2 - \frac{k^2}{k_0^2}\right]$$

$$+ 2ik_0 u[\tfrac{1}{2}\Delta S + \text{grad}\log A\cdot\text{grad}\, S] + e^{ik_0 S}\Delta A = 0 \ .$$

If we now divide through by $k_0^2 u$ and assume that the resulting last term on the left side, namely $\Delta A/k_0^2 A$, remains small as k_0 becomes infinite, then we may satisfy the last equation by requiring that

$$(\text{grad})^2 = n^2 \ , \qquad (3)$$

where $n = k/k_0$, and

$$\text{grad}\log A\cdot\text{grad}\, S + \tfrac{1}{2}\Delta S = 0 \ . \qquad (4)$$

Equation (3) is called the eiconal differential equation and its solutions $S = \text{const.}$ are the wave surfaces or wave fronts of geometrical optics. The second equation can be written in terms of the directional derivative of $\log A$ in the direction of grad S. Since, by (3), $|\text{grad}\, S| = n$, we may write

$$n\frac{\text{grad}\, S}{n}\cdot\text{grad}\log A + \tfrac{1}{2}\Delta S = 0$$

and denoting the directional derivative in the direction of grad S by d/ds, we have

$$n\frac{d(\log A)}{ds} + \tfrac{1}{2}\Delta S = 0 \ . \qquad (5)$$

The direction of grad S is normal to the surface $S = \text{const.}$ and so equation (5) gives us the behavior of $\log A$ along any normal (orthogonal trajectory) to the family of surfaces $S = \text{const.}$ or along a ray.

The fact that equation (3) is derived from the scalar wave equation by letting λ_0 approach 0 and the fact that the equation so obtained is the eiconal equation already known in geometrical optics and from which all of geometrical optics can be derived, provides the argument for concluding that geometrical optics can be derived from Maxwell's equations. Also the fact that the amplitude A travels along the rays is in accord with geometrical optics, though of course A may vary in other directions not revealed by the above derivation.

The Sommerfeld-Runge derivation of geometrical optics is open to many objections. The derivation from the scalar wave equation is not sufficiently general in that not all electromagnetic problems can be reduced to the scalar wave equation. However this criticism has been met in that the same kind of argument has been made for Maxwell's equations. That is, one assumes

$$\begin{aligned} \underline{E}(x,y,z) &= \underline{u}(x,y,z)e^{ik_0 S(x,y,z)} \\ \underline{H}(x,y,z) &= \underline{v}(x,y,z)e^{ikS_0(x,y,z)} \end{aligned} \qquad (6)$$

and one obtains the eiconal equation for S and vector equations for $\underline{u}$ and $\underline{v}$ which are the analogues of (5) above.[8]

Though the Sommerfeld-Runge procedure can be applied to Maxwell's equations as well as the scalar wave equation, it is not a satisfactory derivation of geometrical optics from electromagnetic theory. The assumption (2) represents a very restricted class of fields because it assumes that the function A is independent of k_0. This assumption is fulfilled for plane waves but is not true of the fields encountered even in relatively simple problems of propagation in unbounded media. Hence the argument shows only that a very restricted class of fields gives rise to a geometrical optics field. Secondly, the argument that the A and S determined as solutions of (3) and (4) are limits of the A and S in $u = Ae^{ik_0 S}$ when k_0 is infinite is incomplete. The differential equations (3) and (4) are a limit of the differential equation (1), but this fact must be brought to bear on the solutions. Thirdly, since initial and boundary conditions play no role in the entire derivation the limiting field determined by A and S serves no purpose in representing a geometric optics approximation to some desired field. Finally, the derivation seems to offer no insight into the relationship between wave theory and geometrical optics which might be used to make some gradual transition from one to the other.

Another procedure commonly used to link geometrical optics and Maxwell's theory is to take time harmonic plane wave solutions of Maxwell's equations and to apply the electromagnetic boundary conditions at a plane interface between two homogeneous media. As a consequence one deduces the law of reflection and Snell's law of refraction. Thus the basic laws of geometrical optics are derived. The same procedure is used in homogeneous anisotropic media. As a matter of fact, even the Fresnel formulas for the amplitudes of the reflected and refracted waves are also derivable in this way.

There are several objections to this procedure. Plane waves and plane boundaries are especially simple. There is no indication from such a derivation as to what may happen for curved wave fronts and curved boundaries. The argument is commonly given that the laws of plane waves in homogeneous media suffice for the approximate electromagnetic treatment of such phenomena in which the wave fronts are no longer plane but where the curvature of the wave front can be neglected over domains whose linear dimensions are large compared to the wave length of light. The analogous remark is often made about curved boundaries. But in geometrical optics the laws of reflection and refraction do hold for curved fronts and curved boundaries and even in inhomogeneous media. These facts are not obtained by the argument based on plane waves.

Secondly, in order to use the results obtained from this argument in geometrical optical problems, the practice is to assume that any normal to the wave front is a ray and that each ray behaves at any one point of an interface as though it were independent of all the other rays. But the plane wave argument treats the infinitely extended plane wave and the infinite plane boundary and the argument does not isolate what may happen for any individual ray at a single boundary point. Yet the laws are used thus even at a point on a curved boundary such as the surface of a lens.

Thirdly plane waves have infinite energy and are a highly ideal concept. No physical source sends out plane waves. Finally plane waves have a wave length. Since this fact does not show up in the laws derived, it is ignored.[9]

All one can really say from the study of plane waves is that they obey some of the laws of geometrical optics but they do not suffice to derive geometrical optics from Maxwell's equations.

4. THE RELATIONSHIP OF GEOMETRICAL OPTICS TO ELECTROMAGNETIC THEORY. I should now like to present two new views of geometrical optics from the standpoint of electromagnetic theory. The new viewpoints are valid in both isotropic and anisotropic media, but I shall treat isotropic media. We have Maxwell's equations, which, for simplicity, I shall treat in non-conducting media, namely

$$\begin{aligned} \operatorname{curl} \underline{H} - \frac{\epsilon}{c} \underline{E}_t &= \frac{1}{c} \underline{F}_t \\ \operatorname{curl} \underline{E} + \frac{\mu}{c} \underline{H}_t &= 0 \;. \end{aligned} \tag{7}$$

The term containing $\underline{F}_t$, or strictly the real part of $(1/4\pi)\underline{F}_t$, represents a source current density. In the present discussion its role is irrelevant and one can suppose instead that initial values of $\underline{E}$ and $\underline{H}$, which are functions of x, y, z and t, are specified instead. There may also be boundary conditions.

The first view of geometrical optics is that the geometrical optics field corresponding to any electromagnetic field at any point (x, y, z) of space consists of the singularities of $\underline{E}$ and $\underline{H}$ as functions of time t. By the singularities we mean, of course, the discontinuities of $\underline{E}$ and $\underline{H}$ or of any of their successive time derivatives as functions of t. This definition is, in a sense, too general. If we wish to obtain classical geometrical optics we should restrict ourselves to singularities which are finite discontinuities with respect to time in $\underline{E}$, $\underline{H}$ and their successive time derivatives. There may very well be singularities at which $\underline{E}$ and $\underline{H}$ are continuous, but some time derivative is discontinuous or where the discontinuities of $\underline{E}$ and $\underline{H}$ are finite but those of some time derivative are not.

Before pursuing this concept analytically, let us examine it geometrically. We shall consider two space dimensions. If we suppose that some source located in the plane $t = 0$ begins to act at time $t = 0$, then we know that a field spreads out into space which at a particular time t_0 covers only a bounded region of (x, y, t)-space, the shaded region in Fig. 2. That is, during the time $0 \leq t \leq t_0$ the field will traverse the interior of a cone which lies between $t = 0$ and $t = t_0$. At a point such as P or (x_0, y_0, t_0) the field will be 0 for $t < t_0$ and at $t = t_0$ there will be a jump in the value of $\underline{E}$ and $\underline{H}$ from 0 to a finite value. This finite value of E and $\underline{H}$ is the geometrical optics field at P. Alternatively, the geometrical optics $\underline{E}$ and $\underline{H}$ are the limits approached by $\underline{E}(x_0, y_0, t)$ and $\underline{H}(x_0, y_0, t)$ as t approaches t_0 through values greater than t_0. At times $t > t_0$ the field may continue to be non-zero at the points (x_0, y_0, t) but this field is not a part of the geometrical optics field; it is part of the wave field $\underline{E}(x, y, t)$, $\underline{H}(x, y, t)$ which satisfies Maxwell's equations. Thus the geometrical optics field for all t values is the set of $\underline{E}$ and $\underline{H}$ values which exist only on the surface of the cone.

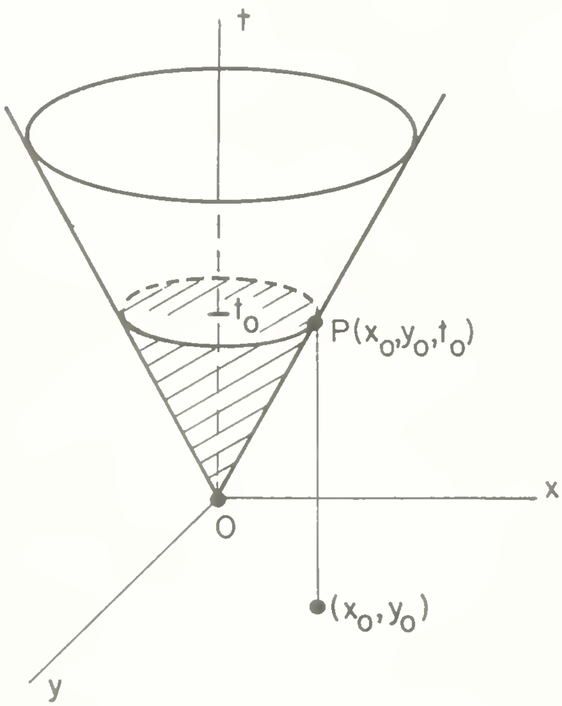

Figure 2

The cone itself is given by some equation $\phi(x, y, t) = 0$ in (x, y, t)-space. One can introduce rays in this space-time picture as the generators of this cone and follow the geometrical optics field along such a ray. (Mathematically these rays have a precise definition as the bicharacteristics of Maxwell's equations.)

There is a second geometrical picture which may be more useful in physical thinking. At each time t the locus of $\phi(x, y, t) = 0$ is a curve. We may plot these curves as a family of curves in (x, y)-space (Fig. 3). These curves are the wave fronts of geometrical optics. Analytically, we suppose that $\phi(x, y, t) = 0$ can be written as $t = \psi(x, y)/c$ and for each value of t there is one curve of this family of wave fronts. The usual rays of geometrical optics are (in isotropic media) the orthogonal trajectories of this family of wave fronts. Insofar as the geometrical optics field is concerned, at each point on a wave front and at the time t_0 given by the equation $\psi/c = t_0$ of this front the values of $\underline{E}$ and $\underline{H}$ change from 0 for $t < t_0$ to some non-zero value. This jump in $\underline{E}$ and $\underline{H}$ is the geometrical optics field at that point. At the same point and at later times $t > t_0$, there may indeed be values of $\underline{E}$ and $\underline{H}$ but these belong to the wave solution of Maxwell's equation and not to the geometrical optics field.

To study the propagation of the geometrical optics field in (x, y)-space one follows it along the rays. Now the wave function $\underline{E}$ is a function of x, y and t. However, for the geometrical optics value of $\underline{E}$, $t = \psi/c$. Hence denoting the geometrical optics $\underline{E}$ by $\underline{E}^*$ we may write

$$\underline{E}^*(x, y) = \underline{E}(x, y, \psi/c) .$$

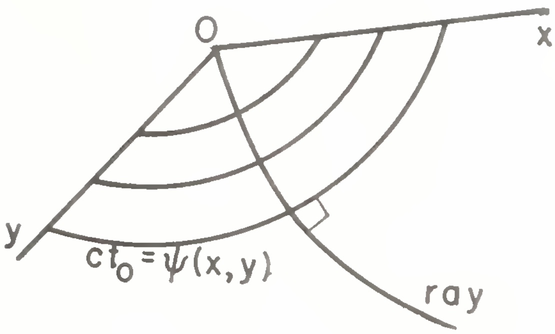

Figure 3

As a function of x and y only, E^* varies continuously.

Now we can show by precise mathematical arguments[10] that

$$\psi_x^2 + \psi_y^2 = n^2(x, y) ,$$

that is, that the wave fronts do indeed satisfy the eiconal equation and that the values of $\underline{E}^*$ and $\underline{H}^*$ along a ray satisfy the vector transport equations

$$2\frac{d\underline{E}^*}{d\tau} + \underline{E}^*\Delta_\mu \psi + \frac{2}{n}(\text{grad } n \cdot \underline{E}^*)\,\text{grad } \psi = 0$$

$$2\frac{d\underline{H}^*}{d\tau} + \underline{H}^*\Delta_\epsilon \psi \quad \frac{2}{n}(\text{grad } n \cdot \underline{H}^*)\,\text{grad } \psi = 0 \qquad (8)$$

where τ is any convenient parameter along the rays $x(\tau)$, $y(\tau)$ and

$$\Delta_\mu \psi = \mu\left(\left(\frac{\psi_x}{\mu}\right)_x + \left(\frac{\psi_y}{\mu}\right)_y\right) = -\text{grad } \psi \cdot \text{grad } \log \mu + \Delta\psi .$$

These transport equations are the vector analogue of (5) above. However the present ones are derived by a precise mathematical argument.

When a front strikes a discontinuity in the medium, then reflected and refracted fronts arise and the discontinuities of $\underline{E}$ and $\underline{H}$, that is $\underline{E}^*$ and $\underline{H}^*$, propagate with the reflected and refracted fronts and satisfy the Fresnel laws at the discontinuity in the medium. The transport equations again describe the propagation of the discontinuities of $\underline{E}$ and $\underline{H}$ along the reflected and refracted rays.

Thus far the approach to geometrical optics is no more than a new mathematical formulation of classical geometrical optics, but indeed one which relates geometrical optics to Maxwell's equations. Classical geometrical optics becomes the behavior of special values of the electromagnetic field. Actually this approach gives more than classical optics, because it gives the vector amplitudes of the geometrical optic fields and the Fresnel laws.

The above-described point of view yields a new insight at once. Let us return to space-time[11]. Consider the field (Fig. 2) at (x_0, y_0, t_0). As t increases beyond t_0 the field $\underline{E}(x_0, y_0, t)$, $\underline{H}(x_0, y_0, t)$ is non-zero. Hence, if $\underline{E}$ and $\underline{H}$ are analytic within the cone, both $\underline{E}$ and $\underline{H}$ should be expressible in power series whose variable is $t-t_0$ which represents the true field for $t > t_0$ [12]. The coefficients of the power series for $\underline{E}$, for example, should be $\underline{E}_t(x_0, y_0, t_0)$, $\underline{E}_{tt}$, ..., where we mean by these derivatives the values assumed by the functions for $t = t_{0+}$ or alternatively the limits approached, for example, by $\underline{E}_t(x_0, y_0, t)$ as t approaches t_0 through values larger than t_0. The values of $\underline{E}$, $\underline{H}$ and their successive time derivatives at $t = t_{0-}$ are 0 because for values of $t < t_0$ the field has not reached (x_0, y_0). The quantities $\underline{E}_t(x_0, y_0, t_0)$, $\underline{E}_{tt}$, ... are then discontinuities of the successive time derivatives of $\underline{E}(x, y, t)$ on the surface $\phi = 0$. Since $t_0 = \psi(x_0, y_0)/c$ each of these discontinuities may be expressed as a function of x_0 and y_0 only.

We may express the thought of the preceding paragraph in terms of the pure space picture (Fig. 3). At any point (x, y) on the wave front $t_0 = \psi(x, y)/c$ and at the time t_0, $\underline{E}$, $\underline{H}$, $\underline{E}_t$, $\underline{H}_t$, ... are discontinuous as functions of t. However, for $t > t_0$ and for points (x, y) on this wave front $\underline{E}$ and $\underline{H}$ are not zero and may be expressed as Taylor's series in powers of $t-t_0$.

Thus under either interpretation we have the expansions

$$\underline{E}(x, y, t) = \underline{E}(x, y, t_0) + \underline{E}_t(x, y, t_0)(t-t_0) + \underline{E}_{tt}(x, y, t_0)\frac{(t-t_0)^2}{2!} + \cdots \quad \text{for } t > t_0 ;$$

$$\underline{E}(x, y, t) = 0 \qquad \text{for } t < t_0 ,$$

and the analogous expansions for $\underline{H}$. Since $t_0 = \psi(x, y,)/c$,

$$\underline{E}(x, y, t) = \underline{E}(x, y, \frac{\psi}{c}) + \underline{E}_t(x, y, \frac{\psi}{c})(t - \frac{\psi}{c}) + \underline{E}_{tt}(x, y, \frac{\psi}{c}) \frac{(t-\psi/c)^2}{2!} + \cdots \quad \text{for } t > \frac{\psi}{c} ;$$

$$\underline{E}(x, y, t) = 0 \qquad \text{for } t < \frac{\psi}{c} . \tag{9}$$

To obtain these power series we must be able to calculate the coefficients. We have already indicated how we can calculate $\underline{E}(x, y, \psi/c) = \underline{E}^*(x, y)$. The method which leads to information about the discontinuities of $\underline{E}$ and $\underline{H}$ themselves, that is, which leads to the transport equations, can be utilized[13] to obtain linear, first order, ordinary differential equations for the discontinuities in $\underline{E}_t$, $\underline{E}_{tt}$, ..., $\underline{H}_t$, $\underline{H}_{tt}$, ... as these propagate along the generators of the cone in the space-time picture or with the wave front or along the rays in the space picture. These differential equations, which we call the higher transport equations, can be solved and so we can obtain the values of these discontinuities at any point (x, y) at the time $t_0 = \psi/c$.

We can then obtain the power series in question and learn something about the time-dependent fields $\underline{E}(x, y, t)$, $\underline{H}(x, y, t)$ in the neighborhood of a wave front, that is, for times t near the time t_0 at which $\underline{E}$ and $\underline{H}$ first become non-zero at (x, y). Stated otherwise, we can obtain the series expansions (9) for $\underline{E}$ and $\underline{H}$ in which the geometrical optics field is the first term.

The second view of geometrical optics to be presented derives from considering time harmonic solutions of Maxwell's equations. The fields we are dealing with then have the form (we now use three space variables)

$$\underline{E}(x, y, z, t) = \underline{u}(x, y, z) e^{-i\omega t}, \quad \underline{H}(x, y, z, t) = \underline{v}(x, y, z) e^{-i\omega t} \tag{10}$$

wherein $\underline{u}$ and $\underline{v}$ are complex vectors. The key result, phrased for simplicity on the assumption that only one family of wave fronts exists, is that

$$\underline{u}(x, y, z) \sim e^{ik\psi(x, y, z)} \left\{ \underline{A}_0(x, y, z) + \frac{\underline{A}_1(x, y, z)}{i\omega} + \frac{\underline{A}_2(x, y, z)}{(i\omega)^2} + \cdots \right\} \tag{11}$$

$$\underline{v}(x, y, z) \sim e^{ik\psi(x, y, z)} \left\{ \underline{B}_0(x, y, z) + \frac{\underline{B}_1(x, y, z)}{i\omega} + \frac{\underline{B}_2(x, y, z)}{(i\omega)^2} + \cdots \right\} \tag{12}$$

wherein the series are asymptotic for large ω and ψ satisfies the eiconal differential equation. The quantity k is ω/c. Thus the functions $\underline{u}$ and $\underline{v}$, which are the amplitudes of the time-harmonic field vectors $\underline{E}$ and $\underline{H}$, may be represented asymptotically by series asymptotic in $1/\omega$ for large ω.

Loosely one can now define the geometrical optics field as the limit for large ω of the field amplitudes $\underline{u}$ and $\underline{v}$. Then the first terms of these two series are the geometrical optics field. The definition as a limit for infinite ω is not quite proper because the first terms of the two series contain the factor $e^{ik\psi}$ and these have no limit as ω becomes infinite. One can however say that the geometrical optics field consists of the first terms of series which are asymptotic for large ω provided we now include in geometrical optics the phase factor $e^{ik\omega}$. This field then is not strictly the classical geometrical optics field but contains an additional and by no means undesirable feature. We also see clearly how this geometrical optics field is related to the full wave solution of Maxwell's equations.

The introduction of this second definition of the geometrical optics field raises the question of whether it is identical, except for the phase factor, with the geometrical optical field previously introduced as the discontinuities of $\underline{E}(x, y, z, t)$ and $\underline{H}(x, y, z, t)$. The answer is that the very derivation of the series (11) and (12) shows that[14]

$$\begin{aligned} \underline{A}_0(x, y, z) &= \underline{E}(x, y, z, \psi/c) = \underline{E}^*(x, y, z) \\ \underline{B}_0(x, y, z) &= \underline{H}(x, y, z, \psi/c) = \underline{H}^*(x, y, z) \\ \underline{A}_1(x, y, z) &= \underline{E}_t(x, y, z, \psi/c) \\ \underline{B}_1(x, y, z) &= \underline{H}_t(x, y, z, \psi/c) \qquad (13) \\ &\cdots\cdots\cdots \end{aligned}$$

Moreover since we know that the above $\underline{E}$, $\underline{H}$, $\underline{E}_t$, $\underline{H}_t \ldots$ satisfy linear, first order differential equations, we know that the same is true for the coefficients of the asymptotic series and so these coefficients can be readily determined. To obtain the geometrical optics field we have but to solve the eiconal equation

$$\psi_x^2 + \psi_y^2 + \psi_z^2 = n^2 ,$$

as must be done in any case, and then solve just the first transport equations, one for $\underline{A}_0$ or $\underline{E}^*$ and the other for $\underline{B}_0$ or $\underline{H}^*$.

The larger mathematical point of interest here is that if one is satisfied to obtain an asymptotic series solution of a time harmonic problem in place of the exact solution, he can replace the solution of Maxwell's partial differential equations by the solution of a series of first order ordinary differential equations. This method must be distinguished from obtaining an exact solution of Maxwell's equations in the form of an integral, say, and then evaluating the integral asymptotically by a method appropriate to the asymptotic evaluation of integrals.

Both views of geometrical optics not only relate this theory directly to Maxwell's equations by precise mathematical connections but accomplish even more. Since one can calculate terms beyond the first ones in the series (11) and (12) this view of the relationship between optics and electromagnetics permits us to improve on geometrical optics approximations to electromagnetic problems. Likewise the Taylor series expansion of the time-dependent $\underline{E}$ and $\underline{H}$ in the neighborhood of $t_0 = \psi(x, y, z)/c$ improves on the geometrical optics field in the direction of the full time-dependent solution. Thus our new views of geometrical optics permit us to make better approximations to wave solutions than geometrical optics itself. We see, incidentally, that we have supplied the mathematical foundation for what Sommerfeld and Runge did.

The theory discussed thus far applies to the direct transmission, reflection and refraction in homogeneous and inhomogeneous isotropic media, and, insofar as geometrical optics as a study of discontinuities is concerned, it has also been carried out for homogeneous and inhomogeneous anisotropic media. Stated otherwise, wherever the rays of classical geometrical optics had been defined, the new theory applies also. For this class of problems one can obtain asymptotic series solutions corresponding to given sources, initial conditions, and boundary conditions.

5. SOME APPLICATIONS OF THE THEORY. The more careful study of the relationship of geometrical optics to electromagnetic theory has stimulated a number of investigations and has thrown new light on older ones within the domain of electromagnetics and outside. We see more clearly that the propagation of discontinuities is the first approximation to aperiodic or time dependent solutions of various equations of mathematical physics and the approximations obtained by letting some parameter approach ∞ are the first terms in asymptotic series developments of time harmonic fields or of solutions of the time free elliptic partial differential equations. I should like to give some indication of the scope of the problems encompassed by the theory presented in article 4.

Since many electromagnetic problems can be treated as scalar problems and since other branches of mathematical physics involve either scalar quantities or different systems of partial differential equations, I should like to point out first that the theory I have sketched for Maxwell's equations has been extended first of all to the general linear second order hyperbolic partial differential equation[15]

$$\sum_{i,j=1}^{n} a^{ij} u_{ij} + \sum_{k=1}^{n} b^{k} u_{k} + cu = f_{tt} \tag{14}$$

wherein u is a function of $(x_1, x_2, \ldots, x_n)$ and the coefficients a^{ij}, b^k, and c are functions of $x_1, x_2, \ldots, x_{n-1}$, and $u_{ij} = \partial^2 u/\partial x^i \partial x^j$. Thus treating x_n as t, one may study the behavior of the discontinuities $[u]$, $[u_t]$, $[u_{tt}]$, ... of u and obtain transport equations for their propagation along what are called the bicharacteristics of (1) or, in the $(x_1, \ldots, x_{n-1})$-space, along the rays. One may also discuss the asymptotic series representation of solutions $u(x_1, \ldots, x_{n-1})e^{-i\omega t}$, that is, of time-harmonic solutions of (1), and all of the relations between the time-dependent solution and the time-independent solution which hold for Maxwell's equations apply here too.

The theory has been further extended[16] to symmetric linear hyperbolic systems of partial differential equations and thus can be applied to more complicated systems of first order partial differential equations than Maxwell's equations.

Insofar as applications to electromagnetic theory are concerned, the applications made in the last ten years have been numerous. A large number of scalar problems involving scattering from the exterior of smooth bodies has been treated by Keller, Seckler and Lewis[17]. Since the method of geometrical optics proper had been available the progress in this work is to obtain improvements over the geometrical optics field by calculating more terms of the asymptotic series solution of steady state problems. Still in the domain of electromagnetic problems I should also like to call attention to the surprising result obtained by Schensted[18]. Schensted calculated the asymptotic series for the vector field diffracted by the exterior of a paraboloid of revolution when a plane wave is incident along the axis (the normal is directed along the axis) and found that the asymptotic series consists only of the first term. In this case, then, the geometrical optics field is also the exact electromagnetic solution.

Our theory has an important bearing on quantum mechanics. In erecting the system of wave mechanics Schrödinger in 1926 gave the following construction[19]. He considered a particle of mass m with momentum p and total energy E in a field of force with potential $V(x, y, z)$. Then Hamilton's partial differential equation for the motion is

$$\frac{\partial W}{\partial t} + H\left(x, y, z, \frac{\partial W}{\partial x}, \frac{\partial W}{\partial y}, \frac{\partial W}{\partial z}\right) = 0$$

where H is the Hamiltonian function for the particle, namely,

$$H = \frac{1}{2m}(p_x^2 + p_y^2 + p_z^2) + V(x, y, z)$$

and W is Hamilton's principal function. Thus the partial differential equation in this case is

$$\frac{\partial W}{\partial t} + \frac{1}{2m}\left\{\left(\frac{\partial W}{\partial x}\right)^2 + \left(\frac{\partial W}{\partial y}\right)^2 + \left(\frac{\partial W}{\partial z}\right)^2\right\} + V(x, y, z) = 0 .$$

In accordance with Hamilton's theory the principal function can be written as

$$W = -Et + S(x, y, z)$$

where S is Hamilton's characteristic function. The equation for S now is

$$\left(\frac{\partial S}{\partial x}\right)^2 + \left(\frac{\partial S}{\partial y}\right)^2 + \left(\frac{\partial S}{\partial z}\right)^2 + 2m(V-E) = 0 .$$

On the basis of heuristic considerations, Schrödinger now introduced a wave function ψ and was led to the time-independent (reduced) Schrödinger equation

$$\nabla^2\psi + \frac{8\pi^2 m}{h^2}(E-V)\psi = 0 \qquad (15)$$

wherein ψ is a function of x, y and z. This derivation of the Schrödinger equation indicated that wave mechanics is in some sense a generalization of classical mechanics in the same vague way that electromagnetic theory appeared in 1926 to be a generalization of geometrical optics. In fact Schrödinger was guided by that analogy and spoke of "working from the Hamiltonian analogy on the lines of undulatory optics."

In 1933[20] Birkhoff suggested that asymptotic series solutions for the function ψ in (15) might be obtained by assuming a series

$$\psi(x, y, z) \sim e^{kS}\left(v_0 + \frac{v_1}{k} + \frac{v_2}{k^2} + \dots\right), \quad k = \frac{2\pi i}{h}, \qquad (16)$$

where S and the v_i are functions of x, y and z. By substitution for ψ in the partial differential equation (15) Birkhoff obtained a first order non-linear partial differential equation for S (which corresponds to our eiconal equation) and showed that the v_n satisfy a system of recursive ordinary differential equations

$$\frac{dv_n}{d\tau} + \Phi v_n = A_{n-1}, \qquad (17)$$

where A_{n-1} is a known linear differential expression in $v_0, v_1, \dots, v_{n-1}$.

The first order partial differential equation (eiconal equation) which Birkhoff obtains is the Hamilton-Jacobi equation from which Schrödinger started. Thus on a purely formal basis Birkhoff showed that classical mechanics is derived from wave mechanics by the introduction of an asymptotic series in practically the same formal way that Sommerfeld and Runge derived the eiconal equation of optics from the scalar wave equation except that Birkhoff assumed a full asymptotic series where Sommerfeld and Runge assumed only the first term. However, the precise mathematical relationship of the Schrödinger equation to the original Hamilton-Jacobi first order partial differential equation remained unclear.

It is now apparent from our theory of asymptotic series solution of partial differential equations that the series (16) adopted purely formally by Birkhoff does indeed provide an asymptotic series solution of the Schrödinger equation (15) and that the corresponding eiconal equation is the Hamilton-Jacobi equation

$$\left(\frac{\partial S}{\partial x}\right)^2 + \left(\frac{\partial S}{\partial y}\right)^2 + \left(\frac{\partial S}{\partial z}\right)^2 = 2m(E-V).$$

Here S is Hamilton's characteristic function. We now know too that this last equation holds precisely in the limit for small h. In other words classical mechanics is indeed the limiting case of quantum mechanics.

But our theory goes farther in the domain of quantum mechanics.

To solve the reduced Schrödinger equation for its eigenfunctions and eigenvalues, physicists used separation of variables and obtained the one-dimensional ordinary differential equation

$$\psi''(x) + \frac{8\pi^2 m}{h^2}(E-V(x))\psi = 0 \ , \qquad -\infty < x < \infty \tag{18}$$

wherein m is a mass, $V(x)$ is the potential energy and E is the total energy (and the eigenvalue parameter) and then (1926) applied the approximation method now known as the WKB method after its innovators Wentzel, Kramers and Brillouin to approximate the eigenvalues. At this time (1926) the precise nature of this approximation remained unclear.

In 1908 Birkhoff had given[21] a theory of asymptotic solution of the n-th order ordinary differential equation

$$\frac{d^n z}{dx^n} + \rho\, a_{n-1}(x,\rho)\frac{d^{n-1} z}{dx^{n-1}} + \cdots + \rho^n a_0(x,\rho) z = 0 \tag{19}$$

and had shown that each solution z_i can be expressed in the asymptotic form

$$z_i(x,\rho) \sim e^{\rho\int_a^x w_i(t)\,dt} \sum_{j=0}^{\infty} z_{ij}(x)\rho^{-j} \tag{20}$$

where the $w_i(t)$ are the solutions of the indicial equation

$$w_n + a_{n-1}(x,0)\,w^{n-1} + \cdots + a_0(x,0) = 0$$

and the z_{ij} can be successively determined by solving a recursive system of rather simple ordinary differential equations. Birkhoff showed in the second of his 1933 papers that his 1908 paper readily covers the one-dimensional Schrödinger equation and the first term of (20) yields the WKB solution of this reduced or time free Schrödinger equation.

Our theory now permits us to say that the first term in the series (16), namely $e^{ks}v_0$, wherein all three variables x, y and z are present, is the direct generalization to partial differential equations of the WKB approximation used in one-dimensional problems.

This latter point may need and warrant elaboration. Let us consider the second order wave equation

$$\Delta u - \frac{n^2}{c^2} u_{tt} = 0 \tag{21}$$

Our theory for the asymptotic series solution of time harmonic solutions of this equation tells us that the time harmonic solution $u = v(x,y,z)e^{-i\omega t}$ can be represented in the form

$$v = e^{ik\psi}\left(v_0 + \frac{v_1}{ik} + \frac{v_2}{(ik)^2} + \cdots\right) \qquad (22)$$

wherein k is ω/c, ψ is a solution of the eiconal equation

$$\psi_x^2 + \psi_y^2 + \psi_z^2 = n^2 \quad , \qquad (23)$$

and the v_i satisfy the recursive system of first order linear ordinary differential equations

$$n \frac{d}{ds} v_n + \frac{1}{2} \Delta\psi\, v_n = -\Delta v_{n-1} \qquad (24)$$

where d/ds is the derivative along the rays in (x, y, z)-space[22].

Now in the case of one space dimension, the equation (21) becomes for $u = v(x)e^{-i\omega t}$

$$v''(x) + \frac{\omega^2}{c^2} n^2 v(x) = 0 \quad . \qquad (25)$$

The eiconal equation is

$$(\psi'(x))^2 = n^2 \qquad (26)$$

and the equation for v_0 becomes (v_{n-1} is 0 for $n = 0$)

$$n \frac{dv_0}{ds} + \frac{1}{2} \psi''(x) v_0 = 0 \quad . \qquad (27)$$

In view of (26)

$$\psi'(x) \frac{dv_0}{ds} + \frac{1}{2}\psi''(x) v_0 = 0 \quad .$$

This equation is readily solvable and gives

$$v_0 = c \frac{1}{\sqrt{\psi'(x)}} \quad .$$

Then by (22) the first term in the asymptotic approximation to v is

$$v \sim C\frac{1}{\sqrt{\psi'(x)}} e^{ik\psi} = C \frac{1}{\sqrt{n}}\, e^{ik\int_0^x n\, dx} \quad .$$

This result for v obtained from the first term of our asymptotic series agrees precisely with the WKB solution of equation (25)[23].

In view of the fact that the first term of our asymptotic series is a generalization to partial differential equations of the WKB method for one-dimensional equations, we can now go a step further in quantum mechanical and related problems. The asymptotic theory of partial differential equations can be used for the three-dimensional Schrödinger equation and other equations when separation of variables is not possible. Thus Keller[24] has derived the half integer quantum numbers for the three-dimensional Schrödinger equation by using the first term of the asymptotic series solution for ψ, that is, by assuming that ψ is represented approximately (for large k or small h) by

$$\psi_0 = \sum_{n=1}^{M} A_n(x, y, z) e^{ikS_n(x, y, z, t)}$$

and the condition that ψ_0 must be single-valued. The summation of terms merely takes care of the fact that there may be many series if S is multiple-valued or in optical terms, if many families of wave fronts pass a given point.

The work described in the preceding paragraph was applied to the Schrödinger equation in unbounded domains. However the same method has been used to find asymptotic values of the large eigenvalues and the corresponding eigenfunctions in bounded domains and indeed for the reduced wave equation[25]. That is, the method is applied to

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

where k is the eigenvalue parameter, the equation is valid in some domain D, and a boundary condition, for example, $\partial u/\partial u = 0$, is imposed on the boundary B of D. The method was also applied to the (reduced Schrödinger equation with a spherically symmetric potential $V(r)$, namely,

$$\Delta^2 u + (-k^2 - V(r))u = 0 \quad ,$$

where $-k^2$ is the eigenvalue parameter. Here the domain B is all of space.

Whereas the application of the theory of article 4 to quantum mechanics utilizes the time-harmonic high frequency point of view other applications recently have made utilized the study of discontinuities. Acoustics had been developed from the wave theory point of view almost from the start of this science. One can however introduce a geometrical acoustics, as Keller[26] and Friedlander[27] have and find that the point of view of discontinuities permits one to study weak shock waves in gases. If one assumes for a fluid motion that the shock waves in the medium are weak and so can ignore the interaction of the shock and the medium behind the shock (the side into which the shock is proceeding) and if one neglects viscosity and heat conduction in the fluid then the shock waves are the discontinuities in the (excess) pressure and the change or discontinuity in pressure at the front is the shock strength. One obtains as in geometrical optics an eiconal equation for the wave or shock front. The rays are orthogonal to the fronts and one derives a transport equation for the variation of the shock strength along a ray. One can also treat the reflection and transmission of the shocks across boundaries as in geometrical optics.

It is also possible to obtain asymptotic series solutions of the linearized acoustic equations for periodic waves of high frequency by using the theory presented earlier for periodic solutions of Maxwell's equations or the general second order scalar equation[28]. Then the theory for weak shocks provides an approximation to the periodic solutions in the same way that geometrical optics is an approximation to wave solutions of Maxwell's equations.

The usefulness of a "geometrical optics" of water wave theory as well as of asymptotic approximation for high frequency periodic water waves has also been favorably considered[29]. For water waves in shallow water the wave amplitude $u(x, y, t)$ satisfies the partial differential equation

$$(ghu_x)_x + (ghu_y)_y = u_{tt}$$

wherein g is the acceleration due to gravity and $h(x, y)$ is the variable depth measured from the equilibrium water surface. In this domain of application the treatment of breakers and surf near a beach can be handled effectively by either the study of the discontinuities of the time dependent equation or by examing the high frequency approximation to periodic waves. We know of course that the first term of either approximation is the same except that the phase factor $e^{ik\psi}$ is present in the latter case[30].

Another class of applications deals with the linearized equations of motion in elastic isotropic media[31]. Here for small amplitude shear and compressional waves one can obtain the propagation of pulses or the asymptotic form of periodic waves of high frequency in both homogeneous and inhomogeneous media. For homogeneous isotropic media the linearized equation of elastic wave motion is

$$\rho \frac{\partial^2 \underline{u}}{\partial t^2} = (\lambda+\mu) \nabla(\nabla \cdot \underline{u}) + \mu \nabla^2 \underline{u} \quad .$$

Here $\underline{u}$ is the displacement vector (in rectangular coordinates), ρ is the density of the medium and λ and μ are Lamé's constants. A more complicated equation holds for inhomogeneous media.

One starts with an asymptotic series solution of the form

$$\underline{u} = \sum_0^\infty \frac{\underline{A}_n}{(i\omega)^n} e^{i\omega(S-t)}$$

where $\underline{A}_n$ and S are functions if x, y and z and ω is the angular frequency of the solutions sought. In this case, one gets two different eiconal equations and two different sets of transport equations, one for compressional and one for transverse waves (because the original differential equation is different from Maxwell's), but the method of obtaining the asymptotic series solutions is that sketched above for Maxwell's equations. The "rays" are the orthogonal trajectories to the solutions of the eiconal equations.

Geometrical optics as the study of discontinuities has application to current problems of magnetohydrodynamics. Here we are definitely in the realm of anisotropic media. For electromagnetic theory proper the geometrical optics of anisotropic media is, as in the case of isotropic media, the transport of the discontinuities of $\underline{E}$ and $\underline{H}$. From this viewpoint we derive first the eiconal equation or the Hamiltonian as it is more commonly called in the case of anisotropic media. In such media the energy of the electromagnetic field does not propagate along the normals to the wave fronts but along distinct curves called rays. The variation of these discontinuities along the rays also satisfy transport equations which prove to be first order ordinary differential equations.

The method of electromagnetics has been applied to plasmas. If one approaches a plasma as a perfect (non-viscous), compressible, infinitely conducting fluid, one applies the equations of fluid dynamics and electromagnetics. For weak shocks the equations may be linearized and one obtains four vector partial differential equations in the velocity vector $\underline{u}$, the magnetic field intensity $\underline{H}$, the density ρ, and the entropy S per unit mass. A discontinuity surface is one across which $\underline{u}$, $\underline{H}$ or ρ is discontinuous. For these equations the study of the propagation of the discontinuities leads to three families of fronts (each with its own speed called Alfvén, slow and fast) in any one normal direction and accordingly three families of rays. The surface of wave normals is accordingly more complicated than the Fresnel surface for crystals. It is then possible to obtain transport equations for each of the discontinuities along each family of rays. The results are extremely useful, for such shock waves can be generated[32].

It is also possible to apply the asymptotic theory to periodic waves of high frequency in plasmas but this has not been carried out as yet.

6. SOME OPEN PROBLEMS. The theory developed for Maxwell's equations permits us to obtain useful approximate solutions for time dependent and for time harmonic problems provided that the corresponding geometrical optics approximation exists, that is, in problems where the wave fronts and rays of classical geometric optics are defined. Physically this limitation means a restriction to propagation, reflection and refraction in homogeneous and inhomogeneous isotropic and anisotropic media. Even in these phenomena, no caustics must be present. In view of the importance of diffraction phenomena and in view of the difficulties encountered in solving diffraction problems it would of course be highly desirable to extend the theory of article 4 to cover

such problems.

The first difficulty one faces in attempting such an extension is that the theory already developed presupposes the existence of geometrical optics; that is, we must be able to obtain the wave fronts as solutions of the eiconal equation and their orthogonal trajectories, the rays. In fact the transport equations describe the behavior of the coefficients of the asymptotic series along the rays. Certainly then when there are no wave fronts and rays, the theory thus far developed has no meaning. Also where the rays form an envelope or come together at a focus, the transport equations break down because the phase function $\psi(x, y, z)$ becomes singular. The first major step in the extension of our theory is to extend geometrical optics itself. This idea has already been tackled by a number of men. It has been developed and systematically handled by J. B. Keller[33] who also suggests the unifying principle that diffracted rays can be obtained from an extension of Fermat's principle. Now Fermat's principle for classical geometrical optics is deducible from Maxwell's equations by the process sketched in article 4 (for the very reason that geometrical optics is deducible.) However, the problem remains as to whether the extended Fermat principle, which encompasses rays and wave fronts not in classical geometrical optics, can be deduced from Maxwell's equations. The deduction already carried out presupposes $\underline{E}(x, y, z, t)$, $\underline{H}(x, y, z, t)$ and their successive time derivatives have finite discontinuities on the wave fronts. This condition limits the wave fronts and rays to those of geometrical optics.

Granted the extension of geometrical optics, the next step in diffraction problems is to derive the form of the asymptotic series solution which is valid in diffraction regions. The theory already available proves that in the case of pure propagation, reflection and refraction the form of the asymptotic series is that of a power series in $1/\omega$ and that the series is truly asymptotic to the time harmonic solutions of Maxwell's equations. The corresponding step is missing for series valid in diffraction regions[34]. At the present writing all that we have been able to do is to assume a form recommended by the asymptotic expansion of solutions obtained in an entirely different manner. Thus the problem of diffraction by a circular cylinder can be solved and its solution expanded asymptotically. The form of this asymptotic series or some generalization of it has been used to solve problems involving other shapes.

The third step would be to learn how to determine the coefficients of the asymptotic series. If it should prove to be the case that these coefficients also satisfy transport equations, then the initial values of the solutions of these transport equations would also have to be determined.

In view of the applicability of the theory already developed for electromagnetics to many other branches of physics, the problems just sketched merit attention. Though some progress has been made beyond what was described in the earlier parts of this paper, the accomplishments are not broad enough to warrant attention in this survey.

FOOTNOTES

1. Huygens, C., Traité de la Luminere (1690). An English translation is available from the University of Chicago Press, Chicago, 1945.

2. Newton, I., Optiks (1704, 1706). An English edition is available from Dover Publications, Inc., N.Y., 1952.

3. Synge, J. L., and Conway, W., The Mathematical Papers of Wm. R. Hamilton, I, Cambridge University Press, London, 1931.

4. A full account of the very long series of efforts to develop an elastic theory of light is given by Whittaker, E.T., History of the Theories of Aether and Electricity, V. I, Rev. Ed., Thomas Nelson and Sons, Ltd., London, 1951.

5. König, W., Electromagnetische Lichtheorie, Handbuch der Physik, Old Ed., XX, p. 147, J. Springer, Berlin, 1928.

6. Kirchhoff's proof is in Ann. der Physik (2), 18, 1883, p. 663, and in his Vorlesungen über Mathematische Physik, Teubner, Leipzig, 1891, V. 2, p. 35. An account of it is given by König, W., Electromagnetische Lichttheorie, Handbuch der Physik, Old. Ed., XX, p. 167 ff., J. Springer, Berlin, 1928. An alternative proof which utilizes the transformation of the Kirchhoff double integral into the incident field plus a line integral is given by Baker, B., and Copson, E.T., The Mathematical Theory of Huygens' Principle, Oxford U. Press, London, 1939, p. 79, and Rubinowicz, A., Die Beugungswelle in der Kirchhoffschen Theorie der Beugung, Warsaw, 1957, p. 166 ff.

7. Sommerfeld, A., and Runge, J., Anwendung der Vektorrechnung auf die Grundlagen der geometrischen Optik, Ann. der Phys. 35, 1911, pp. 277-298. Also in Sommerfeld, A., Optik, 2nd ed., Akademische Verlagsgesellschaft, Leipzig, 1959, p. 187 ff.

8. For isotropic media the derivation is carried out in Born, Max, and Wolf, E., Principles of Optics, Pergamon Press, London, 1959, p. 109. References are given there to original papers and to papers in which the analogous procedure can be employed in anisotropic media.

9. Actually Clemens Schaefer in his Einführung in die Theoretische Physik, III, 1, p. 386 ff., W. De Gruyter, Berlin, 1950, avoids the frequency dependence at least in non-dispersive homogeneous media by using functions such as $F(t - \frac{x \cos \alpha + y \cos \beta + z \cos \gamma}{v})$.

10. This result is due to R.K. Luneberg and can be found in Kline, M., An Asymptotic Solution of Maxwell's Equations, Comm. on Pure

and Appl. Math., 4, 1951, 225-262. Also in Theory of Electromagnetic Waves, A Symposium, Interscience Pub., Inc. N.Y., 1951.

11. We continue to use two space variables in order to illustrate geometrically. However all statements apply when three space variables are present.

12. The series exist and converge for values of $t < t_0$ but do not represent the field.

13. Kline, M., loc. cit.

14. Strictly the identification of the coefficients of the asymptotic series with the discontinuities of the time dependent field presupposes that the same source, say $\underline{g}(x, y, z)$ creates both fields but that the time behavior of the source is $e^{-i\omega t}$ in one case and is $\eta(t)$, the Heaviside unit function, in the other. This fact can be ignored in some applications and is helpful in others.

The existence of the asymptotic series (11) and (12) and the relations (13) are due to R.K. Luneberg. An exposition can be found in Kline, M., loc. cit., 1951.

15. Kline, M., Asymptotic Solution of Linear Hyperbolic Partial Differential Equations, Jour. of Rat. Mech. & Anal., 3, 1954, 315-342.

16. Lewis, Robert M., Asymptotic Expansion of Steady-State Solutions of Symmetric Hyperbolic Linear Differential Equations, Jour. of Math. & Mech., 7, 1958, 593-628.

17. Keller, J.B., Lewis, R.M., and Seckler, B.D., Asymptotic Solution of Some Diffraction Problems, Comm. on Pure and Appl. Math., 9, 1956, 207.

18. Schensted, Craig E., Electromagnetic and Acoustic Scattering by a Semi-Infinite Body of Revolution, Jour. Appl. Phys. 26, 1955, 306-8.

19. Whittaker, E.T., History of the Theories of Aether and Electricity, II, p. 270, Thomas Nelson and Sons, Ltd., London, 1953.

20. Birkhoff, George D., Some remarks concerning Schrödinger's Wave Equation, Proc. Nat'l. Acad. of Sciences, 19, 1933, 339-344, p. 475; Quantum Mechanics and Asymptotic Series, Bull. A.M.S., 39, 1933, 681-700; The Foundations of Quantum Mechanics, Comptes Rendus du Congrès Internat. des Math., Oslo, 1936, 207-225. All of these papers are in Birkhoff's Collected Papers, V. II, American Math. Soc., Providence, 1950.

21. Birkhoff, George D., On the Asymptotic Character of the Solutions of Certain Differential Equations Containing a Parameter, Trans. Amer. Math. Soc., 9, 1908, 219-231.

22. These equations will be found in Kline, M., loc. cit., 1954.

23. For a good description of the WKB method, see Kamke, E., Differential-Gleichungen, Lösungsmethoden und Lösungen, 3rd. ed., Chelsea Pub. Co., N.Y., 1948, p. 138. See also p. 276.

24. Keller, Joseph B., Corrected Bohr-Sommerfeld Quantum Conditions for Nonseparable Systems, Annals of Physics, 4, 1958, 180-188.

25. Keller, J.B., and Rubinow, S.I., Asymptotic Solution of Eigenvalue Problems, Annals of Physics, 9, 1960, pp. 24-75, and 10, 1960, 303-305.

26. Keller, J.B., Geometrical Acoustics I. The Theory of Weak Shock Waves, J. Appl. Phys., 25, 1954, 938-947. Also Friedrichs, K., and Keller, J.B., Geometrical Acoustics II, J. Appl. Phys., 26, 1955, 961-966. Whitham, G.B., On the Propagation of Weak Shock Waves, J. Fluid Mech., 1, 1956, 290-318.

27. Friedlander, F.G., Sound Pulses, Cambridge Univ. Press, London, 1958.

28. See Friedrichs and Keller, loc. cit.

29. Lowell, Sherman C., The Propagation of Waves in Shallow Water, Comm. on Pure and Appl. Math., 2, 1949, pp. 275-291.

30. To handle the time periodic case, Lowell reduced to an ordinary differential equation and applied the WKB method. We now know that the approximate solution obtained by this method is the first term of our asymptotic series.

31. Karal, Frank C., Jr., and Keller, Joseph B., Elastic Wave Propagation in Homogeneous and Inhomogeneous Media, J. Acous. Soc. Amer., 31, 1959, 694-705.

32. Bazer, J., and Fleischman, O., Propagation of Weak Hydromagnetic Discontinuities, Physics of Fluids, 2, 1959, 366-378.

33. Keller's work and the best single discussion of diffracted rays will be found in Keller, J.B., A Geometrical Theory of Diffraction. This paper is in Graves, L.M., ed., Calculus of Variations and its Applications, Proc. of Symposia in Appl. Math., VIII, McGraw-Hill Book Co., N.Y., 1958, pp. 27-52. This paper also contains references to other work on diffracted rays and gives applications.

34. In a number of applications already made of the asymptotic series valid for propagation, reflection and refraction, the authors have assumed the existence of a series $\Sigma_{n=0}^{\infty} v_n(x, y, z)/k^n$ or some more general form and have substituted in Maxwell's equations or the reduced scalar wave equation. This procedure is justified only as a convenience in papers which seek to avoid matters of theory and wish

to get on with applications. It is however no more than a heuristic device and logically is inadequate.

The research reported in this paper has been sponsored by the Electronics Research Directorate of the Air Force Cambridge Research Laboratories, Office of Aerospace Research, under Contract No. AF 19(604)5238, (USAF), Bedford, Mass. Reproduction in whole or in part is permitted for any purpose of the United States Government.